Etudes structurales des assemblages macromoléculaires : Du cytosquelette d'actine aux virus à ARN

Francois Ferron

Résumé des Titres et Travaux Scientifiques

pour l'obtention de

l'Habilitation à Diriger des Recherches

Etudes structurales des assemblages macromoléculaires :

Du cytosquelette d'actine aux virus à ARN.

Jury :

- Dr. Delphine Muriaux, rapporteur
- Dr. Carlo Petosa, rapporteur
- Pr. Patrice Gouet, rapporteur
- Pr. James Sturgis, examinateur
- Dr. Bruno Canard, tuteur

Dr. François-Patrice Ferron

Chargé de Recherches CNRS

Laboratoire CNRS - UMR 7257 (AFMB)
« Réplicases Virales : Structures, Mécanismes & Drug Design »

Table des matières

REMERCIEMENTS

-Je tiens à remercier vivement les membres du jury qui ont accepté d'évaluer ce travail.

-Je tiens à remercier les différentes sources de financements indépendantes et publiques (Nationales, Communautaires ou Internationales) qui ont permis de financer ces travaux.

-Je remercie Yves Bourne, directeur de l'AFMB, de m'accueillir au sein du laboratoire AFMB.

-Je remercie mon mentor Bruno Canard d'avoir toujours cru en moi en me laissant la liberté dans mes projets, d'avoir facilité mon retour sur le vieux continent, et de m'avoir accueilli dans son équipe depuis mon recrutement au CNRS. Merci pour tes nombreux conseils et amitié.

-Je remercie Roberto Dominguez de m'avoir accueilli pour mes séjours post-doctoraux Américains. Je chéris ses précieux conseils et son enseignement en biologie structurale, ainsi que de m'avoir montré qu'avec volonté et acharnement on peut tout réussir.

-Je remercie Julien Lescar de m'avoir accueilli dans son équipe, de m'avoir fait confiance et de m'avoir laissé toujours toute la liberté dans mes projets. Merci pour tous les échanges scientifiques et tes conseils.

-Je remercie Joelle Boretto pour son aide constante, sa patience et sa gentillesse. J'espère continuer à travailler avec toi encore longtemps.

-Je remercie Etienne Decroly pour son soutien constant, son énergie et ses discussions scientifiques.

-Je remercie Isabelle Imbert et Barbara Selisko toujours là avec qui j'aime partager et discuter la science et qui me font l'honneur de leur amitié.

-Je remercie Nicolas Papageorgiou, pour ses discussions scientifiques (et philosophiques), sa curiosité et passion pour les challenges dits « impossibles », sa pédagogie et sa disponibilité.

-Je remercie l'ensemble des collaborateurs (trop nombreux pour être tous cités) passés et présents qui ont contribué et continuent de contribuer à ces travaux.

-Je remercie les membres du laboratoire qui font que ce travail se déroule dans les meilleures conditions possibles, merci pour votre aide, pour tous les bons moments partagés.

-Merci aux relecteurs critiques de ce manuscrit.

-Enfin je remercie ma famille qui a toujours été là pour me soutenir, et en particulier Martha qui toujours me soutient inconditionnellement, merci de croire en moi, toi qui me donne tant d'énergie et de joies.

∴

AVANT-PROPOS

La science c'est un sérieux divertissement nécessitant du temps, du travail, et de la rigueur. Elle procure beaucoup de plaisir en cela que la découverte scientifique c'est d'abord des rencontres, due à un enchainement de circonstances plus ou moins planifiées, et même parfois, au hasard... . Etudiant en biologie cherchant sa voie, je suis entré par hasard dans un laboratoire de biologie structurale un après midi de Juin à Toulouse. J'y ai découvert deux choses : une communauté de chercheurs qui s'attachaient à voir plus loin et au cœur même de la matière afin de donner du sens aux expériences des biologistes; et surtout la fascinante beauté de la complexité des assemblages macromoléculaires. Un monde atomique ordonné formant des « édifices » parfois dynamiques, mais toujours surprenant. Fasciné par l'organisation des atomes dans l'espace et le temps, je me consacre depuis lors à travailler à la compréhension de ces architectures. Un peu plus tard une autre rencontre me fit découvrir la virologie et son incroyable diversité stratégique, qui permet aux virus de survivre et de passer de génération en génération leur matériel génétique. Les implications structurales n'étaient alors pas encore clairement perçues. L'utilisation des diverses approches expérimentales (de la biochimie à la biophysique) m'appris qu'il faut de l'audace, de la détermination et de la patience dans l'entreprise, mais surtout de l'humilité dans l'interprétation. Si dans la cellule il y a prévalence de la stabilité des systèmes ; il y a au contraire dans le monde viral une variété des assemblages qui reflètent des stratégies d'adaptation diverses. Ces descriptions structurales ne cessent de remettre en cause notre perception et notre compréhension des mécanismes du vivant, nul doute qu'il reste encore beaucoup à décrire.

Ce mémoire reprend les travaux auxquels j'ai participé depuis la fin de ma thèse en 2005.

CURRICULUM VITAE

François-Patrice Ferron

Chargé de recherche CNRS 2^{ieme} classe.

Né le 1er Octobre 1976, marié.

Nationalité : Française

Laboratoire :

Laboratoire CNRS - UMR 7257 (AFMB)
Groupe: « Réplication Virale : Structure, Méchanisme & Drug Design »
AFMB - POLYTECH Case 925
163, Av de Luminy
13288 Marseille Cedex 09
France
Tel :+33-491 82 86 28
Courriel :francois.ferron@afmb.univ-mrs.fr

A. *Formation*

- Doctorat de Bio-informatique, biochimie et biologie structurale. Université de la Méditerranée, Aix-Marseille II, 04 Février 2005.
- DEA de Biochimie, Biologie Structurale et Génomique. Université de Provence, Aix-Marseille I. 2001
- Maîtrise de Biochimie Structurale, Université Paul Sabatier, Toulouse III. 2000

B. *Parcours Scientifique*

- Octobre 2010 à ce jour : Chargé de Recherche CNRS (CR2). « Caractérisation structurale des complexes de réplication d'*Arenaviridae* et de *Bunyaviridae* : polymérases et protéines ou molécules associées ; et conception de molécules antivirales. ». UMR 7257, Laboratoire AFMB, Réplication virale, mécanismes, structures et drug-design. Marseille, France Dr. B. Canard.
- Septembre 2008 - Septembre 2010 : 3ème stage post-doctoral (Senior Research Fellow) « Caractérisation structurale de la nucleoprotéine du virus de la Vallée du Rift. ». ATIP-CNRS / PICS Singapore Nanyang Technological University, Singapore. Assist. Prof. : J. Lescar.
- Aout 2006 – Aout 2008 : 2ème stage post-doctoral (Research Fellow) « Etudes de la régulation de la croissance du filament d'actine » University of Pennsylvania, Philadelphia, PA USA. Assist. Prof. : R. Dominguez.
- Mars 2005 – Aout 2006 : 1er stage post-doctoral (Post-Doctoral Fellow) « Etudes des protéines se liant à l'actine ». Boston Biomedical Research Institute, Watertown, MA USA. P. I. : R. Dominguez.

C. *Collaborations*

- En cours d'établissement une coopération avec l'Université d'Addis Abbeba

(Ethiopie), avec l'équipe de **Mitiku Filimon,** pour formation d'étudiants de Master en biochimie, biochimie structurale et virologie.

- En cours d'établissement une coopération avec l'Institut de Recherche Biomédicales des Armées (IRBA), avec les Dr **Sébastien Emonet** et **Isabelle Leparc-Gofard**.
- Collaboration avec **Sylvain Baize** de l'Institut Pasteur de Lyon sur la caracterisation des Nucléoprtoteines d'Arenavirus. (Depuis 2014)
- Dans le cadre de l'ANR AREN_BUNYA collaboration avec **Stephen Cuzack** de l'EMBL à Grenoble.(Depuis 2011)
- Dans le cadre de mon post-doctorat comme chercheur associé, j'ai mis en place une collaboration en Microscopie Electronique avec l'équipe de **Thomas Walz,** Harvard Medical School, Dept. Of Cell Biology. (2009-2010). La collaboration à donné lieu à une publication dans le journal *PloS Pathogens.*
- Dans le cadre d'un PICS, je collabore avec **Julien Lescar,** Nanyang Technological University, Singapore, sur la caractérisation d'inhibiteurs contre les Nucléoprotéines de *Phlebovirus*. (Depuis 2011)

D. Travaux d'expertise

- Relecteur pour les journaux : Plos One, Biochemistry, Embo J., J. Virol, Structure, Antiviral Research, NAR, RNA.
- Relecteur pour la campagne d'évaluation de projet scientifique ANR 2013.
- Co-coordinateur de l'ANR ARENA-BUNYA : « Structures, mécanismes, et conception d’inhibiteurs des protéines L des Bunyavirus et Arenavirus », et depuis Octobre 2013 coordinateur.
- Responsable de la TASK structure de l’ANR Blanc SARS-RNA-SPA: «Activités de synthèse et de modification des ARNs du SARS-Coronavirus ».
- Gestions des missions synchrotrons (ESRF/SOLEIL) pour l'équipe de virologie.
- Membre de l'Association française de Cristallographie (depuis 2012).
- Organisation et gestion du parc informatique pour l'équipe.

E. Encadrements et enseignement

- Enseignement d'un cours de Master1 de bio-informatique (10h) auprès de l'Université de Toulon (2015)
- Encadrement d'un étudiant ingénieur (4^{ieme}) de l'école Polytech. (2015)
- Enseignement d'un cours de Master2 Drug design et produits naturels (20h) auprès de l'Université Science et Technique de Hanoï. (Vietnam) (2014-2015)
- Supervision d'un étudiant de doctorat Chilien de 'Universidad Andrés Bello' en échange (3 mois) (2014)
- Encadrement d'une étudiante en Doctorat, (2013-2016) inscrite à l'EDSVS AMU

et financé par une bourse Infectiopole Sud.

- Encadrement d'une étudiante de Master2 Recherche de Aix-Marseille Université. (AMU) (2014)
- Encadrement d'une étudiante de Master2 Recherche de l'Université de Montpellier. (2013)
- Responsable de la formation des Ingénieurs, doctorants et post-doctorant de l'équipe de Virologie à la cristallogenèse et cristallographie. (Depuis 2010)
- Encadrement d'un ingénieur d'étude. (Depuis 2008)
- Suivi d’un sabbatique, U.Penn (Biologie Moléculaire, biochimie, Calorimétrie) (2007-2008).

F. Publications

i. Articles en préparation

(1): Yekwa E, Page, A, Mateo M, Canard B, Baize S, **Ferron, F**.

Crystal structure and functional analysis of Mopeïa virus Nucleoprotein exonuclease.

(2): Yekwa E, Apibanthamakit C, Khourieh J, Lichière J, Coutard B Canard B, **Ferron, F**.

Structural and functional comparison of old world / new world Nucleoprotein exonuclease.

(3): Sayez-Ayala M, Yekwa E, Emonet, S, Boretto J, Lichière J, Coutard B, Canard B, Alvarez, K, **Ferron, F.**

Structural guided approach to antiviral drug against endonuclease segmented ambisens RNA virus.

ii. Articles publiés

(1): Decroly E, Debarnot C, **Ferron F**, Bouvet M, Coutard B, Imbert I, Gluais L, Papageorgiou N, Sharff A, Bricogne G, Ortiz-Lombardia M, Lescar J, Canard B.

Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex.

PLoS Pathog. 2011 May;7(5):e1002059. doi: 10.1371/journal.ppat.1002059. Epub 2011 May 26.

(2): **Ferron F**, Li Z, Danek EI, Luo D, Wong Y, Coutard B, Lantez V, Charrel R, Canard B, Walz T, Lescar J.

The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes.

PLoS Pathog. 2011 May;7(5):e1002030. doi: 10.1371/journal.ppat.1002030. Epub 2011 May 12.

(3): Debarnot C, Imbert I, **Ferron F**, Gluais L, Varlet I, Papageorgiou N, Bouvet M, Lescar J, Decroly E, Canard B.

Crystallization and diffraction analysis of the SARS coronavirus nsp10-nsp16 complex.

Acta Crystallogr Sect F Struct Biol Cryst Commun. 2011 Mar 1;67(Pt 3):404-8. doi: 10.1107/S1744309111002867. Epub 2011 Feb 25.

(4): Morin B, Coutard B, Lelke M, **Ferron F**, Kerber R, Jamal S, Frangeul A, Baronti C, Charrel R, de Lamballerie X, Vonrhein C, Lescar J, Bricogne G, Günther S, Canard B.

The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription.

PLoS Pathog. 2010 Sep 16;6(9):e1001038. doi: 10.1371/journal.ppat.1001038.

(5): Massé N, Davidson A, **Ferron F**, Alvarez K, Jacobs M, Romette JL, Canard B, Guillemot JC.

Dengue virus replicons: production of an interserotypic chimera and cell lines from different species, and establishment of a cell-based fluorescent assay to screen inhibitors, validated by the evaluation of ribavirin's activity.

Antiviral Res. 2010 Jun;86(3):296-305. doi: 10.1016/j.antiviral.2010.03.010. Epub 2010 Mar 20.

(6): Kim HR, Graceffa P, **Ferron F**, Gallant C, Boczkowska M, Dominguez R, Morgan KG.

Actin polymerization in differentiated vascular smooth muscle cells requires vasodilator-stimulated phosphoprotein.

Am J Physiol Cell Physiol. 2010 Mar;298(3):C559-71. doi: 10.1152/ajpcell.00431.2009. Epub 2009 Dec 16.

(7): Qureshi IA, **Ferron F**, Seh CC, Cheung P, Lescar J.

Crystallographic structure of ubiquitin in complex with cadmium ions.

BMC Res Notes. 2009 Dec 15;2:251. doi:10.1186/1756-0500-2-251.

8: Malet H, Coutard B, Jamal S, Dutartre H, Papageorgiou N, Neuvonen M, Ahola T, Forrester N, Gould EA, Lafitte D, **Ferron F**, Lescar J, Gorbalenya AE, de Lamballerie X, Canard B.

The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket.

J Virol. 2009 Jul;83(13):6534-45. doi: 10.1128/JVI.00189-09. Epub 2009 Apr 22.

(9): Baek K, Liu X, **Ferron F**, Shu S, Korn ED, Dominguez R. Modulation of actin structure and function by phosphorylation of Tyr-53 and profilin binding.

Proc Natl Acad Sci U S A. 2008 Aug 19;105(33):11748-53.

doi: 10.1073/pnas.0805852105. Epub 2008 Aug 8.

(10): **Ferron F**, Rebowski G, Lee SH, Dominguez R.

Structural basis for the recruitment of profilin-actin complexes during filament elongation by Ena/VASP.

EMBO J. 2007 Oct 31;26(21):4597-606. Epub 2007 Oct 4.

(11): Egloff MP, Decroly E, Malet H, Selisko B, Benarroch D, **Ferron F**, Canard B.

Structural and functional analysis of methylation and 5'-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virusNS5.

J Mol Biol. 2007 Sep 21;372(3):723-36. Epub 2007 Jul 12.

(12): Lee SH, Kerff F, Chereau D, **Ferron F**, Klug A, Dominguez R.

Structural basis for the actin-binding function of missing-in-metastasis.

Structure. 2007 Feb;15(2):145-55.

(13): Longhi S, **Ferron F**, Egloff MP.

Protein engineering.

Methods Mol Biol. 2007;363:59-89.

(14): Imbert I, Guillemot JC, Bourhis JM, Bussetta C, Coutard B, Egloff MP, **Ferron F**, Gorbalenya AE, Canard B.

A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus.

EMBO J. 2006 Oct 18;25(20):4933-42. Epub 2006 Oct 5.

(15): Borrego-Diaz E, Kerff F, Lee SH, **Ferron F**, Li Y, Dominguez R.

Crystal structure of the actin-binding domain of alpha-actinin 1: evaluating two competing actin-binding models.

J Struct Biol. 2006 Aug;155(2):230-8. Epub 2006 Apr 25.

(16): Llorente MT, García-Barreno B, Calero M, Camafeita E, López JA, Longhi S, **Ferron F**, Varela PF, Melero JA.

Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization.

J Gen Virol. 2006 Jan;87(Pt 1):159-69.

(17): **Ferron F**, Bussetta C, Dutartre H, Canard B.

The modeled structure of the RNA dependent RNA polymerase of GBV-C virus suggests a role for motif E in Flaviviridae RNA polymerases.

BMC Bioinformatics. 2005 Oct 14;6:255.

(18): Severson W, Xu X, Kuhn M, Senutovitch N, Thokala M, **Ferron F**, Longhi S, Canard B, Jonsson CB.

Essential amino acids of the hantaan virus N protein in its interaction with RNA.

J Virol. 2005 Aug;79(15):10032-9.

(19): **Ferron F**, Rancurel C, Longhi S, Cambillau C, Henrissat B, Canard B.

VaZyMolO:a tool to define and classify modularity in viral proteins.

J Gen Virol. 2005 Mar;86(Pt 3):743-9.

(20): Egloff MP, **Ferron F**, Campanacci V, Longhi S, Rancurel C, Dutartre H, Snijder EJ, Gorbalenya AE, Cambillau C, Canard B.

The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world.

Proc Natl Acad Sci U S A. 2004 Mar 16;101(11):3792-6. Epub 2004 Mar 8.

(21): Karlin D, **Ferron F**, Canard B, Longhi S.

Structural disorder and modular organization in Paramyxovirinae N and P.

J Gen Virol. 2003 Dec;84(Pt 12):3239-52.

(22): Campanacci V, Egloff MP, Longhi S, **Ferron F**, Rancurel C, Salomoni A, Durousseau C, Tocque F, Brémond N, Dobbe JC, Snijder EJ, Canard B, Cambillau C.

Structural genomics of the SARS coronavirus: cloning, expression, crystallization and preliminary crystallographic study of the Nsp9 protein.

Acta Crystallogr D Biol Crystallogr. 2003 Sep;59(Pt 9):1628-31. Epub 2003 Aug 19.

(23): **Ferron F**, Longhi S, Henrissat B, Canard B.

Viral RNA-polymerases – a predicted 2'-O-ribose methyltransferase domain shared by all Mononegavirales.

Trends Biochem Sci. 2002 May;27(5):222-4.

(24): Nicolas A, **Ferron F**, Toker L, Sussman JL, Silman I.

Histochemical method for characterization of enzyme crystals: application to crystals of Torpedo californica acetylcholinesterase.

Acta Crystallogr D Biol Crystallogr. 2001 Sep;57(Pt 9):1348-50. Epub 2001 Aug 23.

iii. Articles méthodologiques, revues, eBooks

(1): Lieutaud P, **<u>Ferron F</u>**, Habchi J, Canard B and Longhi S.

Predicting protein disorder and induced folding: a practical approach.

Advances in Protein and Peptide Sciences, Vol. 1 (2013), *Bentham Science Publishers,* eBooks (chapitre 11)

(2): Bouvet M, Imbert I, **<u>Ferron F</u>**, Canard B, Decroly E.

Structures et fonctions des exoribonucléases d'arénavirus et de coronavirus. Virologie 2013; 17(5) : 317-30. Revue.

(3): **<u>Ferron F</u>**, Decroly E, Selisko B, Canard B.

The viral RNA capping machinery as a target for antiviral drugs.

Antiviral Res. 2012 Oct;96(1):21-31. doi:10.1016/j.antiviral.2012.07.007. Epub 2012 Jul 26. Revue.

(4): Bouvet M, **<u>Ferron F</u>**, Imbert I, Gluais L, Selisko B, Coutard B, Canard B, Decroly E.

Stratégies de formation de la structure coiffe chez les virus à ARN.

Med Sci (Paris). 2012 Apr;28(4):423-9. doi: 10.1051/medsci/2012284021. Epub 2012 Apr 25. Revue.

(5): Decroly E, **<u>Ferron F</u>**, Lescar J, Canard B.

Conventional and unconventional mechanisms for capping viral mRNA.

Nat Rev Microbiol. 2011 Dec 5;10(1):51-65. doi: 10.1038/nrmicro2675. Revue.

(6): Albeck S, Alzari P, Andreini C, Banci L, Berry IM, Bertini I, Cambillau C, Canard B, Carter L, Cohen SX, Diprose JM, Dym O, Esnouf RM, Felder C, **<u>Ferron F</u>**, Guillemot F, Hamer R, Ben Jelloul M, Laskowski RA, Laurent T, Longhi S, Lopez R, Luchinat C, Malet H, Mochel T, Morris RJ, Moulinier L, Oinn T, Pajon A, Peleg Y, Perrakis A, Poch O, Prilusky J, Rachedi A, Ripp R, Rosato A, Silman I, Stuart DI,Sussman JL, Thierry JC, Thompson JD, Thornton JM, Unger T, Vaughan B, Vranken W, Watson JD, Whamond G, Henrick K.

SPINE bioinformatics and data-management aspectsof high-throughput structural biology.

Acta Crystallogr D Biol Crystallogr. 2006 Oct;62(Pt 10):1184-95. Epub 2006 Sep 19. Revue.

(7): **<u>Ferron F</u>**, Longhi S, Canard B, Karlin D.

A practical overview of protein disorder prediction methods.

Proteins. 2006 Oct 1;65(1):1-14. Article méthodologique.

iv. Brevet

(1): Imbert I, Guillemot JC, Canard B & **Ferron F**.

RNA-dependent RNA polymerase from SARS-CoV and its homologs from other coronaviruses and their use in molecular biology and drug screening.

PATENT: EP 1619246 ISSUE DATE: 20060125 APPLICATION: 2004-29735 PP: 27 pp.

G. Communications et Structures déposées

COMMUNICATIONS ORALES (Depuis 2006)

(1): **Wistar Institute**, Philadelphia, PA, April **2008**. Cordon-Bleu (Cobl) : A new class of actin nucleator ?

(2): **Wistar Institute**, Philadelphia, PA, September **2007**. Actin Filament Elongation mechanism by Ena/Vasp.

(3): **Boston Biomedical Research Institute Annual Retreat**, Woods Hole, MA, Avril **2006**. Understanding the role of profilin in cytoskeleton assembly.

COMMUNICATIONS par AFFICHES (Depuis 2006)

(1): **Ferron F**, Danek E.I , Li Z, Wong Y.H., Luo D, Coutard B, Lantez V, Canard B, Walz T, Lescar J (2010)***.***

Structural studies of the rift valley fever virus nucleoprotein.

The Negative Strand Virus Meeting 2010 June 21–25, Brugge, Belgium.

(2): Morin B, Coutard B, Lelke M, **Ferron F**, Kerber R, Jamal S, Frangeul A, Baronti C, Charrel R, de Lamballerie X, Vonrhein C, Lescar J, Bricogne G, Günther S, and Canard B (2010)***.***

The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mrRNA transcription.

The Negative Strand Virus Meeting 2010 June 21–25, Brugge, Belgium.

(3): Kim H.R, **Ferron F**, Boczkowska M, Graceffa P, Gallant C, Leavis P, Dominguez R., Morgan K.G (2009).

Actin polymerization in differentiated vascular smooth muscle cells requires vasodilator-stimulated phosphoprotein (VASP).

The Biophysical Society's 53rd Annual Meeting. 28 Feb. - 4 March, Boston, MA, USA.

(4): **Ferron F**, Rebowski G, Lee S.H, Dominguez R (2008).

Structural Basis For The Recruitment Of Profilin-Actin Complexes During Filament Elongation By Ena/vasp.

52nd Annual Meeting & 16th International Biophysics Congress. 2-6 Février, Long Beach, CA, USA.

(5): **Ferron F**, Li Y, Rebowski G, Chereau D and Dominguez R (2006).

Transition State for Actin Filament Elongation by Ena/VASP.

46th Annual Meeting of the American Society for Cell Biology (ASCB), Dec. 9-13, 2006, San Diego, CA, USA.

STRUCTURES DEPOSEES PUBLIEES

Resolution Å	PDB entry	Title	Release date
2.30	4TU0	rystal structure of chikungunya virus nsp3 macro domain in complex with a 2'-5' oligoadenylate trimer	09/07/2014
2.20	4C5Q	Measles virus phosphoprotein tetramerization domain	09/04/2014
2.10	4BHV	Measles virus phosphoprotein tetramerization domain	09/04/2014
1.25	4C6A	High Resolution Structure of the Nucleoside diphosphate kinase	20/11/2013
2.06	2xyv	Crystal structure of the nsp16 nsp10 sars coronavirus complex	26/10/2011
2.5	2xyr	Crystal structure of the nsp16 nsp10 sars coronavirus complex	26/10/2011
2.0	2xyq	Crystal structure of the nsp16 nsp10 sars coronavirus complex	19/10/2011
2.3	3ouo	Structure of the Nucleoprotein from Rift Valley Fever Virus	25/5/2011
1.6	3ov9	Structure of the Nucleoprotein from Rift Valley Fever Virus	25/5/2011
2.3	3gqe	Crystal structure of macro domain of Venezuelan Equine Encephalitis virus	21/7/2009
2.6	3gqo	Crystal structure of macro domain of Venezuelan Equine Encephalitis virus in complex with ADP-ribose	21/7/2009
1.65	3gpg	Crystal structure of macro domain of Chikungunya virus	21/7/2009
1.9	3gpo	Crystal structure of macro domain of Chikungunya virus in complex with ADP-ribose	21/7/2009

2.0	3gpq	Crystal structure of macro domain of Chikungunya virus in complex with RNA	21/7/2009
3.0	3h1u	Structure of ubiquitin in complex with Cd ions	5/5/2009
2.3	3chw	Complex of Dictyostelium discoideum Actin with Profilin and the Last Poly-Pro of Human VASP	19/8/2008
1.6	3cip	Complex of Dictyostelium Discoideum Actin with Gelsolin	19/8/2008
1.7	3ci5	Complex of Phosphorylated Dictyostelium Discoideum Actin with Gelsolin	19/8/2008
1.5	2pbd	Ternary complex of profilin-actin with the poly-PRO-GAB domain of VASP	13/11/2007
1.8	2pav	Ternary complex of Profilin-Actin with the Last Poly-Pro of Human VASP	23/10/2007
2.2	2p3l	Crystal Structure of Dengue Methyltransferase in Complex with GpppA and S-Adenosyl-L-Homocysteine	28/8/2007
2.7	2p3o	Crystal Structure of Dengue Methyltransferase in Complex with 7MeGpppA and S-Adenosyl-L-homocysteine	28/8/2007
2.75	2p3q	Crystal Structure of Dengue Methyltransferase in Complex with GpppG and S-Adenosyl-L-homocysteine	28/8/2007
2.7	2p40	Crystal Structure of Dengue Methyltransferase in Complex with 7MeGpppG	28/8/2007
1.8	2p41	Crystal Structure of Dengue Methyltransferase in Complex with 7MeGpppG2'OMe and S-Adenosyl-L-homocysteine	28/8/2007
2.5	2d1k	Ternary complex of the WH2 domain of mim with actin-dnase I	12/9/2006
1.85	2d1l	Structure of F-actin binding domain IMD of MIM (Missing In Metastasis)	12/9/2006
1.7	2eyi	Crystal structure of the actin-binding domain of human alpha-actinin 1 at 1.7 Angstrom resolution	29/8/2006
1.8	2eyn	Crystal structure of the actin-binding domain of human alpha-actinin 1 at 1.8 Angstrom resolution	29/8/2006
2.7	1qz8	Crystal structure of SARS coronavirus NSP9	24/2/2004

I. RESUME DU PARCOURS SCIENTIFIQUE

J'ai commencé mon parcours scientifique à l'Université de la Méditerranée à Marseille dans le laboratoire de B. Canard. J'ai effectué ma thèse sous la co-direction de S. Longhi et B. Canard, sur l'étude des réplicases virales par bio-informatique et biologie structurale. Notre objectif était d'une part de développer une base de données se rapportant aux virus à ARN simple brin, et qui gèrent les informations structurales et fonctionnelles des protéines virales, en les classant par module. Le module étant défini comme la fraction (totale ou partielle) de la protéine qui est cristallisable. Cette base de données (VaZyMolO) a été publié en 2005. Cette approche a permis entre autre l'identification d'un domaine méthyltransférase sur la protéine L des Mononegavirales et la définition de la modularité des protéines N et P des *Paramyxovridae*. Par la suite, nous avons construit la première cartographie modulaire du génome du virus du Syndrome Respiratoire Aiguë Sévère Coronavirus lors de l'épidémie de 2003. En dehors des données structurales et d'évolution obtenues à partir de ce travail d'annotation minutieux, il m'a permis de mettre en place une méthode d'annotation particulièrement efficace pour identifier et étudier les régions désordonnées des protéines. Ces régions sont d'un intérêt fondamental dans la mesure où elles jouent un rôle dans le recrutement d'assemblage dynamique des macromolécules du complexe de réplication, du déplacement cytoplasmique ou lors de la stratégie d'évasion à la réponse de l'immunité innée.

En 2005, après la soutenance de ma thèse, J'ai rejoins le groupe de R. Dominguez, d'abord au Boston Biomedical Research Institute à Boston, puis à l'Université de Pennsylvanie à Philadelphie. Lors de ses stages post-Doctoraux, je me suis formé à la cristallographie aux rayons X et aux méthodes biophysiques (fluorescence, calorimétrie), en m'intéressant à la résolution de complexe de macromolécules cellulaires ainsi qu'à leur recrutement dynamique par des régions protéiques désordonnées. Ces travaux ont mené à la résolution de plusieurs structures impliquées dans la régulation du filament d'actine ou de sa structuration près de la membrane. En particulier, nous avons pu mettre en évidence que le recrutement du complexe de profiline-actine par la protéine VASP se fait d'abord par une région désordonnée, qui ensuite transfert ce complexe à une hélice transitoire (WH2) permettant le positionnement correct de l'actine monomérique sur le filament d'actine. Ce mécanisme permet une croissance rapide et contrôlée du filament d'actine.

En 2009, j'ai ensuite rejoins le groupe de J. Lescar à Singapour, qui souhaitait développer dans le cadre d'un Projet International de Coopération Scientifique (PICS) avec le CNRS, une partie de ses projets dans le laboratoire AFMB à Marseille. Grace à une ATIP j'ai pu co-diriger l'équipe de Marseille, et avec l'aide d'un ingénieur d'étude, j'ai développé un projet de recherche sur la caractérisation structurale des Nucléoprotéines de Phlébovirus. Dans ce cadre, j'ai établi une collaboration internationale avec l'équipe de T. Waltz de l'Université de Harvard. Les résultats de ce post-doctorat ont mené à la résolution de la structure à basse et haute résolution de la Nucléoprotéine du virus de la vallée du Rift, sous forme d'anneaux hexamériques.

Après mon recrutement au CNRS en 2010 en tant que CR2, j'ai rejoint le groupe de B. Canard à l'AFMB. Mon objectif était de pouvoir utiliser mes compétences en bio-informatique, biochimie, biologie structurales et biophysique afin d'étudier le processus d'assemblage du complexe de réplication des *Arenaviridae* et *Bunyaviridae*. De plus, je me suis intéressé à la résolution des structures de protéines impliquées dans le processus de la formation de la coiffe chez les virus à ARN étudiés au laboratoire. La formation des structures « coiffe » sur les ARN viraux messagers nécessite le recrutement de complexes macromoléculaire dynamiques impliquant parfois des régions désordonnées et/ou probablement des changements de conformation importants. Nous espérons que la caractérisation de ces différentes interactions pourra mener à l'identification d'inhibiteurs structuraux ciblant l'assemblage de ces complexes.

II. CONTEXTE DE CE MANUSCRIT

A. Les assemblages macromoléculaires

i. Fonction et assemblage dynamiques

L'étape primordiale et obligatoire dans la recherche biologique passe par l'étude de l'activité et de la structure d'une protéine isolée de son contexte biologique. Par ce biais, il est possible de caractériser la fonction de la protéine ou par sa structure de l'utiliser comme outil dans le cadre de design rationnel de molécules. Cependant, ce processus empirique d'accumulation des données ne permet pas toujours une description correcte des processus biologiques ou structuraux qui se passent dans la cellule. La plus part des processus biologiques ; et cela est encore plus vrai pour les virus ; sont issus d'assemblages moléculaires complexes [1]. Ces derniers sont souvent dynamiques et vont accomplir une ou plusieurs tâches impliquant des ré-assemblages structuraux au cours du cycle dans lequel ils sont impliqués. La compréhension de ces mécanismes passe par la détermination des structures correspondant aux différentes étapes et à la description de leurs modes d'action. En plus du problème d'échelle auquel est confronté le biologiste, s'ajoute la stabilisation des complexes dans des états compatibles pour l'étude structurale et garantissant un surplus d'information toujours pertinent. Les premières données qui ont reflété ses grands assemblages structuraux furent celles amenées par l'étude de capside de virus, puis des larges complexes des « machines cellulaires », membranaires ou non [2]. Les études structurales relatives aux protéines de la régulation du cytosquelette s'avèrent être l'un de ces challenges majeurs. Ce sont des protéines dont l'assemblage forme des architectures transitoires voir furtives, mais stable et hautement régulé. « Transitoire mais stable », c'est un paradoxe. Ainsi, la naissance et la croissance d'un filament d'actine, font appel à des partenaires protéiques qui tour à tour peuvent être initiateurs, élongateurs, inhibiteurs et dé-assembleurs. Chacun va former un complexe stable correspondant à l'étape pour lequel il est affecté et dont l'état dépend de l'état structural ou fonctionnel qui le précède. Après maint efforts, la biologie structurale a mis au point des stratégies pour « geler » et étudier les différentes étapes transitoires de ces complexes [3]. Ce qu'il y a de plus extraordinaire dans les protéines de la régulation du cytosquelette, c'est qu'elles ont intégré une panoplie très étendue de techniques pour contrôler l'actine [4]. En particulier, le contrôle d'une activité par l'activation structurale (formation d'un complexe jouant le rôle d'un interrupteur moléculaire), ou encore l'usage de régions désordonnées qui servent de plate-forme de recrutement à tout un panel de protéines régulatrices en plus de l'actine elle-même [5–12]. Ce sont des caractéristiques similaires que l'on retrouve chez les virus. La maturation de la polyprotéine de nombreux virus à ARN (+) permet d'activer ou d'inhiber une activité au cours du cycle virale (*i.e.* Flavivirus / Coronavirus) ; ou l'usage de régions désordonnées chez les virus à ARN (-) pour perturber la réponse à l'interféron ou réguler les étapes de réplication / transcription, en sont de parfait exemples [13–21]. Ainsi, comme nous l'avons montré avec les nucléoprotéines de *Phlébovirus*, les protéines existent dans des états structuraux différents pour répondre à des contraintes structurales mécaniques, mais probablement aussi pour temporiser le cycle réplicatif.

Ce qu'il y a de frappant lorsque l'on étudie et compare les protéines de régulation cellulaire et les protéines virales, c'est leur capacité à modifier profondément leur environnement [1]. En cela, il y a une certaine homologie dans le résultat affectant la cellule. En effet, la cellule génère des micro-environnements lors de processus d'adaptation, de restructuration de la membrane ou lors de transport intracellulaire. De même, lors d'une infection virale, l'une des premières conséquences pour la cellule est la création d'un micro-environnement (usine à virions) créé par des assemblages macromoléculaires dynamiques [22]. Ce micro environnement est imposé par le virus, mais montre combien les virus ont réussi à adapter leurs protéines aux membranes et / ou aux protéines qui s'occupent de l'architecture des membranes. Ainsi, la description au niveau atomique de ces interactions recèlent certainement des réponses sous-jacentes à la réplication tout autant qu'à l'encapsidation. L'un des cas les plus marquant, se trouve chez les

Arenavirus. Le cycle viral des Arenavirus est assez bien connu d'un point de vu de sa localisation cellulaire et de sa cinétique [23,24]. En revanche, les données structurales et fonctionnelles restent extrêmement limitées et ce malgré le fait que les Arenavirus utilisent un nombre limité de protéines pour accomplir leur cycle réplicatif. La difficulté vient de l'obtention de large complexe ribonucléoprotéique qui permettent de comprendre l'assemblage fonctionnel de ces machines virales. Un autre fait surprenant chez les Arenavirus est de retrouver des ribosomes dans leurs virions [25]. Ils ne sont plus actifs dans le virion mais leurs présence dans la particule virale montrent combien ces derniers se trouvent complexés aux protéines virales au point d'être « arraché » à la cellule. Le fait de retrouver systématiquement des ribosomes chez les *Arenaviridae* dans le virion et non pas chez les *Bunyaviridae,* une famille relativement proche, suggère la présence de mécanisme ou d'adaptation propre qui restent à être élucidé.

Les études des complexes impliquant les protéines interagissant avec les acides nucléiques ADN et ARN polymérases, ribosomes... se sont révélées être, et sont encore pour la plupart d'entre eux, plus complexes à obtenir qu'originellement pensé. Les conformations des structures obtenues des protéines seules correspondent souvent à un état stable mais pas forcément à l'état actif. Ainsi si la cristallographie est la seule technique qui donne un résultat absolu, il est d'autant plus nécessaire de le discuter à la lumière de la biologie. Les challenges qui s'offrent aux cristallographes sont donc énormes car il s'agit de résoudre ces assemblages architecturaux qui constituent la nouvelle frontière.

L'étape primordiale et obligatoire dans la recherche biologique passe par l'étude de l'activité et de la structure d'une protéine isolée de son contexte biologique. Par ce biais, il est possible de caractériser la fonction de la protéine ou par sa structure.

ii. Apport de la cristallographie dans la caractérisation des macromolécules biologiques

La contribution de la biologie structurale et de la cristallographie en particulier représente une source de progrès qui génère des retombées non seulement en recherche fondamentale mais aussi en recherche appliquée. Mais les bénéfices de cette contribution va au-delà. A la fin des années 1970, les cristallographes sont confrontés à tellement de défis conceptuels et technologiques, qu'ils mettent en place un projet cristallographique global et collaboratif [26], par cette approche unique ils ont révolutionnés la façon de penser la recherche. En une vingtaine d'année, tout les aspects théoriques et techniques ont été couvert. Ils ont permis de sortir la cristallographie et la diffraction aux rayons x du microcosme d'experts vers une communauté plus large, tout en garantissant un accompagnement de tout les développements techniques [27]. C'est ce modèle qui influence encore aujourd'hui certaines communautés de développeurs et d'universitaires. L'omniprésence de la cristallographie en biologie, s'explique par la valeur ajouté de l'information qu'elle produit éclairant définitivement l'interprétation du biochimiste ou du biologiste cellulaire. Les processus biologiques complexes sont les résultats d’interactions dynamiques soit de macromolécules (protéiques ou nucléiques) entre elles, soit de macromolécules avec de petits substrats cellulaires. La fonction biologique d'une macromolécule est définie d'une part par sa conformation et d'autre part par sa dynamique. Pour être fonctionnelle, une protéine doit être le plus souvent être « correctement » repliée ; ce qui veut dire être dans un état stable (natif), qui lui permet, si nécessaire, de changer de conformation pour accomplir ce pourquoi elle est faite. Afin de pleinement comprendre comment la protéine fonctionne, il est important sinon nécessaire d'accéder à la structure « fine » des protéines. La cristallographie est la méthode par excellence pour définir avec une quasi certitude au niveau atomique la structure de ces macromolécules seul ou en complexe. En combinaison avec d'autres méthodes spectroscopiques elle permet d'atteindre le rêve de tout « voir » au cœur de la matière [28].

Historique

La cristallographie ; dont on vient de célébrer le centenaire en 2014 [29–31]; est la méthode qui a probablement le plus influencé les différents domaines de la science. De la géologie, à la nanotechnologie en passant par la biologie, c'est par elle que les scientifiques ont la possibilité de décrire l'organisation de la matière au niveau atomique. C'est par les recherches appliquées de Max Von Laue et de William Henry et William Lawrence Bragg (père et fils) [32] sur les cristaux de roche que le principe de la diffraction des rayons x pour remonter à la structure atomique fut posé (1913) [32] . C'est dans les année 1930, que l'on démontra la possibilité de cristalliser des protéines [33], en 1937 Max Perutz commence un projet en ayant pour objectif la détermination de la structure cristalline de l'hémoglobine. En 1945 Dorothy Hodgkin et son équipe détermine la structure de la pénicilline [34]. En 1953, Perutz et collègues décrivent l'application de la méthode de remplacement isomorphe multiple (MIR) pour la détermination de structures cristallines de protéine. La même année, Watson J. et Crick F. décrivent la structure en double hélice de l'ADN [35], grâce à la technique de diffraction des rayons X [36] (à partir du travail de Rosalind Elsie Franklin [37,38]). Il faut attendre 1958 pour que la technologie permette la résolution de la première structure atomique d'une protéine, la Myoglobine par Kendrew J. [39]. Max Perutz après 25 années de travaille finit la structure de l'hémoglobine (1962) [40]. La structure du lysozyme (1965) [41] confirmait définitivement que la méthode pouvait être systématisée. Dès 1959, Michael Rossmann commence à développer les premiers codes pour la résolution et l'analyse de structure. En 1973, Rossmann décrit l'un des repliements fondamentaux de la biologie (le *Rossmann fold*) [42], qui est un motif structural protéique présent dans les protéines qui lient les nucléotides, particulièrement dans la nicotinamide adénine dinucléotide, les méthyltransférases, les déshydrogénases ou kinases qui lient les molécules contenant une adénosine triphosphate. En 1978, Gérard Bricogne et Steven Harrisson réussissent pour la première fois à déterminer des structures virales à l'échelle atomique [43]. En 1979, A Jones développe les prémices de ce qui va être l'outil de visualisation des cartes de densité électronique [44]. Bien sur la liste des contributeurs n'est pas exhaustive et bien trop courte pour rendre hommage à tous ceux qui ont contribué au développement de la méthode.

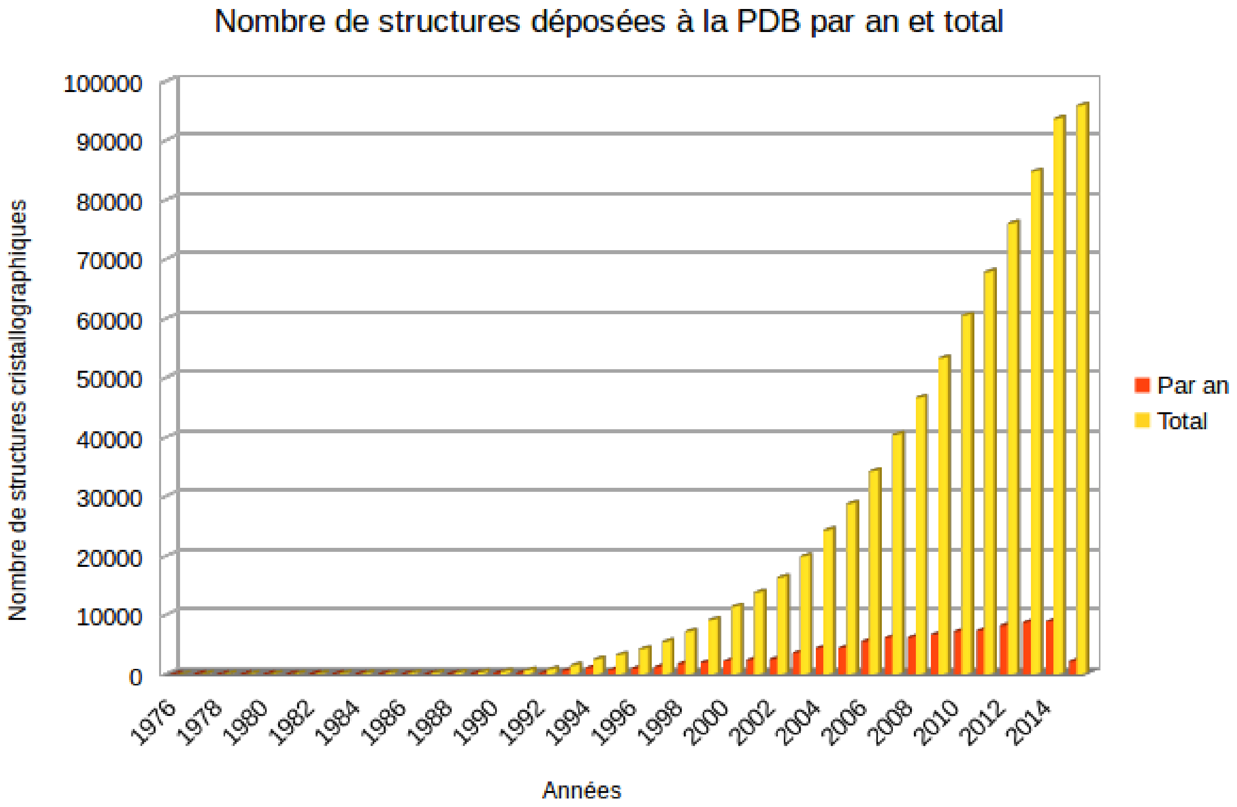


Figure 1: Statistiques des stuctures déposées à la PDB

Enfin, c'est avec la maîtrise et l'amélioration des sources de lumière synchrotron (1995) [45,46] couplées à l'accélération des capacités de calculs, et l'avènement de la biologie moléculaire (2000); qui a permis de produire massivement des protéines recombinantes ; que l'accélération de la détermination de structure a vraiment commencé [47–50]. Aujourd'hui, nous sommes à près de 96.000 structures de protéines déposées et le rythme de dépôts par an ne cesse d'augmenter.

Principes

La cristallographie n'est pas une méthode de visualisation directe qui focalise la lumière des objets à travers des lentilles afin d'obtenir une image magnifiée. Elle nécessite tout d'abord que « l'objet » d'intérêt soit cristallisé. Le cristal étant un assemblage dans lequel un **motif atomique** se répète périodiquement dans les trois directions de l'espace, engendrant ainsi un

réseau. On peut décomposer ce dernier en série de parallélépipèdes dont les arêtes sont parallèles aux axes cristallographiques. Ensuite, elle exploite la caractéristique de la longueur d'onde des rayons X (comprise entre 0,05nm et 5nm) qui fait qu'ils sont dispersés par le nuage électronique des atomes des objets exposés. Comme un réseau optique diffracte la lumière, le réseau cristallin va diffracter les rayons X. Ce phénomène est lié aux interférences qui existent entre les ondes diffusées par les atomes. Dans certaines directions, les interférences seront constructives et une intensité sera diffractée, alors que dans d'autres directions, les interférences seront destructives et aucune réflexion ne sera observée. Les réflexions observées peuvent être collectées ; elles se traduisent sous forme de taches plus ou moins intenses et organisées sous forme de motif. L'interprétation de ces motifs permet de définir les paramètres de la maille et de symétrie du cristal. L'analyse des réflexions permet d'obtenir l'organisation des atomes des molécules qui forment le cristal. Ces réflexions contiennent l'information du contenue de l'unité asymétrique, mais l'information n'est pour autant pas directement accessible. L'information se trouve codée dans la distribution des **intensités** des réflexions dans « l'espace réciproque ». L'espace réciproque est une construction abstraite qui permet de représenter et caractériser des phénomènes complexes en utilisant un espace vectoriel. L'utilisation de la transformation de Fourrier permet de ramener cette information dans « l'espace réel » donnant une image de la molécule cristallisée. La transformation de Fourier est une opération mathématique qui requiert deux termes : (1) les amplitudes (facteur de structure) que l'on peut déterminer de l'intensité des taches de diffraction. (2) l'angle relatif de la phase correspondante à chaque tache de diffraction [51]. La résolution de ce problème, appelé « problème de la phase », est l'étape indispensable à toute détermination de structure cristalline. Cette dernière information doit être déduite soit de phases déjà connus d'un modèle proche déjà résolu (remplacement moléculaire) ; soit de méthodes expérimentales qui permettent de l'identifier.

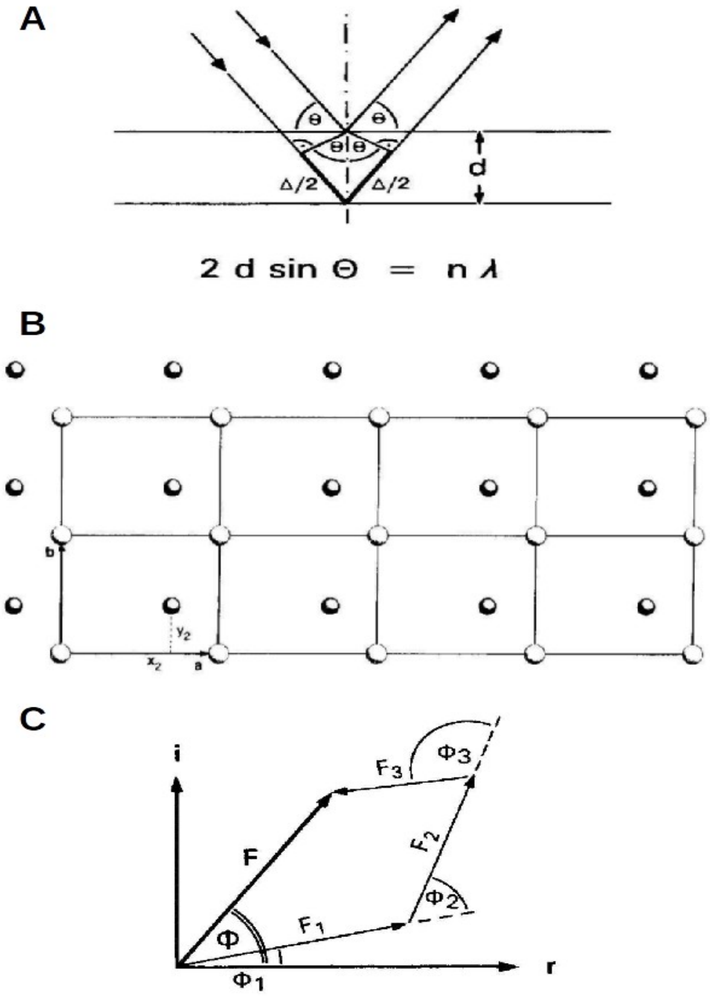


Figure 2: A) Loi de Bragg: les ondes diffusées par deux atomes localisés dans des plans réticulaires adjacents produiront des interférences constructives si la différence de chemin parcourue par ces ondes, Δ , est un multiple entier de la longueur d'onde de la radiation incidente. Elle permet de prévoir l'angle 2 θ entre le faisceau incident et le rayon diffracté par une famille de plan hkl , qui est fonction de la distance réticulaire d_{hkl} et de la longueur d'onde λ de la radiation incidente. B) Modèle simple de structure cristalline à deux atomes. Pour cet atome, le réseau est identique et la géométrie de diffraction restera la même. Lorsque le cristal est exposé au faisceau de rayons X de telle sorte qu'une famille de plans hkl soit en position de diffraction, tous les atomes de type 1 vont diffuser les rayons X en phase l'un avec l'autre. Il en est de même pour tous les atomes de type 2. Toutefois, le second réseau, auquel appartient le second atome, a été déplacé par rapport à l'origine d'un vecteur x_2 , y_2 , z_2 . L'onde résultante diffusée par ces atomes de type 2 présentera par conséquent un déphasage par rapport à l'onde résultante diffusée par le premier atome. Ce déphasage sera différent pour chaque réflexion considérée. La résultante de l'ensemble des ondes diffusées dans la direction de réflexion hkl des atomes de la maille s'appelle le facteur de structure, caractérisé par une amplitude F et une phase φ. C) Addition vectorielle des ondes diffusées par des atomes individuels, aboutissant au facteur de structure, d'amplitude F et de phase φ. Les intensités mesurées fournissent les amplitudes des facteurs de structure, mais pas leurs phases.

- Dans le cas du remplacement moléculaire, on utilise un modèle « proche » pour calculer la phase initiale, qui permet la construction de la densité électronique à partir des données. Cette méthode implique le positionnement correct du modèle dans le cristal puis le calcule de la diffraction théorique que l'on compare à celle qui est observée. Ceci permet de calculer des phases théoriques initiales. La méthode est relativement rapide pour déterminer les phases initiales, mais elle peut aussi produire une phase biaisée. En effet la reconstruction de la densité électronique est dépendante des phases, du coup le modèle initiale reflétera surtout caractéristiques du modèle utilisé pour le remplacement moléculaire et non pas le vrai modèle. Il est important alors d'utiliser des méthodes de correction du biais pour construire le modèle correcte. Cette méthode est massivement utilisé, soit environ les 3/4 des structures déposées [52].

- Dans le cas où il est impossible de trouver un modèle compatible pour déterminer les phases, il faut passer par la détermination expérimentale des phases [53]. Leur détermination *de novo*, passe par la mesure d'une différence d'intensité d'un « marqueur » atomique (Soufre, Sélénium dans des Méthionines modifiées, sels de mercure ou Tantale, *etc*), avec et entre les intensités de jeux de données isomorphes. Ainsi le remplacement isomorphe multiple, consiste à faire diffuser des atomes lourds (riches en électrons) dans un cristal et à comparer la diffraction avec et sans ces éléments lourds. Les « atomes lourds » modifient légèrement les intensités de diffraction, ce qui permet de calculer les phases par triangulation. La limite est d'arriver à positionner les atomes lourds dans la maille du cristal. La diffusion anomale, consiste à faire varier la longueur des rayons X autour du seuil d'absorption d'un des types d'atomes contenu dans la molécule. Le sélénium est fréquemment utilisé car il possède un seuil d'absorption X proche des longueurs d'ondes utilisées (autour de 0,1 nm). Les différences anomales seront les plus significatives lorsque la longueur d'onde des rayons x est légèrement supérieur au seuil d’absorption du marqueur. Dans ces conditions les photons incidents sont alors absorbés et provoquent une excitation électronique de l'atome. Cette excitation génère une diffusion anomale (ou diffusion résonante) ainsi qu'une variation significative des intensités au sein même du jeu de donnée (pour une longueur d'onde d'excitation donnée) ou entre les jeux de données si le cristal est soumis à plusieurs longueur d'ondes. L'information obtenue de ce signal permet alors de déterminer une sous structure des atomes marqueurs à partir de laquelle on peut calculer les phases initiales nécessaire pour reconstruire la densité électronique. L'avantage d'utiliser les marqueurs atomiques des acides aminés naturels ou modifiés (S ou Se-Methionine) c'est de fournir un signal anomal spécifique de la structure et de la séquence. Le phasage expérimentale a été utilisé pour 1/4 des structures résolues. Le phasage anomale ; du fait de sa simplicité et sensibilité ; est largement plus utilisé que les expériences par remplacement isomorphe multiple.

Évaluation des structures cristallographiques

Tout scientifique amené à regarder une structure cristallographique, doit (ou devrait) avoir les moyens d'évaluer la qualité de la structure qu'il observe. Non pas pour la juger ; il n'est pas forcément expert ; mais pour garantir une observation critique de la structure et éviter la surinterprétation. Les structures cristallographiques des protéines sont déposées à la « Worldwide Protein Data Bank (http://www.wwpdb.org/) ». Lors de la soumission d'une structure, la PDB effectue un certain nombre de vérifications pour garantir un contrôle qualité minimal (vérification de la géométrie, distance, angles et torsion des liaisons de la chaîne principale). Ainsi ces informations apparaissent sous la forme du critère de la valeur quadratique moyenne de déviation (ou Root Mean Square Deviation-RMSD) qui décrit de combien en Angström ou en degrés les liaisons ou les angles dévies de la valeur idéale. Ainsi une structure, « fiable »devrait avoir un RMSD de moins 0.015Å pour ses liaisons et de moins 1.5° pour ses angles. Classiquement la qualité de la géométrie de la chaîne principale d'un modèle est présentée sous la forme d'un diagramme de Ramachandran. Ce diagramme définit les valeurs de l'angle diédral φ en abscisse et celle de l'angle diédral ψ en ordonnée, pour chaque acide aminé de la protéine. Un nombre limité de valeurs des angles φ et ψ sont possibles, ainsi les résidus pour lesquels les angles sortent des régions autorisées sont instantanément identifiables [54]. Une structure ne doit avoir aucun résidu dans les régions interdites.

Au delà des critères géométriques, un critère bien plus important est celui de comment le modèle corrèle avec les données expérimentales [55]. En effet, les structures observées ne sont pas la réalité mais une représentation de cette réalité. L'avantage de la cristallographie sur les autres méthodes structurales et la microscopie électronique en particulier, c'est que le cristallographe a les moyens d'évaluer la qualité du modèle proposé par rapport au données collectées. Deux paramètres de fiabilités sont associés aux structures cristallographiques les valeurs $R_{\text{-de travail}}$ et $R_{\text{-libre}}$. La valeur $R_{\text{-de travail}}$ quantifie l'accord global entre les amplitudes expérimentales et les amplitudes théoriques calculées de la structure du modèle. Ainsi un modèle qui est « en accord » avec ses données va avoir des amplitudes calculées très similaires aux amplitudes expérimentales et un $R_{\text{-de travail}}$ bas. Le $R_{\text{-de travail}}$ s'exprime en pourcentage et une valeur correcte est comprise entre 15 % et 25 %. La valeur $R_{\text{-libre}}$ est une valeur de validation croisée. Cette valeur permet d'éviter la sur interprétation des données et d'éviter le biais introduit lors de la construction et l'affinement. Le $R_{\text{-libre}}$ est calculé de la même façon que le $R_{\text{-de travail}}$ mais non pas l'ensemble des amplitudes mais sur un échantillonnage qui a été séparé et qui n'a subit aucune altération durant l'affinement. Ainsi, le comportement du $R_{\text{-libre}}$ suit celui du $R_{\text{-de travail}}$ et lorsque le $R_{\text{-libre}}$ s'écarte du $R_{\text{-de travail}}$, c'est le signe que les données ne supportent plus l'interprétation du modèle. Une différence entre les 2 R de 5 % est acceptable, au-delà c'est une indication qu'il y a un problème avec la structure. Le facteur de température (ou facteur B) est aussi une indication importante quand à la mobilité des atomes présent dans le cristal (protéine, acide nucléique, ligand ou solvant). Un facteur B élevé est une indication d'une mobilité de position dans le cristal. Pour un complexe, il peut être indicatif d'une faible occupation du site de liaison.

Enfin, la résolution à laquelle les réflexions ont pu être mesurées donnent aussi une idée générale au combien les détails sont fiables au regard de la qualité de la carte de densité électronique. Ainsi la majeur partie des structures déposées ont une résolution comprise entre 1.5Å et 2.5Å. A plus haute résolution il est possible d'accéder aux détails fin de la structure tels que : les conformations alternatives des chaînes latérales, la position plus précise des ligands et dans certains cas la déformation de la densité électronique permet d'accéder à l'état de charge des atomes du site actif et de leur potentiel électrostatique. Dans ces conditions il est possible d'envisager, une compréhension au niveau sub-atomique de processus biologiques complexes tels que les transferts d'électrons et de protons, les réactions d'oxydo-réduction faisant intervenir des centres métalliques réactionnels, l'activation du dioxygéne (hémoglobine) [56,57]. A moins de 3Å la structure reste informative, elle fournit une information générale sur le repliement ou sur l'organisation des domaines et les interactions éventuelles, elle ne permet pas d'accéder aux détails fin de la structure ni du positionnement des chaînes latérales ou de petit ligand.

Ainsi la conjonction de progrès réalisés tant au niveau technologique (biologie moléculaire, sources de rayons X, détecteurs de rayons X, supercalculateurs miniatures) qu’au niveau des logiciels de traitement des données de diffraction (collecte, phasage, affinement) et la rigueur de la méthode ont permis à la cristallographie de prendre une place centrale. Cette méthode a permis la détermination des structures tridimensionnelles de plusieurs milliers de macromolécules biologiques dans des gammes de taille et de complexité très variées : petites protéines, oligonucléotides, acides ribonucléiques de transfert, immunoglobulines, assemblage complexe du cytosquelette, complexes nucléoprotéiques, macro-complexe fonctionnels (ribosome) et structuraux (cytoskelette), et bien évidement des virus : de bactéries, d’insectes, de plantes ou de mammifères.

B. Les virus

En 1957, lors de la troisième conférence commémorative de Marjory Stephenson, A. Lwolf conclue sa présentation par la définition suivante : « les virus doivent être considérés comme des virus parce que c'est ce qu'ils sont. » [58]. Lwoff est le père de la virologie moderne en cela qu'il est le premier à comprendre la portée de ses résultats à la lumière des 60 années qui le précède et qui lui permette d'énoncer les caractères fondamentaux qui font des virus des entités originales : *« Les virus sont de petites tailles, infectieux et potentiellement pathogènes ; ce sont des entités nucléo-protéiques possédant un seul type d'acide nucléique (ARN ou ADN) ; ils sont reproduits (par la cellule) à partir de leur matériel génétique ; ils sont incapables de croître et de se diviser ; ils sont dépourvus de système de Lipmann. »*

Cette définition est un peu désuète au regard de la virologie moderne mais révolutionnaire à l'époque où elle fut énoncée. Elle mit définitivement fin au dénigrement issus de la bactériologie qui définissait les virus (non visible au microscope, non cultivable, non filtrable) par opposition aux bactéries et permit de fonder un nouveau domaine de recherche, la Virologie.

Pendant près d'un demi-siècle, les virus ont été perçus comme des structures extrêmement simples dont l’ensemble des éléments protègent quelques « petits bouts » de matériel génétique, et dont la seule raison d'être, est de s’infiltrer dans une cellule pour la parasiter [59]. Les virus ont été et sont encore parfois considérés comme de simples véhicules de gènes ayant joué un rôle mineur dans l'évolution. Seulement ce paradigme a quelque peu été ébranlé, depuis la découverte en 2003 de Mimivirus ; un virus géant [60,61]. La découverte de Mimivirus a démontré l’existence d'une nouvelle famille de virus : les « géants », dont les caractéristiques hors normes ont remis en cause une partie du cloisonnement viral du reste du monde du vivant. Les Mimivirus ont aussi ouvert la voie pour la découverte de nouveaux virus géants (Amphoravirus et Pandoravirus) dont l'analyse, qui ne fait que commencer, laisse entrevoir des perspectives encore mal définies concernant l'évolution et le rôle des virus [62,63]. Ces virus géants sont à l'origine d'une nouvelle façon de percevoir les domaines de la vie et de la virologie. Ils ont fondamentalement bouleversé le champ des connaissances en biologie. En effet, la découverte de Mimivirus, le premier des « géants » ; est un virus en cela qu'il est un parasite intracellulaire, incapable de division binaire et ne possédant pas de système susceptible de subvenir à ses besoins énergétiques. Les Mimivirus ont remis en cause la définition des virus de plusieurs points de vue. D'une part, sa taille (>0.7µm) qui dépasse celle que l'on attendait pour un virus (<0.25µm). D'autre part, sa complexité génomique (plus de 1000 gènes dont 50 % sont des gènes orphelins) [64]. Enfin, la remise en cause de l'unicité du type d'acide nucléique chez les virus (Mimivirus possède de l'ARN et de l'ADN). Ce dernier point avait été déjà démontré en 1998 chez les Cytomegalovirus indiquant que le dogme n'était pas absolu [65].

Ainsi, la remise en cause de trois des cinq critères qui font d'un virus un virus n'est pas encore de nature à exclure les virus géants du monde des virus. La taille des virus et leur cycle de vie sont extraordinairement variables. Le plus petit d'entre eux, l'agent Delta qui ne possède qu'un seul gène, ne peut se multiplier que lors de l'infection d'une cellule eucaryote par le virus de l'hépatite B. L'agent Delta agit alors comme un parasite de virus. A l'inverse, les virus à ADN comportent plusieurs centaines (voire milliers) de gènes (dans le cas des virus géants). Il existe ainsi une variation importante de la taille du génome des virus, qui corrèle assez naturellement à la taille et la structure des virions. Depuis la victoire du virus comme concept rationnel sur « l'observable », le virologiste a appris à se méfier et à remettre en cause les cadres « certains » qui définissent le monde viral. Nous sommes tenus de nous ré-interroger sur la définition des virus, de la vie et son origine. Ainsi, la comparaison des virus avec les cellules, montre que l'un des constituant protéique majeur marqueur du vivant est la présence du ribosome, de ses protéines et ARN spécialisés. Ce dernier critère éclairé de la biologie moléculaire semble être un marqueur plus pertinent que celui de la taille ou la complexité du génome pour définir ce qui est ou non viral. L'une des conséquences de ce raisonnement permet de s'affranchir de la question de définir si les virus sont des « êtres vivants » ou non. En fait dans la comparaison mentionnée plus haut il s'agit de la comparaison des virions et des cellules vivantes. La définition du virus est intiment lié à son virion. Ors si le virion est un « véhicule » pour le virus, il en est aussi la caractéristique, c'est à dire qu'un virus tout aussi complexe qu'il soit, se caractérise par le fait d'avoir une information spécifique pour la formation d'un virion (une capside). Pour autant, le virus ne peut être limité à son virion, la partie du cycle du virus qui implique la cellule, fait aussi partie du virus, ce qui a amené Patrick Forterre à proposer le concept de cellule virale, qui correspond à « l'usine à virions » [22].

Une autre conséquence de la découverte des virus géants a été la prise de conscience du nombre de gènes dits orphelins retrouvés chez les virus. Les données de métagénomiques (qui vont au delà des données sur les virus géants) ont montré un nombre de gènes viraux inconnus bien supérieur à celui relatif aux gènes cellulaires. En ce sens, les virus géants ont été un révélateur de notre ignorance plus qu'un marqueur de divergence. Ces gènes orphelins sont autrement dit, des gènes qui n'ont aucun homologue, et théoriquement qui codent des protéines de fonction et de structure inconnues. Ces gènes ont une histoire détachée du monde vivant étudié et suivent leur propre évolution. Du fait du surclassement du nombre de gènes viraux inconnus, on peut aussi considérer que les virus constituent un formidable réservoir de gènes prêts à être injectés dans le cycle de la vie. De l’étude des virus et de leur hôte, de leur génome, et de leur protéome, il semble bien que les virus aient joués (et si c'est vrai, jouent toujours) un rôle décisif dans l'évolution. L'identification de rétrovirus endogènes dans les génomes de tous les eucaryotes étudiés, va dans le sens qu'au moins un certain nombre de virus participent à la dissémination du matériel génétique. L'intégration de gènes viraux dans le génome des cellules eucaryotes génère parfois des modifications si drastiques, qu'elles contraignent la cellule à diverger de façon irréversible. Ces différentiations peuvent être à l'origine de pathologies tout aussi bien que source d'adaptation et d'évolution. C'est par la séquence et la structure des protéines virales que l'on peut décrire leur filiation et leur évolution.

Qu'il soit grand ou petit, il existe globalement une grande division parmi les virus : ceux dont le virion contient un génome composé d'ADN ou d'ARN. En fonction du type de matériel génétique caractérisant le virus, les stratégies de réplication virales spécifiques vont conditionner les étapes et la complexité du cycle viral ainsi que les assemblages structuraux nécessaires pour les accomplir.

i. Les stratégies de réplication virale

La réplication du matériel génétique est la singularité des êtres vivants et pourtant nul part dans la biosphère ce processus est accompli avec autant d'économie et de simplicité que dans le monde viral. En effet, pour arriver à l'expression, la réplication et la diffusion de leurs gènes, les différentes familles virales ont mis en place des stratégies et des cycles de vie qui vont exploiter la biologie et la biochimie de leur hôte. Comme nous l'avons vu précédemment, les virus ne possède aucun système de production d’énergie, ils ne peuvent se reproduire qu’au sein de cellules vivantes. A l’aide de quelques gènes, ils altèrent et modifient les programmes de fonctions intracellulaires afin d'utiliser la machinerie cellulaire à leur profit pour se répliquer et assurer leur pérennité. Ce sont des parasites obligatoires, et la conséquence la plus importante est la dépendance de leur cycle de vie aux conditions proposées (imposées) par leur hôte. Ainsi comprendre la réplication virale et tous les processus associés, nous permet de comprendre la biologie de l'hôte à tous les niveaux physiologiques (moléculaires, cellulaires, tissulaires, organes, organismes) et sociaux (des individus et de leur population). Chaque infection, représente la mise en contact de matériel génétique exogène à la machinerie de l'hôte et dans certain cas, à son intégration dans le matériel génétique de l'hôte. La persistance de ce mécanisme reflète l'état de la co-évolution entre le virus et l'hôte. Chaque pièce du puzzle que représente la compréhension du mécanisme de réplication virale, nous enseigne sur la biologie de son hôte tout autant que sur celui du virus. L'étude des mécanismes et enzymes virales impliquées dans la réplication virale nous permet d'identifier des points de « faiblesse » du virus. Par exemple, des motifs ou des signatures structurales conservées qui permettent de s'engager dans le design rationnel de molécules antivirales spécifiques.

a) La diversité des génomes viraux et leurs stratégies de réplication

L'un des aspects des plus surprenants lorsque l'on regarde la virosphère, c'est la grande diversité structurale des génomes viraux et leurs stratégies de réplication [66]. Contrairement aux cellules qui possèdent leur génome exclusivement sous la forme d'ADN double brin, les génomes viraux peuvent être composé d'ADN ou d'ARN simple ou double brin, de polarité positive, négative ou les deux (ambisens). Ces génomes peuvent être linéaires ou circulaires, segmentés ou non. Chaque variation impose des stratégies différentes pour le cycle de réplication virale (machinerie réplicative virale, du virion, ou cellulaire) et sa localisation (nucléaire, cytoplasmique, ou les deux).

La plupart des virus à ADN vont se répliquer dans le noyau des cellules infectées, là où se trouve la réplication et la transcription cellulaire. Il y a quelques exceptions à cette règle chez les pox-, irido- et asfivirus pour lesquelles la réplication a lieu tout ou parti dans le cytoplasme. Les virus à ARN ont un cycle réplicatif cytoplasmique à l'exception des rétrovirus (qui intègrent une copie ADN de leur génome dans le génome cellulaire) et des Orhtomyxo- et Bornavirus (des virus à ARN simple brin négatif qui se répliquent dans le noyau).

b) Différentes stratégies de production des complexes ou protéines virales

Pour les virus à ARN, les génomes viraux sont souvent composés d'un seul brin d'ARN, qui souvent code pour plusieurs protéines virales. Paradoxalement, les ARN polymérases ARN-dépendantes sont rarement capable d'accéder des promoteurs internes du génome. Ceci devient un obstacle pour la production de protéines virales, qui doivent être traduites à partir d'ARNm viraux compatibles avec le mode de fonctionnement du ribosome. Pour résoudre ce problème, l'évolution a abouti à trois stratégies : (1) <u>segmentation protéique</u> : la synthèse d'une polyprotéine qui est ensuite découpée par des protéases cellulaire ou virale (*i.e.* picorna-, toga-, flavi-, retrovirus) ; (2) <u>segmentation des ARNm</u> : la production d'ARNm viraux monocistroniques à partir de la même matrice (*i.e*. corona-, arteri-, rhabdo-, paramyxovirus) ; ou enfin (3) <u>segmentation du génome</u> : le génome est segmenté et parfois ambisens. Chaque segment code en général pour un gène ou deux si le segment est ambisens (*i.e*. reo-, orthomyxo-, bunya-, arenavirus).

Pour les virus à ADN, ces stratégies sont plus rarement utilisées (*i.e.* Asfarviridae : deux polyprotéines) car la transcription est prise en charge par la machinerie transcriptionnelle du noyau, elle est donc capable de synthétiser des ARNm à partir de promoteur interne.

ii. Les virus à ARN

Les différences entre les virus à ARN de polarité positive (+) et négative (-) vont bien au delà de la polarité du génome encapsidé dans le virion. En effet, les virus à ARN (+) relarguent leur génome directement dans le cytoplasme. Là, ils sont pris en charge par les protéines du ribosome afin de commencer la synthèse des protéines virales qui vont former le complexe de réplication/ transcription [67]. C'est ce dernier qui va prendre en charge la multiplication virale. Par comparaison, les virus à ARN (-), embarquent en plus de leur matériel génétique, leur machinerie replicative et transcriptionelle [68]. Ainsi, le génome se retrouve complexé en permanence à des protéines virales, dont la fonction est d'autant de protéger le matériel génétique du virus que d'en assurer la transcription et la réplication. L'ensemble de ce complexe ribonucléoprotéique va bien au delà du concept d'empaquetage dans le virion, il est un des moteurs actifs du cycle viral.

Ces adaptations fondamentalement différentes, sont probablement le fruit de contraintes différentes. Les virus à ARN (+) sont contraints sur un critère d'efficacité de la traduction de leurs génome, ce qui implique que les génomes structurellement optimisés pour la traduction doivent être conservés d'une génération à l'autre. Ces contraintes de compatibilités optimisées sur le génome sont dictées par [la dépendance à] la cellule hôte. En revanche, la contrainte sur le génome pour les virus à ARN négatif n'est pas imposée par l'hôte; car ceux-ci ne sont jamais directement traduit. Cependant, le génome des virus à ARN (-) doit rester optimisé pour sa ribonucléoprotéine et en particulier vis-à-vis de sa nucléoprotéine pour son encapsidation et son ARN polymérase ARN-dépendante pour la transcription et la réplication. De ce point de vu, les virus à ARN (-), en ajoutant un niveau intermédiaire de complexité, augmentent leur chance de survie et de dissémination en se donnant la possibilité de faire évoluer leur ARN-polymérase afin que celle-ci puisse répondre à une modification du système de traduction de l'hôte.

Les ARN polymérases ARN dépendantes sont naturellement peu fidèles, ce qui a pour avantage de faciliter l'adaptation virale aux changements de la cellule. Les ARN polymérases virales présentent des taux d'erreurs 10 000 fois plus élevés que leurs homologues cellulaires à ADN, soit en moyenne une mutation par génome et par tour de réplication. A cela, s'ajoute l'absence de mécanisme de validation et de correction d'erreurs (proof reading). Il faut donc considérer que lors de l'infection virale, les virus produits dérivent tous du virus original, et constituent une population dont les séquences génomiques varient autour d'une séquence consensus. Le risque de passer un niveau d'erreurs critiques est possible, mais les virus déviant, s'ils sont produits, ne seront simplement pas sélectionnés lors de la prochaine infection. Pour contrebalancer le risque d'atteindre le taux d'erreurs critique trop rapidement, la taille des génomes des virus à ARN reste limitée. En effet, la plupart des virus à ARN ont une taille compris entre environ 5 et 15kb avec un maximum de 30kb. Les recherches récentes sur les virus appartenant à l'ordre des *Nidovirales*, renforcent ce paradigme tout en introduisant une description plus complète de la contrainte. En effet l'ordre des *Nidovirales* inclus des virus dont la taille des génomes varient entre 12kb à 32kb. Bien que la structure et l'organisation générale du génome soit conservée, les génomes des Coronavirus qui dépassent la taille de 20kb codent en plus une 3'-5' exonucléase, homologue aux exonucléases à ADN impliquées dans l'activité de validation et de correction des erreurs lors de la duplication des chromosomes [69–71]. Il est encore trop tôt pour affirmer la présence d'activité de correction d'erreurs chez les virus à ARN mais les résultats préliminaires montrent que l'inactivation de cette enzyme entraîne une augmentation drastique du taux d'erreur dans la réplication du génome des Coronavirus. Si elle est confirmée, cette découverte est fondamentale en cela qu'elle démontre la contrainte physique imposée sur le génome pour le contrôle de sa taille et montre que l'expansion de la taille d'un génome viral lorsqu'elle se produit, s'accompagne d'un système de contrôle et de validation du génome. Passé une certaine taille, il est possible que l'ARN ne soit plus la molécule de choix pour stabiliser l'information génétique. Dès lors que le support devient de l'ADN, l'économie et l'optimisation du génome n'est plus une contrainte aussi drastique. Ainsi, l'analyse des génomes viraux et de leur machinerie réplicative si variée, nous montre l'aspect fondamental qu'il y a à étudier les protéines des machineries virales et surtout leur assemblage afin de pouvoir saisir les mécanismes mis en place au cours de l'évolution.

III. TRAVAUX DE RECHERCHES

A. Protéines d'assemblage et de régulation du cytosquelette

Boston Biomedical Research Institute, Watertown, MA, USA.
University of Pennsylvania, Philadelphia, PA, USA.
Laboratoire de R. Dominguez (2005-2008)

i. Contexte scientifique

Les cellules possèdent un squelette qui à la fois maintient la forme de la cellule et peut s'adapter et se modifier de façon quasiment instantanée. Cette plasticité est responsable de processus vitaux tels que la morphogenèse embryonnaire, la surveillance immunitaire, angiogenèse, la réparation, la régénération des tissus ou la mise en place des voies de transport intracellulaire des vésicules [72–76]. Ce squelette est composé de filaments d'actine appelés microfilaments [77]. La principale caractéristique de ce squelette est sa capacité dynamique à passer d'une forme monomérique (l'actine-G) à une forme filamenteuse (l'actine-F).

L'actine :

L'actine est l'une des protéines des plus abondante chez les eucaryotes. Très conservée, sa structure s'organise autour de deux domaines se composant chacun de deux sous-domaines (Figure 3A). Les sous-domaines 1 et 2 forment le domaine extérieur (qui sont à la surface du filament), et les sous-domaines 3 et 4 forment le domaine intérieur (domaines qui se font face dans la structure en double hélice que forme le filament) (Figure 3B) [78–80]. L'actine possède une activité ATPase dont le site actif se situe entre les sous-domaines 2 et 4. Son activité peut être régulée par l'occupation de la poche hydrophobe formée par les sous-domaines 3 et 4. L'hydrolyse de l'ATP produit un changement de conformation structurant une boucle désordonnée (D-loop ou DNAse-I loop) en hélice (Figure 1A). Cette hélice joue un rôle structurant dans les interactions actine-actine dans le filament [81,82]. L'hydrolyse de l'ATP est donc un processus essentiel qui est concomitant avec le changement d'état de l'actine G à F.

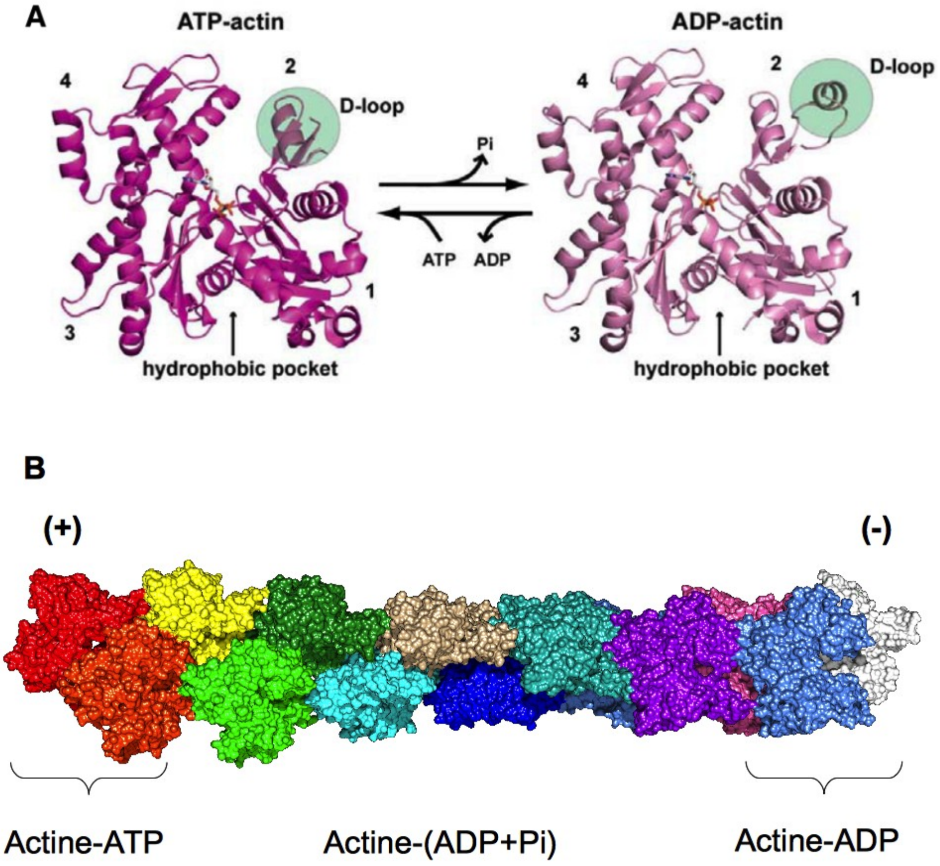


Figure 3: Actine monomérique et filamenteuse. (A) Le monomère d'actine s'organise autour de deux domaines structurellement liés, subdivisés en deux sous-domaines 1-2 (formant le domaine externe), et 3-4 (formant le domaine interne). La combinaison de ces quatre domaines génère deux sillons profonds qui jouent un rôle fonctionnel fondamental: le sillon formé par les sous-domaines 2-4 forment le site de liaison à l'ATP et le sillon formé par les sous-domaines 1-3 forment le site de recrutement/interaction par/avec les protéines régulant la polymérisation de l'actine (PLAs). L'actine est une ATPase, lors de l'hydrolyse de l'ATP en ADP l'actine change de conformation (boucle D dans le cercle vert) et modifie son affinité pour les PLAs. (B) Modèle du filament d'actine (PDB: 3B63). Le filament d'actine est dit polarisé, il possède un coté croissant (+) enrichi en ATP-actine, un coté en décroissance (-) enrichi en ADP-actine et entre les deux il s'agit d'une actine qui a effectué l'hydrolyse de l'ATP et changé de conformation mais qui séquestre encore le pyrophosphate.

Illustration dérivée de [141].

Le filament d'actine :

Le filament d'actine peut se décrire comme deux fils organisés en double hélice droite, composés de monomères d'actine empilés. Le filament est polarisé en ce sens que le coté croissant (+) incorpore de l'actine monomérique liée à une molécule d'ATP, alors que le coté opposé (-) décroit en relâchant de l'actine liée à une molécule d'ADP (Figure 3B) [83]. Pour éviter la décroissance anarchique, le coté (-) est souvent « coiffé » par des protéines qui régulent la structure du filament. Entre les deux extrémités du filament l'actine hydrolyse l'ATP ce qui permet la structuration de la boucle D (Figure 3B) et la stabilisation du filament [79,84–87]. Ce changement d'état énergétique est strictement contrôlé dans les cellules par un ensemble de protéines se liant spécifiquement à l'actine-G ou F (PLA ou plus connus sous le terme anglophone ABPs). Ces protéines sont sous le contrôle de la famille des Rho GTPases [88,89] (voir paragraphe dessous). Les PLAs contrôlent et régulent les différents états du filament d'actine : nucléation, élongation, ramification, réticulation, découpage, plafonnement (coiffe). Certaines sont impliquées dans la séquestration de l'actine monomérique ou sa localisation cellulaire (Table 1 et Figure 4) [88,90,91].

Ce système polarisé et contrôlé permet d'assurer une croissance orientée du filament. C'est grâce à ce système que la cellule peut déformer sa membrane et/ou provoquer une poussée permettant le mouvement en réponse à des signaux spécifiques. (Figure 4)

Contrôle des PLAs par les GTPases :

L'activité des PLAs est sous le contrôle d'une cascade de signalisation. Cette cascade est activée au niveau des récepteurs membranaires et va *in fine* stimuler des GTPases appartenant en particulier à la famille des Rho GTPases (Rho, Rac, CDC42) [92,93]. Ces activateurs vont ensuite répartir et répercuter l'information sur les PLAs déclenchant : les mouvements par la voie acto-myosin [94], la restructuration de la membrane en lamelipode [95] ou en filopode [96]. Le rôle central de cette famille de GTPases a été montré en exprimant des mutants négatifs dans des cellules de mammifères. Ces mutants impactent sévèrement les mouvements et les processus cellulaires dépendant de l'actine tels que : la migration [97], la cytokinèse (ou cytocinèse) [98,99], l'endo- et l'exocytose [100], le guidage axonal [101] et la morphogenèse lors du développement [102,103]. La localisation des GTPases (à la membrane ou autre) semble induire le recrutement spécifique des PLAs. Les PLAs possèdent donc au moins un domaine de recrutement ou d'interaction avec les GTPases. Il a en effet été montré qu'un grand nombre de PLAs (WASP, Formins, IRSp53) [104–106] interagissent directement avec CDC42 et Rac au niveau de motifs conservés [104]

Table 1 : liste des protéines se liant avec l'actine et leur fonction correspondante

Protéine	Fonction cellulaire	Référence
α-actinine	Maintient les filaments liés deux à deux parallèles	[131]
Debrine	Maintient les filaments liés entre eux	[132,133]
Fimbrine	Maintient les filaments liés entre eux	[107]
Fascine	Réticulation (cross-link) des filaments des filopodes	[134]
Tropomyosine	Stabilisation des filaments	[135]
Spectrine	Accroche les filaments à la membrane plasmique	[136,137]
Talin	Adaptateurs aux protéines membranaires	[107]
Vinculin	Adaptateurs aux protéines membranaires	[107]
BAR protéine (MIM/IRSP53/Amphiphysine)	Stabilisation des filaments	[108] *
Cortactin	Nucléation de branche et stabilisation des filaments	[109]
Formine (mDIA)	Nucléation et élongation	[110,111]
CoBL / Spire / VoPL/VopF / Lmod	Nucléation	[112–116]
ARP2/3	Nucléation de branche	[117,118]
WASP	Elongation	[119]
ENA/VASP	Elongation	[120,121]
CAPZ	Coiffe (+)	[122]
EPS8	Coiffe (+)	[123]
Gelsoline	Coiffe (+) et fragmentation des filaments	[124,125]
Tropomoduline	Coiffe (-)	[126]
ADF/Cofiline	Dépolymérisation	[127,128]
Thymosine β4	Séquestration	[129]
Profiline	Séquestration et régénération du pôle d'actine-ATP	[130]

Note : * is a review

Structure modulaire des PLAs et mécanismes de contrôle de l'actine :

Les PLAs sont des protéines à domaines multiples contenant un (parfois plusieurs) domaine de liaison à l'actine monomérique (W, GAB ou WH2), parfois un domaine de liaison à l'actine filamenteuse (FAB), un domaine de liaison protéine-protéine ou d'arrimage membranaire, un domaine d'activation et ou permettant ou non l'auto-inhibition, et un domaine de liaison aux GTPases. Ce dernier domaine induit l'activation et le recrutements de la protéine à des endroits précis de la cellule. Une fois la PLA recrutée et activée, seulement une petite parti de la protéine est impliquée dans le contrôle de la polymérisation de l'actine. En effet le mode de régulation des PLAs sur l'actine s'exerce par un contrôle structural direct.

L'espace formé par les sous-domaines 1 et 3 de l'actine forme une poche hydrophobe, qui si elle est occupée empêche l'hydrolyse de l'ATP (Figure 3A). Ce domaine est un site de liaison privilégié des PLAs régulant l'activité ATPase et incidemment la croissance du filament (Table 1). Un grand nombre de PLAs impliquées dans le contrôle et la séquestration de l'actine non polymérisée, possèdent par conséquent une hélice-α amphiphile dont la fonction est de recruter l'actine par sa poche hydrophobe [138,139]. Cette hélice est annotée comme le domaine WH2 (ou W ou GAB). Malgré son rôle si important, cette hélice présente une variabilité de taille. Elle se compose d'une extrémité amino terminale hydrophobe suivi d'une signature conservée (LKKT(V)) prolongée d'une région hydrophile plus ou moins structurée. Cette dernière peut se fixer le long de l'actine et remonter vers le site actif entre les domaines 2 et 4 [140]. Ce système de contrôle est retrouvé dans de nombreuses structures de PLAs résolues en complexe avec l'actine (revue de Lee et Dominguez [141]

Malgré une faible similitude de la séquence primaire, on observe la conservation structurale. Il y a par conséquent probablement une évolution convergente des systèmes de régulation de l'actine, et un nombre limité de repliements possible des PLAs : Cofilin [142], Wiskott-Aldrich syndrome protein (qui à donné le nom au domaine (WASP)-homology domain 2 (WH2)) [143], gelsolin-homology domain [144], domaine homologue calponine (CH domain) [145], le domaine 2 homolgue de la formine (FH) [146]. A ceux-ci il faut rajouter le groupe particulier de PLAs qui appartiennent à la superfamille des myosines [147], qui utilisent les filaments d'actine comme des rails pour assurer la motilité cellulaire.

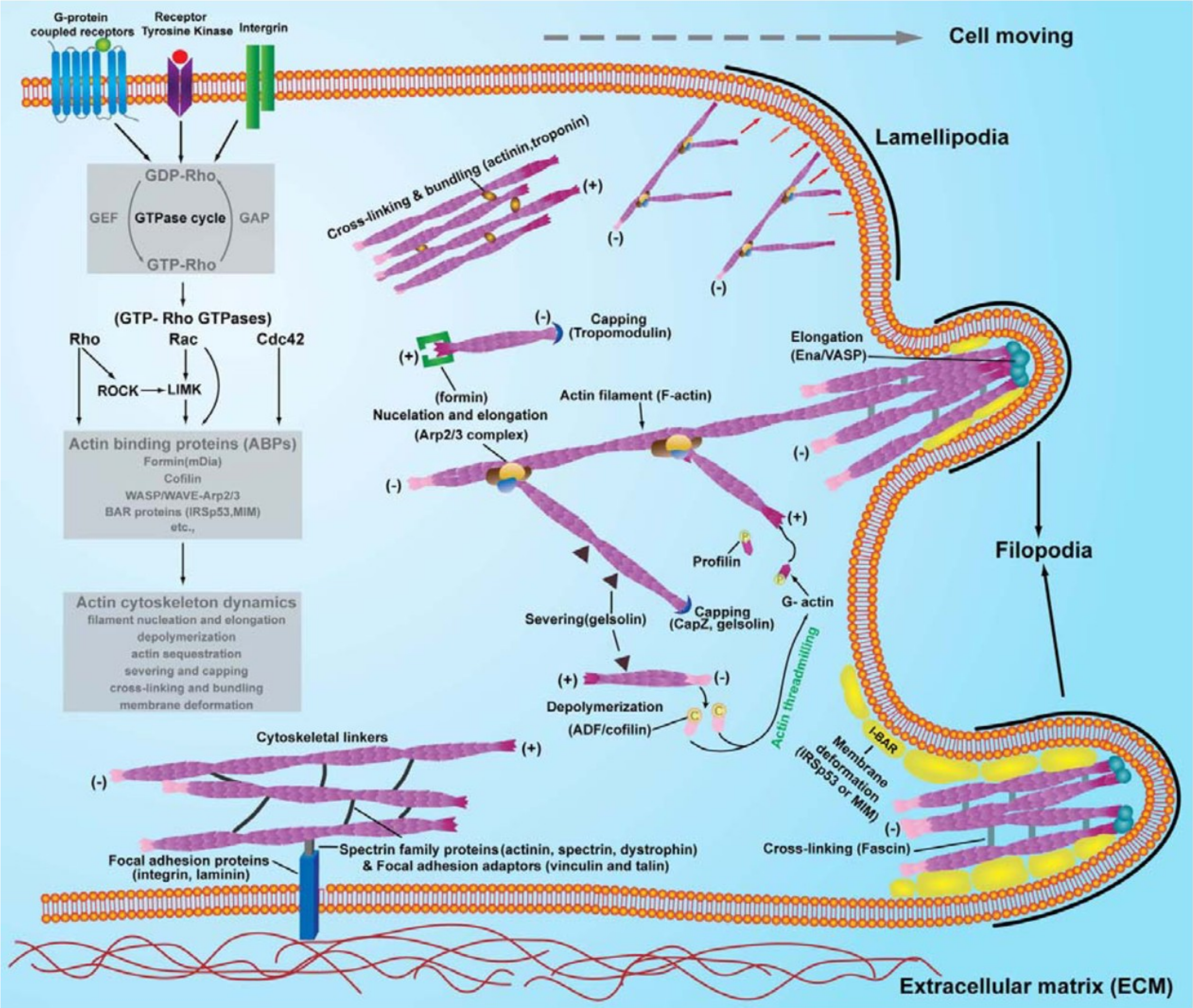


Figure 4: Multiples systèmes de régulation des réseaux de filament d'actine dans la cellule. La famille des Rho GTPases (Rho, Rac, and Cdc42) est la principale voie de régulation de l'actine dans la cellule. Des récepteurs au niveau de la membrane de la cellule activent les protéines de cette famille, engendrant une série de cascades d'hydrolyse et de phosphorylation, interagissant avec les PLAs et déclenchant selon la position dans la cellule des réarrangement du réseau d'actine. Les PLAs peuvent interagir spécifiquement avec de l'actine monomérique (séquestration, dépolymérisation) ou de l'actine filamenteuse (coiffe du filament, réticulation), ou encore les deux (branchage , élongation). Ces régulations peuvent mener à différent type de réseaux (câble de filament ou réseaux dense de soutient) et de phénotypes membranaires (filopodes, lamélipodes, adhésion).

Illustration dérivée de [141].

Le contrôle de la croissance du filament d'actine nécessite la participation d'un autre partenaire protéique : la profiline dont l'un des rôle est d'assurer l'apport exclusif d'actine-ATP aux PLAs. La profiline est une PLA un peu particulière en cela que son interaction avec l'actine ne se fait pas via une hélice mais par une surface d'interaction sous les domaines 1 et 3. De plus, cette interaction n'est pas exclusive, elle permet dans certain cas une liaison simultanée à un domaine WH2 [148]. La profiline a une forte affinité pour l'actine monomérique, et elle joue un rôle capitale dans la régénération du pôle actine-ATP. En effet, du coté (-) du filament, la profiline séquestre l'actine-ADP relarguée et stimule le remplacement de la molécule d'ADP en ATP (1000 fois plus rapide que l'échange naturelle) [149,150]. L'actine-ATP utilisable est donc sous forme de complexe protéique profiline-actine également appelé profilactine. L'autre caractéristique de la profiline est d'avoir une forte affinité pour les régions poly-L-proline que l'on retrouve sur de nombreux PLAs comme les formins, ENA/VASP, WASP....

La thématique du laboratoire s'organise autour de la description et de la compréhension des événements moléculaires régulant l'interaction entre les PLAs et l'actine. En particulier, la description des protéines séquestrant l'actine monomérique ; des mécanismes moléculaires du système WASP/Arp2/3 qui engendre un réseau arborescent de filaments "branchés" (Nucléation de la branche et élongation) ; de l'α-actinine qui réticule le réseaux de filament d'actine, d'ENA/VASP qui permet l'élongation rapide de microfilaments pour former rapidement des filopodes ; et enfin aux protéines BAR qui sont impliquées dans la déformation de la membrane et la consolidation des filaments ainsi formés.

Au cours de mes stages post-doctoraux, je me suis intéressé aux domaines d'interaction de l'actine de l'α-actinine et aux mécanismes de recrutement de l'actine de deux PLAs particulier : celui d'ENA/VASP responsable de la croissance des filaments rapides lors de la formation de filopodes et à MIM la protéine impliquée dans la consolidation de ces structures.

ii. Objectif du travail et stratégies expérimentales

Les objectifs de ces travaux étaient d'identifier les mécanismes de formation des filopodes. Les filopodes sont des déformations membranaires longilignes, formées par l'extension rapide de filaments d'actine. La formation d'un filopode implique d'une part le recrutement et l'assemblage rapide d'actine par la protéine ENA/VASP [151]; et d'autre part la consolidation de cette structure par la protéine MIM. Cette dernière participe à la restructuration de la membrane lors de sa déformation suite à l'élongation rapide des filaments d'actine [152].

La protéine ENA/VASP était identifiée comme une des protéines responsable de l'allongement des filopodes. Sa modularité postulait l'existance de trois domaines : EVH1, une région centrale désordonnée et EVH2. EVH1 se lie aux protéines de la membrane et sert d'ancrage. La région désordonnée sert de site de régulation pour les GTPases et de recrutement des profilactines. Enfin EVH2 semble interagir avec le filament d'actine et est responsable de la tétramérisation. Si la modularité de la protéine était défini en revanche la compréhension des mécanismes moléculaires de recrutement de l'actine et de son incorporation au filament restaient à être décrit. En particulier :

-Identifier la région de recrutement des profilactines.

-Définir les domaines impliqués dans la liaison à l'actine.

-Identifier si des complexes intermédiaires se formaient.

MIM est une protéine modulaire qui recrute l'actine via deux domaines N et C terminaux. Le domaine N-terminal est un domaine de 250 résidus alors que celui retrouvé en c-terminal un domaine de 30 résidus de type WH2. Les 475 résidus intermédiaires forment un domaine partiellement désordonné, riche en résidus Proline, Serine et Thréonine. Ce domaine intermédiaire semble être impliqué dans le recrutement et la régulation de l'activité de la protéine. Le mécanisme d'action de MIM restait à être défini en particulier :

-Comprendre le rôle de la modularité de la protéine MIM.

-Identifier la fonction exacte de MIM dans son rôle de soutien et son mécanisme d'interaction avec l'actine et la membrane plasmique.

iii. Résultats obtenus et perspectives

- *Rôle d'ENA/VASP dans l'élongation du filament d'actine*

ENA/VASP fait parti de la famille de protéines responsable de l'élongation rapide du filament d'actine. Ce processus implique un domaine de liaison à l'actine G (WH2) et un domaine de liaison à l'actine F. Il s'agit d'une protéine à plusieurs domaines, qui tétramérise via son domaine C-terminal constitué par une hélice super enroulée (Coiled-coil). La tétramérisation permet probablement d'assurer l'incorporation efficace de monomères d'actine sur plusieurs filaments regroupés le long de la membrane plasmique. La structure globale de la protéine n'a pas été résolue à cause de la présence de longues régions désordonnées. En revanche, la plus part des structures des différents domaines de régulation ou de multimérisation sont maintenant connus.

Dans un premier temps, nous avons procédé à l'analyse bio-informatique de la protéine, ce qui nous a permis de définir la modularité des différents domaines. Il était notamment nécessaire d'étudier les régions désordonnées (>100aa consécutifs), qui étaient présentées comme le site de recrutement de plusieurs partenaires et dont les caractéristiques se retrouvent dans le domaine EVH2. En appliquant une combinaison d'analyses bio-informatiques [153] nous avons proposé que la région désordonnée présentait trois motifs distincts que nous avons proposé comme sites : de régulation, de recrutement et de charge. La position de ces sites correspondaient au phénotype de délétion décrit dans les études biochimiques (*i.e.* perte du recrutement de la GTPase lors de la délétion d'une région encadrant le site de régulation, perte de du recrutement de profilactine lors de la délétion des autres sites)[154]. L'analyse a également montré que le domaine EVH2 était aussi en parti désordonné. Néanmoins la nature du désordre est différente en raison du nombre d'acides aminés hydrophobes plus élevés. Par ces analyses de composition en acides aminés, nous avons pu positionner et identifier deux courtes régions capables de se structurer. La prédiction de repliement induit se superposait à la composition en acides aminés caractéristique des domaines de recrutement de l'actine G (GAB ou WH2) et F (FAB). Entre le domaine FAB et le Coiled-coil de tétramérisation se trouve une région qui a le potentiel de se structurer en longue hélices. Ainsi nous avons pu décrire l'architecture de cette protéine de la façon suivante : Un domaine N-terminal (EVH1), dont le rôle est l'adressage de la protéine au complexe ancré à la membrane. Une région centrale désordonnée riche en prolines, dont on distingue 3 sites de recrutement, un pour la GTPase qui régule l'activité de la protéine, et deux sites de poly-prolines qui se lient à la profilactine. Enfin, le domaine C-terminal (EVH2) est constitué de 4 sous-domaines dont deux domaines de liaison à l'actine (G et F), un domaine moteur et un domaine de multimerisation.

L'observation de la présence de nombreux sites de poly-prolines dans les protéines de régulation du cytosquelette et en particulier dans celles impliquées dans la nucléation et l'élongation de l'actine, suggérait que leur rôle est de contribuer à l'augmentation de la concentration locale d'actine par le recrutement des profilactines. Nous avons suivi cette hypothèse en obtenant un complexe cristallographique profilactine poly-prolines. Le complexe obtenu avec une ratio molaire ±1/1 demontre que les domaines de poly-prolines peuvent recruter les profilactines par la formation de complexes stables (Figure 3). Nous avons testé les différentes séquences de poly-prolines (suite de prolines, site de recrutement et de charge) et vérifié qu'elles recrutaient toutes les profilactines. Le site de proline le plus proche du site de recrutement de l'actine (site de charge) présente une insertion d'une leucine conservée qui impose le positionnement et l'orientation du complexe profilactine sur la séquence. Nous avons postulé que cette contrainte a pour but de positionner le complexe de profilactine à bonne distance du domaine WH2 afin de transférer l'actine de la profiline sur le domaine WH2 et ensuite sur le filament. Fort de notre prédiction de domaines, nous avons donc synthétisé un peptide incluant le domaine de charge, la région de liaison et le domaine WH2 afin de vérifier cette hypothèse. Nous avons pu obtenir un complexe stable et fait la preuve que le transfert implique un intermédiaire poly-proline profilactine WH2 avant le passage sur le filament (Figure 5). En conclusion, les données structurales et biochimiques suggèrent que l'élongation du filament par ENA/ VASP est processive. Les données structurales renforcent le modèle qui postule que les complexes de profilactines constituent la forme d'actine polarisable la plus abondante à proximité du PLA, et que ceux-ci possèdent un système de recrutement de ces complexes efficace. Une fois l'actine positionnée sur le domaine WH2, l'affinité pour la profiline décroit et permet la dissociation. Dès lors, la proche proximité du filament permet le transfert du monomère d'actine et son incorporation. ENA/VASP glisse alors vers la nouvelle extrémité permettant à un nouveau cycle d'incorporation de prendre place. Enfin la tétramérisation permet également de répondre aussi à la polymérisation

synchronisée de plusieurs filaments. (Figure 5)

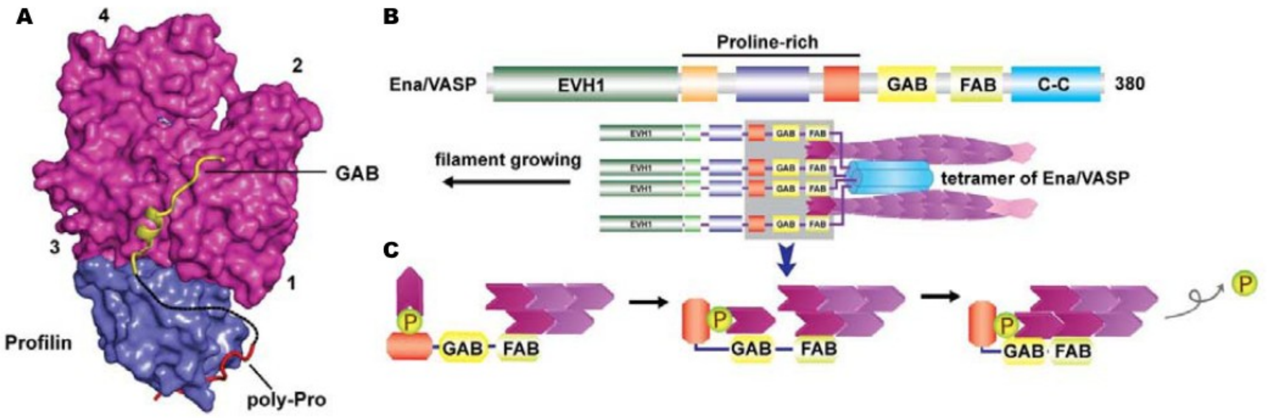


Figure 5: La protéine Ena/VASP est responsable de l'élongation du filament. (A) Structure du complexe profilactine et du domaine de recrutement de VASP (pdb: 2pbd).(B)Schéma de l'organisation modulaire de VASP et de sa probable organisation quaternaire. VASP a une organisation modulaire composé d'un domaine de fixation aux protéines d'encrage de la membrane (WH1), suivit d' une région désordonnées (poly-proline), formant une zone d'activation par les GTPases et de recrutement de profilactin, suivit de deux sites de fixation de l'actine monomérique (GAB) et filamenteuse (FAB), et enfin un domaine de tetramérisation. (C) Modèle mécanistique d'incorporation de l'actine (profilactine) sur le filament par VASP. 1 recrutement 2 charge de l'actine 3 incorporation de l'actine et libération de la profiline.

Illustration dérivée de [141].

Ce modèle reste incomplet, il ne permet pas de comprendre l'ensemble des processus de formation d'un filopode. Notamment, le phénomène intermédiaire de dépolymérisation observé juste après une croissance rapide du filament.

Cependant, les résultats issus de ce travail ont permis de proposer un mécanisme d'élongation efficace impliquant une machinerie protéique dont la dynamique nécessite l'intervention des régions désordonnées et structurées. Ces résultats montrent l'importance de ne pas négliger le rôle des régions désordonnées dans la dynamique de formation de complexes. De plus, la caractérisation des domaines désordonnés riche en poly-prolines comme plate-forme de recrutement pour la profilactine ouvre de nouvelles pistes de recherche concernant le recrutement des monomères d'actine par les autres PLAs. En particulier, ceux impliqués dans la nucléation et le branchage du filament.

- *Rôle de la protéine MIM dans le soutien au filament lors de la création des filopodes*

La déformation de la membrane plasmique par le cytosquelette lors de processus tels que l'endo- ou l'exocytose, la création de filopodes, les déplacements, ou les mouvements intracellulaires [155–157]; nécessite une coopération et une coordination entre la polymérisation de l'actine et les protéines structurant la membrane. Les protéines BAR sont des protéines impliquées lors de ces modifications, à la fois pour consolider les structures du filament nouvellement crées et pour aider la membrane à accepter les forces imposées localement [158]. Afin d'assurer la coordination spatiale et temporelle, ces protéines sont sous le contrôle des mêmes familles de GTPase qui contrôlent les PLAs [157–162].

Les protéines BAR ont une structure modulaire qui leurs permet de se lier au filament ou au monomère d'actine et en même temps à la membrane. Elles tirent leur nom de la présence d'un domaine de liaison à la membrane. Le repliement du domaine BAR consiste en la dimérisation anti-parallèle de longues hélices allongées et repliées sur elles-mêmes. La particularité des différents domaines BAR résulte de la courbure que peuvent prendre ces dimères d'hélices (Figure 6A).

Du fait de la faible similarité de séquence et des variations structurales, les domaines BAR n'étaient pas regroupés en famille et même étaient considérés comme découlant d'une évolution indépendante [163–165].

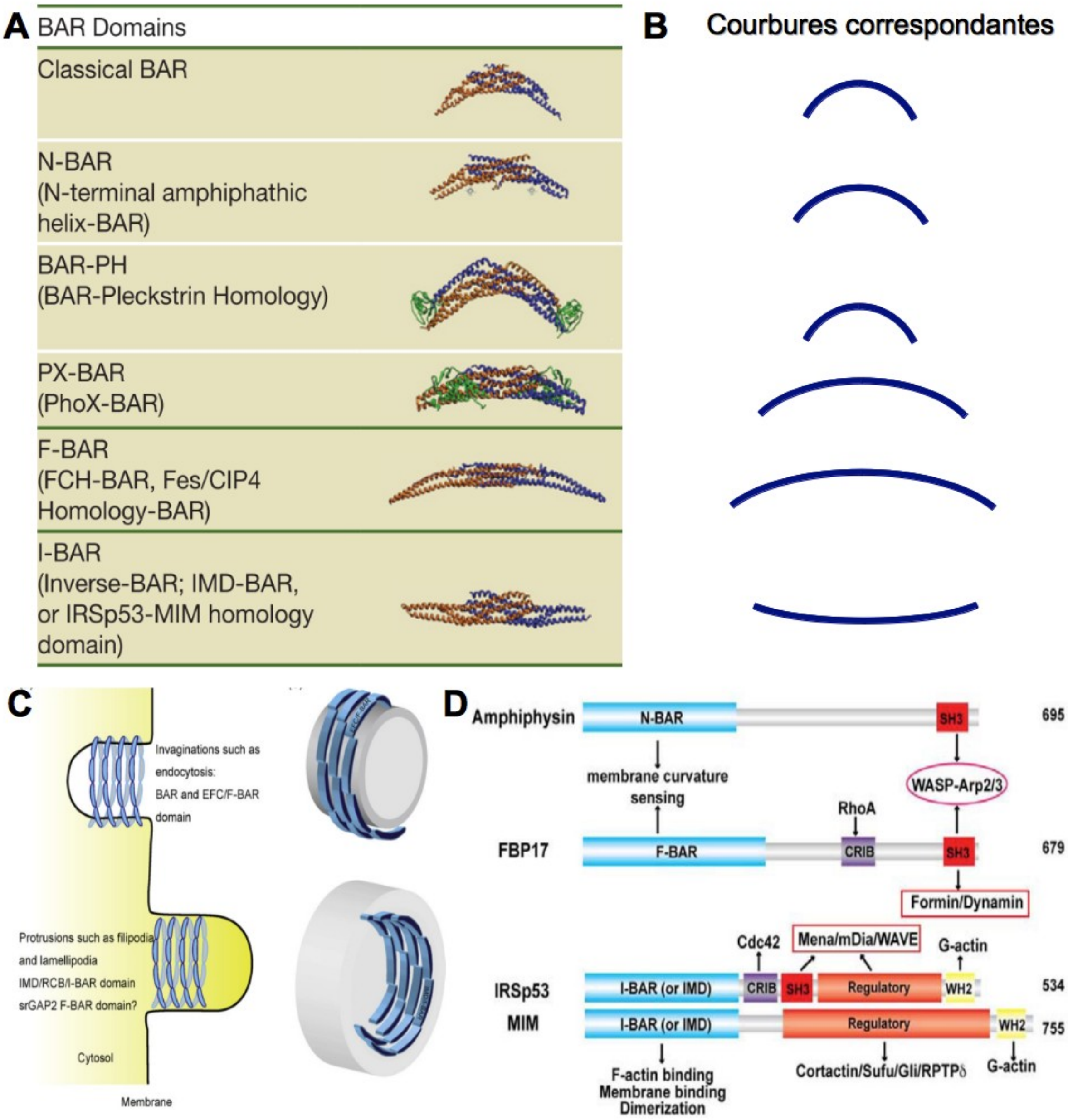

A
BAR Domains
Classical BAR
N-BAR
(N-terminal amphiphathic helix-BAR)
BAR-PH
(BAR-Pleckstrin Homology)
PX-BAR
(PhoX-BAR)
F-BAR
(FCH-BAR, Fes/CIP4 Homology-BAR)
I-BAR
(Inverse-BAR; IMD-BAR, or IRSp53-MIM homology domain)
B
Courbures correspondantes
C
Invaginations such as endocytosis: BAR and EFC/F-BAR domain
Protrusions such as filipodia and lamellipodia IMD/RCB/I-BAR domain srGAP2 F-BAR domain?
Cytosol
Membrane
D
Amphiphysin
N-BAR
SH3
695
membrane curvature sensing
WASP-Arp2/3
RhoA
FBP17
F-BAR
CRIB
SH3
679
Formin/Dynamin
Cdc42
Mena/mDia/WAVE
G-actin
IRSp53
I-BAR (or IMD)
CRIB
SH3
Regulatory
WH2
534
MIM
I-BAR (or IMD)
Regulatory
WH2
755
F-actin binding
Membrane binding
Dimerization
Cortactin/Sufu/Gli/RPTPδ
G-actin

Figure 6: Classement des différents domaines BAR et de leurs effet sur la membrane.(A) Table des différentes protéines contenant les différents types de domaines BAR.(B) Présentation des différentes courbures correspondantes aux types de domaines présentés en (A).(C) Exemple d'utilisation des domaines BAR lors de l'endocytose et de la formation de filopodes.(D) Organisation modulaire des différents domaines BAR. Les domaines BAR font toujours parti de protéines multi-domaines et en particulier des domaines de régulation de l'actine, de liaison au protéine de régulation, de domaine d'activation et de signalement. Ainsi de nombreux domaines BAR ont des domaines de recrutement pour certaines membranes ou protéines d'échafaudages (SH3). Certain domaines BAR présentent les site de recrutement des GTPases (activation/inhibition) (CRIB), ainsi que des domaines de liaisons à l'actine monomérique (WH2).

Illustration dérivée de [141,158,222].

Lorsque nous avons commencé ce travail sur le rôle de la protéine d'adaptation MIM qui contient deux sites indépendant de liaison à l'actine F et G, nous pensions qu'il s'agissait d'une protéine appartenant à une nouvelle famille (IMD) de protéines impliquées à l'interface du cytosquelette et de la membrane, dont l'autre représentant était IRSp53 (protéine tyrosine kinase récepteur de l'insuline). La protéine MIM lie l'actine par l'intermédiaire de deux domaines de nature et de structure différentes, situés aux extrémités de la protéine. En N-terminal par le domaine IMD de 250 acides aminés de structure inconnue et en C-terminal par un domaine de 30 acides aminés dont il est possible par alignement de séquence d'identifier la signature du domaine WH2 de liaison à l'actine-G. La région intermédiaire (± 400 acides aminés) est une région de régulation et de structuration. Ces deux protéines avaient été regroupées en raison de la similarité de leur architecture et des effets similaires qu'elles induisent sur le cytosquelette (Figure 6D).

Nos études structurales ont permis de démontrer que le domaine IMD de MIM formait un dimère antiparallèle. Il est composé d'un système de 3 hélices repliées dont deux sous-unités interagissent et forment le cœur du domaine, ce dimère forme une structure comparable à un étai et présente une très légère courbure inversée.

La comparaison structurale des domaines IMD indique que la seule partie structurellement conservée est la région de dimèrisation du domaine. Les extrémités chargées de la structure de MIM se révèlent être plus flexibles que le cœur du domaine et avec une implication assez faible dans la liaison à l'actine filamenteuse. Ces caractéristiques suggèrent que l'ensemble du domaine est impliqué dans la liaison au filament. Les domaines BAR possèdent également des régions chargées aux extrémités qui ne sont pas impliquées dans le recrutement des filaments mais dans l'ancrage à la membrane. Bien que distant nous avons comparé les structures de MIM et IRSp53 avec les domaines BAR dont plusieurs structures étaient connues (amphiphysine/ arfaptine/ endophiline). L'étude révèle une différence significative entre les extrémités du fait de la courbure, en revanche nous avons observé une superposition quasi parfaite au niveau de la zone de dimérisation. Pour l'arfaptine (protéine de la famille Bar), la zone de dimérisation sert de site de régulation où vient se lier la GTPase. Nous avons pu montrer que la régulation de MIM (ou de IRSp53) par la GTPase semble exclure l'actine au même endroit. Les similarités structurales, fonctionnelles, architecturales ainsi que les modes de recrutement et de régulation nous ont permis de proposer que les domaines IMD soient regroupés dans la famille des domaines BAR. Trois grandes familles de domaine BAR sont maintenant identifiées : BAR-N ; BAR-F ; BAR-I. Les domaines N et F présentent une forme globale en banane ; la différence venant de la courbure prononcée pour BAR-N ; et plus douce mais allongée pour le BAR-F. En revanche pour la famille des BAR-I la courbure est inversée. Pour chaque type de domaine une courbure membranaire spécifique peut être associée (Figure 6B /C).

L'autre site de liaison à l'actine-G de MIM implique un site de type WH2 (Figure 6D). La résolution du complexe actine WH2 a permis de montrer que le domaine WH2 est plus étendu que les domaines WH2 observés chez d'autres PLAs. Ce WH2 interagit dans la poche hydrophobe sous forme d'une hélice et remonte le long de la surface de l'actine sans structure secondaire jusqu'au site catalytique, bloquant le site de liaison à l'ATP.

Si le rôle du domaine BAR-I est pratiquement défini, il reste encore a démontrer le rôle exact du domaine de liaison à l'actine monomérique. Ainsi le modèle à l'étude propose que la combinaison des deux modules, liant deux formes d'actine, séparés par une grande région désordonnée, permet le renforcement de structures tels que les filopodes et les lamelipodes tout en recrutant les actines monomériques afin de les diriger vers l'extrémité (+) du filament.

Les articles des résultats présentés sont dans l'annexe A.

B. Structures des Nucléoprotéines de Phlébovirus

CNRS UMR 6098 – Architectures et Fonctions des Macromolécules Biologiques
Université de la Méditerranée (France) / Nanyang Technological University (Singapour)
Groupe : Protéines des virus émergents et Parasites.
Equipe de J. Lescar (2008-2010)

i. Contexte scientifique

Les Phlébovirus sont des virus qui appartiennent à la famille des *Bunyaviridae* et qui sont transmis par des piqures de petits insectes : les phlébotomes (mouches des sables), les moustiques ou les tiques. Ils sont principalement responsables d'infection touchant le bétails, ce qui en plus du problème sanitaire à pour effet de fragiliser les économies de populations précaires dont l'élevage est l'une des principale ressource [166]. Cependant, dix virus du genre sont liés à des pathologies humaines dont les symptômes varient de la fièvre à l'encéphalite et dans quelques cas à la fièvre hémorragique [167] (Table 2).

Les Phlebovirus ont un génome à ARN segmenté en trois brins (S, M, L) [168,169]. Le brin S est ambisens et code pour une protéine non structurale (NSs) et la Nucléoprotéine (N), qui protège constamment l'ARN génomique (ARNv), et antigénomique. Le polymère de N associé au génome sert de plateforme d'accueil de la protéine L formant le complexe ribonucléoprotéique (RNP). Les segments M et L sont des brins d'ARNs négatifs. M code pour les précurseurs de glycoprotéines d'enveloppe et une protéine non-structurale. L code pour la protéine L qui est une protéine multi-domaines et multi-fonctions qui inclut entre autre la polymérase [170,171] et un domaine endonucléase. Ce dernier est impliqué dans le vol de coiffe, nécessaire à l'acquisition de la structure coiffe des ARN messagers (ARNm) viraux [172–174]. Pour tous les virus à ARN négatif la protéine N est le constituant le plus abondant du complexe RNP. Les travaux réalisés sur ces protéines ces dernières années ont montré une grande variabilité en terme de taille, de mécanisme d'assemblage (polymérisation sur l'ARN) et de mode de protection du génome (extérieur ou intérieur). Contrairement à la plupart des virus à ARN négatif non segmenté, les RNP de Phlébovirus ne présentent pas de super structure tubulaire mais un serpentin super enroulé, suggérant l'existence d'un mécanisme d’encapsidation original [175].

Table 2 : listes des Phlébovirus, de leurs vecteurs associés et des symptômes correspondants

Virus	Vecteur	Symptômes associés
Virus Alenquer	phlébotome	fièvre
Virus Candiru	phlébotome	fièvre
Virus Chagres	phlébotome	fièvre
Virus Heartland	Tique	Fièvre, thrombocytopénie (chute du taux de plaquettes), avec parfois des troubles gastro-intestinaux et parfois une leucocytopénie(chute du taux de globules blancs.
Virus Naples	phlébotome	fièvre
Virus Punta Toro	phlébotome	Anomalie hépatique
Virus de la Fièvre de la Vallée du Rift	Moustique	Fièvre, douleurs musculaire, céphalé, anomalie hépatique,fièvre hémorragique, méningite
Virus du Syndrome de fièvre sévère avec thrombocytopénie	Tique	Fièvre, thrombocytopénie, avec parfois des troubles gastro-intestinaux et parfois une leucocytopénie. Mortalité 30 %
Virus Scilien	phlébotome	fièvre
Virus Toscan	phlébotome	Fièvre, douleurs musculaire, céphalé, méningite

Le virus de la fièvre de la vallée du Rift est un membre du genre Phlebovirus qui représente l'un des cinq genres de la famille des *Bunyaviridae*. La fièvre de la vallée du Rift est transmise par un arthropode (principalement le moustique). C'est une maladie virale aiguë, qui touche les moutons, les bovins, et les chèvres [176–185].

Chez ces espèces, la maladie est caractérisée par des niveaux d'avortement élevés, une mortalité élevée des nouveau-nés, et une nécrose hépatique.

Les humains y sont hautement susceptibles mais les infections sont habituellement asymptomatiques ou relativement bénignes incluant de la fièvre et des anomalies hépatiques. Toutefois, dans certains cas (3%), une forme sévère peut se développer avec des complications incluant des signes cliniques ressemblant à une maladie similaire à la dengue accompagnée d'hémorragie, de méningo-encéphalite, de rétinopathie. Dans certains cas la mort peut survenir. Dans le passé ce virus a fait l'objet de recherche dans le cadre de développement d'arme biologique, et présente dans le cadre de scénari bioterroristes un risque réel pour la sécurité publique. Du fait du manque de connaissance sur le mode de transcription et de réplication de ce genre viral il est urgent d'accumuler de nouvelles données structurales et biochimiques sur les constituants du complexe de réplication/ transcription afin de développer de nouvelles stratégies antivirales.

ii. Objectif du travail et Stratégies expérimentales

Les objectifs de ce travail étaient de comprendre les mécanismes moléculaires de protection de l'ARNv par la nucléoprotéine des phlébovirus. Nous nous sommes intéressés à la Nucléoprotéine du virus de la fièvre de la vallée de Rift, afin de comprendre son mode d'assemblage et de déterminer si les nucléoprotéines forment un polymère autour de l'ARN ou si l'ARN s'embobine autour du polymère. Les données structurales des nucléoprotéines de phlébovirus disponibles dataient des années 1970. Ces données de microscopie électronique présentaient un polymère circularisé très flexible et non tubulaire. Ces structures laissaient penser qu'un mécanisme d'assemblage différent prend place par rapport aux autres virus à ARN tels que les Orthomyxoviridae (virus de la grippe), les Arenaviridae (Virus CML) ou des Mononegavirales (virus de la rougeole, virus de Sendaï, VSV, Virus Ebola... etc), qui utilisent aussi des nucléoprotéines pour protéger leurs matériel génomique.

iii. Résultats obtenus et perspectives

Nous avons résolu la structure de la nucléoprotéine du virus de la fièvre de la vallée du Rift. Il s'agit d'une protéine uniquement composée d'hélices-α et subdivisée en trois sous domaines : un bras amino terminal qui se détend et permet la polymérisation, et un corps bi-globulaire en forme de haricot séparé par un sillon central accommodant l'ARNv (ou complémentaire). Nos structures ont permis de démontrer un système d'assemblage hexamérique permettant une protection optimum de l'ARN. Nos observations indiquent que l'ARN est à l'intérieur du polymère de nucléoprotéine et que le mécanisme d'assemblage conserve pour chaque corps globulaire un certain degré de liberté permettant au polymère la flexibilité observée en microscopie électronique. De toutes les structures de Nucléoprotéines résolues, celle de Phlébovirus est la plus petite (Figure 7), mais conserve les caractéristiques générales des nucléoprotéines : un corps globulaire central creusé d'un sillon préservant l'ARN et au moins une extension plus ou moins structurée permettant la polymérisation. La comparaison de notre structure avec celle précédemment publié par Raymond et collègues (2010) montre que le corps globulaire reste conservé mais que le bras peut se replier sur le sillon central en le couvrant. Cette différence structurale suggère que cette conformation fermée peut être une conformation d'attente ; qui permet aux monomères de N d'être en solution avant leur recrutement par les ARNv ou ARNc. Il se peut aussi qu'il s'agisse d'un artefact structural due à la purification. Cette question reste à débattre car si les tenants de la conformation fermée y voient un avantage sélectif, ils font fi des données biochimiques qui suggèrent que les N (en attentes) sont sous forme de dimères ou trimères [186] . De plus, cette hypothèse laisse en suspens le mécanisme d'ouverture du sillon pour le recrutement de l'ARNv. Il est d'autant plus important de se pencher sur cette question que de nouvelles structures de Nucléoprotéines d'Arenavirus ont été résolues [187,188]. Ces nouvelles données structurales proposent un nouveau mécanisme d'encapsidation. Cependant en observant les structures en prenant en compte les données de caractérisation de N de LCMV [189], il est possible de découvrir un sillon qui peut accueillir l'ARNv et de proposer un système de multimérisation similaire à celui de la Nucléoprotéine de RVFV. Dans cette hypothèse les structures résolues peuvent être considérés comme des systèmes repliés comme le montre le modèle proposé (Figure 7).

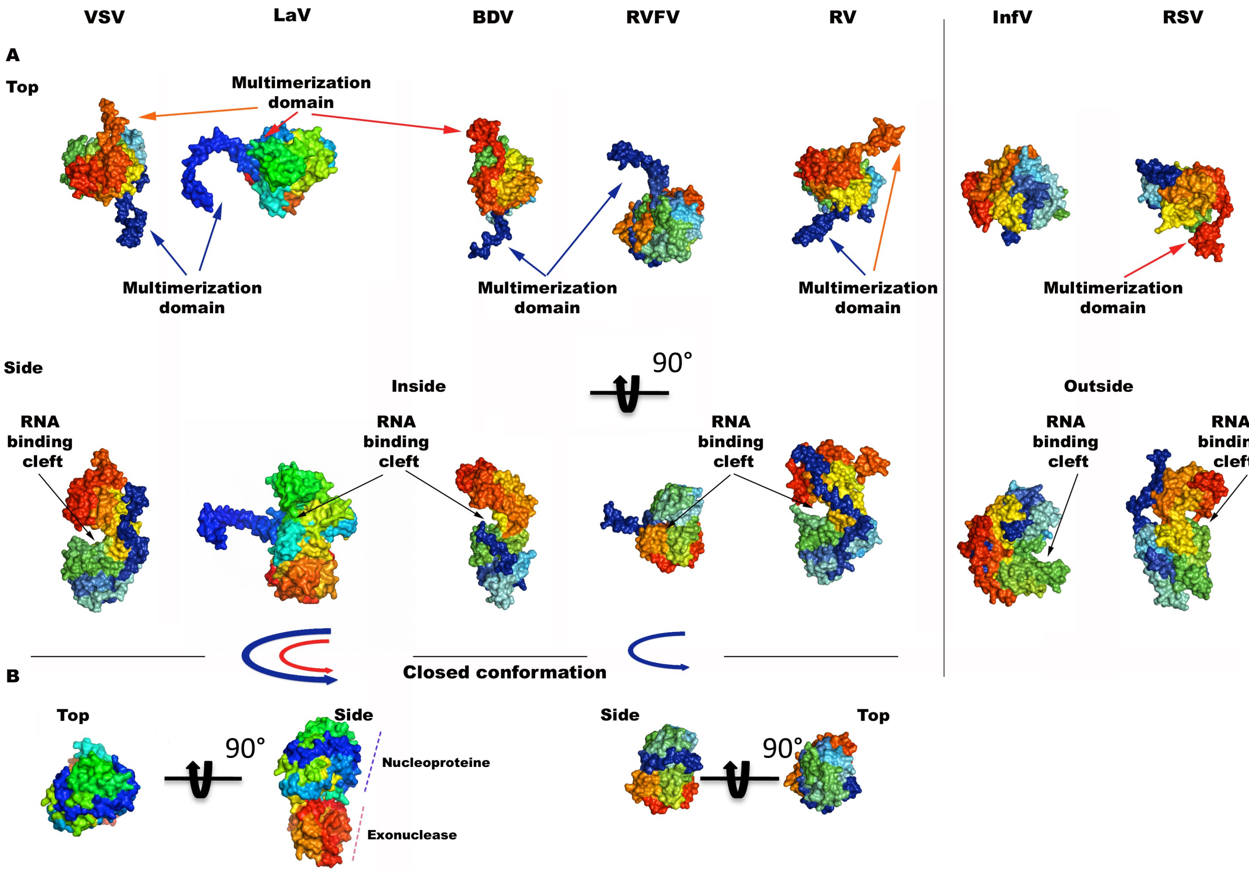


Figure 7: Comparaison structurales des différentes Nucléoprotéines de virus à ARN négatifs. Présentation des différentes structures de Nucléoprotéines de virus à ARN négatifs : Vesicular Stomatitis Virus (VSV)pdb :3pto / Lassa Virus (LaV) modèle / Borna Disease Virus (BDV) pdb :1n93 / Rift Valley Fever virus (RVFV)pdb :3OV9 / Rabies Virus (RV)pdb :2gtt / Influenza Virus (InfV)pdb :3zdp / Respiratory Syncitial Virus(RSV)pdb :2wJ8. Pour toutes les structures sont représenté sous forment de surface colorée selon la nomenclature arc-en-ciel (bleu vers rouge/ Nterminal vers Cterminal). (A) Top, vue de dessus de VSV/BDV/RVFVRV/InfV/RSV modèle cristallographique des monomères de nucléoprotéines. LaV est une modélisation reposant sur des observations de [169]. Les régions de multimérisation se retrouvent à l'une ou l'autre des extrémités . Side, vue de coté correspondante, mise en évidence de la région de liaison à l'ARN. (B) structures (LaV)pdb :3mwt / (RVFV)pdb :3lyf observées par cristallographie avec un bras de multimérisation replié et ne permettant pas la polymérisation de la nucléoprotéine, ni la protection de l'ARN.

L'article des résultats présentés est dans l'annexe B.

C. Structures et fonctions de protéines de la formation de la coiffe et du complexe de transcription/réplication des virus à ARN.

CNRS UMR 7257 – Architectures et Fonctions des Macromolécules Biologiques
Aix-Marseille Université (France)
Groupe : Structure, mécanismes et drug design des réplicases virales.
Laboratoire de B. Canard (2010-Présent)

i. Contexte scientifique

Le génome d'une cellule eucaryote est constitué d'ADN localisé dans le noyau. Il contient l'information génétique qui gouverne l'essentiel des aspects du fonctionnement cellulaire. Cette information doit ou peut être traduite sous la forme de protéines afin de répondre aux besoins fonctionnels et métaboliques de la cellule. La traduction de l'information génétique en protéines est assurée par les ribosomes, localisés dans le cytoplasme. Le passage de l'information du noyau au cytoplasme nécessite un ARN messager (ARNm) qui assure le transport de l'information génétique d'un compartiment cellulaire à l'autre. Cet ARNm est une copie transitoire de la partie de l'ADN contenant les instructions de synthèse d'une protéine. Sa présence dans le cytoplasme est contrôlée et strictement régulée selon les besoins cellulaires qui peuvent varier en fonction : des conditions environnementales, du type cellulaire, du stade de développement, de l'âge de la cellule. Lorsque la cellule régule négativement l'expression d'une protéine, la transcription nucléaire du gène s'arrête et au niveau du cytoplasme l'ARNm est progressivement dégradé par des ribonucléases (ou RNases). Ainsi la production de protéine peut être stimulée ou réprimée en fonction des besoins.

Les cellules codent pour des senseurs cytoplasmiques tel que RIG-1 (retinoic acid inducible gene 1 protein) et MDA5 (melanoma-differentiation-associated gene 5). Ces molécules sont des acteurs de l’immunité innée capables de reconnaitre l’ARN de pathogènes infectant une cellule. L'ARNm cellulaire n'est pas reconnu par les senseurs de l'immunité (RIG-1 et MDA5) de la cellule car il répond à certains critères structuraux : coiffé en 5' de l'ARN (Figure 8A) et polyadénylé en 3'. La coiffe joue également un rôle dans le recrutement du facteur eIF4E (eukaryotic translation initiation factor 4E), qui initie la traduction des ARNm au niveau des ribosomes. Les mécanismes de synthèse de la coiffe des eucaryotes sont détaillés Figure 8B. L'inactivation de la voie de synthèse de la coiffe (enzymes impliquées dans la formation de la coiffe) est létale pour les cellules, ce qui démontre l'importance de la coiffe pour la survie cellulaire.

Lorsque une cellule est infectée par un virus à ARN, les ARNs viraux relargués dans le cytoplasme (génomique et/ou messager), déclenchent une série de réponses cellulaires qui résultent entre autre par l'activation de la RNAse L. Cette ribonucléase assure la dégradation des ARN viraux. Afin de retarder la détection de leurs ARNs par les senseurs de l'immunité innée; et d'assurer la traduction de leurs ARNs en protéines, certains virus ont acquis au cours de l'évolution la capacité de coiffer leurs ARNs. La formation de la coiffe chez les virus peut résulter de différentes stratégies : détournement des mécanismes céllulaires, synthèse ou vol de coiffe médiés par des enzymes virales spécifiques. Notre intérêt se porte sur la compréhension et la caractérisation structurale de ces systèmes viraux de formation de coiffe.

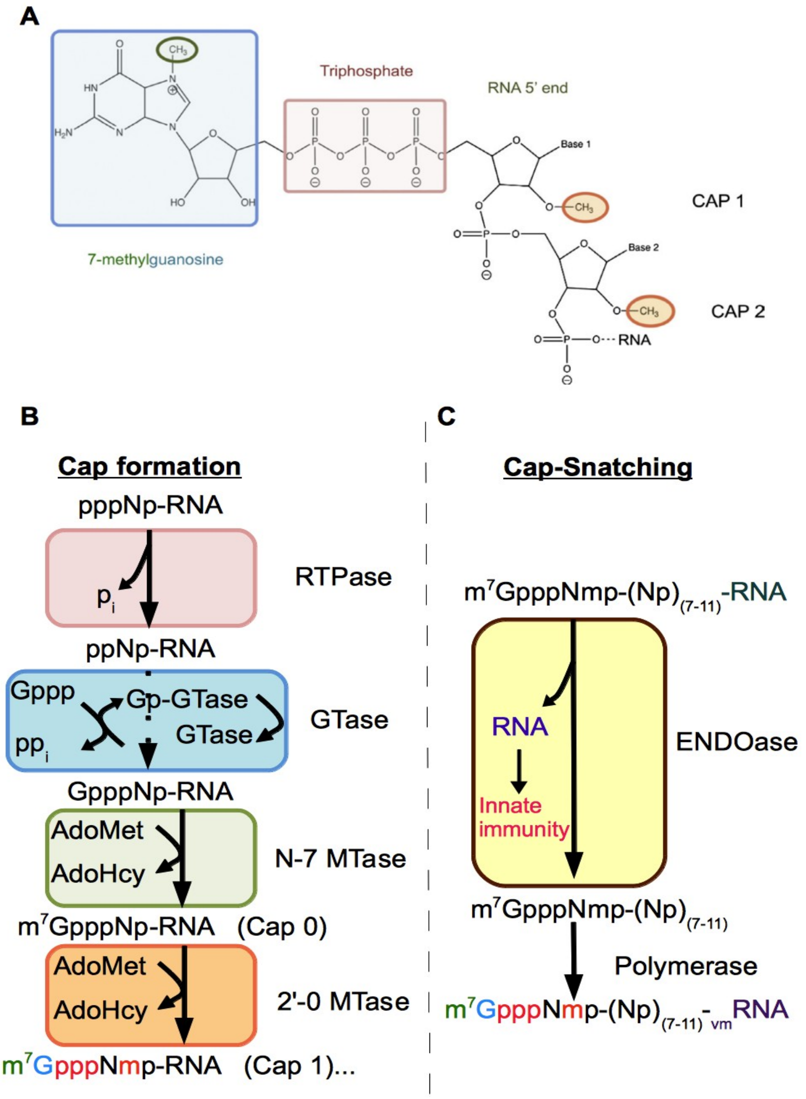


Figure 8: Structure de la coiffe des ARNm et mécanismes de leur formation. (A) Structure de la coiffe des ARN messagers : La structure coiffe-0 (fond bleu) est composée d'un résidu guanosine méthylé en position N7 (cercle vert à gauche) qui est lié par une liaison 5'-5' triphosphate (fond rose) au nucléotide en aval de la coiffe. Un groupement méthyl (cercle orange à droite) peut être ajouté à la position 2'O du ribose du premier et du deuxième nucléotide de l'ARNm pour former la coiffe-1 et -2.(B) Le phosphate γ À l'extrémité 5' de l'ARN pré-messager est hydrolysé par l'ARN triphosphatase (RTPase). L'ARN guanylyltransférase (GTase) fixe ensuite une molécule de GMP à partir du GTP (Gppp) en libérant du pyrophosphate inorganique (PPi). La GTase transfère le GMP sur l'extrémité diphosphate de l'ARN prémessager formant une liaison 5'-5' triphosphate. La synthèse de la coiffe se termine par la méthylation du résidu guanosine en position N7 par la N7-MTase à partir du donneur de méthyl (SAM). Enfin, chez les eucaryotes supérieurs, la formation de la structure coiffe-1 fait intervenir une 2'O-MTase qui méthyle la position 2'O du ribose du premier nucléotide de l'ARN.(C)Dans le mécanisme de vol de coiffe (cap-snatching), une endonucléase viral coupe la coiffe de l'ARNm de l'hôte, qui va servir d'amorce pour la synthèse de l'ARNm viral.

ii. Objectifs et stratégies expérimentales

La structure de la coiffe se compose d’une guanosine méthylée en position N7 liée par une liaison 5’-5’ triphosphate au premier nucléotide transcrit de l’ARNm (7MeGpppN), elle peut être éventuellement décorée de méthylation sur les positions 2'O des riboses des premières bases de l'ARN. Ainsi chez les eucaryotes supérieurs une coiffe peut se décomposer en trois niveaux (Figure 8A): coiffe-0 (fond bleu) est composée d’un résidu guanosine méthylée en position N7 (cercle vert à gauche) qui est liée par une liaison 5’-5’ triphosphate (fond rose) au nucléotide en aval de la coiffe. Un groupement méthyle (cercle vert à droite) peut être ajouté à la position 2’O du ribose du premier nucléotide de l’ARNm pour former la coiffe-1, voir sur le second ribose et dans ce cas là, la structure est appelée coiffe-2.

Pour les ARNm de la cellule, la coiffe est synthétisée par l'action cordonnée et séquentielle d'une série d'enzymes présentent dans le noyau. Cette synthèse est concomitante au début de la synthèse de l'ARNm (Figure 6B). La première étape consiste en l’hydrolyse du phosphate γ de l’extrémité 5’ triphosphate de l’ARNm naissant par une ARN triphosphatase (RTPase). L’ARN guanylyltransférase (GTase) recrute une molécule de GTP et forme un intermédiaire covalent via sa lysine catalytique, avant de transférer le groupement GMP sur l’extrémité 5’ de l’ARN diphosphate. Enfin, une N7-méthyltransférase (N7-MTase) méthyle la guanosine sur la position N7 en utilisant comme donneur de méthyle le S-adénosyl-L-méthionine (SAM ou AdoMet).

Tous les virus utilisent la machinerie de la cellule infectée afin de traduire leur ARNm en protéines. Bien que certains virus comme les virus de l'hépatite C ou Poliovirus, possèdent des structures IRES (internal ribosome entry site) qui permettent la traduction de leur ARNm en l'absence de structure coiffe, la plupart des virus coiffent l’extrémité 5’ de leurs ARNm afin d'assurer leur traduction par les ribosomes cellulaires. Trois stratégies virales d’acquisition de coiffe sont connues :

- <u>Stratégie de détournement cellulaire.</u> Dans cette stratégie, les ARNm sont synthétisés par l’ARN Polymérase II et la coiffe par les enzymes cellulaires . C’est le cas des rétrovirus et de la plupart des virus à ADN (*Hepadnaviridae* et *Parvoviridae*) à l’exception des *Poxviridae* et *Herpesviridae*.

- Stratégie de synthèse embarquée . Cette stratégie est utilisée par la plupart des virus à cycle entièrement cytoplasmique et consiste à coder et synthétiser l'intégralité de la machinerie enzymatique nécessaire à la formation de leur structure coiffe. Cette stratégie est retrouvée chez les Nidovirales, Flavivirus et Mononegavirales. Il est à noter une certaine diversité structurale dans ces machineries virales impliquées dans la formation de la coiffe, en revanche leurs fonctions et mécanismes réactionnels sont similaires aux enzymes effectuant les coiffes cellulaires.

- Stratégie du vol de la coiffe des ARNm cellulaires (cap-snatching). Cette stratégie (Figure 8C) est utilisée par les virus à ARN segmenté ambisens de polarité négative dont le cycle est cytoplasmique (*Orthomyxoviridae*, *Arenaviridae* et *Bunyaviridae*). Dans ce cas, le vol de coiffe précède l'initiation de la synthèse d'ARNm et implique un système de capture des ARNm cellulaires, constitué entre autre d'une endonucléase dont la fonction est de couper l'ARNm cellulaire en aval de la structure coiffe. Ces petits fragments d’ARN sont ensuite utilisés comme amorces par l’ARN polymérase ARN-dépendante virale pour synthétiser les ARNm viraux.

Nous nous sommes intéressés à la caractérisation structurale de deux méthyltransférases de deux familles virales différentes (Flavivirus et Coronavirus) et impliquées dans la formation de structure coiffe-1 ainsi qu'à l'endonucléase de la protéine L d'Arenavirus impliquée dans le mécanisme de vol de coiffe [190–192].

iii. Résultats obtenus et perspectives

- Méthyltransférase du virus du Syndrome Respiratoire Aigu Sévère (SRAS-CoV) :

Les coronavirus sont à l’origine de nombreuses maladies respiratoires et entériques chez le bétail et les animaux de compagnie et constituent de ce fait un problème économique d’importance. Ils causent également une part des rhumes communs chez l’homme avec les rhinovirus. Ils ont été largement médiatisés en 2003 puis en 2013 lors des épidémies de syndrome respiratoire aiguë sévère (SRAS) dont les agents étiologiques sont le SRAS-coronavirus (SRAS-CoV) et le MERS-coronavirus (Middle East Respiratory Syndrome Coronavirus). Ces derniers appartiennent à l'ordre des Nidovirales et à la famille des *Coronaviridae*. Les coronavirus sont des virus à ARN simple brin de polarité positive qui possèdent un génome d’une taille surprenante (27-32 kb), au moins deux fois supérieure à celle des génomes des autres virus à ARN. Leur génome encode une partie si ce n'est l'intégralité de la machinerie de formation de la coiffe. Chez ces virus la conversion des structures coiffe-0 en coiffe-1 est assurée par la protéine nsp16 (sous forme de complexe activé par nsp10). Dès 2003, nous avions et d'autre identifié par bio-informatique, le motif caractéristique d'une activité 2'O-Méthyltransférase sur cette protéine et l'activité fut caractérisée au laboratoire sur un coronavirus félin en 2008. Les études d'interactions ont montrés qu'une petite protéine nsp10 interagissait fortement avec nsp16. La caractérisation biochimique a montré que nsp10 joue le rôle d'activateur de la protéine nsp16 suggérant un rôle d'interrupteur moléculaire. Utilisant ces données nous avons pu résoudre la structure de nsp16 en complexe avec nsp10. Nsp16 présente globalement un repliement typique des repliements des S-Adenosyl-Methionine (SAM) 2'O-Méthyltransférase, constitué d'un feuillet β central de 7 brins et entouré de 5 hélices α. Cependant la topologie diverge du repliement standard des S-Adenosyl-Methionine (SAM) 2'O-Méthyltransférase (Figure 9). En effet, la structure révèle qu'il manque 2 hélices des 7 hélices α pour avoir un repliement canonique et que ce manque est compensé par la présence de nsp10 qui vient stabiliser et rigidifier le site de liaison au SAM (donneur de méthyle) et du sillon potentiel de liaison à l'ARN. Nous avons également identifié un ion magnésium dans la structure qui stabilise une partie de nsp16, et prouvé son rôle indispensable pour l'activité de nsp16. Enfin, nous avons aussi résolu la structure du complexe en présence d'un inhibiteur (Sinefungin) analogue de SAM, montrant ainsi que ce dernier est capable de cibler le site de liaison par compétition. *In vitro*, l'acide aurintricarboxylique (ATA) a été montré comme étant un inhibiteur de l'activité méthyltransférase. Nous n'avons pas encore obtenu de cristaux en présence de cet inhibitueur, il reste à comprendre comment il fonctionne et son mode de fixation. La prochaine étape de caractérisation du mécanisme de méthylation de l'ARN par nsp16 consiste en l'obtention d'une structure de nsp10/16 en complexe avec de l'ARN.

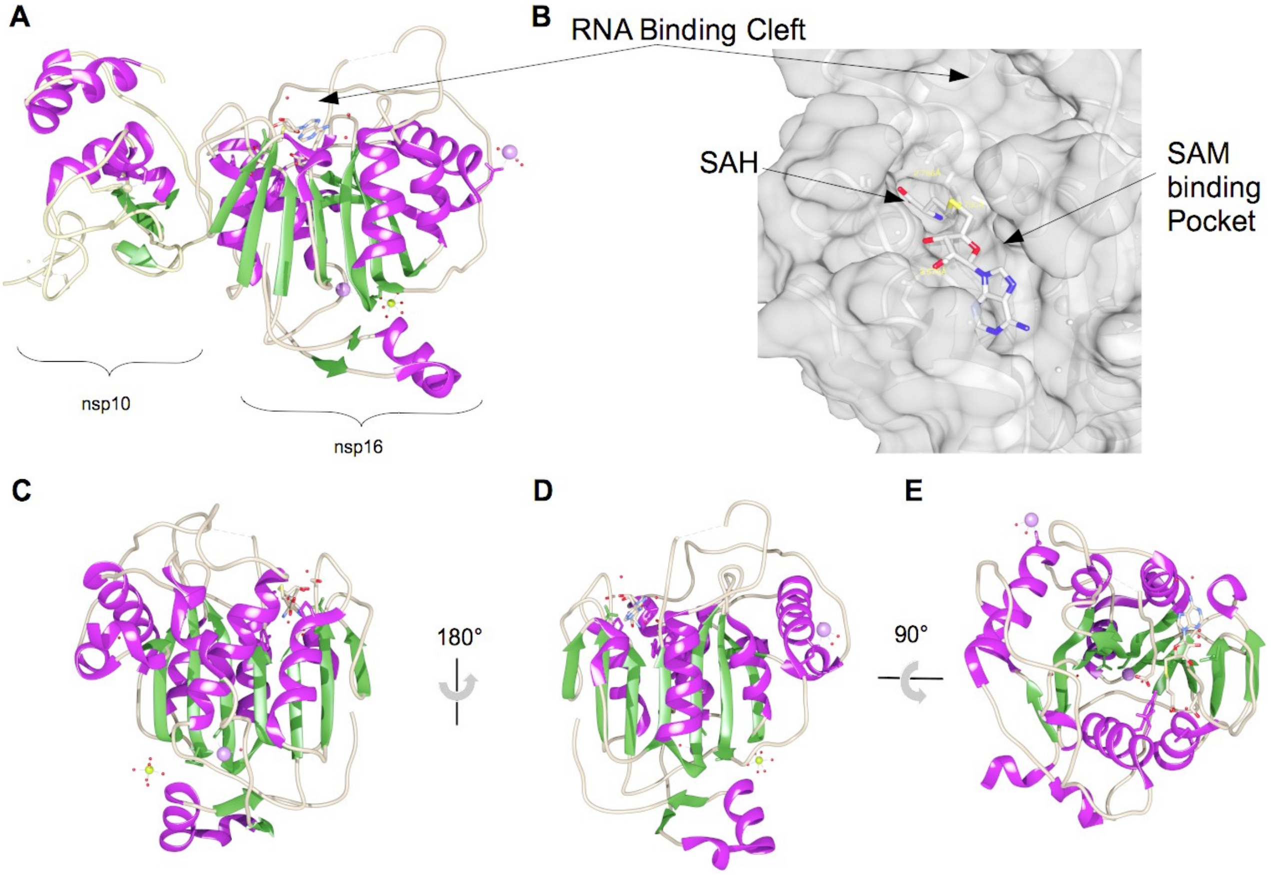


Figure 9: Structure du complexe nsp10/16. (A) Représentation cartoon du complexe nsp10/16 avec le produit de réaction SAH et un ion métallique. nsp16 se lie à nsp10 par la face qui n'implique pas les ions zincs. Un ion métallique se retrouve sur la face opposée du site de liaison au SAM. (B) Vue plongeante du site de liaison du SAM en représentation surface, avec le site de liaison potentiel de l'ARN. (C-E) Représentation cartoon de la structure de nsp16 dans différentes orientations afin de mettre en évidence le repliement typique des 2'O-Methyltransférases : C/D de coté et E vue de dessus. Le code couleur s'appliquant pour toute la figure : les brins β sont en vert, les hélices α sont en rose, et les boucles en blanc; ions (Na violet, Mg jaune, Zn crème), l'eau en rouge, les représentations du SAH suivent le code couleurs des atomes (N bleu, O rouge, S jaune, C blanc), surface en gris.

- Méthyltransférase du virus de la Dengue (Flavivirus) :

La dengue est une arbovirose, transmise à l'homme par l'intermédiaire de la piqûre d'un moustique diurne du genre Aedes (ægypti ou albopictus), lui-même porteur du virus de la dengue (flavivirus). Le génome du virus de la Dengue est constitué d'un ARN simple brin de polarité positive coiffé en 5' mais non polyadénylé en 3'. Le génome de 11kb code pour 7 protéines non structurales (NS1-5) et 3 protéines structurales (C, prM et E). Le complexe de Transcription/Réplication associe 7 NS d'une manière dynamique impliquant des réarrangements structuraux au cours du cycle. Parmis les 7 protéines du complexe, NS3 et NS5 ont fait l'objet de caractérisations avancées. Elles portent respectivement les activités RTPase/Hélicase et Méthyltransférase/Polymérase.

NS5 est une protéine organisée en deux domaines methyltransférase (NS5-Mtase) et polymérase (NS5-Pol) séparés par un domaine de liaison flexible.

NS5-MTase est responsable des N7 et 2'O méthylations lors de la formation de la coiffe des ARNm viraux [193]. L'étude structurale du domaine NS5-MTase a été effectuée en présence d'analogues de coiffe. Cette étude vient compléter une série d'études biochimiques sur les *flaviviridae* dont l'objet était la caractérisation des intermédiaires réactionnels des étapes de méthylation lors de la formation de la coiffe. Nous avons résolu au cours de cette étude 5 structures en complexe avec GpppA, N^7GpppA, GpppG, N^7GpppG, N^7GpppG$_{2'O}$. Dans ces structures, nous avons pu observer que l'analogue de coiffe est lié dans la poche de fixation du GTP, à une distance trop éloigné du site catalytique pour comprendre le mécanisme de la N7 méthylation. En revanche cette étude a permis de définir la poche de fixation du GTP et les contacts entre la protéine et la guanine, le ribose et le phosphate. La molécule se positionne dans cette poche du fait de la présence d'une phénylalanine (F25) qui permet par un effet d'empilement avec le cycle, la stabilisation de la base (Figure 10). Il apparaît que lorsqu'on ajoute une base supplémentaire l'effet d'empilement augmente, la guanine de la coiffe se trouve pris en sandwich entre la F25 et la première base.

Les structures obtenues en complexe avec le GpppA et le GpppG semblent présenter des caractéristiques structurales compatibles avec celles d'états intermédiaires de la réaction de coiffe. En effet l'analogue de coiffe GpppA se retrouve dans une conformation repliée sur lui-même. Cette conformation ne peut pas correspondre à une conformation permettant la N7 ou 2'O methylation. En effet, ni la base ni le sucre se trouvent à bonne distance du site AdoMet pour permettre à la réaction de se faire. En revanche, nous avons proposé l'hypothèse que cette conformation mime le produit de réaction de guanylyltransfert qui précède les réactions de méthylation. Cette structure est compatible avec la présence d'un ARN (Figure 10). La structure en complexe avec le GpppG présente la molécule étendue. Seule la première base et le phosphate sont actuellement visibles. Le degré de liberté trop grand de la seconde base laisse supposer la conformation du GpppG dans laquelle le ribose de la seconde base est à bonne distance du site AdoMet pour permettre la 2'O méthylation. Nous avons proposé un mécanisme réactionnel reconstituant les différentes étapes de la formation de la coiffe. (I) fixation du GTP au site de liaison du GTP de la méthyltransférase. (II) Réaction de guanylyl transfert sur l'ARN (structure GpppA). (III) Ouverture de la conformation empilé des deux bases pour permettre la N7 méthylation. (IV) repositionnement de l'ARN pour méthylation 2'O. Yap et collaborateurs [194] ont depuis résolu la structure de la méthyltransférase en complexe avec un ARN coiffé qui se positionne comme le GpppA. Cette structure montre que la guanine est à une distance compatible avec l'étape (III) du modèle que nous avons proposé.

Sur la base de la conformation de ces structures, nous avons postulé que nous avions reconstitué des complexes structuraux correspondant au produit d'une réaction du transfert coiffe. Nous avons proposé une implication direct de NS5 dans cette étape, et cette nouvelle hypothèse ouvre la voie vers de nouvelles approches antivirales.

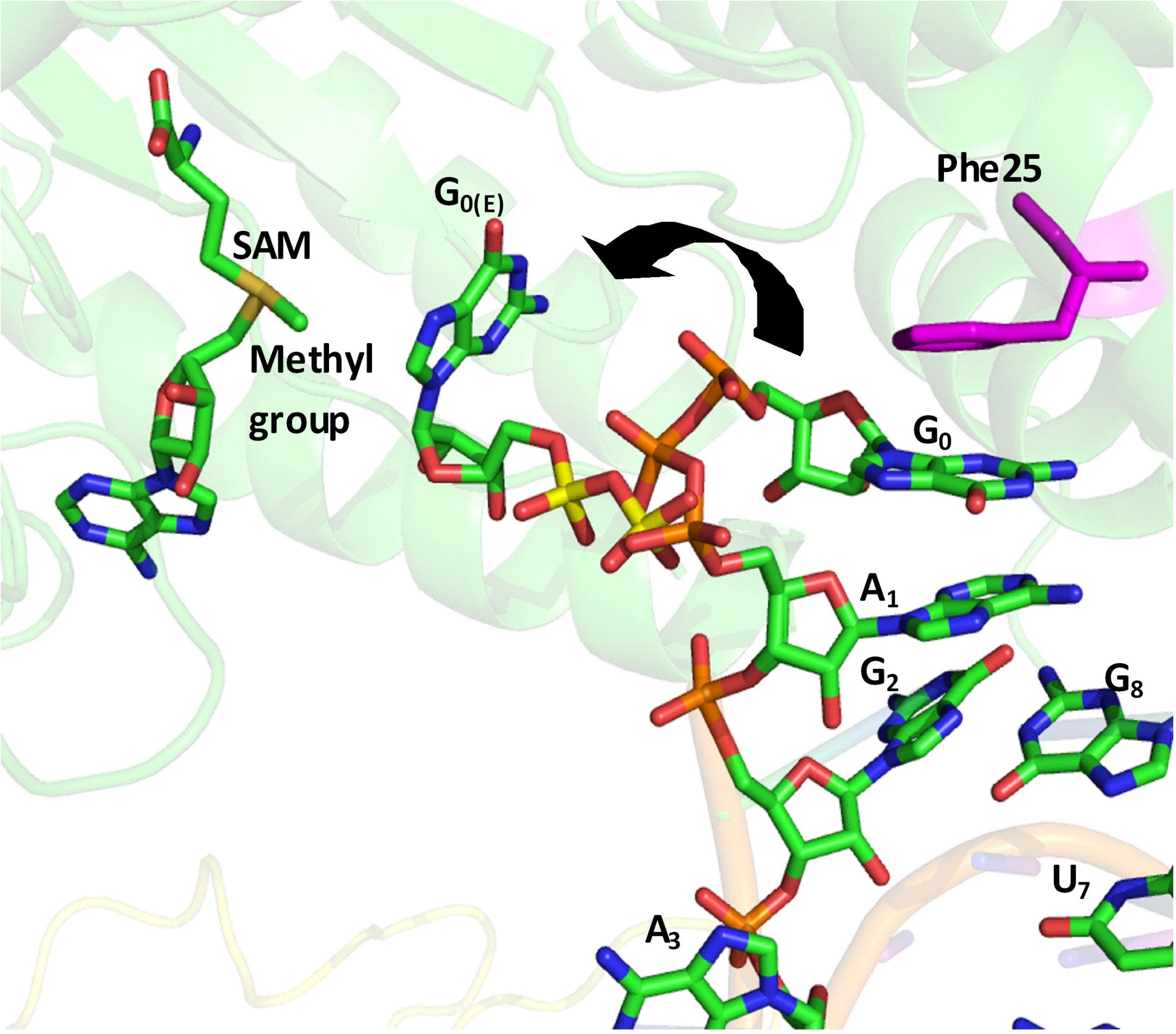


Figure 10: Modèle hypothètique pour la méthylation en N7 de la première base, suite à l'observation du positionnement de plusieurs complexes de Méthyltransférase en présence de GTP ou GpppA ou GpppG et de SAH. La position observé dans la structure (G0) montre un empilement de base avec la phénylalanine 25. L'extension du triphosphate ramène la base de la coiffe à bonne distance du SAM pour la méthylation en position N7(G0$_{(E)}$).

La liaison phosphate observé dans la structure est en orange celle modélisé est en jaune.

Illustration dérivée de [194]

- Endonucléase du Virus de la Chorioméningite Lymphocytaire (VCML) :

Le virus de la Chorioméningite Lymphocytaire est un Arenavirus, qui utilise une stratégie de vol de coiffe pour ses ARNm (Figure 11A).Cette activité est portée par la Large (L) protéine une protéine multidomaines de plus de 2000 résidus. Contrairement aux L de Mononégavirales qui possèdent en plus du domaine polymérase une ligne de production de coiffe intégrée, la L des Arenavirus reste d'architecture et fonctions inconnues. Les études bio-informatiques ont identifié précocement la présence du domaine ARN polymérase, plus récemment et par analogie avec la polymérase du virus de la grippe un domaine endonucléase a été postulé en amino terminal de la L. Nous avons résolu la structure de ce domaine et réalisé une étude structure fonction. Il s'agit d'une protéine qui présente un repliement typique des endonucléases de type II, constituée d'un feuillet central entouré d'un coté de deux hélices-α parallèles et de l'autre d'une seule grande hélice-α mais dont l'extrémité regroupe 4 hélices-α (Figure 11B). L'analyse structurale a démontré que le repliement est globalement conservé au seins des trois familles virales qui pratiquent le vol de coiffe ainsi que le site catalytique. Cela étant dit, le domaine endonucléase de VCML ne présente pas les résidus qui permettent de définir une cavité catalytique capable de prépositionner le substrat. Malgré cette absence surprenante nous avons pu démontrer l'activité endonucléase. Notre étude montre que l'activité endonucléase est ions manganèse dépendante. Toutefois nous n'avons pas pu obtenir de structure en complexe (Figure 11C) et ce malgré des trempages répétés en présence d'ions. Cela suggère qu'il manque encore des « pièces » au complexe nécessaire et suffisant pour reconstituer le mécanisme de vol de coiffe complet. Il sera nécessaire en particulier de comprendre le mécanisme de recrutement des ARN coiffés.

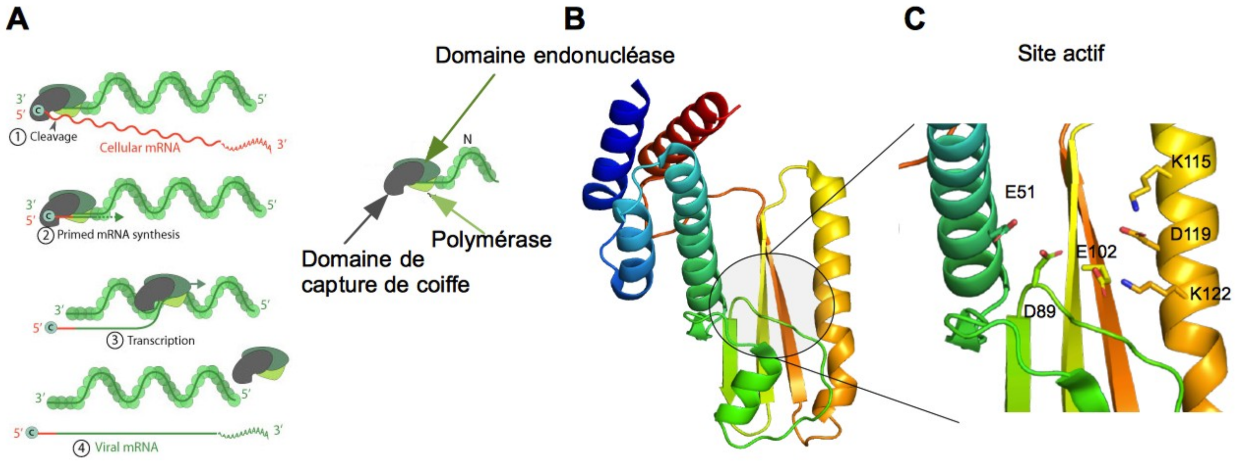


Figure 11: Machinerie virale du vol de la coiffe d'ARNm céllulaire. (A) Le complexe de réplication/transcription, formé de Nucléoprotéines polymérisés sur l'ARNv et de la protéine L, possède un domaine de capture d'ARNm qui recrute un ARNm de la cellule. Le domaine endonucléase présent sur L vient couper l'ARN en aval de la coiffe, l'amorce ainsi récupéré est utilisé par le domaine polymérase (toujours sur L) pour commencer la synthèse de l'ARNm. (B) Structure cartoon du domaine endonucléase du VCML. (C) Vue plongeante du site actif présumé, les résidus impliqués sont en en représentation bâton.

Illustration (A) dérivée de [221]

Les articles des résultats présentés sont dans l'annexe C.

IV. PERSPECTIVES DE RECHERCHE

Jusqu’en 2003 les virus émergents étaient considérés d’intérêts mineurs. Suite aux épidémies du virus du « Syndrome Respiratoire Aigu Sévère » (SRAS) en 2003, de grippe aviaire en 2004 et 2006 puis de Chikungunya en 2006, 2007 et 2013, et en 2012 et 2013 de l'émergence de nouveaux pathogènes Coronavirus (MERS-CoV) et d'Arénavirus [195,196]; le monde a pris conscience de la vulnérabilité des sociétés (moderne ou non) face à ces nouvelles menaces. Ces pandémies ont montré que les autorités sanitaires mondiales étaient dépourvues de réponses thérapeutiques adéquates et ont mis en évidence l’importance de la recherche fondamentale dans ce domaine. Le projet de recherche que je développe et qui fait écho à mon parcours scientifique, est l'étude des complexes structuraux des protéines du complexe de transcription/réplication chez les *Arenaviridae*. Cette famille constitue l'un des principaux réservoirs de virus émergents à ARN segmenté simple brin et qui inclus des pathogènes responsables entre autres : de la fièvre de Lassa, de la fièvre de Lujo et de la fièvre de Machupo.

Si l’on veut disposer d'un arsenal thérapeutique contre l'émergence des *Arenaviridae,* il est nécessaire de mieux comprendre leurs architectures moléculaires et leurs modes de fonctionnement.

A. Rappels sur la biologie des Arenaviridae

Les *Arenaviridae* sont des virus enveloppés, à ARN simple brin ambisens de polarité négative. Ils sont d'abord prévalents chez les rongeurs (avec une exception pour le virus Tacaribe qui se retrouve chez la chauve souris) et occasionnellement ils sont transmis à l'homme. Les *Arenaviridae* sont des virus au genre simple Arenavirus. Composé seulement de 24 virus, leur classement se fait sur la base de leur propriété antigénique. Le genre est ainsi divisé en deux sous-groupes, ceux de l'ancien monde (Europe/ Asie/ Afrique) et du nouveau monde (Amériques / Antilles / Océanie). Les virus du nouveau monde sont répartis en trois sous clades (A, B et C). L'organisation des clades repose

sur l'analyse des séquences des gènes des Nucléoprotéines [24].
Le virus de la chorioméningite lymphocytaire (VCML) ; virus de l'ancien monde ; sert comme comme virus modèle pour l'analyse des Arenavirus . En effet, les Arenavirus comptent des virus hautement pathogènes pour l'homme, comme les virus Lassa ou Lujo, qui sont responsables de fièvres hémorragiques chez l'homme et que l'on retrouve en afrique de l'ouest ou au sud du continent. Parmi les virus du nouveau monde, les virus responsables de fièvres appartiennent tous à la clade B (Sabia, Junin, Machupo, Guanarito, Amapari, Tacaribe, Cupixi) [197].

i. Structure du virion

Les Arenavirus sont des virus sphériques, de diamètre compris entre 110 et 130nm, enveloppés d'une couche lipidique dans laquelle s'insère les glycoprotéines d'enveloppe en forme de club de golf. Sous la membrane se trouve une capside (composé de protéines Z assemblées) à laquelle s'associe deux ribonucleoprotéines pseudo-circulaires (Figure 12A). Ces ribonucléoprotéines correspondent au génome sous forme de deux brins d'ARN. L pour le grand brin (Large ±7,5kb) et S pour le petit (Small ±3,5kb). L'extrémité 3' des brins (±19-30 nucléotides) est complémentaire de l'extrémité 5' ce qui permet l'appariement des extrémités dans le virion et de donner la forme circulaire au génome. En plus d'être complémentaire, la séquence se trouve être conservée entre les brins et au sein de la famille. Les virions contiennent aussi des ribosomes, qui donnent l'aspect granuleux en microscopie électronique d'où cette famille tire son nom :*arena*= sable en latin (Figure 12) [24].

ii. Description du génome et des protéines

Le génome code au total pour quatre protéines NP / L / GP / Z qui sont réparties sur le génome de façon à produire les protéines de façon séquentielle lors des différentes étapes du cycle viral.

Le segment L comprend pour deux gènes L en 3' et Z en 5' alors que le segment S porte NP en 3' et GP en 5'. Les gènes de chaque segment sont séparés par une région intergénique qui arrête la polymérase lors de la transcription (Figure12 B).

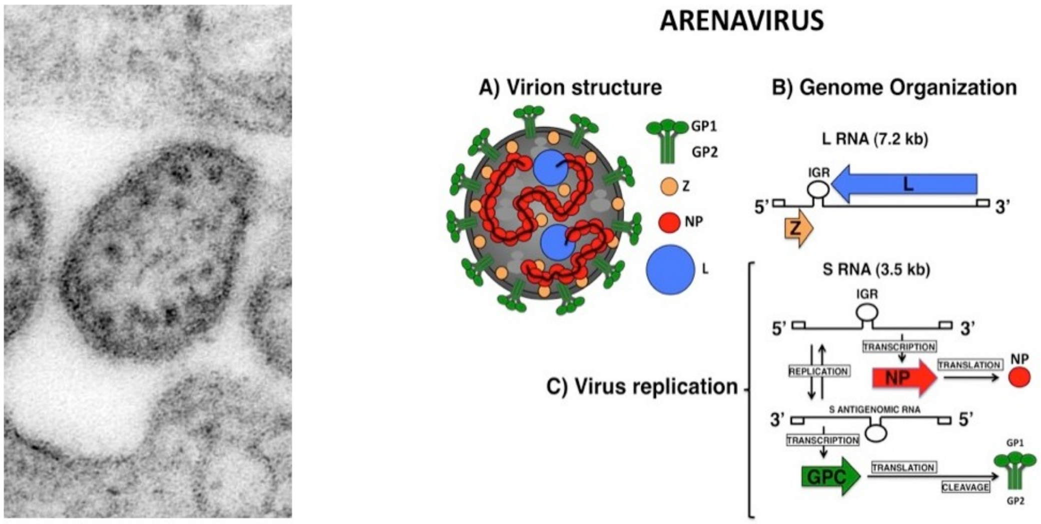


Figure 12: Structure du virion d'Arenavirus. Panneau de gauche présente une image de microcscopie éléctrionique à transmission du virus de Lassa. (Photo du domaine publique du CDC auteur C. S. Goldsmith ID#: 8699). Panneau de droite présente ***A****) le schéma d'une coupe de virion et la position dans le virion des protéines virales (GP/Z/L/NP).* ***B****) organisation du segment L du génome* ***C****) organisation du segment S et cycle de réplication en deux temps.*

L est une protéine de plus de 200kDa qui porte entre autre les activités polymérase et endonucléase. La protéine L est divisées en 4 sous domaines. Le premier domaine en amino-terminal de la protéine L est un domaine endonucléase impliqué dans le mécanisme de vol de coiffe. C'est le seul domaine de la protéine L qui a été exprimé et dont la structure et l'activité ont été caractérisés. Le troisième domaine correspond à l'ARN polymérase ARN-dépendante. Ce domaine a été identifié par bioinformatique grâce à la présence des motifs ARN polymérase ARN-dépendante conservés et l'activité a été confirmée par génétique inverse. Les autres domaines de la protéine sont de fonctions et de structures inconnues .

Le second gène en 5' de l'ARN génomique est Z : une petite RING protéine 10 kDa (Really INtersting Gene) qui possède un corps structuré par deux doigts de Zinc et dont le reste est principalement désordonné. Z joue le rôle de protéine de matrice : lors du bourgeonnement Z a un rôle structural majeur dans l'assemblage du virion par le recrutement de NP et des glycoprotéines. Bien que Z ne soit pas essentiel à la réplication, elle joue un rôle régulateur dans l'activité de la protéine L. La protéine Z interfère aussi avec les facteurs cellulaires PML (promyelocytic protéine) pour prévenir l'apoptose et PRH (proline-riche homeodomaine protéine) qui joue un rôle dans les processus de régénération des cellules. Enfin Z joue un rôle dans le recrutement de eIF4E ; facteur permettant le recrutement de la sous-unité ribosomale 40S sur la coiffe de l'ARN viral [19].

NP est la nucléoprotéine, une protéine de 63 kDa qui à la fois protège l'ARN génomique et antigénomique et qui possède un domaine exonucléase impliqué dans l'inhibition de la réponse à l'interferon de type I [198,199]. NP est la protéine la plus abondante du virus, elle polymérise sur l'ARN et forme un cofacteur essentiel à la protéine L. La structure de la NP du virus de Lassa a été résolu seul et sous forme de domaine en complexe. Ces études présentent des résultats contradictoires quant à l'organisation possible du polymère de NP avec l'ARN viral. La région C-terminale de la NP contient un domaine 3'-5' exonucléasique spécifique des ARN double brins (db), similaire à celui présent dans les exonucléases de la famille DEDDh . La NP peut digérer les ARN db marqueurs de l'infection virale, empêchant ainsi la cascade d'activation initiée via la reconnaissance de ces ARN par l'hélicase RIG-I (et donc l'activation des réponses IFN) [200]. Les structures apo et en complexe avec le substrat ont été résolues pour plusieurs virus. Les acides aminés présents au niveau du site catalytique ont été identifiés et leur mutation abolit la capacité de la NP à inhiber la réponse antivirale. Cependant, le virus Mopeia possède également ce domaine DEDDh et est probablement capable d'inhiber la réponse IFN, bien que de façon moins efficace. D'ailleurs, des virus Lassa recombinants contenant des mutations dans ce domaine induisent des quantités d'IFN de type I bien supérieures à celles sécrétées lors de l'infection par le virus Mopeïa [201–203]. La capacité de lutter contre les défenses cellulaires ne se limite pas seulement à la fonction exonucléase de NP. En effet, la NP des arénavirus empêche la phosphorylation d’IRF3 en se liant à la kinase IKKε [204]. La NP séquestre IKKε dans une forme inactive en inhibant son activité auto-catalytique. Ce complexe inhibant le changement de conformation de la kinase empêche l’induction des réponses immunes innées en amont des cascades de régulation cellulaire. Enfin, la NP est capable d’inhiber la translocation nucléaire et par extension l’activité transcriptionnelle du facteur de transcription NFκB ce qui a pour effet d'inhiber la production de cytokines [205]. Cette observation corrèle en particulier avec l’absence d’activation des cellules présentatrices d’antigènes et l’absence de production des cytokines pro-inflammatoires observées lors de l’infection par le virus Lassa. Ainsi, les Arenavirus ont développé des stratégies efficaces pour empêcher l’induction de l’immunité innée dont la pierre d'angle est la Nucléoprotéine.

GPC (75 kDa) est un précurseur des deux glycoprotéines GP1, GP2 et d'un petit peptide signal (SSP) de 58 acides aminés, responsable de la structure du virion et de sa pénétration dans la cellule. Un fois produit, le peptide signal adresse GPC à la membrane, GPC trimérise et forme un complexe tripartite SSP/GP1/GP2. Après maturation dans le Golgi, par la protéine site-1-protease/subtilisin-like kexin isozyme-1 (S1P/SKI-1), deux protéines et un peptide sont formés : GP1 est la protéine d'attachement au récepteur, GP2 est la protéine de fusion et enfin SSP joue un rôle de transmetteur. SSP se réarrange après modification afin de traverser deux fois la membrane et d'avoir son C-terminal du coté cytosolique. Les deux domaines transmembranaires sont séparés par une boucle hydrophyle exposée qui joue un rôle dans la détection de l'acidification de l'endosome.

Chez les virus de l'ancien monde, GP1 cible les récepteurs d'entrées α-dystroglycan, alors que pour les virus du nouveau monde, GP1 s'accroche aux récepteurs transferrin receptor-1 (TfR1).

GP2 est une protéine transmembranaire avec un petit domaine cytosolique contactant l'extrémité C-terminal de SSP. GP2 contient un peptide bipartite en N-terminal qui se retrouve exposé lors d'une modification de pH et qui assure la fusion.

L'interaction de GP1 avec le récepteur, déclenche l'invagination de la membrane et la formation par la cellule d'un endosome autour du virion.

L'acidification de l'endosome est détectée par la boucle de SSP qui va induire le changement de conformation de la protéine de fusion GP2 déclenchant la fusion des membranes et l'injection du complexe ribonucléoprotéique dans le cytoplasme [206].

iii. Le cycle réplicatif.

Du fait, de l'organisation du génome à ARN simple brin ambisens et de polarité (-), le cycle des Arenavirus se déroule en deux phases : précoce et tardive. La phase précoce correspond à :(1) la synthèse des ARNm des protéines NP et L par la protéine L. La synthèse de l'ARNm commence par le vol de coiffe d'ARNm cellulaires qui vont servir d'amorce pour la synthèse d'ARNm viraux. Le mécanisme de recrutement de ces ARNm est encore peu clair, en revanche c'est le domaine endonucléase en N-terminal de L qui génère la coupure et probablement avec l'aide d'autres domaines de L, commence l'initiation de la synthèse de l'ARNm viral. Ainsi les ARNm sont coiffés en 5' et possèdent 5 nucléotides n'appartenant pas à la matrice virale. L'extrémité 3' de l'ARNm n'est pas poly-adényleé, en revanche elles sont stabilisée par une boucle correspondant à la région intergénique du génome. (2) La traduction des ARNm viraux en protéines. Après accumulation de NP, la L commence la (3) la synthèse de l'antigénome et la formation d'un complexe ribonucléoptéique antigénomique. (4) Début de la réplication (Figure 12C) [24].

Pour assurer le bon positionnement de la polymérase pour initier la réplication viral, il faut une contrainte particulière sur les extrémités du génome. Cette contrainte assure en même temps la spécificité de la séquence pour la polymérase au sein de la région conservée des 19 nucléotides en 3' et la conservation de l'intégrité structurale de la structure en queue de poêle formée par l'assemblage complémentaire des extrémités 5' et 3' du génome viral. Une des traces de cette contrainte se retrouve au niveau de la conservation d'un motif GC au niveau de l'extrémité 5' du génome et de l'antigénome. Pour expliquer cette conservation, un mécanisme d'initiation et de réalignement (prime and realign) a été proposé (Figure 13) [220].

Dans ce mécanisme, la réplication du génome des Arenavirus commence par l'initiation avec un GTP en position +2 de la matrice au lieu de la base se situant à l'extrémité 3' du brin d'ARN. Après formation d'un 5'pppGpC, il y a réalignement de l'amorce sur la matrice suivi par son élongation. Ce mécanisme de réarrangement a pour effet de laisser une base non appariées. Cependant comme la longueur du brin d'ARN reste constant cela suppose qu'en 3' une base soit enlevée. Ce mécanisme présume aussi que la polymérase peut initier la synthèse d'ARN *de novo* avec un GTP.

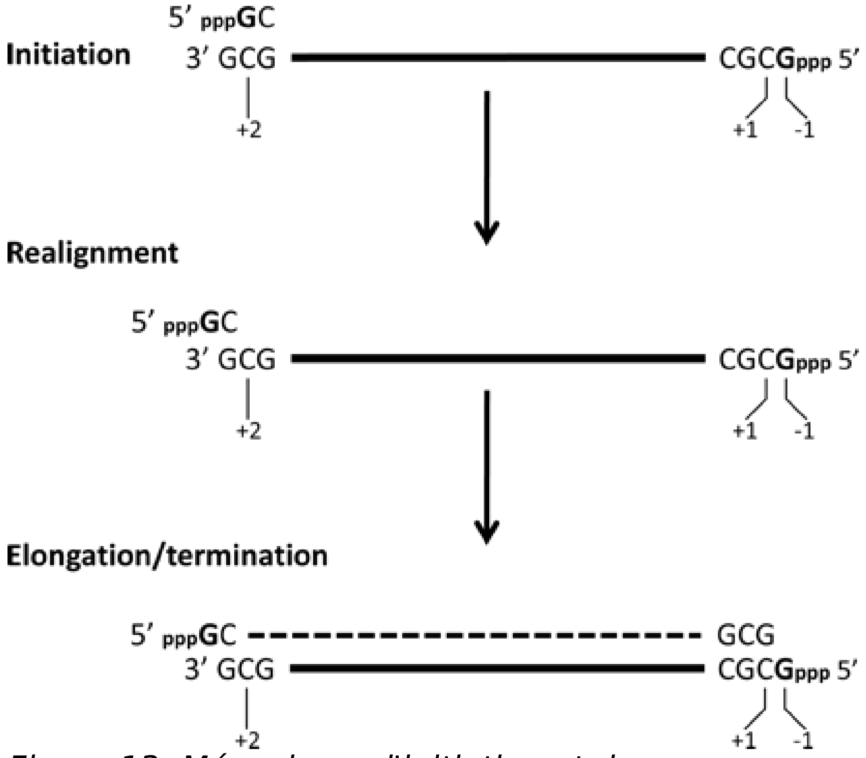


Figure 13: Mécanisme d'initiation et de réalignement de la polymérase. Illustration dérivée de [220]

La phase tardive correspond à : (5) la synthèse des ARNm des protéines Z et GPc par la L à partir du complexe antigénomique. (6) La traduction des ARNm viraux en protéines. Les protéines Z et GPc ne sont pas essentielles à la transcription ni à la réplication.

iv. La réponse Immunitaire innée

Les infections virales sont régulées chez les vertébrés par l'action du système immunitaire qui détecte l'infection virale et initie les réponses antivirales [207–210]. Les mécanismes de l'immunité innée constituent la première ligne de défense et joue un rôle particulièrement important dans le contrôle des infections. Au cours de l'infection précoce, les composés du système de l'immunité innée perçoivent des déterminants viraux appelés « PAMPs » (Pathogen-Associated Molecular Patterns) [211–215]. Parmi les PAMPs, l'ARN db est un signal clé de l'infection par des virus à ARN (-). L'ARN double brin est détecté dans le cytoplasme de la cellule infectée par les récepteurs d'ARN viral double brin comme le récepteur RIG-I (Retinoic acid-inductible gene) et MDA-5 (Melanoma differenciation-associated gene). RIG-I reconnait spécifiquement des ARNb db courts portant une extrémité 5'-triphosphate [216] alors que MDA-5 reconnaît plutôt des ARN db plus longs indépendamment de leur extrémité 5' [217]. Après la reconnaissance de leurs cibles, ces récepteurs activent plusieurs facteurs de transcription (comme IRF-3, IRF-7 et NFκB). IRF-3 et IRF7 sont des facteurs de transcription latents, phosphorylés par les kinases IKKε ou TBK-1 lors de la transduction du signal [218]. IRF-3 et IRF-7 ainsi phosphorylés s'homo- ou s'hétéro-dimérisent et s'accumulent dans le noyau conduisant à la production de l'interféron (IFN) de type I (α et β) et de cytokines pro-inflammatoires activant donc la réponse antivirale de la cellule [219].

Les réponses précoces du système d'immunitaire innée peuvent être capables de contrôler ou ralentir la réplication virale et permettre *a posteriori* à l'immunité adaptative de contrôler le virus et d'empêcher une ré-infection. Les stratégies de survie sont mises en place non seulement par l'hôte mais aussi par les virus. Ainsi, les Arenavirus ont également développé comme nous l'avons vu précédemment, des systèmes qui leurs permettent de bloquer, contourner ou réguler les défenses des cellules hôtes. NP et Z sont les deux protéines les plus impliquées dans ces mécanismes et dérèglent les cascades conduisant à la réponse immunitaire innée.

Cette stratégie de contre, permet aux virus de se répliquer plus tôt et plus vite avant que la réponse de l'immunité adaptative soit mise en place.

B. Projet de recherche à moyen terme

Mon projet vise à acquérir de nouvelles connaissances dans ce domaine en proposant d’une part, une étude structurale pour un certain nombre de virus représentatifs de ces deux familles. Le but est de caractériser au niveau moléculaire, les enzymes impliquées dans la transcription/réplication virale. D’autre part, la mise en place d’une stratégie d’inhibition des mécanismes de la transcription des ARNm viraux, et des enzymes impliquées dans la réponse innée.

Pour cette famille, le virus a développé une série de mécanismes permettant d'échapper à la détection et aux défenses de la cellule hôte. On peut classer ces mécanismes comme moyens de défense « passif », d'autres sont de nature « actif ». Parmi les stratégies d'évasion passive on peut nommer par exemple la formation de structures polymérisées pour cacher l'ARNv, le recrutement de protéines du ribosome près des complexes de réplication afin de favoriser les ARNm viraux, le détournement de kinases perturbant les cascades d'activation, ou l'utilisation d'ARNm utilisant une coiffe cellulaire. Les stratégies actives impliquant une activité enzymatique, sont le vol de la coiffe des ARNm cellulaires, l'activité de dégradation des ARN db marqueur de l'infection virale.

Ainsi, je souhaite caractériser et comprendre les stratégies de protection des ARNs viraux (génomiques, complémentaires et messagers,) acquis ou mis en place par les *Arenaviridae afin d*'échapper à la réponse antivirale de la cellule. Ce projet implique de décrire un modèle mécanistique qui repose sur l'étude structurale et fonctionnelle des protéines impliquées L et NP. Plus spécifiquement se pose les questions des différentes étapes de recrutement des partenaires et leur régulation ? Des changements dynamiques impliquées dans l'assemblage et l'ouverture des NP pour laisser L accéder à l'ARN ? De la régulation des activités nucléases (endo / exo) pour ne pas affecter les ARNs viraux ? Ces questions sont fondamentales dans la compréhension du fonctionnement de la machinerie de transcription/réplication et passent par la :

* Caractérisation biochimique et structurale du mécanisme de recrutement de l'ARNm coiffé par l'endonucléase afin de voler la coiffe, ainsi que de

caractériser le mécanisme qu'utilise la polymérase pour initier la transcription avec cette amorce. J'espère pouvoir valider mes résultats par génétique inverse grâce à notre collaborateur (S. Emonet) de l'unité de virologie de l' Institut de Recherches Biomédicales des Armées.

* Caractérisation de la structure à haute résolution de la NP seul et en complexe avec de l'ARN, de façon à comprendre le mécanisme d'assemblage de ce polymère. A cet effet je développe une stratégie similaire à celle développée avec succès pour la N de RVFv. Il est particulièrement important de comprendre ce processus de polymérisation et décrire les changements de conformations possibles chez les *Arenaviridae* afin de déterminer si la polymérisation induit ou non une activation ou une inhibition du domaine exonucléase.

* Caractérisation de la structure à haute résolution de la NP en complexe avec la kinase IKKε afin de mieux comprendre quel est le mécanisme d'inhibition de la cascade de signalisation. Il sera intéressant d'identifier les résidus impliqués dans cette interaction et d'identifier peut être une surface utile pour développer une stratégie pour bloquer la formation de ce complexe.

* Caractérisation biochimique et structurale des domaines exonucléases de différents Arenavirus (virus CML, Machupo et Mopeïa) ainsi que l'identification d'une classe de molécules inhibant spécifiquement cette activité. Je souhaite vérifier son rôle dans l'inactivation de la réponse à l'interféron. Aussi sur la base de résultats obtenus au laboratoire sur une exonucléase de coronavirus qui est impliquée dans la stabilité du génome et qui présente un comportement similaire face aux substrats ARN ; je souhaite vérifier que cette exonucléase participe à un mécanisme de correction d'erreurs. Grâce à notre collaborateur de l'Institut Pasteur (Dr. Baize) j'espère valider par génétique inverse mes hypothèses chez ces différents virus.

V. REFERENCES

[1] T.D. Pollard, W.C. Earnshaw, Cell Biology, Saunders, 2004. http://books.google.fr/books?id=NrlpAAAAMAAJ.

[2] F. a Rey, W.I. Sundquist, Macromolecular assemblages., Curr. Opin. Struct. Biol. 21 (2011) 221–2. doi:10.1016/j.sbi.2011.03.001.

[3] M. Mueller, S. Jenni, N. Ban, Strategies for crystallization and structure determination of very large macromolecular assemblies., Curr. Opin. Struct. Biol. 17 (2007) 572–9. doi:10.1016/j.sbi.2007.09.004.

[4] N. Ban, E.H. Egelman, Structure and function of large cellular assemblies., Curr. Opin. Struct. Biol. 20 (2010) 207–9. doi:10.1016/j.sbi.2010.02.003.

[5] P. Tompa, D. Kovacs, Intrinsically disordered chaperones in plants and animals., Biochem. Cell Biol. 88 (2010) 167–174. doi:10.1139/o09-163.

[6] N. Pietrosemoli, R. Pancsa, P. Tompa, Structural Disorder Provides Increased Adaptability for Vesicle Trafficking Pathways, PLoS Comput. Biol. 9 (2013). doi:10.1371/journal.pcbi.1003144.

[7] D. Kovacs, B. Szabo, R. Pancsa, P. Tompa, Intrinsically disordered proteins undergo and assist folding transitions in the proteome, Arch. Biochem. Biophys. 531 (2013) 80–89. doi:10.1016/j.abb.2012.09.010.

[8] P. Tompa, P. Csermely, The role of structural disorder in the function of RNA and protein chaperones., FASEB J. 18 (2004) 1169–75. doi:10.1096/fj.04-1584rev.

[9] P. Tompa, Intrinsically disordered proteins: a 10-year recap., Trends Biochem. Sci. 37 (2012) 509–16. doi:10.1016/j.tibs.2012.08.004.

[10] R. Pancsa, P. Tompa, Structural disorder in eukaryotes., PLoS One. 7 (2012) e34687. doi:10.1371/journal.pone.0034687.

[11] K. Namba, Roles of partly unfolded conformations in macromolecular self-assembly., Genes Cells. 6 (2001) 1–12. http://www.ncbi.nlm.nih.gov/pubmed/11168592 (accessed June 21, 2011).

[12] A. Garcia-Pino, S. Balasubramanian, L. Wyns, E. Gazit, H. De Greve, R.D. Magnuson, et al., Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity., Cell. 142 (2010) 101–11. doi:10.1016/j.cell.2010.05.039.

[13] A. Kentsis, R.E. Gordon, K.L.B. Borden, Control of biochemical reactions through supramolecular RING domain self-assembly., Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 15404–9. doi:10.1073/pnas.202608799.

[14] P.J. Kranzusch, S.P.J. Whelan, Arenavirus Z protein controls viral RNA synthesis by locking a polymerase-promoter complex., Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 19743–8. doi:10.1073/pnas.1112742108.

[15] A. Kentsis, R.E. Gordon, K.L.B. Borden, Self-assembly properties of a model RING domain., Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 667–72.

doi:10.1073/pnas.012317299.

[16] D. Karlin, S. Longhi, V. Receveur, B. Canard, The N-terminal domain of the phosphoprotein of Morbilliviruses belongs to the natively unfolded class of proteins., Virology. 296 (2002) 251–62. doi:10.1006/viro.2001.1296.

[17] A. Kentsis, E.C. Dwyer, J.M. Perez, M. Sharma, A. Chen, Z.Q. Pan, et al., The RING domains of the promyelocytic leukemia protein PML and the arenaviral protein Z repress translation by directly inhibiting translation initiation factor eIF4E., J. Mol. Biol. 312 (2001) 609. doi:10.1006/jmbi.2001.5003.

[18] R. Jácamo, N. López, M. Wilda, M.T. Franze-Fernández, R. Jacamo, N. Lopez, et al., Tacaribe virus Z protein interacts with the L polymerase protein to inhibit viral RNA synthesis., J. Virol. 77 (2003) 10383–93. doi:10.1128/JVI.77.19.10383.

[19] S. Fehling, F. Lennartz, T. Strecker, Multifunctional Nature of the Arenavirus RING Finger Protein Z, Viruses. 4 (2012) 2973–3011. doi:10.3390/v4112973.

[20] L. Fan, T. Briese, W.I. Lipkin, Z proteins of New World arenaviruses bind RIG-I and interfere with type I interferon induction., J. Virol. 84 (2010) 1785–91. doi:10.1128/JVI.01362-09.

[21] S. Longhi, V. Receveur-Bréchot, D. Karlin, K. Johansson, H. Darbon, D. Bhella, et al., The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein., J. Biol. Chem. 278 (2003) 18638–48. doi:10.1074/jbc.M300518200.

[22] P. Forterre, The virocell concept and environmental microbiology, ISME J. 7 (2013) 233–236. doi:10.1038/ismej.2012.110.

[23] N.L. Baird, J. York, J.H. Nunberg, Arenavirus infection induces discrete cytosolic structures for RNA replication., J. Virol. 86 (2012) 11301–10. doi:10.1128/JVI.01635-12.

[24] M.J. Buchmeier, J.-C. de la Torre, C.J. Peters, J. De Torre, Arenaviridae: The Viruses and Their Replication, in: D.M. Knipe, P.M. Howley (Eds.), Fields Virol., 5th ed., Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2007: pp. 1791–1827.

[25] I.R. Pedersen, E.P. Konigshofer, Characterization of ribonucleoproteins and ribosomes isolated from lymphocytic choriomeningitis virus., J. Virol. 20 (1976) 14–21.

[26] N. 4 Collaborative Computational Project, The CCP4 suite: programs for protein crystallography., Acta Crystallogr. D. Biol. Crystallogr. 50 (1994) 760–3. doi:10.1107/S0907444994003112.

[27] E.F. Garman, Developments in x-ray crystallographic structure determination of biological macromolecules., Science. 343 (2014) 1102–8. doi:10.1126/science.1247829.

[28] H. van den Bedem, J.S. Fraser, Integrative, dynamic structural biology at atomic resolution—it’s about time, Nat. Methods. 12 (2015) 307–318.

doi:10.1038/nmeth.3324.

[29] T. Sumner, Dazzling History, Science (80-.). 343 (2014) 1092–1093. doi:doi: 10.1126/science.343.6175.1092.

[30] Crystallography at 100, Science. 343 (2014) 1049–1168.

[31] CRYSTALLOGRAPHY, Nat. Milestones. (2014) 1–33.

[32] W.L. Bragg, The Structure of Some Crystals as Indicated by Their Diffraction of X-rays, Proc. R. Soc. A Math. Phys. Eng. Sci. 89 (1913) 248–277. doi:10.1098/rspa.1913.0083.

[33] J.B. Sumner, A.L. Dounce, CRYSTALLINE CATALASE., Science. 85 (1937) 366–367. doi:10.1126/science.85.2206.366.

[34] D.C. HODGKIN, The X-ray analysis of the structure of penicillin., Adv. Sci. 6 (1949) 85–89.

[35] J.D. WATSON, F.H. CRICK, The structure of DNA., Cold Spring Harb. Symp. Quant. Biol. 18 (1953) 123–131. doi:10.1101/SQB.1953.018.01.020.

[36] R.E. FRANKLIN, R.G. GOSLING, Molecular Configuration in Sodium Thymonucleate, Nature. 171 (1953) 740–741. doi:10.1038/171740a0.

[37] R. Olby, Quiet debut for the double helix., Nature. 421 (2003) 402–405. doi:10.1038/nature01397.

[38] B. Maddox, The double helix and the “wronged heroine”., Nature. 421 (2003) 407–408. doi:10.1038/nature01399.

[39] J.C. KENDREW, G. BODO, H.M. DINTZIS, R.G. PARRISH, H. WYCKOFF, D.C. PHILLIPS, A three-dimensional model of the myoglobin molecule obtained by x-ray analysis., Nature. 181 (1958) 662–666. doi:10.1038/181662a0.

[40] M.F. PERUTZ, Relation between structure and sequence of haemoglobin., Nature. 194 (1962) 914–917.

[41] L.N. Johnson, D.C. Phillips, Structure of some crystalline lysozyme-inhibitor complexes determined by X-ray analysis at 6 Angstrom resolution., Nature. 206 (1965) 761–763. doi:10.1038/206761a0.

[42] S.T. Rao, M.G. Rossmann, Comparison of super-secondary structures in proteins., J. Mol. Biol. 76 (1973) 241–256. doi:10.1016/0022-2836(73)90388-4.

[43] S.C. Harrison, A.J. Olson, C.E. Schutt, F.K. Winkler, G. Bricogne, Tomato bushy stunt virus at 2.9 A resolution., Nature. 276 (1978) 368–373. doi:10.1038/276368a0.

[44] T.A. Jones, A graphics model building and refinement system for macromolecules, J. Appl. Crystallogr. 11 (1978) 268–272. doi:10.1107/S0021889878013308.

[45] J.L. Laclare, Target Specifications and Performance of the ESRF Source., J. Synchrotron Radiat. 1 (1994) 12–18. doi:10.1107/S0909049594006564.

[46] J.P. Quintana, M. Hart, D. Bilderback, C. Henderson, D. Richter, T. Setterston, et al., Adaptive Silicon Monochromators for High-Power Insertion Devices. Tests at CHESS, ESRF and HASYLAB., J. Synchrotron

Radiat. 2 (1995) 1–5. doi:10.1107/S090904959400957X.

[47] B. Canard, J.S. Joseph, P. Kuhn, International research networks in viral structural proteomics: Again, lessons from SARS, Antiviral Res. 78 (2008) 47–50. doi:10.1016/j.antiviral.2007.09.007.

[48] I.M. Berry, O. Dym, R.M. Esnouf, K. Harlos, R. Meged, A. Perrakis, et al., SPINE high-throughput crystallization, crystal imaging and recognition techniques: Current state, performance analysis, new technologies and future aspects, Acta Crystallogr. Sect. D Biol. Crystallogr. 62 (2006) 1137–1149. doi:10.1107/S090744490602943X.

[49] A.R. Aricescu, R. Assenberg, R.M. Bill, D. Busso, V.T. Chang, S.J. Davis, et al., Eukaryotic expression: Developments for structural proteomics, Acta Crystallogr. Sect. D Biol. Crystallogr. 62 (2006) 1114–1124. doi:10.1107/S0907444906029805.

[50] P.M. Alzari, H. Berglund, N.S. Berrow, E. Blagova, D. Busso, C. Cambillau, et al., Implementation of semi-automated cloning and prokaryotic expression screening: The impact of SPINE, Acta Crystallogr. Sect. D Biol. Crystallogr. 62 (2006) 1103–1113. doi:10.1107/S0907444906029775.

[51] B. Rupp, Biomolecular crystallography: principles, practice and applications to structural biology., Garland Science, Taylor & Francis Group LLC., Abingdon, New York, 2010. http://scripts.iucr.org/cgi-bin/paper?pf0075\npapers2://publication/uuid/4D6F6384-B11F-4C6E-9560-E1D1E5C49B25.

[52] V. Krishnan, B. Rupp, Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy, eLS. (2012). doi:10.1002/9780470015902.a0002716.pub2.

[53] G. Rhodes, Obtaining Phases, in: Crystallogr. Made Cryst. Clear. 3rd Ed., 3rd ed., Elsevier/Academic Press, 2006: pp. 109–143.

[54] G.N. RAMACHANDRAN, C. RAMAKRISHNAN, V. SASISEKHARAN, Stereochemistry of polypeptide chain configurations., J. Mol. Biol. 7 (1963) 95–99. doi:10.1016/S0022-2836(63)80023-6.

[55] F.M. Buttner, M. Renner-Schneck, T. Stehle, X-ray crystallography and its impact on understanding bacterial cell wall remodeling processes, Int J Med Microbiol. (2014). doi:10.1016/j.ijmm.2014.12.018.

[56] C. Jelsch, M.M. Teeter, V. Lamzin, V. Pichon-Pesme, R.H. Blessing, C. Lecomte, Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin., Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 3171–3176. doi:10.1073/pnas.97.7.3171.

[57] V.S. Lamzin, R.J. Morris, Z. Dauter, K.S. Wilson, M.M. Teeter, Experimental observation of bonding electrons in proteins, J. Biol. Chem. 274 (1999) 20753–20755. doi:10.1074/jbc.274.30.20753.

[58] A. Lwoff, The Concept of Virus, J. Gen. Microbiol. 17 (1957) 239–253.

[59] D. Raoult, P. Forterre, Redefining viruses: lessons from Mimivirus., Nat. Rev. Microbiol. 6 (2008) 315–9. doi:10.1038/nrmicro1858.

[60] B. La Scola, S. Audic, C. Robert, L. Jungang, X. de Lamballerie, M.

Drancourt, et al., A giant virus in amoebae., Science. 299 (2003) 2033. doi:10.1126/science.1081867.

[61] D. Raoult, S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, et al., The 1.2-megabase genome sequence of Mimivirus., Science. 306 (2004) 1344–1350. doi:10.1126/science.1101485.

[62] N. Philippe, M. Legendre, G. Doutre, Y. Couté, O. Poirot, M. Lescot, et al., Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes., Science. 341 (2013) 281–6. doi:10.1126/science.1239181.

[63] C. Abergel, J.-M. Claverie, [Pithovirus sibericum: awakening of a giant virus of more than 30,000 years]., Med. Sci. (Paris). 30 (2014) 329–31. doi:10.1051/medsci/20143003022.

[64] D. Raoult, S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, et al., The 1.2-megabase genome sequence of Mimivirus., Science. 306 (2004) 1344–1350. doi:10.1126/science.1101485.

[65] M.N. Prichard, S. Jairath, M.E. Penfold, S. St Jeor, M.C. Bohlman, G.S. Pari, Identification of persistent RNA-DNA hybrid structures within the origin of replication of human cytomegalovirus., J. Virol. 72 (1998) 6997–7004.

[66] A.L. Ball, Virus Replication Strategies, in: D.M. Knipe, P.M. Howley (Eds.), Fields Virol., 5th ed., Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2007: pp. 120–140.

[67] E. V Koonin, V. V Dolja, Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences., Crit. Rev. Biochem. Mol. Biol. 28 (1993) 375–430. doi:10.3109/10409239309078440.

[68] J. Ortín, J. Martín-benito, The RNA synthesis machinery of negative-stranded RNA viruses, Virology. (2015) 1–13. doi:10.1016/j.virol.2015.03.018.

[69] E.C. Smith, M.R. Denison, Coronaviruses as DNA wannabes: a new model for the regulation of RNA virus replication fidelity., PLoS Pathog. 9 (2013) e1003760. doi:10.1371/journal.ppat.1003760.

[70] M.C. Freeman, R.L. Graham, X. Lu, C.T. Peek, M.R. Denison, Coronavirus replicase-reporter fusions provide quantitative analysis of replication and replication complex formation., J. Virol. 88 (2014) 5319–27. doi:10.1128/JVI.00021-14.

[71] M.R. Denison, R.L. Graham, E.F. Donaldson, L.D. Eckerle, R.S. Baric, Coronaviruses: An RNA proofreading machine regulates replication fidelity and diversity, RNA Biol. 8 (2011) 270–279. doi:10.4161/rna.8.2.15013.

[72] P.J. Hussey, T. Ketelaar, M.J. Deeks, Control of the actin cytoskeleton in plant cell growth., Annu. Rev. Plant Biol. 57 (2006) 109–25. doi:10.1146/annurev.arplant.57.032905.105206.

[73] G. Itoh, S. Yumura, A novel mitosis-specific dynamic actin structure in Dictyostelium cells., J. Cell Sci. 120 (2007) 4302–9. doi:10.1242/jcs.015875.

[74] H.T. McMahon, J.L. Gallop, Membrane curvature and mechanisms of dynamic cell membrane remodelling., Nature. 438 (2005) 590–6. doi:10.1038/nature04396.

[75] A. Puppo, J.T. Chun, G. Gragnaniello, E. Garante, L. Santella, Alteration of the cortical actin cytoskeleton deregulates Ca2+ signaling, monospermic fertilization, and sperm entry., PLoS One. 3 (2008) e3588. doi:10.1371/journal.pone.0003588.

[76] H. Yamaguchi, J. Condeelis, Regulation of the actin cytoskeleton in cancer cell migration and invasion., Biochim. Biophys. Acta. 1773 (2007) 642–52. doi:10.1016/j.bbamcr.2006.07.001.

[77] M. Bailly, J. Condeelis, Cell motility: insights from the backstage., Nat. Cell Biol. 4 (2002) E292–4. doi:10.1038/ncb1202-e292.

[78] P. Graceffa, R. Dominguez, Crystal structure of monomeric actin in the ATP state. Structural basis of nucleotide-dependent actin dynamics., J. Biol. Chem. 278 (2003) 34172–80. doi:10.1074/jbc.M303689200.

[79] K.C. Holmes, D. Popp, W. Gebhard, W. Kabsch, Atomic model of the actin filament., Nature. 347 (1990) 44–9. doi:10.1038/347044a0.

[80] L.R. Otterbein, P. Graceffa, R. Dominguez, The crystal structure of uncomplexed actin in the ADP state., Science. 293 (2001) 708–11. doi:10.1126/science.1059700.

[81] W. Kabsch, H.G. Mannherz, D. Suck, E.F. Pai, K.C. Holmes, Atomic structure of the actin:DNase I complex., Nature. 347 (1990) 37–44. doi:10.1038/347037a0.

[82] R. Dominguez, K.C. Holmes, Actin structure and function., Annu. Rev. Biophys. 40 (2011) 169–86. doi:10.1146/annurev-biophys-042910-155359.

[83] T.D. Pollard, L. Blanchoin, R.D. Mullins, Molecular mechanisms controlling actin filament dynamics in nonmuscle cells., Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 545–76. doi:10.1146/annurev.biophys.29.1.545.

[84] L.D. Belmont, A. Orlova, D.G. Drubin, E.H. Egelman, A change in actin conformation associated with filament instability after Pi release., Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 29–34. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=15087&tool=pmcentrez&rendertype=abstract (accessed November 24, 2013).

[85] E.H. Egelman, Actin Filament Structure: The ghost of ribbons past, Curr. Biol. 4 (1994) 79–81. doi:10.1016/S0960-9822(00)00020-8.

[86] D. Scoville, J.D. Stamm, D. Toledo-Warshaviak, C. Altenbach, M. Phillips, A. Shvetsov, et al., Hydrophobic loop dynamics and actin filament stability., Biochemistry. 45 (2006) 13576–84. doi:10.1021/bi061229f.

[87] C.E. Schutt, J.C. Myslik, M.D. Rozycki, N.C. Goonesekere, U. Lindberg, The structure of crystalline profilin-beta-actin., Nature. 365 (1993) 810–6. doi:10.1038/365810a0.

[88] T.D. Pollard, G.G. Borisy, Cellular Motility Driven by Assembly and

Disassembly of Actin Filaments, Cell. 112 (2003) 453–465. doi:10.1016/S0092-8674(03)00120-X.

[89] M. Raftopoulou, A. Hall, Cell migration: Rho GTPases lead the way., Dev. Biol. 265 (2004) 23–32. http://www.ncbi.nlm.nih.gov/pubmed/14697350 (accessed November 13, 2013).

[90] S.M. Rafelski, J.A. Theriot, Crawling toward a unified model of cell mobility: spatial and temporal regulation of actin dynamics., Annu. Rev. Biochem. 73 (2004) 209–39. doi:10.1146/annurev.biochem.73.011303.073844.

[91] S.H. Zigmond, Beginning and ending an actin filament: control at the barbed end., Curr. Top. Dev. Biol. 63 (2004) 145–88. doi:10.1016/S0070-2153(04)63005-5.

[92] R. Foster, K.Q. Hu, Y. Lu, K.M. Nolan, J. Thissen, J. Settleman, Identification of a novel human Rho protein with unusual properties: GTPase deficiency and in vivo farnesylation., Mol. Cell. Biol. 16 (1996) 2689–99. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=231259&tool=pmcentrez&rendertype=abstract (accessed November 24, 2013).

[93] A. Hall, Small GTP-binding proteins and the regulation of the actin cytoskeleton., Annu. Rev. Cell Biol. 10 (1994) 31–54. doi:10.1146/annurev.cb.10.110194.000335.

[94] A.J. Ridley, A. Hall, The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors., Cell. 70 (1992) 389–99. http://www.ncbi.nlm.nih.gov/pubmed/1643657 (accessed November 26, 2013).

[95] A.J. Ridley, H.F. Paterson, C.L. Johnston, D. Diekmann, A. Hall, The small GTP-binding protein rac regulates growth factor-induced membrane ruffling, Cell. 70 (1992) 401–410. doi:10.1016/0092-8674(92)90164-8.

[96] R. Kozma, S. Ahmed, A. Best, L. Lim, The Ras-related protein Cdc42Hs and bradykinin promote formation of peripheral actin microspikes and filopodia in Swiss 3T3 fibroblasts., Mol. Cell. Biol. 15 (1995) 1942–52. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=230420&tool=pmcentrez&rendertype=abstract (accessed November 24, 2013).

[97] C.D. Nobes, A. Hall, Rho GTPases control polarity, protrusion, and adhesion during cell movement., J. Cell Biol. 144 (1999) 1235–44. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2150589&tool=pmcentrez&rendertype=abstract (accessed November 24, 2013).

[98] I. Mabuchi, Y. Hamaguchi, H. Fujimoto, N. Morii, M. Mishima, S. Narumiya, A rho-like protein is involved in the organisation of the contractile ring in dividing sand dollar eggs., Zygote. 1 (1993) 325–31. http://www.ncbi.nlm.nih.gov/pubmed/8081830 (accessed November 24, 2013).

[99] S.N. Prokopenko, R. Saint, H.J. Bellen, Untying the Gordian knot of

cytokinesis. Role of small G proteins and their regulators., J. Cell Biol. 148 (2000) 843–8. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2174545&tool=pmcentrez&rendertype=abstract (accessed November 24, 2013).

[100] E. Caron, A. Hall, Identification of two distinct mechanisms of phagocytosis controlled by different Rho GTPases., Science. 282 (1998) 1717–21. http://www.ncbi.nlm.nih.gov/pubmed/9831565 (accessed November 17, 2013).

[101] L. Luo, L.Y. Jan, Y.-N. Jan, Rho family small GTP-binding proteins in growth cone signalling, Curr. Opin. Neurobiol. 7 (1997) 81–86. doi:10.1016/S0959-4388(97)80124-9.

[102] Y. Lu, J. Settleman, The role of rho family GTPases in development: lessons from Drosophila melanogaster., Mol. Cell Biol. Res. Commun. 1 (1999) 87–94. doi:10.1006/mcbr.1999.0119.

[103] J. Settleman, Rho GTPases in development., Prog. Mol. Subcell. Biol. 22 (1999) 201–29. http://www.ncbi.nlm.nih.gov/pubmed/10081071 (accessed November 24, 2013).

[104] P.D. Burbelo, D. Drechsel, A. Hall, A conserved binding motif defines numerous candidate target proteins for both Cdc42 and Rac GTPases., J. Biol. Chem. 270 (1995) 29071–4. http://www.ncbi.nlm.nih.gov/pubmed/7493928 (accessed November 24, 2013).

[105] S.H. Zigmond, Formin-induced nucleation of actin filaments., Curr. Opin. Cell Biol. 16 (2004) 99–105. doi:10.1016/j.ceb.2003.10.019.

[106] A. Yamagishi, M. Masuda, T. Ohki, H. Onishi, N. Mochizuki, A novel actin bundling/filopodium-forming domain conserved in insulin receptor tyrosine kinase substrate p53 and missing in metastasis protein., J. Biol. Chem. 279 (2004) 14929–36. doi:10.1074/jbc.M309408200.

[107] P. Mangeat, K. Burridge, Actin-membrane interaction in fibroblasts: what proteins are involved in this association?, J. Cell Biol. 99 (1984) 95s–103s. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2275575&tool=pmcentrez&rendertype=abstract.

[108] M. Masuda, N. Mochizuki, Structural characteristics of BAR domain superfamily to sculpt the membrane., Semin. Cell Dev. Biol. 21 (2010) 391–8. doi:10.1016/j.semcdb.2010.01.010.

[109] J.T. Trachtenberg, B.E. Chen, G.W. Knott, G. Feng, J.R. Sanes, E. Welker, et al., Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex., Nature. 420 (n.d.) 788–94. doi:10.1038/nature01273.

[110] J.W. Copeland, R. Treisman, The diaphanous-related formin mDia1 controls serum response factor activity through its effects on actin polymerization., Mol. Biol. Cell. 13 (2002) 4088–99. doi:10.1091/mbc.02-06-0092.

[111] F. Li, H.N. Higgs, The mouse Formin mDia1 is a potent actin nucleation factor regulated by autoinhibition., Curr. Biol. 13 (2003) 1335–40.

http://www.ncbi.nlm.nih.gov/pubmed/12906795 (accessed November 23, 2013).

[112] X. Chen, F. Ni, X. Tian, E. Kondrashkina, Q. Wang, J. Ma, Structural basis of actin filament nucleation by tandem W domains., Cell Rep. 3 (2013) 1910–20. doi:10.1016/j.celrep.2013.04.028.

[113] A.D.B. Liverman, H.-C. Cheng, J.E. Trosky, D.W. Leung, M.L. Yarbrough, D.L. Burdette, et al., Arp2/3-independent assembly of actin by Vibrio type III effector VopL., Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 17117–22. doi:10.1073/pnas.0703196104.

[114] D. Chereau, M. Boczkowska, A. Skwarek-Maruszewska, I. Fujiwara, D.B. Hayes, G. Rebowski, et al., Leiomodin is an actin filament nucleator in muscle cells., Science. 320 (2008) 239–43. doi:10.1126/science.1155313.

[115] R. Ahuja, R. Pinyol, N. Reichenbach, L. Custer, J. Klingensmith, M.M. Kessels, et al., Cordon-bleu is an actin nucleation factor and controls neuronal morphology., Cell. 131 (2007) 337–50. doi:10.1016/j.cell.2007.08.030.

[116] M.E. Quinlan, J.E. Heuser, E. Kerkhoff, R.D. Mullins, Drosophila Spire is an actin nucleation factor., Nature. 433 (2005) 382–8. doi:10.1038/nature03241.

[117] L. Blanchoin, K.J. Amann, H.N. Higgs, J.B. Marchand, D.A. Kaiser, T.D. Pollard, Direct observation of dendritic actin filament networks nucleated by Arp2/3 complex and WASP/Scar proteins., Nature. 404 (2000) 1007–11. doi:10.1038/35010008.

[118] R.D. Mullins, J.A. Heuser, T.D. Pollard, The interaction of Arp2/3 complex with actin: nucleation, high affinity pointed end capping, and formation of branching networks of filaments., Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 6181–6. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=27619&tool=pmcentrez&rendertype=abstract (accessed November 23, 2013).

[119] F. Castellano, P. Montcourrier, J.C. Guillemot, E. Gouin, L. Machesky, P. Cossart, et al., Inducible recruitment of Cdc42 or WASP to a cell-surface receptor triggers actin polymerization and filopodium formation., Curr. Biol. 9 (1999) 351–60. http://www.ncbi.nlm.nih.gov/pubmed/10209117 (accessed November 23, 2013).

[120] M. Geese, J.J. Loureiro, J.E. Bear, J. Wehland, F.B. Gertler, A.S. Sechi, Contribution of Ena/VASP proteins to intracellular motility of listeria requires phosphorylation and proline-rich core but not F-actin binding or multimerization., Mol. Biol. Cell. 13 (2002) 2383–96. doi:10.1091/mbc.E02-01-0058.

[121] J.E. Bear, T.M. Svitkina, M. Krause, D.A. Schafer, J.J. Loureiro, G.A. Strasser, et al., Antagonism between Ena/VASP proteins and actin filament capping regulates fibroblast motility., Cell. 109 (2002) 509–21. http://www.ncbi.nlm.nih.gov/pubmed/12086607 (accessed November 23, 2013).

[122] J.E. Caldwell, S.G. Heiss, V. Mermall, J.A. Cooper, Effects of CapZ, an actin

capping protein of muscle, on the polymerization of actin., Biochemistry. 28 (1989) 8506–14. http://www.ncbi.nlm.nih.gov/pubmed/2557904 (accessed November 23, 2013).

[123] G. Scita, J. Nordstrom, R. Carbone, P. Tenca, G. Giardina, S. Gutkind, et al., EPS8 and E3B1 transduce signals from Ras to Rac., Nature. 401 (1999) 290–3. doi:10.1038/45822.

[124] W. Morishita, H. Marie, R.C. Malenka, Distinct triggering and expression mechanisms underlie LTD of AMPA and NMDA synaptic responses., Nat. Neurosci. 8 (2005) 1043–50. doi:10.1038/nn1506.

[125] A. Disanza, M.-F. Carlier, T.E.B. Stradal, D. Didry, E. Frittoli, S. Confalonieri, et al., Eps8 controls actin-based motility by capping the barbed ends of actin filaments., Nat. Cell Biol. 6 (2004) 1180–8. doi:10.1038/ncb1199.

[126] A. Weber, C.R. Pennise, G.G. Babcock, V.M. Fowler, Tropomodulin caps the pointed ends of actin filaments., J. Cell Biol. 127 (1994) 1627–35. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2120308&tool=pmcentrez&rendertype=abstract (accessed November 23, 2013).

[127] K.-I. Okamoto, R. Narayanan, S.H. Lee, K. Murata, Y. Hayashi, The role of CaMKII as an F-actin-bundling protein crucial for maintenance of dendritic spine structure., Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 6418–23. doi:10.1073/pnas.0701656104.

[128] S. Raghavachari, J.E. Lisman, Properties of quantal transmission at CA1 synapses., J. Neurophysiol. 92 (2004) 2456–67. doi:10.1152/jn.00258.2004.

[129] A. Weber, V.T. Nachmias, C.R. Pennise, M. Pring, D. Safer, Interaction of thymosin beta 4 with muscle and platelet actin: implications for actin sequestration in resting platelets., Biochemistry. 31 (1992) 6179–85. http://www.ncbi.nlm.nih.gov/pubmed/1627561 (accessed November 23, 2013).

[130] W. Witke, The role of profilin complexes in cell motility and other cellular processes., Trends Cell Biol. 14 (2004) 461–9. doi:10.1016/j.tcb.2004.07.003.

[131] Z. Parnass, A. Tashiro, R. Yuste, Analysis of spine morphological plasticity in developing hippocampal pyramidal neurons., Hippocampus. 10 (2000) 561–8. doi:10.1002/1098-1063(2000)10:5<561::AID-HIPO6>3.0.CO;2-X.

[132] J. Grutzendler, N. Kasthuri, W.-B. Gan, Long-term dendritic spine stability in the adult cortex., Nature. 420 (2002) 812–6. doi:10.1038/nature01276.

[133] M. Maletic-Savatic, R. Malinow, K. Svoboda, Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity., Science. 283 (1999) 1923–7. http://www.ncbi.nlm.nih.gov/pubmed/10082466 (accessed November 23, 2013).

[134] R.A. Edwards, J. Bryan, Fascins, a family of actin bundling proteins., Cell Motil. Cytoskeleton. 32 (1995) 1–9. doi:10.1002/cm.970320102.

[135] B.W. Bernstein, J.R. Bamburg, Tropomyosin binding to F-actin protects the F-actin from disassembly by brain actin-depolymerizing factor (ADF)., Cell Motil. 2 (1982) 1–8. http://www.ncbi.nlm.nih.gov/pubmed/6890875 (accessed November 23, 2013).

[136] J.C. Pinder, E. Ungewickell, D. Bray, W.B. Gratzer, The spectrin-actin complex and erythrocyte shape., J. Supramol. Struct. 8 (1978) 439–45. doi:10.1002/jss.400080406.

[137] S.R. Goodman, D. Branton, Spectrin binding and the control of membrane protein mobility., J. Supramol. Struct. 8 (1978) 455–63. doi:10.1002/jss.400080408.

[138] R. Dominguez, Actin-binding proteins--a unifying hypothesis., Trends Biochem. Sci. 29 (2004) 572–8. doi:10.1016/j.tibs.2004.09.004.

[139] R. Dominguez, The beta-thymosin/WH2 fold: multifunctionality and structure., Ann. N. Y. Acad. Sci. 1112 (2007) 86–94. doi:10.1196/annals.1415.011.

[140] D. Chereau, F. Kerff, P. Graceffa, Z. Grabarek, K. Langsetmo, R. Dominguez, Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly., Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 16644–9. doi:10.1073/pnas.0507021102.

[141] S.H. Lee, R. Dominguez, Regulation of actin cytoskeleton dynamics in cells., Mol. Cells. (2010) 311–325. doi:10.1007/s10059-010-0053-8.

[142] P. Lappalainen, M.M. Kessels, M.J. Cope, D.G. Drubin, The ADF homology (ADF-H) domain: a highly exploited actin-binding module., Mol. Biol. Cell. 9 (1998) 1951–9. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=25446&tool=pmcentrez&rendertype=abstract (accessed November 24, 2013).

[143] E. Paunola, P.K. Mattila, P. Lappalainen, WH2 domain: a small, versatile adapter for actin monomers, FEBS Lett. 513 (2002) 92–97. doi:10.1016/S0014-5793(01)03242-2.

[144] A.M. McGough, C.J. Staiger, J.-K. Min, K.D. Simonetti, The gelsolin family of actin regulatory proteins: modular structures, versatile functions, FEBS Lett. 552 (2003) 75–81. doi:10.1016/S0014-5793(03)00932-3.

[145] M. Gimona, K. Djinovic-Carugo, W.J. Kranewitter, S.J. Winder, Functional plasticity of CH domains, FEBS Lett. 513 (2002) 98–106. doi:10.1016/S0014-5793(01)03240-9.

[146] B.L. Goode, M.J. Eck, Mechanism and function of formins in the control of actin assembly., Annu. Rev. Biochem. 76 (2007) 593–627. doi:10.1146/annurev.biochem.75.103004.142647.

[147] J.R. Sellers, Myosins: a diverse superfamily., Biochim. Biophys. Acta. 1496 (2000) 3–22. doi:10.1016/S0167-4889(00)00005-7.

[148] E.G. Yarmola, S. Parikh, M.R. Bubb, Formation and implications of a ternary complex of profilin, thymosin beta 4, and actin., J. Biol. Chem. 276 (2001) 45555–63. doi:10.1074/jbc.M105723200.

[149] S.C. Mockrin, E.D. Korn, Acanthamoeba profilin interacts with G-actin to increase the rate of exchange of actin-bound adenosine 5’-triphosphate., Biochemistry. 19 (1980) 5359–62. http://www.ncbi.nlm.nih.gov/pubmed/6893804 (accessed November 26, 2013).

[150] E. Nishida, Opposite effects of cofilin and profilin from porcine brain on rate of exchange of actin-bound adenosine 5’-triphosphate., Biochemistry. 24 (1985) 1160–4. http://www.ncbi.nlm.nih.gov/pubmed/4096896 (accessed November 26, 2013).

[151] F. Ferron, G. Rebowski, S.H. Lee, R. Dominguez, Structural basis for the recruitment of profilin-actin complexes during filament elongation by Ena/VASP., EMBO J. 26 (2007) 4597–606. doi:10.1038/sj.emboj.7601874.

[152] S.H. Lee, F. Kerff, D. Chereau, F. Ferron, A. Klug, R. Dominguez, Structural basis for the actin-binding function of missing-in-metastasis., Structure. 15 (2007) 145–55. doi:10.1016/j.str.2006.12.005.

[153] F. Ferron, S. Longhi, B. Canard, D. Karlin, A practical overview of protein disorder prediction methods., Proteins. 65 (2006) 1–14. doi:10.1002/prot.21075.

[154] S. Grenklo, M. Geese, U. Lindberg, J. Wehland, R. Karlsson, A.S. Sechi, A crucial role for profilin-actin in the intracellular motility of Listeria monocytogenes., EMBO Rep. 4 (2003) 523–9. doi:10.1038/sj.embor.embor823.

[155] A.E.Y. Engqvist-Goldstein, D.G. Drubin, Actin assembly and endocytosis: from yeast to mammals., Annu. Rev. Cell Dev. Biol. 19 (2003) 287–332. doi:10.1146/annurev.cellbio.19.111401.093127.

[156] M. Kaksonen, C.P. Toret, D.G. Drubin, Harnessing actin dynamics for clathrin-mediated endocytosis., Nat. Rev. Mol. Cell Biol. 7 (2006) 404–14. doi:10.1038/nrm1940.

[157] G. Scita, S. Confalonieri, P. Lappalainen, S. Suetsugu, IRSp53: crossing the road of membrane and actin dynamics in the formation of membrane protrusions., Trends Cell Biol. 18 (2008) 52–60. doi:10.1016/j.tcb.2007.12.002.

[158] A. Frost, V.M. Unger, P. De Camilli, The BAR domain superfamily: membrane-molding macromolecules., Cell. 137 (2009) 191–6. doi:10.1016/j.cell.2009.04.010.

[159] J.C. Dawson, J.A. Legg, L.M. Machesky, Bar domain proteins: a role in tubulation, scission and actin assembly in clathrin-mediated endocytosis., Trends Cell Biol. 16 (2006) 493–8. doi:10.1016/j.tcb.2006.08.004.

[160] A. Frost, R. Perera, A. Roux, K. Spasov, O. Destaing, E.H. Egelman, et al., Structural basis of membrane invagination by F-BAR domains., Cell. 132 (2008) 807–17. doi:10.1016/j.cell.2007.12.041.

[161] T. Itoh, P. De Camilli, BAR, F-BAR (EFC) and ENTH/ANTH domains in the regulation of membrane-cytosol interfaces and membrane curvature., Biochim. Biophys. Acta. 1761 (2006) 897–912.

doi:10.1016/j.bbalip.2006.06.015.

[162] W.M. Henne, H.M. Kent, M.G.J. Ford, B.G. Hegde, O. Daumke, P.J.G. Butler, et al., Structure and analysis of FCHo2 F-BAR domain: a dimerizing and membrane recruitment module that effects membrane curvature., Structure. 15 (2007) 839–52. doi:10.1016/j.str.2007.05.002.

[163] A. Shimada, H. Niwa, K. Tsujita, S. Suetsugu, K. Nitta, K. Hanawa-Suetsugu, et al., Curved EFC/F-BAR-domain dimers are joined end to end into a filament for membrane invagination in endocytosis., Cell. 129 (2007) 761–72. doi:10.1016/j.cell.2007.03.040.

[164] B.J. Peter, H.M. Kent, I.G. Mills, Y. Vallis, P.J.G. Butler, P.R. Evans, et al., BAR domains as sensors of membrane curvature: the amphiphysin BAR structure., Science. 303 (2004) 495–9. doi:10.1126/science.1092586.

[165] B. Habermann, The BAR-domain family of proteins: a case of bending and binding?, EMBO Rep. 5 (2004) 250–5. doi:10.1038/sj.embor.7400105.

[166] P. Little, T. Teka, A. Azeze, Cross-Border Livestock Trade and Food Security in the Horn of Africa: An Overview, Washington, D.C., 2001.

[167] C. Schmaljohn, J.W. Hooper, Bunyaviridae: the viruses and their replication, in: D.M. Knipe, P.M. Howley, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman, et al. (Eds.), Fields Virol., 4th ed., Lippincott, Williams and Wilkins, Philadelphia, Pa., 2001: pp. 1581–1602. http://lww.com.

[168] R. Elliott, Molecular biology of the Bunyaviridae, J. Gen. Virol. 71 (1990) 501–522. http://vir.sgmjournals.org/cgi/content/abstract/71/3/501 (accessed August 3, 2010).

[169] J.N.J.N. Barr, G.W.G.W. Wertz, Role of the conserved nucleotide mismatch within 3’- and 5'-terminal regions of Bunyamwera virus in signaling transcription, J. Virol. 79 (2005) 3586–3594. doi:10.1128/JVI.79.6.3586.

[170] M. Delarue, O. Poch, N. Tordo, D. Moras, P. Argos, An attempt to unify the structure of polymerases., Protein Eng. 3 (1990) 461–7. http://www.ncbi.nlm.nih.gov/pubmed/2196557 (accessed November 28, 2013).

[171] J. a. Bruenn, A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases, Nucleic Acids Res. 31 (2003) 1821–1829. doi:10.1093/nar/gkg277.

[172] J.L. Patterson, B. Holloway, D. Kolakofsky, La Crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependent endonuclease., J. Virol. 52 (1984) 215–22. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=254508&tool=pmcentrez&rendertype=abstract (accessed November 28, 2013).

[173] D.H.M.Y. Bishop, M.E. Gay, Y. Matsuoko, Non Viral heterogeneous sequences are present at the 5’ ends of one species of snowshoe hare bunyavirus S complementary RNA, Nucleic Acids Res. 11 (1983) 6409 – 6418. doi:10.1093/nar/gkn942.

[174] J. Reguera, F. Weber, S. Cusack, Bunyaviridae RNA polymerases (L-

protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription., PLoS Pathog. 6 (2010). doi:10.1371/journal.ppat.1001101.

[175] R.F. Pettersson, C.H. von Bonsdorff, Ribonucleoproteins of Uukuniemi virus are circular., J. Virol. 15 (1975) 386–92. http://www.ncbi.nlm.nih.gov/pubmed/1167604.

[176] C. Peters, K. Linthicum, Rift Valley fever, Handb. Zoonoses, Sect. B. Viral, 2nd Ed. CRC Press. Boca Raton, Fla. (1994) 125–138. http://books.google.com/books?hl=en&amp;lr=&amp;id=hDxeXDisYKEC&amp;oi=fnd&amp;pg=PA125&amp;dq=Rift+Valley+fever&amp;ots=cW9K7plGKt&amp;sig=TKCrWf-fXAmMEYAiIN7w59TzaEE.

[177] a a Sall, P.M. Zanotto, P. Vialat, O.K. Sène, M. Bouloy, Molecular epidemiology and emergence of Rift Valley fever., Mem. Inst. Oswaldo Cruz. 93 (1998) 609–14. http://www.ncbi.nlm.nih.gov/pubmed/9830526 (accessed July 26, 2011).

[178] L. de Spiez, Fièvre de la vallée du Rift, Spiez, 2006. www.labor-spiez.ch.

[179] C.F.D. Control, Prevention, Update: outbreak of Rift Valley fever, 2000.

[180] Rift Valley Fever could spread with movement of animals from East Africa, 2007. http://www.fao.org/DOCREP/006/Y4611E/Y4611E00.HTM.

[181] Ungerer, D. Klerk, Prinsloo, Rift Valley fever virus : An evaluation of the outbreaks in South africa, Vet. Res. 41 (n.d.) 19. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2896810&tool=pmcentrez&rendertype=abstract.

[182] D. Sissoko, Rift Valley Fever, Mayotte, 2007–2008, Emerg. Infect. Dis. 15 (2009) 568–570. doi:10.3201/eid1504.081045.

[183] G.E. Sciences, Système sous régional d’alerte et de contrôle de la fièvre de la vallée du Rift en Afrique de l'Ouest., 2005.

[184] M. Pépin, M. Bouloy, B.H. Bird, A. Kemp, J. Paweska, M. Pepin, Rift Valley fever virus (Bunyaviridae: Phlebovirus): an update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention., Vet. Res. 41 (2010) 61. doi:10.1051/vetres/2010033.

[185] C.F.D. Control, Prevention, Outbreak of Rift Valley fever, Yemen, August-October. 49 (2000) 1065–1066.

[186] N. Le May, N. Gauliard, The N Terminus of Rift Valley Fever Virus Nucleoprotein Is Essential for Dimerization, J. Virol. 79 (2005) 11974–11980. doi:10.1128/JVI.79.18.11974.

[187] L. Brunotte, R. Kerber, W. Shang, F. Hauer, M. Hass, M. Gabriel, et al., Structure of the Lassa virus nucleoprotein revealed by X-ray crystallography, small-angle X-ray scattering, and electron microscopy., J. Biol. Chem. 286 (2011) 38748–56. doi:10.1074/jbc.M111.278838.

[188] X. Qi, S. Lan, W. Wang, L.M. Schelde, H. Dong, G.D. Wallat, et al., Cap binding and immune evasion revealed by Lassa nucleoprotein structure., Nature. 468 (2010) 779–83. doi:10.1038/nature09605.

[189] E. Ortiz-Riaño, B.Y.H. Cheng, J.C. de la Torre, L. Martínez-Sobrido, Self-association of Lymphocytic Choriomeningitis Virus Nucleoprotein is mediated by its N-terminal region and is not required for its anti-interferon function., J. Virol. (2012) JVI.05503–11–. doi:10.1128/JVI.05503-11.

[190] M.-P. Egloff, E. Decroly, H. Malet, B. Selisko, D. Benarroch, F. Ferron, et al., Structural and functional analysis of methylation and 5’-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5., J. Mol. Biol. 372 (2007) 723–36. doi:10.1016/j.jmb.2007.07.005.

[191] E. Decroly, C. Debarnot, F. Ferron, M. Bouvet, B. Coutard, I. Imbert, et al., Crystal Structure and Functional Analysis of the SARS-Coronavirus RNA Cap 2’-O-Methyltransferase nsp10/nsp16 Complex., PLoS Pathog. 7 (2011) e1002059. doi:10.1371/journal.ppat.1002059.

[192] B. Morin, B. Coutard, M. Lelke, F. Ferron, R. Kerber, S. Jamal, et al., The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription., PLoS Pathog. 6 (2010) e1001038. doi:10.1371/journal.ppat.1001038.

[193] A.D. Davidson, Chapter 2. New insights into flavivirus nonstructural protein 5., in: Adv. Virus Res., 1st ed., Elsevier Inc., 2009: pp. 41–101. doi:10.1016/S0065-3527(09)74002-3.

[194] L.J. Yap, D. Luo, K.Y. Chung, S.P. Lim, C. Bodenreider, C. Noble, et al., Crystal structure of the dengue virus methyltransferase bound to a 5’-capped octameric RNA., PLoS One. 5 (2010). http://www.ncbi.nlm.nih.gov/pubmed/20862256 (accessed September 29, 2010).

[195] B. Hijawi, M. Abdallat, A. Sayaydeh, S. Alqasrawi, A. Haddadin, N. Jaarour, et al., Novel coronavirus infections in Jordan, April 2012: epidemiological findings from a retrospective investigation., East. Mediterr. Health J. 19 Suppl 1 (2013) S12–8. doi:23888790.

[196] K.C. Kronmann, S. Nimo-Paintsil, F. Guirguis, L.C. Kronmann, K. Bonney, K. Obiri-Danso, et al., Two novel arenaviruses detected in pygmy mice, Ghana., Emerg. Infect. Dis. 19 (2013) 1832–5. doi:10.3201/eid1911.121491.

[197] R.N. Charrel, X. De Lamballerie, Arenaviruses other than Lassa virus, Antiviral Res. 57 (2003) 89–100.

[198] J.M. Levingston Macleod, A. D’Antuono, M.E. Loureiro, J.C. Casabona, G. a Gomez, N. Lopez, Identification of two functional domains within the arenavirus nucleoprotein., J. Virol. 85 (2011) 2012–23. doi:10.1128/JVI.01875-10.

[199] K.M. Hastie, C.R. Kimberlin, M.A. Zandonatti, I.J. MacRae, E.O. Saphire, Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3’ to 5' exonuclease activity essential for immune suppression., Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 2396–401. doi:10.1073/pnas.1016404108.

[200] L. Martínez-Sobrido, S. Emonet, P. Giannakas, B. Cubitt, A. García-Sastre, J.C. de la Torre, Identification of amino acid residues critical for the anti-

interferon activity of the nucleoprotein of the prototypic arenavirus lymphocytic choriomeningitis virus., J. Virol. 83 (2009) 11330–11340. doi:10.1128/JVI.00763-09.

[201] M. Russier, S. Reynard, N. Tordo, S. Baize, NK cells are strongly activated by Lassa and Mopeia virus-infected human macrophages in vitro but do not mediate virus suppression., Eur. J. Immunol. 42 (2012) 1822–32. doi:10.1002/eji.201142099.

[202] D. a Moshkoff, M.S. Salvato, I.S. Lukashevich, Molecular characterization of a reassortant virus derived from Lassa and Mopeia viruses., Virus Genes. 34 (2007) 169–76. doi:10.1007/s11262-006-0050-3.

[203] D. Pannetier, S. Reynard, M. Russier, A. Journeaux, N. Tordo, V. Deubel, et al., Human dendritic cells infected with the nonpathogenic Mopeia virus induce stronger T-cell responses than those infected with Lassa virus., J. Virol. 85 (2011) 8293–306. doi:10.1128/JVI.02120-10.

[204] C. Pythoud, W.W.S.I. Rodrigo, G. Pasqual, S. Rothenberger, L. Martínez-Sobrido, J.C. de la Torre, et al., Arenavirus nucleoprotein targets interferon regulatory factor-activating kinase IKK{varepsilon}, J. Virol. 86 (2012) 7728–38. doi:10.1128/JVI.00187-12.

[205] W.W.S.I. Rodrigo, E. Ortiz-Riaño, C. Pythoud, S. Kunz, J.C. de la Torre, L. Martínez-Sobrido, Arenavirus nucleoproteins prevent activation of nuclear factor kappa B., J. Virol. 86 (2012) 8185–97. doi:10.1128/JVI.07240-11.

[206] S. Urata, J. Yasuda, Molecular Mechanism of Arenavirus Assembly and Budding, Viruses. 4 (2012) 2049–2079. doi:10.3390/v4102049.

[207] O. Takeuchi, S. Akira, MDA5/RIG-I and virus recognition., Curr. Opin. Immunol. 20 (2008) 17–22. doi:10.1016/j.coi.2008.01.002.

[208] P. Luthra, D. Sun, R.H. Silverman, B. He, Activation of IFN-β expression by a viral mRNA through RNase L and MDA5, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 2118–23. doi:10.1073/pnas.1012409108.

[209] M. Yoneyama, K. Onomoto, T. Fujita, Cytoplasmic recognition of RNA., Adv. Drug Deliv. Rev. 60 (2008) 841–6. doi:10.1016/j.addr.2007.12.001.

[210] E. Meylan, J. Tschopp, M. Karin, Intracellular pattern recognition receptors in the host response., Nature. 442 (2006) 39–44. doi:10.1038/nature04946.

[211] K.M. Hastie, S. Bale, C.R. Kimberlin, E.O. Saphire, Hiding the evidence: two strategies for innate immune evasion by hemorrhagic fever viruses., Curr. Opin. Virol. 2 (2012) 151–6. doi:10.1016/j.coviro.2012.01.003.

[212] C.A. Biron, G.C. Sen, Innate Responses to Viral Infections, in: D.M. Knipe, P.M. Howley (Eds.), Fields Virol., 5th ed., Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2007: pp. 250–278.

[213] C. Wilkins, M. Gale, Recognition of viruses by cytoplasmic sensors., Curr. Opin. Immunol. 22 (2010) 41–7. doi:10.1016/j.coi.2009.12.003.

[214] T. Kawai, S. Akira, Innate immune recognition of viral infection., Nat. Immunol. 7 (2006) 131–7. doi:10.1038/ni1303.

[215] S. Akira, S. Uematsu, O. Takeuchi, Pathogen recognition and innate immunity., Cell. 124 (2006) 783–801. doi:10.1016/j.cell.2006.02.015.

[216] A. Pichlmair, O. Schulz, C.P. Tan, T.I. Näslund, P. Liljeström, F. Weber, et al., RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates., Science. 314 (2006) 997–1001. doi:10.1126/science.1132998.

[217] A. Pichlmair, O. Schulz, C.-P. Tan, J. Rehwinkel, H. Kato, O. Takeuchi, et al., Activation of MDA5 requires higher-order RNA structures generated during virus infection., J. Virol. 83 (2009) 10761–9. doi:10.1128/JVI.00770-09.

[218] H. Häcker, M. Karin, Regulation and function of IKK and IKK-related kinases., Sci. STKE. 2006 (2006) re13. doi:10.1126/stke.3572006re13.

[219] X.-L. Li, H.J. Ezelle, T.Y. Hsi, B.A. Hassel, A central role for RNA in the induction and biological activities of type 1 interferons, Wiley Interdiscip. Rev. RNA. 2 (2011) 58–78. doi:10.1002/wrna.32.

[220] J.-B. Marq, D. Kolakofsky, D. Garcin, Unpaired 5' ppp-nucleotides, as found in arenavirus double-stranded RNA panhandles, are not recognized by RIG-I., J. Biol. Chem. 285 (2010) 18208–18216.

[221] SIB Swiss Institute of Bioinformatics, Viral Zone : Cap Snatching, (n.d.). http://viralzone.expasy.org/all_by_protein/839.html (accessed October 1, 2013).

[222] S. Suetsugu, K. Toyooka, Y. Senju, Subcellular membrane curvature mediated by the BAR domain superfamily proteins., Semin. Cell Dev. Biol. 21 (2010) 340–9. doi:10.1016/j.semcdb.2009.12.002.

VI. ABREVIATIONS

ADP :Adénosine Di-Phosphate

ARN:Acide Ribonucléique

ARNm : ARN messager

Arp2/3 complexe :Actine Related Protein

ATP :Adénosine Tri-Phosphate

BAR protéine ou domaine : acronyme des protéines Bin/Amphiphysin/Rvsp

ENA (Protéine) :Enable homologue protéine

FAB : Domaine de liaison à l'actine F

GAB : Domaine de liaison à l'actine G (autre nom du domaine WH2)

IMD : IRSp53/MIM Domaine homologue

IRSp53 :Insulin receptor substrate protein of 53 kDa

MIM : Missing in Metastasis

N / NP : Nucléoprotéine

NSP/nsp : non-structural protein

SAM : S-Adenosyl-Methionine

VASP :vasodilator-stimulated phosphoprotéine

VCML : Virus de la Chorioméningite Lymphocytaire

VFVR : Virus de la Fièvre de la Vallée du Rift

VSV : *Virus de la stomatite vésiculaire*

WASP :Wiskott-Aldrich Syndrome Protéine

Domaine WH2 : Domaine Wiskott-Aldrich Homologue 2

VII. ANNEXES

A. Protéines d'assemblage et de régulation du cytosquelette

The EMBO Journal (2007) 26, 4597–4606 | © 2007 European Molecular Biology Organization | All Rights Reserved 0261-4189/07
www.embojournal.org

THE
EMBO
JOURNAL

Structural basis for the recruitment of profilin–actin complexes during filament elongation by Ena/VASP

François Ferron, Grzegorz Rebowski, Sung Haeng Lee and Roberto Dominguez*

Department of Physiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Cells sustain high rates of actin filament elongation by maintaining a large pool of actin monomers above the critical concentration for polymerization. Profilin–actin complexes constitute the largest fraction of polymerization-competent actin monomers. Filament elongation factors such as Ena/VASP and formin catalyze the transition of profilin–actin from the cellular pool onto the barbed end of growing filaments. The molecular bases of this process are poorly understood. Here we present structural and energetic evidence for two consecutive steps of the elongation mechanism: the recruitment of profilin–actin by the last poly-Pro segment of vasodilator-stimulated phosphoprotein (VASP) and the binding of profilin–actin simultaneously to this poly-Pro and to the G-actin-binding (GAB) domain of VASP. The actin monomer bound at the GAB domain is proposed to be in position to join the barbed end of the growing filament concurrently with the release of profilin.

The EMBO Journal (2007) **26,** 4597–4606. doi:10.1038/sj.emboj.7601874; Published online 4 October 2007
Subject Categories: cell & tissue architecture; structural biology
Keywords: actin; cytoskeleton dynamics; filament elongation; profilin

Introduction

The enabled/vasodilator-stimulated phosphoprotein (Ena/VASP) family is implicated in diverse cellular processes involving dynamic actin assembly, such as fibroblast migration, axon guidance and the movement of the bacterial pathogen *Listeria monocytogenes* (Krause *et al*, 2003). Ena/VASP proteins share a tripartite domain organization (Figure 1), consisting of N- and C-terminal Ena/VASP homology 1 and 2 (EVH1 and EVH2) domains and a central Pro-rich region. The EVH1 domain targets Ena/VASP proteins to focal adhesions, filopodia and lamellipodia by binding to target proteins containing the consensus sequence 'FPPPP', such as vinculin (Brindle *et al*, 1996; Reinhard *et al*, 1996), lamellipodin (Krause *et al*, 2004), zyxin (Reinhard *et al*, 1995b), migfilin (Zhang *et al*, 2006) and palladin (Boukhelifa *et al*, 2004). The Pro-rich region of Ena/VASP binds profilin and the SH3 and WW domains of various signaling and scaffolding proteins, including Abl, Src and IRSp53 (Gertler *et al*, 1995; Ahern-Djamali *et al*, 1999; Krugmann *et al*, 2001). The EVH2 domain comprises globular and filamentous actin-binding sites (GAB and FAB), as well as a C-terminal coiled-coil (CC) region that mediates VASP tetramerization (Bachmann *et al*, 1999; Walders-Harbeck *et al*, 2002).

Ena/VASP functions primarily as an actin filament elongation factor, whereas it is a relatively weak nucleator (Bear *et al*, 2002; Barzik *et al*, 2005). Thus, depletion of Ena/VASP produces shorter and more densely branched filament networks, whereas its overexpression causes the opposite effect (Skoble *et al*, 2001; Bear *et al*, 2002). Since profilin–actin constitutes the major pool of polymerization-competent actin in eukaryotic cells (Pollard and Borisy, 2003), filament elongation by Ena/VASP will depend upon its ability to 'process' profilin–actin complexes for their incorporation onto the barbed end of growing filaments. Consistent with this notion, the recruitment of profilin–actin by the central Pro-rich region of VASP enhances *Listeria* motility (Chakraborty *et al*, 1995; Kang *et al*, 1997; Geese *et al*, 2000, 2002; Auerbuch *et al*, 2003; Grenklo *et al*, 2003). Furthermore, a covalently cross-linked profilin–actin complex, which cannot function as a source for filament elongation, markedly reduces *Listeria* motility, and this effect depends on the ability of profilin to interact with the poly-Pro region of Ena/VASP (Grenklo *et al*, 2003). Profilin also appears to enhance actin polymerization by increasing the anti-capping activity of VASP (Barzik *et al*, 2005). The role of profilin–actin in filament elongation is not unique to Ena/VASP. Most formins also recruit profilin–actin via their Pro-rich FH1 domain, which increases the rate of filament elongation (Kovar *et al*, 2006). Despite the importance of profilin–actin in filament elongation, the molecular bases for this function are poorly understood. Here, we investigate the energetics and structural bases for the interactions of profilin–actin with the last poly-Pro and the poly-Pro-GAB regions of human VASP. Based on the results, we propose a model for the recruitment and processing of profilin–actin complexes during filament elongation by Ena/VASP.

Results

The Pro-rich region of Ena/VASP consists of three distinct poly-Pro sites

To better understand the modular organization of VASP, and in particular that of its Pro-rich region, we analyzed the distribution of clusters of hydrophobic amino acids using the program HCA (Callebaut *et al*, 1997) (Figure 1).

*Corresponding author. Department of Physiology, University of Pennsylvania School of Medicine, A507 Richards Building, 3700 Hamilton Walk, Philadelphia, PA 19104-6058, USA.
Tel.: +1 215 573 4559; Fax: +1 215 573 5851;
E-mail: droberto@mail.med.upenn.edu

Received: 18 July 2007; accepted: 10 September 2007; published online: 4 October 2007

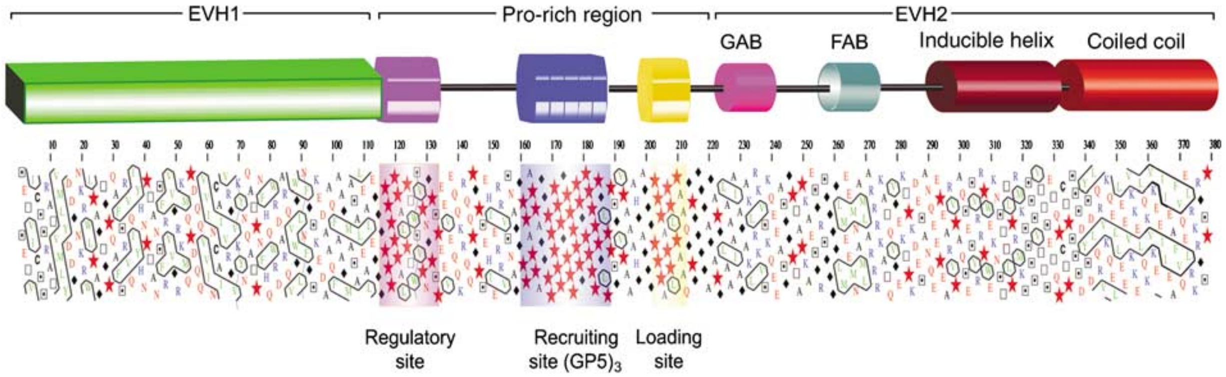


Figure 1 Modular organization of VASP. The diagram shown on the upper part of the figure represents the modular organization of the human VASP sequence (UniProt P50552) based on the distribution of clusters of hydrophobic amino acids determined with the program HCA (Callebaut *et al*, 1997). Symbols are as follows: Pro, ★; Gly, ◆; Thr, □ and Ser, ▣. Amino acids are colored according to their chemical characteristics: green, hydrophobic; red, negatively charged; blue, positively charged; black, Ala and Cys. The EVH1 and CC domains, which are stably folded and whose structures have been determined (Prehoda *et al*, 1999; Kuhnel *et al*, 2004), are characterized by a higher density of clusters of hydrophobic amino acids (contoured black). In contrast, the extended Pro-rich region (amino acids 116–213) lacks hydrophobic clusters, and is predicted to be mostly unfolded. Within this region, we identify three distinct groups of Pro residues, which are present in all members of the Ena/VASP family: regulatory site (116–135), recruiting site (160–194) and loading site (201–211). Finally, the region preceding the CC domain has the typical HCA pattern of an inducible α-helix.

Predictably, the EVH1 and CC regions, whose crystal structures have been determined (Prehoda *et al*, 1999; Kuhnel *et al*, 2004), are characterized by a higher density of clusters of hydrophobic amino acids. Two small clusters, displaying a characteristic helical pattern (Callebaut *et al*, 1997), correspond to the GAB and FAB domains. The extended Pro-rich region located between the EVH1 and GAB domains lacks hydrophobic clusters and is therefore predicted to be mostly unfolded. Although this region is more divergent, we identify three distinct groups of proline residues that are present in all members of the Ena/VASP family (Figure 1). The first group, amino acids 116–135 (human VASP sequence numbers), consists of a mixture of proline and hydrophobic amino acids, a signature feature of SH3- and WW-binding sequences. Thus, programs such as iSPOT (Brannetti and Helmer-Citterich, 2003) and SCANSITE (Obenauer *et al*, 2003) predict that this poly-Pro site binds the SH3 domains of a number of signaling proteins, including Abl, Nck, Crk and cortactin. The interactions of Ena/VASP with some of these proteins have been demonstrated (Gertler *et al*, 1995; Sparks *et al*, 1996; Coppolino *et al*, 2001), although not always mapped specifically to this site. Because this site does not correspond to a canonical profilin-binding sequence (Perelroizen *et al*, 1994; Petrella *et al*, 1996; Kang *et al*, 1997), and since it is most likely involved in regulation, we identify it as the 'regulatory' poly-Pro site. The second group of proline residues (amino acids 160–194) comprises three repeats of the specific profilin-binding sequence GPPPPP (or GP5), and has therefore the potential to bind multiple profilin–actin complexes (Mahoney *et al*, 1999). For this reason, we identify this site as the 'recruiting' poly-Pro site. Most of the studies on the profilin–poly-Pro interaction have focused on this GP5 region or plain poly-Pro sequences (Perelroizen *et al*, 1994; Reinhard *et al*, 1995a; Petrella *et al*, 1996; Kang *et al*, 1997; Lambrechts *et al*, 1997; Jonckheere *et al*, 1999). The third group of proline residues is located between the recruiting poly-Pro site and the GAB domain, and is flanked on both sides by short flexible linkers rich in Gly residues. Because of its location, immediately N-terminal to the GAB domain, and its apparent role in delivering profilin–actin from the poly-Pro region to the GAB domain (see below), we identify this site as the 'loading' poly-Pro site. The loading site is highly conserved among all members of the Ena/VASP family, and always presents a hydrophobic amino acid at the penultimate position, most commonly Leu (Supplementary Figure 1). The loading poly-Pro site and the GAB domain are the central focus of this study.

Binding of profilin and profilin–actin to the loading poly-Pro site of VASP

As mentioned above, various studies have investigated the binding of profilin to plain poly-Pro sequences and GP5 sequences of the kind found in the recruiting site of VASP (Perelroizen *et al*, 1994; Reinhard *et al*, 1995a; Petrella *et al*, 1996; Kang *et al*, 1997; Lambrechts *et al*, 1997; Jonckheere *et al*, 1999). In particular, profilin has been reported to bind the GP5 sequence of VASP with relatively low affinity (K_D of 84–250 μM) (Petrella *et al*, 1996; Kang *et al*, 1997). However, neither the binding of profilin to the highly conserved loading poly-Pro site nor the binding of profilin–actin to any poly-Pro sequence has yet been investigated. Here, we used two different methods, isothermal titration calorimetry (ITC) and intrinsic tryptophan fluorescence, to measure the binding affinities of both profilin and profilin–actin for a 16-aa synthetic peptide (^{198}GAGGGPPPAPPLPAAQ^{213}) containing the loading poly-Pro site of human VASP (Figure 2; Supplementary Figure 2). Both methods consistently showed that the binding of profilin–actin for this peptide in high ionic strength buffer is 6- to 11-fold higher affinity than that of profilin alone (K_D of 7.5–8 μM for profilin–actin vs 50–84 μM for profilin alone). Although not specifically addressed here, similar results were obtained for a GP5 triplet peptide, resulting in a K_D of 23 μM for profilin–actin vs 230 μM for profilin alone (data not shown). We also found that the affinity of profilin for actin in G-buffer increases twofold when profilin is bound to the loading poly-Pro peptide of VASP (K_D of 0.8 vs 1.6 μM) (Supplementary Figure 3, and

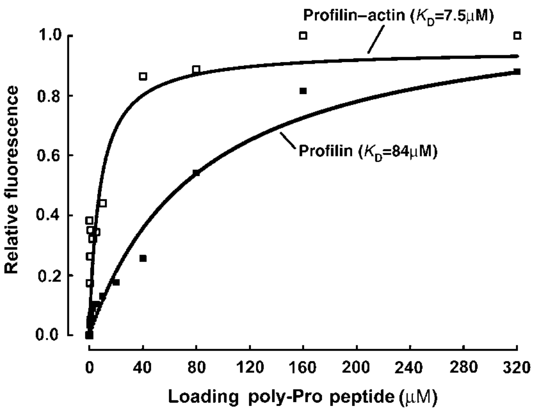


Figure 2 Binding of profilin and profilin–actin to the loading poly-Pro site of VASP. Binding of the VASP peptide (^{198}GAGGGPPP APPLPAAQ^{213}) produces a significant change in the intrinsic tryptophan fluorescence of profilin alone (■) and profilin–actin (□). In this experiment, the concentration of profilin and profilin–actin was 5 μM and the concentration of the VASP peptide varied for 0–320 μM (see Materials and methods). Each data point corresponds to the average of five independent measurements. Least-square fitting of the data, using a single-site binding model, resulted in dissociation constant (K_D) estimates of 84 and 7.5 μM for profilin and profilin–actin, respectively.

the legend). Together, these results suggest that poly-Pro sequences bind preferentially profilin–actin than profilin alone. This finding may have important functional implications, since profilin–actin constitutes the major pool of polymerization-competent actin in cells, and the affinities determined here for poly-Pro sequences fall within the intracellular range of concentrations of profilin–actin (5–40 μM) (Condeelis, 1993; Stossel, 1993; Kang *et al*, 1999; Dickinson and Purich, 2002; Dickinson *et al*, 2002; Pollard and Borisy, 2003).

Structure of profilin–actin bound to the loading poly-Pro site of human VASP

Crystal structures of profilin bound to decameric and pentadecameric L-Pro peptides have been determined (Mahoney *et al*, 1997, 1999). These structures revealed that profilin, like SH3 domains (Feng *et al*, 1994), binds plain poly-Pro sequences in two distinct backbone orientations. It was proposed at the time that non-Pro residues in profilin ligands may dictate the actual polarity and register of binding (Mahoney *et al*, 1999). These two factors, polarity and register, become particularly important if there is, as we speculate, a mechanism whereby profilin–actin complexes are transferred directly from the last poly-Pro site to the GAB domain of Ena/VASP during filament elongation. To test this hypothesis, we determined the structure of the ternary complex of profilin–actin bound to the 16-aa peptide corresponding to the loading poly-Pro site of human VASP. In the structure, determined to 1.8-Å resolution (Table I), actin and the VASP peptide bind on opposite sides of the profilin molecule (Figure 3A; Supplementary Figure 4A). The conformation of skeletal α-actin is very similar to that of β-actin in the structure of its complex with profilin determined previously to 2.55-Å resolution (Schutt *et al*, 1993). In particular, in both structures the nucleotide cleft is in the closed conformation, different from the open cleft conformation suggested by another study (Chik *et al*, 1996). Compared to other actin structures, the two profilin–actin complexes also display a slightly closed target-binding cleft, consisting of the cleft between actin subdomains 1 and 3 (Dominguez, 2004). This effect, which appears to be directly linked to the binding of profilin, may explain the increased affinity of the GAB domain for profilin–actin compared with actin alone (Chereau and Dominguez, 2006).

Table I Crystallographic data and refinement statistics

	Profilin–actin–poly-Pro	Profilin–actin–poly-Pro-GAB
Diffraction statistics		
Space group	P 2_1 2_1 2_1	C 2
Cell parameters		
a, *b*, *c* (Å)	37.47, 75.46, 180.74	119.09, 56.59, 75.23
α, β, γ (deg)	90.0, 90.0, 90.0	90.0, 104.75, 90.0
Resolution		
Total (Å)	47.09–1.8	35.05–1.5
Outer shell (Å)	1.86–1.8	1.55–1.5
Completeness (%)	98.1 (97.7)	99.1 (98.2)
Redundancy	7.0 (5.9)	8.3 (7.5)
Unique reflections	47 041 (4579)	76 002 (7370)
R-merge[a] (%)	7.6 (34.5)	7.1 (29.4)
Average *I*/σ	28.1 (5.7)	32.44 (7.8)
Refinement statistics		
R-factor[b] (%)	15.8	15.6
R-free[c] (%)	20.1	19.0
R.m.s. deviations		
Bond length (Å)	0.013	0.015
Bond angles (deg)	1.402	1.701
Average *B*-factor		
All atoms (Å^2)	14.92	10.84
Actin/profilin/VASP (Å^2)	13.18/14.26/28.82	9.13/7.17/12.34
Solvent atoms (Å^2)	23.29	22.5
PDB code	2PAV	2PBD

GAB, G-actin-binding domain; VASP, vasodilator-stimulated phosphoprotein.
Values in parentheses correspond to highest-resolution shell.
[a] *R*-merge $=\sum(I-\langle I\rangle)/\sum I$; *I* and $\langle I\rangle$ are the intensity and the mean value of all the measurements for an individual reflection.
[b] *R*-factor $=\sum|Fo-Fc|/\sum|Fo|$; *F*o and *F*c are the observed and calculated structure factors.
[c] *R*-free, *R*-factor calculated for a randomly selected subset of the reflections (5%) that were omitted during the refinement.

One noticeable difference between the two profilin–actin structures occurs in the DNase I-binding loop, which is disordered in the ternary complex studied here, but was visualized in the original structure of profilin–β-actin (Schutt *et al*, 1993). Also, the C-terminal Phe 375 of actin is rotated around the main chain backbone ∼180° between the two structures. This amino acid is at the center of the profilin–actin interface, an area that is strictly conserved between α- and β-actin, and which is very well defined in the electron density map (Supplementary Figure 4B). The conformation of profilin is also very similar in the two structures, except for the binding interface of the VASP peptide (Figure 3C), where the side chains of residues His 133, Ser 137 and Tyr 139 adopt different rotamer orientations. The N-terminus of profilin also forms part of the binding interface and moves in the current structure toward the VASP peptide, making contacts with it.

The first three amino acids of the loading poly-Pro peptide are disordered in the structure, whereas the remaining 13 (Gly 201–Gln 213) are all very well defined in the electron

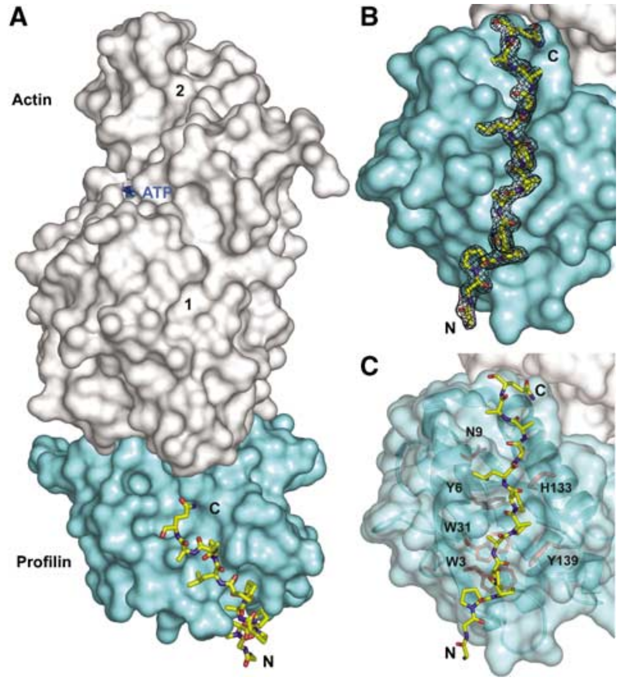


Figure 3 Crystal structure at 1.8-Å resolution of the ternary complex of profilin–actin with the loading poly-Pro site of human VASP. (**A**) General view of the structure (actin, gray; profilin, cyan; VASP peptide, all-atom representation). The view is rotated ~90° relative to the 'classical' orientation of actin shown in Supplementary Figure 3A. (**B**) Electron density map around the VASP peptide, contoured at 1.0σ. (**C**) Illustration of the main interactions of the VASP peptide with profilin amino acids (colored red, under a transparent surface representation of profilin).

density map (Figure 3B). The orientation of the VASP peptide is both unambiguous and unique, with the C-terminal end of the peptide pointing toward actin (Figure 3A). The interactions with profilin are limited to the eight amino acids between Pro 203 and Pro 210, and include stacking and hydrogen bonding interactions with profilin residues Trp 3, Asn 9, Trp 31, His 133 and Tyr 139 (Figure 3C). Profilin Tyr 6 also plays a critical role in the interaction, making both a hydrogen-bonding contact with a main-chain atom and a hydrophobic stacking interaction with Leu 209 of the peptide.

The conformation of the VASP peptide is significantly different from the two backbone conformations of the decameric L-Pro peptide, but somewhat similar to that of the pentadecameric L-Pro peptide studied previously (Mahoney *et al*, 1997; Mahoney *et al*, 1999) (Supplementary Figure 5). However, important differences occur, which correlate with substitutions of some of the Pro residues in the all-Pro pentadecameric peptide by non-Pro residues in VASP. Thus, although the N-terminal portions of the two peptides superimpose well, the chains diverge slightly after Ala 206, and this effect becomes more pronounced after Leu 209 of the VASP peptide. In about 50% of Ena/VASP sequences, Ala 206 is replaced by Pro (Supplementary Figure 1) and this substitution is unlikely to have any major effect in the general conformation of the complex. However, Leu 209 is strictly conserved and is also found in WASP (Supplementary Figure 1). The role of this Leu residue appears to be twofold; it determines the orientation of the C-terminal portion of the loading poly-Pro region (toward actin), and also determines the register and polarity of the interaction. As we demonstrate next, these factors are in turn critical in defining the position of the GAB domain C-terminal to the loading poly-Pro site so that it can bind to actin.

Structure of profilin–actin bound to the loading poly-Pro-GAB region of VASP

We had previously shown that profilin–actin binds the GAB domain of VASP with higher affinity than actin alone, leading to the proposal that profilin–actin complexes might transition from the loading poly-Pro site to the GAB domain during filament elongation (Chereau and Dominguez, 2006). We had also proposed that the GAB domain is related to WH2 and would therefore be expected to bind in the cleft between actin subdomains 1 and 3 (Chereau *et al*, 2005). However, an important question remained; the structures of WH2–actin indicated that the binding sites for profilin and WH2 partially overlap and a conformational change would be necessary if the two were to bind actin simultaneously (Chereau *et al*, 2005). To test this hypothesis we determined the 1.5-Å resolution crystal structure of the ternary complex of profilin—actin, with a 43-aa peptide comprising the loading poly-Pro and GAB domains of human VASP (residues Gly 202–Ser 244) (Figure 4A**)**. The length of the VASP peptide was defined on the basis of the structures of the loading poly-Pro region described above (Figure 3), and that of the WH2 of WASP bound to actin (Chereau *et al*, 2005), ensuring that none of the amino acids involved in interactions with profilin and actin were excluded.

This ternary complex crystallized in a different space group than that of the loading poly-Pro site (Table I), yet the profilin–actin portions of the two structures are very similar. Differences in three loops of actin (amino acids 199–203, 241–247 and 322–329) can all be attributed to different crystal packing contacts. The electron density map is generally very well defined, and includes two portions of the VASP peptide, amino acids Gly 202–Ala 212 (loading poly-Pro site) and Ala 222–Ser 238 (GAB domain) (Figures 4B and C). The conformation and interactions of the loading poly-Pro site are very similar to those described above (Figures 3C and 4B). The only noticeable difference is that in this structure the loop Tyr 24–Pro 28 of profilin moves toward the peptide and makes one additional hydrogen-bonding contact with it.

As anticipated, the GAB domain binds roughly at the same site as the WH2 of WASP (Figure 4D). Both domains consist of an N-terminal helix that binds in the cleft between actin subdomains 1 and 3, and a C-terminal extended region that binds along the actin surface, climbing toward the pointed end of the actin monomer. Based on this observation, the GAB domain of Ena/VASP can be unambiguously classified as a WH2 domain. However, comparison of the structure with those of WH2 and Tβ domains determined in the absence of profilin (Hertzog *et al*, 2004; Chereau *et al*, 2005; Lee *et al*, 2007) also reveals important differences. Thus, the helix of the GAB domain is rotated ~45° and shifted forward half a helical turn (Figure 4D). These changes add to a displacement of the GAB domain away from profilin, whereas profilin is unaffected. It is unclear if these changes can all be attributed to the presence of profilin, since the sequence of the GAB domain contains some important differences compared to classical WH2 domains. In particular, the helix of the GAB domain is only two turns long, interrupted at the N-terminus by a combination of Gly and Pro residues, whereas in most

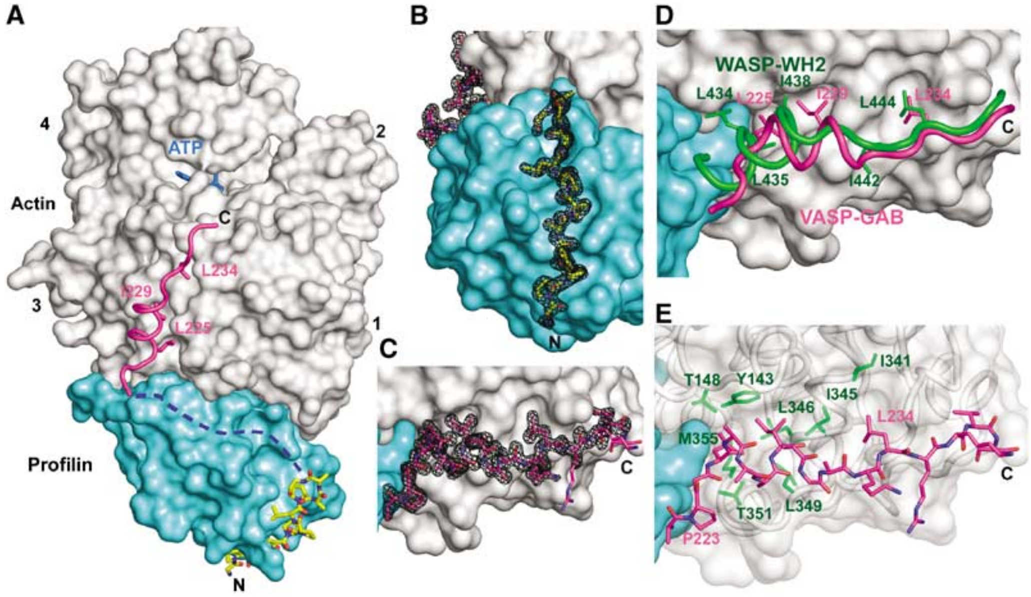


Figure 4 Crystal structure at 1.5-Å resolution of the ternary complex of profilin–actin with the loading poly-Pro-GAB region of human VASP. (**A**) General view of the structure (actin, gray; profilin, cyan). The two portions of the peptide visualized in the electron density map are colored differently (corresponding to the colors used in Figures 1 and 3): poly-Pro, all-atom representation with carbon atoms in yellow and GAB in magenta. The linker between these two sites is missing in the structure (represented by a discontinuous blue line), but can be modeled along the proposed path (see also Supplementary Movie 6). (**B**, **C**) Electron density map contoured at 1.0σ around the poly-Pro and GAB, respectively. (**D**) Superimpositions of the structures of the GAB of VASP (magenta) and the WH2 of WASP (green) (Chereau *et al*, 2005). Only the side chains of hydrophobic amino acids that interact with actin are shown. Note that the helix of the GAB is both shifted forward and rotated ∼45° relative to that of WH2, which is at least in part due to the presence of profilin in the current structure. The LKKT portions of the two structures, however, superimpose well. (**E**) Illustration of the main interactions of the GAB (magenta) with hydrophobic amino acids (green) of the cleft between actin subdomains 1 and 3.

WH2s the helix is at least one turn longer and would be expected to conflict with profilin (Figure 4D). Despite these differences the interaction of the helix of the GAB domain presents the same hydrophobic character as the helices of other actin-binding proteins that commonly bind in this cleft (Dominguez, 2004). Thus, VASP residues Leu 225 and Ile 229, on the hydrophobic side of this helix, face the hydrophobic cleft in actin (Figure 4E).

C-terminal to the helix of the GAB domain is the sequence ^{234}LRKV^{237}. This sequence, conserved in all members of the Ena/VASP family (Supplementary Figure 1), corresponds to the so-called ‘LKKT’ motif found in a number of actin-binding proteins, including thymosin β (Hertzog *et al*, 2004), gelsolin (Irobi *et al*, 2003) and WH2 (Paunola *et al*, 2002; Chereau *et al*, 2005). The interactions of the LKKT motif with actin are conserved in these proteins as well as in the GAB domain of VASP. The most important element of this interaction is the binding of VASP residue Leu 234 in a hydrophobic pocket on the actin surface, surrounded by actin residues Ile 341 and Ile 345 (Figure 4E). In the current structure, as well as in the structure of the WH2 of WASP (Chereau *et al*, 2005), the amino acids C-terminal to the LKKT motif are disordered and do not appear to interact with actin.

Also disordered in the structure is the linker between the poly-Pro and GAB domains, not a surprising result given the high content of Gly residues in this linker (^{213}QGPGGGGAG^{221}). Since the VASP peptide is not observed as a continuous chain, the possibility exists that two different peptides are bound. However, this is unlikely since the complex was crystallized using an actin–profilin peptide stoichiometry of 1:1.1:1.1 (see Materials and methods), and no traces of density are observed in the 1.5-Å resolution map that would correspond to the significant unbound portions of two peptides. Another possibility is domain swapping of the VASP peptide between symmetry related molecules, although again there is no direct evidence of this in the electron density map. Nevertheless, none of these possibilities change the fact that profilin–actin binds the poly-Pro and GAB domains simultaneously in the structure. Furthermore, the missing amino acids of the linker can be easily modeled along a groove at the profilin–actin interface, so that they span the distance between the two sites (Supplementary Movie 6). Note also that GAB is followed by FAB, which tethers Ena/VASP to the elongating filament, thereby directing the incorporation of profilin–actin from the poly-Pro-GAB onto the barbed end of a specific filament (Figure 5).

Discussion

Two monomeric actin-binding proteins, profilin and thymosin-β4, contribute to maintaining a large fraction (∼50%) of the cellular actin in the unpolymerized pool at a concentration (∼20–100 μM) that is 200- to 1000-fold higher than the critical concentration for barbed-end polymerization (∼0.1 μM) (Pollard and Borisy, 2003). While free ATP–actin is present in low amounts (0.1–1 μM) and thymosin-β4–actin is polymerization incompetent, profilin–actin (present at 5–40 μM intracellular concentration) constitutes the main source of actin monomers for polymerization (Dickinson and Purich, 2002; Dickinson *et al*, 2002). Multiple properties allow profilin to play such a central role in filament assembly (dos Remedios *et al*, 2003; Witke, 2004; Yarmola and Bubb, 2006); (1) profilin catalyzes the exchange of ADP for ATP on actin, which replenishes the pool of ATP–actin monomers

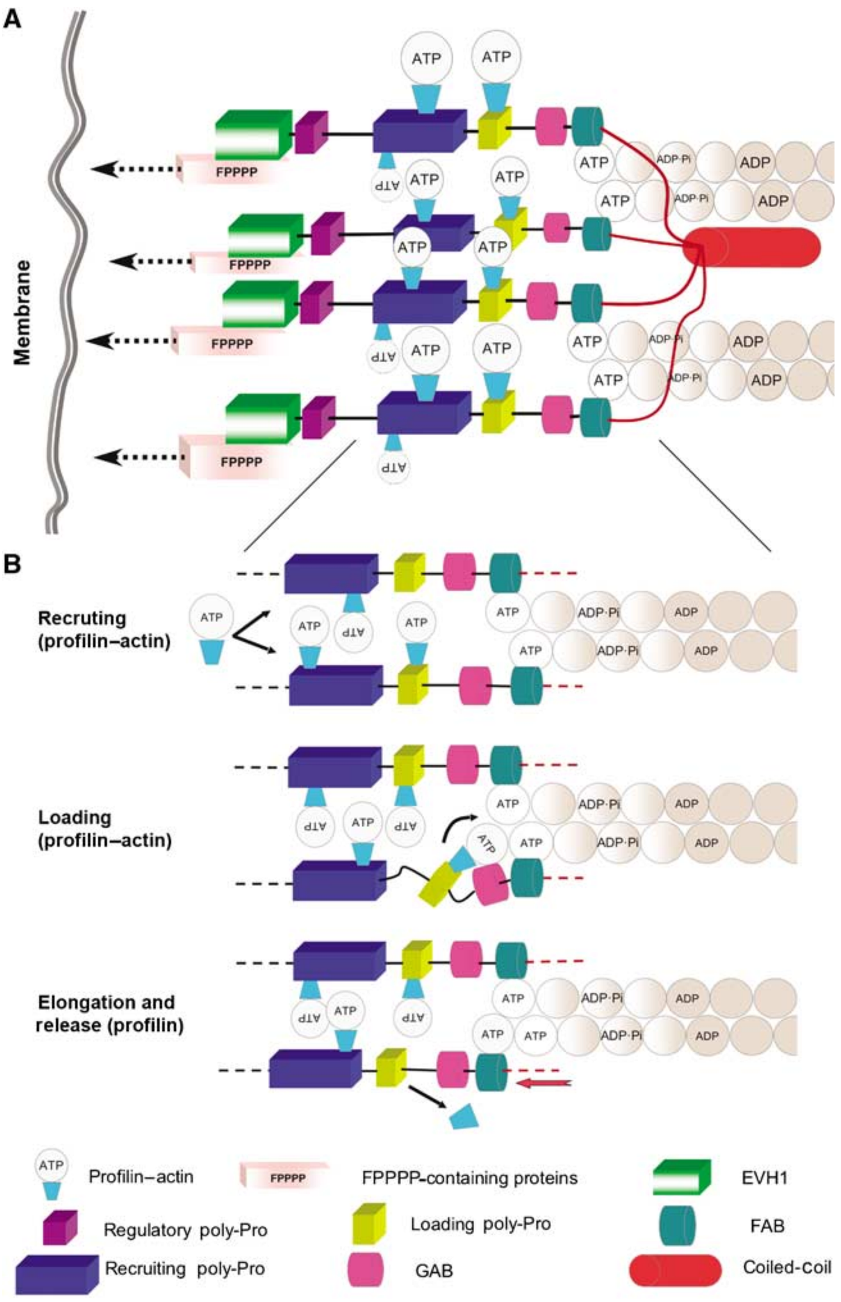


Figure 5 Model of filament elongation by Ena/VASP. (**A**) Ena/VASP tetramers are recruited to the sites of active filament assembly via interactions of the EVH1 domain with FPPPP-containing scaffolding proteins such as zyxin (Reinhard *et al*, 1995b), vinculin (Brindle *et al*, 1996; Reinhard *et al*, 1996), lamellipodin (Krause *et al*, 2004), migfilin (Zhang *et al*, 2006) and palladin (Boukhelifa *et al*, 2004). These proteins are connected to the plasma membrane directly or via interactions with membrane-binding scaffolding proteins (indicated by dashed arrows). VASP tetramerization and tethering via the EVH1 domain are likely important for processivity in cells (see Discussion). The recruiting poly-Pro site contains multiple profilin–actin binding modules, which may serve to increase the local concentration of actin monomers for assembly. It is unclear if each subunit of Ena/VASP supports the elongation of one filament or, as shown in the figure, two subunits per filament are required (one per each long-pitch helix). (**B**) Close view of the elongation machinery showing three steps of the elongation mechanism (indicated by black arrows): (1) recruitment of profilin–actin to the poly-Pro regions, (2) loading of profilin–actin from the last poly-Pro site onto the GAB domain and (3) filament elongation (red arrow) resulting from the addition of the actin subunit at the barbed end and the release of profilin.

ready for polymerization, (2) it inhibits filament nucleation, which explains in part the need for nucleation promoting factors, (3) profilin-bound actin monomers cannot add to pointed ends, but can elongate filament barbed ends at approximately the same rate as free actin monomers, (4) profilin competes effectively with thymosin-β4 for actin, which is due to its greater affinity for actin than thymosin-β4 and high dissociation constants of their respective complexes with actin, and (5) profilin binds poly-Pro sequences in various cytoskeletal proteins, including the Arp2/3 complex activator WASP and the elongation factors Ena/VASP and formin. Thus, recruitment of profilin–actin increases the elongation rates of formin (Kovar *et al*, 2006) and Ena/VASP-dependent *Listeria* motility (Chakraborty *et al*, 1995; Kang *et al*, 1997; Geese *et al*, 2000, 2002; Machner *et al*, 2001; Auerbuch *et al*, 2003; Grenklo *et al*, 2003).

How filament elongation factors process profilin–actin for its incorporation at the barbed end of the filament has remained a mystery. A view commonly held is that the role of the Pro-rich region of Ena/VASP (as well as formin) is to increase the local concentration of profilin–actin available for assembly (Kang *et al*, 1997; Geese *et al*, 2000; Machner *et al*, 2001; Dickinson and Purich, 2002; Grenklo *et al*, 2003; Kovar *et al*, 2006). This work provides evidence for a more active role of the Pro-rich region of Ena/VASP in elongation, by directing the transition of profilin–actin to the GAB domain, from where the actin monomer can join the barbed end of the growing filament (Figure 5). Thus, we have shown here that the Pro-rich region of Ena/VASP has a complex organization, consisting of three discrete modules: the regulatory, recruiting and loading poly-Pro sites (Figure 1). While the binding of profilin to poly-Pro sequences of the kind found in the recruiting site had been previously studied (Petrella *et al*, 1996; Kang *et al*, 1997), the loading site had not yet been investigated. This site is particularly important since it is highly conserved among all the members of the Ena/VASP family and WASP, and is connected to the GAB domain by a short Gly-rich linker (Supplementary Figure 1).

We found that the loading poly-Pro site binds profilin–actin with significantly higher affinity than profilin alone (Figure 2; Supplementary Figure 2). The implication of this finding is that Ena/VASP has the ability to discriminate between polymerization-competent profilin–actin complexes and profilin alone, which could jam the elongation machinery. We had previously reported that the GAB domain also has higher affinity for profilin–actin than actin alone (Chereau and Dominguez, 2006). Therefore, both the loading poly-Pro and GAB domains have a clear preference for profilin–actin. However, the two sites must cooperate, and not compete, for the transfer of profilin–actin from the cellular pool onto the barbed end of the elongating filament. To test this hypothesis, we determined the high-resolution structures of profilin–actin bound to the loading poly-Pro (Figure 3) and poly-Pro-GAB of human VASP (Figure 4). Among the main results of the structures are the following: (1) the conformation of profilin–α-actin in the two structures is very similar to that of profilin–β-actin determined previously (Schutt *et al*, 1993), (2) the presence of non-Pro amino acids in the loading poly-Pro site, and in particular Leu 209 that is conserved in Ena/VASP and WASP, define the register, polarity and orientation of binding, (3) these factors are in turn essential in determining the position of the GAB domain, so that it can access its binding site on actin within the profilin–actin complex, (4) the GAB domain is related to the WH2 of WASP and makes similar interactions with actin, (5) the GAB domain can bind actin concomitantly with profilin, which requires significant repositioning compared with WH2 domains crystallized in the absence of profilin, and most importantly (6) the loading poly-Pro and GAB domains of Ena/VASP do not compete for profilin–actin, which they can bind simultaneously (Figure 4).

The structures do not provide clear insights as to why the loading poly-Pro binds more tightly to profilin–actin (Figure 2; Supplementary Figure 2). However, compared with other actin structures, profilin produces a slight closure of the cleft between subdomains 1 and 3, known as the target-binding cleft (Dominguez, 2004), which may explain the increased affinity of profilin–actin for the GAB domain observed previously (Chereau and Dominguez, 2006). Another possibility is the mutual stabilization of profilin and actin within their complex, making the binding of both the GAB and poly-Pro domains more favorable. This latter effect appears to work both ways, since the affinity of actin for profilin also increases when profilin is bound to the loading poly-Pro peptide of VASP (Supplementary Figure 3).

What is the next step after profilin–actin has been recruited to the loading poly-Pro and GAB domains? We would like to suggest that the actin monomer bound to the GAB domain is already in contact with the barbed end of the FAB-tethered filament (Figure 5B). This idea is based strictly on structural constraints imposed by the short length of the linker between the GAB and FAB domains (Supplementary Figure 1), independently of where on F-actin the latter binds. However, as noted before (Chereau and Dominguez, 2006), the FAB domain is related to the C region of WASP and the two appear to be related to WH2. In other words, the GAB-FAB of Ena/VASP is related to the WH2-C of WASP, a relationship that extends at both ends to the loading poly-Pro and, to a lesser extent, to the acidic region (Supplementary Figure 1). WH2 domains often occur in tandem repeats (Paunola *et al*, 2002) and, if this proposal is correct, GAB-FAB and WH2-C constitute specialized forms of tandem WH2 domains, involved in filament elongation and Arp2/3 activation, respectively. FAB would then be expected to bind in the cleft between subdomains 1 and 3 of the last actin subunit at the barbed end of the elongating filament. The transition of an actin monomer from the GAB domain to the barbed end of the elongating filament would then trigger a series of events, including the release of profilin and the stepping of Ena/VASP (Figure 5B).

What is the driving force of these events? The mechanism proposed here is generally in agreement with the direct-transfer polymerization model of Dickinson and Purich (2002) (Dickinson *et al*, 2002). According to this model, ATP hydrolysis by actin is the driving force for the processive stepping of actin polymerization-based motors. However, the precise role of ATP hydrolysis in elongation remains controversial. While it has been suggested that ATP hydrolysis is required for profilin dissociation (Romero *et al*, 2007), other experiments show that filament elongation rates can exceed the ATP hydrolysis rate by at least 20-fold (Blanchoin and Pollard, 2002). Furthermore, phosphate dissociation after hydrolysis is slow (Carlier and Pantaloni, 1986), implying that a significant fraction of the filament consists of ADP-Pi-actin, which is considered to be structurally and functionally indistinguishable from ATP–actin (Pollard and Borisy, 2003). Therefore, based on the evidence to date, it is unclear whether ATP hydrolysis or a nucleotide-independent conformational change due to the G- to F-actin transition triggers profilin dissociation and Ena/VASP stepping.

How does this mechanism apply to all four subunits of the full-length Ena/VASP tetramer? Although the core of the elongation machinery consists of the poly-Pro-GAB-FAB region (Figure 5), it is very likely that all the domains of Ena/VASP are necessary for processive elongation in cells (the exception may be the motility of *Listeria*, which has evolved different recruitment mechanisms). Ena/VASP tetramers are recruited to sites of rapid assembly by interactions of the EVH1 domains with FPPPP-containing proteins such as zyxin (Reinhard *et al*, 1995b), vinculin (Brindle *et al*, 1996; Reinhard *et al*, 1996), lamellipodin (Krause *et al*, 2004), migfilin (Zhang *et al*, 2006) and palladin (Boukhelifa *et al*, 2004). Filaments associated with Ena/VASP tend to be unbranched, longer and form crosslinked bundles (Skoble *et al*, 2001; Bear *et al*, 2002; Svitkina *et al*, 2003). A tetramer may be more efficient at generating such actin bundles, in particular if the subunits of Ena/VASP work cooperatively, both within and between neighboring tetramers. But more importantly, tetramerization via the CC domain and tethering via the EVH1 domain may enable Ena/VASP to hold on to a particular filament during processive stepping, by sequentially allowing each subunit to release and advance while the others remain attached. Consistent with this idea, recent evidence shows that all the domains of Ena/VASP are required for proper localization and continued polymerization at filopodial tips (Applewhite *et al*, 2007). In particular, the GAB domain was found to play a key role in maintaining filopodial tip localization, which is consistent with recent evidence that the WH2 of N-WASP transiently tethers actin filament to the membrane (Co *et al*, 2007).

Many questions will require further investigation. For instance, it is unclear whether a single subunit of Ena/VASP can support the elongation of a whole filament, or whether two subunits per filament are required (one for each of the long-pitch helices of the filament). It is also possible that cycles of elongation by Ena/VASP alternate with short periods of depolymerization, or backward steps. Finally, we have stressed here the striking resemblance between the loading poly-Pro-GAB-FAB of VASP and the equivalent region in WASP (Supplementary Figure 1). Some of the results obtained here may therefore be applicable to WASP-Arp2/3-mediated branch nucleation. However, there also are important differences, notably that the helix of WH2 is longer than that of the GAB domain (Figure 4D), which may preclude the transfer of profilin–actin from the last poly-Pro of WASP to WH2.

Materials and methods

Preparation of proteins and peptides

The cDNA encoding for human profilin I was purchased from ATCC, amplified by PCR and inserted into vector pET29/T7 (Novagen). BL21(DE3) competent cells (Invitrogen) were transformed with this constructs and grown in LB medium at 37°C until the OD at 600 nm reached a value of 0.6. Expression was induced by the addition of 1 mM isopropyl-β-D-1-thiogalactopyranoside and carried out for 4 h at 37°C. Cells were harvested by centrifugation, resuspended in 10 mM Tris pH 7.5, 100 mM glycine, 100 mM NaCl, 1 mM DTT and lysed using a French press. The soluble lysate was purified on an affinity poly-L-proline sepharose column. Profilin was eluted from the column using 30% (v/v) DMSO, dialysed against 10 mM Tris pH 7.5, 50 mM NaCl and concentrated to ~14 mg/ml. Actin was prepared from rabbit skeletal muscle as described (Graceffa and Dominguez, 2003). Peptides corresponding to human VASP residues 198–213 (loading poly-Pro) and 202–244 (loading poly-Pro and GAB domain) were synthesized on an ABI431 peptide synthesizer and purified by HPLC (see Supplementary Figure 1 for sequence information). The concentrations of the peptides were determined by amino-acid analysis (Dana-Farber Cancer Institute, Boston, MA, USA).

Crystallization, data collection and structure determination

The complexes of profilin–actin with the two VASP peptides were prepared by mixing 66 μM actin in G-buffer (2 mM HEPES pH 7.4, 0.2 mM $CaCl_2$ and 0.2 mM ATP) with 73 μM profilin (in 10 mM Tris pH 7.5, 50 mM NaCl), followed by the addition of 73 μM VASP peptide (same buffer as profilin). The complexes were stored on ice and used at this concentration during crystallization trials at 4 and 20°C. The best crystals of profilin–actin with the loading poly-Pro peptide were obtained at 4°C in 4 μl sitting drops, consisting of a 1:1 (v/v) mixture of protein solution with 200 mM sodium formate and 20% (w/v) PEG 3350 well solution. The best crystals of profilin–actin with the poly-Pro-GAB peptide were obtained at 4°C in 3 μl sitting drops, consisting of a 1:2 (v/v) mixture of the protein solution with 150 mM DL-malic acid pH 7.0 and 18% (w/v) PEG 3350 well solution. The crystals were flash-frozen in liquid nitrogen, using 20% glycerol as cryoprotectant. X-ray data sets were collected at the BioCARS beamline 14-BM-C (Advance Photon Source, Argonne, IL, USA). The data sets were indexed and scaled with program HKL-2000 (HKL Research Inc.). The structures were determined by molecular replacement using the CCP4 program AMoRe and the structure of profilin–β-actin (Schutt *et al*, 1993) as a search model. Model building and refinement were performed with the program Coot (Emsley and Cowtan, 2004) and CCP4 program Refmac (Table I).

Binding of VASP peptide measured by tryptophan fluorescence

Binding of the loading poly-Pro peptide of human VASP to profilin and profilin–actin was measured by the change of intrinsic tryptophan fluorescence using a Cary Eclipse Fluorescence spectrophotometer (Varian). The excitation wavelength was set to 295 nm and the emission spectra were recorded from 300 to 400 nm. The experiments were performed at 20°C in 10 mM sodium phosphate pH 7.4 and 150 mM NaCl. The concentration of profilin and profilin–actin in the cell was 5 μM. The VASP peptide was added at varying concentrations (0, 0.15, 0.30, 0.6, 1.25, 2.5, 5, 10, 20, 40, 80, 160 and 320 μM). Each data point represents the average of five independent measurements (Figure 2). Least-square fitting of the data, using a single-site binding model, resulted in dissociation constant (K_D) estimates of 84 and 7.5 μM for profilin and profilin–actin, respectively.

Binding of VASP peptide measured by ITC

Binding of the VASP peptide to profilin and profilin–actin was also measured using ITC on a VP-ITC instrument (MicroCal, Northampton, MA, USA). To determine ΔH and K_a of association, the VASP peptide, at a concentration of 1 mM, was titrated in 10-μl injections, into 1.44 ml of 100 μM profilin (or 80 μM profilin–actin). The experiments were all performed at 25°C in 10 mM sodium phosphate pH 7.4 and 150 mM NaCl. The duration of each injection was 10 s, with an interval of 3 min between injections. The heat of binding was corrected for the small exothermic heat of injection, determined by injecting VASP peptides into buffer. Data were analyzed using the MicroCal's Origin program.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

This work was supported by NIH grant HL086655. We thank Paul Leavis and Elizabeth Gowell for the synthesis of VASP peptides. Use of the Advanced Photon Source was supported by DOE contract no. W-31-109-Eng-38. Use of the BioCARS facilities was supported by NIH grant RR07707.

References

Ahern-Djamali SM, Bachmann C, Hua P, Reddy SK, Kastenmeier AS, Walter U, Hoffmann FM (1999) Identification of profilin and src homology 3 domains as binding partners for *Drosophila* enabled. *Proc Natl Acad Sci USA* **96:** 4977–4982

Applewhite DA, Barzik M, Kojima SI, Svitkina TM, Gertler FB, Borisy GG (2007) Ena/VASP proteins have an anti-capping independent function in filopodia formation. *Mol Biol Cell* **18:** 2579–2591

Auerbuch V, Loureiro JJ, Gertler FB, Theriot JA, Portnoy DA (2003) Ena/VASP proteins contribute to *Listeria monocytogenes* pathogenesis by controlling temporal and spatial persistence of bacterial actin-based motility. *Mol Microbiol* **49:** 1361–1375

Bachmann C, Fischer L, Walter U, Reinhard M (1999) The EVH2 domain of the vasodilator-stimulated phosphoprotein mediates tetramerization, F-actin binding, and actin bundle formation. *J Biol Chem* **274:** 23549–23557

Barzik M, Kotova TI, Higgs HN, Hazelwood L, Hanein D, Gertler FB, Schafer DA (2005) Ena/VASP proteins enhance actin polymerization in the presence of barbed end capping proteins. *J Biol Chem* **280:** 28653–28662

Bear JE, Svitkina TM, Krause M, Schafer DA, Loureiro JJ, Strasser GA, Maly IV, Chaga OY, Cooper JA, Borisy GG, Gertler FB (2002) Antagonism between Ena/VASP proteins and actin filament capping regulates fibroblast motility. *Cell* **109:** 509–521

Blanchoin L, Pollard TD (2002) Hydrolysis of ATP by polymerized actin depends on the bound divalent cation but not profilin. *Biochemistry* **41:** 597–602

Boukhelifa M, Parast MM, Bear JE, Gertler FB, Otey CA (2004) Palladin is a novel binding partner for Ena/VASP family members. *Cell Motil Cytoskeleton* **58:** 17–29

Brannetti B, Helmer-Citterich M (2003) iSPOT: a web tool to infer the interaction specificity of families of protein modules. *Nucleic Acids Res* **31:** 3709–3711

Brindle NP, Holt MR, Davies JE, Price CJ, Critchley DR (1996) The focal-adhesion vasodilator-stimulated phosphoprotein (VASP) binds to the proline-rich domain in vinculin. *Biochem J* **318** (Part 3)**:** 753–757

Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon JP (1997) Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci* **53:** 621–645

Carlier MF, Pantaloni D (1986) Direct evidence for ADP-Pi-F-actin as the major intermediate in ATP-actin polymerization. Rate of dissociation of Pi from actin filaments. *Biochemistry* **25:** 7789–7792

Chakraborty T, Ebel F, Domann E, Niebuhr K, Gerstel B, Pistor S, Temm-Grove CJ, Jockusch BM, Reinhard M, Walter U, Wehland J (1995) A focal adhesion factor directly linking intracellularly motile *Listeria monocytogenes* and *Listeria ivanovii* to the actin-based cytoskeleton of mammalian cells. *EMBO J* **14:** 1314–1321

Chereau D, Dominguez R (2006) Understanding the role of the G-actin-binding domain of Ena/VASP in actin assembly. *J Struct Biol* **155:** 195–201

Chereau D, Kerff F, Graceffa P, Grabarek Z, Langsetmo K, Dominguez R (2005) Actin-bound structures of Wiskott–Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. *Proc Natl Acad Sci USA* **102:** 16644–16649

Chik JK, Lindberg U, Schutt CE (1996) The structure of an open state of beta-actin at 2.65 Å resolution. *J Mol Biol* **263:** 607–623

Co C, Wong DT, Gierke S, Chang V, Taunton J (2007) Mechanism of actin network attachment to moving membranes: barbed end capture by N-WASP WH2 domains. *Cell* **128:** 901–913

Condeelis J (1993) Life at the leading edge: the formation of cell protrusions. *Annu Rev Cell Biol* **9:** 411–444

Coppolino MG, Krause M, Hagendorff P, Monner DA, Trimble W, Grinstein S, Wehland J, Sechi AS (2001) Evidence for a molecular complex consisting of Fyb/SLAP, SLP-76, Nck, VASP and WASP that links the actin cytoskeleton to Fcgamma receptor signalling during phagocytosis. *J Cell Sci* **114:** 4307–4318

Dickinson RB, Purich DL (2002) Clamped-filament elongation model for actin-based motors. *Biophys J* **82:** 605–617

Dickinson RB, Southwick FS, Purich DL (2002) A direct-transfer polymerization model explains how the multiple profilin-binding sites in the actoclampin motor promote rapid actin-based motility. *Arch Biochem Biophys* **406:** 296–301

Dominguez R (2004) Actin-binding proteins—a unifying hypothesis. *Trends Biochem Sci* **29:** 572–578

dos Remedios CG, Chhabra D, Kekic M, Dedova IV, Tsubakihara M, Berry DA, Nosworthy NJ (2003) Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiol Rev* **83:** 433–473

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132

Feng S, Chen JK, Yu H, Simon JA, Schreiber SL (1994) Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3–ligand interactions. *Science* **266:** 1241–1247

Geese M, Loureiro JJ, Bear JE, Wehland J, Gertler FB, Sechi AS (2002) Contribution of Ena/VASP proteins to intracellular motility of *Listeria* requires phosphorylation and proline-rich core but not F-actin binding or multimerization. *Mol Biol Cell* **13:** 2383–2396

Geese M, Schluter K, Rothkegel M, Jockusch BM, Wehland J, Sechi AS (2000) Accumulation of profilin II at the surface of *Listeria* is concomitant with the onset of motility and correlates with bacterial speed. *J Cell Sci* **113** (Part 8)**:** 1415–1426

Gertler FB, Comer AR, Juang JL, Ahern SM, Clark MJ, Liebl EC, Hoffmann FM (1995) enabled, a dosage-sensitive suppressor of mutations in the *Drosophila* Abl tyrosine kinase, encodes an Abl substrate with SH3 domain-binding properties. *Genes Dev* **9:** 521–533

Graceffa P, Dominguez R (2003) Crystal structure of monomeric actin in the ATP state: structural basis of nucleotide-dependent actin dynamics. *J Biol Chem* **278:** 34172–34180

Grenklo S, Geese M, Lindberg U, Wehland J, Karlsson R, Sechi AS (2003) A crucial role for profilin–actin in the intracellular motility of *Listeria monocytogenes*. *EMBO Rep* **4:** 523–529

Hertzog M, van Heijenoort C, Didry D, Gaudier M, Coutant J, Gigant B, Didelot G, Preat T, Knossow M, Guittet E, Carlier MF (2004) The beta-thymosin/WH2 domain; structural basis for the switch from inhibition to promotion of actin assembly. *Cell* **117:** 611–623

Irobi E, Burtnick LD, Urosev D, Narayan K, Robinson RC (2003) From the first to the second domain of gelsolin: a common path on the surface of actin? *FEBS Lett* **552:** 86–90

Jonckheere V, Lambrechts A, Vandekerckhove J, Ampe C (1999) Dimerization of profilin II upon binding the (GP5)3 peptide from VASP overcomes the inhibition of actin nucleation by profilin II and thymosin beta4. *FEBS Lett* **447:** 257–263

Kang F, Laine RO, Bubb MR, Southwick FS, Purich DL (1997) Profilin interacts with the Gly–Pro–Pro–Pro–Pro–Pro sequences of vasodilator-stimulated phosphoprotein (VASP): implications for actin-based *Listeria* motility. *Biochemistry* **36:** 8384–8392

Kang F, Purich DL, Southwick FS (1999) Profilin promotes barbed-end actin filament assembly without lowering the critical concentration. *J Biol Chem* **274:** 36963–36972

Kovar DR, Harris ES, Mahaffy R, Higgs HN, Pollard TD (2006) Control of the assembly of ATP- and ADP–actin by formins and profilin. *Cell* **124:** 423–435

Krause M, Dent EW, Bear JE, Loureiro JJ, Gertler FB (2003) Ena/VASP proteins: regulators of the actin cytoskeleton and cell migration. *Annu Rev Cell Dev Biol* **19:** 541–564

Krause M, Leslie JD, Stewart M, Lafuente EM, Valderrama F, Jagannathan R, Strasser GA, Rubinson DA, Liu H, Way M, Yaffe MB, Boussiotis VA, Gertler FB (2004) Lamellipodin, an Ena/VASP ligand, is implicated in the regulation of lamellipodial dynamics. *Dev Cell* **7:** 571–583

Krugmann S, Jordens I, Gevaert K, Driessens M, Vandekerckhove J, Hall A (2001) Cdc42 induces filopodia by promoting the formation of an IRSp53:Mena complex. *Curr Biol* **11:** 1645–1655

Kuhnel K, Jarchau T, Wolf E, Schlichting I, Walter U, Wittinghofer A, Strelkov SV (2004) The VASP tetramerization domain is a right-handed coiled coil based on a 15-residue repeat. *Proc Natl Acad Sci USA* **101:** 17027–17032

Lambrechts A, Verschelde JL, Jonckheere V, Goethals M, Vandekerckhove J, Ampe C (1997) The mammalian profilin isoforms display complementary affinities for PIP2 and proline-rich sequences. *EMBO J* **16:** 484–494

Lee SH, Kerff F, Chereau D, Ferron F, Klug A, Dominguez R (2007) Structural basis for the actin-binding function of missing-in-metastasis. *Structure* **15:** 145–155

Machner MP, Urbanke C, Barzik M, Otten S, Sechi AS, Wehland J, Heinz DW (2001) ActA from *Listeria monocytogenes* can interact with up to four Ena/VASP homology 1 domains simultaneously. *J Biol Chem* **276:** 40096–40103

Mahoney NM, Janmey PA, Almo SC (1997) Structure of the profilin–poly-L-proline complex involved in morphogenesis and cytoskeletal regulation. *Nat Struct Biol* **4:** 953–960

Mahoney NM, Rozwarski DA, Fedorov E, Fedorov AA, Almo SC (1999) Profilin binds proline-rich ligands in two distinct amide backbone orientations. *Nat Struct Biol* **6:** 666–671

Obenauer JC, Cantley LC, Yaffe MB (2003) Scansite 2.0: proteome protwide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* **31:** 3635–3641

XI

Paunola E, Mattila PK, Lappalainen P (2002) WH2 domain: a small, versatile adapter for actin monomers. *FEBS Lett* **513:** 92–97

Perelroizen I, Marchand JB, Blanchoin L, Didry D, Carlier MF (1994) Interaction of profilin with G-actin and poly(L-proline). *Biochemistry* **33:** 8472–8478

Petrella EC, Machesky LM, Kaiser DA, Pollard TD (1996) Structural requirements and thermodynamics of the interaction of proline peptides with profilin. *Biochemistry* **35:** 16535–16543

Pollard TD, Borisy GG (2003) Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **112:** 453–465

Prehoda KE, Lee DJ, Lim WA (1999) Structure of the enabled/VASP homology 1 domain–peptide complex: a key component in the spatial control of actin assembly. *Cell* **97:** 471–480

Reinhard M, Giehl K, Abel K, Haffner C, Jarchau T, Hoppe V, Jockusch BM, Walter U (1995a) The proline-rich focal adhesion and microfilament protein VASP is a ligand for profilins. *EMBO J* **14:** 1583–1589

Reinhard M, Jouvenal K, Tripier D, Walter U (1995b) Identification, purification, and characterization of a zyxin-related protein that binds the focal adhesion and microfilament protein VASP (vasodilator-stimulated phosphoprotein). *Proc Natl Acad Sci USA* **92:** 7956–7960

Reinhard M, Rudiger M, Jockusch BM, Walter U (1996) VASP interaction with vinculin: a recurring theme of interactions with proline-rich motifs. *FEBS Lett* **399:** 103–107

Romero S, Didry D, Larquet E, Boisset N, Pantaloni D, Carlier MF (2007) How ATP hydrolysis controls filament assembly from profilin–actin: IMPLICATION FOR FORMIN PROCESSIVITY. *J Biol Chem* **282:** 8435–8445

Schutt CE, Myslik JC, Rozycki MD, Goonesekere NC, Lindberg U (1993) The structure of crystalline profilin–beta-actin. *Nature* **365:** 810–816

Skoble J, Auerbuch V, Goley ED, Welch MD, Portnoy DA (2001) Pivotal role of VASP in Arp2/3 complex-mediated actin nucleation, actin branch formation, and *Listeria monocytogenes* motility. *J Cell Biol* **155:** 89–100

Sparks AB, Rider JE, Hoffman NG, Fowlkes DM, Quillam LA, Kay BK (1996) Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2. *Proc Natl Acad Sci USA* **93:** 1540–1544

Stossel TP (1993) On the crawling of animal cells. *Science* **260:** 1086–1094

Svitkina TM, Bulanova EA, Chaga OY, Vignjevic DM, Kojima S, Vasiliev JM, Borisy GG (2003) Mechanism of filopodia initiation by reorganization of a dendritic network. *J Cell Biol* **160:** 409–421

Walders-Harbeck B, Khaitlina SY, Hinssen H, Jockusch BM, Illenberger S (2002) The vasodilator-stimulated phosphoprotein promotes actin polymerisation through direct binding to monomeric actin. *FEBS Lett* **529:** 275–280

Witke W (2004) The role of profilin complexes in cell motility and other cellular processes. *Trends Cell Biol* **14:** 461–469

Yarmola EG, Bubb MR (2006) Profilin: emerging concepts and lingering misconceptions. *Trends Biochem Sci* **31:** 197–205

Zhang Y, Tu Y, Gkretsi V, Wu C (2006) Migfilin interacts with vasodilator-stimulated phosphoprotein (VASP) and regulates VASP localization to cell–matrix adhesions and migration. *J Biol Chem* **281:** 12397–12407

Structural Basis for the Actin-Binding Function of Missing-in-Metastasis

Sung Haeng Lee,[1] Frederic Kerff,[2,3] David Chereau,[2,4] François Ferron,[1] Alexandra Klug,[2,5] and Roberto Dominguez[1,*]
[1] Department of Physiology, University of Pennsylvania School of Medicine, 3700 Hamilton Walk, Philadelphia, PA 19104, USA
[2] Boston Biomedical Research Institute, 64 Grove Street, Watertown, MA 02472, USA
[3] Present address: Centre d'Ingenierie des Proteines, Universite de Liege, Institut de Physique B5, B-4000 Liege, Belgium.
[4] Present address: Elan Pharmaceuticals, South San Francisco, CA 94080, USA.
[5] Present address: Wyeth Pharmaceuticals, Cambridge, MA 02140, USA.
*Correspondence: droberto@mail.med.upenn.edu

DOI 10.1016/j.str.2006.12.005

SUMMARY

The adaptor protein missing-in-metastasis (MIM) contains independent F- and G-actin binding domains, consisting, respectively, of an N-terminal 250 aa IRSp53/MIM homology domain (IMD) and a C-terminal WASP-homology domain 2 (WH2). We determined the crystal structures of MIM's IMD and that of its WH2 bound to actin. The IMD forms a dimer, with each subunit folded as an antiparallel three-helix bundle. This fold is related to that of the BAR domain. Like the BAR domain, the IMD has been implicated in membrane binding. Yet, comparison of the structures reveals that the membrane binding surfaces of the two domains have opposite curvatures, which may determine the type of curvature of the interacting membrane. The WH2 of MIM is longer than the prototypical WH2, interacting with all four subdomains of actin. We characterize a similar WH2 at the C terminus of IRSp53 and propose that in these two proteins WH2 performs a scaffolding function.

INTRODUCTION

Missing-in-metastasis (MIM) and insulin receptor tyrosine kinase substrate p53 (IRSp53) form part of a new family of actin cytoskeleton adaptor proteins (Bompard et al., 2005; Funato et al., 2004; Miki et al., 2000; Woodings et al., 2003). Like most actin binding proteins, MIM and IRSp53 are multidomain proteins, containing protein-protein interaction modules, involved in signaling and localization, and structurally conserved actin binding motifs.

A gene coding for a 356 aa C-terminal fragment of MIM was originally isolated using mRNA differential display, and this fragment was identified as a protein whose expression appeared to be downregulated in certain bladder cancer cell lines (Lee et al., 2002). Full-length MIM was subsequently cloned and shown to contain 759 aa (Woodings et al., 2003). Although it was initially proposed that MIM might function as a metastasis suppressor protein (Lee et al., 2002), this role has not been confirmed (Bompard et al., 2005; Nixdorf et al., 2004). Instead, MIM seems to play a role in cytoskeleton remodeling (Lin et al., 2005; Mattila et al., 2003; Yamagishi et al., 2004), possibly downstream of tyrosine kinase signaling (Gonzalez-Quevedo et al., 2005; Woodings et al., 2003) and Rho-family GTPases (Bompard et al., 2005). MIM localizes to areas of dynamic actin assembly, and its overexpression induces the formation of actin-rich protrusions resembling surface ruffles and microspikes (Woodings et al., 2003). MIM has also been identified as a sonic hedgehog inducible protein that potentiates Gli transcription (Callahan et al., 2004).

MIM is a modular protein (Figure 1A). Its actin binding function can be attributed to two spatially separated actin-binding domains: an N-terminal 250 aa IRSp53/MIM homology domain (IMD) (Yamagishi et al., 2004) and a C-terminal 30 aa WASP-homology domain 2 (WH2) (Mattila et al., 2003). The 475 aa central region sandwiched in between these two actin-binding domains is rich in Pro, Ser, and Thr residues. This region appears to play regulatory/scaffolding roles; it binds receptor protein tyrosine phosphatase δ (RPTPδ) (Gonzalez-Quevedo et al., 2005; Woodings et al., 2003), the transcription factor Gli and the tumor suppressor Sufu (Callahan et al., 2004), and the SH3 domain of cortactin (Lin et al., 2005), a protein implicated in the nucleation and stabilization of Arp2/3-mediated filament branches (Uruno et al., 2001; Weaver et al., 2001).

The relationship between MIM and IRSp53 first emerged from the discovery that the two proteins share similar N-terminal IMDs, an actin-binding domain that has also been implicated in actin bundling (Yamagishi et al., 2004). Like MIM, IRSp53 is an adaptor protein that plays a role in actin cytoskeleton remodeling by linking Rho-family GTPases, such as Rac and Cdc42, to effector proteins, such as Mena (Krugmann et al., 2001) and the Arp2/3 complex activator protein WAVE (Miki et al., 2000). The crystal structure of the IMD of IRSp53 has been determined, consisting of a dimer, with each subunit forming an extended four-helix bundle (Millard et al., 2005).

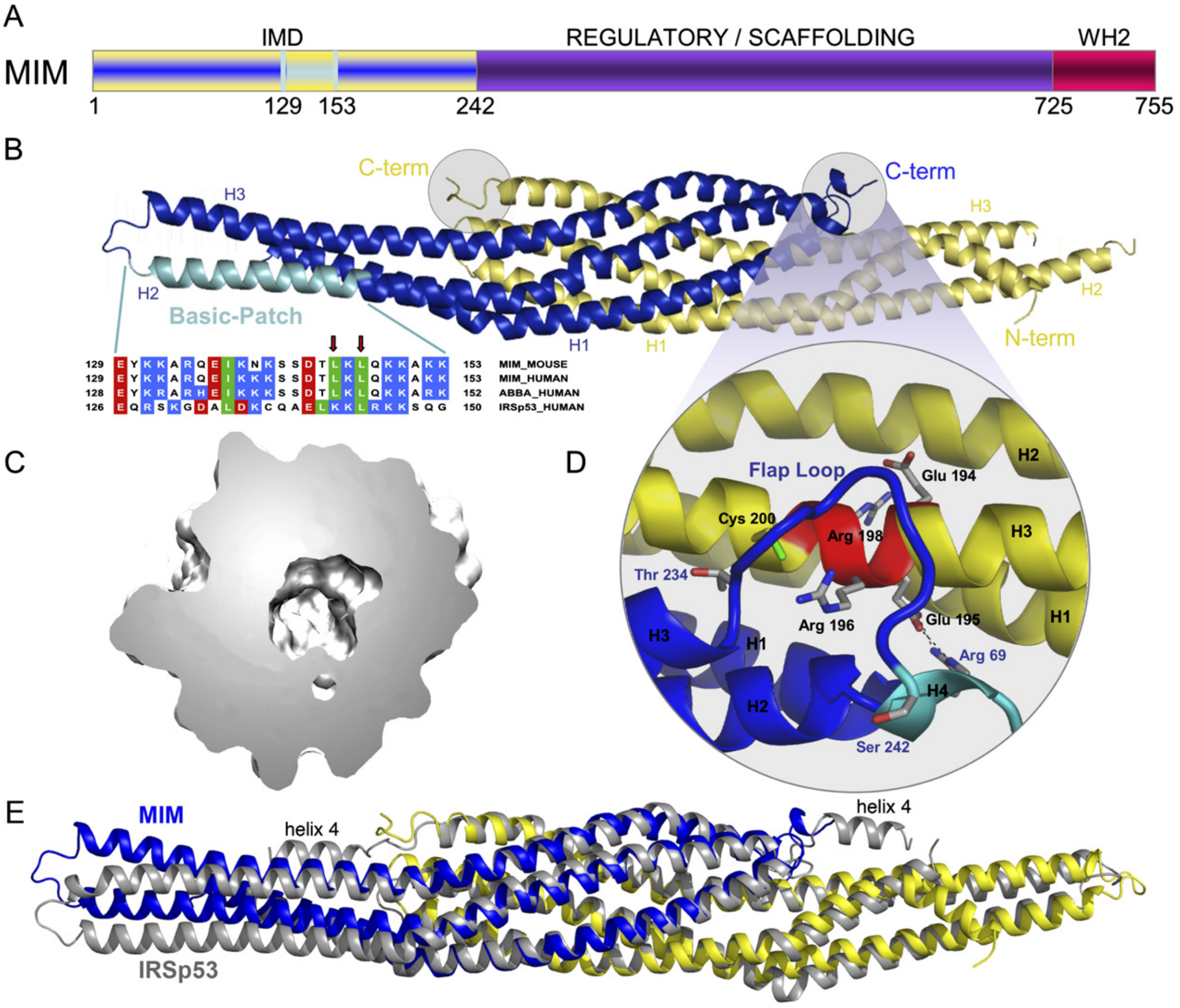

A
IMD
REGULATORY / SCAFFOLDING
WH2
MIM
1
129
153
242
725
755
B
C-term
C-term
H3
H3
H2
H2
Basic-Patch
H1
H1
N-term
129 E Y K K A R Q E I K N K S S D T L K L Q K K A K K 153 MIM_MOUSE
129 E Y K K A R Q E I K K K S S D T L K L Q K K A K K 153 MIM_HUMAN
128 E Y K R A R H E I K K K S S D T L K L Q K K A R K 152 ABBA_HUMAN
126 E Q R S K G D A L D K C Q A E L K K L R K K S Q G 150 IRSp53_HUMAN
C
D
Flap Loop
Glu 194
H2
Cys 200
Arg 198
H3
Thr 234
Glu 195
H1
H3
H1
Arg 196
Arg 69
H4
H2
Ser 242
E
MIM
helix 4
helix 4
IRSp53

Figure 1. Crystal Structure of the IMD of Mouse MIM

(A) Schematic representation of MIM (yellow/blue, IMD; purple, middle regulatory/scaffolding region; red, WH2).

(B) Ribbon representation of the structure of the IMD dimer (figure made with the program PyMOL, http://www.pymol.org). The two subunit of the dimer are colored blue and yellow. Helices 1 to 3 of each subunit are labeled H1, H2, and H3. Also shown is a sequence alignment corresponding to the conserved basic cluster at the symmetric ends of the IMD dimer (highlighted cyan in one of the subunits of the structure). In this alignment, red, blue, and green represent negatively charged, positively charged, and hydrophobic conserved amino acids, respectively. Accession numbers are: MIM_MOUSE, Q8R1S4; MIM_HUMAN, O43312; ABBA_HUMAN, Q765P7; IRSp53_HUMAN, Q9UQB8. Red arrows point to amino acids Leu 145 and Leu 147, which were mutated in this study.

(C) Slice cut through the middle of the molecular surface of the IMD dimer revealing the interior cavity.

(D) Close-up view of the "flap" loop between helices 3 and 4 that covers the "signature sequence" (Yamagishi et al., 2004) of the IMD, which is a charged and conserved sequence that is buried in the structure (red colored area of helix 3).

(E) Superimposition of the structures of the IMDs of MIM and IRSp53. The two structures were superimposed based on the best overlapping central region (amino acids 26–68, 72–110 and 24–66, 69–107 of both chains of MIM and IRSp53, respectively). The view is as in (B) and Figure 3. This orientation highlights the differences between the A chains of the two proteins (blue). Although not well seen from this angle, similar differences occur between the B chains (yellow). Notice that helix 4 of the IMD of IRSp53 is missing in MIM.

Here we describe the crystal structures of the IMD of MIM and that of its WH2 bound to actin. Despite low sequence similarity, the IMDs of MIM and IRSp53 are structurally similar and, therefore, may bind actin and Rac in a similar fashion. The structure of the IMD is generally related to that of the BAR (Bin-Amphiphysin-Rvs) domain, a fold involved in membrane binding (Peter et al., 2004). However, the overall shape of the two domains is markedly different, which probably explains their different roles in membrane curvature sensing. The WH2 of MIM is unusual, both because of its localization in the protein and the way in which it interacts with actin. We characterize a similar WH2 in IRSp53, further expanding the relationship between these two adaptor proteins.

Table 1. Crystallographic Data and Refinement Statistics

		IMD	
	WH2	High-Resolution	Se-Peak
Diffraction Statistics			
Space group	P 2_1 2_1 2_1	P 2_1	P 2_1
Cell parameters			
a, b, c (Å)	42.1, 75.5, 229.0	53.5, 37.3, 129.0	53.5, 37.3, 128.9
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 94.07, 90.0	90.0, 94.08, 90.0
Resolution			
Total (Å)	41.4–2.5	48–1.85	48–2.1
Highest shell (Å)	2.59–2.5	1.92–1.85	2.17–2.1
Completeness (%)	88.7 (80.1)	98.7 (86.7)	99.4 (99.1)
Redundancy	8.9 (5.1)	10.7 (4.5)	6.8 (6.5)
Unique reflections	23,589 (2,069)	43,479 (3,873)	30,686 (3,335)
R merge[a] (%)	7.2 (23.2)	5.7 (38)	8.6 (33.3)
Average I/σ	25.8 (7.2)	30.4 (3.6)	11.8 (3.5)
Refinement Statistics			
R factor[b] (%)	21.7	18.4	
R free[c] (%)	28.4	22.9	
Rms deviations			
Bond length (Å)	0.012	0.014	
Bond angles (°)	1.38	1.13	
Average B factor			
Protein atoms (Å^2)	59.1	33.2	
Solvent atoms (Å^2)	43.7	39.9	
PDB code	2D1K	2D1L	

Values in parentheses correspond to highest resolution shell.

[a] R merge=$\Sigma(I - <I>)/\Sigma I$; I and $<I>$ are the intensity and the mean value of all the measurements of an individual reflection.

[b] R factor=$\Sigma|F_o - F_c|/\Sigma\ |F_o|$; F_o and F_c are the observed and calculated structure factors.

[c] R free; R factor calculated for a randomly selected subset of the reflections (5%) that were omitted during the refinement.

RESULTS

Structure of the IMD of MIM

The crystal structure of the IMD of MIM (N-terminal 250 amino acids) was determined to 1.85 Å resolution, by using the single anomalous dispersion method and X-ray data collected from a Se-Met-substituted crystal (Experimental Procedures and Table 1). The IMD forms a dimer (Figure 1B). The structure is well defined in the electron density map, except for three areas, which are disordered: the last six amino acids of chain A, the last eight amino acids of chain B, and amino acids Asp 155 to Ser 168 of chain B. The electron density map also reveals three amino acids from the expression vector (Ala-Gly-His) at the N-terminal ends of both chains.

Each chain is folded as an extended (~135 Å) antiparallel three-helix bundle (Figure 1B). The two subunits in the dimer are oriented opposite to one another and interact extensively. Thus, the contact area between subunits is 2941 Å^2 (calculated with CCP4 program AreaIMOL, using a 1.4 Å probe). The six helices that comprise the IMD dimer form a twisted ellipsoid ~183 Å in length and ~30 Å in diameter (at the widest point). Despite extensive contacts between adjacent helices, the IMD cannot be classified as a coiled-coil structure. Indeed, an analysis of the structure using the program Socket (Walshaw and Woolfson, 2001) reveals that there exist short, scattered regions of coiled-coil between pairs of helices, but not a single region of the six-helix bundle displays the classical knobs-into-holes layer extending through all the helices.

The dimer features a sizable 1396 Å^3 (calculated with the program Swiss-PDB using a 1.4 Å probe) cavity in the middle (Figure 1C). This cavity contains a number of water molecules. Although the side chains that are directed toward the cavity are predominantly hydrophobic, a number of polar amino acids, including Thr 47, His 86, Glu 213, and Glu 195, also point toward this cavity. These amino acids are involved in interactions among them. Thus, the Oγ of Thr 47 of one chain is hydrogen bonded to the same atom from the other chain. Similarly, Glu 213 and Glu 195 of one chain form salt bridges with His 86 and Arg 69 from the other chain.

The structure of the IMD of IRSp53 has also been determined (Figure 1E) (Millard et al., 2005). The fact that this structure could not be used as a molecular replacement model to determine the current structure suggested from the beginning that important differences were to be expected. Indeed, an alignment of the sequences based on a superimposition of the structures reveals that the IMDs of MIM and IRSp53 share only ~19.3% sequence identity, and although generally similar, the structures superimpose with a relatively large rms deviation of 2.8 Å. The differences remain important within the core (or middle) region (rms deviation 1.64 Å), defined as the region where the two subunits of the dimer overlap (MIM residues 22–119 and 192–235). However, the two structures differ more significantly toward the N- and C termini and the distal ends of the ellipsoid (corresponding to the loop between helices 2 and 3). These differences may be ascribed mainly to increased flexibility in these regions, since the identical molecules that form the IMD dimers also display large rms deviations (2.0 Å for MIM and 1.5 Å for IRSp53).

The IMD of IRSp53 presents a short helix at the C terminus (helix 4). In the structure of MIM's IMD, only one turn of this helix is observed for chain A (Figure 1E), whereas the helix is fully missing in chain B, possibly due to local disorder in the structure (Figure 1B). In contrast, the loop Thr 234–Ser 242 preceding helix 4 (Figure 1D) is well defined in the electron density map for both chains. The conformation of this loop is very similar between the two chains of MIM, as well as between the two chains of IRSp53. However, the conformation of the loop differs quite significantly between the two proteins, which may have functional implications. This loop forms a "flap" that covers the so-called signature sequence of the IMD (Yamagishi et al., 2004) of the other molecule in the dimer (Figure 1D). The signature sequence, EER[R/G]R (MIM residues Glu 194–Arg198), is located within helix 3. MIM presents a Gly at the fourth position of this motif, whereas a bulkier residue (Arg 192) at this position in IRSp53 is directed toward the flap loop and affects its conformation. Another important difference is that a putative disulfide bond in IRSp53 (Millard et al., 2005), between Cys 195 of the signature sequence and Cys 230 of the flap loop, is missing in MIM, which lacks the latter Cys (corresponding to Thr 234 in MIM).

Because the signature sequence forms part of a helix, some of the charged amino acids in this sequence are directed inward, while others are covered by the flap loop (Figure 1D). As a result, MIM amino acids Glu 195, Arg 196, and Arg 198 are all buried in the structure and make electrostatic contacts with main chain atoms, as well as a salt bridge between Glu 195 of one chain and Arg 69 of the other chain. Charged amino acids are rarely buried, and their occurrence typically points to important regions of the structure. The occurrence of buried and charged side chains within the signature sequence and the interaction with the "flap" loop, which connects the IMD to other domains of the protein, suggest an important role for this region of the IMD, possibly in the control or protein-protein interactions involving the IMD.

Actin-Binding and -Bundling Activities of the IMD of MIM

The IMD was originally described as an actin-binding and -bundling domain (Yamagishi et al., 2004). However, the ability of the IMD of human MIM to bundle actin in vitro has lead to conflicting results, ranging from significant bundling (Bompard et al., 2005) to weak (Yamagishi et al., 2004) or no bundling activity (Gonzalez-Quevedo et al., 2005). We decided to test the actin-binding and -bundling activities of the IMD of mouse MIM, which presents a four amino acid insert (^{154}VDAQ^{157}) (Mattila et al., 2003) near what has been described as the actin-binding site (Bompard et al., 2005; Millard et al., 2005). F-actin binding by the IMD of mouse MIM was confirmed using a high-speed cosedimentation experiment (Figure 2A and Experimental Procedures). A similar experiment, carried out at varying IMD concentrations, resulted in a K_d estimate of ~17 μM (Figure 2B). This value is similar to that obtained for IRSp53 (5 μM) (Millard et al., 2005).

A cluster of basic amino acids at the distal ends of the IMD dimer has been implicated in actin binding in IRSp53 (Millard et al., 2005). Although mutations of individual amino acids in this cluster had no effect in actin binding, a construct where Lys residues 142, 143, 146, and 147 were simultaneously mutated to Glu showed somewhat reduced actin binding (10 μM versus 5 μM for the wild-type IMD) (Millard et al., 2005). This result was interpreted as evidence that the basic cluster is involved in actin binding. However, the change in binding affinity appears minor, in particular considering the effect that a substitution of four positive charges by negative charges could have on the general stability and electrostatic properties of the IMD dimer. The IMD of MIM also presents positively charged clusters at the extremes of the dimer (Figure 3A), featuring a total of ten positively charged side chains in the region between Lys 131 and Lys 153 (Figures 1B). The same group reported a similar reduction in the actin binding affinity of MIM's IMD when Lys residues 149, 150, 152, and 153 in this cluster were simultaneously replaced by Asp (Bompard et al., 2005).

Amino acids Ile 137, Leu 145, and Leu 147 (equivalent to IRSp53 Leu 134, Lys 142, and Leu 144) are the only hydrophobic amino acids within the basic cluster of MIM (Figure 1B). While Ile 137 is buried in the structure, Leu 147 is partially exposed and Leu 145 is fully exposed. Given the general importance of hydrophobic amino acids in protein-protein recognition (Jones and Thornton, 1996), we decided to mutate these two Leu residues. Ten mutations were generated, with both leucine residues being replaced individually to Ala, Trp, Arg, and Glu, and simultaneously to Ala or Trp. These mutations were designed to change the electrostatic character of the basic patch, as well as to increase or reduce the size of the two exposed

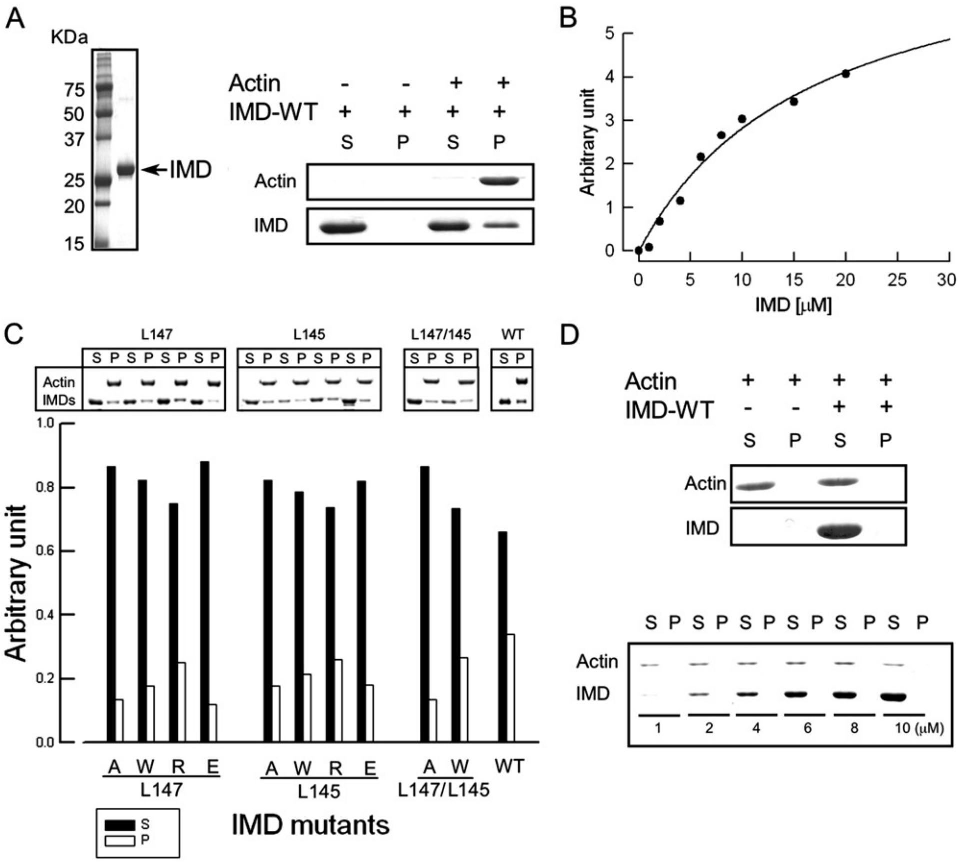


Figure 2. Testing the Actin Binding and Bundling Activities of the IMD of MIM

(A) Purified IMD (left panel) and high-speed F-actin binding assay (right panel). 10 μM aliquots of IMD were ultracentrifuged at 400,000 × g in the presence (+) and the absence (–) of 5 μM F-actin. Equal aliquots of supernatant (S) and pellet (P) were analyzed on a SDS-PAGE gel.

(B) Quantification of the F-actin binding affinity of the IMD. 2 μM F-actin aliquots were incubated with increasing amounts of IMD (0, 1, 2, 4, 6, 8, 10, 15, and 20 μM) and analyzed on gel as in (A). Each point corresponds to the average densitometric reading of three independent experiments. The line represents the best fit of the data to the Michaelis-Menten equation.

(C) High-speed cosedimentation analysis of IMD mutants. 10 μM aliquots of each mutant were incubated with 5 μM F-actin and analyzed on gel (upper panel). The percentage of F-actin-bound IMD was calculated as the IMD fraction in pellet. Before each experiment the mutants were ultracentrifuged at 400,000 × g to eliminate potential aggregates. The lower panel illustrates the percentage of bound (empty bars) versus unbound (black-filled bars) mutants from three independent experiments.

(D) Low-speed analysis of the F-actin bundling activity of the IMD. 5 μM F-actin was centrifuged at 10,000 × g in the absence (–) and the presence (+) of 10 μM wild-type IMD, and the supernatant (S) and pellet (P) were analyzed on gel (upper panel). F-actin stays in the supernatant, indicating the lack of bundling activity. This result was confirmed by experiments carried out at increasing ratios IMD to F-actin (1, 2, 4, 6, 8, 10) (lower panel).

hydrophobic side chains within this area. In addition, we expected these mutations to alter the local conformation of the basic patch, and thereby actin binding. To our surprise, none of the mutations affected actin binding significantly (Figure 2C). More importantly, the double mutations did not contribute additively to a lesser actin-binding efficiency. These results, which are quantitatively similar to those obtained for IRSp53 (Millard et al., 2005), suggest that the basic cluster is unlikely to form a major (or a single) actin-binding site. Thus, although the actin-binding interface appears to involve the region around the basic cluster, other parts of the IMD are most likely also involved. Note further that the structurally related BAR domain (see below) presents a similar cluster of basic amino acids at the distal ends of the dimer (Peter et al., 2004), and yet the BAR domain is not typically implicated in actin binding. Instead, the basic cluster of the BAR domain binds negatively charged phospholipid membranes (Peter et al., 2004). Similarly, the basic cluster of the IMD has been recently implicated in membrane binding (Suetsugu et al., 2006).

We further tested the ability of the IMD of MIM to bundle actin at physiological salt concentration. First, the quality of F-actin for this experiment was checked using low-speed sedimentation (10,000 × g) and rotary-shadowing electron microscopy, to guarantee that no bundles were formed in the absence of the IMD. In contrast with two previous reports (Bompard et al., 2005; Yamagishi et al., 2004), we found that the IMD of MIM did not bundle F-actin under any of the conditions tested (Figure 2D and

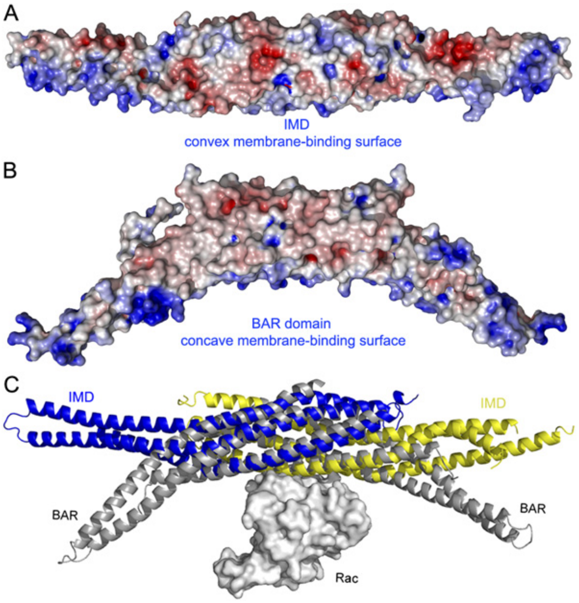

A
IMD
convex membrane-binding surface
B
BAR domain
concave membrane-binding surface
C
IMD
IMD
BAR
BAR
Rac

Figure 3. Structural and Functional Relationship between the IMD and BAR Domains

(A) Electrostatic surface representation of the IMD dimer calculated with the program APBS (Baker et al., 2001) and displayed with the program PyMOL (http://www.pymol.org). Red and blue indicate negatively and positively charged regions, respectively (red, −6 kTe^{-1}; blue +6 kTe^{-1}). Note the positively charged and slightly convex surface, which is thought to mediate the interactions with membranes of the IMD (Suetsugu et al., 2006).

(B) Similar electrostatic representation of the BAR domain of amphiphysin (Peter et al., 2004). The orientation is the same as in (A). Note that the shape of the positively charged membrane binding surface of the BAR domain is concave.

(C) Superimposition of the structures of the IMD of MIM (blue, yellow) with that of the BAR domain of arfaptin complexed with Rac (Tarricone et al., 2001) (gray). The orientation is the same as in (A) and (B). The two folds have different curvatures, but superimpose well in the middle section where the dimers overlap, suggesting that this region may also mediate the binding of Rac in MIM and IRSp53.

Experimental Procedures). This result is in agreement, however, with another report that a slightly longer MIM construct (amino acids 1–277) showed markedly reduced bundling activity compared to the full-length protein (Gonzalez-Quevedo et al., 2005). The disagreement between different laboratories concerning the bundling activity of the IMD may have resulted from nonspecific aggregation of the IMD at low ionic strengths, or the use of F-actin preparations that appear to sediment even in the absence of the IMD construct.

However, it could be also questioned whether the IMD construct studied here is dimeric in solution. To answer this question, we determined the molecular mass of the IMD in solution using multi-angle light scattering combined with particle separation by asymmetric field flow fractionation (Figure S1 and legend; see the Supplemental Data available with this article online). The molar mass of the elution peak determined by this method was 56.7 ± 3.4 kDa, which is in excellent agreement with the expected theoretical mass of the dimer (56.3 kDa). Combined, the lack of bundling activity and the finding that the IMD is a dimer in solution would suggest that the actin binding surface of the IMD spans over the two subunits of the dimer.

Structure of the WH2 of MIM Complexed with Actin

MIM presents a second actin-binding site at the C terminus, consisting of a WH2 domain (Mattila et al., 2003). We made a synthetic peptide corresponding to this WH2, comprising amino acids Asp 724 to the C terminus (Ser 755) of human MIM. With the exception of the first two amino acids, which display conservative mutations, this sequence is identical to that of mouse MIM (Figure 4A). The crystal structure of this WH2 (Figure 4B) was determined as a ternary complex with actin-DNase I. DNase I, which was necessary in order to prevent actin polymerization during crystallization, does not appear to have a significant effect on the actin-binding affinity of WH2 (Chereau et al., 2005), and makes no contacts with the WH2 peptide in the current structure. However, the last two amino acids of the peptide (Phe 754 and Ser 755) are disordered in the structure, and we cannot distinguish whether this is due to a local effect of DNase I, or that these amino acids do not normally interact with actin and therefore become disordered. Also disordered in the structure is the first amino acid of the WH2 peptide, which does not appear to be important for actin binding, since the interactions with actin start after residue Gly 728 of the WH2 peptide.

The structure of the WH2 of MIM can be conceptually subdivided into two parts, an N-terminal amphiphilic helix, comprising amino acids Gly 728 to Gly 738, and a C-terminal extended region from Val 739 to Arg 753 (Figure 4B). As previously shown (Chereau et al., 2005), the most important contribution to the interaction with actin comes from the N-terminal helix that binds in the hydrophobic cleft between actin subdomains 1 and 3. In the current structure, amino acids Met 731, Leu 732, and Ile 735, on the hydrophobic side of this helix, are embedded within the hydrophobic cleft in actin (Figure 4B). The extended portion of the WH2 of MIM follows a path alongside the actin surface, climbing to the top of actin subdomains 2 and 4. Amino acids Val 739 and Leu 741 within this region bind in a hydrophobic pocket on the actin surface formed by amino acids Ile 341, Ile 345, and Leu 349. Residue Leu 741 of the WH2 peptide forms part of the canonical LKKT sequence, found in other actin-binding proteins, such as thymosin β4 (Paunola et al., 2002) and the linker region between gelsolin domains 1 and 2 (Irobi et al., 2003). Interestingly, the two Lys residues of this canonical sequence bind atop actin residues Asp 24 and Asp 25, but interact with actin only via main chain atoms. Thus, the main-chain nitrogen and oxygen atoms of Lys 742 are hydrogen bonded to the oxygen and nitrogen atoms of actin residues Gly 23 and Asp 25, respectively. The other important element of the interaction for this part of the WH2 peptide involves the segment Thr 746 to Arg 749, which is incorporated as an additional β strand into a β sheet in actin subdomain 1 (running parallel to actin β strand Arg 28 to Phe 31). Finally, Arg 749 of the WH2

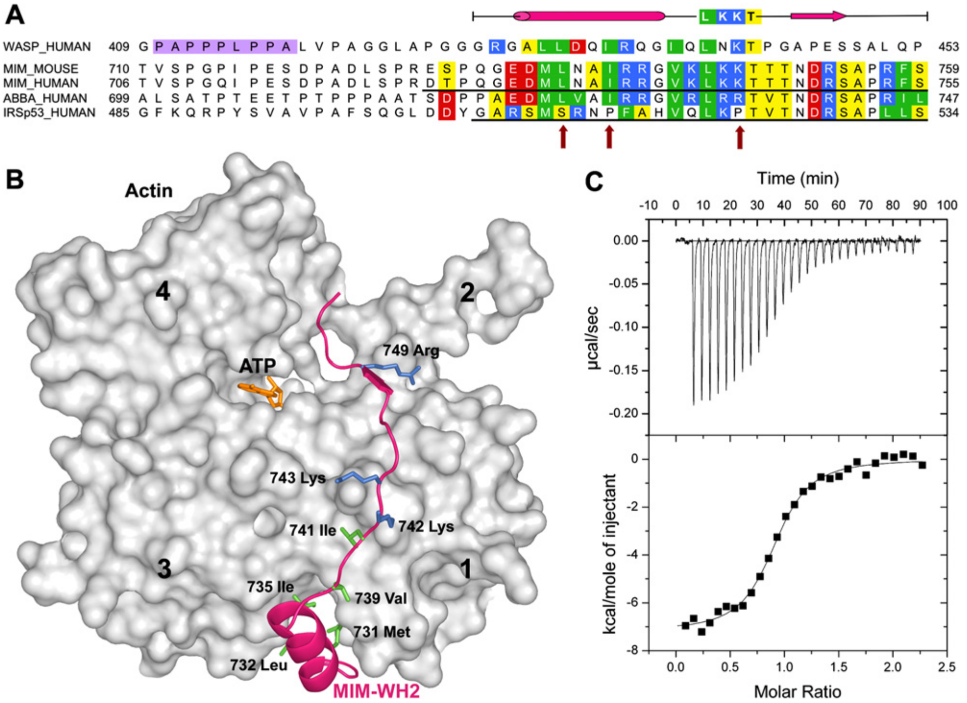


Figure 4. The WASP-Homology Domain 2 of MIM and IRSp53

(A) Comparison of a classical WH2 (represented by WASP, Wiskott-Aldrich syndrome protein) with the WH2s of MIM, ABBA, and IRSp53. Red, blue, green, and yellow correspond to negatively charged, positively charged, hydrophobic, and small (Thr, Val, Ser, Ala) conserved amino acids, respectively. The diagram above the sequences represents a secondary structure assignment based on the structure determined here (cylinder, α helix; arrow, β strand). Accession numbers are as in Figure 1, and WASP_HUMAN, P42768. Red arrows point to noncanonical amino acids present in the WH2 of IRSp53.

(B) Structure of the WH2 of MIM (red ribbon) bound to actin (gray surface). Numbers 1–4 indicate actin's four subdomains. The side chains of some of the amino acids involved in interactions with actin are shown (green, hydrophobic; blue, positively charged).

(C) Binding of the WH2 of IRSp53 to actin measured by ITC. The upper graph corresponds to the heat evolved upon repeated 10 µl injections of a 100 µM solution of the WH2 peptide into a 10 µM solution of actin in G buffer. The lower graph shows the binding isotherm produced by integration of the heat for each injection. The line represents a nonlinear least squares fit to the data using a single-site binding model. The following thermodynamic parameters were determined from the fitting: dissociation constant K_d = 0.28 ± 0.04 µM; molar enthalpy ΔH = −7.2 ± 0.1 kcal.mol^{-1}; and stoichiometry n = 0.9.

peptide forms a salt bridge with actin residue Glu 93. The remaining portion of WH2 extends across the cleft between actin subdomains 2 and 4, and appears to interact only weakly with actin, which is consistent with the limited contribution of this portion of WH2 to the actin binding affinity (Chereau et al., 2005).

An Alternatively Spliced WH2 at the C Terminus of IRSp53

IRSp53 is another actin cytoskeleton adaptor protein, which like MIM presents an N-terminal IMD. Six isoforms of IRSp53 have been identified. In addition to the IMD, all six isoforms present identical CRIB, SH3, and WW protein-protein interaction modules. The differences between isoforms occur at the C termini. Two of the isoforms present a C-terminal extension consisting of a PDZ binding sequence (Soltau et al., 2004), another two present WH2-related extensions, and the remaining two appear to have no functionally identifiable extensions. The WH2 of IRSp53 is unusual (Figure 4A). While the C-terminal portion of this WH2 is nearly identical to that of MIM, the N-terminal helix, known to play a critical role in actin binding (Chereau et al., 2005), has a noncanonical sequence. Indeed, the periodicity of hydrophobic amino acids in the segment corresponding to the helix is altered, and there are conserved Pro and Gly residues in this region, which could prevent the formation of a helix. In addition, one of the Lys residues in the canonical LKKT sequence is replaced by Pro in IRSp53 (Figure 4A). Taken together, these observations raised doubts about the capacity of this WH2 to bind actin. We decided to study the binding of this WH2 to actin using ITC (Figure 4C). A peptide corresponding to the WH2 of human IRSp53 isoform 2, amino acids Gly 506 to Ser 534, was synthesized (Figure 4A). The actin-binding affinity of this peptide (K_d = 0.28 µM) was found to be surprisingly similar to that measured previously by us under identical conditions for the WH2 of MIM (K_d = 0.23 µM) (Chereau et al., 2005). Therefore, we conclude that, like MIM, certain isoforms of IRSp53 present two independent actin-binding domains at the N- and

C-terminal ends, which in both proteins enclose a large central region featuring various protein-protein interaction modules.

DISCUSSION

The IMD of MIM is an all α-helical structure, which dimerizes to form a twisted ellipsoid ~183 Å in length, with a large cavity in the middle (Figure 1). Despite low sequence similarity, the structures of the IMDs of MIM and IRSp53 (Millard et al., 2005) are generally similar. The loop following helix 3 of MIM's IMD forms a "flap" that covers the so-called "signature sequence" of the IMD, a conserved and charged sequence that is conspicuously buried in the structure (Figure 1D).

While we were able to confirm that the IMD binds F-actin with ~17 μM affinity (Figures 2A and 2B), we found that the symmetric patches of basic amino acids at the distal ends of the dimer (Figure 3A) play only a limited role in this interaction (Figure 2C). Furthermore, the IMD is also a dimer in solution (Figure S1), but it does not bundle actin (Figure 2D), as would have been expected if the symmetric ends of the dimer were solely responsible for actin binding. If, as previously suggested (Yamagishi et al., 2004), MIM is an actin-bundling protein, this function may require other parts of the molecule that lie outside the IMD. Gonzalez-Quevedo et al. (2005) reached a similar conclusion by studying various fragments of MIM. They showed that most of the bundling activity could be restored by a construct comprising amino acids 1–408 of MIM. Another possibility is that bundling is regulated (or potentiated) in vivo by still unknown factors.

The IMD of IRSp53 interacts with Rac, possibly functioning as an intermediate for the activation of WAVE, which is recruited by the SH3 domain of IRSp53 (Miki et al., 2000). Similarly, the IMD of MIM has been shown to bind and activate Rac, suggesting that MIM could link Rac to effector proteins involved in lamellipodia formation, such as WAVE (Bompard et al., 2005). The structural basis for the Rac-IMD interaction is unknown. Interestingly, the structure of the IMD resembles that of the BAR domain, which also binds small GTPases (Habermann, 2004). The crystal structures of various BAR-domain proteins, including arfaptin (Tarricone et al., 2001), amphiphysin (Peter et al., 2004) and endophilin (Weissenhorn, 2005), have been determined. Although the BAR domain is curved and the IMD is relatively straight, the two folds superimpose remarkably well in the middle section, where the two subunits that conform these two domains overlap (Figure 3C). It is via this well-overlapping middle section that the binding of small GTPases appears to take place. Indeed, the structure of arfaptin was also determined bound to Rac (Tarricone et al., 2001). One molecule of Rac sits at the midpoint of the arfaptin BAR dimer. It is likely that the IMDs of MIM and IRSp53 bind Rac in a similar fashion, as illustrated by a superimposition of the structures of MIM and arfaptin-Rac (Figure 3C). Note, however, that this superimposition does not represent an accurate model of the interaction, since there is no obvious sequence similarity between the IMD and BAR domains and local changes are likely.

The binding of Rac and actin by the IMD of MIM appear to be mutually exclusive (Bompard et al., 2005). Although this study did not determine the total extent of the actin-binding interface, the lack of bundling activity (Figure 2D) and the fact that the distal ends of the IMD dimer do not constitute a major actin-binding site (Figure 2C) would suggest that the middle section of the IMD dimer also participates in actin binding. As suggested by the analogy with the BAR domain, the binding of Rac may also involve the middle section of the IMD dimer (Figure 3C), possibly explaining why actin and Rac bind in a mutually exclusive manner.

We have stressed here the striking resemblance between the IMD and BAR folds, including their shared ability to bind small GTPases. In addition, both domains present similar clusters of positively charged amino acids (Figures 3A and 3B), which in the BAR domain coincide with the concave surface of the dimer and are involved in phospholipid membrane binding (Peter et al., 2004). The most noticeable difference between the two folds is that the IMD forms relatively straight dimers (Millard et al., 2005), whereas the BAR domain forms curved, "banana-shaped" dimers (Peter et al., 2004; Tarricone et al., 2001; Weissenhorn, 2005). However, the curvature of the BAR domain varies from protein to protein (arfaptin > amphiphysin > endophilin), which may facilitate the binding to membranes with different curvatures. The IMD was discovered independently and due to the lack of sequence similarity was not originally considered a member of the BAR domain family (Yamagishi et al., 2004). A comparison of the structures of the IMD and BAR domains would now suggest that the two domains are not only structurally but also functionally related to each other (Figure 3). Indeed, it was recently reported that, like the BAR domain, the IMD also binds membranes and that this function is mediated by the clusters of basic amino acids at the distal ends of the dimer (Suetsugu et al., 2006; P. Lappalainen, personal communication). Interestingly, the directionality of membrane deformation by the IMD (outward) was found to be opposite to that produced by the BAR domain (inward). The structures may provide an explanation for this observation, because the concave and positively charged surface implicated in membrane binding in the BAR domain adopts a somewhat convex shape in the IMD (Figure 3). Therefore, the evidence to date suggests that the IMD is a multifunctional module, linking the actin cytoskeleton to the formation of membrane protrusions by direct interactions with both F-actin and membranes, all under the control of the small GTPase Rac.

The WH2 of MIM interacts with all four subdomains of actin (Figure 4B). It consists of an N-terminal amphiphilic helix that binds in the cleft between actin subdomains 1 and 3 and a C-terminal extended region that binds along the actin surface and the nucleotide cleft, reaching the top of actin subdomains 2 and 4. Note that the end of this WH2 coincides with the C terminus of the MIM protein. The prototypical WH2 found among WASP-family proteins tends to be shorter (Figure 4A) and presents few or no interactions with actin after the LKKT sequence (Chereau et al., 2005).

We demonstrated here that certain isoforms of IRSp53 present a C-terminal WH2 that binds actin with similar affinity to that of the WH2 of MIM, further extending the relationship between these two actin-cytoskeleton scaffolding proteins. WH2 is the smallest actin-binding motif known. Based on their sequences and structures, we have identified two types of WH2s: long and short (Chereau et al., 2005). Short WH2s consist solely of the N-terminal helix and the LKKT-related sequence (for example, WASP's WH2; Figure 4A). Long WH2s present an additional ~10 amino acids at the C terminus. The extra amino acids of long WH2s share sequence similarity with Tβ4 and make similar contacts with actin (Irobi et al., 2004), supporting a previously proposed relationship between the WH2 and Tβ families (Paunola et al., 2002). However, it remains unclear whether the extra amino acids of long WH2s play any specific role, since they do not seem to contribute significantly to the actin binding affinity nor the nucleotide exchange inhibition by actin (Chereau et al., 2005).

What is the role of WH2 in MIM and IRSp53? WH2 could serve two possible functions: recruit actin monomers, or recruit a protein to a specific actin cytoskeletal network. Actin filament nucleation and elongation factors, including WASP, Ena/VASP and spire, form the main group of WH2-containing proteins. These proteins present short WH2s, typically positioned C-terminal to Pro-rich sequences (Figure 4A). In WASP, WH2 is followed by the central (or C) region that binds one of the subunits of Arp2/3 complex, whereas, in VASP, WH2 is known as the G-actin binding domain (GAB) and is followed by the F-actin binding domain (FAB). The C region of WASP and the FAB domain of VASP are related to each other, and both constitute specialized forms of WH2 (Chereau and Dominguez, 2006). Spire, on the other hand, contains four WH2s in tandem (Quinlan et al., 2005). We have proposed that in these proteins WH2 becomes involved in nucleation and elongation by bridging actin subunits along a single filament strand and by mediating the incorporation of profilin-actin at the barbed end of growing filaments (Chereau and Dominguez, 2006; Chereau et al., 2005). So far, we have identified long WH2s in actobindin, WIP, MIM (Chereau et al., 2005), and now in IRSp53. It appears that, in MIM and IRSp53, WH2 occurs within a different domain organization than in most cytoskeletal proteins (Figure 4A). Thus, in MIM and IRSp53, WH2 is found in isolation at the C-terminal end; i.e. not immediately preceded by Pro-rich sequences nor followed by other WH2s (or WH2-related sequences). Unlike the actin monomer-trapping molecule Tβ4 and the nucleation-elongation factors described above, MIM and IRSp53 function as scaffolding proteins. It is therefore likely that WH2 helps recruit MIM and IRSp53, as well as their multiple binding partners, to specific cytoskeletal networks. Consistent with this idea, images of cells overexpressing full-length MIM show a significant loss of stress fibers (Gonzalez-Quevedo et al., 2005; Mattila et al., 2003; Woodings et al., 2003), but this effect appears diminished for MIM constructs lacking the WH2 region (Bompard et al., 2005; Gonzalez-Quevedo et al., 2005).

What is the spatial relationship between the IMD and WH2 domains? Hydrophobic cluster analysis (Callebaut et al., 1997) suggests that the region sandwiched in between the IMD and WH2 of MIM is mostly unstructured, with only two segments with predicted globular or inducible folding (Figure S2). Given these characteristics and the antiparallel organization of the IMD dimer, the two WH2s could be located far apart from each other in the protein, which would imply a lack of communication between them. More likely, however, the various domains of MIM and IRSp53 fold back into a more compact structure, possibly mediated by autoregulatory interactions involving the IMD and other parts of the molecule.

EXPERIMENTAL PROCEDURES

Preparation of Proteins and Peptides

The cDNA encoding for full-length mouse MIM was purchased from ATCC. Amino acids 1–250, corresponding to MIM's IMD, were amplified by PCR and inserted into vector pTYB12 (New England Biolabs). This vector comprises a chitin affinity purification tag and an intein self-cleavage domain. IMD mutants (Leu 145 to Ala, Trp, Asp, Arg; Leu 147 to Ala, Trp, Asp, Arg; and the double mutants Leu 145 and Leu 147 to Ala; Leu 145 and Leu 147 to Trp) were generated using the QuikChange Site-Directed Mutagenesis Kit (Stratagene).

BL21(DE3) cells (Invitrogen) were transformed with the various IMD constructs and grown in LB medium at 37°C until the OD at 600 nm reached a value of 0.8. Expression was induced by addition of 1 mM isopropylthio-β-D-galactoside (IPTG) and carried out overnight at 20°C. Cells were harvested by centrifugation and resuspended in chitin-affinity-column equilibration buffer (20 mM Tris [pH 7.5], 500 mM NaCl, 1 mM EDTA, 100 μM PMSF), followed by standard purification on a chitin affinity column at 4°C (New England Biolabs manual). The proteins were eluted from this column following DTT-induced self-cleavage of the intein. The proteins were then dialyzed against 20 mM Tris (pH 7.5), 50 mM NaCl, and 1 mM DTT and purified to homogeneity on a MonoQ column (Pharmacia). A Se-Met-substituted IMD protein was obtained following a similar procedure by growing cells in M9 media, supplemented with 70 mg/ml Se-Met. Actin was prepared from rabbit muscle as described (Graceffa and Dominguez, 2003). Ultra-pure-grade bovine pancreatic DNase I was purchased from BioWorld. WH2 domains, corresponding to human $MIM_{724–755}$ and human $IRSp53_{506–534}$, were synthesized on an ABI431 peptide synthesizer and purified by HPLC. The concentrations of the peptides were determined by amino acid analysis (Dana-Farber Cancer Institute, Boston, MA).

F-Actin-Binding Assay

For this experiment actin underwent additional purification through a gel filtration Sephacryl S300HR column (Pharmacia). Actin in G buffer (2 mM Tris [pH 7.4], 0.2 mM $CaCl_2$, 0.2 mM ATP, 1 mM DTT, 1 mM NaN_3) was polymerized by addition of 100 mM KCl, 2 mM $MgCl_2$, and 2 mM EGTA (F buffer). IMD and IMD mutants were dialyzed against the same F buffer and centrifuged at 400,000 × g for 30 min before the experiments, to remove potential aggregates. 10 μM IMD (or IMD mutants) was mixed with 5 μM F-actin on ice and incubated for 30 min. The protein mixtures were then centrifuged at 400,000 × g for 30 min. Equal volumes of supernatant and pellet were analyzed on a 15% SDS-PAGE gel. A quantification of the F-actin binding affinity was obtained by adding increasing amounts of IMD (0, 1, 2, 4, 6, 8, 10, 15, and 20 μM) to 2 μM F-actin and analyzed as described above.

The Coomassie blue-stained gels were then scanned at high resolution, and the intensities of the bands were quantified using the program ImageJ version 1.34S (http://rsb.info.nih.gov/).

F-Actin-Bundling Assay

Actin was purified and polymerized as described above. Before each bundling experiment, both F-actin and the IMD were centrifuged for 30 min at 10,000 and 400,000 × g, respectively. Note that this step is important in order to remove potential high molecular weight aggregates. F-actin samples were then visualized on a Philips EM 300 electron microscope to ensure that actin bundles were not present prior to the addition of the IMD. 10 μM IMD was then mixed with 5 μM F-actin and incubated for 1 hr at room temperature. Next, the mixture was centrifuged at low speed (10,000 × g) for 30 min at room temperature. Supernatant and pellet were analyzed by SDS-PAGE gel as described above. A quantitative bundling assay was done in a similar way by adding increasing amounts of IMD (1, 2, 4, 6, 8, and 10 μM) to 1 μM F-actin.

Crystallization, Data Collection and Structure Determination

Actin-DNase I complex at a 1:1 molar ratio was mixed with MIM's WH2 domain peptide at 1.5 molar excess. The ternary complex was then dialyzed against G-buffer (2 mM Tris [pH 7.5], 0.2 mM $CaCl_2$, 0.2 mM ATP, 1 mM NaN_3) and concentrated to ~10 mg/ml using a Centricon device (Millipore). The complex was crystallized at 20°C, using the hanging drop vapor diffusion method. The 2 μl hanging drops consisted of a 1:1 (v/v) mixture of protein solution and a well solution containing 13%–14% PEG 3350, 50 mM Na cacodylate (pH 6.8–7.2), and 100 mM Na-formate. Before crystallization MIM's IMD was dialyzed against 20 mM Tris (pH 7.4), 50 mM NaCl, and 5 mM DTT and concentrated to ~20 mg/ml. Crystals were obtained at 4°C in 0.1 M Tris (pH 7.4), 0.2 M LiCl, and 16% PEG 2000 MME. The crystals were flash-frozen in propane, using 25% glycerol as cryoprotectant. X-ray data sets were collected at the BioCARS beamlines 14-BM-D (IMD) and 14-BM-C (actin-WH2) at the Advance Photon Source (Argonne, IL). The data sets were indexed and scaled with program HKL-2000 (HKL Research, Inc.). The structure of the actin-WH2 complex was determined by molecular replacement using CCP4 program AMoRe and the structure of WIP-actin-DNase I as a search model (Chereau et al., 2005). Model building and refinement were done with the program Coot (Emsley and Cowtan, 2004) and CCP4 program Refmac (Table 1).

The structure of the IMD was determined from the anomalous signal of a Se-Met-substituted crystal, using the single anomalous dispersion (SAD) method and data collected to 2.1 Å resolution at the absorption peak wavelength of the Se atoms (Table 1). The positions of 11 out of 18 Se atoms in the structure were found with the program SnB (Weeks and Miller, 1999). These positions were then refined and phases were calculated with the program Solve (Terwilliger and Berendzen, 1999). About 70% of the model was built automatically with the program ARP/wARP (Morris et al., 2003) and using a 1.85 Å resolution X-ray dataset collected from the same Se-Met substituted crystal. Further model building and refinement were carried out with the program Coot (Emsley and Cowtan, 2004) and CCP4 program Refmac. Data collection and refinement statistics are given in Table 1.

Isothermal Titration Calorimetry

These measurements were done using a VP-ITC (MicroCal, Northampton, MA). To determine Δ*H* and Ka of WH2-actin association, the WH2 peptide of IRSp53 at a concentration of 100 μM was titrated, in 10 μl injections, into 1.44 ml of 10 μM actin in G buffer at 25°C. The duration of each injection was 10 s, with an interval of 3 min between injections. The heat of binding was corrected for the small exothermic heat of injection, determined by injecting WH2 peptides into buffer. Data were analyzed using MicroCal's Origin program.

Supplemental Data

Supplemental Data include two figures and are available at http://www.structure.org/cgi/content/full/15/2/145/DC1/.

ACKNOWLEDGMENTS

This work was supported by NIH grant GM073791. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract No. W-31-109-Eng-38. Use of the BioCARS facilities was supported by NIH grant RR07707. The authors thank Andrei Lupas for important comments about coiled-coil structures and Pekka Lappalainen for helpful discussions.

Received: October 30, 2006
Revised: December 8, 2006
Accepted: December 15, 2006
Published: February 13, 2007

REFERENCES

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA *98*, 10037–10041.

Bompard, G., Sharp, S.J., Freiss, G., and Machesky, L.M. (2005). Involvement of Rac in actin cytoskeleton rearrangements induced by MIM-B. J. Cell Sci. *118*, 5393–5403.

Callahan, C.A., Ofstad, T., Horng, L., Wang, J.K., Zhen, H.H., Coulombe, P.A., and Oro, A.E. (2004). MIM/BEG4, a Sonic hedgehog-responsive gene that potentiates Gli-dependent transcription. Genes Dev. *18*, 2724–2729.

Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B., and Mornon, J.P. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell. Mol. Life Sci. *53*, 621–645.

Chereau, D., and Dominguez, R. (2006). Understanding the role of the G-actin-binding domain of Ena/VASP in actin assembly. J. Struct. Biol. *155*, 195–201.

Chereau, D., Kerff, F., Graceffa, P., Grabarek, Z., Langsetmo, K., and Dominguez, R. (2005). Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. Proc. Natl. Acad. Sci. USA *102*, 16644–16649.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Funato, Y., Terabayashi, T., Suenaga, N., Seiki, M., Takenawa, T., and Miki, H. (2004). IRSp53/Eps8 complex is important for positive regulation of Rac and cancer cell motility/invasiveness. Cancer Res. *64*, 5237–5244.

Gonzalez-Quevedo, R., Shoffer, M., Horng, L., and Oro, A.E. (2005). Receptor tyrosine phosphatase-dependent cytoskeletal remodeling by the hedgehog-responsive gene MIM/BEG4. J. Cell Biol. *168*, 453–463.

Graceffa, P., and Dominguez, R. (2003). Crystal structure of monomeric actin in the atp state: structural basis of nucleotide-dependent actin dynamics. J. Biol. Chem. *278*, 34172–34180.

Habermann, B. (2004). The BAR-domain family of proteins: a case of bending and binding? EMBO Rep. *5*, 250–255.

Irobi, E., Aguda, A.H., Larsson, M., Guerin, C., Yin, H.L., Burtnick, L.D., Blanchoin, L., and Robinson, R.C. (2004). Structural basis of actin sequestration by thymosin-beta4: implications for WH2 proteins. EMBO J. *23*, 3599–3608.

Irobi, E., Burtnick, L.D., Urosev, D., Narayan, K., and Robinson, R.C. (2003). From the first to the second domain of gelsolin: a common path on the surface of actin? FEBS Lett. *552*, 86–90.

Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. Proc. Natl. Acad. Sci. USA *93*, 13–20.

Krugmann, S., Jordens, I., Gevaert, K., Driessens, M., Vandekerckhove, J., and Hall, A. (2001). Cdc42 induces filopodia by promoting the formation of an IRSp53:Mena complex. Curr. Biol. *11*, 1645–1655.

Lee, Y.G., Macoska, J.A., Korenchuk, S., and Pienta, K.J. (2002). MIM, a potential metastasis suppressor gene in bladder cancer. Neoplasia *4*, 291–294.

Lin, J., Liu, J., Wang, Y., Zhu, J., Zhou, K., Smith, N., and Zhan, X. (2005). Differential regulation of cortactin and N-WASP-mediated actin polymerization by missing in metastasis (MIM) protein. Oncogene *24*, 2059–2066.

Mattila, P.K., Salminen, M., Yamashiro, T., and Lappalainen, P. (2003). Mouse MIM, a tissue-specific regulator of cytoskeletal dynamics, interacts with ATP-actin monomers through its C-terminal WH2 domain. J. Biol. Chem. *278*, 8452–8459.

Miki, H., Yamaguchi, H., Suetsugu, S., and Takenawa, T. (2000). IRSp53 is an essential intermediate between Rac and WAVE in the regulation of membrane ruffling. Nature *408*, 732–735.

Millard, T.H., Bompard, G., Heung, M.Y., Dafforn, T.R., Scott, D.J., Machesky, L.M., and Futterer, K. (2005). Structural basis of filopodia formation induced by the IRSp53/MIM homology domain of human IRSp53. EMBO J. *24*, 240–250.

Morris, R.J., Perrakis, A., and Lamzin, V.S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. Methods Enzymol. *374*, 229–244.

Nixdorf, S., Grimm, M.O., Loberg, R., Marreiros, A., Russell, P.J., Pienta, K.J., and Jackson, P. (2004). Expression and regulation of MIM (Missing In Metastasis), a novel putative metastasis suppressor gene, and MIM-B, in bladder cancer cell lines. Cancer Lett. *215*, 209–220.

Paunola, E., Mattila, P.K., and Lappalainen, P. (2002). WH2 domain: a small, versatile adapter for actin monomers. FEBS Lett. *513*, 92–97.

Peter, B.J., Kent, H.M., Mills, I.G., Vallis, Y., Butler, P.J., Evans, P.R., and McMahon, H.T. (2004). BAR domains as sensors of membrane curvature: the amphiphysin BAR structure. Science *303*, 495–499.

Quinlan, M.E., Heuser, J.E., Kerkhoff, E., and Mullins, R.D. (2005). Drosophila Spire is an actin nucleation factor. Nature *433*, 382–388.

Soltau, M., Berhorster, K., Kindler, S., Buck, F., Richter, D., and Kreienkamp, H.J. (2004). Insulin receptor substrate of 53 kDa links postsynaptic shank to PSD-95. J. Neurochem. *90*, 659–665.

Suetsugu, S., Murayama, K., Sakamoto, A., Hanawa-Suetsugu, K., Seto, A., Oikawa, T., Mishima, C., Shirouzu, M., Takenawa, T., and Yokoyama, S. (2006). The RAC-binding domain/IRSP53-MIM homology domain of IRSP53 induces RAC-dependent membrane deformation. J. Biol. Chem. *281*, 35347–35358.

Tarricone, C., Xiao, B., Justin, N., Walker, P.A., Rittinger, K., Gamblin, S.J., and Smerdon, S.J. (2001). The structural basis of Arfaptin-mediated cross-talk between Rac and Arf signalling pathways. Nature *411*, 215–219.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Uruno, T., Liu, J., Zhang, P., Fan, Y., Egile, C., Li, R., Mueller, S.C., and Zhan, X. (2001). Activation of Arp2/3 complex-mediated actin polymerization by cortactin. Nat. Cell Biol. *3*, 259–266.

Walshaw, J., and Woolfson, D.N. (2001). Socket: a program for identifying and analysing coiled-coil motifs within protein structures. J. Mol. Biol. *307*, 1427–1450.

Weaver, A.M., Karginov, A.V., Kinley, A.W., Weed, S.A., Li, Y., Parsons, J.T., and Cooper, J.A. (2001). Cortactin promotes and stabilizes Arp2/3-induced actin filament network formation. Curr. Biol. *11*, 370–374.

Weeks, C.M., and Miller, R. (1999). The design and implementation of SnB v2.0. J. Appl. Crystallogr. *32*, 120–124.

Weissenhorn, W. (2005). Crystal structure of the endophilin-A1 BAR domain. J. Mol. Biol. *351*, 653–661.

Woodings, J.A., Sharp, S.J., and Machesky, L.M. (2003). MIM-B, a putative metastasis suppressor protein, binds to actin and to protein tyrosine phosphatase delta. Biochem. J. *371*, 463–471.

Yamagishi, A., Masuda, M., Ohki, T., Onishi, H., and Mochizuki, N. (2004). A novel actin bundling/filopodium-forming domain conserved in insulin receptor tyrosine kinase substrate p53 and missing in metastasis protein. J. Biol. Chem. *279*, 14929–14936.

Accession Numbers

Atomic coordinates have been deposited in the Protein Data Bank, http://www.pdb.org (PDB ID codes 2D1L and 2D1K).

B. Structures des Nucléoprotéines de Phlébovirus

The Hexamer Structure of the Rift Valley Fever Virus Nucleoprotein Suggests a Mechanism for its Assembly into Ribonucleoprotein Complexes

François Ferron[1], Zongli Li[2,3], Eric I. Danek[2], Dahai Luo[4], Yeehwa Wong[4], Bruno Coutard[1], Violaine Lantez[1], Rémi Charrel[5], Bruno Canard[1], Thomas Walz[2,3], Julien Lescar[1,4]*

1 Architecture et Fonction des Macromolécules Biologiques, Marseille, France, 2 Harvard Medical School, Department of Cell Biology, Boston, Massachusetts, United States of America, 3 Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, United States of America, 4 Nanyang Technological University, School of Biological Sciences, Singapore, 5 Unité des Virus Emergents, Université Aix-Marseille II et Institut de Recherche pour le Développement, Marseille, France

Abstract

Rift Valley fever virus (RVFV), a *Phlebovirus* with a genome consisting of three single-stranded RNA segments, is spread by infected mosquitoes and causes large viral outbreaks in Africa. RVFV encodes a nucleoprotein (N) that encapsidates the viral RNA. The N protein is the major component of the ribonucleoprotein complex and is also required for genomic RNA replication and transcription by the viral polymerase. Here we present the 1.6 Å crystal structure of the RVFV N protein in hexameric form. The ring-shaped hexamers form a functional RNA binding site, as assessed by mutagenesis experiments. Electron microscopy (EM) demonstrates that N in complex with RNA also forms rings in solution, and a single-particle EM reconstruction of a hexameric N-RNA complex is consistent with the crystallographic N hexamers. The ring-like organization of the hexamers in the crystal is stabilized by circular interactions of the N terminus of RVFV N, which forms an extended arm that binds to a hydrophobic pocket in the core domain of an adjacent subunit. The conformation of the N-terminal arm differs from that seen in a previous crystal structure of RVFV, in which it was bound to the hydrophobic pocket in its own core domain. The switch from an intra- to an inter-molecular interaction mode of the N-terminal arm may be a general principle that underlies multimerization and RNA encapsidation by N proteins from *Bunyaviridae*. Furthermore, slight structural adjustments of the N-terminal arm would allow RVFV N to form smaller or larger ring-shaped oligomers and potentially even a multimer with a super-helical subunit arrangement. Thus, the interaction mode between subunits seen in the crystal structure would allow the formation of filamentous ribonucleocapsids *in vivo*. Both the RNA binding cleft and the multimerization site of the N protein are promising targets for the development of antiviral drugs.

Citation: Ferron F, Li Z, Danek EI, Luo D, Wong Y, et al. (2011) The Hexamer Structure of the Rift Valley Fever Virus Nucleoprotein Suggests a Mechanism for its Assembly into Ribonucleoprotein Complexes. PLoS Pathog 7(5): e1002030. doi:10.1371/journal.ppat.1002030

Editor: Félix A. Rey, Institut Pasteur, France

Received December 13, 2010; **Accepted** March 3, 2011; **Published** May 12, 2011

Funding: This work was supported by BMRC grants (08/1/22/19/589) a CRP grant CRP4-2008-2 and an ATIP from the CNRS to the laboratory of JL. This work was also supported by the European Virus Archive (EVA) project (European FP7 Capacities Project no 228292, http://www.european-virus-archive.com/). TW is an investigator in the Howard Hughes Medical Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Julien.Lescar@afmb.univ-mrs.fr

Introduction

The *Bunyaviridae* family comprises more than 330 viruses that affect vertebrates and plants. La Crosse virus, a member of the *Orthobunyavirus* genus, causes pediatric viral encephalitis in North America. The *Bunyaviridae* family also includes several other emerging human pathogens, such as the Hantaan and Sin Nombre viruses (genus *Hantavirus*) and the Crimean-Congo hemorrhagic fever virus (genus *Nairovirus*). Viruses of the *Tospovirus* genus infect plants [1]. *Bunyaviridae* have either arthropods- or rodent-borne vectors and are amplified by vertebrate hosts. The Rift Valley fever virus (RVFV), a *Phlebovirus* within the *Bunyaviridae* family, is transmitted by *Aedes* and *Culex* mosquitoes and is a medically and agriculturally important cause of epizootics in Africa. Although this virus primarily affects livestock, humans can be infected as well, and infections can lead to several syndromes ranging from a febrile illness to blindness, encephalitis and lethal hemorrhagic fever. The virus is currently found in the sub-Saharan area, as well as in Egypt, Yemen, Saudi-Arabia, Mayotte and Madagascar [2]. The continuing geographical expansion of RVFV draws concern for Europe, where the virus is considered to be an emerging threat [3,4]. Current vaccines to prevent RVFV epizootics are only partially attenuated, expensive and only induce short-lived immunity [5]. No specific drugs are available to cure an infection, and preventive efforts to avoid new outbreaks are mostly based on weather monitoring [6].

The genome of RVFV consists of three single-stranded RNA segments of either negative or ambisense polarity designated as L (6,404 nucleotides [nt]), M (3,885 nt), and S (1,690 nt). Within each of these three segments, coding regions are flanked at their 5′ and 3′ termini by non-translated regions that comprise two stretches of complementary nucleotides, leading to the formation of RNA panhandle structures [7]. The L and M segments are of negative polarity, while S has ambisense polarity, encoding the nucleoprotein (N) in antisense and the non-structural protein NSs in sense orientation. The L segment expresses a multifunctional

Author Summary

The Rift Valley fever virus (RVFV), a negative strand RNA virus spread by infected mosquitoes, affects livestock and humans who can develop a severe disease. We studied the structure of its nucleoprotein (N), which forms a filamentous coat that protects the viral RNA genome and is also required for RNA replication and transcription by the polymerase of the virus. We report the structure of the RVFV N protein at 1.6 Å resolution, which reveals hexameric rings with an external diameter of 100 Å that are formed by exchanges of N-terminal arms between the nearest neighbors. Electron microscopy of recombinant protein in complex with RNA shows that N also forms rings in solution. A reconstruction of the hexameric ring at 25 Å resolution is consistent with the hexamer structure determined by crystallography. We propose that slight structural variations would suffice to convert a ring-shaped oligomer into subunits with a super-helical arrangement and that this mode of protein-protein association forms the basis for the formation of filamentous ribonucleocapsids by this virus family. Both the RNA binding cleft and the multimerization site of the N protein can be targeted for the development of drugs against RVFV.

protein that comprises an N-terminal endonuclease [8] and a large RNA-dependent RNA polymerase domain [1]. The M segment codes for glycoproteins G_N and G_C that are inserted in the virus lipid envelope and are responsible for cell tropism and membrane fusion. The endodomain of G_N interacts with N, and this interaction is critical for genome packaging into infectious virus particles.

As in other negative-stranded viruses, the genomic RNA (vRNA) in RVFV is packaged with two virally expressed proteins, N and L, into a ribonucleoprotein (RNP) complex that is competent for (+)RNA synthesis and transcription. Contrary to RNPs of Mononegavirales, RVFV N does not assemble into a tube-like structure [9–14] but rather forms a flexible serpentine-like structure [15]. The precise organization of RNA, N and L in this macrostructure is unknown. In addition to its critical role in protecting the vRNA and the antigenome (cRNA), the N protein also plays an active role in RNA transcription and replication [1], as well as in virion assembly [16]. Biochemical studies have shown that RVFV N forms dimers through aromatic residues located in the N terminus of the protein [17]. Recently, a crystal structure was reported for the RVFV N protein [15], which revealed the basic fold of the protein, but raised a number of questions. For example, the crystal structure provided little insight into the mechanism of N multimerization into an RNP complex, and it was unclear how to relate the crystal structure to EM images of N polymers. Furthermore, the RNA binding site identified in the crystal structure of RVFV N differed from that seen in other viral N proteins. Here, we present the crystal structure of RVFV N forming a hexameric ring. The structure reveals the likely binding site for vRNA, and comparison with the previous crystal structure of RVFV N allows us to speculate on the mechanism that underlies the multimerization of N and its encapsidation of viral RNA.

Results

Recombinant N protein forms oligomers and can bind RNA

To produce sufficient amounts of protein for structural studies in the absence of other viral proteins, we expressed RVFV N in *E. coli* with an N-terminal cleavable thioredoxin tag and purified it under non-denaturing conditions to preserve its structural integrity. The final gel filtration column showed two peaks, denoted as N_1 and N_2 (**Figure 1A**). SDS-PAGE analysis revealed that both peaks contained a protein of the size expected for N (27 kDa), suggesting that N was the only protein present and ruling out protein contaminants that could have influenced the oligomeric state of N (**Figure 1A**, inset). The position of peak N_1 corresponds to a protein species with an apparent molecular mass of 300 kDa, suggesting that N formed higher-order oligomers. The position of peak N_2 corresponds to a protein species with an apparent molecular mass of 94 kDa and would thus suggest the presence of smaller oligomers. This notion was confirmed by cross-linking experiments that indicated the presence of dimers, trimers and tetramers in fraction N_2 (**Figure 1B**). We also measured the OD_{260nm}/OD_{280nm} ratio for the two peak fractions to test for the presence of bound nucleic acids. Peak N_1 had an OD_{260nm}/OD_{280nm} ratio of 1.19, clearly indicating that the higher-order N oligomers co-eluted with nucleic acids [18], presumably RNA from the expression host. In contrast, the OD_{260nm}/OD_{280nm} ratio of the N_2 peak was 0.72, showing that this fraction contained much less RNA than fraction N_1 [18]. The variability in the oligomeric state of N expressed in *E. coli* is consistent with previous studies that used either N purified from infected cells or recombinant N expressed in insects cells, although in the latter case multimers with a higher MW were observed [16,17].

We next used surface plasmon resonance experiments to test whether the recombinant protein in fraction N_2 retained its capacity for non-specific RNA binding. We measured the interaction of N with a 20-nucleotides-long RNA, and determined that the Kd of N for RNA is 3.8 μM (**Figure S1**). This result demonstrated that the recombinant N protein present in peak N_2 can still bind RNA and therefore presumably has the native fold. We therefore used this protein for 3D crystallization and structure determination.

Structure determination, crystal packing and structure of the RVFV N protein

Using the N_2 fraction, crystals were obtained in the P6 space group with unit cell parameters of a = b = 180.9 Å and c = 47.4 Å. The selenomethionyl protein crystallized in the same space group with similar unit cell parameters, a = b = 175.5 Å and c = 47.4 Å (**Table 1**). The structure was determined using the SAD technique with data recorded at the Se absorption edge from crystals of the selenomethionyl protein that diffracted to 2.3 Å

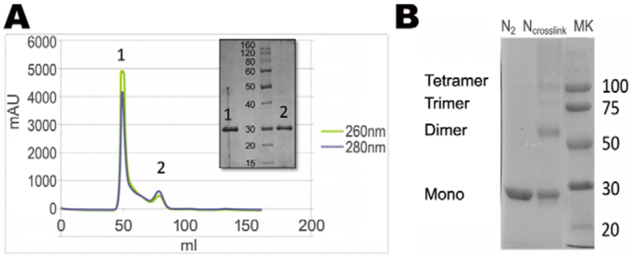


Figure 1. Oligomeric state of recombinant N protein. **(A)** Chromatogram of N run on an S200 size exclusion column. The grey line shows the absorbance at 280 nm, and the green line the absorbance at 260 nm. The inset shows a 12.5% SDS-PAGE gel of peaks 1 (N_1) and 2 (N_2). **(B)** A 12.5% SDS-PAGE gel showing the presence of dimers, trimers and tetramers after cross-linking fraction N_2 with 0.05% glutaraldehyde for one hour at room temperature.
doi:10.1371/journal.ppat.1002030.g001

Table 1. Data collection and refinement statistics.

Data collection and refinement statistics of RVFV nucleoprotein	Se-MET[a]	Native[b]
Instrument	ESRF - ID14–4	ESRF - ID14–4
Wavelength	0.9790	0,9770
Space group	P6	P6
Cell dimensions *a*, *b*, *c* (Å)/α, β, γ (°)	175.5, 175.5, 47.4/90, 90, 120	180.9, 180.9, 47.7/90, 90, 120
Resolution range (Å)	47.44– 2.30 (2.42–2.30)*[&]	47.1–1.6 (1.69 – 1.6)*
Total number of reflections	524146 (73999)*	575561 (39900)*
Number of unique reflections	37702 (5378)*	111024 (13148)*
Completeness (%)	99.6 (97.6)*	93.8 (76.5)*
I/Σ(*I*)	27.5 (15.6)*[&]	10.9 (2.0)*
R_{sym} **	0.072 (0.130)*[&]	0.084 (0.510)*
Multiplicity	13.9 (13.8)*	5.2 (3.0)*
Refinement		
R***	0.2015	0.2221
R_{free} *****	0,2505	0.2536
No. atoms		
Protein	5781	5901
Water	347	965
B-Factors		
Protein	14.14	19.16
Water	19.14	33.2
R.m.s. deviations		
Bonds lengths (Å)	0,0078	0.0061
Bonds angles (°)	1.023	0.883

[a] PDB code: 3OUO

[b] PDB code: 3OU9

[&] A resolution cutoff of 2.3 Å was chosen to retain only the most reliable data for phasing purposes.

*Values in parentheses give the values in the highest resolution shell (1.69–1.6 Å).

**R_{sym} = Σ |I-<I>|/Σ I

***R = Σ||Fo|-|Fc||/Σ |Fo|

*****Rfree = R factor calculated using 5% of reflections not included in refinement.

doi:10.1371/journal.ppat.1002030.t001

resolution. The structure was subsequently refined using a native data set that extended to 1.6 Å resolution (**Table 1**).

The asymmetric unit contains three N molecules, labeled α, β and γ in **Figure 2A**, that form two distinct hexameric rings in the crystal, labeled I and II in **Figure 2B**. Hexamer I is formed by six copies of subunit α that surround the crystallographic 6-fold axis, whereas hexamer II is formed by three β,γ dimers that surround the crystallographic 3-fold axis (**Figure 2B**). The two sets of hexamers, which face in opposite directions and are offset by 10 Å in the direction of the crystallographic *c* axis (**Figure 2C**), form layers along the [*a*,*b*] plane of the crystal. Stacking of the layers in the crystal results in the formation of two sets of tubes, one set formed by hexamers I and the other by hexamers II, that both run along the crystallographic *c* axis but in opposite directions (**Figure S2**).

The crystal structure reveals that the N monomer consists of an orthogonal bundle of thirteen α-helices (**Figure 3**). The structure can be divided into three domains. Residues 1–32 form a flexible N-terminal arm containing two α-helical segments that extends away from the globular core of the protein. The globular core itself consists of two domains, one formed by six helices spread over residues 36–90, 110–122, 211–220 and the other one formed by four helices spread over residues 103–110 and 130–204 (**Figure 3**). The core domain in our structure is virtually identical to that in the previously reported crystal structure of N [15] with an rms deviation between the backbone atoms of the two structures of~0.7 Å (**Figure S3B**). The position and conformation of the N-terminal arm, however, are very different in the two structures (**Figure S3**), a finding that will be discussed below. The fold of RVFV N is currently unique in the PDB, but considering the high level of conservation in their amino-acid sequences (average identity>30%) (**Figure S4**), other *Phlebovirus* N proteins are likely to adopt a similar fold.

Multimerization of the N protein

In our crystals, N forms ring-shaped hexamers (the subunits are denoted A to F as shown in **Figure S2**) that have a thickness of 45 Å, an external diameter of approximately 100 Å, and a central funnel-like aperture with a diameter that narrows from 50 to 30 Å (**Figure 2C**). Multimerization appears to be driven by the extended N-terminal arm, which wraps around the external surface of the globular core of the adjacent subunit, fitting snuggly into a hydrophobic groove and burying a surface of 1456 Å^2 (**Figure 4A**). In particular, the aromatic rings of residues Y3, F11, W24, F28 and Y30 and the aliphatic side chains of residues L7, V9, V16, I21 and V25 project from the N-terminal arm and fill up the hydrophobic groove formed by regions 36–82, 108–126, and 207–210 of the core domain of the adjacent molecule (**Figure 4B/C**). This arm-core interaction is repeated in a directional manner, such that the arm of subunit A extends into the hydrophobic groove of subunit B, B into C, C into D, D into E, E into F and F into A, creating the hexameric rings seen in the crystal. This mode of multimerization is consistent with mutagenesis data that mapped the interacting domain of the *Phlebovirus* N protein to its N-terminal arm [17].

Hexamers I and II in the crystals of the native protein superimpose very well, but the two rings in the crystals of the selenomethionyl protein are slightly different (**Figure S5A**). While hexamer II formed by seleniated N is the same as the two hexamers formed by native N (**Figure S5B**), the subunits in hexamer I are more closely packed about the 6-fold symmetry axis. The domain of N that is near the center of the ring, comprising the loop connecting helices α10 and α11, occludes part of the central aperture, suggesting a twist in the assembly of the ring subunits (**Figure S5A**). Superimposition of native and seleniated hexamers I based on subunits A, creates an 11° deviation between the planes of the two rings (**Figure S5C**). Furthermore, comparison of the subunits in hexamers I formed by native and seleniated protein shows that the contraction of the ring is due to a lateral slippage between adjacent subunits (**Figure S6**). As a result of the slightly different subunit organization, the asymmetric unit is shorter in crystals of the seleniated protein and the length of the crystallographic *a* and *b* axes is decreased by about 5 Å (**Table 1**). The existence of two types of rings in the crystals demonstrates the natural ability of N to form oligomers with different subunit organizations, providing a basis for the formation of serpentine-like RNP structures.

RNA binding

The core of the N protein has a concave crescent shape and the relative orientation of its two domains is reminiscent of a head of pliers, suggestive of a role in grabbing genomic RNA (**Figure 3C**).

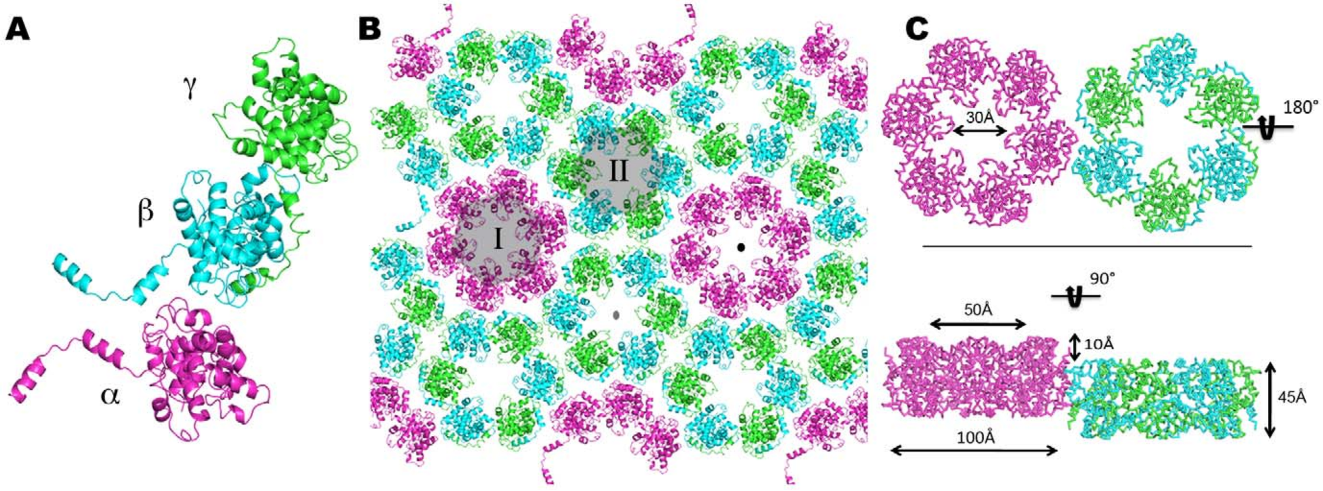


Figure 2. Hexamers formed by N in the crystal. (A) The three N monomers in the asymmetric unit are shown in pink, cyan and green and labeled α, β and γ, respectively. **(B)** Subunit packing in the crystal layer corresponding to the crystallographic [*a*,*b*] plane. The six copies of subunit α that surround the crystallographic 6-fold symmetry axis (black dot) form one hexamer (labeled I), and the three βγ dimers that surround the crystallographic 3-fold symmetry axis (grey dot) form a second hexamer (labeled II). **(C)** The dimensions of the hexameric ring are labeled. Hexamers I and II face in opposite directions and are offset by approximately 10 Å in the direction of the *c* axis.
doi:10.1371/journal.ppat.1002030.g002

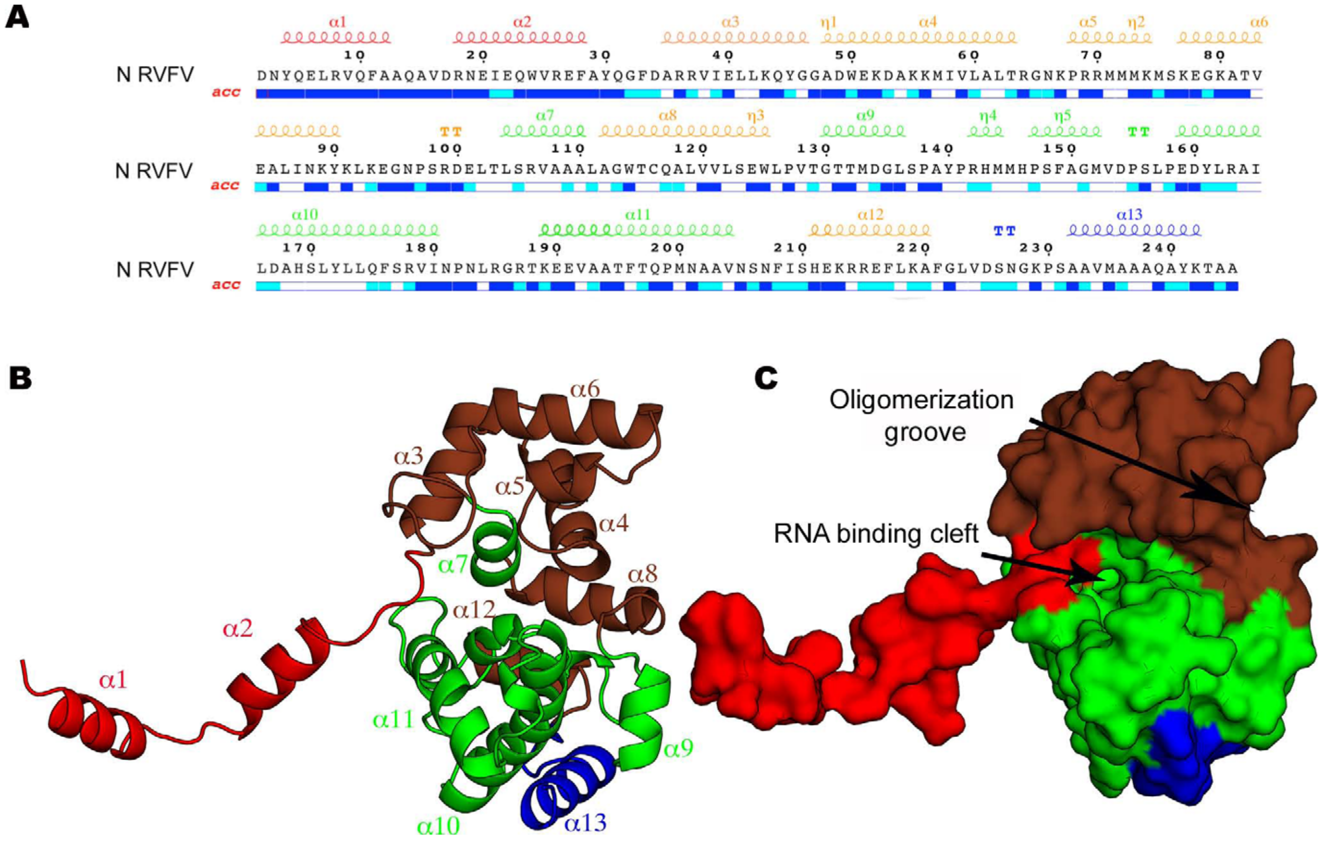

A
α1
α2
α3
η1
α4
α5
η2
α6
10
20
30
40
50
60
70
80
N RVFV
DNYQELRVQFAAQAVDRNEIEQWVREFAYQGFDARRVIELLKQYGGADWEKDAKKMIVLALTRGNKPRRMMMKMSKEGKATV
acc
TT
α7
α8
η3
α9
η4
η5
TT
90
100
110
120
130
140
150
160
N RVFV
EALINKYKLKEGNPSRDELTLSRVAAALAGWTCQALVVLSEWLPVTGTTMDGLSPAYPRHMMHPSFAGMVDPSLPEDYLRAI
acc
α10
α11
α12
TT
α13
170
180
190
200
210
220
230
240
N RVFV
LDAHSLYLLQFSRVINPNLRGRTKEEVAATFTQPMNAAVNSNFISHEKRREFLKAFGLVDSNGKPSAAVMAAAQAYKTAA
acc
B
α6
α3
α5
α4
α7
α8
α12
α2
α1
α11
α9
α10
α13
C
Oligomerization groove
RNA binding cleft

Figure 3. Structure of the N monomer. (A) Sequence of the RVFV N protein with secondary structure elements indicated above. The colors correspond to those used to color the different sub-domains in the crystal structure of N shown in panels B and C. **(B)** Ribbon representation of the crystal structure of the RVFV N protein showing that the N terminus forms an arm (red) that extends from the globular core domain (brown and green). The C terminus, which is not involved in RNA binding, is shown in blue. The α-helices are labeled. **(C)** View of the RVFV N protein in surface representation. The orientation and color code are the same as in panel B.
doi:10.1371/journal.ppat.1002030.g003

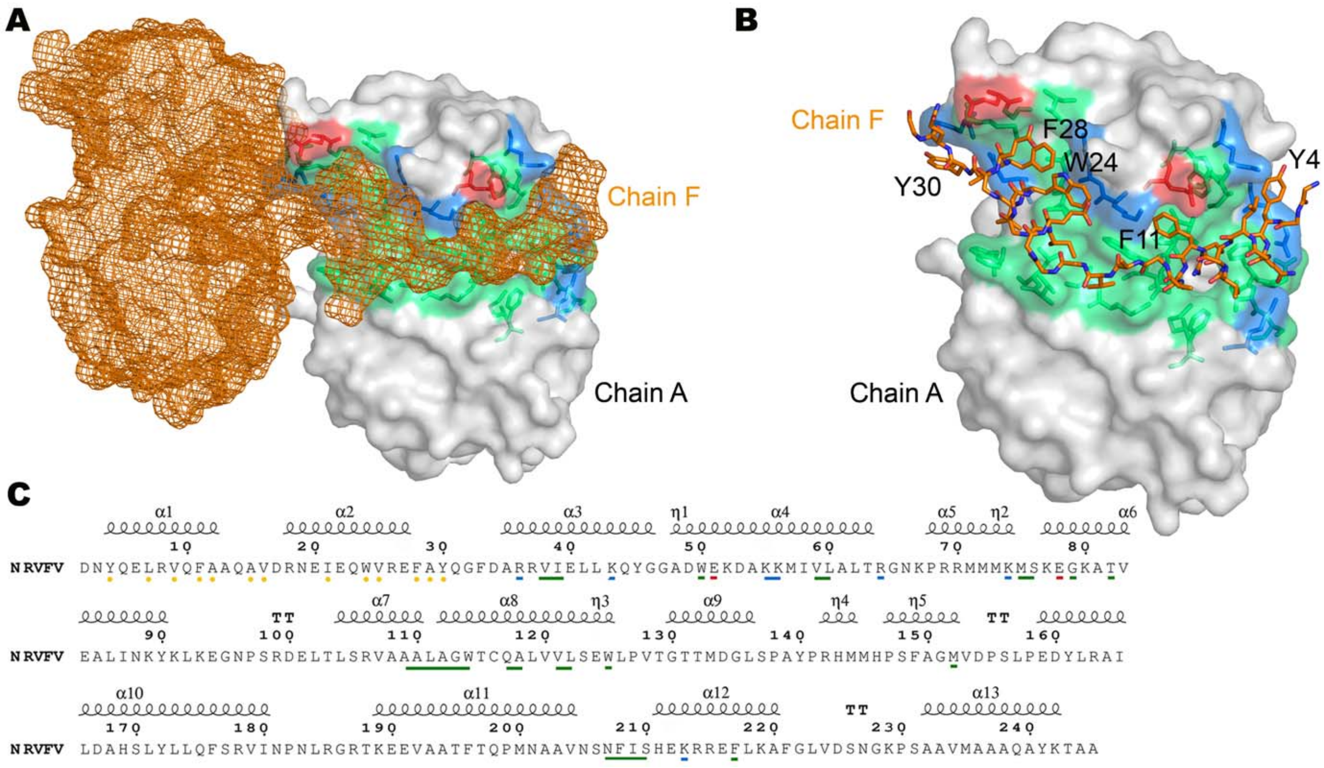

A
B
Chain F
Chain A
Chain F
F28
W24
Y4
Y30
F11
Chain A
C
α1 α2 α3 η1 α4 α5 η2 α6
10 20 30 40 50 60 70 80
NRVFV DNYQELRVQFAAQAVDRNEIEQWVREFAYQGFDARRVIELLKQYGGADWEKDAKKMIVLALTRGNKPRRMMMKMSKEGKATV
α7 α8 η3 α9 η4 η5 α10
TT TT
90 100 110 120 130 140 150 160
NRVFV EALINKYKLKEGNPSRDELTLSRVAAALAGWTCQALVVLSEWLPVTGTTMDGLSPAYPRHMMHPSFAGMVDPSLPEDYLRAI
α10 α11 α12 α13
TT
170 180 190 200 210 220 230 240
NRVFV LDAHSLYLLQFSRVINPNLRGRTKEEVAATFTQPMNAAVNSNFISHEKRREFLKAFGLVDSNGKPSAAVMAAAQAYKTAA

Figure 4. Interaction between adjacent N subunits in the hexamer. (A) Overview of how the N-terminal arm of one subunit, shown in orange mesh, fits into the hydrophobic pocket in the surface of the adjacent subunit, shown as grey surface with the hydrophobic pocket highlighted in green. **(B)** Magnified view of the interaction. The N-terminal arm is shown in ball-and-stick representation and the surface of the hydrophobic pocket is shown transparent to reveal details of the hydrophobic interactions. (**C**) Amino acid sequence of an RVFV N polypeptide, showing above the secondary structure elements derived from the crystal structure. Below the sequence, residues are labeled that are involved in inter-subunit interactions in the hexamer. The yellow dots indicate residues of the arm that interact with residues in the oligomerization groove. The colored bars indicate the character of the residues in the oligomerization groove: green, hydrophobic; blue, positively charged; red, negatively charged.

doi:10.1371/journal.ppat.1002030.g004

This cleft is sandwiched between three helices on one side (α4, α5, α7) and two 3_{10}-helices (η4, η5) followed by three α-helices on the other (α9, α10, α11), a fold in accordance with the "(5H+3H)" structural motif for RNA binding [19]. Furthermore, analysis of the electrostatic surface potential reveals a positively charged patch located within the inner part on one side of the hexamers (**Figure 5**). This patch includes residues R64, K67 and K74 that are evolutionary conserved across *Phleboviruses* (**Figure S4**). To test whether this positively charged patch indeed constitutes the RNA binding site, we expressed and purified a triple RVFV N mutant (R64D, K67D, K74D). The triple mutant eluted from the gel filtration column as a single peak, corresponding to N_2 (**Figure S7A**). SDS-PAGE analysis revealed that the peak fractions contained a protein of the size expected for N (27 kDa) (**Figure S7A,** inset), and mass spectrometry confirmed the protein to be RVFV N. The OD_{260nm}/OD_{280nm} ratio of the peak fraction was 0.52, indicating that this fraction contained only protein [20]. Binding studies using surface plasmon resonance spectroscopy with a 20-nucleotides-long RNA showed that the triple mutant lost its ability to bind RNA, supporting the notion that the positively charged patch serves as the RNA binding cleft (**Figure S7B**).

Taking as a guide the structure of the rabies virus N protein bound to single-stranded RNA (PDB code: 2GTT [11]), we could position an RNA molecule in the concave surface between the two core domains of the RVFV N protein, such that the RNA sugar phosphate backbone interacts with the positive charges in the basic cleft. The model of RVFV N protein bound to RNA further showed that each N subunit can accommodate approximately six RNA bases (**Figure S8**).

Electron microscopy of N-RNA complexes

The crystal structure showed that RVFV N forms hexameric rings. To assess whether N also forms hexamers in solution, we prepared negatively stained samples for analysis by electron microscopy (EM). EM images of fraction N_2, which contained only protein and was used for 3D crystallization, did not show any ring-shaped complexes (data not shown), consistent with the SEC result that showed that this fraction contains only small oligomers. By contrast, EM images of fraction N_1, which contained both protein and RNA, revealed distinct circular structures with diameters ranging from~70 to 100 Å, which were stable over a period of one month (**Figure 6A**). The images thus suggest that the formation of stable higher-order N oligomers requires the protein to associate with RNA and that the resulting higher-order oligomers have a ring-shaped structure.

To obtain a better understanding of the structure of N-RNA complexes, we calculated 3D reconstructions of the ring-shaped complexes seen in fraction N_1. The small size of the complexes prevented us from using vitrified specimens for EM imaging, and we therefore prepared samples by cryo-negative staining. This specimen preparation method provides the high contrast of stain but minimizes the artifacts associated with conventional negative staining [21]. The N-RNA complexes adsorbed to the carbon support film preferentially with the flat side of the ring, making it necessary to use the random conical tilt approach to calculate 3D

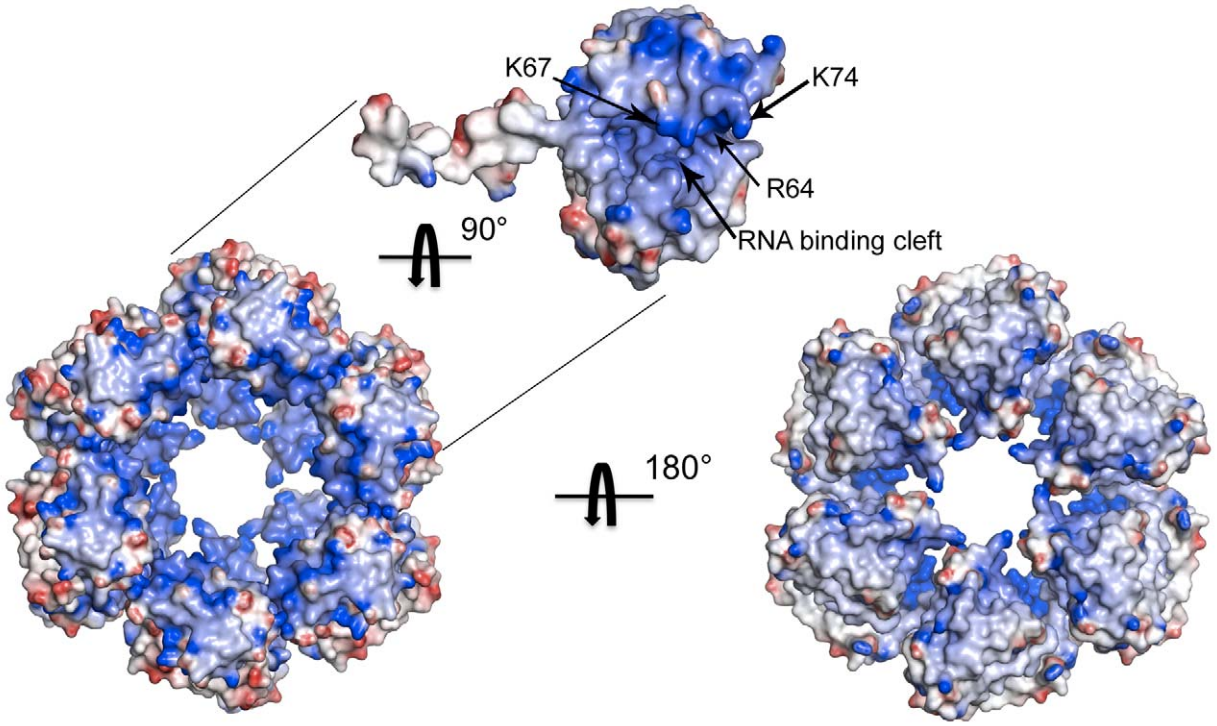


Figure 5. Electrostatic surface potential of the N hexamer. Mapping of the electrostatic surface potential, from −10 kT in red to+10 kT in blue, onto the surface of hexamer I formed by native N protein reveals a patch of positive charges in the inner part of the ring, which likely accommodates the vRNA. Key residues in the RNA binding site are labeled on the electrostatic surface of a single monomer.
doi:10.1371/journal.ppat.1002030.g005

reconstructions [22]. We recorded a total of 30 image pairs at tilt angles of 50° and 0°, from which we selected 10,764 particle pairs. The particles from the images of the untilted specimen were classified into 100 classes, which revealed a variety of oligomers, ranging from tetramers to octamers (**Figures 6B and S9**). About 57% of all the particles were ring-shaped oligomers. The hexamer was the most abundant species with 24%, followed by the pentamer (22%), the heptamer (7%), and finally the octamer (4%). The averages of the various oligomers revealed a large variability in the ring shape, pointing to structural flexibility in the various N-RNA complexes. Because the hexamer was most prevalent and because N alone formed hexamers in the 3D crystals, we focused on calculating a 3D reconstruction of the hexameric N-RNA complex. We combined the particles from classes that produced the most similar averages (399 particles from 2 classes) and calculated a 3D density map using the particles selected from the images of the tilted specimen and the best 10% of particles selected from the untilted specimen. According to the Fourier shell correlation (FSC) = 0.5 criterion, the final density map had a resolution of 25 Å (**Figure S10**).

With a diameter of about 100 Å and a thickness of about 45 Å (**Figure 6C**), the EM density map of the N-RNA complex has virtually identical dimensions as the crystal structures of the N hexamer. Accordingly, the EM density map nicely accommodated the crystal structure of the hexamer, illustrating that the hexamers, and by extension also the other ring-shaped oligomers, are compatible with RNA binding (**Figure 6D**).

Discussion

The N protein is the most abundant viral protein in the *Phlebovirus* virion and plays a key role in encasing vRNA in a protective coat. We have determined the crystal structure of RVFV N in a hexameric form, which shows largely the same fold that was previously seen in a crystal structure of monomeric N [15]. The two structures differ, however, in the position of the N-terminal arm. In the previous structure, the N-terminal arm packs closely against the core domain, while it extends away from it in our structure (**Figure S3**). Extension of the N-terminal arm is crucial for the oligomerization of N, as it mediates the interaction with the adjacent subunit in the crystallographic hexamer. We believe that the RVFV N hexamer is biologically relevant, because oligomers have been observed for many other N proteins [9,10,11,12,13,14] and EM revealed that the RVFV N-RNA complex also forms ring-shaped oligomers in solution (**Figure 6**).

The different position of the N-terminal arm in our and the previous structure is intriguing as it may reflect the structural change that has to occur for N to multimerize, thus potentially providing a clue to the mechanism underlying the formation of a ribonucleocapsid. In the hexamer, the N-terminal arm lies in a hydrophobic pocket of the adjacent subunit, thus mediating an inter-molecular interaction (**Figure 4**). By contrast, in monomeric N, the N-terminal arm makes an intra-molecular interaction and binds to the same hydrophobic pocket but in its own core domain, burying a surface area of 1179 Å^2 (**Figure S11**). The inter- and intra-molecular interactions with the N-terminal arm are mediated largely by the same residues of the core domain (**Figure 4C** and **Figure S11B**). Interestingly, the intra-molecular interaction of the N-terminal arm not only fills the hydrophobic pocket of its own core domain, thus preventing oligomerization, but also covers the RNA binding cleft, so that N in this conformation is incapable of binding RNA. For a monomer, the closed conformation is presumably more favorable, because it reduces the hydrophobic surfaces on both the N-terminal arm and oligomerization groove.

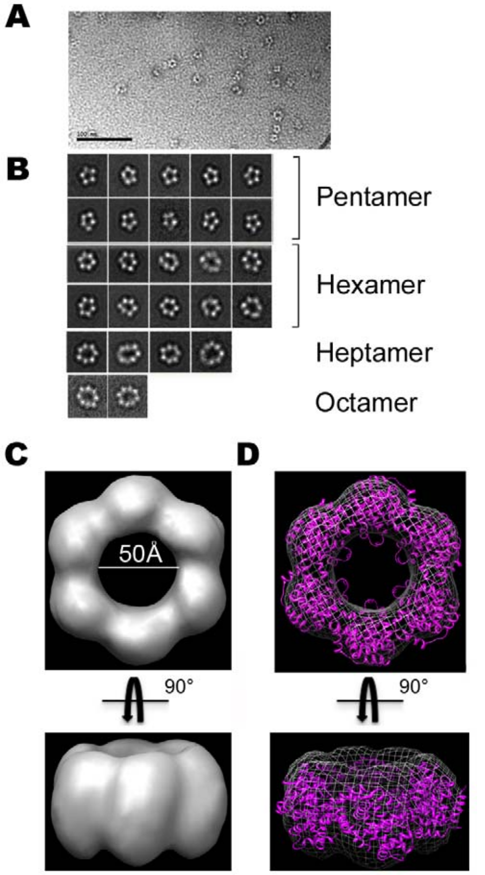


Figure 6. Electron microscopy of N-RNA complexes. **(A)** Representative electron micrograph of the N_1 fraction in negative stain, revealing ring-shaped particles of different sizes. Scale bar is 100 nm. **(B)** Representative class averages of N-RNA rings. **(C)** 3D reconstruction of a hexameric N-RNA complex obtained with cryo-negatively stained samples. **(D)** Docking of the crystal structure of hexamer I formed by native N into the EM density map of the hexameric N-RNA complex.
doi:10.1371/journal.ppat.1002030.g006

In case the closed conformation is a "waiting" conformation before oligomerization; residues involved in the molecular interaction would have to compete for the oligomerization groove and expose the hydrophobic side of the arm (**Video S1**).

Our SEC analysis shows that peak N_2, which lacks RNA, contains only small oligomers, suggesting that the inter-molecular interactions mediated by the N-terminal arm are not very strong on their own, potentially because the intra-molecular interactions outcompete the inter-molecular interactions, and thus do not support large oligomer formation. The weak interactions between N proteins would allow easy addition and removal of subunits. The fact that we see hexamers in our crystals may be explained by the high protein concentration used for crystallization trials that drives the small units of nucleoproteins to assemble into larger stable oligomers. In solution, however, stabilization of the oligomers may require the additional association of the subunits with RNA. Binding to RNA would align N proteins to each other and increase their local concentration, thus stabilizing the inter-molecular interactions of the N-terminal arms and resulting in the stable, ring-shaped oligomers seen in SEC peak N1 (**Figure 6**). This model of RNA-stabilized oligomers provides an elegant molecular explanation for why N proteins can have an inherent tendency to multimerize without forming undesired, large oligomers in the absence of RNA.

With only six subunits and a diameter of 100 Å, the RVFV N ring is the smallest one among the ring-shaped oligomers seen in crystal structures of N proteins from negative strand viruses (**Figure 7**). Although there are clearly common structural features in the oligomers, the mode of how the subunits interact with each other varies. In rabies virus (RV), vesicular stomatitis virus (VSV), respiratory syncytial virus (RSV) and influenza virus, extensions at both the N and C termini of the polypeptide are involved in organizing adjacent subunits into an ordered assembly [11,12,14]. By contrast, it is only the interaction of the N-terminal arm of RVFV N with the hydrophobic pocket of the neighboring subunit that mediates the contacts between adjacent subunits in oligomers. While this interaction appears sufficient to promote efficient protein polymerization, it leaves a significant degree of freedom at the level of lateral interactions. This plasticity is illustrated by the slightly different positions of the N-terminal arm on the core domain of the neighboring subunit seen in hexamers I and II in the crystals of the native and seleniated proteins (**Figure S5**). As a result, like the N proteins of RV and RSV, RVFV N can form rings with deformed shapes and a variable number of subunits (**Figure 6B**) and, although not yet observed, N may even have the capacity to form oligomers with a superhelical arrangement of the subunits.

Although EM of RVFV N-RNA complexes also showed ring-shaped oligomers (**Figure 6**), it is not clear whether rings are the building block of the native ribonucleocapsid. The RNPs of several Mononegavirales have a superhelical subunit arrangement, including those of RV [11], VSV [12], RSV [14], measles [9,10] and mumps [13]. However, the RNPs of *Phleboviruses* do not assemble into a highly ordered structure, but rather into flexible filamentous assemblies [15,23,24,25,26]. In particular, EM images of RNPs from RVFV [15] and other *Bunyaviridae* [23,24,25] display an extended filament-like structure, but they do not rule out some degree of symmetry in the way the vRNA is packaged. While it thus remains uncertain whether the RVFV ribonucleocapsid is formed by stacked rings or a superhelical oligomer or even a mixture thereof, the flexibility in the interaction between adjacent subunits would allow great variability in the architecture of the ribonucleocapsid. Flexibility in the contacts between N subunits allows the assembly to readily adapt to distortions introduced by external constraints or signals within the infected cells, while maintaining the connectivity of the RNA.

The crystal structure of the hexamer reveals that several positively charged residues are clustered in a cleft that can accommodate a single molecule of RNA (**Figures 5 and S8**). This finding is in agreement with the proposal of Luo and collaborators that N RNA binding site is formed by two domains that contain a "(5H+3H)" structural motif [19]. Genomic RNA would thus run like a belt inside the ring and be completely concealed from the innate immune system of the host cell, in a manner similar to the ribonucleocapsid of the rabies virus [11]. Each N subunit can accommodate up to six bases, so that one turn of the RNA inside the hexameric ring would translate to~36 bases. Our single-particle EM reconstruction of the hexameric RVFV N-RNA complex is consistent with the crystallographic hexamer of N, but it does not show the RNA inside the ring (**Figure 6D**). Considering the limited resolution of the EM density map of 25 Å, the small mass of 30 RNA bases and the negative charge of RNA, which would favor positive staining of the RNA, it is not surprising that the RNA is not visible in the EM map. By considering that the thickness of the hexameric N-RNA complex is about 45 Å (**Figure 6C**) and making the simplifying assumption

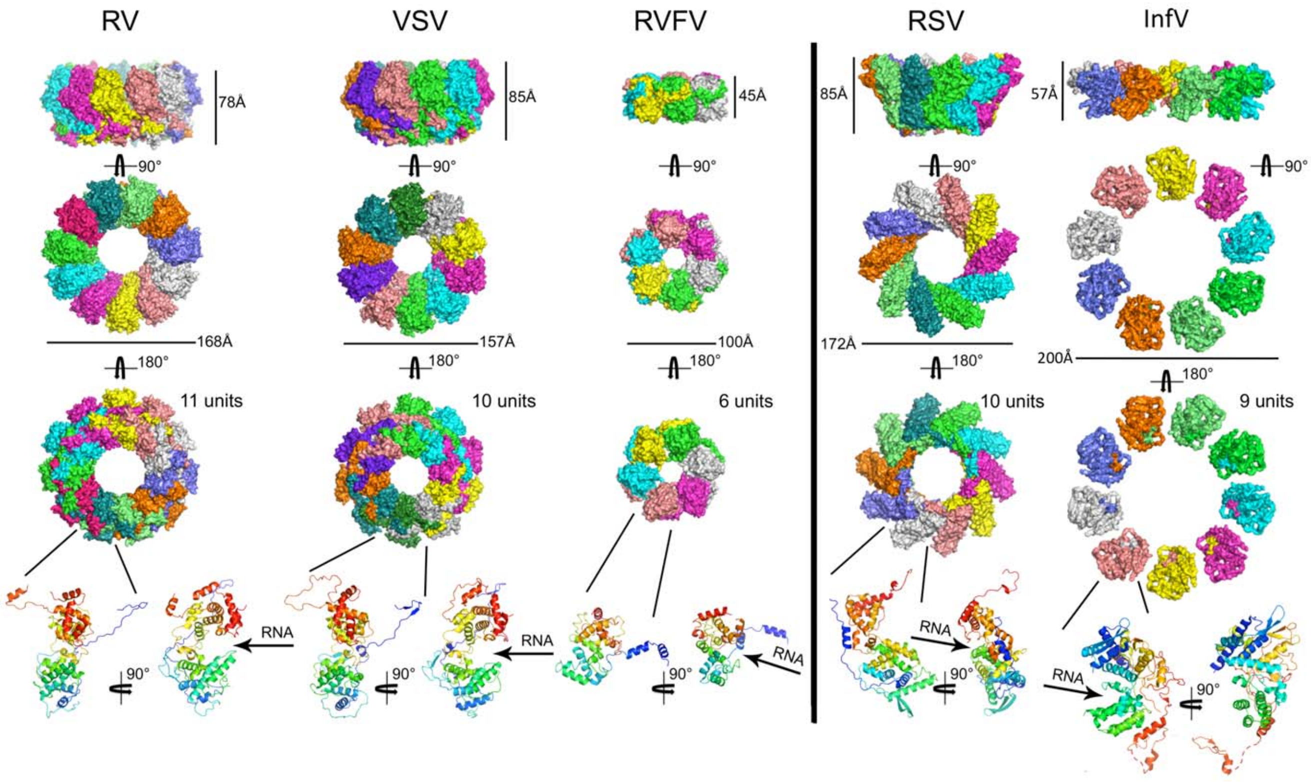


Figure 7. Gallery of crystal structures of N proteins from different negative strand RNA viruses. The structures shown are those of the N proteins from the rabies virus (RV; PDB code: 2GTT), the vesicular stomatitis virus (VSV; PDB code: 2GIC), the Rift Valley fever virus (RVFV; this work), the respiratory syncytial virus (RSV; PDB code: 2WJ8), and the influenza virus (InfV; PDB code: 2IQH). The model fro the influenza NP ring is derived from an EM reconstruction into which the crystal structure of the monomer was modeled (PDB code: 2WFS). The top three rows show different views of the ring structures in surface representation, in which each subunit is shown in a different color. The fourth row shows two views of the monomers in ribbon representation. The arrows indicate the cavity that binds the vRNA.
doi:10.1371/journal.ppat.1002030.g007

that the entire nucleocapsid is formed by stacked hexamers, the ribonucleocapsid of the S segment of 1690 nt would span a total linear distance of about 0.25 µm. This value is consistent with the size of 0.27 µm of the ribonucleocapsid of Uukuniemi virus seen in EM images [27], whose genome is only slightly larger than that of RVFV.

Earlier studies showed that transcription and replication require not only the polymerase L but also the N protein [28], implying that naked vRNA cannot be transcribed [1]. While it has been established that the two proteins are positioned in close proximity to each other, as L is recruited to the vRNA through a panhandle structure [29] and N through a short region in the 5′ region of the ORF [30], how the two proteins interact with each other remains unclear. A recent study found a conserved region in the second domain of N consisting of helices α4, α5, α6 [31]. This domain may well play a role in the stacking of N subunits in the oligomer, but it could also mediate a transient interaction with L and promote a temporary release of N, thus liberating the RNA to become accessible for transcription by L. Additionally, helices α1, α12 and α13 are located at the periphery of the hexameric ring, and residues projecting from these helices are also likely to form a significant part of the L-binding surface on the ribonucleocapsid. Given the substantial contribution of the N-terminal arm to the buried interface (1456 Å^2 out of a total of 1640 Å^2 buried surface between two adjacent N subunits), it is conceivable that interactions of α1 of the N-terminal arm with L could lead to a local unwinding of the filament structure and the exposure of vRNA while avoiding complete disassembly of the ribonucleocapsid.

In conclusion, the structure of the hexamer formed by the RVFV N protein presented here shows that oligomerization is mediated by a flexible N-terminal arm, which binds a hydrophobic pocket in the adjacent subunit. The different hexamers seen in our crystals and the variability in the oligomeric state of N-RNA complexes seen in EM images demonstrate substantial flexibility in the interaction between subunits. Furthermore, comparison with a previous structure of the RVFV N protein suggests an elegant mechanism that allows the formation of stable N oligomers only in the presence of RNA. Finally, the nucleoprotein structure identifies potential sites that could be targeted for drug development. For instance, compounds blocking ribonucleocapsid assembly either by interfering with RNA binding or by trapping the N-terminal arm of N in a conformation that is not compatible with oligomerization, could serve as starting points to design specific antiviral molecules.

Materials and Methods

Cloning, mutagenesis, protein production and purification

cDNA corresponding to the RVFV N protein (strain Smithburn DQ380157.1) was cloned by recombination (Gateway, Invitrogen) into the pETG20A vector (kindly provided by Dr. Arie Geerlof), which adds a cleavable N-terminal thioredoxin-hexahistidine tag, and used to transform *E. coli* strain C41 (Avidis) carrying the pRARE plasmid (Novagen). Bacteria were grown in TB medium (Athena Enzyme) at 37°C to an OD$_{600nm}$ of 0.5. Expression was induced with 0.5 mM IPTG, and bacteria were grown overnight at 17°C. Cells were pelleted, resuspended in 30 ml of lysis buffer (50 mM Tris, pH 8, 300 mM NaCl, 5 mM imidazole, 5% glycerol, 0.1% Triton X-100, 2 mM EDTA), frozen, and stored at -80°C.

N was purified at 4°C. A frozen pellet was melted on ice, sonicated, and the lysate was cleared by centrifugation at 20,000 rpm for 30 min. The protein was first purified by metal affinity chromatography using a 5 ml HisPrep column (GE Healthcare). The tag was then removed by cleavage with TEV protease, and the protein was further purified with a second metal affinity column followed by size exclusion chromatography (SEC) using a Superdex 200 column (GE Healthcare) in 10 mM HEPES, pH 7.5, 300 mM NaCl.

The R64D/K67D/K74D triple mutant was generated by first simultaneously introducing the R64D and K67D mutations followed by introducing the K74D mutation into the RVFV N cDNA using the QuickChange Site-Directed Mutagenis Kit (Agilent). The sequences of the primers used to introduce the point mutations were:

R64D/K67D forward: CTGGCTCTAACTgaTGGCAACg-AcCCCCGGAGGATG,

R64D/K67D reverse: CATCCTCCGGGGgTcGTTGCCAt-cAGTTAGAGCCAG,

K74D forward: CGGAGGATGATGATGgAcATGTCAAAA-GAAGGC, and

K74D reverse: GCCTTCTTTTGACATgTcCATCATCATC-CTCCG.

The complete coding region of each mutant was sequenced to confirm the desired modification. The triple mutant was expressed and purified analogous to the wild-type protein, and the expressed protein was verified by mass spectrometry.

Analytical SEC was performed on a KW 803 column (Shodex) using a High Pressure Liquid Chromatography Alliance 2695 system (Waters), and absorbance was measured at both 260 nm and 280 nm. The SEC column was calibrated with Kit LMW markers (GE Healthcare). The protein eluted in two peaks, with apparent molecular weights of~300 kDa (N_1) and 94 kDa (N_2). The N_1 peak was used for EM analysis of N-RNA complexes and the N_2 peak for 3D crystallization screens of N.

Characterization of N-RNA interactions by surface plasmon resonance spectroscopy

Binding affinities of wild-type and mutant N protein for ssRNA were determined using a ProteOn XPR36 instrument (Bio-Rad Laboratories, Inc). NeutrAvidin (Thermo Scientific) was amine-coupled to a carboxylated sensor surface (GLM sensor chip) to a final immobilized level of 6000 RU. To test non-specific binding by N, biotin-labeled ssRNA oligonucleotides with a non-relevant sequence from the dengue virus 5′ non-translated region (RNA_{20}-3′biotin: GAGUUGUUAAUCUUUUUUUU-biotin; Sigma) were diluted to 10 nM in sodium acetate (pH 5.5) and injected for two minutes at a flow rate of 25 µl/min. Association and dissociation phases were measured for 240 sec and 600 sec, respectively. Measurements were performed in buffer containing 10 mM HEPES, pH 7.5, 300 mM NaCl, 0.005% NP-20. Data were analyzed in ProteOn Manager version 2.0.

Crystallization

N protein in 10 mM HEPES, pH 7.5, 300 mM NaCl collected from fraction N_2 was concentrated to 7.8 mg/ml, and 2 µl of protein solution was mixed with 2 µl of reservoir solution containing 200 mM $MgNO_3$ and 17% (w/v) PEG 3350 for crystallization at 20°C using the hanging drop method. SDS-PAGE analysis of dissolved crystals confirmed that they contained full-length N protein. The crystals were flash-frozen in liquid nitrogen using 5% glycerol as cryo-protectant.

Data collection, structure determination and refinement

The RVFV N protein crystallizes in space group P6 with unit cell parameters of a = b = 180.9 Å, c = 47.7 Å for the native protein and a = b = 175.5 Å, c = 47.4 Å for the seleniated protein. A native data set extending to 1.6 Å resolution and a Se-Met data set extending to 2.3 Å resolution were collected on beamline ID14–4 at the ESRF (Grenoble, France). The Se-Met data set was collected at the Se absorption edge. Data were processed using the program XDS [32]. Of a total of 33 Se sites for the three monomers in the asymmetric unit, the position of 27 sites were identified using the program SHELXD [33] to analyze anomalous data ranging from 10 to 2.3 Å. After initial phase calculation and modifications with the SHELX suite, a readily interpretable map was obtained with an overall figure of merit of 0.62. The program ARP/wARP [34] was used to generate an initial model, and a complete model for the three independent monomers was built using COOT [35]. Using this model, the native data set was subsequently solved by molecular replacement using the program Phaser [36]. The program REFMAC5 with the TLS option was used for crystallographic refinement [37]. The final models were assessed with PROCHECK [38]. Surface electrostatics were calculated using DELPHI [39]. Sequences were aligned using Muscle [40] and seaview [41]. Intermediate structures for the morphing were generated using LSQman [42]. Figures and movie were generated with the programs ENDscript, ESPript [43] and PyMOL (http://www.pymol.org).

Sample preparation, electron microscopy and image processing

Samples were prepared by negative staining and cryo-negative staining with uranyl formate as described [44]. For specimens prepared by conventional negative staining, images were taken using Philips CM10 electron microscope equipped with a tungsten filament and operated at an acceleration voltage of 100 kV. Images were recorded on a 1 k×1 k Gatan CCD camera at a magnification of 52,000×using a defocus value of –1.5 µm. For cryo-negative staining specimens, images were recorded using a Tecnai F20 electron microscope (FEI), equipped with a field emission gun and operated at an acceleration voltage of 200 kV. Grids of cryo-negatively stained specimens, used to collect image pairs of specimens tilted to 50° and 0°, were loaded on an Oxford cryo-transfer holder and maintained at liquid nitrogen temperature during image acquisition. Images were taken at a magnification of 50,000×, with a defocus value of –2.0 µm for images of untilted specimens and –1.8 µm for specimens tilted to 50°. All images were recorded using low-dose procedures on Kodak film SO163 and developed for 12 min with full-strength Kodak D-19 developer at 20°C.

Electron micrographs were digitized with a SCAI scanner (Zeiss) using a step size of 7 µm, and 3×3 pixels were averaged to obtain a pixel size of 4.2 Å on the specimen level for cryo-negatively stained specimens. 3D reconstructions from the cryo-negatively stained preparations were calculated using the SPIDER software package [45]. 10,764 particle pairs were interactively selected from a total of 30 image pairs using WEB, the display program associated with SPIDER, and windowed into small images of 60×60 pixels. The particles from the images of the untilted specimens were classified over 10 cycles of *K* means classification and multi-reference alignment specifying 100 output classes. 3D density maps of individual classes were calculated with the corresponding particles selected from the images of the tilted specimen and using the back-projection, back-projection refinement, and angular refinement procedures implemented in SPIDER. The final 3D reconstruction of the hexameric N-RNA complex included 439 particles (399 particles from images of tilted specimens and 40 particles from images of untilted specimens) and its resolution was estimated by Fourier shell correlation (FSC) to be 25 Å according to the FSC = 0.5 criterion. The crystal structure of hexamer I formed by native RVFV N was first manually docked into the EM density map and then refined using the program UCSF Chimera [46].

Supporting Information

Figure S1 Surface plasmon resonance spectroscopy analysis of RNA binding by RVFV N.A 20-nucleotides-long RNA was immobilized on a NeutrAvidin chip, and association and dissociation phases were measured for 240 sec and 600 sec, respectively, using the indicated concentrations of RVFV N. Data are representative of two independent experiments.
(TIF)

Figure S2 Hexamers in the crystals form tubes in the direction of the *c* axis. Top and side view of the tubes formed by hexamers I and II (same color code as in Figure 2) in the direction of the *c* axis. The six subunits in hexamer II are labeled A to F. The arrows indicate that the two tubes run in opposite directions.
(TIF)

Figure S3 Comparison of the two crystal structures of RVFV N. **(A)** Ribbon representation of the two crystal structures of an RVFV N subunit in the hexamer (PDB code: 3OU9) and as monomer (PDB code: 3LYF). Color code is the same as in Figure 3. **(B)** Side and top view of a superimposition of the two crystal structures based on the core domain. The hexamer structure is shown in orange (PDb code: 3OU9) and the monomer structure in purple (PDB code: (PDB code: 3LYF [15]). The rmsd between the backbone atoms of the core domains is~0.7 Å. The U-shaped arrow indicates the movement of the N-terminal arm.
(TIF)

Figure S4 Multiple sequence alignment of N proteins from the *Phlebovirus* family. Invariant residues are shown in white with red background, conserved residues are shown in red with white background, and variable residues are shown in black with white background. The secondary structure elements are indicated above the alignment with the same color code used in Figure 3. The sequence alignment was generated with ClustalW, and secondary structure was assigned with ESPript. The sequences and their database accession numbers are: Rift Valley fever virus (GI 9632367), Phlebovirus sp. Be Ar 371637 (GI 146336853), Phlebovirus sp. VP-161A (GI 146336850), Phlebovirus sp. PAN 483391 (GI 146336925) Phlebovirus sp. Pa Ar 2381 (GI 146336916), Punta Toro virus (GI 146336898), Phlebovirus sp. GML 902878 (GI 146336904), Phlebovirus sp. VP-366G (GI 146336907), Phlebovirus sp. Co Ar 171616 (GI 146336901), Buenaventura virus (GI 146336910), Sandfly fever sicilian virus (GI 146336868), Corfou virus (GI 146336856), Massilia virus (GI 208610196), Sandfly fever Naples virus (GI 146336886), and Uukuniemi virus (GI 38371708).
(TIF)

Figure S5 Comparison of hexamers formed by native and seleniated RVFV N. **(A)** The left panel shows hexamers I and II formed by native protein in pink and cyan, respectively, and the right panel shows hexamers I and II formed by seleniated protein in marine and yellow, respectively. Hexamer I formed by seleniated N has a different organization from all the other hexamers. **(B)** Superimposition of hexamer I formed by native N (pink) with hexamers II formed by native and seleniated N (cyan and yellow), showing that the subunits in these rings have an identical arrangement. **(C)** Superimposition of hexamer I formed by native N (pink) with hexamer I formed by seleniated N (marine), revealing an 11° rotation between the planes of the two rings.
(TIF)

Figure S6 Structural variability in the N protein and in N-N interactions. **(A)** N-N interaction in hexamer I formed by the native protein. In panels A to C, the subunit shown in grey surface representation is fixed, with the hydrophobic groove shown in color, and the interacting subunit is shown in wire mesh. **(B)** N-N interaction in hexamer I formed by the seleniated protein. **(C)** Comparison of the N-N interactions shown in panels A and B. The superimposition reveals a shift of the protein core with respect to the N-terminal arm (lateral slippage). **(D)** Comparison of the relative position of the N-terminal arm in hexamer I formed by the native protein (top panel) and hexamer I formed by the seleniated protein (bottom panel). In panels D to G, the native proteins are shown in cyan and yellow and the seleniated proteins in marine and brown. **(E)** Comparison of the angle between two subunits in hexamer I formed by the native protein (top panel) and hexamer I formed by the seleniated protein (bottom panel), showing that the angle is identical. **(F)** Comparison of the relative position of two subunits in hexamer I formed by the native protein (left panel) and hexamer I formed by the seleniated protein (right panel), showing a deviation of 11° between the two monomers. **(G)** The arrangement of the cores of two subunits shows a lateral slippage of 2.3 Å.
(TIF)

Figure S7 The R64D/K67D/K74D triple mutant fails to bind RNA. **(A)** Elution profile of the R64D/K67D/K74D triple N mutant from a S200 size exclusion column. The blue line shows the absorbance at 280 nm. The inset shows a 12.5% SDS-PAGE gel of the elution fractions of the peak, revealing a protein band at 27 kDa. **(B)** Surface plasmon resonance profile for binding of a 20-nucleotides-long RNA by wild-type N (blue line) and the triple mutant (red line). Association and dissociation phases were measured for 100 sec and 500 sec, respectively.
(TIF)

Figure S8 Modeling of an RNA molecule into the basic cleft of an RVFV N dimer. The RNA molecule was positioned based on the RNA seen in the structure of the N protein from RV (PDB code: 2GTT; [11]) with some manual adjustments in order to fit the RNA molecule into the cavity. The RNA is colored according to the atoms, with carbon in white, oxygen in red, phosphate in orange, and nitrogen in blue. Positively charged residues that were substituted in the triple mutant are shown as sticks and labeled.
(TIF)

Figure S9 Class averages of cryo-negatively stained N-RNA oligomers. The 100 class averages, obtained from the classification of 10,764 particles, are arranged according to particle number such that the upper-left panel shows the average with the most particles and the lower-right panel shows the average with the least particles. The side length of the individual panels is 24 nm.
(TIF)

Figure S10 Fourier shell correlation (FSC) curve of the single-particle EM reconstruction of the hexameric N-RNA complex. The actual FSC curve (dashed line) and a smoothened

representation (continuous line), suggesting that the density map has a resolution of 25 Å according to the FSC = 0.5 criterion.
(TIF)

Figure S11 Intra-molecular interaction of the N-terminal arm with its own core domain. **(A)** Surface representation of the N protein in monomeric form (PDB code: 3LYF; [15]). The surface is shown transparent and in purple except the hydrophobic residues that interact with the N-terminal arm, which are shown in green. The Cα backbone of N is shown in red and the side chains of the residues of the N-terminal arm that interact with the hydrophobic groove in the core domain are shown as yellow sticks. **(B)** Amino acid sequence of the RVFV N polypeptide, showing above the secondary structure elements derived from the crystal structure. Below the sequence, residues are labeled that are involved in intra-subunit interactions. Yellow dots indicate residues of the N-terminal arm interacting with the core domain and green bars indicate residues of the core domain interacting with the N-terminal arm in the monomer (PDB: 3LYF). For reference, residues involved in intermolecular interactions with the N-terminal arm of an adjacent molecule are indicated by an orange dot below the sequence (hexameric structure).
(TIF)

Video S1 Transition of the N-terminal arm between the positions seen in the two crystal structures.The crystal structures of monomeric N (PDB codes: 3LYF; [15]) and hexameric N (PDB codes: 3OU9) were overlayed based on the core domain, and the movement of the N-terminal arm between the positions seen in the two crystal structures was simulated using the program LSQMAN. The N-terminal arm is shown in red, the globular core domain in brown and green, and the C terminus in blue.
(MOV)

Acknowledgments

We are grateful to Janet Smith for insightful discussions. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081). We thank Biorad for access to their Proteon instrument and the ESRF for allocation of beam time. We thank Joelle Boretto and Ramya Chandrasekaran for their assistance with mutagenesis and surface plasmon resonance experiments.

Author Contributions

Conceived and designed the experiments: F. Ferron, J. Lescar. Performed the experiments: F. Ferron, Z. Li, E. Danek, D. Luo, Y. Wong, B. Coutard, V. Lantez, R. Charrel, J. Lescar. Analyzed the data: F. Ferron, Z. Li, E. Danek, B. Canard, T. Walz, J. Lescar. Contributed reagents/materials/analysis tools: R. Charrel, B. Coutard. Wrote the paper: F. Ferron, T. Walz, J. Lescar.

References

1. Schmaljohn C, Hooper JW (2001) Bunyaviridae: the viruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, et al. (2001) Field Virol 4th ed. PhiladelphiaPa.: Lippincott, Williams and Wilkins. pp 1581–1602.
2. Balkhy HH, Memish ZA (2003) Rift Valley fever: an uninvited zoonosis in the Arabian peninsula. Int J Antimicrob Agents 21: 153–157.
3. Chevalier V, Pepin M, Plee L, Lancelot R (2010) Rift Valley fever—a threat for Europe? Euro Surveill 15: 19506.
4. Weaver SC, Reisen WK (2010) Present and future arboviral threats. Antiviral Res 85: 328–345.
5. Ikegami T, Makino S (2009) Rift valley fever vaccines. Vaccine 27(Suppl 4): D69–72.
6. Anyamba A, Chretien JP, Small J, Tucker CJ, Formenty PB, et al. (2009) Prediction of a Rift Valley fever outbreak. Proc Natl Acad Sci U S A 106: 955–959.
7. Barr JN, Wertz GW (2005) Role of the conserved nucleotide mismatch within 3′- and 5′-terminal regions of Bunyamwera virus in signaling transcription. J Virol 79: 3586–3594.
8. Morin B, Coutard B, Lelke M, Ferron F, Kerber R, et al. (2010) The N-terminal domain of the Arenavirus L protein is an RNA endonuclease essential in mRNA transcription. PLoS Pathog 6(9): e1001038. doi:10.1371/journal.ppat.1001038.
9. Bhella D, Ralph A, Yeo RP (2004) Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction. J Mol Biol 340: 319–331.
10. Schoehn G, Mavrakis M, Albertini A, Wade R, Hoenger A, et al. (2004) The 12 A structure of trypsin-treated measles virus N-RNA. J Mol Biol 339: 301–312.
11. Albertini AA, Wernimont AK, Muziol T, Ravelli RB, Clapier CR, et al. (2006) Crystal structure of the rabies virus nucleoprotein-RNA complex. Science 313: 360–363.
12. Green TJ, Zhang X, Wertz GW, Luo M (2006) Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Science 313: 357–360.
13. Cox R, Green TJ, Qiu S, Kang J, Tsao J, et al. (2009) Characterization of a mumps virus nucleocapsidlike particle. J Virol 83: 11402–11406.
14. Tawar RG, Duquerroy S, Vonrhein C, Varela PF, Damier-Piolle L, et al. (2009) Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. Science 326: 1279–1283.
15. Raymond DD, Piper ME, Gerrard SR, Smith JL (2010) Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. Proc Natl Acad Sci U S A 107: 11769–11774.
16. Liu L, Celma CC, Roy P (2008) Rift Valley fever virus structural proteins: expression, characterization and assembly of recombinant proteins. Virol J 5: 82.
17. Le May N, Gauliard N, Billecocq A, Bouloy M (2005) The N terminus of Rift Valley fever virus nucleoprotein is essential for dimerization. J Virol 79: 11974–11980.
18. Iseni F, Barge A, Baudin F, Blondel D, Ruigrok RW (1998) Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. J Gen Virol 79(Pt 12): 2909–2919.
19. Luo M, Green TJ, Zhang X, Tsao J, Qiu S (2007) Structural comparisons of the nucleoprotein from three negative strand RNA virus families. Virol J 4: 72.
20. Ruigrok RW, Baudin F (1995) Structure of influenza virus ribonucleoprotein particles; II. Purified RNA-free influenza ribonucleoprotein froms structures that are indistinguishable from the intact influenza virus ribonucleoprotein particles. J Gen Virol 76(Pt 4): 1009–1014.
21. Cheng Y, Wolf E, Larvie M, Zak O, Aisen P, et al. (2006) Single particle reconstructions of the transferrin-transferrin receptor complex obtained with different specimen preparation techniques. J Mol Biol 355: 1048–1065.
22. Radermacher M, Wagenknecht T, Verschoor A, Frank J (1987) Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. J Microsc 146: 113–136.
23. Pettersson RF, von Bonsdorff CH (1975) Ribonucleoproteins of Uukuniemi virus are circular. J Virol 15: 386–392.
24. Saikku P, von Bonsdorff CH, Brummer-Korvenkontio M, Vaheri A (1971) Isolation of non-cubical ribonucleoprotein from Inkoo virus, a Bunyamwera supergroup arbovirus. J Gen Virol 13: 335–337.
25. Samso A, Bouloy M, Hannoun C (1975) [Circular ribonucleoproteins in the virus Lumbo (Bunyavirus)]. C R Acad Sci Hebd Seances Acad Sci D 280: 779–782.
26. Samso A, Bouloy M, Hannoun C (1976) [Demonstration of circular ribonucleic acid in the Lumbo virus (Bunyavirus)]. C R Acad Sci Hebd Seances Acad Sci D 282: 1653–1655.
27. Hewlett MJ, Pettersson RF, Baltimore D (1977) Circular forms of Uukuniemi virion RNA: an electron microscopic study. J Virol 21: 1085–1093.
28. Lopez N, Muller R, Prehaud C, Bouloy M (1995) The L protein of Rift Valley fever virus can rescue viral ribonucleoproteins and transcribe synthetic genome-like RNA molecules. J Virol 69: 3972–3979.
29. Flick R, Elgh F, Pettersson RF (2002) Mutational analysis of the Uukuniemi virus (Bunyaviridae family) promoter reveals two elements of functional importance. J Virol 76: 10849–10860.
30. Osborne JC, Elliott RM (2000) RNA binding properties of bunyamwera virus nucleocapsid protein and selective binding to an element in the 5′ terminus of the negative-sense S segment. J Virol 74: 9946–9952.
31. Rancurel C, Khosravi M, Dunker AK, Romero PR, Karlin D (2009) Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. J Virol 83: 10719–10736.
32. Kabsch W (2010) Xds. Acta Crystallogr D Biol Crystallogr 66: 125–132.
33. Sheldrick GM (2008) A short history of SHELX. Acta Crystallogr A 64: 112–122.
34. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS (2001) ARP/wARP and molecular replacement. Acta Crystallogr D Biol Crystallogr 57: 1445–1450.
35. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126–2132.
36. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674.

37. Winn MD, Isupov M, Murshudov GN (2000) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr D Biol Crystallogr 57: 122–133.
38. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26: 283–291.
39. Rocchia W, Alexov E, Honig B (2001) Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. The J Phys Chem B 105: 6507–6514.
40. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.
41. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27: 221–224.
42. Kleywegt GJ (1996) Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr D Biol Crystallogr 52: 842–857.
43. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 31: 3320–3323.
44. Ohi M, Li Y, Cheng Y, Walz T (2004) Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. Biol Proced Online 6: 23–34.
45. Frank J, Radermacher M, Penczek P, Zhu J, Li Y, et al. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol 116: 190–199.
46. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25: 1605–1612.

C. Structures et fonctions de protéines de la formation de la coiffe et du complexe de transcription/réplication

Crystal Structure and Functional Analysis of the SARS-Coronavirus RNA Cap 2′-O-Methyltransferase nsp10/nsp16 Complex

Etienne Decroly[1]*, Claire Debarnot[1], François Ferron[1], Mickael Bouvet[1], Bruno Coutard[1], Isabelle Imbert[1], Laure Gluais[1], Nicolas Papageorgiou[1], Andrew Sharff[2], Gérard Bricogne[2], Miguel Ortiz-Lombardia[1], Julien Lescar[1,3], Bruno Canard[1]*

1 Centre National de la Recherche Scientifique and Université de la Méditerranée, UMR 6098, Architecture et Fonction des Macromolécules Biologiques, Marseille, France, **2** Global Phasing Ltd., Sheraton House, Castle Park, Cambridge, United Kingdom, **3** School of Biological Sciences, Nanyang Technological University, Singapore, Republic of Singapore

Abstract

Cellular and viral S-adenosylmethionine-dependent methyltransferases are involved in many regulated processes such as metabolism, detoxification, signal transduction, chromatin remodeling, nucleic acid processing, and mRNA capping. The Severe Acute Respiratory Syndrome coronavirus nsp16 protein is a S-adenosylmethionine-dependent (nucleoside-2′-O)-methyltransferase only active in the presence of its activating partner nsp10. We report the nsp10/nsp16 complex structure at 2.0 Å resolution, which shows nsp10 bound to nsp16 through a ~930 Å^2 surface area in nsp10. Functional assays identify key residues involved in nsp10/nsp16 association, and in RNA binding or catalysis, the latter likely through a SN2-like mechanism. We present two other crystal structures, the inhibitor Sinefungin bound in the S-adenosylmethionine binding pocket and the tighter complex nsp10(Y96F)/nsp16, providing the first structural insight into the regulation of RNA capping enzymes in (+)RNA viruses.

Citation: Decroly E, Debarnot C, Ferron F, Bouvet M, Coutard B, et al. (2011) Crystal Structure and Functional Analysis of the SARS-Coronavirus RNA Cap 2′-O-Methyltransferase nsp10/nsp16 Complex. PLoS Pathog 7(5): e1002059. doi:10.1371/journal.ppat.1002059

Editor: Félix A. Rey, Institut Pasteur, France

Received December 7, 2010; **Accepted** March 23, 2011; **Published** May 26, 2011

Funding: This work was supported, in its initial phase, by the Euro-Asian SARS-DTV Network (SP22-CT-2004-511064) from the European Commission specific research and technological development Programme "Integrating and strengthening the European Research area." It was subsequently supported by the VIZIER integrated project (LSHG-CT-2004-511960) of the European Union 6th Framework and the French National Research agency, under reference "ANR-08-MIEN-032," and subsequently the Fondation pour la Recherche Médicale (Programme Équipe FRM), and the Direction Générale de l'Armement (contrat n°07co404). MB has a fellowship from the Direction Générale de l'Armement. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: etienne.decroly@afmb.univ-mrs.fr (ED); bruno.canard@afmb.univ-mrs.fr (BC)

Introduction

Most eukaryotic cellular and viral mRNAs are modified by the addition of a polyadenine tail at the 3′- terminal and a cap structure at the 5′-terminal. The RNA cap protects mRNA from degradation by 5′ exoribonucleases, ensures efficient mRNA translation, and prevents recognition of viral RNA *via* innate immunity mechanisms[1,2,3,4]. The RNA cap is made of an N7-methylated guanine nucleotide connected through a 5′-5′ triphosphate bridge to the first transcribed nucleotide, generally an adenine. Through 2′-O methylation of the latter, this cap-0 structure (^{7Me}GpppA…) may be converted into a cap-1 structure (^{7Me}GpppA$_{2'\text{-O-Me}}$…). In the eukaryotic cell, the cap is added co-transcriptionally in the nucleus by three sequential enzymatic reactions[1,5]: (i) an RNA triphosphatase (RTPase) removes the 5′ γ-phosphate group of the nascent mRNA; (ii) a guanylyltransferase (GTase), dubbed capping enzyme, catalyses the attachment of GMP to the 5′-diphosphate mRNA; and (iii) an S-adenosylmethionine (SAM)-dependent (N7-guanine)-methyltransferase (N7MTase) methylates the cap onto the N7-guanine, releasing S-adenosylhomocysteine (SAH). In general, a SAM-dependent (nucleoside-2′-O-)-methyltransferase (2′-O-MTase) further intervenes, in higher eukaryotes, to yield a cap-1 structure.

The viral RNA capping machinery is structurally and mechanistically diverse, and RNA viruses often deviate from the paradigmic eukaryotic mRNA capping scheme. For example, alphaviruses methylate GTP onto the N7-guanine before the presumed attachment of ^{7Me}GMP to the nascent viral 5′-diphosphate mRNA[6]. In the case of single-stranded negative-sense (-)RNA viruses, such as the vesicular stomatitis virus, the L polymerase attaches GDP rather than GMP to a nascent viral 5′-monophosphate mRNA, covalently linked to the viral capping enzyme[7]. Other viruses, such as influenza virus capture a short capped RNA oligonucleotide from host cell mRNAs and use it as an RNA synthesis primer. This process is known as « cap snatching »[8].

In 2003, a novel coronavirus named Severe Acute Respiratory Syndrome coronavirus (SARS-CoV[9]) was responsible for the first viral pandemic of the new millennium with ~8000 cases globally and a 10 % case-fatality rate. Coronaviruses encode an unusually large membrane-associated RNA replication/transcription machinery comprising at least sixteen proteins (nsp1-to-16)[10]. For

Author Summary

A novel coronavirus emerged in 2003 and was identified as the etiological agent of the deadly disease called Severe Acute Respiratory Syndrome. This coronavirus replicates and transcribes its giant genome using sixteen non-structural proteins (nsp1-16). Viral RNAs are capped to ensure stability, efficient translation, and evading the innate immunity system of the host cell. The nsp16 protein is a RNA cap modifying enzyme only active in the presence of its activating partner nsp10. We have crystallized the nsp10/16 complex and report its crystal structure at atomic resolution. Nsp10 binds to nsp16 through a ~930 Å^2 activation surface area in nsp10, and the resulting complex exhibits RNA cap (nucleoside-2′-O)-methyltransferase activity. We have performed mutational and functional assays to identify key residues involved in catalysis and/or in RNA binding, and in the association of nsp10 to nsp16. We present two additional crystal structures, that of the known inhibitor Sinefungin bound in the SAM binding pocket, and that of a tighter complex made of the mutant nsp10(Y96F) bound to nsp16. Our study provides a basis for antiviral drug design as well as the first structural insight into the regulation of RNA capping enzymes in (+)RNA viruses.

SARS-CoV, the RNA cap structure likely corresponds to a cap-1 type[11,12,13]. As in many other (+)RNA viruses, the RTPase activity is presumably embedded in the RNA helicase nsp13, whereas the GTase remains elusive. RNA cap 2′-O-MTase activity was first discovered in the feline coronavirus (FCoV) nsp16[14]. Shortly after, SARS-CoV nsp14 was shown to methylate RNA caps in their N7-guanine position[15]. Curiously, although closely homologous to that of FCoV, recombinant SARS-CoV nsp16 alone was devoid of enzymatic activity. It was demonstrated[16,17,18,19] that nsp10 interacts with nsp16, conferring 2′-O-MTase activity to nsp16 on N7-methyl guanine RNA caps selectively[16]. The latter selectivity implies that RNA cap methylation obeys an ordered sequence of events during which nsp14-mediated N7-guanine methylation precedes nsp10/nsp16 RNA 2′-O methylation. Nsp10 is a double zinc finger protein of 148 residues whose crystal structure is known[20,21]. Together with nsp4, nsp5, nsp12, nsp14, and nsp16, nsp10 has been found to be essential in the assembly of a functional replication/transcription complex[22]. Drawing on these observations, nsp10 has been proposed to play pleiotropic roles in viral RNA synthesis[23] and polyprotein processing through interaction with the main protease nsp5[24].

SAM-dependent MTases belong to a large class of enzymes present in all life forms. These enzymes catalyze the transfer of the SAM methyl group to a wide spectrum of methyl acceptors, indicating that a common chemical reaction is used on a variable active-site environment able to activate the methyl acceptor atom. Although SAM-dependent MTases share little sequence identity, 2′O-MTases exhibit a KDKE catalytic tetrad and a very conserved folding made of a seven-stranded β-sheet surrounded by one to three helices on each side[25], always similar to the paradigmatic catechol-O-MTase[26]. The SAM binding site general location is conserved, suggesting that evolutionary pressure on the MTase fold has maintained the same SAM-binding region whilst accommodating the versatile chemistry of the methyltransfer reaction.

Structural and functional studies of viral MTases involved in RNA capping is an expanding research area, since these enzymes show unexpected diversity relative to their cellular counterparts, and thus constitute attractive antiviral targets. Crystal structures of viral RNA cap MTases exist for only three viral families, namely *Poxviridae*, *Reoviridae*, and *Flaviviridae*. The Vaccinia virus VP39 crystal structure was the first to be elucidated in 1996[27]. The structure of this DNA virus RNA 2′-O-MTase revealed a conserved MTase fold similar to that of RrmJ (also named FtsJ), the canonical reference folding for RNA cap MTases[26]. More recently, the crystal structure of a second Vaccinia virus N7-guanine RNA cap MTase domain (D1) was determined in complex with its activator protein D12[28]. The study revealed that D12 also bears an MTase fold, but has lost catalytic capability due to truncation of its SAM binding site. In turn, *Reoviridae* provided the first RNA cap MTase structures at 3.6 Å resolution as forming part of the reovirus core[29]. Another RNA cap machinery was more recently described for the non-turreted orbivirus Bluetongue virus VP4 protein at 2.5 Å resolution[30], which revealed a three-domain protein, with a "head" guanylyltransferase domain, a central N7-guanine MTase, and a "bottom" 2′-O-MTase domain. This architecture illustrates the sequence of three out of the four chemical reactions involved in RNA capping described above.

Regarding (+)RNA viruses, MTase structural information at the atomic level is only available for a single genus. The flavivirus N-terminus domain (residues 1–265) of the NS5 RNA-dependent RNA polymerase harbors an RrmJ fold with an N-terminus extension able to accommodate RNA cap structures[31,32]. This enzyme carries both N7-guanine MTase and 2′-O-MTase activities on a single domain with one shared active site[33]. Homologous domains have been crystallized for a number of flaviviruses, revealing a conserved fold and activity[34], suggesting that MTases might represent interesting targets for drug design. No other (+)RNA virus RNA cap MTase crystal structures have as yet been defined.

In 2003, the identification of the 2′-O-MTase signature sequence in the SARS-CoV genome added nsp16 to the list of putative targets for antiviral drugs[35]. Several compounds have been shown to inhibit viral MTases, such as the co-product of the MTase reaction SAH, Sinefungin, and aurintricarboxylic acid (ATA)[14,36,37,38,39]. In this paper, we report the crystal structure of the SARS-CoV 2′-O-MTase nsp16 in complex with its activator, the zinc finger protein nsp10, at 2.0 Å resolution, in conjunction with mutagenesis experiments, binding and activity assays. These results lay down the structural basis for the nsp10 function as an activator of nsp16-mediated 2′-O-MTase. We identify residues playing key roles in the nsp10/nsp16 interaction, as well as other residues involved in 2′-O-MTase catalysis and RNA binding. We also report the crystal structure of the nsp10/nsp16 complex bound to the inhibitor Sinefungin. Comparison with known cellular SAM binding sites points to the nsp16 nucleobase binding pocket as a possible target for the design of selective antiviral molecules.

Results

Crystallization and Structure Determination of an Active nsp16 2′-O-MTase

We observed that purified nsp16 was unstable in solution, impeding crystallogenesis. Yeast double-hybrid and co-immunoprecipitation experiments on purified SARS-CoV nsp10 and nsp16 have uncovered the reciprocal interaction of these two proteins[16,17,18]. Indeed, SARS-CoV nsp16 exhibits 2′-O-MTase activity only when complemented with SARS-CoV nsp10, raising the interesting possibility that nsp10 acted as a scaffold for nsp16. Co-expression of nsp10 and nsp16 using a bi-cistronic prokaryotic expression vector facilitated affinity chromatography purification and crystallization of the complex[16,40]. Crystals diffracted to ~1.9 Å. The position of the nsp10 protein was determined using molecular replacement with the SARS-CoV nsp10 protein structure[20] as a search model. Strong peaks in both the residual and anomalous Fourier maps confirmed the presence of two zinc ions. Nsp16 was well defined by its electron density except for two flexible loops (residues 19–35 and 135–137) with high B factors and weak or missing electron density. These loops are solvent-exposed at each side of the putative RNA-binding groove (see below). Structure determination data and refinement statistics are reported in Table 1.

Structure of the nsp10/16 Heterodimer

The heterodimer can be conveniently viewed as nsp16 sitting on top of a nsp10 monomer (Fig. 1A). The nsp10 overall structure in the complex remains essentially unchanged relative to published structures of nsp10 alone, with its N-terminus comprising two α-helices, a central β-sheet domain, and a C-terminus domain containing various loops and helices (see[20,21], Fig. 1B). Comparison with existing crystal structures of nsp10 using DaliLite[41] rendered nsp10 atomic coordinates very similar to those of nsp10 in our nsp10/nsp16 complex. The average RMSD is about 0.77 Å in 118 residues (PDB codes 2FYG, 2G9T and 2GA6[20,21]). This indicates that neither significant conformational change nor surface modification occurs in nsp10 when binding to nsp16. The nsp10 structural Zn^{2+} ions are not directly involved in the nsp10/nsp16 interface (Fig. 1A).

Nsp16 adopts a canonical SAM-MT fold (Figs. 1B, 2A and B), as defined initially for the catechol O-MTase[25]. The seven-stranded β-sheet MTase fold has been described as having a secondary structure topology defining two binding domains, one for SAM and the other for the methyl acceptor substrate (Fig. 2A). The nsp16 topology matches those of dengue virus NS5 N-terminal domain and of vaccinia virus VP39 MTases[27,31]. Nsp16 lacks several elements of the canonical MTase fold, such as helices B and C (Fig. 2B).

S-adenosylhomocysteine- and Sinefungin-Binding

Electron density corresponding to one molecule of S-adenosylhomocysteine (SAH), the co-product of the methylation reaction, was identified in the putative SAM-binding site (Figs. 1A and 3A). Neither SAM nor SAH was added to the purification or crystallization buffers, therefore it must have been captured from the medium by nsp16 during bacterial growth. The SAH molecule is found with its adenine in an *anti* conformation and the ribose pucker in a *southern* (2′-endo/3′-exo) conformation. All the residues involved in SAM/SAH binding are absolutely conserved in coronavirus np16s (Fig. S1). Binding specificity for SAM/SAH is achieved by holding distal SAM/SAH carboxylic and amino groups through five hydrogen bonds (G81, N43, Y47, G71, and D130) (Fig. 3A and Fig. S2A). The ribose moiety is held by three hydrogen bonds involving Y132, G73, and D99. As in the case of other MTases[25], the SAH binding cleft is globally positively charged. However, an aspartic acid (D99) acts as the ribose-sensing residue with its side chain carboxyl making strong hydrogen bonds with both ribose hydroxyls (Fig. S2A). Binding of the adenine base involves few contacts. The nucleobase occupies a loose hydrophobic pocket engaging two hydrogen bonds of moderate strength with side chain and main chain atoms of conserved residues D114 and C115, respectively.

Soaking the crystals into a Sinefungin-containing buffer captured this MTase inhibitor in the SAH binding site almost perfectly superimposable on SAH (Fig. 3B and Fig. S2B). Binding involved the same residues and contacts as SAH. Inhibition of the MTase reaction by Sinefungin therefore probably occurs competitively. The Sinefungin amino group quasi-isosteric to the donated SAM methyl group indicates a cavity where the 2′-hydroxyl of the capped RNA is expected to bind. Lining this empty substrate

Table 1. Crystal, collection, structure determination data and refinement statistics.

DATA	wild-type (SAH)	nsp10(Y96F)/nsp16	wild-type (SFG)
Instrument	SOLEIL, Proxima I	ESRF ID14-1	ESRF ID23-1
Wavelength	0.9792	0.9334	0.9792
Space group	$C222_1$	$C222_1$	$C222_1$
Cell dimensions *a*, *b*, *c* (Å)	68.07 184.62 128.83	68.15 184.80 129.01	68.42 185.04 129.46
Resolution range (Å)	37.52-2.00 (2.11 – 2.00)*	45.51-2.05 (2.17 – 2.05)	45.58 -2.50 (2.64 – 2.50)
Total number of reflections	201703 (29338)	187379 (27241)	211213 (30837)
Number of unique reflections	54947 (7963)	51033 (7362)	28921 (4163)
Completeness (%)	99.7 (100)	99.6 (100)	100 (100)
I/σ(*I*)	7.1 (2.7)	7.5 (3.2)	12.5 (4.3)
Rsym **	0.114 (0.430)	0.112 (0.350)	0.102 (0.400)
Multiplicity	3.7 (3.7)	3.7 (3.7)	7.3 (6.7)
Refinement (20 cycles)			
R***	0.213	0.200	0.201
Rfree	0.227	0.234	0.215
RMSD bond length (Å)	0.007	0.007	0.011
RMSD bond angle (Å)	1.015	0.968	1.250

*Values in parentheses give the high resolution shell values.
**Rsym = Σ |I-<I>|/Σ I.
***R = Σ||Fo|-|Fc||/Σ |Fo|.
doi:10.1371/journal.ppat.1002059.t001

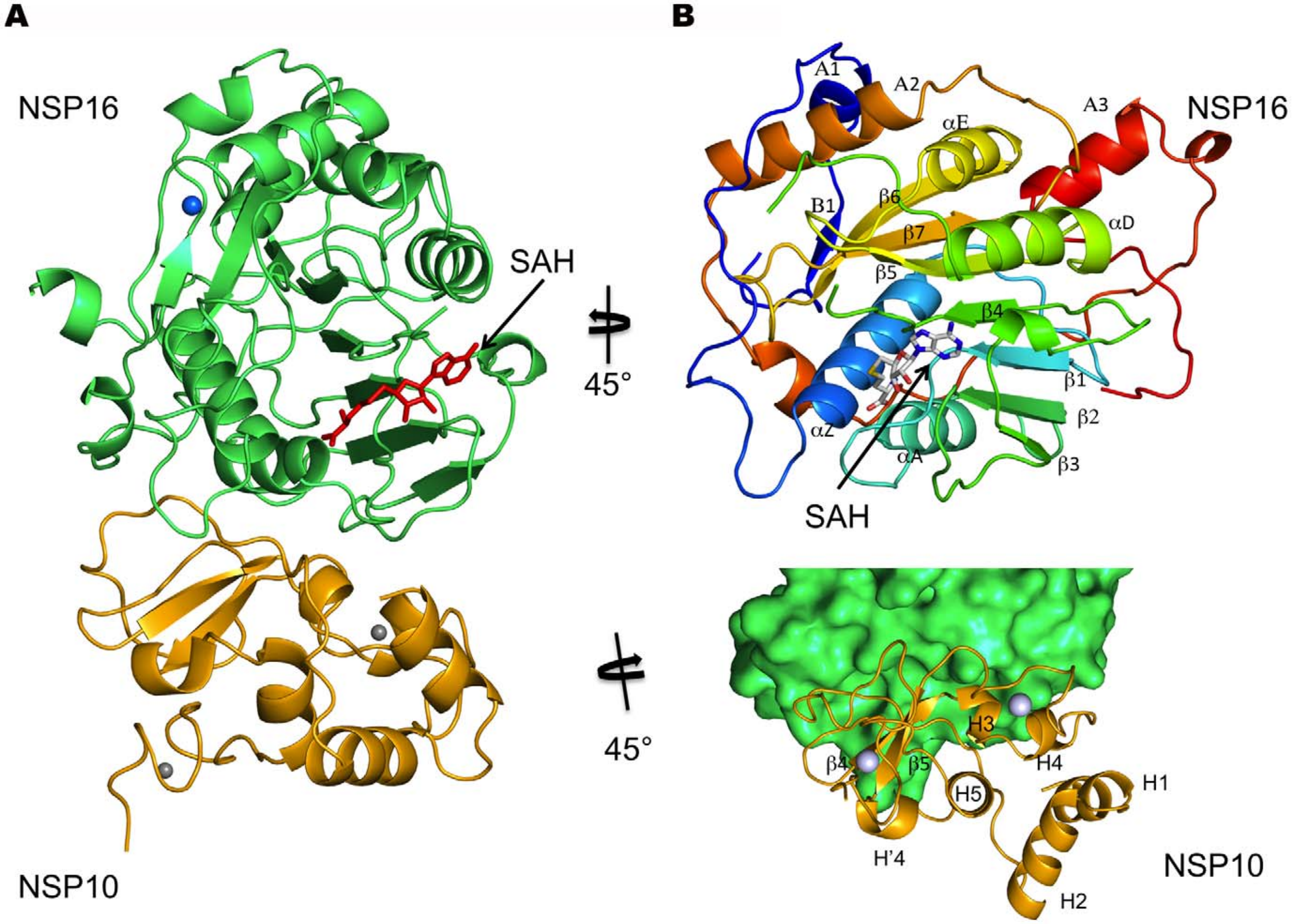


Figure 1. Cartoon representation of the nsp10/nsp16 complex with the reaction product SAH and metal ions. A) The nsp16 protein (green) is bound to nsp10 (yellow) through an interface which does not involve Zn ions (grey spheres) present in nsp10. One metal ion (blue sphere) is found in nsp16 on the opposite face from the active site to which a SAH molecule is found (red sticks). B) Ribbon representation of nsp16, rainbow colors from N- to C-terminus. Top: Each secondary structure element is labeled according to[25] (see also Fig. S3). The SAH molecule shown in sticks colored following atom type. Bottom: View of interface involving nsp16 (green surface) and nsp10 (yellow ribbons) showing the nsp10 secondary structure elements involved in the interface.
doi:10.1371/journal.ppat.1002059.g001

cavity are the residues proposed to be involved in the catalytic reaction: K46, D130, K170, and E203[16]. Alanine substitutions in the catalytic tetrad (K46, D130, K170, or E203) almost completely block 2′-O-MTase activity without jeopardizing binding to nsp10 (Table 2, and [16]). Several SAM-binding residues (N43, G73, D99 and Y132, Fig. S2A) were substituted by alanine. Although they conserve their specific nsp10 binding properties, indicating that they are correctly folded, they all show a drastically reduced MTase activity (Table 2), validating the structural description of the nsp10/nsp16/SAH ternary complex.

A Mg^{2+} Cation is Present in nsp16, Outside the Active Site

We recently reported[16] that nsp10/nsp16 MTase activity requires Mg^{2+}. Although the crystallization buffer contains Mg^{2+}, we were unable to locate any such cation in the nsp16 active site. In enzyme activity assays, the Mg^{2+} ion can be substituted by Mn^{2+} or $Ca^{2+,}$ but not Zn^{2+} (data not shown, see also[16]). A peak of electron density presumably corresponding to Mg^{2+} is localized onto nsp16, distant from the SAH-binding cavity. The Mg^{2+} coordination mode is through six first-shell water molecules in an octahedral geometry (Fig. 4). Binding *via* water molecules, involves T58 and S188 side chain hydroxyls and the main chain carbonyl of E276. Since there are no carboxylic acids involved in binding this cation, it was suspected that its presence resulted from the crystallization procedure[42], with no biological relevance. However, the T58A, T58N, T58E and S188A substitutions show 43, 70, 99 and 72% loss of activity, respectively (Table 2), with no significant effect on the stability of the nsp10/nsp16 complex except for T58E whose association was 54 % that of wild-type. These residues are located on three distinct structural elements at the C-terminus of helix Z (T58), the N-terminus of β6 (S188), and in the central part of helix A3 (E276), respectively (Fig. 4 and Fig. S3). The cation may thus hold these elements together.

The Putative RNA Binding Site: Mutagenesis and Effect on Activity

The nsp10/nsp16 complex absolutely requires an N7-methyl guanine capped RNA substrate to exhibit MTase activity[16]. The structural basis for the preferential binding to methylated N7-guanine versus non-methylated caps has been elucidated in four cases, those of VP39[27], eIF4E[43], CBC[44], and PB2[45] proteins (PDB codes 1AV6, 1EJ1, 1H2T, and 2VQZ, respectively)

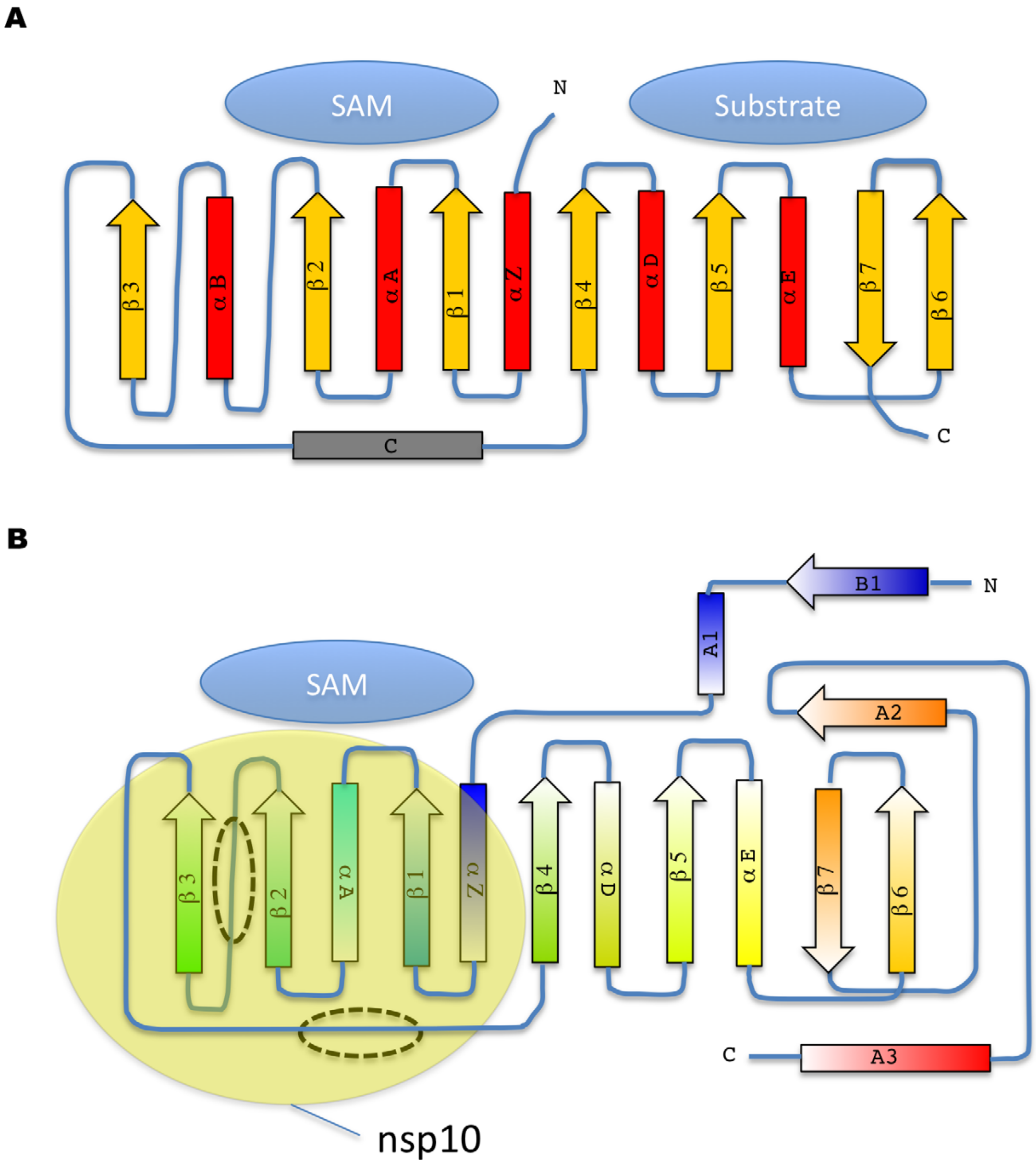


Figure 2. Topology diagram of MTase secondary structure elements. A) The consensus topology diagram is shown with a two domain organization for methyl acceptor substrate and SAM, as defined[25] with the catechol O-MTase and its canonical seven strand beta sheet. B) Topology diagram of nsp16 colored according to rainbow colors from N-to C-terminus as in Fig. 1B. The missing helix B and helix C are indicated by dashed ovals. The approximate general binding site of nsp10 is shown (yellow oval).
doi:10.1371/journal.ppat.1002059.g002

bound to cap analogues or capped RNAs. In these cases, the methylated base specificity is achieved through increased binding energy resulting from the stacking of the N7-methyl guanine between parallel aromatic residues of the cap binding protein. The presence of the methyl group greatly enhances π-π stacking, providing a dominant effect over unmethylated guanine[46]. Despite numerous attempts, cap analogues (m7GpppA, GpppA, m7GpppG, GpppG) and short capped RNA substrates (m7GpppA(C)$_n$) could neither be co-crystallized with nsp10/nsp16 nor soaked and bound onto preformed nsp10/nsp16 crystals. However, the atomic coordinates of the N7-methyl guanine RNA oligomer in complex with VP39[47] provided data from which a model of RNA binding to the nsp16 protein was derived. SAM molecules identified in both structures were superimposed, and the VP39-bound RNA was positioned onto the nsp16 structure. After minimal manual adjustments not exceeding 5 Å, the VP39 RNA was a reasonably good fit into an nsp16 hydrophobic groove radiating from the catalytic site (Fig. 5), establishing very few contacts with nsp10. We note that the protein side diametrically opposite

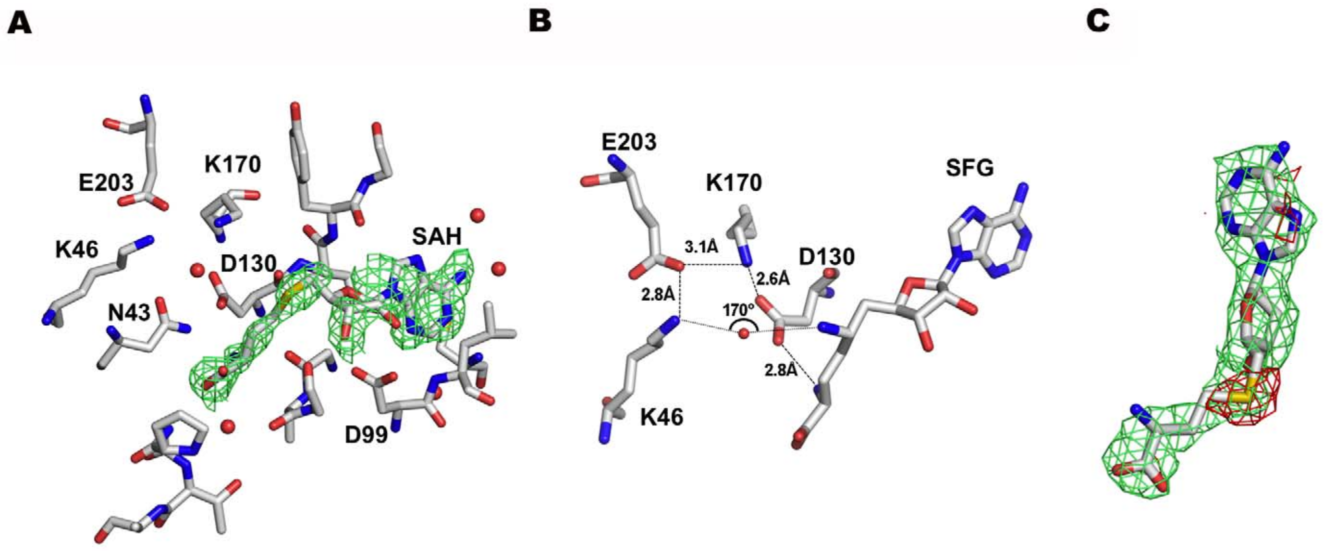


Figure 3. The SAM binding site of nsp16. A) SAH modeled in a simulated annealing Omit map contoured at 1σ. Carbons, oxygens, sulfur, and nitrogens are in grey, red, yellow, and blue sticks, respectively. Water molecules are shown as red spheres. B) Sinefungin bound in the SAM/SAH binding site, with main catalytic residues and the water molecule indicative of a catalytic mechanism. Colors as in A). C) SAH modeled in a 2Fo-Fc map (green) of Sinefungin contoured at 1σ with a Fo-Fc difference map (red) contoured at 3σ. A peak of negative density appears clearly on the sulfur atom of the SAH molecule showing that SAH was indeed replaced by Sinefungin. Colors as in A).
doi:10.1371/journal.ppat.1002059.g003

to the proposed hydrophobic RNA binding groove is highly positively charged (not shown), an observation that may account for the difficulty of achieving experimental RNA binding in the proposed RNA binding site. In the absence of robust data to guide docking of the guanine cap, the m7Gpp cap structure was not positioned in the structure but two possible N7-methylated cap guanine binding areas are indicated by arrows (Fig. 5). The first transcribed nucleotide together with its ribose receiving the methyl group fit well in the active site (Fig. 5, panel B) as predicted in the proposed mechanism. The same holds for the immediately preceding three nucleotides. The base of the first transcribed nucleotide may be held by contact with P134 and Y132, bending the extending RNA cap structure. Accordingly, the substitution of Y132 greatly depresses MTase activity (Table 2). We also note that Y132 is located in the vicinity of a highly mobile loop (residues 135-138) not always visible in our crystal structures suggesting that this loop may move in order to wrap the triphosphate moiety of the RNA cap and/or the RNA cap itself. The solvent exposed side chain of Y30 may also participate in RNA binding. In the model, the highly mobile side chain of Y30 was flipped out in an alternative conformation in order to open the groove. In that position, Y30 should specifically contact the third transcribed nucleotide. Our mutagenesis data confirms the importance of Y30 since its replacement with either Ala or Phe severely impairs MTase activity without affecting the interaction with nsp10 (Table 2).

Interface of the Heterodimer

All nsp10 secondary structure elements but helices 2 and 5 contact nsp16 (Fig. 1). The nsp10 contact points can be viewed as 5 small patches A to E (residues 40–47, 57–59, 69–72, 77–80, and 93–96, respectively, Fig. 6A). In turn, these five patches contact most of the nsp16 SAM-binding structural elements in 4 areas, I to IV (Fig. 6B, Fig. S3)), mainly involving β2, β3, αA, αZ, and for area IV, the loop connecting helices A2 and A3 at the C-terminus (Figs. 1 and 6). In total, the interface of the heterodimer involves 53 residues, 23 and 30 from nsp10 and nsp16, respectively. In nsp10, a single residue (Asn10, at the edge of the interaction surface) is not conserved out of 23 (4.3 %), whereas in nsp16 there are 8 non-conserved residues out of 30 (26.7 %) (Fig. S1). The interface has a buried surface area of 1820 Å^2, with nsp10 contributing to 930 Å^2 and nsp16 to 890 Å^2.

Four nsp10 patches included in the 5 interaction patches identified here were recently mapped using reverse yeast two-hybrid methods coupled to bioluminescence resonance energy transfer and *in vitro* pull-down assays (see[18] and below). To probe the observed crystal structure of the interface further, we engineered 5 new nsp10 alanine mutants (N40A, L45A, T58A, G69A, and H80A, see Table S1) sitting in patches A, B, C and D. Whereas T58A, G69A and H80A showed limited effect on nsp16 binding, N40A reduced it to 64% of wild type affinity, and L45A almost abrogated it. The crystal structure indicates that the nsp10 interface proposed by Lugari *et al.*[18], is a correct and conservative estimation, as the interface also includes L45 belonging to patch A. We also confirm the positive co-relation of the detected nsp10/nsp16 interaction with MTase activity. In no instance can nsp16 be active in the absence of nsp10/16 complex formation.

The Y96 position is of particular interest. Alanine substitution (Y96A) abrogates interaction whereas a phenylalanine (Y96F) increases both interaction and MTase activity[18]. In order to understand how residue 96 plays such a pivotal role, we determined at 2.0 Å resolution the crystal structure of this nsp10(Y96F)/nsp16 complex (Table 1). Strikingly, the absence of the hydroxyl group does not alter the topology of the interface. Wild-type and Y96F residues superimpose without significant difference at all atomic positions (not shown). Either Y96 or F96 is in direct contact with nsp16 helix αZ, which carries the catalytic residue K46. Detailed surface analysis using PISA indicates that the position of K46 in Y96F nsp10 is identical to that of K46 in wild-type nsp10, ruling out a better alignment of catalytic residues of the Y96F mutant. The nsp10(Y96F)/nsp16 differs from wild-type nsp10/nsp16 in the SAH binding site, though. When compared to that of wild-type, the SAH occupancy is much lower

Table 2. Mutational analysis, complex formation, and enzyme activity of the nsp10/nsp16 complex.

Number	Function*	Mutant	% associated**	% MTase act. ***
1 = Wild-Type	Wild-TypE	Wild-Type	100	100
2	Catalytic	K46A	97	1
3	Catalytic	D130A	104	2
4	Catalytic	K170A	95	0
5	Catalytic	E203A	93	0
6	SAM-BS	N43A	144	11
7	SAM-BS	G73A	119	18
8	SAM-BS	D99A	57	0
9	Mg^{2+}-BS	T58A	123	57
10	Mg^{2+}-BS	T58N	110	30
11	Mg^{2+}-BS	T58E	54	1
12	Mg^{2+}-BS	S188A	102	28
13	RNA/SAM-BS	Y132A	86	5
14	RNA/SAM-BS	Y132T	95	5
15	RNA/SAM-BS	Y132F	123	9
16	RNA/SAM-BS	Y132H	106	0
17	RNA-BS	Y30A	89	1
18	RNA-BS	Y30F	113	6
Interface mutants				
Number	**Location***	**Mutant**	**% associated****	**% MTase act. *****
19	Patch I	I40A	84	8
20	Patch I	M41A	122	4
21	Patch I	V44A	113	1
22	Patch I	T48A	123	20
23	Patch II	V78A	3	0
24	Patch II	R86A	61	0
25	Patch II	Q87A	114	62
26	Patch III	V104G	6	4
27	Patch III	D106A	101	38
28	Patch IV	L244A	30	0
29	Patch IV	M247A	4	0

*as inferred from the crystal structure.
**nsp10/nsp16 complex formation as determined using pull-down experiments, SDS-PAGE analysis, and quantitation (see Methods).
***The % of MTase activity was determined using filter binding assays (see Methods) relative to wild-type.
SAM-BS: S-adenosylmethionine binding site.
Mg^{2+-}BS: Magnesium binding site.
RNA-BS: RNA binding site.
doi:10.1371/journal.ppat.1002059.t002

(~0.3 versus ~1), leading to poor density definition. SAH is known to be a fairly good inhibitor of the methylation reaction. Therefore, a lower binding affinity might translate into less end-product inhibition, and account for the observed increased activity. We measured the affinity of nsp16 for SAH using fluorescence spectroscopy, but no significant differences were found (not shown). Likewise, the MTase inhibition pattern by SAH was identical for wild-type and nsp10(Y96F)/nsp16 (not shown). We therefore infer that the previously observed ~10-fold increased stability of the heterodimer[18] may be responsible for the increased activity. A more hydrophobic character of the interaction may appear upon the loss of the tyrosine hydroxyl which, in the wild-type protein, was not engaged in any polar contact. We thus attribute the increased activity of the nsp10(Y96F)/nsp16 complex relative to wild-type to a stronger equilibrium association of nsp10(Y96F) with nsp16 than that of wild-type nsp10 with nsp16.

Mutation analysis was also conducted on nsp16 residues presumably involved in the interface and interfacial activation (Tables 2 and S2). Several mutants (V78A, V104A, L244A, M247A) in patches II, III and IV completely disrupt the nsp10/nsp16 complex and annihilate nsp16 MTase activity. Interestingly, we also identified nsp16 mutants still interacting with nsp10, but with a strongly reduced 2′-O-MTase activity (I40A, M41A, V44A, T48A, Q87A, D106A) suggesting, that these mutations in the nsp10/nsp16 interface may alter the fine positioning of catalytic residues without any significant effect on nsp10 binding. Accordingly, most of these mutants are localized in αZ helix of

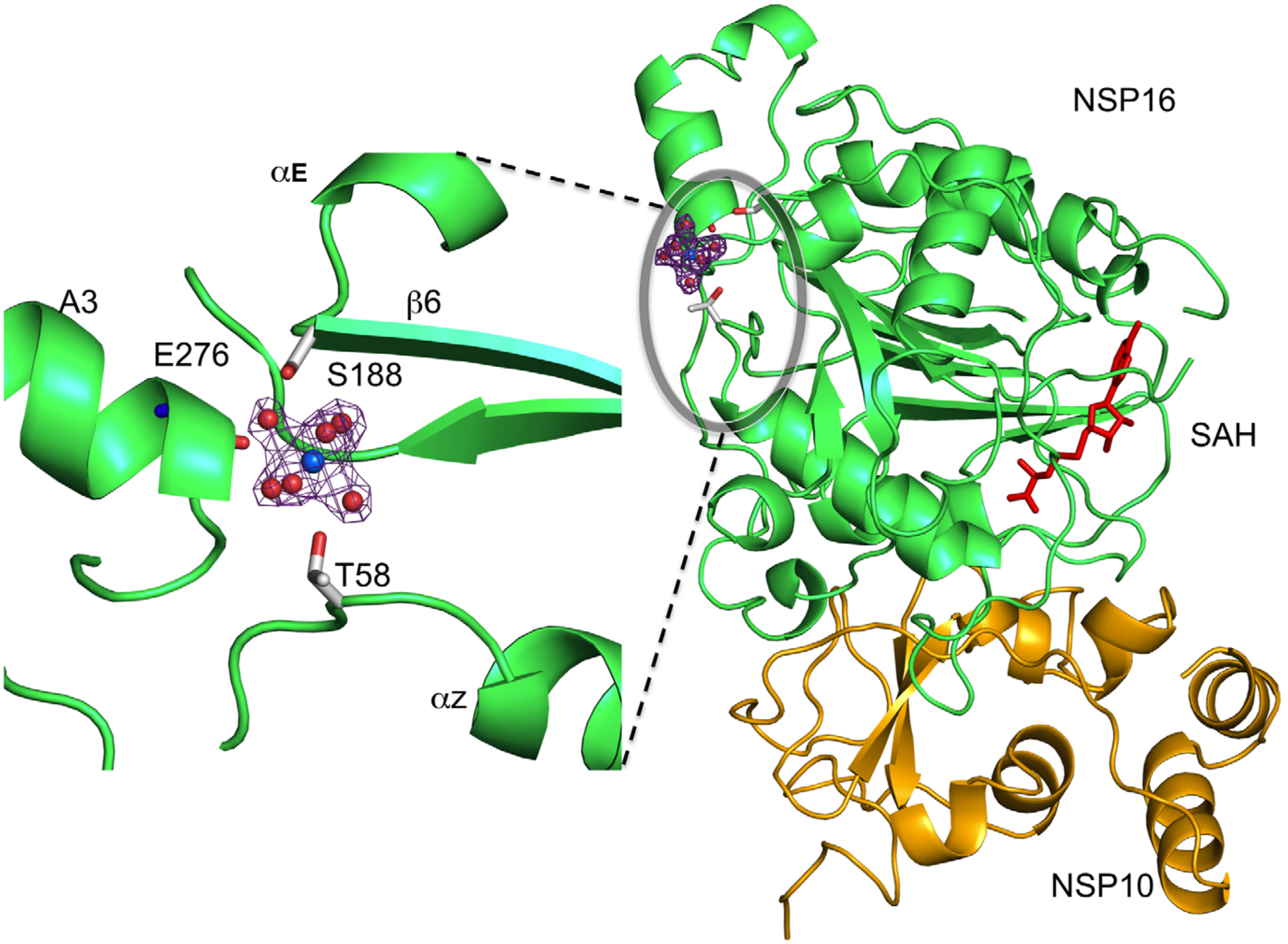


Figure 4. Position and coordination of a metal ion. Left, the putative Mg^{2+} ion (blue sphere) is shown solvated in its first atomic shell by six water molecules (red spheres). The corresponding 2Fo-Fc electron density map (contoured 1σ) with a cross shape is shown in purple. Residues of Nsp16 involved in coordination of Mg^{2+} ion via water molecules are labeled. Right, a global view of the bound metal ion in its electron density on the opposite side from the SAH molecule (red sticks). Nsp16 is shown as green ribbons, nsp10 as yellow ribbons with its two bound Zn^{2+} as grey spheres as in Fig. 1.
doi:10.1371/journal.ppat.1002059.g004

patch I which contains the K46 catalytic residue. On the other hand, patch II and III mutants tend to have more mitigated phenotypes, yielding to full-blown interaction with only about half of the expected activity. Finally, patch IV mutants were totally inactive. Using all mutants reported in Table 2, a plot (Fig. 6C) of interaction versus activity shows that the nsp10/nsp16 interaction is strictly required to obtain significant nsp16 MTase activity.

Discussion

Mechanism of 2′-O-Methyltransfer

The SARS-CoV RNA cap 2′-O-MTase is a heterodimer comprising SARS-CoV nsp10 and nsp16. When bound to nsp10, nsp16 is active as a type-0 RNA cap-dependent 2′-O-MTase, ie., active only when the cap guanine is methylated at its N7 position[16]. The nsp10/nsp16 crystal structure shows that nsp16 adopts a typical fold of the S-adenosylmethionine-dependent methyltransferase family as defined initially for the catechol O-MTase[25]. A good alignment (170°) is found between the SAH sulfur atom, a water molecule present in both SAH- and sinefungin-bound nsp16 structures, and the K46 ε-amino group (Fig. 3A and B). This geometry provides interesting hints for a catalytic mechanism, as the positions of the catalytic residues (K46, D130, K170, E203) match spatially those of the vaccinia virus VP39 2′-O-MTase[47]. At the initial stage of the reaction the 2′-hydroxyl of the capped RNA substrate would occupy the position of the water molecule. In turn, E203 and K170 decrease the pKa of the K46 ε-amino group that becomes a deprotonated general base (-NH_2) able to activate the RNA 2′-hydroxyl at neutral pH. In VP39, K175 has been identified as the general base catalyst[47] with a pKa depressed by ~ 2 pH units by the neighbouring D138 and R209 residues[48]. These findings indeed suggest a related mechanism: once K46 has activated the 2′-hydroxyl group, the 2′-oxygen would produce an *in line* attack through a SN2-like mechanism onto the electrophilic SAM methyl group. The methyl group would pass through a pentavalent intermediate with the 2′-O and sulfur at apical positions. D130 is positioned to stabilize the transient positive charge on the donated methyl atom of SAM before the sulfur recovers a neutral electric charge during SAH generation (Fig. 3B).

Role of Mg^{2+} on MTase Activity

Unlike most SAM-dependent MTases, the SARS-CoV nsp10/nsp16 enzyme requires a divalent cation, either magnesium, manganese or calcium[25]. We have found that this cation does

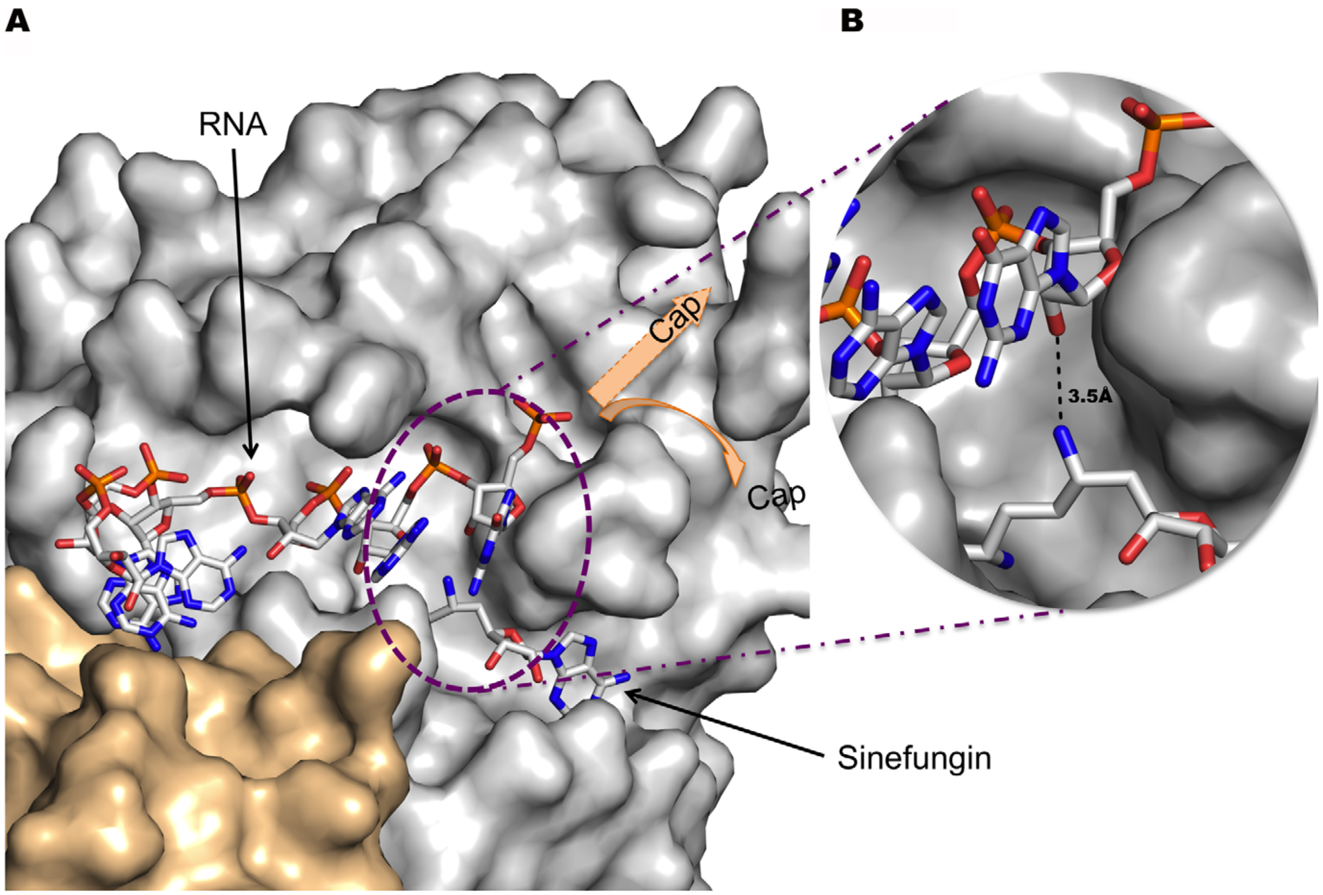


Figure 5. Stick model of RNA bound to the nsp16 RNA binding groove and Sinefungin in the methyltransferase active site. A) In this representation, Sinefungin was preferred over SAH because one of its NH_2 groups approximates the direction of a transferred CH_3 group from the SAM substrate. Carbon is white, oxygen is red, nitrogen is blue and phosphorous is orange. Nsp16 and nsp10 are rendered as a solvent-accessible surface colored grey and wheat respectively. The Sinefungin molecule defines the methyltransferase active site. Missing residues in the 135-137 loop (see text) are indicated by a shaded blue dotted box. Position of Y30 and Y132 are indicated. Y30 generated poor electron density (see « Methods ») and its aromatic ring position has been manually adjusted before generation of this image. B) Close caption of the methyltransferase active site showing distance between the NH2 of Sinefungin to the 2′-O of the ribose of the first base, thus mimicking the position of the methyl of the S-adenosylmethionine.
doi:10.1371/journal.ppat.1002059.g005

not reside in the active site. Instead, the cation is coordinated through water molecules by three residues located on three distinct structural elements. It is thus possible that one divalent cation, presumably Mg^{2+}, present in the host cell at millimolar levels, plays a structural role in holding these three nsp16 structural elements together and so regulate the enzyme activity. It is intriguing that T58A is more active than T58N or T58E that can still bind the water that chelates to the metal. Alternatively, it is possible that divalent cations such as Mg^{2+} or Ca^{2+} act as a phosphodiester charge shield to allow RNA binding in the hydrophobic binding groove[49].

MTase Activity is Regulated via Protein-Protein Interactions

The main regulation mechanism of nsp16 is through its physical association with nsp10. Nsp16 is unstable in solution, and nsp10 acts as a scaffold for nsp16, yielding a stable dimer active as an RNA cap-dependent (nucleoside-2′-O)-MTase. The complex is assembled through a ~890 Å^2 contact surface in nsp16, an area typically in the intermediate zone differentiating strongly from weakly associated dimers[50]. This finding is consistent with a Kd estimated at ~0.8 µM[18] that qualifies the nsp10/nsp16 complex as a rather weak heterodimer. The nsp10 interaction surface identified in the crystal structure was confirmed by site-directed mutagenesis and overlaps that previously identified by indirect methods[18]. Remarkably, the nsp10 surface in the nsp10/nsp16 complex is essentially identical to that of uncomplexed nsp10 crystallized alone by others[20,21](Fig. S4). It is therefore reasonable to see this heterodimer as a non-permanent species which would tolerate nsp10 or nsp16 engaging in interactions with other partners. This notion is actually in line with the involvement of nsp10 in a network of protein-protein interactions that we and others have proposed[17,19]. Donaldson *et al.*[23] have engineered mutations in nsp10 using reverse genetics. Out of eight mutations that turned out to be in the nsp10/nsp16 interface (this work), five, two and one rendered lethal, debilitated, and viable phenotypes, respectively[23]. Interestingly, the nsp10(Q65E) mutant providing a temperature-sensitive phenotype[22,23] does not map in the nsp10/nsp16 interface, confirming that nsp10 has a pleïotropic role.

Our mutagenesis analysis shows that the formation of an nsp10/nsp16 complex is a pre-requisite for MTase activity (Fig. 6C) indicating that physical association of nsp10 and nsp16 is essential to activate nsp16 2′-O-MTase activity and foster efficient virus

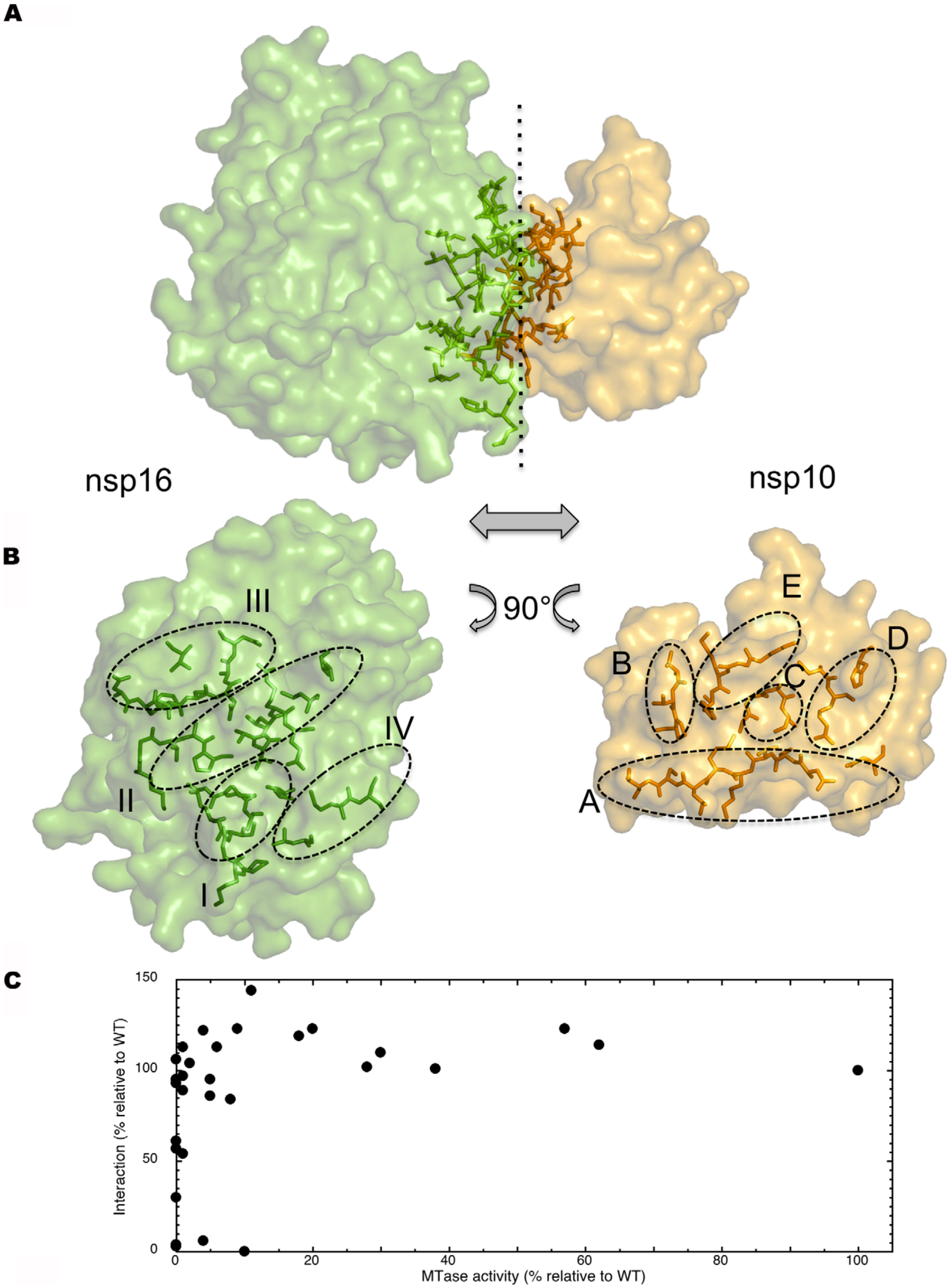

A
nsp16
nsp10
B
III
II
IV
I
90°
E
B
C
D
A
C
150
100
50
0
Interaction (% relative to WT)
0
20
40
60
80
100
MTase activity (% relative to WT)

Figure 6. Detailed definition of nsp16 and nsp10 elements involved in the interface. A) nsp16 (left) and nsp10 (right) are rendered as a solvent-accessible surface colored green and yellow, respectively. Residues involved in the interface are rendered as sticks with the same color code, in transparency into respective proteins. B) Separate representation of the nsp16 (left) and nsp10 (right) interface. Nsp16 residues defining patch I, II, III, and IV are rendered as sticks as above in transparency into nsp16. Nsp10 residues defining patch A, B, C, D, and E are rendered as sticks as above in transparency into nsp10. C) A plot of MTase activity as a function of nsp10/nsp16 complex formation using the data reported in Table 2.
doi:10.1371/journal.ppat.1002059.g006

replication. We note that most interface mutants exhibit a severe loss of their 2′-O-MTase activity, whereas the apparent association affinity is often only modestly affected. That minor changes in the interface translate into potent effects is also dramatically illustrated by the Y96F mutation, where the loss of a single hydroxyl provokes a significant change in affinity[18]. Remarkably, it is not the most active complex that was selected in nature, since the nsp10(Y96F)/nsp16 complex is both more stable and more active than the wild-type heterodimer (this work and[18]). This is yet another observation hinting at the involvement of nsp10 in protein-protein interaction networks including other partners than nsp16, such as nsp5 and nsp14[17,19]. In most other coronaviruses, the nsp10 residue at position 96 is a phenylalanine. It would be interesting to determine whether this polymorphism is relevant to the SARS-CoV pathogenicity at any (direct or indirect) level, or if compensating polymorphisms in other coronaviral nsp10 (or nsp16) restore a weaker nsp10/nsp16 association equivalent to that of the SARS-CoV pair. Since a *bona fide* viral RNA cap is key in evading the host cell innate immunity[4,51], a minimal level of 2′-O-MTase activity would be expected to be critical to virus survival.

MTase activation through dimerisation of two viral protein partners has already been reported in the case of the vaccinia virus D1/D12 N7-guanine MTase[28]. However, the activating D12 subunit does not contact the D1 subunit through a homologous surface mainly defined by canonical αA and αZ helices. Rather, the D1/D12 activation surface would be located at a 90° clockwise rotation relative to the nsp10/nsp16 interface depicted in Fig. 1A. In the case of dengue virus, the bi-functional N7-guanine and 2′-O-MTase is part of the N-terminus of the dengue NS5 protein. Based on reverse genetic data and modeling[52], the MTase domain would be associated with the Pol domain through an interface topologically similar to that of nsp10/nsp16, *i.e.*, involving mainly helices αA, αZ and strands β2 and β3 as depicted in Fig. 1A.

We have previously shown that the nsp10/nsp16 is only active as N7-guanine methylated capped RNA, implying that RNA cap methylation obeys to an ordered sequence of events where nsp14-mediated N7-guanine methylation precedes nsp10/nsp16 RNA 2′-O methylation[16]. In the absence of data regarding the RNA substrate, we built a model of RNA binding based on that of the vaccinia virus VP39 ternary complex structure. Interestingly, our model proposes that the RNA interacts only with nsp16 residues, in keeping with what was recently suggested based on RNA binding assays[14]. Although the position of the cap structure on the nsp16 surface remains to be determined, our model suggests a well-defined position for the ribose of the first transcribed nucleotide in the active site. In agreement with mutagenesis analysis, the model also suggests that the transcribed RNA 5′-end stacks between Y132 and Y30. Furthermore, this model is consistent with the observation that coronavirus MTase requires RNA substrates of at least 3 transcribed nucleotides in length[14]. It is also worth to know that a comparison of nsp16 and VP39 electrostatic surfaces reveals that the putative RNA-binding groove of nsp16 is mostly hydrophobic, whereas the VP39 RNA-binding groove is positively charged. This variation would imply a change in the nature of the RNA/protein interaction.

Binding of Sinefungin to nsp16/nsp10 Suggests Feasibility of a Drug Design Approach to Inhibit MTase Activity

Viral MTases are increasingly evaluated as potential drug design targets[34,37,53]. We have crystallized the inhibitor Sinefungin with the nsp10/nsp16 complex. Sinefungin exhibits an IC_{50} of 0.74 μM, 16-fold lower than that of SAH as reported by Bouvet *et al.*[16] using purified nsp10/nsp16. Analysis of the structure suggests a likely mechanism of action that also accounts for the observed inhibitory effect of this drug. We note that the adenine nucleobase does not fit snugly into its binding pocket, raising interest regarding structure-based drug design. Preliminary examination of eukaryotic non-viral MTase structures from main classes as defined in Martin and McMillan[25] indicates that the SAH adenine is bound tighter in any of the latter enzymes than in the nsp16 SAM-binding site, indicating a possible breach to achieve anti-coronavirus selectivity with a small molecule inhibitor of nsp16.

In conclusion, the crystal structures presented here extend our general understanding of the mechanism and regulation of viral RNA cap MTases in (+)RNA viruses, and point to both the nsp10/nsp16 interface and the substrate binding sites as putative antiviral targets.

Methods

Crystallization and Data Collection

Both nsp10 and nsp16 were expressed from the same dual expression vector pmCOX [16]. Nsp10 had a N-terminal strep-tag (WSHPQFEK), and nsp16 a N-terminal hexa-histidine tag. The purification and crystallogenesis of the nsp10/nsp16 complex was performed as described in [40]. Typical crystals of the wild-type nsp10/nsp16 appear in hanging drops after 24 h at 20°C in 0.1 M CHES pH 9, 1.52 M $MgCl_2$ hexahydrate. Crystals (a = 68.53 Å, b = 184.74 Å, c = 129.01 Å, $C222_1$) contain one nsp10/nsp16 complex per asymmetric unit, with a solvent content of 70 % and V_m of 4.17 $Å^3$/Da. Crystals of nsp10(Y96F)/nsp16 were grown in 67 mM CHES pH 8.5, 0.99 M $MgCl_2$ hexahydrate, 33 mM Tris-HCl, 8.3 % PEG 8000. Both crystallization conditions yielded crystals diffracting to 1.9 Å when exposed to synchroton radiation at the ID14-1 beamline of the European Synchrotron Radiation Facility, Grenoble, France. Crystals were cryo-cooled in the same buffer supplemented with 15 % glycerol. Crystal soaking was performed in the same buffer supplemented with 5 mM SAH or Sinefungin during 24 h.

Structure Determination and Refinement

The position of the nsp10 protein was unambiguously determined by molecular replacement using the program PHASER[54] with the nsp10 protein (2FYG), as search probe[20]. Strong peaks in both the residual and anomalous Fourier maps confirmed the presence of two Zinc ions at the expected positions within the nsp10 protein, thus giving confidence in the validity of the MR solution. Phases calculated from this partial model were combined with SAD phases from the Zn atoms using PHASER. To ameliorate the resulting low quality density map, phases were improved with PARROT[55]. An initial model, comprising both nsp10 and nsp16, was automatic built by successive use of BUCCANEER[56] and ARP/wARP[57]. The resulting model was subject to several cycles of manual rebuilding using COOT [58] and refinement with REFMAC [59]. The protein structure model could be built, except the strep and hexahistidine tags. In nsp16, density was too weak for the mobile, solvent exposed nsp16 loop 136–139, Y30 (see "Results"), and 2 and 6 residues in N- and C-terminus, respectively. Likewise, nsp10 solvent exposed 9 and 8 residues in N- and C-terminus were missing, respectively.

Overall, the chain traces are unambiguous, with clear electron density including for a single SAH residue bound to the nsp16 protein. Solvent accessible surfaces were calculated using program AREAIMOL [60] with a 1.7 Å radius sphere as the probe (Table 1) and values rounded to the nearest 5 Å^2. Conformational differences were analyzed using the DynDom server (http://www.cmp.uea.ac.uk/dyndom/main.jsp). Figures were created using PYMOL (http://www.pymol.org). The coordinates of the wild-type/SAH, mutant, and wild-type/Sinefungin structures have been deposited at the Protein Data bank under PDB codes 2XYQ, 2XYV, and 2XYR, respectively.

RNA Cap Structure Modeling

The modeling of the RNA cap structure in the nsp10/nsp16 complex structure is derived from the analysis of the structure of the vaccinia virus methyltransferase VP39 crystallized in complex with a capped RNA and a S-Adenosylhomocysteine[47] (SAH) (pdb code: 1AV6). The two structures are manually aligned using COOT[58] based on the position of SAH binding sites, as well as SAH, and Sinefungin (SFG) molecules. The RNA binding site of VP39 is only partly overlapping that of nsp16 whilst the shape of the cavity is similar; thus local adjustments necessary to accommodate the RNA molecule in its binding groove were done manually using COOT. The side chain of tyrosine 30 of nsp16initially pointed to the putative RNA binding site, preventing any *bona fide* modeling. In order to fit the RNA molecule in the cavity, an alternative conformation was sought for this side chain. The second most common conformation for the tyrosine side chain was selected. Due to the biochemistry data and surface electrostatic analysis, it is not possible to describe with certainty the final position of the cap, thus the cap was removed and replaced by arrows symbolizing possible positions. No other modification was performed on the RNA, the Sinefungin molecule or the nsp16 structure.

Plasmids

The SARS-CoV nsp10 and nsp16-coding sequences were amplified by RT-PCR from the genome of SARS-CoV Frankfurt-1 (accession number AY291315) as previously described[16]. The nsp10 and nsp16 genes (encoding residues 4231–4369, 5903–6429, and 6776–7073 of replicase pp1ab) were cloned into a Gateway modified dual-promotor expression plasmid and in the gateway pDest 14 expression vector. In this backbone, SARS CoV nsp10 can be expressed under a tet promoter and encodes a protein in fusion with a N-terminal strep tag, whereas nsp16 is expressed under a T7 promoter and encodes a protein in fusion with a N-terminal hexahistidine tag. The mutants were generated by PCR using the Quickchange site–directed mutagenesis kit (Stratagene), according to the manufacturer's instructions.

Reagents

AdoMet and cap analogs GpppA and ^{7Me}GpppA were purchased from New England BioLabs, the[^{3}H]-AdoMet was purchased from Perkin Elmer and Sinefungin (adenosylornithine) from Sigma-Aldrich.

Expression and Purification of SARS-CoV nsp10, nsp16 and nsp10/nsp16 Complex

E. coli C41 (DE3) cells (Avidis SA, France), containing the pLysS plasmid (Novagen), were transformed with nsp10 or nsp16 cloned in pDest14, or nsp10/nsp16 cloned in pmCox, and grown in 2YT medium supplemented with appropriate antibiotics. The expression of strep-tagged nsp10 or 6His-tagged nsp16 mutants was induced (DO_{600} = 0.6) by adding 50 μM IPTG, and the expression of the nsp10/nsp16 complex by adding 50 μM IPTG and 200 μg/L of anhydrotetracycline. After an incubation for a 16 h at 24°C, the cell were pellets, frozen and resuspended in lysis buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 5 mM $MgSO_4$, 5 mM β-mercaptoethanol (only for nsp10) supplemented with 1 mM PMSF, 40 mM imidazole, 10 μg/ml DNase I, and 0.5% Triton X-100. After sonication and clarification, proteins were purified either by IMAC (HisPurTM Cobalt Resin; Thermo Scientific) chromatography[16] (nsp10 mutants and nsp16 mutants), and the nsp10/nsp16 complex was purified by using Strep-Tactin sepharose (IBA Biotagnology) as previously described[16]. All purified proteins were analyzed by SDS-PAGE. The binding of wild-type nsp10 to mutant nsp16, and that of mutant nsp10 to wild-type nsp16 was quantified using ImageJ as described[16].

Radioactive Methyltransferase Assay

MTase activity assays were performed in 40 mM Tris-HCl, pH 8.0, 5 mM DTT, 1 mM $MgCl_2$, 2 μM $^{7Me}GpppAC_5$ or $GpppAC_5$, 10 μM AdoMet, and 0.03 μCi/μl [^{3}H]AdoMet (GE Healthcare). Short capped RNAs ($^{7Me}GpppAC_5$, $GpppAC_5$, were synthesized *in vitro* using bacteriophage T7 DNA primase and were purified by high-performance liquid chromatography (HPLC) as previously described[61]. In the standard assay, nsp10 and nsp16 were added at final concentrations of 600 nM, and 200 nM, respectively, and the amount of $^3H\text{-}CH_3$ transferred onto $^{7Me}GpppAC_5$ substrates was determined by filter binding assay as previously described[16].

Supporting Information

Figure S1 Sequence alignment and amino acid conservation in coronavirus nsp16. The alignment of coronavirus nsp16 sequences was generated with Muscle program (http://www.ebi.ac.uk/Tools/msa/muscle/), and the resulting alignment converted using the ESPript program, (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi). Residues that are conserved in all or >70% sequences are boxed in red and yellow, respectively. National Center for Biotechnology Information (NCBI) accession numbers for replicase polyprotein sequences that include nsp16 are as follows: Severe acute respiratory syndrome virus SARS-CoV Frankfurt isolate (SARF), Severe acute respiratory syndrome virus SARS-CoV Tor2 isolate (SART), NP_828873.2; Turkey coronavirus (Turk), YP_001941189; Infectious Bronchitis Virus (IBV), NP_066134; Feline Coronavirus (Feli), YP_239426;Porcine Transmissible Gastroenteritis Coronavirus (PTGC), P18457; Transmissible Gastroenteritis coronavirus (TGEV), NP_840013; Porcine epidemic diarrhea virus CV777 (PEDV), NP_839969; Bat coronavirus 512/2005 (BATC), YP_001351683; Human coronavirus NL63 (NL63), AAS58176; Human coronavirus 229E (229E), NP_073549; Murine hepatitis virus strain JHM (MurJ), YP_209243; Mouse Hepatitis virus strain A59 (MA59), NP_740613; Human coronavirus HKU1 genotype A (HKU1), YP_460023; Human enteric coronavirus 4408 (4408), ACJ35483; Bovine coronavirus (BCoV), NP_742142; Human coronavirus OC43 (OC43), AAT84359. Numbering was made using SARF as a reference.
(TIF)

Figure S2 Binding determinants in the SAM/SAH/Sinfungin binding site. A) LIGPLOT diagram (www.ebi.ac.uk/thornton-srv/software/LIGPLOT/) of the SAH ligand molecule interacting with the nsp16 binding site. Ligand bonds are in purple, neighbor residue (non-ligand) bonds are in light brown, hydrogen bonds are green dashed lines. Ligand atoms surrounded by a yellow circle are highly accessible. Non-ligand residues in hydrophobic contacts with the ligand are presented by red semi-circles with radiating spokes. B) Superimposition of SAH and Sinefungin molecules. The nsp16 residues are green sticks, the SAH is represented in red sticks, and Sinefungin and water molecules are in sticks and spheres as in Fig. 3, respectively.
(TIF)

Figure S3 Primary and secondary structure elements of nsp16. Helices and sheets are colored according to rainbow colors from N- to C-terminus as in Fig. 1. Patches I to IV are boxed in grey, labeled above the sequences, and correspond to residues represented in Figs. 6A and B.
(TIF)

Figure S4 Nsp10 interface comparison between the nsp10/nsp16 complex and nsp10 homo multimer. A) For the corresponding nsp10/nsp16 interface nsp10 (cyan) interacts with 2 monomers (green and orange). Left: Cartoon representation of 3 nsp10 monomer part of the dodecamer structure (PDB: 2G9T). Right: Cartoon representation of the nsp10(cyan)-nsp16(dark red) complex. Nsp10 molecules in cyan are in the same orientation. B) Nsp10 sequence is presented with above the corresponding secondary structure elements. Below dark red colored dots indicate residues involved in the interaction between nsp10 and nsp16 while green and orange dots indicate residues involved in the homomultimer in the dodecamer complex.
(TIF)

Table S1 Effect of alanine mutations in nsp10 interface residues. Residues were identified using PISA (http://www.ebi.ac.uk/msd-srv/prot_int/pi_link.html). Bold on grey background: strictly conserved residues amongst coronaviruses; Bold on white background: conserved a.a. (>70%) amongst coronaviruses (see Fig. S1). The % of Bioluminescence Resonance energy Transfer (BRET) signal was previously reported[18]. The interaction of each nsp10 or nsp16 mutant was determined using strep-tactin pull-down experiments of strep-tagged nsp10 co-expressed with nsp16 followed by SDS-PAGE analysis, and quantitation (see Methods). The interaction of wild-type nsp10 with wild-type nsp16 was normalized to 100%. The % of MTase activity was determined using filter binding assays (see Methods) relative to wild-type.
(DOC)

Table S2 Effect of alanine mutations in nsp16 interface residues. Residues were identified using PISA (http://www.ebi.ac.uk/msd-srv/prot_int/pi_link.html). Same legend as Table S1 except that no BRET experiments were performed.
(DOC)

Acknowledgments

We thank Barbara Selisko for helpful discussion, comments and support along this work, and Joëlle Boretto-Soler for help in mutagenesis and protein purification. We are indebted to Karen Dalle, Violaine Lantez for excellent technical assistance and General Hugues Toulou for his support.

Author Contributions

Conceived and designed the experiments: E. Decroly, B. Canard. Performed the experiments: E. Decroly, C. Debarnot, F. Ferron, M. Bouvet, B. Coutard, I. Imbert, L. Gluais, N. Papageorgiou, B. Canard. Analyzed the data: E. Decroly, F. Ferron, I. Imbert, N. Papageorgiou, A. Sharff, G. Bricogne, M. Ortiz-Lombardia, J. Lescar, B. Canard. Wrote the paper: E. Decroly, B. Canard.

References

1. Gu M, Lima CD (2005) Processing the message: structural insights into capping and decapping mRNA. Curr Opin Struct Biol 15: 99–106.
2. Shuman S (2001) Structure, mechanism, and evolution of the mRNA capping apparatus. Prog Nucleic Acid Res Mol Biol 66: 1–40.
3. Shuman S (2002) What messenger RNA capping tells us about eukaryotic evolution. Nat Rev Mol Cell Biol 3: 619–625.
4. Yoneyama M, Fujita T (2010) Recognition of viral nucleic acids in innate immunity. Rev Med Virol 20: 4–22.
5. Shuman S (2001) The mRNA capping apparatus as drug target and guide to eukaryotic phylogeny. Cold Spring Harb Symp Quant Biol 66: 301–312.
6. Ahola T, Kaariainen L (1995) Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. Proc Natl Acad Sci U S A 92: 507–511.
7. Ogino T, Banerjee AK (2007) Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. Mol Cell 25: 85–97.
8. Plotch SJ, Bouloy M, Ulmanen I, Krug RM (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell 23: 847–858.
9. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, et al. (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300: 1394–1399.
10. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ (2006) Nidovirales: evolving the largest RNA virus genome. Virus Res 117: 17–37.
11. Lai MM, Patton CD, Stohlman SA (1982) Further characterization of mRNA's of mouse hepatitis virus: presence of common 5′-end nucleotides. J Virol 41: 557–565.
12. Lai MM, Stohlman SA (1981) Comparative analysis of RNA genomes of mouse hepatitis viruses. J Virol 38: 661–670.
13. van Vliet AL, Smits SL, Rottier PJ, de Groot RJ (2002) Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. EMBO J 21: 6571–6580.
14. Decroly E, Imbert I, Coutard B, Bouvet M, Selisko B, et al. (2008) Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2′O)-methyltransferase activity. J Virol 82: 8071–8084.
15. Chen Y, Cai H, Pan J, Xiang N, Tien P, et al. (2009) Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7-methyltransferase. Proc Natl Acad Sci U S A 106: 3484–3489.
16. Bouvet M, Debarnot C, Imbert I, Selisko B, Snijder EJ, et al. (2010) In vitro reconstitution of SARS-coronavirus mRNA cap methylation. PLoS Pathog 6: e1000863.
17. Imbert I, Snijder EJ, Dimitrova M, Guillemot JC, et al. (2008) The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. Virus Res 133: 136–148.
18. Lugari A, Betzi S, Decroly E, Bonnaud E, Hermant A, et al. (2010) Molecular mapping of the RNA Cap 2′-O-methyltransferase activation interface between SARS coronavirus nsp10 and nsp16. J Biol Chem 285: 33230–33241.
19. Pan J, Peng X, Gao Y, Li Z, Lu X, et al. (2008) Genome-wide analysis of protein-protein interactions and involvement of viral proteins in SARS-CoV replication. PLoS One 3: e3299.
20. Joseph JS, Saikatendu KS, Subramanian V, Neuman BW, Brooun A, et al. (2006) Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. J Virol 80: 7894–7901.
21. Su D, Lou Z, Sun F, Zhai Y, Yang H, et al. (2006) Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10. J Virol 80: 7902–7908.
22. Sawicki SG, Sawicki DL, Younker D, Meyer Y, Thiel V, et al. (2005) Functional and genetic analysis of coronavirus replicase-transcriptase proteins. PLoS Pathog 1: e39.
23. Donaldson EF, Sims AC, Graham RL, Denison MR, Baric RS (2007) Murine hepatitis virus replicase protein nsp10 is a critical regulator of viral RNA synthesis. J Virol 81: 6356–6368.
24. Donaldson EF, Graham RL, Sims AC, Denison MR, Baric RS (2007) Analysis of murine hepatitis virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in polyprotein processing. J Virol 81: 7086–7098.
25. Martin JL, McMillan FM (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr Opin Struct Biol 12: 783–793.
26. Vidgren J, Svensson LA, Liljas A (1994) Crystal structure of catechol O-methyltransferase. Nature 368: 354–358.
27. Hodel AE, Gershon PD, Shi X, Quiocho FA (1996) The 1.85 A structure of vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends. Cell 85: 247–256.

28. De la Pena M, Kyrieleis OJ, Cusack S (2007) Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap N7 methyl-transferase. EMBO J 26: 4913–4925.
29. Reinisch KM, Nibert ML, Harrison SC (2000) Structure of the reovirus core at 3.6 A resolution. Nature 404: 960–967.
30. Sutton G, Grimes JM, Stuart DI, Roy P (2007) Bluetongue virus VP4 is an RNA-capping assembly line. Nat Struct Mol Biol 14: 449–451.
31. Egloff MP, Benarroch D, Selisko B, Romette JL, Canard B (2002) An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. EMBO J 21: 2757–2768.
32. Egloff MP, Decroly E, Malet H, Selisko B, Benarroch D, et al. (2007) Structural and functional analysis of methylation and 5′-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. J Mol Biol 372: 723–736.
33. Ray D, Shah A, Tilgner M, Guo Y, Zhao Y, et al. (2006) West Nile virus 5′-cap structure is formed by sequential guanine N-7 and ribose 2′-O methylations by nonstructural protein 5. J Virol 80: 8362–8370.
34. Bollati M, Alvarez K, Assenberg R, Baronti C, Canard B, et al. (2010) Structure and functionality in flavivirus NS-proteins: perspectives for drug design. Antiviral Res 87: 125–148.
35. Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, et al. (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol 331: 991–1004.
36. Balzarini J, De Clercq E, Serafinowski P, Dorland E, Harrap KR (1992) Synthesis and antiviral activity of some new S-adenosyl-L-homocysteine derivatives. J Med Chem 35: 4576–4583.
37. Dong H, Zhang B, Shi PY (2008) Flavivirus methyltransferase: a novel antiviral target. Antiviral Res 80: 1–10.
38. Pugh CS, Borchardt RT, Stone HO (1977) Inhibition of Newcastle disease virion messenger RNA (guanine-7-)-methyltransferase by analogues of S-adenosylhomocysteine. Biochemistry 16: 3928–3932.
39. Pugh CS, Borchardt RT, Stone HO (1978) Sinefungin, a potent inhibitor of virion mRNA(guanine-7-)-methyltransferase, mRNA(nucleoside-2′-)-methyl-transferase, and viral multiplication. J Biol Chem 253: 4075–4077.
40. Debarnot C, Imbert I, Ferron F, Gluais L, Varlet I, et al. (2011) Crystallization and diffraction analysis of the SARS coronavirus nsp10/nsp16 complex. Acta Crystallogr Sect F Struct Biol Cryst Commun 67: 404–8.
41. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. Bioinformatics 16: 566–567.
42. Dudev T, Lim C (2003) Principles governing Mg, Ca, and Zn binding and selectivity in proteins. Chem Rev 103: 773–788.
43. Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK (1997) Cocrystal structure of the messenger RNA 5′ cap-binding protein (eIF4E) bound to 7-methyl-GDP. Cell 89: 951–961.
44. Mazza C, Segref A, Mattaj IW, Cusack S (2002) Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. EMBO J 21: 5548–5557.
45. Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, et al. (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. Nat Struct Mol Biol 15: 500–506.
46. Quiocho FA, Hu G, Gershon PD (2000) Structural basis of mRNA cap recognition by proteins. Curr Opin Struct Biol 10: 78–86.
47. Hodel AE, Gershon PD, Quiocho FA (1998) Structural basis for sequence-nonspecific recognition of 5′-capped mRNA by a cap-modifying enzyme. Mol Cell 1: 443–447.
48. Li C, Xia Y, Gao X, Gershon PD (2004) Mechanism of RNA 2′-O-methylation: evidence that the catalytic lysine acts to steer rather than deprotonate the target nucleophile. Biochemistry 43: 5680–5687.
49. Sigel RK, Sigel H (2010) A stability concept for metal ion coordination to single-stranded nucleic acids and affinities of individual sites. Acc Chem Res 43: 974–984.
50. Dey S, Pal A, Chakrabarti P, Janin J (2010) The subunit interfaces of weakly associated homodimeric proteins. J Mol Biol 398: 146–160.
51. Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, et al. (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nat Immunol 5: 730–737.
52. Malet H, Egloff MP, Selisko B, Butcher RE, Wright PJ, et al. (2007) Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. J Biol Chem 282: 10678–10689.
53. Dong H, Liu L, Zou G, Zhao Y, Li Z, et al. (2010) Structural and functional analyses of a conserved hydrophobic pocket of flavivirus methyltransferase. J Biol Chem 285: 32586–32595.
54. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. Journal of applied crystallography 40: 658–674.
55. Cowtan K (2010) Recent developments in classical density modification. Acta Crystallogr D Biol Crystallogr 66: 470–478.
56. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr 62: 1002–1011.
57. Cohen SX, Morris RJ, Fernandez FJ, Ben Jelloul M, Kakaris M, et al. (2004) Towards complete validated models in the next generation of ARP/wARP. Acta Crystallogr D Biol Crystallogr 60: 2222–2229.
58. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta crystallographica Section D, Biological crystallography 60: 2126–2132.
59. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50: 760–763.
60. Saff EB, Kuijlaars ABJ (1997) Distributing Many Points on a Sphere. Math Intell 19: 5–11.
61. Peyrane F, Selisko B, Decroly E, Vasseur JJ, Benarroch D, et al. (2007) High-yield production of short GpppA- and 7MeGpppA-capped RNAs and HPLC-monitoring of methyltransfer reactions at the guanine-N7 and adenosine-2′O positions. Nucleic Acids Res 35: e26.

doi:10.1016/j.jmb.2007.07.005 *J. Mol. Biol.* (2007) **372**, 723–736

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Structural and Functional Analysis of Methylation and 5′-RNA Sequence Requirements of Short Capped RNAs by the Methyltransferase Domain of Dengue Virus NS5

Marie-Pierre Egloff, Etienne Decroly, Hélène Malet, Barbara Selisko Delphine Benarroch, François Ferron and Bruno Canard*

Architecture et Fonction des Macromolécules Biologiques CNRS and Universités d'Aix-Marseille I et II UMR 6098, ESIL Case 925 13288 Marseille, France

**Corresponding author*

The N-terminal 33 kDa domain of non-structural protein 5 (NS5) of dengue virus (DV), named NS5MTase$_{DV}$, is involved in two of four steps required for the formation of the viral mRNA cap ^{7Me}GpppA$_{2'OMe}$, the guanine-N7 and the adenosine-2′O methylation. Its *S*-adenosyl-L-methionine (AdoMet) dependent 2′O-methyltransferase (MTase) activity has been shown on capped $^{7Me\pm}$GpppAC$_n$ RNAs. Here we report structural and binding studies using cap analogues and capped RNAs. We have solved five crystal structures at 1.8 Å to 2.8 Å resolution of NS5MTase$_{DV}$ in complex with cap analogues and the co-product of methylation *S*-adenosyl-L-homocysteine (AdoHcy). The cap analogues can adopt several conformations. The guanosine moiety of all cap analogues occupies a GTP–binding site identified earlier, indicating that GTP and cap share the same binding site. Accordingly, we show that binding of ^{7Me}GpppAC$_4$ and ^{7Me}GpppAC$_5$ RNAs is inhibited in the presence of GTP, ^{7Me}GTP and ^{7Me}GpppA but not by ATP. This particular position of the cap is in accordance with the 2′*O*-methylation step. A model was generated of a ternary 2′*O*-methylation complex of NS5MTase$_{DV}$, ^{7Me}GpppA and AdoMet. RNA-binding increased when $^{7Me\pm}$GpppAGC$_{n-1}$ starting with the consensus sequence GpppAG, was used instead of $^{7Me\pm}$GpppAC$_n$. In the NS5MTase$_{DV}$–GpppA complex the cap analogue adopts a folded, stacked conformation uniquely possible when adenine is the first transcribed nucleotide at the 5′ end of nascent RNA, as it is the case in all flaviviruses. This conformation cannot be a functional intermediate of methylation, since both the guanine-N7 and adenosine-2′O positions are too far away from AdoMet. We hypothesize that this conformation mimics the reaction product of a yet-to-be-demonstrated guanylyltransferase activity. A putative *Flavivirus* RNA capping pathway is proposed combining the different steps where the NS5MTase domain is involved.

Keywords: methyltransferase; capping; flavivirus; dengue; guanylyltransferase

Introduction

Despite a wide diversity in genome organization and replication mechanism, many eukaryotic cellular and viral RNAs are modified by addition of a cap structure consisting of a guanine connected through an 5′–5′ triphosphate bridge to the first transcribed nucleoside and methylated at the N7 position. This cap 0 structure (^{7Me}GpppN...) is often converted in a cap 1 structure (^{7Me}GpppN$_{2'OMe}$...) by methylation of the 2′-O position of the first nucleotide. The main role of the cap structure is to protect mRNA from degradation by 5′ exoribonucleases and to enhance

Present addresses: D. Benarroch, Memorial Sloan-Kettering Cancer Center, Department of Molecular Biology 1275 York Avenue, Box 73 New York, NY 10021, USA; F. Ferron, Department of Physiology, University of Pennsylvania School of Medicine, A501 Richards Building, 3700 Hamilton Walk, Philadelphia, PA 19104-6085, USA.

Abbreviations used: DV, dengue virus; AdoMet, *S*-adenosyl-L-methionine; AdoHcy, *S*-adenosyl-L-homocysteine; RTPase, RNA triphosphatase; GTase, guanylyltransferase; 2′OMTase, nucleoside-2′-O-methyltransferase; N7MTase, guanine-N7-methyltransferase; NS, non-structural protein.

E-mail address of the corresponding author: bruno.canard@afmb.univ-mrs.fr

initiation of mRNA translation.[1–3] The interest in viral capping is reinforced by the fact that inhibiting the cap formation may prevent viral replication.[4–7] Consequently, structural and biochemical work deciphering viral mRNA capping mechanisms may add some other names than that of proteases and polymerases to the enzyme target list for antiviral drug research.

In eukaryotic cells, the cap is added co-transcriptionally in the nucleus on nascent transcripts and is completed by three sequential enzymatic activities[1,3]: (i) an RNA triphosphatase (RTPase) removes the 5′ γ-phosphate group of mRNA; (ii) a guanylyltransferase (GTase), or capping enzyme, catalyzes the transfer of GMP to the remaining 5′-diphosphate end of mRNA; and (iii) a *S*-adenosyl-L-methionine (AdoMet)-dependent (guanine-N7)-methyltransferase (N7MTase) methylates the cap at the N7 position. Whereas lower eukaryotes including yeast contain a cap 0 structure, in higher eukaryotes a nuclear AdoMet-dependent (nucleoside-2′-O-)-methyltransferase (2′OMTase) completes the cap 1 structure. In the case of viruses, which replicate in the cytoplasm and code for their own RNA capping machinery, cap formation may follow the sequential three or four-step strategy of eukaryotic mRNA cap formation. Nonetheless, some of them have acquired specific capping strategies, as shown for alphaviruses that methylate GTP prior to the transfer of ^{7Me}GMP to the 5′-diphosphate end of the RNA.[8] Another example is the negative-sense (−) single-stranded (ss) RNA vesicular stomatitis virus, which transfers GDP rather than GMP onto the 5′-monophosphate end of RNA.[9] In addition to viruses replicating in the cytoplasm, some replicate in the nucleus and acquire their RNA cap by hijacking the host cell capping apparatus (i.e. positive-sense (+) ssRNA retroviruses such as HIV[10]), or by stealing a cap structure from cellular mRNAs in a process called cap snatching (i.e. negative strand (-) ssRNA viruses such as influenza virus[11]).

Despite the cap structures being conserved in many organisms, there is a large diversity in the molecular organization of the RNA capping machinery. In metazoans and plants the RTPase and GTase reactions are performed by a single two-domain protein, while the MTase activities reside in two separate single-domain enzymes.[1] In yeast, three separate proteins are each responsible for one of the three catalytic activities. For viral capping machineries a variety of combinations have been described. In the case of the DNA vaccinia virus, one protein bears RTPase, GTase and N7MTase activities on separate domains and another protein bears the 2′OMTase activity.[1] For double-stranded (ds) RNA orthoreovirus, the RTPase activity resides in one domain of either protein λ1[12] or μ2,[13] and the other three activities in three separate domains of protein λ2.[14] For (+) ssRNA alphaviruses the RTPase activity is found on one domain of non-structural protein 2 (nsp2), while the GTase and N7MTase are located on separate domains of protein nsp1.[15]

The (+) ssRNA genome of viruses of the genus *Flavivirus* bears a cap 1 structure ^{7Me}GpppA$_{2'OMe}$G where the first two nucleotides are strictly conserved.[16] There are over 70 flaviviruses, including important human pathogens, i.e. dengue, yellow fever and West Nile virus. They seem to code for their own capping machinery although it has not yet been entirely established. The multifunctional non-structural protein NS3 was shown to carry RTPase activity within the C-terminal helicase domain.[17–19] The GTase activity has not been identified. The 2′OMTase activity was first demonstrated for the 33 kDa N-terminal domain of dengue virus (DV) protein NS5 (NS5MTase$_{DV}$) using small capped RNA substrates $^{7Me\pm}$GpppAC$_n$ (meaning both methylated and non-methylated at guanine-N7).[20,21] Recent works demonstrated that the same MTase domain of dengue virus, and also of the cognate West Nile virus (WNV) NS5 and yellow fever virus (YFV), bears both the N7 and the 2′O-MTase activities when longer capped RNAs were used with the 5′ sequence of the WNV genome.[7,22,23] Thus, the *Flavivirus* MTase domain seems to have an active center that is able to conduct a methyl transfer reaction onto two distinct acceptor positions, which are rather different in their chemical characteristics and conformational context. The only other example, which has been reported recently, is the MTase domain of the large protein of the (−) ssRNA vesicular stomatitis virus.[24] To accomplish this task the mRNA cap substrate has to be accommodated in two very different positions: one with the guanine-N7 position and the other with the nucleoside-2′O position facing the AdoMet methyl group. Thus, it has to be repositioned between the two methylation steps. The authors also showed that the N7-methyltransfer precedes the 2′ O-methyltransfer.[7,22]

The crystal structure of NS5MTase$_{DV}$ was solved with an *S*-adenosyl-L-homocysteine (AdoHcy), the co-product of the methyltransfer, in the AdoMet-binding site.[20] When a non-hydrolyzable GTP analogue was soaked into the crystals, it did not fix in the active site but went exclusively to a second binding pocket at 12 Å from the AdoMet binding site (complex structure deposited under PDB code 2P1D).[20] This site binds selectively GTP and was interpreted as a site that accommodates the cap during the methyltransfer to the nucleoside-2′O position of nascent RNA.

In order to further understand the molecular mechanism of RNA cap binding and methylation by NS5MTase$_{DV}$, we conducted binding and structural studies using cap-analogues and small uncapped or capped RNAs. Here we report the crystal structures of complexes of NS5MTase$_{DV}$ with five different cap analogues. NS5MTase$_{DV}$ accommodated these molecules in three different conformations, two of which are proposed to have relevance in the *Flavivirus* RNA capping pathway. The binding assays with specific 2′O-methylation substrates $^{7Me\pm}$GpppAC$_n$ showed that NS5MTase$_{DV}$ binds exclusively the capped form and binding increases with chain length. Competition experiments with GTP analogues demonstrate that the GTP-binding site of NS5MTase$_{DV}$ serves effectively as a RNA-cap binding site. Binding can be further increased when the RNA sequence is changed to $^{7Me\pm}$GpppAGC$_{n\text{-}1}$. Finally, we present an hypothesis that combines different steps of *Flavivirus* RNA cap formation where the NS5MTase domain is involved.

Results

Structural analysis of NS5MTase$_{DV}$ complexes with cap analogues

In order to get insight into structural requirements of cap and RNA recognition by NS5MTase$_{DV}$, we soaked protein crystals in solutions of different cap analogue molecules (GpppA, ^{7Me}GpppA, GpppG, ^{7Me}GpppG, ^{7Me}GpppG$_{2'OMe}$). Additionally, small capped RNAs ^{7Me}GpppAC$_3$ and ^{7Me}GpppAC$_5$ were produced using DNA-dependent RNA primase of bacteriophage T7 (T7 DNA primase) and purified using an optimized protocol.[21] Subsequently we used these small capped RNAs for soaking experiments. Unfortunately, they either destroyed the crystals or the soaked crystals did not diffract. Co-crystallization assays were not successful either. In contrast, we could solve the crystal structures of all cap analogue-NS5MTase$_{DV}$ complexes using the co-ordinates of NS5MTase$_{DV}$ (PDB code 1L9K) as a starting model. No conformational change of the protein was observed upon cap analogue binding. The rmsd value on C$^{\alpha}$ atoms 7 to 264 between the different structures is between 0.16 Å and 0.5 Å. After the first round of refinement, two distinct strong residual (F_o–F_c) electron densities were revealed in each complex, and were assigned to one molecule of AdoHcy and one molecule of the cap analogue (Figure 1(a)). In one of the five structures, that of NS5MTase$_{DV}$ in complex with ^{7Me}GpppG, the AdoHcy was absent, which did not cause any significant change of the protein conformation in general or the AdoMet binding site in particular. In all cases, there is no ambiguity about the identification of the nucleosides (G, ^{7Me}G, A or G$_{2'OMe}$). A guanosine analogue is always bound at the GTP-binding site, which has been proposed to accommodate the cap of nascent RNA upon 2′O-methylation.[20] The electron density is very well defined for the ligands in the GTP-binding site. As shown in Figure 1(b), this site and the AdoHcy binding site are connected by a positively charged surface groove, which continues below AdoHcy. This continuation corresponds topologically to the RNA-binding region of a complex between the vaccinia virus mRNA cap 2′OMTase VP39[25] and a capped RNA substrate. It may thus represent the NS5MTase$_{DV}$ RNA binding groove.

A common guanosine–binding mode in the GTP-binding site

The positioning of the guanosine of GpppN and ^{7Me}GpppN in the GTP-binding site is similar in all complexes. There are a number of common features with the GTP-binding mode observed in the complex of NS5MTase$_{DV}$ with the GTP analogue,[20] consistent with the competitive binding of these cap analogues with GTP.[20] Figure 1(c) shows as an example the ^{7Me}G-moiety of ^{7Me}GpppA. The guanine base stacks against the aromatic ring of residue F25. The ribose adopts a 3′-endo configuration. Its 2′-hydroxyl group interacts *via* hydrogen bonding with the amino group of the K14 side-chain and the oxygen of the N18 side-chain. The 3′-hydroxyl group is hydrogen-bonded to the main-chain carbonyl group of S151 and to the side-chain amino group of K14. As for GTP, specificity for guanine is ensured by hydrogen bonds between main-chain carbonyl groups of residues L17, N18 and L20 and the 2-amino-group of the nucleobase. Apart from aromatic stacking, these hydrogen bonds are the only contacts established by the base moiety with the protein, whether or not the base is methylated at its N7-position.

Three different conformations for the second nucleotide

The five crystal structures presented here revealed that the protein can accommodate a cap analogue in at least three different manners. Figure 2 represents a summary of the observed binding modes.

Four cap analogues adopt two different stacked conformations (referred to as S1 and S2), both displaying an overall hairpin-like shape in which the two nucleobases stack against each other. The S1 conformation is found in the complexes between NS5MTase$_{DV}$ and N7-methylated cap analogues, i.e. ^{7Me}GpppA, ^{7Me}GpppG, and ^{7Me}GpppG$_{2'OMe}$ (see Figure 2(b)–(d) and Figure 3(a)). In these crystal structures, the ^{7Me}G is sandwiched between F25 and the second base moiety of the cap analogue. The aromatic side-chain of F25 and the two base moieties of the cap analogue molecules are arranged in a nearly perfect parallel alignment, with an average interplanar spacing of 3.4 Å between F25 and ^{7Me}G, and of 3.70 Å between ^{7Me}G and the second base moiety. The A face of the guanine interacts with the B face of the second nucleotide. The triphosphate linker adopts a U shape; the first phosphate is hydrogen-bonded to the amino group of the K29 side-chain (as observed for the GTP analogue), the second interacts with the O$^{\gamma}$ atom of the S150 side-chain and the third is free of any hydrogen bond, as is the ribose of the second nucleotide of the cap. This ribose adopts a C2′-endo form with its 3′-hydroxyl position pointing towards the protein loop (P150 to S152) connecting β-strand 4 to α-helix D. The 3′-hydroxyl position is thus blocked so that no longer RNA chain could be accommodated. The adenine of ^{7Me}GpppA does not establish

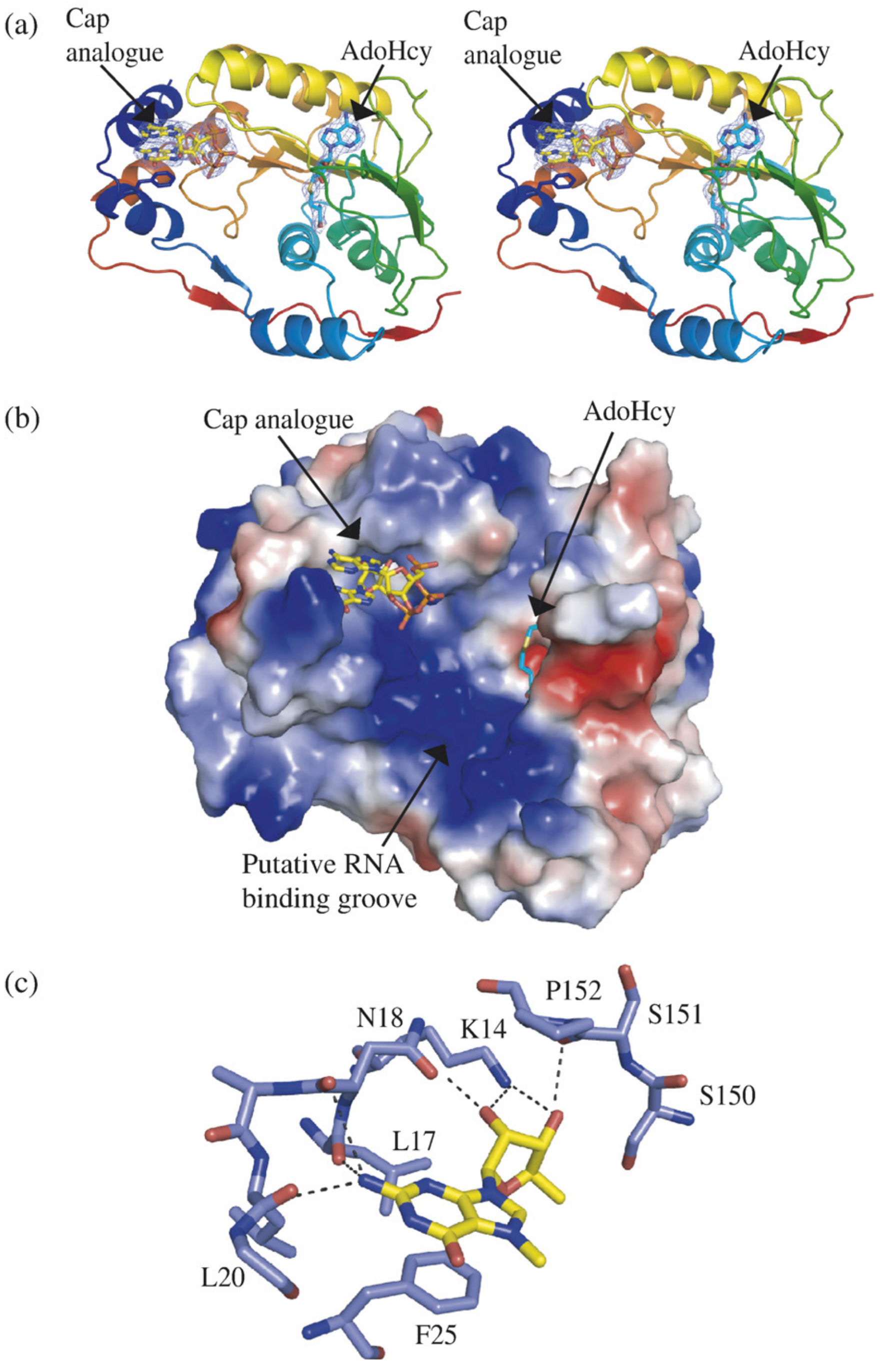


Figure 1. Ligand binding to NS5 MTase$_{DV}$. (a) A ribbon representation in rainbow color of NS5MTase$_{DV}$ in complex with AdoHcy (AdoHcy colored with carbon in cyan, nitrogen in blue, oxygen in red and sulfur in orange) and GpppA (colored with carbon in yellow, nitrogen in blue, oxygen in red and phosphorus in orange). The F_o–F_c electron density map, contoured at 3σ, was calculated at 2.2 Å resolution from a model in which the ligands were omitted. (b) Electrostatic surface representation of NS5MTase$_{DV}$ in complex with AdoHcy and GpppA. Both ligands are colored in as in (a). The positively charged (blue) surface region connecting the cap analogue binding site and the AdoHcy and continuing towards the bottom might be the RNA binding groove. (c) Coordination of the moiety of ^{7Me}GpppA in the GTP-binding pocket (colored as GpppA in (a)). Amino acids of NS5MTase$_{DV}$ that interact with ^{7Me}G are numbered and colored in slate, blue and red for C, N and O, respectively. Hydrogen bonds between the protein and the ^{7Me}G are represented by black dotted lines. F25 engages into a stacking interaction with guanine.

any interaction with the protein, whereas the N2-position of the second guanine in ^{7Me}GpppG and ^{7Me}GpppG$_{2'OMe}$ analogues interacts with both the main-chain and the side-chain oxygen atoms of N18.

The S2 conformation is adopted by cap analogue GpppA (Figures 2(e) and 3(b)). The overall hairpin-like conformation and the interplanar distances between the two stacks (the first one between the

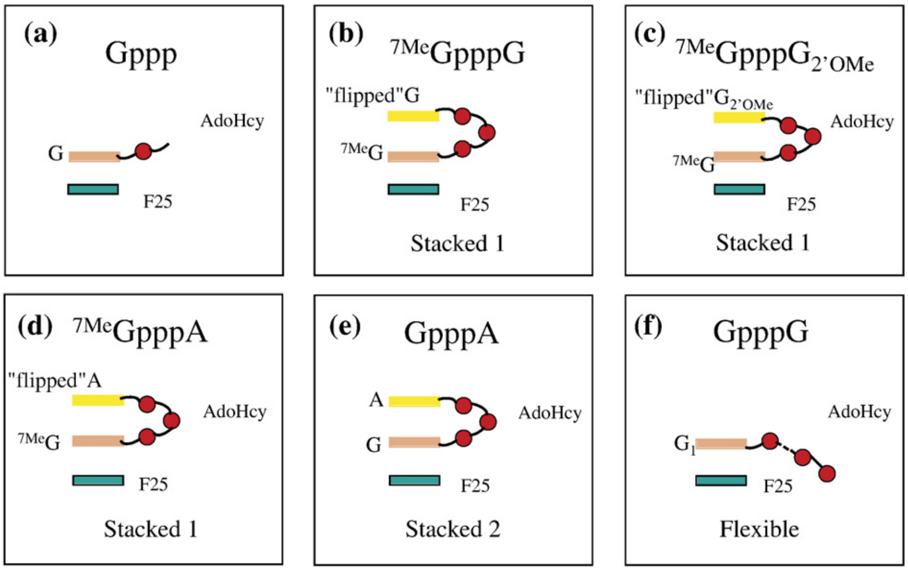


Figure 2. Schematic representations of GTP and cap analogue binding. Green rectangles represent F25, yellow and orange rectangles represent A, G, ^{7Me}G, G$_{2'OMe}$ bases as indicated, and red circles represent phosphate groups. (a) GTP complex reported earlier[20]; (b)–(d) first stacked conformation (S1) adopted by ^{7Me}GpppN cap analogues; (e) the second stacked conformation (S2) adopted by GpppA; (f) flexible conformation (F) adopted by GpppG.

aromatic ring of residue F25 and the guanine nucleobase, and the second between the guanine and the adenine nucleobases) are similar to those observed in the S1 conformation. However, the triphosphate linker and the adenosine positioning are different and the adenine base is in the same plane but flipped upside down. As a result, the A face of the guanine interacts with the A face of the adenine. The S2 conformation allows specific interactions of N6-position of the adenine, which is hydrogen-bonded to both the main-chain and side-chain oxygen groups of N18. Only the amino-group at position 6 of a purine base (i.e. an adenine) can ascertain these interactions. Thus exclusively an adenine can be positioned on top of the guanosine in that manner which corresponds, again, to the absolute conservation of an adenosine as the first transcribed nucleotide of the *Flavivirus* RNA genome. Finally, in S2 the 3′-hydroxyl group of the adenosine ribose points down to the positively charged zone where nascent RNA may be bound (see Figure 1(b)).

The third binding mode is that observed for GpppG (Figures 2(f) and 3(c)). As for the others and the GTP analogue,[20] the density is well defined in the cap-binding site, up to the first phosphate. Then, there is a clear interruption in the density. Additional density is found following the groove towards the AdoMet binding site, which cannot be clearly assigned. Thus the second guanosine seems to be rather flexible and this conformation is referred to as flexible (F).

Binding of small RNAs to NS5MTase$_{DV}$

To complement our efforts to obtain complexes of NS5MTase$_{DV}$ with cap analogues and small capped RNAs, binding studies were conducted. In continuation of earlier binding assays of GTP and cap analogues to NS5MTase$_{DV}$ using UV cross-linking,[20] here we explored the binding of small capped and non-capped RNAs of different lengths and sequences but without UV cross-linking. In analogy to the *Flavivirus* (+) RNA genome, all RNAs contained an adenosine at the first position. We performed *in vitro* binding assays of [^{32}P]-radiolabelled RNAs on NS5MTase$_{DV}$ immobilized on nickel agarose beads. Short oligonucleotides carrying at their 5′end ^{7Me}Gppp, Gppp or ppp were produced using ^{7Me}GpppA, GpppA or ATP, respectively, in conjunction with [α-^{32}P]CTP and the T7 DNA primase to synthesize NpppAC$_n$. Likewise *Escherichia coli* primase was used with ^{7Me}GpppA, GpppA or ATP, [α-^{32}P]GTP and CTP to synthesize NpppAGC$_n$ RNAs, which contain a guanine at the second position as observed in all *Flavivirus* (+) RNA genomes. His-tagged NS5MTase$_{DV}$ was expressed in *E.coli* and immobilized on the nickel beads. NS5MTase$_{DV}$-beads were incubated with capped or non-capped [^{32}P]-radiolabelled RNAs. After incubation, bound [^{32}P]-RNAs were extracted, analyzed by PAGE, and quantitated by autoradiography. As illustrated in Figure 4(a), RNAs do not bind to control nickel beads. NS5MTase$_{DV}$ fails to bind the non-capped pppAC$_{3to5}$ RNA but binds significantly $^{7Me\pm}$GpppAC$_{3to5}$. When the second transcribed nucleotide is replaced by a G, even the non-capped RNA binds significantly to NS5MTase$_{DV}$albeit less than capped $^{7Me\pm}$GpppAGC$_n$. Figure 4(b) shows quantitated data, i.e. the percentage of bound RNA for single band products. A comparison of the percentage of bound $^{7Me\pm}$GpppAC$_n$ (panels 1 and 2) and $^{7Me\pm}$GpppAGC$_{n-1}$ (panels 5 and 6), i.e. substrates of the same chain length, clearly shows that the presence of G as the second transcribed

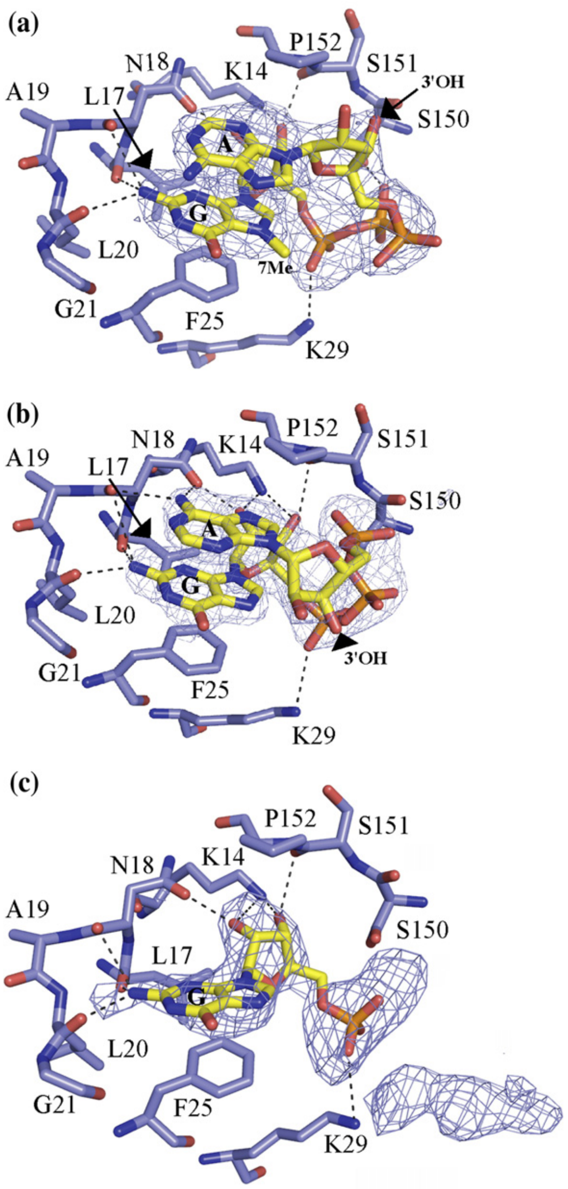


Figure 3. Cap analogue conformations. (a) Binding of ^{7Me}GpppA in the first stacked (S1) conformation in the same orientation as in Figure 1. Colors of the cap analogue and of the interacting amino acids are the same as in Figure 1(c). Hydrogen bonds are represented in black dotted lines. The F_o–F_c electron density map, contoured at 3σ, was calculated at 2.8 Å resolution from a model in which the ligand was omitted. In this conformation the 3′-OH of the adenosine cannot engage into a 3′ to 5′ phosphosdiester bond as it points towards the protein loop P150 to S152. (b) Binding of GpppA adopting the second stacked (S2) conformation. The representation is the same as in (a)). No steric clash would preclude the 3′-hydroxyl group to engage into a 3′ to 5′ phosphodiester bond. Specific interactions of the N6-position of the adenine to both the main-chain and side-chain oxygen groups of N18 are shown. (c) Binding of GpppG in a flexible (F) conformation. A GMP can be modelled without any ambiguity in the electron density map. A clear interruption in the electron density map is visible between the GMP and an additional density in the continuity of the positively charged groove joining the AdoMet binding site.

nucleotide increases binding significantly. The binding of ^{7Me}GpppAC$_{3\text{-}4}$ to NS5MTase$_{DV}$ is enhanced compared to that of GpppAC$_{3\text{-}4}$ (panels 1 and 2) whereas methylation of the guanine-N7 position does not influence significantly the binding of GpppAGC to GpppAGC$_3$ (panels 4 to 6). We also note an influence of chain length on binding. $^{7Me\pm}$GpppAC$_4$ (panel 2) binds stronger than of $^{7Me\pm}$GpppAC$_3$ (panel 1), $^{7Me\pm}$GpppAGC (panel 4) binds stronger than $^{7Me\pm}$GpppAG (panel 3) but then increasing chain length does not further increase binding.

In order to determine if capped RNA binds to the specific GTP-binding site, we performed interference assays, studying binding of ^{7Me}GpppAC$_4$ and ^{7Me}GpppAC$_5$ in the presence of GTP, ^{7Me}GTP, ATP and ^{7Me}GpppA. As illustrated in Figure 5, GTP, ^{7Me}GTP, and ^{7Me}GpppA interfered with binding of ^{7Me}GpppAC$_4$ rendering apparent IC$_{50}$ values of 0.34 mM, 0.23 mM and 1.70 mM, respectively; whereas ATP had no significant binding inhibition effect until 10 mM. Similar IC$_{50}$ values were obtained for the inhibition of ^{7Me}GpppAC$_5$ binding.

Discussion

Here, we address structural pre-requisites for cap and RNA-binding as well as mRNA cap methylation by the MTase domain of protein NS5 of dengue virus (NS5MTase$_{DV}$). Our crystallographic studies rendered structures of NS5MTase$_{DV}$ in complex with five cap analogues. Three different conformations were captured, two of them are proposed to have biological relevance.

The first stacked conformation S1, was adopted by all cap analogues methylated at the N7-position. We consider it an artifactual conformation for two reasons. (i) The 3′-hydroxyl position of the adenosine is pointing towards the protein and is thus not compatible with longer RNA substrates where it would be engaged in a regular 3′ to 5′ phosphodiester bond. (ii) The adenine of ^{7Me}GpppA does not establish any interaction with the protein, whereas the N2-position of the second guanine in ^{7Me}GpppG and ^{7Me}GpppG$_{2'OMe}$ interacts with both the main-chain and the side-chain oxygen atoms of N18. This denies the importance of the conservation of adenosine as the first transcribed nucleotide in all *Flavivirus* RNA genomes. Why do all the N7-methylated cap analogues adopt the S1 conformation with the ^{7Me}G as the central part of the stack although there are no specific interactions involving the N7-methyl group in this position? All crystal structures were determined at pH 5.8, a value at which the N7-methylated guanine is predominantly positively charged. The overall sandwich-like conformation with a positively charged ^{7Me}G nucleobase in the middle is similar to the Phe-Tyr and Trp-Trp ^{7Me}G-cap binding slot of vaccinia virus 2′ OMTase VP39 and eukaryotic initiation factor 4E (eIF4E), respectively.[26,27] There, the presence of the central positive charge was proposed to strengthen the interactions within the stack.[28] Since the pK_a

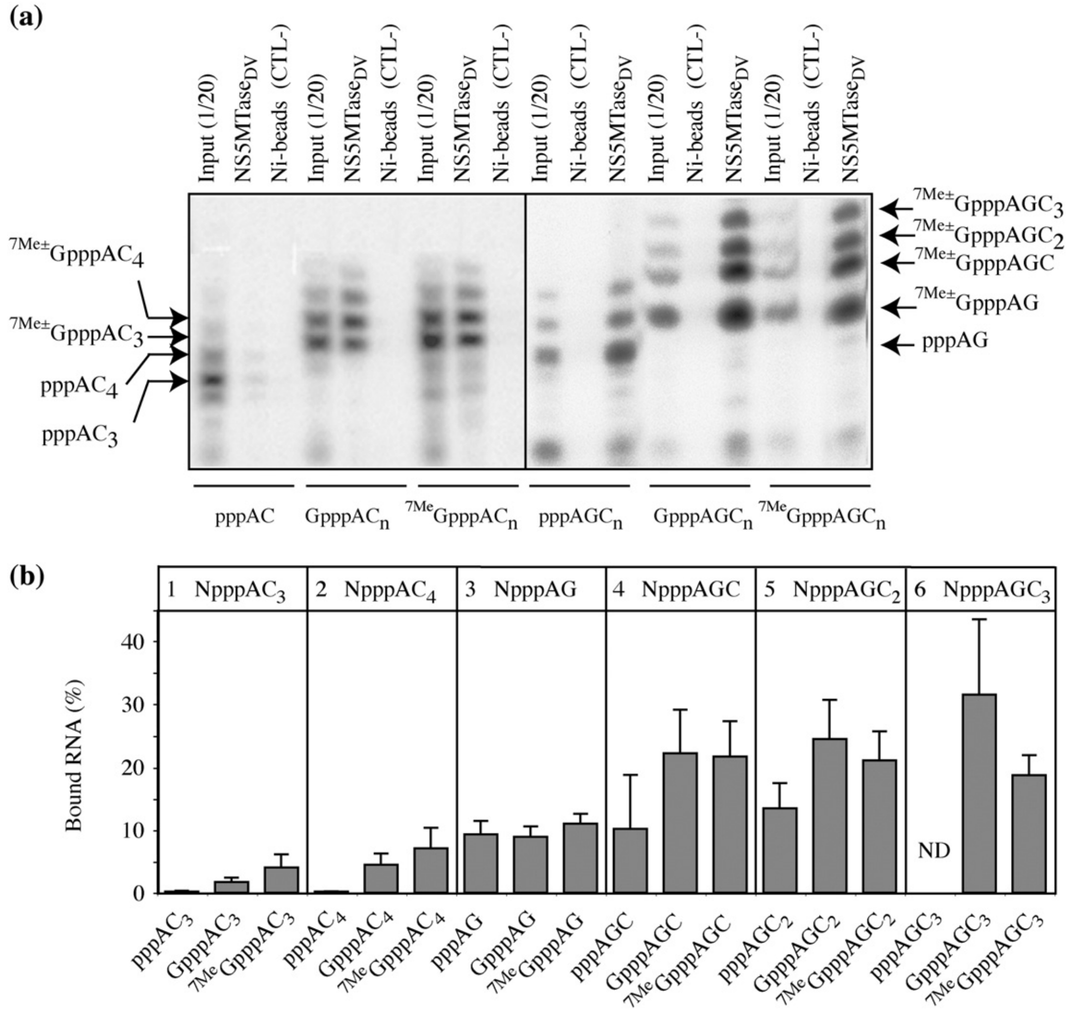
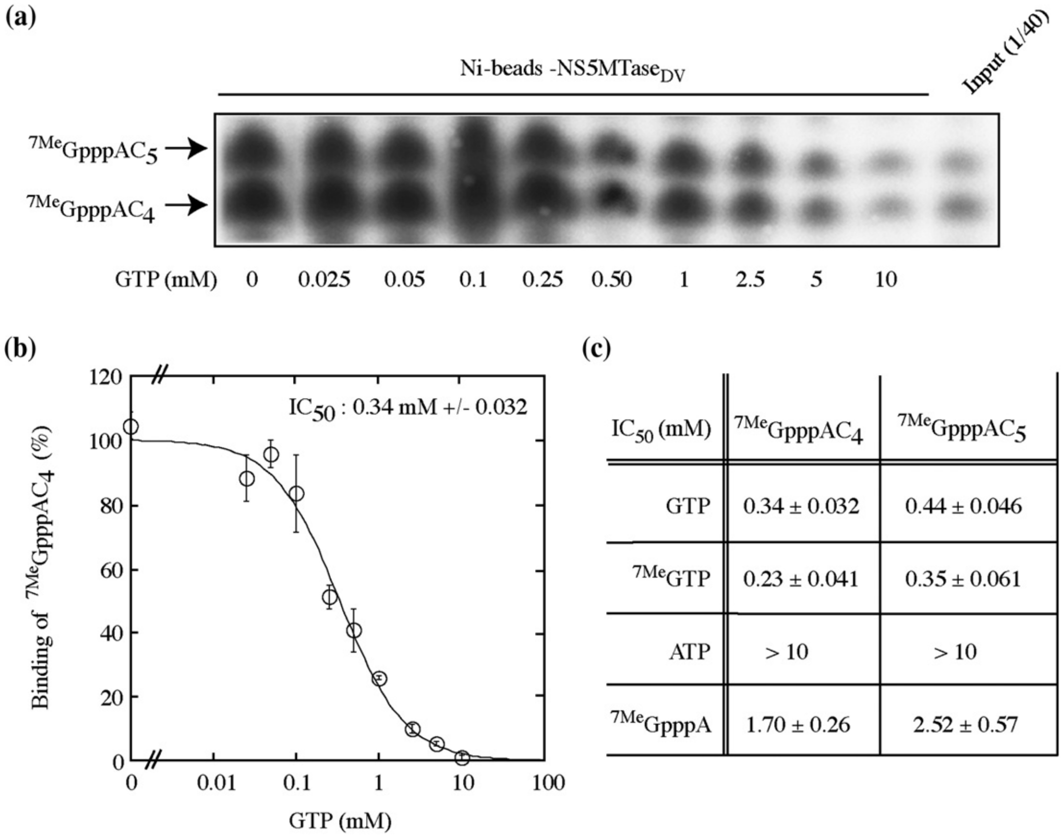


Figure 4. Binding of capped or non-capped RNAs to NS5MTase$_{DV}$. Short [^{32}P]-radiolabelled RNAs bearing a cap $^{7Me\pm}$Gppp or being uncapped were incubated with NS5MTase$_{DV}$ immobilized on Sepharose beads (see Experimental Procedures). After three washes, bound [^{32}P]-labelled RNAs were separated by PAGE and quantitated by phosphor-imager analysis. (a) PAGE analysis of non-capped and capped RNA binding of NpppAC$_n$ (left) and NpppAGC$_n$ (right). Input (1/20) indicates that input RNA was diluted 20 times before loading it on the gel. Ni-beads (CTL-) stands for negative control using beads loaded with Ni^{2+}. (b) The percentage of bound RNA was calculated comparing the amount of RNA bound to the beads with the corresponding RNA present in the input lane analyzing three independent experiments.

value of the N1 atom of free ^{7Me}G nucleobase is 7.5,[28] at pH values higher than 7.5, the positive charge at N7 is neutralized by a negative charge at the N1 atom. Alternate conformations might therefore exist at a more physiological pH value. This contributes to our conclusion that the S1 conformation observed at pH 5.8 is an artifact.

The GpppA cap analogue adopts the second conformation, the stacked conformation S2 in which the guanosine is positioned in the GTP-binding site. Two arguments suggest that this conformation might be biologically relevant. (i) Longer mRNA molecules may be appended without problems because the 3′-position of the ribose is oriented towards the positively charged surface region that may accommodate the nascent RNA (see Figure 1(b)). (ii) Only an adenine base can be used in this binding mode involving specific hydrogen bonding with residue N18. This would coincide with the absolute conservation of an adenosine as the first nucleotide of the *Flavivirus* genome. Note that the N18A mutant does not bind GpppA (K_d>1 mM,[20]). Concerning the non-physiological pH value used in the soaking experiments, for GpppA (in contrast to ^{7Me}GpppA) no change of conformation is expected at more physiological pH values because the N1 and N7 atoms of G have pK_a values of 9.4 and 3.3,[29] respectively. It is obvious that conformation S2 does not represent a functional conformation corresponding to the MTase activities of NS5MTase$_{DV}$, because both the guanine-N7 or the adenosine-2′O position of the GpppA substrate are too far away from the AdoMet binding site. In fact NS5MTase$_{DV}$ harbored only the MTase activities there would be no need for the GTP-binding site to accommodate non-methylated GTP. We propose the hypothesis that the S2 conformation of GpppA mimics the reaction product of



IC$_{50}$ (mM)	^{7Me}GpppAC$_4$	^{7Me}GpppAC$_5$
GTP	0.34 ± 0.032	0.44 ± 0.046
^{7Me}GTP	0.23 ± 0.041	0.35 ± 0.061
ATP	> 10	> 10
^{7Me}GpppA	1.70 ± 0.26	2.52 ± 0.57

Figure 5. Inhibition of ^{7Me}GpppAC$_4$ and ^{7Me}GpppAC$_5$ binding to NS5MTase$_{DV}$ by GTP, ^{7Me}GTP, and ^{7Me}GpppA. (a) PAGE analysis of ^{7Me}GpppAC$_4$ and ^{7Me}GpppAC$_5$ binding was performed in presence of 0 to 10 mM GTP. The input was diluted 40 times. (b) Inhibition curve of ^{7Me}GpppAC$_4$ binding by GTP. Binding in absence of GTP was set to 100 %. The standard deviation of three independent experiments is shown. The IC$_{50}$ value was calculated as given in Experimental Procedures. (c) Interference of GTP, ^{7Me}GTP, ATP and ^{7Me}GpppA with ^{7Me}GpppAC$_4$ and ^{7Me}GpppAC$_5$ binding. IC$_{50}$ values were calculated as given in Experimental Procedures. ATP does not inhibit binding up to 10 mM.

a putative *Flavivirus* guanylyltransferase (GTase) activity: pppG + ppAGN leading to GpppAGN + pyrophosphate. A guanylyltransfer reaction preceding the MTase reactions would be the only occasion where the GTP-binding site of NS5MTase$_{DV}$ needs to accommodate a non-methylated GTP.

The third conformation, adopted by GpppG, shows no stacked cap analogue. One guanosine is located in the cap-binding site and density is interrupted after the first phosphate group. The second guanosine is not visible. One reason for this might be a cleavage of GpppG into GMP and non-identified fragments. The observed residual density (Figure 3(c)) might correspond to either a pyrophosphate, which can be fitted in one part of the density (not shown), or residual density of a GDP fragment. In order to test the possibility of a cleavage, we incubated the concentrated protein with 2 mM GpppG in the crystallization buffer for 30 min at room temperature. The subsequent HPLC analysis of the mixture in comparison with control GpppG did not show any cleavage (data not shown). A second possible explanation is that the second guanosine adopts an extended orientation compatible with 2′O-methylation but with an increased mobility, since (i) it does not correspond to the conserved, first transcribed adenosine nucleotide of *Flavivirus* (+) RNA genomes, and/or (ii) a longer RNA is needed to stabilize an extended conformation. This extended conformation would correspond to NS5MTase$_{DV}$ in the 2′O-methylation mode.

Binding experiments were conducted using small capped and non-capped RNAs containing the strictly conserved first nucleotide, adenosine, and in some cases also the second strictly conserved nucleotide, guanine, of the *Flavivirus* (+) RNA genome (Figure 4). In agreement with methyltransferase activity assays,[20] NS5MTase$_{DV}$ does not bind the non-capped pppAC$_{3to5}$ RNA. In contrast, it binds significantly to capped $^{7Me\pm}$GpppAC$_{3-5}$, which have been shown to be efficient substrates of NS5MTase$_{DV}$ being exclusively methylated at the 2′O-position.[20,21] We observe that an increase of chain length from ^{7Me}GpppAC$_3$ and ^{7Me}GpppAC$_4$ leads to an increase of affinity. Competition experiments using GTP, ^{7Me}GTP, ^{7Me}GpppA and ATP show that binding of ^{7Me}GpppAC$_4$ and ^{7Me}GpppAC$_5$ is specifically inhibited by guanine-containing compounds (Figure 5). This observation indicates that the guanine-cap of these RNAs preferentially occupies the GTP-binding site observed for the cap analogues in all NS5MTase$_{DV}$–complex structures and that this site serves effectively as the cap-binding site upon 2′O methylation. There was no difference in inhibition of ^{7Me}GpppAC$_4$ and ^{7Me}GpppAC$_5$-binding by GTP and ^{7Me}GTP. The higher IC$_{50}$ value of ^{7Me}GpppA in comparison to $^{7Me\pm}$GTP is in accordance with the lower affinity of ^{7Me}GpppA compared to GTP, ^{7Me}GTP and GpppA observed earlier.[20] The specific interactions between the adenine of GpppA and NS5MTase$_{DV}$ in S2, which are missing for ^{7Me}GpppA in S1 may explain the higher affinity of GpppA compared to ^{7Me}GpppA. The binding and affinity measurements were done at pH values 7.5 and 7.6, respectively, indicating that the differences observed between S1 (^{7Me}GpppA) and S2 (GpppA), may still be maintained at these pH values.

Binding improves when the RNAs contain the correct first and second nucleotide, guanosine, of the *Flavivirus* (+) RNA genome. In this case even non-capped RNA binds to some extent, albeit weaker than their capped counterparts (Figure 4). This observation is in agreement with the finding that a capped RNA corresponding to the exact (+) RNA genomic 5′ sequence served as efficient substrate for both MTase activities of *Flavivirus* NS5MTase.[7,22,23] Thus, these substrates might bind in different modes mimicking the guanylyltransfer as well as the N7 and 2′O-methyltransfer. One or several of these positions may provide specific interactions for the conserved guanosine. As in the case of $^{7Me\pm}$GpppAC$_n$, for $^{7Me\pm}$GpppAGC$_n$ RNAs an increase of chain length leads to an increase of affinity, here the plateau is reached at $^{7Me\pm}$GpppAGC. In conclusion, $^{7Me\pm}$GpppAGC should be an excellent candidate to be tested in soaking or co-crystallization experiments with NS5MTase$_{DV}$. This has not yet been possible because yields of producing $^{7Me\pm}$GpppAGC$_n$ by *E.coli* primase are considerably lower in comparison to the optimized production of $^{7Me\pm}$GpppAC$_n$ by T7 DNA primase.[21] Optimization experiments are under way.

In a modeling approach we built an extended ^{7Me}GpppA–complex representing NS5MTase$_{DV}$ in 2′O-methylation mode. The model (Figure 6(a)) is based on (i) the existent complex structures, (ii) the known catalytic tetrad KDKE, which plays a pivotal role in the 2′O methyl transfer reaction of NS5MTase$_{WNV}$[7] and is situated next to the 2′O position in vaccinia virus VP39 complexed to a short RNA substrate (see Figure 6(b) for comparison).[25] The position of the guanine nucleobase, the ribose and the phosphorus atom of the α-phosphate were maintained in the common position that is found in all the complexes. The remaining torsion angles of the ^{7Me}GpppA molecule were released allowing the β and γ-phosphate groups to be positioned in the positively charged crevice connecting the GTP binding site and the AdoHcy binding site. In particular, the γ-phosphate was positioned in the same position as a sulfate ion present in several complexes. This sulfate ion originates from the crystallization condition and could mimic a phosphate group. The ribose of the adenosine nucleoside was placed in such a way that it allows the reaction to take place and the cap analogue to be continued via a hypothetical 3′–5′ linkage pointing towards the putative RNA-binding region of NS5MTase$_{DV}$. The AdoHcy was completed to AdoMet using chemical restraints and superposition with VP39 in complex with AdoMet (PDB code 1VPT[30]). The position of the ribose 2′O atom allows its nucleophilic attack of

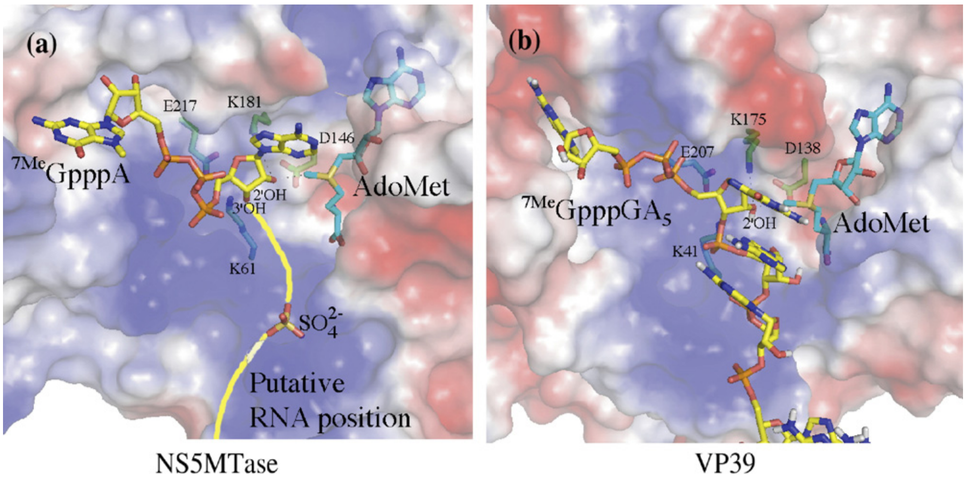


Figure 6. Model of the ternary complex of NS5MTase of DV in 2′O methylation mode compared to the 2′OMTase VP39 of vaccinia virus. (a) Model of NS5MTase$_{DV}$ in complex with ^{7Me}GpppA in the proper conformation for 2′O-methylation. The electrostatic surface representation of NS5MTase$_{DV}$ is shown. ^{7Me}GpppA and AdoMet are represented and colored as for Figure 1(a). The G moiety of ^{7Me}GpppA binds to the GTP-binding site. The 2′O position of the ribose of the adenosine lies within 2.6 Å from the carbon atom of the methyl group of AdoMet. The catalytic tetrad K61, D146, K181 and E217 is shown in the catalytic center around the 2′O position. A sulfate ion is shown in the putative RNA binding groove. (b) Crystallographic structure of 2′OMTase VP39 in complex with a capped RNA substrate ^{7Me}GpppGA$_5$ (PDB code 1AV6). The catalytic tetrad in the catalytic center is indicated and located around the 2′O position of the second guanosine of the RNA substrate. This second base stacks to the following adenine bases, as shown in the center. The first three nucleobases GAA are bound without apparent specificity.

the C atom of the AdoMet methyl group following a proposed in-line S_N2 reaction.[31] The angle between S^{δ}–C^{ε} of AdoMet and the ribose 2′O is 162°, and the distance between the AdoMet-C^{ε} and the ribose 2′O is 2.6 Å, comparable with the values found in VP39 complexed with short RNA.[25] This places the ribose 2′O atom 3.2 Å away from the catalytic lysine K181 in similar conformation to the VP39 complex. The adenine base was positioned in a way to avoid any steric clash with AdoMet and the protein. The ribose 3′O atom of the ^{7Me}GpppA model points towards the positively charged groove where the nascent RNA chain might be accommodated in a similar way as observed in VP39 (Figure 5(b)). A sulfate group, which originates from the crystallization conditions and is present in all the refined structures could mimic one of the inter-nucleotidic phosphates in an actual RNA strand as indicated in Figure 6(a). Since the RNA is expected to be repositioned between the N7- and 2′O-methylation step, binding to the RNA-binding groove of NS5MTase$_{DV}$ should not be sequence-specific but based on ionic interactions with the phosphate backbone. Apart from binding the cap in the GTP-binding site, base stacking of the second, third and fourth nucleotide might be necessary to stabilize the conformation of the 2′O-methylation ternary complex explaining why we were not able to crystallize cap analogues in the conformation of the model. The adenine ring could stack with the following bases of the nascent

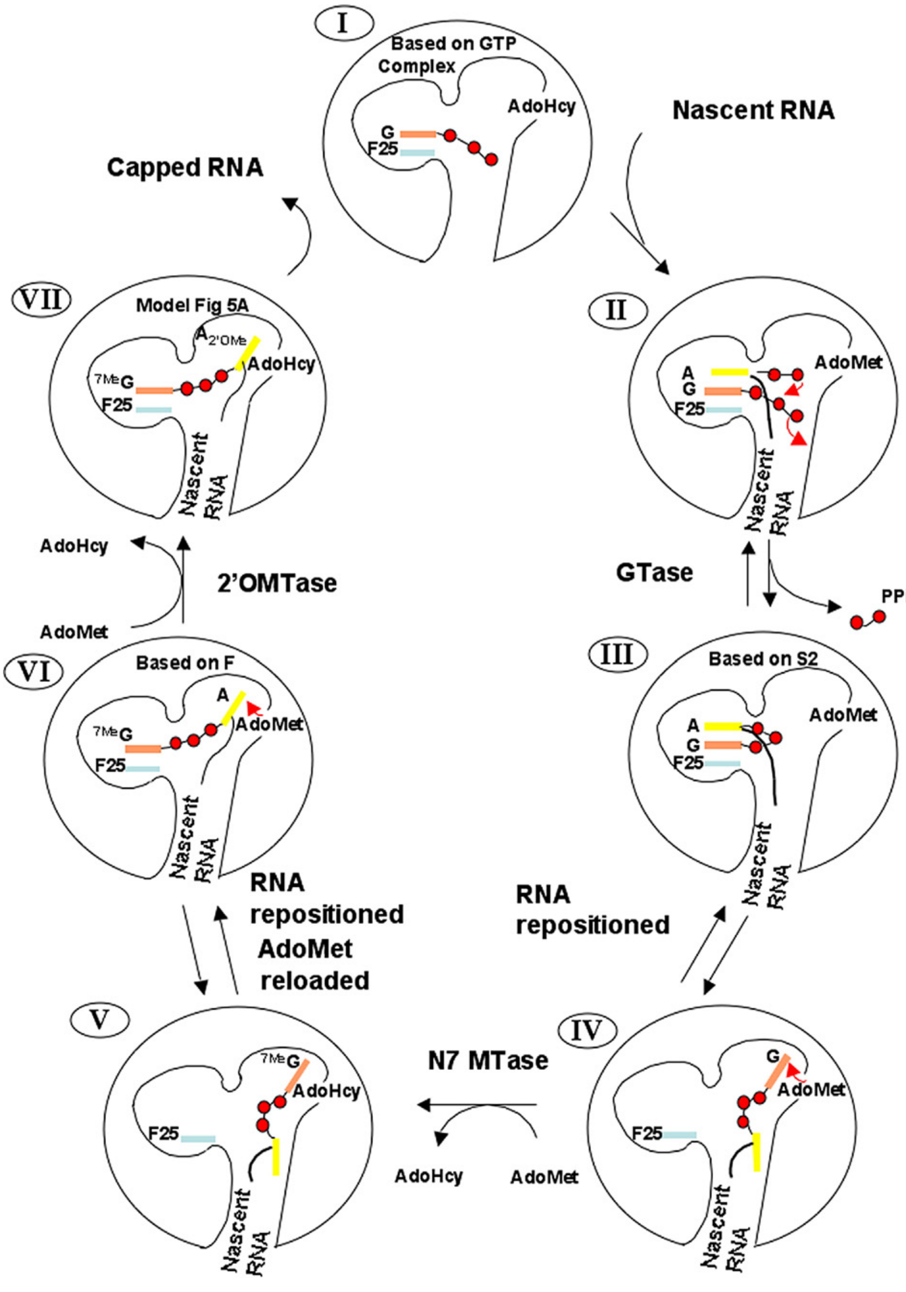


Figure 7. A proposition for the different activities of the *Flavivirus* NS5MTase domain involved in RNA cap formation (see Discussion).

RNA strand, as it was observed for VP39. Considering the length of the positively charged crevice (approximately 17 Å), a maximum of five consecutive nucleotides with stacked bases could be accommodated after the adenosine (mean of 3.4 Å between the stacked bases), indicating why $^{7Me\pm}$GpppAC$_4$ binds stronger than $^{7Me\pm}$GpppAC$_3$.

In conclusion, a hypothetical scheme of the *Flavivirus* capping pathway is presented in Figure 7 that combines different steps where the NS5MTase domain seems to be involved. This model hypothesizes both that conformation S2 represents the product of the GTase activity and that NS5MTase$_{DV}$ methylates the guanine-N7 position[7] before methylating the adenosine-2′O position.[20,21] (I) GTP binds to the GTP-binding site of NS5MTase[20] followed by a hypothetical guanylyltransfer on nascent diphosphate RNA (II to III). Repositioning of the capped RNA (III to IV) allows subsequent methylation at the guanine-N7 position[22] (IV to V). During N7-methyltransfer, the binding of the RNA substrate cannot be supported by the accommodation of the cap guanine in the GTP-binding site. We suspect that stacking interactions within the RNA substrate might be important for the formation of the ternary complex. Stacking interactions between purine bases are stronger than interactions between purine and pyrimidine bases.[32] This might explain the stronger binding of GpppAG compared to GpppAC$_3$ observed here and the absence of N7MTase activity using GpppAC$_n$[20] compared to GpppAGN substrates.[22,23] After the N7-methyltransfer cofactor AdoMet is reloaded and the capped RNA is repositioned with the methylated cap being in the GTP binding site (V to VI). Finally, 2′O-methylation occurs[20–22] (VI to VII) and the capped and methylated nascent RNA dissociates.

The expected multiple movements of the RNA substrate between various positions on NS5MTase$_{DV}$, necessary to accomplish its functions, might explain the difficulties to obtain complex structures with small capped RNAs. Nevertheless, more biologically relevant complex structures of *Flavivirus* NS5MTase domains with its substrates and/or products are needed to better understand the *Flavivirus* capping machinery. Likewise, more biochemical data are required on the mechanism of both methyltransfer reactions, and especially data on the putative guanylyltransferase activity being presented here as a hypothesis. These studies will pave the way into the further exploration of *Flavivirus* capping as a target of antiviral therapy.

Experimental Procedures

Expression and purification of NS5MTase$_{DV}$

Recombinant NS5MTase$_{DV}$ corresponds to residues 1 to 296 of NS5. The protein used for crystallization was expressed and purified as described.[20] For the NS5MTase$_{DV}$ binding assay, the following modifications were applied. NS5MTase$_{DV}$ was expressed in *E.coli* Rosetta(pLysS) transformed with pQE30/NS5MTase$_{DV}$. Cells were grown in Luria-Bertani medium containing ampicillin (100 μg/ml) and chloramphenicol (17 μg/ml). At an A_{600} of 0.6, isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 50 μM, and expression was allowed to proceed for 4 h at 30 °C. The cellular pellet was resuspended in lysis buffer (10 mM Bicine (pH 7.5), 300 mM NaCl, 1 mM phenylmethylsulfonyl fluoride (PMSF)) supplemented with 1 mM benzamidine, 10 mM imidazole, 100 μg/ml of lysozyme and 1 μg/ml of DNase I before sonication. After clarification, the His-tagged soluble protein was immobilized on chelating Sepharose fast flow resin (Amersham Biosciences) loaded with Ni^{2+}. After several washes with Bicine buffer containing 25 mM imidazole, 20 μg of bead-bound NS5MTase$_{DV}$ was used for pull-down assay.

Crystallization

Crystals of NS5MTase$_{DV}$ were grown at room temperature in hanging drops, by mixing 1 μl of protein solution with the same volume of a reservoir solution containing 0.4 M ammonium sulfate, 0.1 M sodium citrate (pH 5.8), and 1.2 M lithium sulfate. Crystals were soaked for 3 h at room temperature in the mother liquor supplemented with 25% (v/v) glycerol for cryo-protection and 2 mM of cap analogue (New England Biolabs) and then flash frozen in a nitrogen stream. Crystals belong to the space group $P3_121$. Data were collected in ESRF on the beamline ID14-EH2 (Table 1). Images were processed using DENZO[33] and MOSFLM,[34] and intensities were merged with SCALA.[35]

Structure determination

The cap-analogue soaked crystals have the same space group but slightly different cell parameters compared to the native NS5MTase$_{DV}$. The structures were solved by molecular replacement using AMoRE[36] with the native NS5MTase$_{DV}$ (PDB code 1L9K) as a search model. After an initial rigid body refinement of the MR solution, F_o–F_c SIGMAA-weighted electron density map clearly indicated the presence of AdoHcy (excepted for the data of the ^{7Me}GpppG–complex), sulfate ions and cap analogue in three different conformations depending on the nature of the cap (see Results). Citrate and glycerol were also seen in some electron density maps. Refinement was carried out with REFMAC using the maximum likelihood approach and bulk solvent correction.[37] Reflections omitted from refinement for R_{free} calculation,[38] where those used for the R_{free} calculation of the native NS5MTase$_{DV}$ (PDB code 1L9K). Manual rebuilding between refinement cycles were done using Coot.[39] The final models include residues 7 to 264 for all the models. The residue 6 can be seen in some electron density maps. Residues up to 269 are ordered and can be refined in the electron density map of the NS5MTase$_{DV}$–^{7Me}GpppG complex only. Stereochemistry was evaluated using PROCHECK.[40] Statistics of data collection and refinement for all five complexes are provided in Table 1.

RNA synthesis

Capped and non-capped RNA were synthesized *in vitro* using either bacteriophage T7 or *E. coli* DNA primase. The T7 primase fragment was produced in *E. coli* BL21(DE3) cells (500 ml) transformed with pET19b/PrDT as

Table 1. Statistics of data collection and refinement

	^{7Me}GpppA	^{7Me}GpppG	^{7Me}GpppG$_{2'OMe}$	GpppA	GpppG
Space group	$P\ 3_121$	$P\ 3_121$	$P\ 3_121$	$P\ 3_121$	$P\ 3_121$
Cell dimensions	$a=b=111.56$	$a=b=114.72$	$a=b=108.35$	$a=b=108.72$	$a=b=111.5$
a,b,c (Å)	$c=56.33$	$c=56.43$	$c=55.87$	$c=55.92$	$c=56.53$
α,β,γ (°)	$\alpha=\beta=90$	$\alpha=\beta=90$	$\alpha=\beta=90$	$\alpha=\beta=90$	$\alpha=\beta=90$
	$\gamma=120$	$\gamma=120$	$\gamma=120$	$\gamma=120$	$\gamma=120$
X-ray source	ID14-EH2	ID14-EH2	ID14-EH2	ID14-EH2	ID14-EH2
Wavelength (Å)	0.933	0.933	0.933	0.933	0.933
Resolution range (Å)	27.96–2.8 (2.95–2.8)	35.00–2.70 (2.85–2.70)	38.9–1.80 (1.90–1.80)	35.0–2.20 (2.32–2.20)	35–2.75 (2.90–2.75)
Total reflections	52,349 (7375)	97,271 (14,569)	24,1820 (20,960)	66,764 (8763)	54,461 (7458)
Unique reflections	10,025 (1448)	12,029 (1750)	35,115 (4976)	19,614 (2825)	10,782 (1551)
Completeness (%)	97.1 (97.1)	99.9 (99.9)	99.5 (97.3)	99.9 (99.8)	99.9 (99.7)
$<I/\sigma(I)>$	27.4 (3.4)	21.7 (5.1)	37.6 (6.1)	15.5 (2.7)	21.7 (2.7)
R_{sym} (%)[a]	8.3 (51.5)	8.8 (33.0)	3.4 (21.5)	6.1 (38.2)	5.6 (50.8)
Multiplicity	5.2 (5.1)	8.2 (8.3)	6.9 (4.2)	3.4 (3.1)	5.1 (4.8)
R_{work} (%)[b]	19.5	20.4	16.3	19.7	20.8
R_{free} (%)[c]	24.4	25.5	19.6	22.8	27.1
Number of atoms					
Protein	2022	2077	2105	2055	2025
Water molecules	27	42	203	98	8
Cap analogue	51	52	64 (alternative conformation)	50	25 (Gp)
AdoHcy	26	0	30 (alternative conformation)	26	26
SO_4	45 (9 SO_4)	55 (11 SO_4)	35 (7 SO_4)	25 (5 SO_4)	35 (6 SO_4)
Glycerol	0	0	18 (3 glycerol molecules)	6 (1 glycerol)	0
Citrate	13 (1 citrate)	0	13 (1 citrate)	13 (1 citrate)	0
rms deviations					
Bond lengths (Å)	0.013	0.010	0.012	0.012	0.015
Bond angles (°)	1.615	1.433	1.542	1.476	1.779
Ramachandran analysis (%)					
Most favored	86.3	90.6	93.6	95.0	87.2
Additionally allowed	12.8	9.0	5.9	4.5	12.3
Generously allowed	0.9	0.4	0.0	0.0	0.5
Disallowed	0.0	0.0	0.5	0.5	0.0
PDB code	2P3O	2P40	2P41	2P3L	2P3Q

[a] $R_{\mathrm{sym}}=\sum|I-<I>|/\sum I$, where I is the observed intensity and $<I>$ is the average intensity. Values in parentheses refer to the highest-resolution shell.

[b] $R=\sum||F_{\mathrm{o}}|-|F_{\mathrm{c}}||/\sum|F_{\mathrm{o}}|$.

[c] A 5% of all reflections that are never used in crystallographic refinement are used to calculate the R_{free} (same calculation as for R factor).

detailed.[21] The *E. coli* DNA primase (*dnaG* gene) cDNA was amplified by PCR using primers 5′-CCATGAAACATCACCATCACCATCACGCTGGACGAATCCCACGCGTATTC-3′ and 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACTTTTTCGCCAGCTCCTGG 3′ and cloned into the pDest14 (Invitrogen) expression vector using the Gateway technology. *E. coli* Rosetta(pLysS) cells were transformed with pDest14/6His-DNAG and grown in Luria-Bertani medium containing ampicillin and chloramphenicol. At an A_{600} of 0.6, IPTG was added to a final concentration of 200 μM, and expression was allowed to proceed for 4 h at 37 °C. The cellular pellet was resuspended in 10 ml lysis buffer (50 mM Tris (pH 8.5), 300 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol, antiprotease cocktail (complete, Roche)) supplemented with 10 mM imidazole, 100 μg/ml of lysozyme, 1 μg/ml of DNase I and 0.5% (v/v) Triton X100. After lysis by sonication and clarification, immobilized-metal-affinity chromatography was used for one-step purification (chelating Sepharose fast flow resin (Amersham Biosciences) loaded with Ni^{2+}). The *E. coli* DNA primase was eluted with lysis buffer (pH changed to pH7.5) containing 250 mM imidazole, adjusted to 50% glycerol and stored at −20 °C.

The ^{32}P-labelled pppAC$_n$ or $^{7Me\pm}$GpppAC$_n$ RNAs were produced as described[21] with the following minor modifications: in a reaction volume of 10 μl, 10 μM DNA template CCCCGGGTCT$_{25}$ was incubated with 4 μM T7 DNA primase, 5 mM CTP (containing 5 μCi of [α-^{32}P]CTP; Amersham Biosciences) and 1 mM ATP or cap analogue (New England Biolabs) in 40 mM Tris (pH 7.5), 10 mM $MgCl_2$, 50 mM potassium glutamate, 1 μM $ZnCl_2$, 10 mM dithiothreitol (DTT) and 50 μg/ml of bovine serum albumin (BSA; New England Biolabs) for 4 h at 37 °C. The ^{32}P-labelled pppAGC$_n$ or $^{7Me\pm}$GpppAGC$_n$ RNAs were produced by incubating purified *E. coli* DNA primase in a reaction volume of 20 μl with 10 μM DNA template GGGGCTGCAAAGCTGG, 0.5 mM ATP or cap analogue, 0.5 mM CTP and 0.5 mM GTP (containing 2 μCi of [α-^{32}P] GTP; Amersham Biosciences) in 50 mM Hepes (pH 7.5), 10 mM magnesium acetate, 50 mM potassium glutamate, 10 mM DTT and 50 μg/ml of BSA for 20 h at 30 °C. Reactions were stopped by the addition of RNase-free DNase I (Amersham Biosciences; 500 units/ml, 30 min at 37 °C) and proteinase K (Invitrogen; 100 μg/ml, 16 h at 37 °C) in order to eliminate the DNA template and the primase present in the reaction mixture. Proteinase K was inactivated at 70 °C during 5 min and after clarification, the RNA were, respectively, diluted to 100 μl of 10 mM Tris (pH 7.5), 50 mM NaCl, 2.5% glycerol, BSA (500 μg/ml) and stored at −20 °C.

NS5MTase binding and competition assays

Binding reactions were performed during 2 h at 4 °C in 10 mM Tris (pH 7.5), 50 mM NaCl, 2.5% glycerol and BSA (500 μg/ml) in a total volume of 150 μl. For competition experiments incubation mixtures contained increasing concentrations of GTP, ^{7Me}GTP, ATP, ^{7Me}GpppA (see Figure 5). NS5MTase$_{DV}$ linked to Ni beads was produced as described above. For each experiment, we used 30 μl of NS5MTase$_{DV}$ beads (≈4 μg/μl) and 10 μl of ^{32}P-radiolabelled RNA. Accordingly, if we suppose that 100% of the cap analogues were incorporated during capped RNA synthesis, which is the case for the optimized T7 DNA primase reactions, 1 nmol of RNA was incubated with around 4 nmol of protein linked to Ni beads. After three washes with binding buffer, the RNA bound to the beads was resuspended in 10 μl of formamide/EDTA gel-loading buffer. Bound RNA was separated by polyacrylamide gel electrophoresis (PAGE, 14% acrylamide/bisacrylamide (19:1), 7 M urea) using TTE buffer (89 mM Tris (pH 8.0), 28 mM taurine, 0.5 mM EDTA). RNA bands (input corresponds to 1/20 or 1/40 of the RNA incubated with NS5MTase$_{DV}$-beads) were visualized using photo-stimulated plates (Fluorescent Image Analyzer FLA3000 (Fuji)). Quantification of single bands was performed using Image Gauge (Fuji) software. Three independent experiments were done for each RNA and two for each inhibitor except for GTP inhibition where three independent experiments were conducted. The IC$_{50}$ (inhibitor concentration at 50% binding) values were determined using Kaleidagraph. Data were adjusted to a logistic dose–response function (% activity = 100/(1+ [*I*] /IC$_{50}$)^b, where *b* corresponds to the slope factor that determines the slope of the curve.[41]

Protein Data Bank accession codes

The atomic coordinates and structure factors (codes 2P3L, 2P3O, 2P3Q, 2P40 and 2P41) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ†.

Acknowledgements

This work was supported by a grant from the Direction Générale de l'Armement, and by the VIZIER integrated project (LSHG-CT-2004-511960) of the European Union 6th Framework.

References

1. Furuichi, Y. & Shatkin, A. J. (2000). Viral and cellular mRNA capping: past and prospects. *Adv. Virus Res.* **55**, 135–184.
2. Gu, M. & Lima, C. D. (2005). Processing the message: structural insights into capping and decapping mRNA. *Curr. Opin. Struct. Biol.* **15**, 99–106.
3. Shuman, S. (2001). Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog. Nucl. Acid Res. Mol. Biol.* **66**, 1–40.
4. Bougie, I. & Bisaillon, M. (2004). The broad spectrum antiviral nucleoside ribavirin as a substrate for a viral RNA capping enzyme. *J. Biol. Chem.* **279**, 22124–22130.
5. Graci, J. D. & Cameron, C. E. (2006). Mechanisms of action of ribavirin against distinct viruses. *Rev. Med. Virol.* **16**, 37–48.
6. Magden, J., Kaariainen, L. & Ahola, T. (2005). Inhibitors of virus replication: recent developments and prospects. *Appl. Microbiol. Biotechnol.* **66**, 612–621.
7. Ray, D., Shah, A., Tilgner, M., Guo, Y., Zhao, Y., Dong, H. *et al.* (2006). West Nile virus 5′-cap structure is formed by sequential guanine N-7 and ribose 2′-O methylations by nonstructural protein 5. *J. Virol.* **80**, 8362–8370.
8. Ahola, T. & Kaariainen, L. (1995). Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc. Natl Acad. Sci. USA*, **92**, 507–511.
9. Ogino, T. & Banerjee, A. K. (2007). Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol. Cell*, **25**, 85–97.
10. Zhou, M., Deng, L., Kashanchi, F., Brady, J. N., Shatkin, A. J. & Kumar, A. (2003). The Tat/TAR-dependent phosphorylation of RNA polymerase II C-terminal domain stimulates cotranscriptional capping of HIV-1 mRNA. *Proc. Natl Acad. Sci. USA*, **100**, 12666–12671.
11. Engelhardt, O. G. & Fodor, E. (2006). Functional association between viral and cellular transcription during influenza virus infection. *Rev. Med. Virol.* **16**, 329–345.
12. Bisaillon, M. & Lemay, G. (1997). Characterization of the reovirus lambda1 protein RNA 5′-triphosphatase activity. *J. Biol. Chem.* **272**, 29954–29957.
13. Kim, J., Parker, J. S., Murray, K. E. & Nibert, M. L. (2004). Nucleoside and RNA triphosphatase activities of orthoreovirus transcriptase cofactor mu2. *J. Biol. Chem.* **279**, 4394–4403.
14. Reinisch, K. M., Nibert, M. L. & Harrison, S. C. (2000). Structure of the reovirus core at 3.6 Å resolution. *Nature*, **404**, 960–967.
15. Kaariainen, L. & Ahola, T. (2002). Functions of alphavirus non-structural proteins in RNA replication. *Prog. Nucl. Acid Res. Mol. Biol.* **71**, 187–222.
16. Cleaves, G. R. & Dubin, D. T. (1979). Methylation status of intracellular dengue type 2 40 S RNA. *Virology*, **96**, 159–165.
17. Bartelma, G. & Padmanabhan, R. (2002). Expression, purification, and characterization of the RNA 5′-triphosphatase activity of dengue virus type 2 nonstructural protein 3. *Virology*, **299**, 122–132.
18. Benarroch, D., Selisko, B., Locatelli, G. A., Maga, G., Romette, J. L. & Canard, B. (2004). The RNA helicase, nucleotide 5′-triphosphatase, and RNA 5′-triphosphatase activities of dengue virus protein NS3 are Mg^{2+}-dependent and require a functional Walker B motif in the helicase catalytic core. *Virology*, **328**, 208–218.
19. Wengler, G. & Wengler, G. (1993). The NS 3 nonstructural protein of flaviviruses contains an RNA triphosphatase activity. *Virology*, **197**, 265–273.
20. Egloff, M. P., Benarroch, D., Selisko, B., Romette, J. L. & Canard, B. (2002). An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J.* **21**, 2757–2768.

† http://www.rscb.org/

21. Peyrane, F., Selisko, B., Decroly, E., Vasseur, J. J., Benarroch, D., Canard, B. & Alvarez, K. (2007). High-yield production of short GpppA- and 7MeGpppA-capped RNAs and HPLC-monitoring of methyltransfer reactions at the guanine-N7 and adenosine-2′O positions. *Nucl. Acids Res.* **35**, e26.
22. Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z. *et al.* (2007). Structure and function of flavivirus NS5 methyltransferase. *J. Virol.* **81**, 3891–3903.
23. Dong, H., Ray, D., Ren, S., Zhang, B., Puig-Basagoiti, F., Takagi, Y. *et al.* (2007). Distinct RNA elements confer specificity to flavivirus RNA cap methylation events. *J. Virol.* **81**, 4412–4421.
24. Li, J., Wang, J. T. & Whelan, S. P. (2006). A unique strategy for mRNA cap methylation used by vesicular stomatitis virus. *Proc. Natl Acad. Sci. USA*, **103**, 8493–8498.
25. Hodel, A. E., Gershon, P. D. & Quiocho, F. A. (1998). Structural basis for sequence-nonspecific recognition of 5′-capped mRNA by a cap-modifying enzyme. *Mol. Cell*, **1**, 443–447.
26. Hodel, A. E., Gershon, P. D., Shi, X., Wang, S. M. & Quiocho, F. A. (1997). Specific protein recognition of an mRNA cap through its alkylated base. *Nature Struct. Biol.* **4**, 350–354.
27. Matsuo, H., Li, H., McGuire, A. M., Fletcher, C. M., Gingras, A. C., Sonenberg, N. & Wagner, G. (1997). Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. *Nature Struct. Biol.* **4**, 717–724.
28. Hu, G., Gershon, P. D., Hodel, A. E. & Quiocho, F. A. (1999). mRNA cap recognition: dominant role of enhanced stacking interactions between methylated bases and protein aromatic side chains. *Proc. Natl Acad. Sci. USA*, **96**, 7149–7154.
29. Fasman, G. D. (1975). *Handbook of Biochemistry and Molecular Biology: Nucleis Acids*, 1, CRC, Cleveland.
30. Hodel, A. E., Gershon, P. D., Shi, X. & Quiocho, F. A. (1996). The 1.85 Å structure of vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends. *Cell*, **85**, 247–256.
31. Li, C., Xia, Y., Gao, X. & Gershon, P. D. (2004). Mechanism of RNA 2′-O-methylation: evidence that the catalytic lysine acts to steer rather than deprotonate the target nucleophile. *Biochemistry*, **43**, 5680–5687.
32. Saenger, W. (1984). *Principles of Nucleic Acid Structure.* Springer, New York.
33. Otwinowski, Z. & Minor, W. (1997). Processing of X-Ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
34. Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallog. sect. D*, **62**, 48–57.
35. CCP4. (1994). The CCP4 suite: programs for protein cristallography. *Acta Crystallog. sect. D*, **50**, 760–763.
36. Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallog. sect. D*, **57**, 1367–1372.
37. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
38. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
39. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallog. sect. D*, **60**, 2126–2132.
40. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of a protein structure. *J. Appl. Crystallog.* **26**, 283–291.
41. DeLean, A., Munson, P. J. & Rodbard, D. (1978). Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. *Am. J. Physiol.* **235**, E97–E102.

Edited by J. Doudna

(Received 13 March 2007; received in revised form 2 July 2007; accepted 5 July 2007)
Available online 12 July 2007

The N-Terminal Domain of the Arenavirus L Protein Is an RNA Endonuclease Essential in mRNA Transcription

Benjamin Morin[1], Bruno Coutard[1], Michaela Lelke[2], François Ferron[1], Romy Kerber[2], Saïd Jamal[1], Antoine Frangeul[1], Cécile Baronti[3], Rémi Charrel[3], Xavier de Lamballerie[3], Clemens Vonrhein[4], Julien Lescar[1,5], Gérard Bricogne[4], Stephan Günther[2], Bruno Canard[1]*

1 Architecture et Fonction des Macromolécules Biologiques, CNRS and Universités d'Aix-Marseille I et II, UMR 6098, ESIL Case 925, Marseille, France, **2** Department of Virology, Bernhard-Nocht-Institute for Tropical Medicine, Hamburg, Germany, **3** Unité des Virus Emergents UMR190, Université de la Méditerranée & Institut de Recherche pour le Développement, Marseille, France, **4** Global Phasing Ltd., Cambridge, United Kingdom, **5** School of Biological Sciences, Nanyang Technological University, Singapore, Singapore

Abstract

Arenaviridae synthesize viral mRNAs using short capped primers presumably acquired from cellular transcripts by a 'cap-snatching' mechanism. Here, we report the crystal structure and functional characterization of the N-terminal 196 residues (NL1) of the L protein from the prototypic arenavirus: lymphocytic choriomeningitis virus. The NL1 domain is able to bind and cleave RNA. The 2.13 Å resolution crystal structure of NL1 reveals a type II endonuclease α/β architecture similar to the N-terminal end of the influenza virus PA protein. Superimposition of both structures, mutagenesis and reverse genetics studies reveal a unique spatial arrangement of key active site residues related to the PD…(D/E)XK type II endonuclease signature sequence. We show that this endonuclease domain is conserved and active across the virus families *Arenaviridae*, *Bunyaviridae* and *Orthomyxoviridae* and propose that the arenavirus NL1 domain is the *Arenaviridae* cap-snatching endonuclease.

Citation: Morin B, Coutard B, Lelke M, Ferron F, Kerber R, et al. (2010) The N-Terminal Domain of the Arenavirus L Protein Is an RNA Endonuclease Essential in mRNA Transcription. PLoS Pathog 6(9): e1001038. doi:10.1371/journal.ppat.1001038

Editor: Michael J. Buchmeier, University of California Irvine, United States of America

Received April 7, 2010; **Accepted** July 12, 2010; **Published** September 16, 2010

Funding: Toscana virus and H5N1 influenza endonuclease domains were produced as part of the European Virus Archive (EVA) project (European FP7 Capacities Project no 228292, http://www.european-virus-archive.com/). This work was supported in part by the VIZIER integrated project (LSHG-CT-2004-511960) of the European Union 6th Framework Programme (FP6), the Direction Générale de l'Armement (DGA 09cà402), the Fondation pour la Recherche Médicale (FRM-Aide aux équipes), the German Research Foundation grant GU 883/1-1, and a grant from the "Vereinigung der Freunde des Tropeninstituts Hamburg e.V." to ML. JL was supported by grant 08/1/22/19/589 from BMRC and an ATIP from CNRS. We gratefully acknowledge the ID14-4 beamline staff and equipment at the European Syncrotron Radiation facility, Grenoble, France. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: bruno.canard@afmb.univ-mrs.fr

Introduction

The *Arenaviridae* family includes 22 viral species into a single genus Arenavirus, with new species awaiting classification [1,2]. They cause chronic and asymptomatic infections in rodents, and occasional transmission to man may result in life-threatening meningitis and/or hemorrhagic fever. *Lymphocytic choriomeningitis virus* (LCMV) is the prototypic species and first arenavirus isolated in 1933. Because its natural host is the common house mouse (*Mus musculus*), LCMV is the only known arenavirus presumably exhibiting a worldwide distribution. LCMV is a human pathogen of significant clinical relevance, causing central nervous system disease, congenital malformation, choriomeningitis, and systemic and highly fatal infection in immuno-compromised, organt transplant recipient patients [3,4,5,6]. Humans are generally infected through the respiratory tract after exposure to aerosols, or by direct contact with infectious material.

Arenaviruses are enveloped viruses with a bisegmented negative single-strand RNA genome. Each RNA segment, called large (L; ~7.2 kb) and short (S; ~3.5 kb), contains two open reading frames in mutually opposite orientations and use an ambisense coding strategy to direct the synthesis of two polypeptides [7]. Between the two open reading frames of each segment resides a non-coding intergenic region (IGR), composed of a sequence predicted to form a stable hairpin structure [8]. The S RNA encodes the viral nucleoprotein (NP; ~63 kDa) and glycoprotein precursor (GPC; ~75 kDa), whereas the L RNA encodes a small RING finger protein (Z; ~11 kDa) and a large protein (L; ~250 kDa) which is the viral RNA-dependent RNA polymerase (RdRp). The two RNA genomes are encapsidated by the NP, which is the most abundant protein in virions and infected cells, and act as templates for two fundamentally different processes, RNA replication and transcription. During RNA replication, the L protein first binds to the 3′-end of RNA templates and reads them from end to end to direct the synthesis of encapsidated full-length anti-genomes. During transcription, the RdRp stops RNA synthesis at a pause site located near the IGR [7]. The newly synthesized mRNA molecules have a non-polyadenylated 3′-end with a heterogeneous sequence mapped within the predicted hairpin in the IGR [9]. Furthermore, non template-directed sequences have been identified at the 5′-end of the subgenomic mRNA [10]. These sequences are variable in length [9,10,11] and terminate with a 5′-cap structure, which suggests the presence of a cap-snatching mechanism for arenaviruses. In this process, originally described for influenza viruses [12,13] and

Author Summary

The *Arenaviridae* virus family includes several life-threatening human pathogens that cause meningitis or hemorrhagic fever. These RNA viruses replicate and transcribe their genome using an RNA synthesis machinery for which no structural data currently exist. They synthesize viral mRNAs using short capped primers presumably acquired from cellular transcripts by a 'cap-snatching' mechanism thought to involve the large L protein, which carries RNA-dependent RNA polymerase signature sequences. Here, we report the crystal structure and functional characterization of an isolated N-terminal domain of the L protein (NL1) from the prototypic arenavirus: lymphocytic choriomeningitis virus. The NL1 domain is able to bind and cleave RNA. The 2.13 Å resolution crystal structure of NL1 reveals a type II endonuclease α/β architecture similar to the N-terminal end of the influenza virus PA protein. Superimposition of both structures and mutagenesis studies reveal a unique spatial arrangement of key active site residues related to the PD…(D/E)XK type II endonuclease signature sequence. Reverse genetic studies show that mutation of active site residues selectively abolish transcription, not replication. We show that this endonuclease domain is conserved and active across the virus families: *Arenaviridae*, *Bunyaviridae* and *Orthomyxoviridae* and propose that the arenavirus NL1 domain is the *Arenaviridae* cap-snatching endonuclease.

bunyaviruses [14], the viral RdRp binds cellular mRNAs caps and 'steals' them using an endonuclease activity, located in the influenza PA subunit [15,16], and presumably in L protein of bunyaviruses. These short capped RNAs are then used as primers for mRNA synthesis. The arenavirus L protein is an essential element in genome replication and transcription [17]. It is the largest viral protein composed of approximately 2200 amino-acid (aa) residues, and sequence analysis using homologous proteins led to the prediction of several conserved domains [18,19]. A biological function can be inferred for the L3 domain containing conserved and typical RdRp signature sequence motifs [19,20]. For Tacaribe virus, both domains L1 and L3 interact with the Z protein [21]. By analogy with influenza and bunyaviruses, the L protein may also carry activities and domains responsible for a cap-snatching mechanism that would account for the sequence diversity found at the 5′-end of RNA transcripts. The expression and purification of such a large viral polymerase is problematic and has not been documented.

We report here the first crystal structure of an *Arenaviridae* L protein domain at 2.13 Å resolution, that of the N-terminus domain of the LCMV L protein. We show that this domain is able to bind nucleotides, with a preference for UTP, and RNA. Structural comparison with the N-terminal part of the influenza virus PA protein characterizes unambiguously the domain as an endonuclease. Sequence and secondary structure analysis of L proteins from various *Bunyaviridae* family members predict that their N-terminal end carries a similar endonuclease activity, that we demonstrate for Toscana virus (TOSV) (genus *Phlebovirus*, family *Bunyaviridae*). Activity assays and mutagenesis show that the arenavirus endonuclease exhibits sequence-specificity with a preference for uracil-containing substrates. Lastly, reverse genetics studies correlate expression of endonuclease activity with the selective production of mRNA, making the N-terminus domain of the L protein a likely candidate to be involved in the cap-snatching mechanism of arenaviruses.

Results

Delineation of an Arenavirus L Protein Domain and its Crystal Structure

Based on aa sequence conservation across arenaviruses and on the presence of a potential nucleotide-binding site, we designed cDNA constructs encoding aa residues 1 to ~250 for the N-terminal end of four arenavirus (Pirital virus (PIRV), Lassa fever virus (LASV), Parana virus (PARV), and LCMV) L proteins. All four domains were expressed as soluble recombinant proteins. We observed a self-limited proteolysis of the Parana arenavirus N-terminus L domain which prompted us to refine boundaries into a shorter 196 aa form, hereafter named "NL1", fully included in the previously predicted arenavirus L1 domain (1–250 aa) [19]. The construct was expressed in *E.coli* and purified, but yielded crystals diffracting to 8 Å. However, the homologous 196 residues domain of LCMV yielded well-diffracting crystals. The atomic structure of NL1 was first determined by the SAD technique with seleno-methionylated crystals that diffracted to 3.4 Å. The structure was refined using a native data set at 2.13 Å resolution (Table 1). Two NL1 molecules are present within the asymmetric unit. Residues 1–191 are visible for one molecule whilst only 1–175 could be modelled for the other NL1 molecule owing to high mobility of the C-terminal end of helix α7.

The LCMV NL1 Domain Exhibits a Type II Endonuclease Fold

The LCMV NL1 monomer structure has approximate dimensions of 59 Å ×37 Å ×27 Å. It features four mixed β-strands forming a twisted plane surrounded by seven α-helices (Figure 1A). The two anti-parallel strands β1 and β2 are connected by helix α4, whereas the two parallel strands β3 and β4 are

Table 1. Data Collection and Refinement Statistics.

Data collection and refinement statistics	LCMV NL1
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	145.0, 159.3, 52.6
Resolution (Å)	107.25−2.13 (2.19−2.13)*
Measured/unique reflections	231,821/33,999
R_{merge}	3.9 (45.2)
$<I/\sigma I>$	27.1 (1.9)
Completeness (%)	97.9 (97.8)
Redundancy	6.8 (2.9)
Resolution (Å)	72.50−2.13
No. reflections	33,931
R_{work}/R_{free}(%)	19.86/22.21
No. atoms	3,297
Protein	3,011
Water	253
B-factors	
Protein	57.6
Water	68.2
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.14

*Values in parentheses are for the highest-resolution shell.

doi:10.1371/journal.ppat.1001038.t001

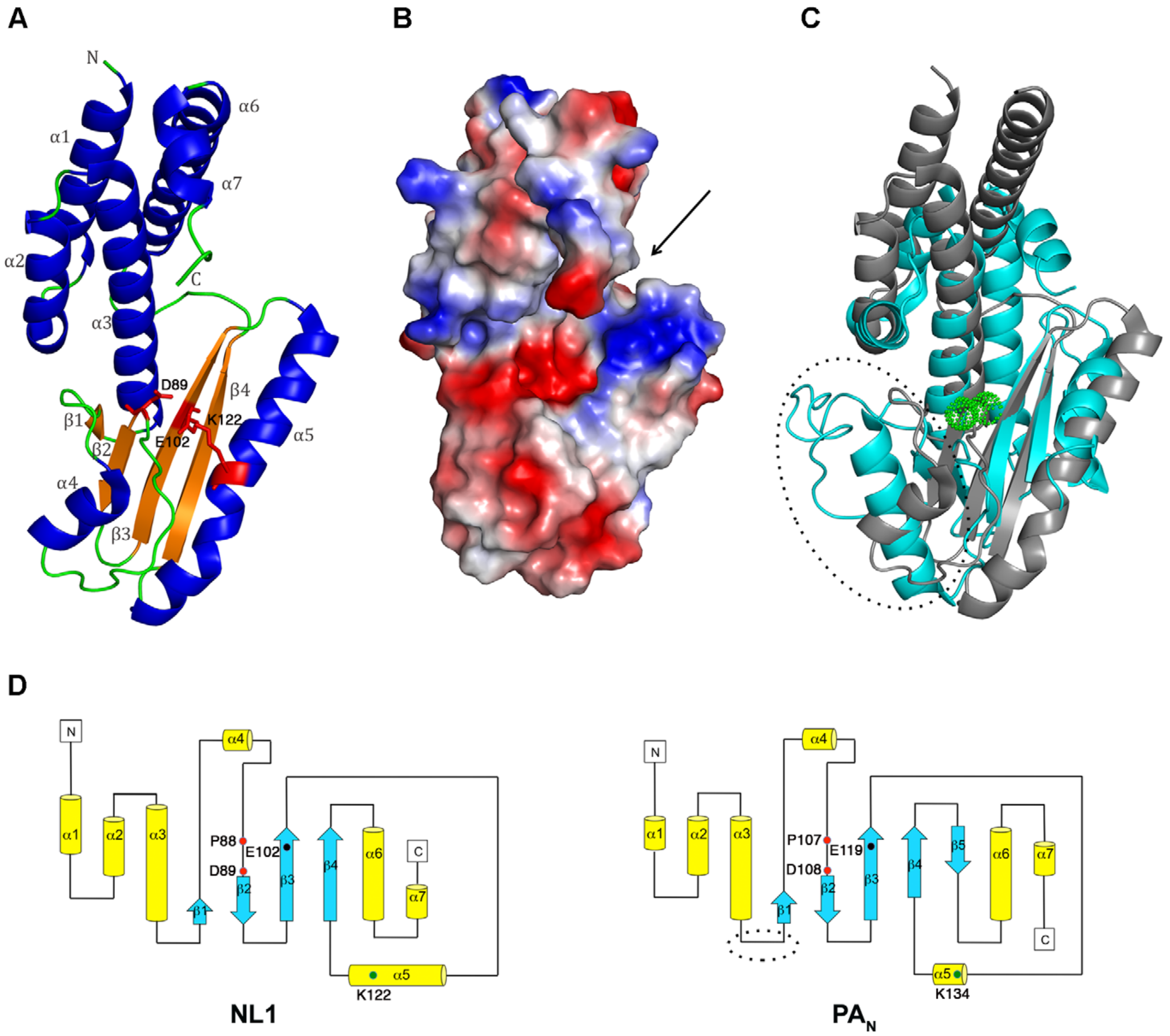


Figure 1. NL1 structure and comparisons with the influenza PA$_N$ structure. A, Cartoon-representation of the NL1 structure. Secondary-structure elements are labelled and colored as follows: α-helices, blue, β-strands, orange, and loops, green. Side-chains for residues within the NL1 endonuclease active site are shown as red sticks and labelled. **B**, Electrostatic surface representation showing NL1 in the same orientation as in panel (**A**). The arrow indicates the putative RNA binding groove and the active site crevice. Negative charges are in red and positive charges in blue and neutral in white. **C**, Superimposition (view in the same orientation as in **A**) of the structures of NL1 (grey) and PA$_N$ (PDB code:2W69, cyan) highlighting their shared structural core as well as variations in the form of an extra loop only present in the PA$_N$ structure (circled). The two Mn^{2+} ions in the PA$_N$ structure active site are depicted as green spheres. **D**, Topology diagrams of the NL1 (left) and PA$_N$ (right) structures. α-helices are represented as yellow tubes and β-strands are blue arrows. The extra-loop of PA$_N$ protein is circled as in panel **C**. Key residues from the endonuclease active site (PD, E/D, and K), are schematically depicted by colored dots and labelled, highlighting the fact that they project from conserved structural elements between the influenza PA$_N$ protein and the arenavirus NL1 domain.
doi:10.1371/journal.ppat.1001038.g001

connected by the long helix α5. These two helices run parallel to the central β-sheet and are disposed at the same side of the latter. On the opposite side of the β-sheet, helix α3 is surrounded at its extremity by two N-terminal (α1 and α2) and C-terminal helices (α6 and α7). A search for similar protein folds using the DALI server [22] returned the PA N-terminal domain structure that was recently identified as a type II endonuclease domain [15,16]. The structural match with published molecular structures of the influenza PA N-terminal domains (PA$_N$) returns a Z-score of 5.7 and an r.s.m.d. of 3.9 Å for 121 superposed aa (PDB code 3EBJ) and Z-score 5.2, r.m.s.d. 4 Å for 122 aa (PDB code 2W69). As was the case for PA$_N$, other type II endonuclease proteins are also recovered: the Tt1808 hypothetical protein from *Thermus Thermophilus* HB88 (PDB code 1WDJ, Z-score 3.8, r.m.s.d. 3.4 Å for 81 aa), and the restriction endonuclease *SdaI* (PDB code 2IXS, Z-score 3.6, r.m.s.d. 6.3 Å for 104 aa).

The β-sheet forms a negatively charged cavity creating a binding site for divalent cations, whilst above that cavity, the C-terminal end of helix α5 forms a positively charged patch and a concave surface that is likely to accommodate the RNA substrate (Figure 1B, arrow). The PA protein constitutes one subunit that associates with PB1 and PB2 to form the heterotrimeric influenza virus polymerase. Its N-terminal domain PA_N hosts the RNA cap-snatching endonuclease activity [15,16]. Both NL1 and PA_N share a similar core structure. Except for the absence of a fifth β-strand in NL1, all other secondary structure elements are conserved (Figure 1C) and the overall topology of these two structures is very similar (Figure 1D), albeit with interesting differences in the vicinity of the PA_N active site (discussed below). At the aa sequence level, NL1 shares the conserved active site sequence motif characteristic of type II endonucleases: PD…(D/E)XK. In NL1, the corresponding residues are P88, D89…E102, and either K115 or K122 (Figure S1A, B). The identity of the distal lysine is not certain since it is found at different positions in the primary sequence, as is the case for influenza virus. The influenza PA_N domain was crystallized either in the presence of magnesium or manganese ions in the active site which comprises five conserved catalytic residues: H41, E80, D108, E119 and K134. A structural superimposition of the arenavirus NL1 and influenza PA_N active sites shows that the side-chains of three evolutionary-conserved residues within arenaviruses (P88, D89 and E102) closely superimpose with P107, D108 and E119 of the influenza virus PA_N protein, pointing to a common function for these residues (Figure 2A and Figure S1B). Upon superimposition with PA_N, one Mn^{2+} ion needed for the enzymatic reaction coordinated by D108 in the PA_N active site, falls at right distances to be coordinated by the carboxylate side-chains of D89 and E102. NL1 was crystallized without metal ions and a water molecule is found close to the position that should be occupied by the divalent metal. Interestingly, no close structural match is found neither for H41 nor K134 of the influenza virus PA_N. This points to differences between the two active sites since His41 was proposed to play a catalytic role in the influenza PA_N. However, we note that another possible contributor could be NL1 C103 main-chain carbonyl as it superimposes quite well with PA_N I120 main-chain carbonyl (Figure 2B). The triad made of K115, D119, and K122 in NL1 is spatially equivalent to K134 in PA_N. In summary, despite no aa sequence homology, the active site structures of the influenza PA_N and LCMV NL1 domains are clearly related but not identical (Figure 1C, 2), strongly suggesting that these two domains exhibit closely related enzymatic activities (see below).

The NL1 Endonuclease Fold is Conserved Amongst *Bunyaviridae*

In addition to *Arenaviridae* and *Orthomyxoviridae*, *Bunyaviridae* is the other family of virus to possess a segmented negative-strand RNA genome. It contains four genera of animal viruses (*Orthobunyavirus, Phlebovirus, Nairovirus, Hantavirus*) and one genus of plant virus (*Tospovirus*) [23]. Although the genomic organisation differs between these three virus families, *Bunyaviridae* are also thought to use a cap-snatching mechanism to prime mRNA synthesis [24]. Arenaviruses, and *Bunyaviridae* share a conserved RdRp motif within their large L protein, as well as a conserved N-terminus domain [18]. Amino-acid sequence alignments, assisted by secondary structured prediction, of the N-terminal part of LCMV and *Bunyaviridae* L protein reveal that the latter also possesses the conserved active site motifs characteristic of type II endonucleases (Figure S2A). However, we could identify the catalytic motif within the L protein N-terminal end for only four out of the five bunyavirus genera: *Orthobunyavirus*, *Phlebovirus*, *Hantavirus* and *Tospovirus*. The L protein of *Nairovirus* is much larger (~4000 aa) than the L protein of other members of the *Bunyaviridae* family (~2200 aa). The putative endonuclease catalytic motif was located after aa ~700, the N-terminal of *Nairovirus* L protein being assigned as a so-called OTU-like domain [25].

Secondary structure predictions were used to draw the topology diagram of the NL1-like domain for each genera (Figure S2B). As expected from the sequence alignment, each genus seems to share a β-sheet with a variable number of β-strands. Furthermore, the PD catalytic motifs are in each case located in a loop before a β-strand, as expected. The PUMV, HLCV and RVFV NL1-like domains are more closely related to LCMV NL1 than are the TOMV and CCGV. The TOMV NL1-like domain contains 6 β-strands and shares the PD motif just upstream the first β-strand,

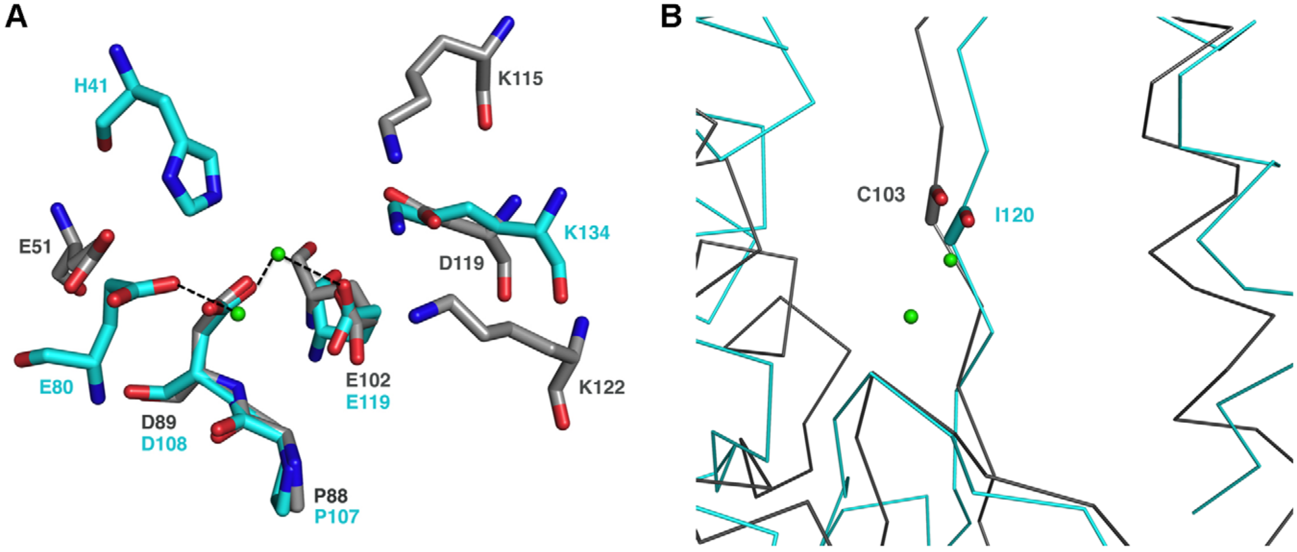


Figure 2. The endonuclease active site. A, Structure-based superimposition of the endonuclease active site from the influenza PA_N protein and the arenavirus NL1 domain. Putative active site residues of NL1 are shown as grey sticks and the active site of PA_N (PDB code 2W69) in cyan. The two Mn^{2+} ions present in the PA_N structure (but not in the present NL1 domain crystal structure) are shown as light green spheres with their closest ligand indicated by a dashed line. **B**, C-alpha trace ribbon-representation of the superimposition of the endonuclease active site from the influenza PA_N (cyan) protein and the arenavirus NL1 domain (grey). The carbonyl main-chain of PA_N I120 and NL1 C103 are shown in sticks. The metal ions are shown as light green spheres.
doi:10.1371/journal.ppat.1001038.g002

whereas it is just upstream the second β-strand in the case of NL1 and PA_N. Finally, the structural organization of the putative CCGV endonuclease domain seems to diverge even further from the others. Indeed, whereas the conserved lysine is shared by the same helix for all the domains, that of *Nairovirus* may be located at the end of the β4 strand (Figure S2B). Thus we conclude that the endonuclease motif is conserved across four animal virus genera *Orthobunyavirus*, *Phlebovirus*, *Nairovirus* and *Hantavirus*.

NL1 is a Mn^{2+}-Dependent RNA Endonuclease

Recent crystal structures of complexes of PA_N with three different nucleoside monophosphates show that PA_N binds nucleotides [26]. The ability of NL1 to bind nucleotides was investigated using UV-crosslink experiments. We observe that NL1 binds NTPs, preferably UTP and GTP, whereas ATP and CTP show a weaker association (Figure 3A). The PA_N structures were determined in complex with ATP, CTP and UTP but not GTP [26] whereas NL1 bind GTP in a stronger fashion than ATP or CTP. The crystal structure relatedness to the endonuclease fold would suggest that the NL1 domain is able to bind RNA rather than nucleotides. We tested RNA binding by NL1, and found that indeed, NL1 binds RNA (Figure 3B). The band shift assay is also suggestive that the RNA substrate is cleaved under the assay conditions, as judged by degradation products at the bottom of the gel under the labeled RNA oligo (Figure 3B). Therefore, we surmise that nucleotide binding properties observed here reflect the ability of NL1 to bind RNA with some sequence specificity in the cap-snatching pathway (see below).

Several synthetic RNA oligonucleotides were used to characterize the endonuclease activity (Figure 4). NL1 is able to cleave ssRNA having no stable secondary structure at specific sites indicating a preference for the presence of uracil (Figure 4A, B), and adenosine to a lesser extent. Likewise, a moderately stable RNA hairpin containing uracil ($\Delta G = -3.4$ kcal/mole) is cleaved down to a 14/15-mer product whereas a stable ($\Delta G = -14.7$ kcal/mole) RNA hairpin devoid of uracil remains unattacked even in its single stranded regions (Figure 4A, B). PolyU RNA is cleaved randomly down to a 8-mer product with a better efficiency than polyA, whereas polyC is not a substrate for NL1 (not shown). A 5′-terminal nucleoside uracil or adenosine 5′-monophosphate is also cleaved and the 5′-monophosphate RNA end apparently competes for internal cleavage. A 5′-capped RNA of 264 nucleotides in length also acts as a substrate. It is cleaved at several specific positions indicated by the sequential appearance of band products over time (Figure 4B). This indicates that the cap structure does not seem to be a direct RNA binding determinant. A *Phlebovirus* (Toscana) virus endonuclease domain was prepared according to bio-informatic predictions described above. Its endonuclease activity was compared to both that of arenavirus NL1 and the influenza H5N1 endonuclease [16]. The enzymes were equally active using short RNA substrates, although it is apparent that sequence-specific cleavage is different for each enzyme: the influenza enzymes prefers cleavage at puric sites, Toscana virus and LCMV enzymes prefer adenosine- and uracil-containing sites (Figure 4B). NL1 is ~90-fold more active in the presence of Mn^{2+} than Mg^{2+}, and shows background activity with Ca^{2+} and Zn^{2+}(Figure 4C and not shown). The Mn^{2+} ion has also a significant stabilizing effect as judged by thermostability studies, whereas Zn^{2+} has a deleterious effect.

Mutagenesis analysis of most residues identified as part of the active site (Figure 2A) impaired the endonuclease activity. The most drastic effect was observed for D119, but residual activity was scored for E51, D89, and less for E102 (Figure 4D). As these three residues might coordinate metal ions as proposed above, defective

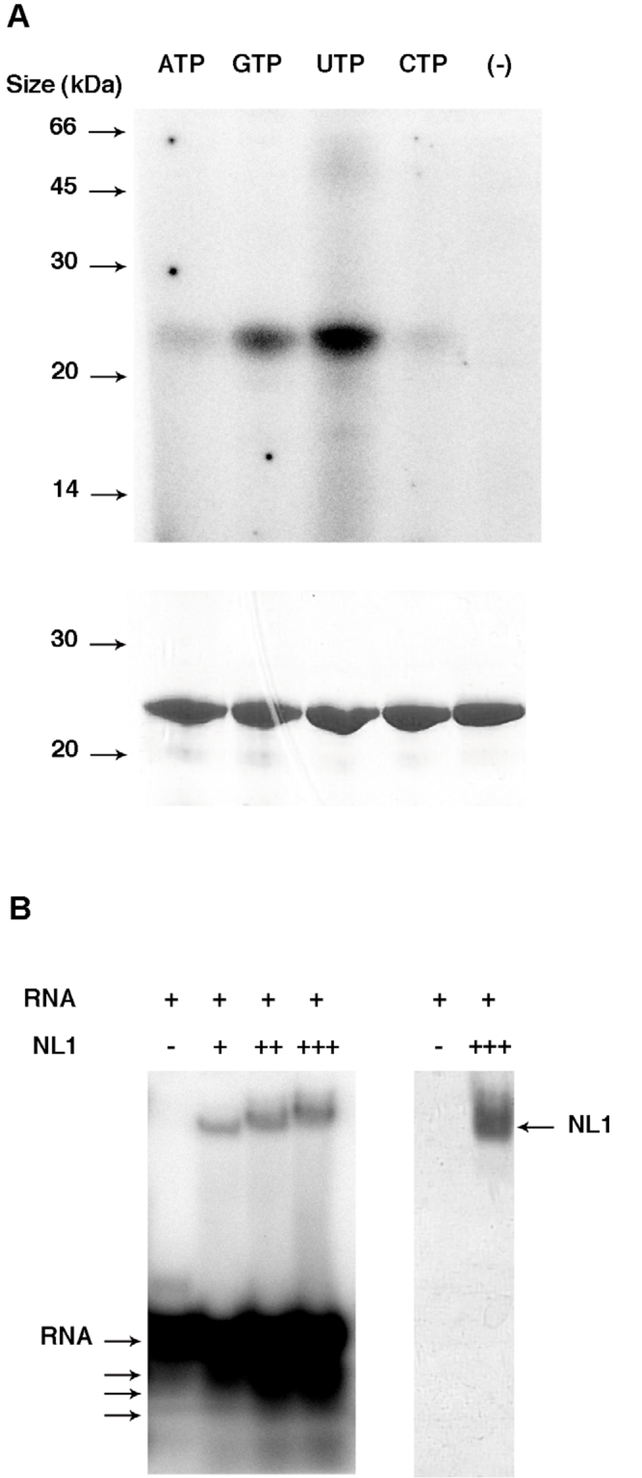


Figure 3. Nucleotide and RNA binding assays of LCMV NL1 domain. A, Cross-linking assay. 7 µg of purified protein were incubated in the absence (−) or presence of each indicated radiolabelled NTP. The mixture was then UV-irradiated and loaded onto a denaturing polyacrylamide gel. The latter was analyzed by autoradiography (top) and Coomassie blue staining (bottom). **B**, Band shift assay. Radiolabelled RNA was incubated with increasing quantities (1.4 µg (+), 4.2 µg (++) and 7 µg (+++)) of NL1 protein. Reaction mixture was then analyzed by PAGE, and the gel was visualized by autoradiography (left) and Coomassie blue staining (right). Apparent degradation products are indicated by arrows under the RNA input arrow.
doi:10.1371/journal.ppat.1001038.g003

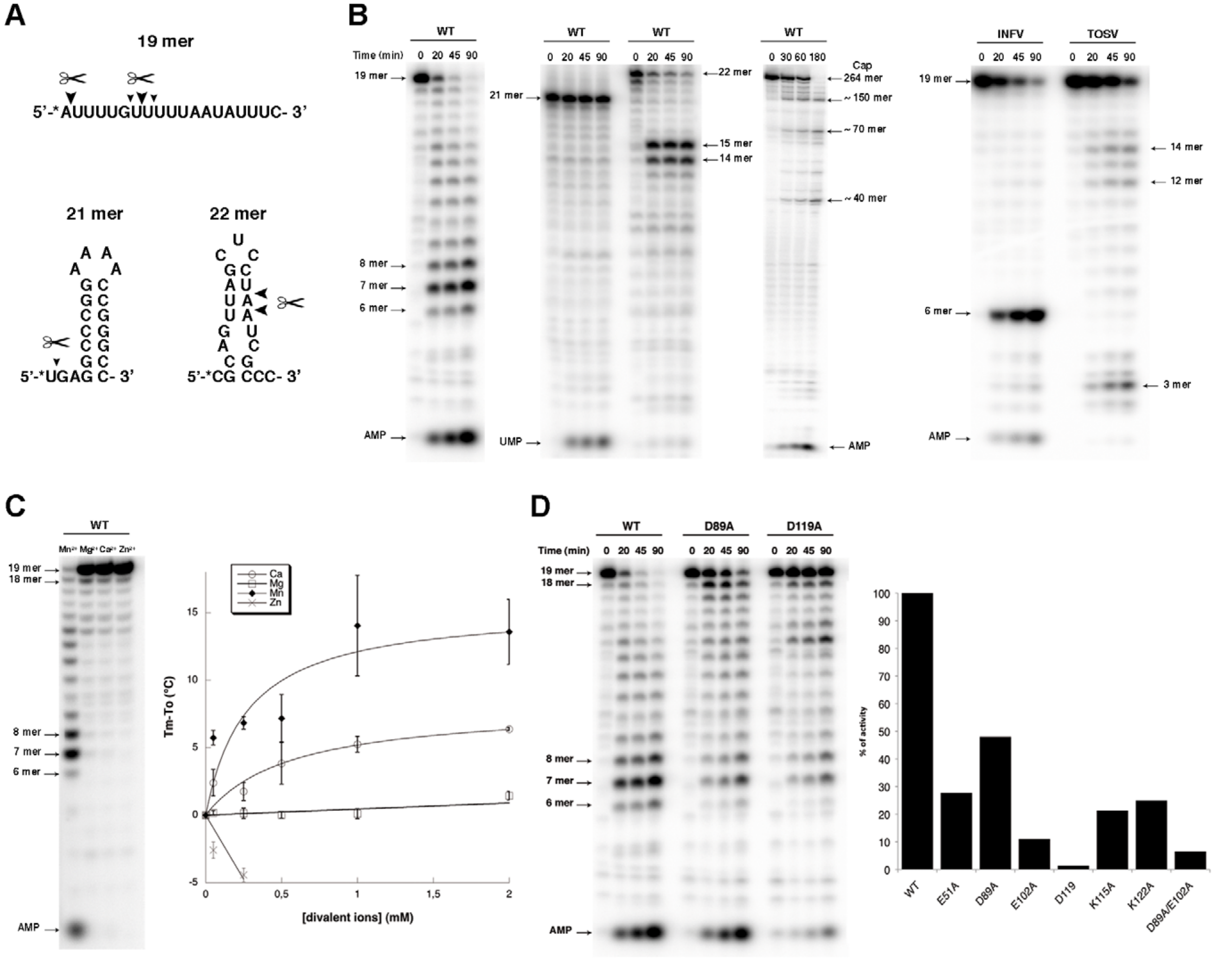


Figure 4. Endonuclease activity of NL1. A, Nucleotide sequence of the radiolabelled RNA used for the activity assays. The * indicates the radiolabelled nucleotide. Big and small triangles indicates the primary and secondary cleavage site by wild-type (WT) NL1, respectively. **B**, Kinetics of endonuclease activity of WT NL1 on different substrates (left), and for Influenza and Toscana virus (INFV, TOSV) endonuclease domain (right). Activity assays were performed as described in Materials and Methods, using 3.3 μM of RNA and 1 μM of protein. Reactions were quenched by the addition of EDTA/formamide, and analyzed using 20% polyacrylamide/7M urea gel. Substrate and degradation product sizes are indicated. **C**, Divalent cations effect on the NL1 activity. The reaction was allowed to proceed during 45 min as described in Materials and Methods. The divalent cation assay (left) was run during 45 min without intermediate points. Titration of divalent ions on NL1 by thermal shift assay (right). T_m is the melting temperature of NL1 with the divalent ions ; T_o is the melting temperature of the protein alone. **D**, Mutational analysis of NL1 domain on the endonuclease activity. Kinetics were performed as described above, with WT, D89A and D119A mutants (Left). Graph showing the % of endonuclease activity determined using FujiImager normalized quantitation for the different mutants (right).
doi:10.1371/journal.ppat.1001038.g004

metal-binding due to a point mutation might be compensated by the presence of the remaining two adjacent acidic residues. A double mutant D89A/E102A shows further reduced but not abolished activity. Mutations K115A and K122A generated strongly altered activity, but the similar level of residual activity does not allow the identification of which lysine is predominant in catalysis.

The Endonuclease Activity is Essential for RNA Transcription, not Replication

The effect of 33 mutations in L1 on virus RNA and protein expression was studied in a cell-based mini-replicon system. The LCMV L protein mediates the synthesis of two RNA species: first, capped mRNA terminating within the intergenic region, and second, antigenomic RNA being a full-length copy of the genomic RNA template [9,27]. This dual role in RNA synthesis is recapitulated in the mini-replicon system. It contains all trans-acting factors (L protein and NP) required for transcription and replication of a genome analogue containing Renilla luciferase as a reporter gene (mini-genome). Reporter gene expression was measured in luciferase assay (Table 2), while RNA synthesis was measured in Northern blot (Figure 5), in which luciferase mRNA and antigenome can easily be distinguished due to their size difference. Wild-type (WT) L protein led to expression of high levels of Renilla luciferase (2–3 log units signal-to-noise ratio) as well as Renilla luciferase mRNA and antigenome in a ratio of about 1:1. Expression of mutant L protein was verified by immunoblotting (Figure S3).

The phenotype of mutants E41A, E41Q, K44A, S54A, C60A, T108S, F116A, D142N, and W155A is similar to that of wild-type.

Table 2. Functional Analysis of L Protein Mutants in LCMV Mini-Replicon System.

Mutant[1]	Renilla luciferase activity (sRLU)		RNA expression level (Northern blot signal)	
	% of wild-type[2]	Signal-to-noise ratio[3]	Antigenome level, % of wild-type[4]	mRNA-to-antigenome ratio, relative to wild-type[5]
E41A	44.3	243.2	21.9	2.01
E41Q	42.4	310.7	25.5	1.95
K44A	30.2	215.8	23.6	0.93
E51A	**1.3**	**8.5**	**28.5**	**0.39**
E51Q	**0.5**	**3.6**	**32.9**	**0.26**
S54A	23.5	175.6	28.3	0.80
C60A	47.9	357.7	62.4	1.13
D89A	**0.5**	**3.5**	**107.6**	**0.20**
D89N	**0.5**	**3.8**	**138.5**	**0.19**
E102A	**0.9**	**3.5**	**119.3**	**0.22**
E102N	**3.3**	**13.6**	**78.2**	**0.26**
F104A	0.6	3.8	—	—
R106A	4.4	32.5	—	—
T108S	29.7	212.7	48.0	0.89
F112A	0.3	1.8	—	—
K115A	0.9	6.8	—	—
F116A	27.4	209.4	39.1	1.79
D119A	**2.0**	**12.5**	**47.5**	**0.06**
D119N	**8.0**	**43.0**	**46.4**	**0.07**
K122A	**1.6**	**9.3**	**55.3**	**0.04**
D129A	**1.3**	**6.5**	**56.5**	**0.04**
D129N	**1.3**	**7.4**	**54.1**	**0.05**
D142A	2.8	17.6	—	—
D142N	42.8	270.6	66.1	0.93
R144A	2.3	10.2	—	—
W155A	34.7	203.3	91.0	0.91
R161A	3.4	17.6	—	—
E179A	40.0	239.4	109.9	0.46
E179Q	25.4	174.6	105.0	0.44
E182A	32.6	201.3	154.9	0.40
E182Q	29.3	184.9	103.1	0.42
Y183A	16.9	98.8	82.3	0.48
R185A	0.3	1.4	—	—

[1] Mutants with selective defect in mRNA synthesis are shown in boldface.

[2] Standardized relative light unit (sRLU) value (wild-type = 100%). Mean of ≥2 independent transfection experiments.

[3] sRLU value of mutant divided by sRLU value of negative control mutant containing a mutation in the catalytic site of the RNA-dependent RNA polymerase. Mean of ≥2 independent transfection experiments.

[4] Antigenome signals in Northern blots were quantified via intensity profiles (wild-type = 100%).

[5] RNA signals in Northern blots were quantified and the mRNA-to-antigenome signal ratio was calculated. The wild-type ratio was set at 1 for each experiment (i.e. the signal ratio of a mutant was normalized with the wild-type ratio) to render independent blots comparable.

doi:10.1371/journal.ppat.t002

Mutants E179A, E179Q, E182A, E182Q, and Y183A also express luciferase and RNA at high level, but the steady-state level of mRNA relative to that of antigenome is reduced by about 50%. Mutants F104A, R106A, F112A, K115A, D142A, R144A, R161A, R185A neither express Renilla luciferase nor any RNA species, indicating that global functions of L protein are affected.

The most interesting phenotype is observed with mutants D89A, D89N, E102A, E102N, D119A, D119N, K122A, D129A, and D129N. They synthesize antigenome close or equal to wild-type level, but are defective in mRNA and, thus, reporter gene expression (Figure 5 and Table 2, shown in boldface). A similar phenotype is seen with mutants E51A and E51Q, though associated with reduced antigenome level. These data indicate that residues E51, D89, E102, D119, K122, and D129A are essential for viral mRNA synthesis, but not required for expression of uncapped RNA species. With the exception of the D129 residue located at the surface of the protein remote from the endonuclease active site, it is remarkable that these transcription-null mutants form the catalytic site (Figure 2) and match precisely those of the PD…(D/E)XK endonuclease type II signature sequence.

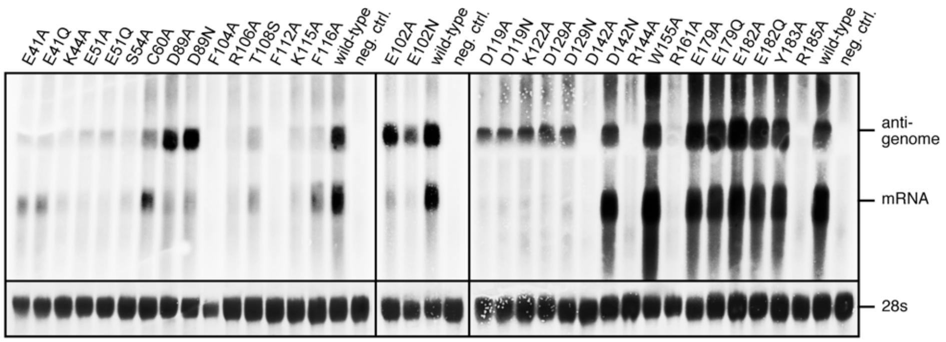


Figure 5. Mutational analysis of the L protein in the context of the LCMV replicon system. Synthesis of antigenomic RNA and Renilla luciferase mRNA was analyzed by Northern blotting. Negative control cells (neg. ctrl.) expressed mini-genome, NP, and an L protein mutant with a mutation in the catalytic site of the RNA-dependent RNA polymerase. The methylene blue-stained 28S rRNA is shown below the blots as a marker for gel loading and RNA transfer. Each panel represents an independent experiment with separate controls. Careful examination of the blots revealed residual signals at the mRNA position for some mutants negative in Renilla luciferase assay. Thus, these signals do not correspond to functional mRNA, but may be prematurely terminated antigenome.

doi:10.1371/journal.ppat.1001038.g005

Discussion

The structural and functional results presented here show that the LCMV NL1 domain is an RNA endonuclease. The uncoupling of RNA replication from transcription and selective disappearance of mRNA when NL1 active site residues are mutated strongly suggests that this activity is involved in cap-snatching.

The identification of the arenavirus endonuclease is in line with the recent discovery of the PA_N endonuclease domain of influenza virus. Whereas the active site of influenza virus features a cluster of three acidic residues, the active site of arenavirus contains four acidic residues (E51, D89, E102 and D119), as well as two important lysine residues K115 and K122 neighboring D119 (Figure 2A). The NL1 active site resembles but is clearly distinct from that of influenza PA_N. Indeed, there is no histidine in the catalytic center, and the arenavirus NL1 nuclease has some specific features both upstream and downstream of the PD signature sequence. We define the arenavirus endonuclease motif as $E\text{-}X_{38}\text{-}P\text{-}D\text{-}X_{(11,13)}\text{-}E\text{-}X_{12}\text{-}K\text{-}X_3\text{-}D\text{-}X_2\text{-}K$. The most obvious difference with the only known related RNA endonuclease, that of influenza virus PA_N, is a divergence upstream the PD motif in structural elements carrying the E51 residue (Figure 1C), and the presence of a triad K…D…K at the distal side of the latter signature sequence (Figure 2A). Contrary to PA_N which shares a conserved and essential histidine involved in the binding of both the metal ion and a nucleotide onto helix $\alpha 3$ [15,16,26], NL1 does not possess this conserved histidine residue. Instead, NL1 has a glutamic acid residue E51, which might reflect a different nucleobase specificity as detected in our nuclease assays (Figure 4). Likewise, residues downstream the PD motifs are distinct from the consensus sequence, and differently organized into a triad including two lysines. The presence of water molecules and previous structural models for influenza PA_N allows to propose putative positions of metal ions, coordinated by D89 and E102.

The first step in the general mechanism for phosphodiester hydrolysis is the preparation of the attacking nucleophile by deprotonation, usually involving a general base deprotonating a water molecule. Lysine is often considered as this general base candidate in endonucleases but is not strictly conserved [28,29]. Here, there are no indications against D119 being this general base. Alternately, it could well be either lysine K115 or K122. Both are oriented towards the active site, and they could well have their pKa lowered by D119 in order to initiate the reaction. Reverse genetic studies provide evidence for K122, not K115. Indeed, mRNA production is selectively abolished and clearly uncoupled from RNA synthesis in the case of K122A mutant, while the K115A mutant was completely defective preventing interpretation of its role in the endonuclease catalytic site. Although it is not known if uncapped mRNAs are synthesized and degraded for the transcription-null mutants, the most plausible scenario is that primer shortage prevents significant capped mRNA synthesis. Overall, the replicon data presented here closely match those obtained on the closely related Lassa arenavirus using a similar replicon system [30]. Arenaviruses may thus use two clearly independent and distinct RNA synthesis priming mechanisms: one is dependent on an active endonuclease carried by the N-terminus of the L protein, and the other might be linked to the observation that an extra G residue is found at the 5′-end of arenavirus genomes and antigenomes. The latter G bases would thus reflect a yet-uncharacterized priming mechanism unrelated to the U/A cleavage sequence preference of NL1.

NL1 also binds nucleotides, but the NTP binding site should differ from that of PA_N. Indeed, the influenza PA_N histidine 41 is involved in binding the nucleobase of the presumed incoming RNA substrate. The NL1 endonuclease does not share the same sequence specificity, and E51 is positioned at a spatially equivalent position.

The cap structure does not seem to be a direct RNA binding determinant (Figure 4B), as endonucleolytic cleavage is not directed to cleavage sites preferentially in the vicinity of the cap. We thus infer that an independent cap-binding site way exist elsewhere in viral proteins to bind and select cellular mRNAs, a possibility reminiscent of influenza for which PA carries the endonuclease activity and PB2 the cap binding site [15,16,31].

Structure and sequence alignment studies show that the N-terminal endonuclease domain of the L protein is also conserved in the *Bunyaviridae* family, although the *Nairovirus* endonuclease domain is not located into the N-terminal end of the protein. These findings were confirmed by the endonuclease activity of the N-terminal end of the L protein of TOSV (Figure 4B). Thus, we provide evidence that all three segmented negative single-strand RNA virus species share an endonuclease domain probably involved in the cap-snatching process during the viral life cycle. These data raise the question of a possible common ancestor for these viruses. Indeed, these three virus families use a cap-snatching mechanism involving binding and cleavage of cellular mRNA caps subsequently used by a large primer-dependent RNA-dependent RNA polymerase. It seems more plausible that the L gene has evolved by divergence over time, rather than by multiple acquisitions of several activities converging into a common structure, at least in the case of the endonuclease. Furthermore, our study raises the interesting possibility that other activities involved in RNA replication/transcrition might be discovered by comparative analysis of *Orthomyxoviridae* PB1, PB2, PA and *Arenaviridae*/*Bunyaviridae* L proteins.

To our knowledge, a single crystal structure of a functional arenavirus protein is currently available, that of the Machupo virus glycoprotein GP1 in complex with its human receptor, TfR1 [32]. Our results provide an arenavirus L domain structure, with a role consistent with the hypothesis of a cap-snatching mechanism suggested for arenaviruses [9,10]. The strategy used here to produce individually active domains might be useful to further characterize the *Arenaviridae*/*Bunyaviridae* large L protein which had so far resisted all biochemical characterization attempts.

The influenza, *Arenaviridae* and *Bunyaviridae* endonucleases are so far the only three examples of RNA endonucleases similar to type II DNA restriction endonucleases. The presence of such an endonuclease suggests that it could serve as a fruitful target for antiviral strategies against these two families, since such kind of inhibitors have been reported in the case of the influenza virus [33,34,35].

Materials and Methods

Cloning, Expression and Purification of LCMV NL1 Domain

The LCMV NL1 cDNA (Armstrong strain, aa 1 to 196) was cloned into pDest14 with a N-terminus hexa-histidine tag and expressed in *E.coli* Rosetta (DE3) pLysS (Novagen), at 17°C in 2YT medium overnight after induction with 500 μM IPTG. Cell pellets from harvested cultures were resuspended in 50 mM Tris buffer, pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.1% Triton, 5% Glycerol. Lysozyme (0.25 mg/ml), PMSF (1 mM), DNase I (2 μg/ml), and EDTA free protease cocktail (Roche) were added before sonication. IMAC chromatography of clarified lysates was performed on a 5 ml His prep column (Akta Xpress FPLC system, GE Healthcare) eluted with imidazole. Size exclusion chromatography was performed on preparative Superdex 200 column (GE Healthcare) pre-equilibrated in 10 mM Imidazole, pH 8.0, 50 mM NaCl, 2 mM DTT. Protein was concentrated (28 mg/ml) using a centrifugal concentrator. For enzymatic studies, WT and mutants were express in the *E.coli* BL21 star strain (Invitrogen) and further purified on HiTrap Q sepharose 1 ml column (GE Healthcare) to remove *E. coli* RNase contaminants. Proteins eluted in a linear gradient from 50 mM to 1 M NaCl in 10 mM Hepes buffer, pH 7.5, 2 mM DTT. A synthetic gene of the H5N1 PA_N endonuclease was designed as described [16]. The Toscana virus (strain France AR_2005, aa 2 to 233) cDNA was obtained from infected cell cultures. Both ORFs were cloned as a N-terminal Thioredoxin-Hexahistidine fusion in pETG20A. The tag was cleaved using TEV protease before a final gel filtration.

Crystallization

Crystals grew in $LiSO_4$ 250 mM, citrate 50 mM, isopropanol 5.5%, using the hanging drop vapor diffusion method in Linbro plates by mixing 1 μl of protein solution with 1 μl of reservoir solution. Crystals were cryoprotected by dipping in a solution containing 65% of crystallization buffer and 35% of a buffer made of size exclusion chromatography buffer/glycerol (50/50). Crystals were cryo-cooled in liquid N_2. The crystals belong to space group $C222_1$ and have two molecules per asymmetric unit. Despite repeated attempts, crystal soaked into the above buffer supplemented with various concentrations of $MnCl_2$ yielded crystals diffracting to >4 Å.

Data Collection and Structure Determination

Diffraction intensities were recorded on the ID14-4 beamline at the European Synchrotron Radiation facility, Grenoble, France. Data were processed and integrated with MOSFLM [36]. Scaling and merging of the intensities was performed with SCALA and programs from the Collaborative Computational Project, No. 4 (CCP4) suite [37]. The structure was determined using SAD data from one selenomethionylated protein crystal diffracting to 3.4 Å resolution with SHARP/autoSHARP, followed by density modification with SOLOMON and DM. An initial model was built using BUCCANEER and completed in COOT, followed by refinement using BUSTER (see Text S1). Details of structure determination are given as supplemental material. Data from a native crystals diffracting to a 2.13-Å resolution were collected on an ADSC QUANTUM 315r at a wavelength of 0.9835 Å. The structure was refined with BUSTER and COOT using this data set (Table 1) [38]. The atomic coordinates have been deposited at the PDB (3JSB).

Sequence Retrieval

A PHI-BLAST search using the sequence corresponding to the L1 domain and the signature of the *Arenaviridae* endonuclease motif *i.e.* P-D-$_{x(11,13)}$-E-$_{x(12)}$-K-$_{x(3)}$-D-$_{x(2)}$-K ; was performed against non-redundant databases [39]. After 3 iterations, Batai and Kairi viruses both belonging to *Orthomyxoviridae*, appears in the section with an E-Value below threshold. A fourth iteration including these two sequences allows retrieving the entire family of orthomyxoviruses, with E-value comprised between $3e^{-18}$ and $2e^{-4}$.

A standard CDD search from the sequence of Tensaw virus allows retrieving all the L of the *Bunyaviridae* family hitting the pfam 04196 [40].

Sequence Comparison

A multiple sequence alignment of the N-terminal end of the L protein from LCMV, HLCV, BUNV, HANV, PUMV, RVFV, TOSV, TOMV, WTMV, CCGV, DUGV, was first performed with the T-coffee algorithm (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi). Using the secondary structure prediction of the endonuclease domain of L proteins, the putative conserved active site residues were identified and placed correctly in the alignment.

UV-Crosslink Experiments

7 μg of purified protein were incubated for 15 min at 25°C, with 0.5 μl of the various α-^{32}P NTP (0.4 μCi/μl) in 10 μl of reaction buffer containing 10 mM Imidazole, pH 8.0, 50 mM NaCl, 2 mM DTT. The reaction mixtures were then exposed to UV light (254 nm) for 6 min at 5 mm distance. The crosslinked species were separated in a 15% polyacrylamide denaturing gel and visualized by autoradiography using photo-stimulated plates and a FujiImager (Fuji).

RNA Binding Experiments

The RNA 5′-AUUUUGUUUUUAAUAUUUC-3′ (Ambion) was [^{32}P] 5′-end labeled, and 0.4 μM of radiolabelled RNA was incubated 20 min at 25°C without and with 1.4 μg, 4.2 μg and 7 μg of protein in 10 μl of 10 mM Imidazole, pH 8.0, 50 mM NaCl, 2 mM DTT. Reaction mixtures was analyzed by PAGE and visualized by autoradiography.

Ion Binding Assays

Titration curves with $CaCl_2$, $MnCl_2$, $MgCl_2$ and $ZnCl_2$ were performed at 1 mg/ml protein in gel filtration buffer using thermal shift assay. Technical details can be be found in [41].

Endonuclease Assays

Endonuclease activity was assayed using 4 different heteromeric RNA substrates: an unstructured 19 mer as described above, a 21 mer stable hairpin ($5'$-UGAGGCCCGGAAACCGGGGCC-$3'$ (Ambion), $\Delta G = -14.7$ Kcal/mole), a 22 mer moderately stable hairpin ($5'$- CGCAGUUAGCUCCUAAUCGCCC-$3'$ (Ambion), $\Delta G = -3.4$ Kcal/mole), and a long 264 mer RNA corresponding to the SARS-CoV $5'$-genome sequence. The latter was radiolabelled with a cap structure at its $5'$-end using the ScriptCap m7G Capping System (Epicentre *Biotechnologies*) with $[\alpha^{32}P]$GTP. Endonuclease assays were carried out using 3.3 μM of radiolabeled RNA in a buffer containing 40 mM Tris-base, pH 7.5, 100 mM NaCl, 10 mM β-Mercaptoethanol and 2 mM $MnCl_2$. Reactions were initiated by the addition of 1 μM of protein and incubated at 37°C, and stopped by the addition of EDTA/formamide. Reactions products were analyzed using denaturing polyacrylamide gel electrophoresis (20% polyacrylamide, 7 M urea in TTE buffer (89 mM Tris, 28 mM taurine, 0.5 mM EDTA) and analyzed by autoradiography.

Mutagenesis and Reverse Genetics Assays Using a LCMV Mini-Replicon System

The LCMV replicon system is based on strain Armstrong clone 13 and has been established in analogy to the Lassa virus replicon described previously [42]. BSR T7/5 cells constitutively expressing T7 RNA polymerase [43] were transiently transfected with T7 promoter-driven expression constructs for L protein, nucleoprotein (NP), mini-genome (MG) containing Renilla luciferase reporter gene, and firefly luciferase as a transfection control. L protein mutants were generated as described [44]. One day after transfection, total RNA was prepared for Northern blotting and cell lysate was assayed for firefly and Renilla luciferase activity. Renilla luciferase levels were normalised with firefly luciferase levels resulting in standardized relative light units (sRLU). Northern blot was performed using an antisense ^{32}P-labeled riboprobe targeting the Renilla luciferase gene. Autoradiography was quantified on a PhosphorImager (Amersham Biosciences). To verify protein expression, hemagglutinin (HA)-tagged L protein was expressed in BSR T7/5 cells inoculated with modified vaccinia virus Ankara expressing T7 RNA polymerase (MVA- T7) [45] and detected in immunoblot using anti-HA antibody.

Supporting Information

Figure S1 Sequence Alignment of Viral Endonuclease Domains. **A**, Structure-based sequence alignment of NL1 with the two influenza PAN (3EBJ, 2W69), Tt1808 (1WDJ) and SdaI (2IXS), showing the structurally-conserved endonuclease motif (Highlighted in red). **B**, Sequence alignment of arenavirus NL1: Lymphocytic choriomeningitis virus (LCMV), Dandenong virus (DANV), Mopeia virus (MOPV), Morogoro virus (MORV), Mobala virus (MOBV), Ippy virus (IPPV), Lassa virus (LASV), Lujo Virus (LUJV), Parana virus (PARV), Pichinde virus (PICV), Allpahuayo virus (ALLV), Chapare virus (CHAV), Tamiami virus (TAMV), Whitewater Arroyo virus (WWAV), Bear Canyon virus (BCNV), Flexal virus (FLEV), Pirital virus (PIRV), Amapari virus (AMAV), Guanarito virus (GTOV), Cupixi virus (CPXV), Machupo virus (MACV), Junin virus (JUNV), Tacaribe virus (TCRV), Sabia virus (SABV), Oliveros virus (OLVV), Latino virus (LATV). Residues in a solid red background are strictly conserved. The blue point indicates the key active site residues.

Found at: doi:10.1371/journal.ppat.1001038.s001 (4.29 MB TIF)

Figure S2 Conservation of NL1 Between Arenaviruses and Bunyaviruses. **A**, Sequence alignment showing endonuclease motif (highlighted in red) of the NL1 domain from LCMV and the five genus of the *Bunyaviridae* families. Each genera is represented by two viruses: *Orthobunyavirus*: Human La Cross Virus (HLCV) and Bunyamwera virus (BUNV), *Hantavirus*: Hantaan virus (HANV) and Puumala virus (PUMV), *Phlebovirus*: Rift valley fever virus (RVFV) and Toscana virus (TOSV), *Tospovirus*: Tomato virus (TOMV) and Watermelon silver mottle virus (WTMV), *Nairovirus*: Crimean-Congo hemorrhagic fever virus (CCGV) and Dugbe virus (DUGV). Position of the start of the motif is labelled in blue for each virus. **B**, Based on secondary structure predictions topology diagrams were drawn for the NL1 domains from PUMV, RVFV, TOMV, HANV, CCGV. Colours are the same as in Figure 1D. N-terminal and C-terminal position are labelled with the aa position. The two red dots, the black dot and the green dot indicate the PD, E/D, and K residues, respectively, from the key active site. Principal conserved secondary structures are labelled as for LCMV NL1 domain.

Found at: doi:10.1371/journal.ppat.1001038.s002 (8.97 MB TIF)

Figure S3 Verification of L protein expression by immunoblotting. L protein used for analysis in replicon assay was tagged with HA tag, expressed under T7 promoter control in cells inoculated with modified vaccinia virus Ankara expressing T7 RNA polymerase (MVA-T7), and detected in immunoblot using anti-HA antibody. MVA, cells inoculated with MVA-T7 but not transfected; neg. ctrl., L mutant containing a mutation in the catalytic site of the RdRp.

Found at: doi:10.1371/journal.ppat.1001038.s003 (1.13 MB TIF)

Text S1 Supplementary Methods.

Found at: doi:10.1371/journal.ppat.1001038.s004 (0.08 MB DOC)

Acknowledgments

We thank Joëlle Boretto, Violaine Lantez, Karen Dalle for excellent technical assistance in preparing cDNAs and proteins used throughout, Lukas Flatz and Daniel Pinschewer for providing L and NP gene plasmids to facilitate construction of the LCMV mini-replicon system, Boris Klempa for helpful discussion and support on Bunyavirus work, Rolf Hilgenfeld and Hugues Tolou for sharing information and brainstorming at the initial phase of the project.

Author Contributions

Conceived and designed the experiments: B. Morin, R. Charrel, S. Günther, B. Canard. Performed the experiments: B. Morin, B. Coutard, M. Lelke, F. Ferron, R. Kerber, S. Jamal, A. Frangeul, C. Baronti, R. Charrel, X. de Lamballerie, J. Lescar. Analyzed the data: B. Morin, B. Coutard, M. Lelke, F. Ferron, R. Kerber, C. Vonrhein, G. Bricogne, S. Günther, B. Canard. Contributed reagents/materials/analysis tools: G. Bricogne, B. Canard. Wrote the paper: B. Morin, J. Lescar, B. Canard. Coordinated the study: B. Canard.

References

1. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, et al. (2009) Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. PLoS Pathog 5: e1000455.
2. Charrel RN, de Lamballerie X, Emonet S (2008) Phylogeny of the genus Arenavirus. Curr Opin Microbiol 11: 362–368.
3. Barton LL, Budd SC, Morfitt WS, Peters CJ, Ksiazek TG, et al. (1993) Congenital lymphocytic choriomeningitis virus infection in twins. Pediatr Infect Dis J 12: 942–946.
4. Barton LL, Hyndman NJ (2000) Lymphocytic choriomeningitis virus: reemerging central nervous system pathogen. Pediatrics 105: E35.

5. Fischer SA, Graham MB, Kuehnert MJ, Kotton CN, Srinivasan A, et al. (2006) Transmission of lymphocytic choriomeningitis virus by organ transplantation. N Engl J Med 354: 2235–2249.
6. Palacios G, Druce J, Du L, Tran T, Birch C, et al. (2008) A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med 358: 991–998.
7. Meyer BJ, de la Torre JC, Southern PJ (2002) Arenaviruses: genomic RNAs, transcription, and replication. Curr Top Microbiol Immunol 262: 139–157.
8. Salvato MS, Shimomaye EM (1989) The completed sequence of lymphocytic choriomeningitis virus reveals a unique RNA structure and a gene for a zinc finger protein. Virology 173: 1–10.
9. Meyer BJ, Southern PJ (1993) Concurrent sequence analysis of 5′ and 3′ RNA termini by intramolecular circularization reveals 5′ nontemplated bases and 3′ terminal heterogeneity for lymphocytic choriomeningitis virus mRNAs. J Virol 67: 2621–2627.
10. Raju R, Raju L, Hacker D, Garcin D, Compans R, et al. (1990) Nontemplated bases at the 5′ ends of Tacaribe virus mRNAs. Virology 174: 53–59.
11. Polyak SJ, Zheng S, Harnish DG (1995) 5′ termini of Pichinde arenavirus S RNAs and mRNAs contain nontemplated nucleotides. J Virol 69: 3211–3215.
12. Plotch SJ, Bouloy M, Krug RM (1979) Transfer of 5′-terminal cap of globin mRNA to influenza viral complementary RNA during transcription in vitro. Proc Natl Acad Sci U S A 76: 1618–1622.
13. Plotch SJ, Bouloy M, Ulmanen I, Krug RM (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell 23: 847–858.
14. Bishop DH, Gay ME, Matsuoko Y (1983) Nonviral heterogeneous sequences are present at the 5′ ends of one species of snowshoe hare bunyavirus S complementary RNA. Nucleic Acids Res 11: 6409–6418.
15. Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, et al. (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature 458: 914–918.
16. Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, et al. (2009) Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. Nature 458: 909–913.
17. Lopez N, Jacamo R, Franze-Fernandez MT (2001) Transcription and RNA replication of tacaribe virus genome and antigenome analogs require N and L proteins: Z protein is an inhibitor of these processes. J Virol 75: 12241–12251.
18. Muller R, Poch O, Delarue M, Bishop DH, Bouloy M (1994) Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. J Gen Virol 75(Pt 6): 1345–1352.
19. Vieth S, Torda AE, Asper M, Schmitz H, Gunther S (2004) Sequence analysis of L RNA of Lassa virus. Virology 318: 153–168.
20. Lukashevich IS, Djavani M, Shapiro K, Sanchez A, Ravkov E, et al. (1997) The Lassa fever virus L gene: nucleotide sequence, comparison, and precipitation of a predicted 250 kDa protein with monospecific antiserum. J Gen Virol 78(Pt 3): 547–551.
21. Wilda M, Lopez N, Casabona JC, Franze-Fernandez MT (2008) Mapping of the tacaribe arenavirus Z-protein binding sites on the L protein identified both amino acids within the putative polymerase domain and a region at the N terminus of L that are critically involved in binding. J Virol 82: 11454–11460.
22. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. Bioinformatics 24: 2780–2781.
23. Nichol STBBJ, Elliott RM (2005) Virus Taxonomy, VIIIth Report of the ICTV. In: Fauquet CM, Mayo AM, Maniloff J et al eds London: Elsevier Academic Press. pp 695–716.
24. Gro MC, Di Bonito P, Accardi L, Giorgi C (1992) Analysis of 3′ and 5′ ends of N and NSs messenger RNAs of Toscana Phlebovirus. Virology 191: 435–438.
25. Frias-Staheli N, Giannakopoulos NV, Kikkert M, Taylor SL, Bridgen A, et al. (2007) Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. Cell Host Microbe 2: 404–416.
26. Zhao C, Lou Z, Guo Y, Ma M, Chen Y, et al. (2009) Nucleoside monophosphate complex structures of the endonuclease domain from the influenza virus polymerase PA subunit reveal the substrate binding site inside the catalytic center. J Virol 83: 9024–9030.
27. Garcin D, Kolakofsky D (1990) A novel mechanism for the initiation of Tacaribe arenavirus genome replication. J Virol 64: 6196–6203.
28. Newman M, Strzelecka T, Dorner LF, Schildkraut I, Aggarwal AK (1994) Structure of restriction endonuclease BamHI and its relationship to EcoRI. Nature 368: 660–664.
29. Pingoud A, Fuxreiter M, Pingoud V, Wende W (2005) Type II restriction endonucleases: structure and mechanism. Cell Mol Life Sci 62: 685–707.
30. Lelke M, Brunotte L, Busch C, Gunther S (2010) An N-terminal region of Lassa virus L protein plays a critical role in transcription but not replication of the virus genome. J Virol 84: 1934–1944.
31. Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, et al. (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. Nat Struct Mol Biol 15: 500–506.
32. Abraham J, Corbett KD, Farzan M, Choe H, Harrison SC. Structural basis for receptor recognition by New World hemorrhagic fever arenaviruses. Nat Struct Mol Biol.
33. De Clercq E, Neyts J (2007) Avian influenza A (H5N1) infection: targets and strategies for chemotherapeutic intervention. Trends Pharmacol Sci 28: 280–285.
34. Hsieh HP, Hsu JT (2007) Strategies of development of antiviral agents directed against influenza virus replication. Curr Pharm Des 13: 3531–3542.
35. Parkes KE, Ermert P, Fassler J, Ives J, Martin JA, et al. (2003) Use of a pharmacophore model to discover a new class of influenza endonuclease inhibitors. J Med Chem 46: 1153–1164.
36. Powell HR (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. Acta Crystallogr D Biol Crystallogr 55: 1690–1695.
37. (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50: 760–763.
38. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126–2132.
39. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.
40. Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. The Pfam protein families database. Nucleic Acids Res 38: D211–222.
41. Malet H, Coutard B, Jamal S, Dutartre H, Papageorgiou N, et al. (2009) The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. J Virol 83: 6534–6545.
42. Hass M, Gölnitz U, Müller S, Becker-Ziaja B, Günther S (2004) Replicon system for Lassa virus. J Virol 78: 13793–13803.
43. Buchholz UJ, Finke S, Conzelmann KK (1999) Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. J Virol 73: 251–259.
44. Hass M, Lelke M, Busch C, Becker-Ziaja B, Günther S (2008) Mutational evidence for a structural model of the Lassa virus RNA polymerase domain and identification of two residues, Gly1394 and Asp1395, that are critical for transcription but not replication of the genome. J Virol 82: 10207–10217.
45. Sutter G, Ohlmann M, Erfle V (1995) Non-replicating vaccinia vector efficiently expresses bacteriophage T7 RNA polymerase. FEBS Lett 371: 9–12.
46. Kabsch W Xds. Acta Crystallogr D Biol Crystallogr 66: 125–132.
47. Evans P (2006) Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr 62: 72–82.
48. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. Methods Mol Biol 364: 215–230.
49. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 58: 1772–1779.
50. Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr D Biol Crystallogr 59: 2023–2030.
51. Cowtan K (1994) An automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 31: 34–38.
52. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr 62: 1002–1011.
53. Abrahams JP, Leslie AG (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr D Biol Crystallogr 52: 30–42.
54. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486–501.
55. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, et al. (2010) BUSTER version 2.X. Global Phasing Ltd. Cambridge, United Kingdom, .
56. Leslie AG (1992) MOSFLM - Recent changes and future developments. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 35: 18–19.
57. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53: 240–255.
58. Smart O, Brandl M, Flensburg C, Keller P, Paciorek W, et al. (2008) Refinement with Local Structure Similarity Restraints (LSSR) Enables Exploitation of Information from Related Structures and Facilitates use of NCS. Annual Meeting of the American Crystallographic Association.

Nature Reviews Microbiology | AOP, published online 5 December 2011; doi:10.1038/nrmicro2675

Conventional and unconventional mechanisms for capping viral mRNA

Etienne Decroly[1], *François Ferron*[1], *Julien Lescar*[1,2] *and Bruno Canard*[1]

Abstract | In the eukaryotic cell, capping of mRNA 5′ ends is an essential structural modification that allows efficient mRNA translation, directs pre-mRNA splicing and mRNA export from the nucleus, limits mRNA degradation by cellular 5′–3′ exonucleases and allows recognition of foreign RNAs (including viral transcripts) as 'non-self'. However, viruses have evolved mechanisms to protect their RNA 5′ ends with either a covalently attached peptide or a cap moiety (7-methyl-Gppp, in which p is a phosphate group) that is indistinguishable from cellular mRNA cap structures. Viral RNA caps can be stolen from cellular mRNAs or synthesized using either a host- or virus-encoded capping apparatus, and these capping assemblies exhibit a wide diversity in organization, structure and mechanism. Here, we review the strategies used by viruses of eukaryotic cells to produce functional mRNA 5′-caps and escape innate immunity.

The cap structure found at the 5′ end of eukaryotic mRNAs consists of a 7-methylguanosine (m^7G) moiety linked to the first nucleotide of the transcript via a 5′–5′ triphosphate bridge[1] (FIG. 1a). The cap has several important biological roles, such as protecting mRNA from degradation by 5′ exoribonucleases and directing ` pre-mRNA splicing and mRNA export from the nucleus[2]. In addition, the cap confers stability to mRNAs and ensures their efficient recognition by eukaryotic translation initiation factor 4E (eIF4E) for translation[3,4]. Conversely, RNA molecules with unprotected 5′ ends are degraded in cytoplasmic granular compartments called processing bodies (P-bodies)[5]. Uncapped RNAs, such as nascent viral transcripts, may also be detected as 'non-self' by the host cell, triggering (in mammalian cells) an antiviral innate immune response through the production of interferons[6,7].

Pre-mRNA splicing
A post-transcriptional modification of pre-mRNA, in which introns are excised and exons are joined in order to form a translationally functional, mature mRNA.

During virus–host co-evolution, viral RNA-capping pathways that lead to the same cap structure as that of host mRNAs have been selected for as efficient mechanisms to ensure both escape from detection by the innate immune system and efficient production of viral proteins. Compared with canonical eukaryotic mRNA capping, viral mRNA capping is highly diverse in terms of its genetic components, protein domain organization, enzyme structures and reaction mechanisms. Here, we review what is known about the various mRNA-capping pathways used by viruses that infect eukaryotes, paying particular attention to human pathogens. We also attempt to connect the pathways, machineries, structures and reactions involved in the viral RNA-capping process, and the specific cellular factors that trigger a response from the innate immune system.

Capping, decapping and turnover of host RNA

Nascent cellular mRNAs are generally produced in the nucleus in a 5′-triphosphate form and are modified co-transcriptionally by a set of cap-synthesizing enzymes. These enzymes are recruited by the DNA-dependent RNA polymerase RNA pol II on pausing, once the transcript is approximately 20–25 nucleotides long. Interestingly, two viruses — vaccinia virus (from the *Poxviridae* family of double-stranded DNA (dsDNA) viruses) and mammalian orthoreovirus (from the dsRNA *Reoviridae* family) — played a major part in the discovery of the RNA cap[1,8–11], because they produce high levels of capped viral mRNAs and encode their own capping machinery, allowing bona fide RNA cap synthesis *in vitro*. Mammalian orthoreovirus mRNAs and dsRNA genomes were first shown to be blocked at the 5′ end, preventing phosphorylation by polynucleotide kinase[9,10]. Purified mammalian orthoreovirus preparations were then demonstrated to be able to methylate mRNA 5′ ends, and the full mRNA cap structures were deciphered for both dsRNA orthoreoviruses[9,10,12] (those targeting humans and insects) and the dsDNA species vaccinia virus[11], followed by many other viral species[1,13,14].

The three canonical capping reactions responsible for the synthesis of the cap structure are outlined in

[1]*Centre National de la Recherche Scientifique and Aix-Marseille Université, UMR 6098, Architecture et Fonction des Macromolécules Biologiques, 163 avenue de Luminy, 13288 Marseille cedex 09, France.*
[2]*School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Republic of Singapore.*
Correspondence to B.C.
e-mail: Bruno.Canard@afmb.univ-mrs.fr

doi:10.1038/nrmicro2675
Published online
5 December 2011

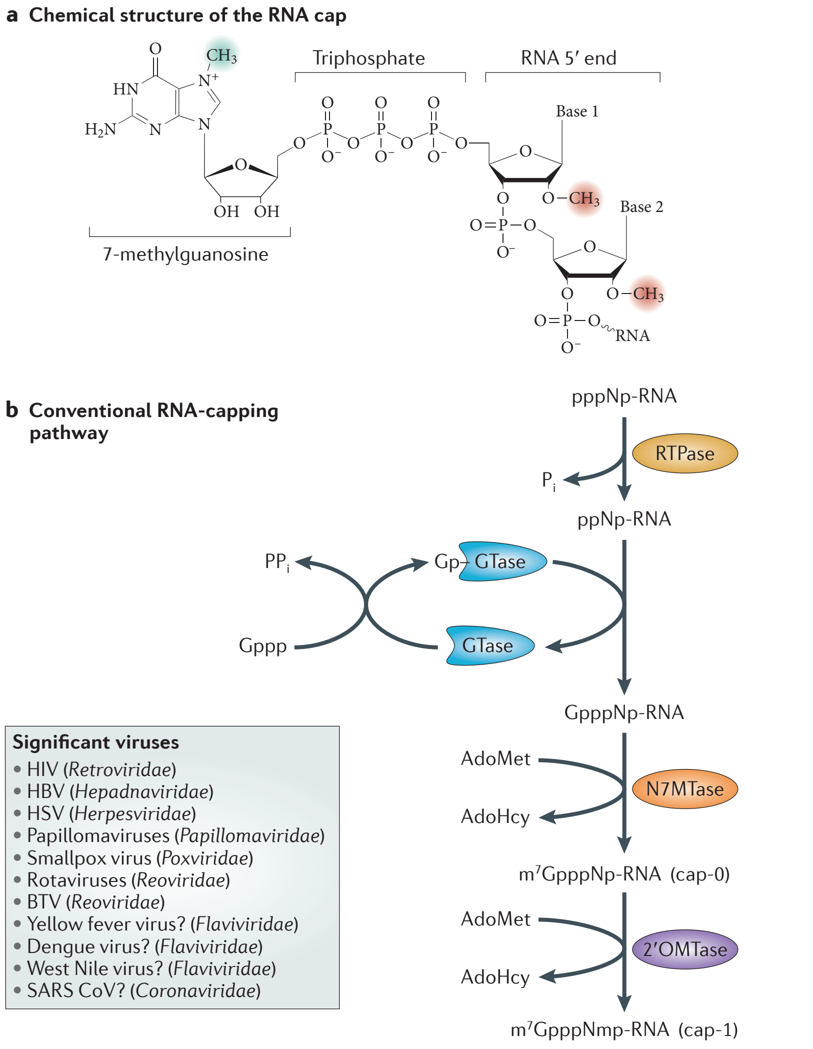


Figure 1 | **RNA cap structure and canonical capping mechanisms. a** | The mRNA cap consists of a 7-methylguanosine linked to the 5′ nucleoside of the mRNA chain through a 5′–5′ triphosphate bridge. The methyl group at the N7 position of the guanosine is shaded green, and the 2′-*O*-methyl groups of the first and second nucleotide residues, forming the cap-1 and the cap-2 structures, respectively, are shaded red. **b** | The cap-0 structure is formed on nascent RNA chains by the sequential action of three enzymes. First, the RNA triphosphatase (RTPase) hydrolyses the γ-phosphate of the nascent RNA (pppNp-RNA, in which N denotes the first transcribed nucleotide and p denotes a phosphate group) to yield a diphosphate RNA (ppNp-RNA) and inorganic phosphate (P_i). Then, guanylyltransferase (GTase) reacts with the α-phosphate of GTP (Gppp), releasing pyrophosphate (PP_i) and forming a covalent enzyme–guanylate intermediate (Gp–GTase). The GTase then transfers the GMP molecule (Gp) to the 5′-diphosphate RNA to create GpppNp-RNA. In the final step, (guanine-N7)-methyltransferase (N7MTase) transfers the methyl group from *S*-adenosyl-l-methionine (AdoMet) to the cap guanine to form the cap-0 structure, 7-methyl-GpppNp (m^{7}GpppNp), and releases *S*-adenosyl-l-homocysteine (AdoHcy) as a by-product. The capping reaction is completed by methylation of the ribose-2′-*O* position of the first nucleotide by the AdoMet-dependent (nucleoside-2′-*O*)-methyltransferase (2′OMTase), generating the cap-1 structure (m^{7}GpppNm$_{2'\text{-}O}$p). The box contains examples of viruses that acquire their cap structures using the cellular capping machinery or encode their own viral capping machineries that adopt the canonical pathway. Question marks indicate viruses that are likely to follow this conventional pathway. The RNAs capped by viral enzymes are indistinguishable from cellular mRNA and can thus be translated into proteins by the cellular ribosomal machinery. BTV, bluetongue virus; HBV, hepatitis B virus; HSV, herpes simplex viruses; SARS CoV, severe acute respiratory syndrome coronavirus.

FIG. 1b. This pathway is found in all eukaryotic species, as well as in most DNA viruses and members of the family *Reoviridae*, and consists of the following reactions: hydrolysis of the γ-phosphate of the primary transcript by an RNA 5′-triphosphatase (RTPase); transfer of GMP to the 5′-diphosphate RNA (ppNp-RNA; in which N is the first transcribed nucleotide and p is a phosphate group) by a guanylyltransferase (GTase) through a 'ping-pong' mechanism, leading to the formation of GpppNp-RNA; and methylation of the guanosine moiety by a cap-specific *S*-adenosyl-l-methionine (AdoMet)-dependent (guanine-N7)-methyltransferase (N7MTase), providing the minimal RNA cap chemical structure, named cap-0 (m^{7}GpppNp), which is recognized by the translation factor eIF4E[15]. Further methylation reactions catalysed by (nucleoside-2′-*O*)-methyltransferases (2′OMTases) can occur on the first and second nucleotides 3′ to the triphosphate bridge, yielding cap-1 (m^{7}GpppNm$_{2'\text{-}O}$), and cap-2 (m^{7}GpppNm$_{2'\text{-}O}$pNm$_{2'\text{-}O}$p) structures, respectively (FIG. 1). Cap-4 structures (m^{7}GpppNm$_{2'\text{-}O}$pNm$_{2'\text{-}O}$pNm$_{2'\text{-}O}$pNm$_{2'\text{-}O}$p) were also identified on parasite mRNAs[16].

Following translation, the lifespan of an mRNA molecule is controlled by two main processes in eukaryotic cells: first, the removal of its poly(A) tail and subsequent 3′-to-5′ exonucleolytic degradation, and second, an mRNA-decapping step that allows 5′-to-3′ exonucleolytic degradation (see REFS 17,18 for a review). Interestingly, decapped RNA may apparently be re-capped by the combined action of an as-yet-unknown 5′-monophosphate kinase[19] interacting with a host cell GTase that is also present in minor amounts in the cytoplasm (see REF. 20 for a review). Cytoplasmic decapping is catalysed by DCPS, a member of the NUDIX hydrolase superfamily, and stimulated by decapping enhancer proteins. Following RNA decapping in P-bodies, transcripts are quickly degraded by 5′-to-3′ exonucleases such as XRN1. Thus, 5′-triphosphate mRNAs are almost completely absent in the cytoplasm. The eukaryotic cell has consequently evolved mechanisms to sense triphosphate RNA as non-self and uses these RNA species to trigger an innate immune response[6,7]. Viruses are the most common cell invaders to produce cytoplasmic mRNAs and have therefore been under a selective pressure to evolve

γ-phosphate
The third phosphate attached at the 5′ end of the ribose moiety of a nucleotide.

'Ping-pong' mechanism
A two-step mechanism in which a substrate molecule first forms a (covalent) link with the enzyme and is then transferred to an acceptor molecule to yield a product.

Poly(A) tail
A string of AMP that is added to the 3′ end of mRNA.

NUDIX hydrolase superfamily
A family of proteins that hydrolyse a wide range of organic pyrophosphates, including NDPs, NTPs, dinucleoside and diphosphoinositol polyphosphates, nucleotide sugars and RNA caps, with varying degrees of substrate specificity.

Box 1 | **Getting around the lack of capping**

Some viruses (such as single-stranded positive-sense RNA (ss(+)RNA) viruses from the families *Picornaviridae*, *Caliciviridae* and *Astroviridae*) do not have a cap structure at the 5′ end of their mRNAs or genomic RNAs. Rather, they covalently attach to the 5′ RNA end a protein termed VPg[21] and/or carry an internal ribosome entry site (IRES) structure in the 5′ untranslated region[154]. IRESs have now been found in many different cellular and viral RNAs, including those of ss(+)RNA viruses from the families *Picornaviridae* and *Dicistroviridae*, the genus Lentivirus and the *Flaviviridae* genera Hepacivirus and Pestivirus. In pestiviruses, genomic RNA remains in a 5′-triphosphate form and thus promotes high levels of expression of host interferon-stimulated genes. However, these viruses also trigger several pathways that limit the antiviral response[155,156], resulting in a competitive advantage for IRES-dependent translation of viral genes[157].

counteracting mechanisms to conceal their RNA 5′ ends from the innate immunity machineries of the host cell.

Some viruses, such as those in the family *Picornaviridae* (single-stranded positive-sense RNA viruses (ss(+)RNA viruses); for example, polioviruses and encephalomyocarditis virus (EMCV)) recruit the 43S pre-initiation complex in a 5′-cap-independent manner (BOX 1). Other viruses (for example, viruses of the family *Caliciviridae*, which are also ss(+)RNA viruses) covalently attach their RNA 5′ end to a VPg-like protein, which directly interacts with the cap-binding protein eIF4E and initiates translation of viral mRNAs[21]. In members of the *Picornaviridae*, VPg may be lost before translation[22,23].

However, by far the most common viral mechanism for ensuring efficient translation of viral proteins and avoiding immune surveillance mechanisms is through the acquisition of a cap structure. The remarkable diversity of mechanisms that lead to an RNA cap structure identical to that of the host cell mRNAs is described below.

Conventional capping of viral RNA

Even when the viral replication cycle includes a nuclear phase, any virus entering a host cell must express its genes in the cytoplasm using the host translation machinery. Viral genomes can be made of single-stranded or double-stranded nucleic acids, either RNA or DNA, and the corresponding strategies used for protecting their RNA transcripts are outlined in FIG. 2.

Cap structures can be added to viral RNAs by one of the three following mechanisms. In the first mechanism, most viruses that synthesize their mRNA using cellular RNA pol II use the cellular capping machinery (FIGs 1b,2), as exemplified by most DNA viruses (except for those from the dsDNA virus family *Poxviridae*) and by RNA viruses such as those of the family *Retroviridae* (ss(+)RNA viruses) and the family *Bornaviridae*

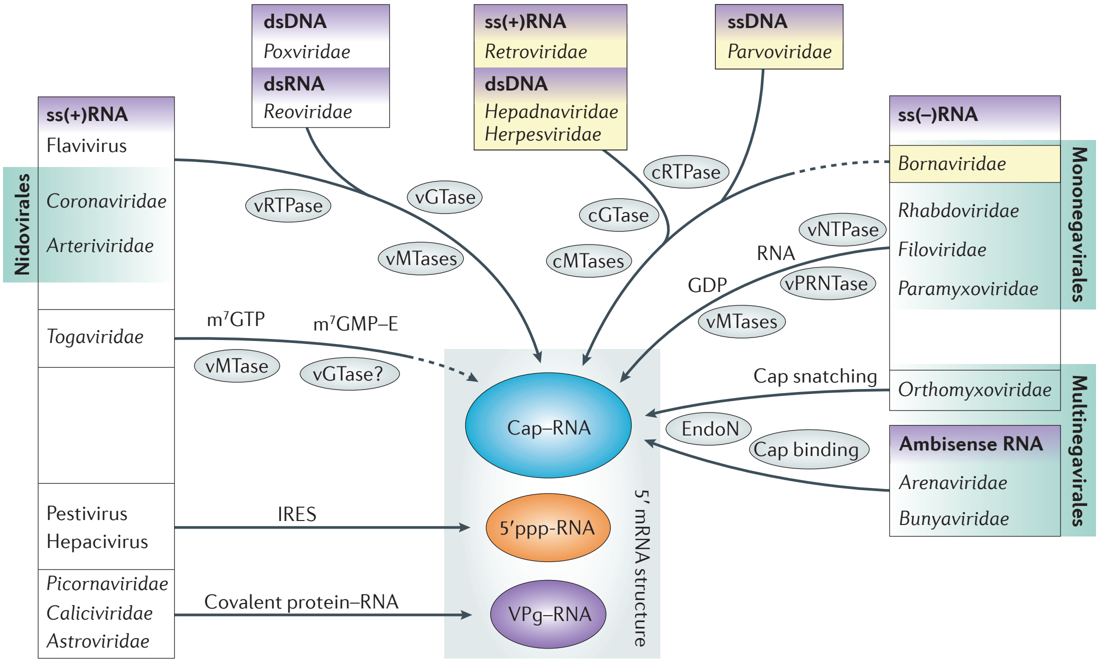


Figure 2 | **RNA 5′ ends in the mammalian-virus world.** Mammalian viruses, with the exception of those from the single-stranded positive-sense RNA (ss(+)RNA) virus genera Pestivirus and Hepacivirus, use strategies to chemically modify their mRNA 5′ ends through either covalent attachment of a protein (VPg for ss(+)RNA viruses from the families *Picornaviridae*, *Caliciviridae* and *Astroviridae*) or covalent attachment of an RNA cap structure (all other viruses). Arrows indicate the type of RNA 5′ end protection that is used by these viral groups, and the enzymes and mechanisms involved are indicated. Dashed arrows indicate a likely but incompletely demonstrated pathway. Viral and cellular proteins are distinguished with the prefixes v and c, respectively. Yellow shading highlights viral groups for which the life cycle includes a nuclear phase in the host cell. This list of viral taxa is non exhaustive and used as an example only. ds, double-stranded; E, enzyme; EndoN, endonuclease; GTase, guanylyltransferase; IRES, internal ribosome entry site; m^7, 7-methyl; MTase, methyltransferase; NTPase, nucleotide 5′-triphosphatase; p, phosphate group; PRNTase, polyribonucleotidyl transferase; RTPase, RNA triphosphatase; ss(–)RNA, single-stranded negative-sense RNA.

Single-stranded positive-sense RNA viruses
(ss(+)RNA viruses). Viruses that have or produce mRNAs that are co-linear to their genomic RNA.

43S pre-initiation complex
A multiprotein complex composed of eukaryotic translation initiation factor 3 (eIF3), eIF4A, eIF4E and eIF4G associated with the small ribosomal subunit. This pre-initiation complex scans the mRNA towards the 'start' codon (typically AUG), where translation is initiated.

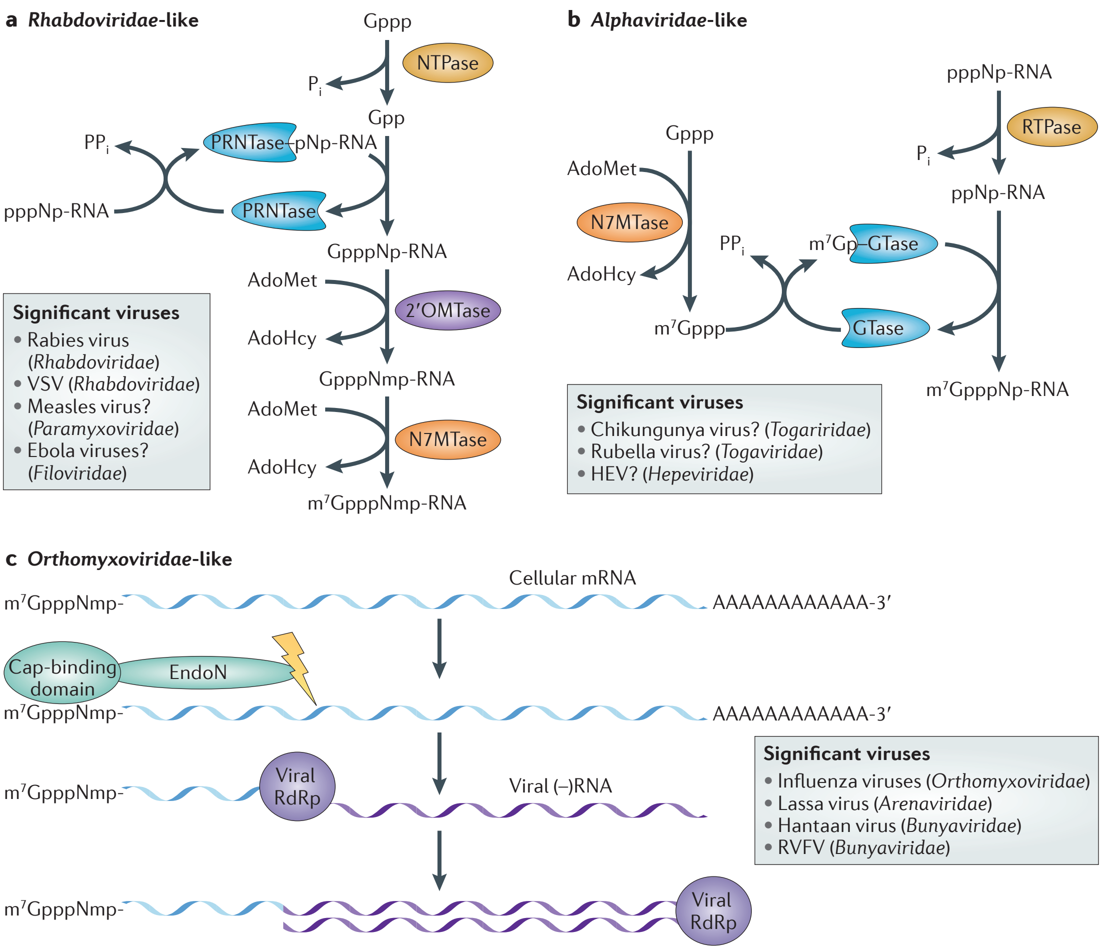


Figure 3 | **Unconventional capping pathways. a** | The RNA-capping mechanism of negative-sense RNA ((–)RNA) viruses such as those of the family *Rhabdoviridae*. The NTPase hydrolyses the γ-phosphate of GTP (Gppp) to yield a GDP (Gpp) and inorganic phosphate (P_i). Polyribonucleotidyl transferase (PRNTase) reacts with the nascent viral RNA (pppNp-RNA; in which N denotes the first transcribed nucleotide), releasing pyrophosphate (PP_i) and forming a covalent PRNTase–pNp-RNA intermediate. The PRNTase then transfers the RNA molecule to the GDP to create GpppNp-RNA. (Nucleoside-2′-*O*)-methyltransferase (2′OMTase) transfers the methyl group from *S*-adenosyl-L-methionine (AdoMet) to the first nucleotide of the RNA. The capping reaction is then completed by methylation of the cap by the AdoMet-dependent (guanine-N7)-methyltransferase (N7MTase). The box lists examples of viruses that acquire their cap structures using such a capping pathway. Question marks indicate viruses that are likely to follow this conventional pathway. **b** | The RNA-capping mechanism of positive-sense RNA ((+)RNA) viruses such as those of the family *Alphaviridae*. The RNA triphosphatase (RTPase) hydrolyses the γ-phosphate of the viral RNA to yield a diphosphate RNA (ppNp-RNA) and P_i. A GTP molecule is methylated at its N7 position by the AdoMet-dependent N7MTase. Guanylyltransferase (GTase) then binds the N7-methyl-GTP (m^7Gppp), forming a covalent link with a catalytic histidine (m^7Gp–GTase) and releasing PP_i. The GTase then transfers the m^7GMP molecule (m^7Gp) to the 5′-diphosphate RNA to create m^7GpppNp-RNA. The box indicates viruses that acquire their cap structures using such a capping pathway. **c** | the RNA-capping mechanism of (–)RNA viruses such as those of the family *Orthomyxoviridae*; this mechanism is referred to as cap snatching. The PB2 subunit of the viral RNA-dependent RNA polymerase (RdRp) binds to the 5′ end of cellular capped mRNAs (which are enriched in the processing (P)-bodies), and the PA subunit then releases short capped RNAs by using its endonuclease (EndoN) activity. These capped RNAs are used as primers by the viral RdRp in viral transcription to generate viral mRNA using the viral (–)RNA as a template. The RdRp then synthesizes the complementary negative-sense strand. The box provides example of viruses that acquire their cap structures using a similar capping pathway. Note that most of the mRNAs that are capped by viral enzymes are indistinguishable from cellular mRNAs and can be translated into proteins by the cellular ribosomal machinery. AdoHcy, *S*-adenosyl-L-homocysteine; HEV, hepatitis E virus; RVFV, Rift Valley fever virus; VSV, vesicular stomatitis virus.

Single-stranded negative-sense RNA viruses
(ss(–)RNA viruses). Viruses that have or produce mRNAs that are complementary to their genomic RNA.

Ambisense RNA viruses
Viruses (such as members of the families *Arenaviridae* and *Bunyaviridae*) that have or produce both mRNAs that are co-linear to and mRNAs that are complementary to their genomic RNA, although most mRNAs are complementary in polarity.

(single-stranded negative-sense RNA viruses (ss(–)RNA viruses)). A second strategy consists of acquiring cap structures from cellular mRNAs by 'cap snatching' (FIGS 2,3). RNA viruses such as members of the families *Orthomyxoviridae*, *Arenaviridae* and *Bunyaviridae* — which are ss(–)RNA viruses, with the latter two families also referred to as ambisense RNA viruses — steal short, capped cellular mRNAs through endonucleolytic cleavage. The stolen short, capped mRNAs are then used by the viral polymerase to prime synthesis of viral RNA.

For the third method, many viruses encode their own capping machinery and have evolved a diverse set of dedicated enzymes and mechanisms to carry out capping. Accordingly, most viruses with an ssRNA genome synthesize or acquire the RNA cap using their own set of enzymes (FIG. 2). Within this diversity, the capping of viral mRNA can be classified as either 'conventional', when it follows the mRNA-capping pathways used by eukaryotes and DNA viruses (that is, the sequential action of RTPase, GTase and MTases) (FIGS 1b,2), or 'unconventional', when it does not (see below and FIG. 3).

The best characterized conventional RNA-capping system is that exemplified by the dsDNA virus vaccinia virus, which expresses a multifunctional mRNA cap-synthesizing enzyme (D1) containing RTPase, GTase and N7MTase domains[24]. The 5′-triphosphate of the nascent mRNA is first hydrolysed by the RTPase to yield 5′-diphosphate RNA, which is then sequentially transferred to other internal domains[25,26] to be capped and methylated, the latter reaction with allosteric stimulation through direct association with viral D12 protein[25,26]. Cap assembly is completed by the viral VP39, a bifunctional protein that catalyses the 2′-*O*-methylation of the cap-0 structure and also acts as an elongation factor for poly(A) polymerase[27]. Viruses from the dsRNA virus family *Reoviridae* share the same pathway as vaccinia virus. The enzymes remain physically associated in an inner capsid (or 'transcriptionally active core'), which constitutes a molecular machine or 'assembly line' that is able to transcribe the genome, synthesize the cap and inject the resulting mRNA into the cytoplasm of the host cell for translation. In addition, further assignment and characterization of the GTase activity in viruses of the genus Flavivirus and those of the family *Coronaviridae* may join these viruses, which are ss(+)RNA viruses, to the conventional RNA-capping pathway group.

Conventional viral RNA-cap-synthesizing enzymes

RNA triphosphatases. RTPases catalyse the cleavage of the interphosphate bond between the β-phosphate and the γ-phosphate of 5′-triphosphorylated mRNA (FIG. 1b). A range of enzyme structures exist (FIG. 4a), indicating that during evolution several independent solutions evolved for this initial step of the capping reaction, which is often found in association with RNA-binding and strand separation activities. Most RTPases are also able to hydrolyse NTPs[28–30].

RTPases can be categorized into four groups, further defined by the metal dependency of the hydrolytic reaction mechanism that they use to cleave the interphosphate bond. Metal-dependent RTPases can be organized into three structural groups: the histidine triad (HIT)-like fold (α–β complex), the triphosphate tunnel metalloenzyme (TTM) and the viral RNA helicase-like fold (FIG. 4a).

The HIT-like fold is found in NSP2, from the dsRNA rotaviruses, and so far it is the only RTPase identified with such a fold. It has an amino-terminal helical domain and a carboxyl terminus with an α–β fold that resembles the ubiquitous cellular HIT group of nucleotidyl hydrolases[31]. The nucleotide-binding site is located in the cleft between the two domains, which contains a histidine residue involved in binding and in Mg^{2+}-dependent hydrolysis of both NTP and RNA substrates. NSP2 self-assembles into a doughnut-shaped octamer, a quaternary structural organization that creates several RNA-binding sites, which are presumably needed to destabilize RNA duplexes during genome replication and packaging[30,32,33].

TTM enzymes hydrolyse NTPs in the presence of Mn^{2+} or Co^{2+} and are found in fungi and protozoa, and in most DNA viruses that encode an RTPase, including poxviruses, chlorella virus, baculoviruses and mimiviruses[34–37]. The RTPase Cet1 from *Saccharomyces cerevisiae* is nearly identical to that of mimiviruses and serves as a paradigm for the TTM group. Its structure features a characteristic tunnel lined by eight antiparallel β strands (FIG. 4a) and a double glutamate motif[36,38]. The mRNA supposedly sits in the tunnel, which harbours several charged and hydrophilic side chains that coordinate Mn^{2+} and SO^{2-} in the crystal structure. SO^{2-} is thought to indicate the position of the γ-phosphate of mRNA.

The third group comprises proteins with NTPase and helicase activity, belonging to the large helicase superfamilies SF1 and SF2 (REFS 39–41). Such NTPase–helicase family members are found in flaviviruses (for example, NS3 proteins of dengue virus, yellow fever virus, West Nile virus and Japanese encephalitis virus), coronaviruses (for example, nsp13 of severe acute respiratory syndrome coronavirus (SARS CoV), although for this member of the SF1 helicases the crystal structure is not known), alphaviruses (for example, the protein nsP2) and potexviruses (for example, protein 1A of bamboo mosaic virus), all of which are ss(+)RNA viruses, and also in viruses of the family *Reoviridae* (for example, the protein λ1), which are dsRNA viruses. The fold adopted by SF2 helicases features two RecA-like subdomains between which a cleft accommodates the nucleotide or 5′-triphosphate RNA substrate[42]. The RecA subdomains I and II carry the Walker A and B motifs. Residues belonging to the Walker A motif (also named motif I) form the P-loop, which stabilizes the terminal phosphate moiety of the substrate, whereas acidic residues from the Walker B motif (also named motif II or DEXD box) coordinate the Mg^{2+} needed for hydrolysis[42]. Structural and biochemical studies revealed that the RTPase and NTPase activities of helicase enzymes from flaviviruses (ss(+)RNA) have a common catalytic site[28,42–47]. Similar biochemical studies have mapped an associated helicase–RTPase activity in other viral families, including alphaviruses (ss(+)RNA)[48]. However, structural data are needed to better understand how these multifunctional viral proteins coordinate their various activities.

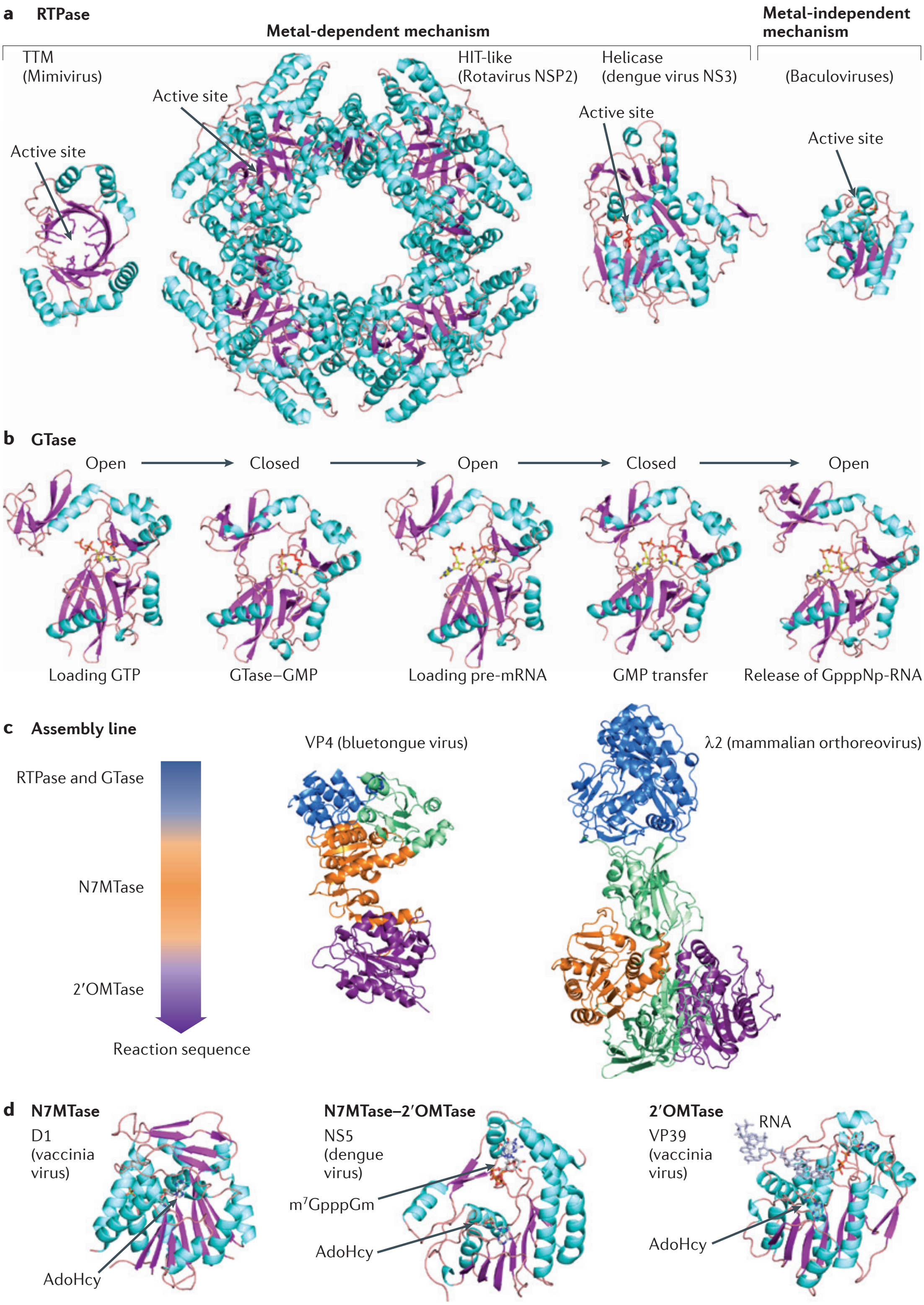
Metal-independent RTPases are found in plants, metazoa and viruses such as baculoviruses (dsDNA)[49,50]. They belong to the cysteine phosphatases superfamily. The RTPase reaction proceeds in two steps, starting with entry of the 5′-triphosphate RNA into the active-site tunnel and formation of a covalent cysteinyl-*S*-phosphoester adduct. The catalytic cysteine responsible for the nucleophilic attack on the γ-phosphate of the RNA 5′ end belongs to a P-loop motif (HCXXXXXR(T/S)). The second step releases inorganic phosphate. Cysteine

β-phosphate
The second phosphate attached at the 5′ end of the ribose moiety of a nucleotide.

Walker A and B motifs
Motifs that are present in nucleotide-binding proteins but also in a range of proteins with widely varying functions, including ATP synthases, myosins, transducins, helicases, kinases and RecA proteins. The Walker A motif contains a phosphate-binding loop (P-loop) motif with the consensus sequence GXXXGK(T/S), and the Walker B motif contains the consensus sequence (R/K)XXXXGXXXXLhhhhD (in which h refers to any hydrophobic residue).

Nucleophilic attack
Generally, a starting point for a chemical reaction; a doublet of electrons selectively attacks the positive or partially positive charge of the atomic nucleus in order to create a new chemical bond.

LXXXIII



phosphatases adopt a characteristic α–β fold with a central twisted, five-stranded parallel β-sheet flanked by six α-helices[49]. The catalytic cysteine that specifically recognizes the γ-phosphate of the RNA 5′ end resides at the bottom of the substrate-binding pocket, whereas other conserved residues from the P-loop, together with surrounding residues, form a positively charged channel that can accommodate the α-phosphate and the β-phosphate of the RNA 5′ end. The shape of the binding pocket dictates the selectivity for triphosphate RNA and seems too deep to grant diphosphate RNA access to the active site.

α-phosphate
The first phosphate attached at the 5′ end of the ribose moiety of a nucleotide.

◂ Figure 4 | **Structural constituents of viral capping machineries, folds and mechanisms. a** | Different RNA triphosphatase (RTPase) folds. For the metal-dependent mechanism, the triphosphate tunnel metalloenzyme (TTM) fold is exemplified by the structure of the RTPase (Protein Data Bank (PDB) accession code 2QZE) from the genus Mimivirus, which consists of double-stranded DNA (dsDNA) viruses; the histidine triad (HIT)-like fold of the RTPase from dsRNA rotaviruses (PDB accession code 1L9V) and the helicase fold of the RTPase from the single-stranded positive-sense RNA (ss(+)RNA) virus dengue virus (PDB accession code 2BHR) are also shown. For the metal-independent mechanism, the fold of the RTPase from the dsDNA baculoviruses (PDB accession code 1YN9) is the only viral structure available so far. Structures are coloured cyan for α-helices and pink for β-strands. **b** | Structure-based model of a guanylyl transfer mediated by the dsDNA virus *Paramecium bursaria* chlorella virus 1 guanylyltransferase (GTase). Presented are the five stages of the reaction: loading of GTP onto the active site, with the GTase in an open conformation (PDB accession code 1CKO); the GTase in a closed conformation and the creation of covalent intermediate GTase–GMP through a lysine residue (model derived from the structure in PDB accession code 1CKN); the intermediate GTase–GMP reopening to bind a pre-mRNA molecule (model derived from the structure in PDB accession code 1CKO); the GTase closing again to complete the transfer (model derived from the structure in PDB accession code 1CKN); and, transfer completed, the GTase opening to release the GpppNp-RNA (in which p is a phosphate group and N is the first transcribed nucleotide) (PDB accession code 1CKM). Structures are coloured cyan for α-helices and pink for β-strands. **c** | Assembly line structures. Protein VP4 (PDB accession code 2JHC) of bluetongue virus (a dsRNA virus from the genus Orbivirus) and protein λ2 of mammalian orthoreovirus (a dsRNA virus) (PDB accession code 1EJ6). The colour code correlates with the colour of each domain as represented in the assembly sequence arrow to the left. Domains in green are extra domains that do not take part directly in the capping mechanism. **d** | Methyltransferase structures. Left: the (guanine-N7)-methyltransferase (N7MTase) domain of protein D1 from the dsDNA virus vaccinia virus (PDB accession code 2VDW) in complex with a molecule of *S*-adenosyl-L-homocysteine (AdoHcy). Middle: the N7MTase–(nucleoside-2′-*O*)-methyltransferase (2OMTase) domain of NS5 from dengue virus (ss(+)RNA) in complex with the cap analogue 7-methyl-G (m^7G)-pppGm$_{2'-O}$ and AdoHcy (PDB accession code 2P41). Right: VP39, the 2′OMTase of the dsDNA virus vaccinia virus, in complex with a capped RNA and AdoHcy (PDB accession code 1AV6). All figures were prepared using PyMOL.

pK_a value
The acid dissociation constant, a quantitative measurement of the strength of a chemical group as an acid in solution. It corresponds to the pH value at which half of the ionizable group is either protonated or deprotonated.

ε-amino group
A positively charged group found at the extremity of a lysine side chain. The ε-amino group is a primary amine and, owing to its high pK_a value, it is reactive and often participates in reactions at the active site of enzymes.

Guanylyltransferases. GTases of DNA viruses contain two domains: a nucleotidyltransferase (NTase) domain that is conserved in capping enzymes, RNA ligases and DNA ligases[51], and a C-terminal oligonucleotide-binding domain that is observed in capping enzymes and several DNA ligases. Sequence alignments aided by structural information for several family members identified conserved residues and motifs both in the nucleotide-binding site and in the NTase site[51,52]. A lysine-containing motif, KXDG(I/L), is conserved among the GTases encoded by several DNA viruses (vaccinia virus, Shope fibroma virus and African swine fever virus) and the yeasts *S. cerevisiae* and *Schizosaccharomyces pombe*. The lysine in this motif was shown to be the catalytic residue in the GTase of vaccinia virus[53–55] and in the yeast capping enzyme[56]. Moreover, this motif is conserved in the active site of polynucleotide ligases, which, like capping enzymes, catalyse an enzymatic reaction via the formation of a covalent Lys–NMP intermediate[57].

The structure of the GTases from *Paramecium bursaria* chlorella virus 1 (dsDNA), humans[58] and the yeast *Candida albicans* (the protein Cgt1) have been solved individually and, for the chlorella virus GTase and Ctg1, in complexes with cap analogues, GTP or as a covalent GTase–GMP intermediate[59–61]. These structures suggest the following reaction scheme, during which the oligonucleotide-binding and NTase domains, acting as a clamp, undergo a series of large conformational changes. These domains open up to load GTP, close to catalyse the formation of a covalent enzyme–GMP complex through the action of the catalytic lysine, open again to release pyrophosphate and bind the pre-mRNA substrate, close again to catalyse nucleotidyl transfer to the RNA and finally re-open to release the GpppNp-RNA product[51,60] (FIG. 4b).

The dsRNA viruses of the family *Reoviridae* use one multifunctional capsid protein for the capping reaction. λ2 of the dsRNA virus mammalian orthoreovirus[62], VP4 of bluetongue virus (in the genus Orbivirus[63]) and VP3 from cytoplasmic polyhedrosis virus (in the genus Cypovirus[64]) act as assembly lines, such that the RNA substrate is shuttled from one domain to the next (see below). The GTase domains of these proteins feature different folds (see below and FIG. 4c) to DNA virus GTases. Despite these structural differences, viruses from the genera Orthoreovirus and Aquareovirus use a conserved lysine residue (although not part of the signature sequence found in DNA viruses) to form a covalent intermediate with GMP[65,66]. In flaviviruses (ss(+)RNA), the 2′OMTase domain has been proposed to act as a GTase[67,68]. However, the proposed catalytic lysine is not conserved in flaviviruses[69].

Methyltransferases. Despite the limited overall amino acid sequence identity in the large family of AdoMet-dependent MTases, most of these enzymes share a common structural core made of seven β-strands flanked by three α-helices on each side of the sheet, similar to the core found in the catechol-*O*-MTase (a class I AdoMet-dependent MTase[70,71]). This core catalytic domain has evolved extensions that consist of structurally non-conserved domains and allow these MTases to accommodate a range of methyl acceptors[70]. The MTases exist either as isolated proteins (for example, nsp16 viruses from the *Coronaviridae*; ss(+)RNA[72]), or as domains of larger proteins (such as NS5 of viruses from the ss(+)RNA virus genus Flavivirus[69,73] or the multidomain cap assembly lines of dsRNA viruses from the *Reoviridae*[62]). In some instances the same protein domain can have dual N7MTase and 2′OMTase activities (for example, the MTase domain in NS5 from flaviviruses[74,75]), sharing the same cofactor-binding site. When this is the case, repositioning of the RNA substrate must occur.

Both N7MTases and 2′OMTases share the class I family fold (FIG. 4d). N7-methyl transfer is thought to be promoted by optimal positioning of the reacting groups, mediated by several aromatic residues, and by an electrostatic environment that is favourable for the reaction[25,76]. By contrast, 2′OMTases rely on a conserved catalytic tetrad, Lys-Asp-Lys-Glu[69,77,78]. The catalytic reaction was deciphered by several structural studies of the 2′OMTase VP39 of vaccinia virus[79,80]. It was suggested that residues in the vicinity of the catalytic Lys175 (the second lysine in the Lys-Asp-Lys-Glu motif) decrease the pK_a value of its ε-amino group. This orientates the 2′-hydroxyl group of the ribose for a nucleophilic in-line *sn*-2 attack on the AdoMet methyl group[81].

Using cap analogues, the molecular basis for recognition of GpppNp-RNA versus cap-0-RNA has also been characterized for VP39 (REFS 80,82). The m^7G is stacked between two aromatic residues, and electron delocalization and electrostatically enhanced stacking owing to N7 methylation favours the recognition of the cap over GTP (FIG. 4d).

One interesting aspect of MTase activity relates to its regulation: in three cases, interfacial activation is achieved either by a cofactor protein (such as the vaccinia virus D12 subunit, which enhances the N7MTase activity of D1 through an allosteric mechanism[25,26], and the SARS CoV (ss(+)RNA) metalloprotein nsp10, which acts as a cofactor to activate the 2′OMTase nsp16 but not the N7MTase nsp14 (REF. 83)) or by binding to lipid membranes (as is the case for nsP1 of the ss(+)RNA alphaviruses).

What determines the sequence of the two methylation steps? For flaviviruses[74,75] or coronaviruses[83,84], which are ss(+)RNA viruses, the order in which methylations are performed is not encoded in the global protein architecture. Rather, variations in kinetics and affinity may dictate the order in which reactions occur[83,85]. In the case of flaviviruses, RNA secondary structures also seem to be important: whereas the N7MTase activity of the bifunctional NS5 MTase domain requires a long substrate encompassing a specific stem loop RNA structure, christened stem loop A (SLA), the 2′OMTase activity is able to transfer a methyl group to short RNA acceptors[75].

Cap assembly lines. Several dsRNA viruses encode structural proteins that are packaged with their genome in the viral particle and are able to perform the four reactions needed to synthesize a cap-1 structure, much like an assembly line (FIG. 4c). The key molecular components of the RNA-capping machinery in members of the dsRNA virus family *Reoviridae* are RNA-directed RNA polymerase (named VP1 in orbiviruses and rotaviruses, and λ3 in orthoreoviruses) and a multifunctional cap-synthesizing enzyme (named VP4 in orbiviruses, VP3 in rotaviruses and λ2 in orthoreoviruses). Both λ2 and VP4 are composed of four domains that were identified as RTPase, GTase, N7MTase and 2′OMTase, respectively[62,63]. The spatial arrangement of the different protein domains reflects the time sequence of the enzymatic reactions that are required for mRNA capping following synthesis by the viral polymerase (FIG. 4c). Although it is unclear how and when RTPase activity occurs, a complete pathway has been proposed in which guanylyl transfer occurs near the base of the pentameric 'turret' (formed by λ2 in orthoreoviruses), followed by N7-methylation and 2′-*O*-methylation of the mRNA[62]. Therefore, in dsRNA viruses, the sequence of steps in the cap synthesis pathway should remain identical to the sequence in the capping pathway for cellular mRNAs and for DNA viruses such as vaccinia virus, for which the capping machinery is embedded in a multidomain protein complex.

Stem loop
A hairpin structure in single-stranded RNA or DNA, resulting from intramolecular base-pairing when two regions of the same strand contain partial or perfect anti-complementary nucleotide sequences.

The GTase domain and both MTase domains of bluetongue virus (a dsRNA virus in the genus Orbivirus) were unambiguously mapped on the VP4 structure, but the position of the RTPase domain remains uncertain[63]. However, it is believed that both the RTPase and GTase activities reside in the same C-terminal domain of VP4, a unique architecture that is reminiscent of, but distinct from, double-domain RTPase–GTase proteins found in metazoans and plants. The enzymatic activity requires Mg^{2+} (REFS 63,86,87). However, VP4 does not adopt a typical metal-dependent RTPase fold. A cysteine residue is found in a deep cavity similar to that harbouring the catalytic motif of the cysteine phosphatase superfamily[63]. Thus, these assembly line enzymes seem to have incorporated features from various phylogenetic origins.

Unconventional cap synthesis pathways

The first indication that there are deviations from the conventional RNA-capping pathway for viral mRNAs came in the early 1970s, around the time of the discovery of the RNA cap structure. Since then, is has been demonstrated that the ss(–)RNA virus vesicular stomatitis virus (VSV) and ss(+)RNA alphaviruses (from the family *Togaviridae*) can synthesize a viral RNA cap that is identical to a cellular RNA cap, albeit constructed through a completely different mechanism. Although alphaviruses do not proceed further than synthesizing a cap-0 structure, the fact that divergent biosynthetic pathways converge to the consensus cap structure indicates that the selective pressure to maintain this structure must be high.

The Mononegavirales RNA-capping pathway. Mononegavirales is a viral order of ss(–)RNA viruses with unsegmented genomes, such as VSV and rabies virus (in the family *Rhabdoviridae*), measles virus (from the family *Paramyxoviridae*), bornavirus (from the family *Bornaviridae*), and Ebola viruses and Marburg viruses (from the family *Filoviridae*). These viruses encode a multifunctional L protein that carries RNA-dependent RNA polymerase (RdRp) and RNA cap synthesis activities. These enzymes have evolved independently from other known eukaryotic cap-synthesizing enzymes, and the L proteins of VSV[88,89], spring viraemia of carp virus[90], human respiratory syncytial virus[91] and Chandipura virus[92] transfer GDP rather than GMP to the RNA 5′ end. Part of a domain in the conserved region V of L protein contains the GDP polyribonucleotidyl transferase (PRNTase) activity, and forms a covalent enzyme–pNp–RNA intermediate (FIG. 3a) with the nascent viral RNA. The covalent bond with RNA involves a conserved histidine residue present in an 'HR' motif instead of the lysine residue used by conventional GTases[92]. The 5′-monophosphorylated viral mRNA start sequence then receives GDP generated from GTP[93] by an as-yet-unknown NTPase. The VSV MTase, present in domain VI of L protein[78,94], subsequently methylates the core cap structure at the ribose-2′-*O* position of the first nucleotide, followed by methylation at the guanine-N7 position, generating $GpppAm_{2'-O}$-RNA and $m^7GpppAm_{2'-O}$-RNA, respectively[95–97]. The capping reaction seems to be dependent on RNA length, indicating a possible spatial rearrangement in L protein[98]. Although no crystal structure is available yet, the MTase activities

Box 2 | The RNA-decapping pathway of viruses

Viruses cap and decap RNA, and many viruses regulate the decapping pathway in order to control the ratio of viral and cellular mRNAs.

Decapping of cellular mRNA by *Saccharomyces cerevisiae* L-A virus (from the *Totiviridae* family of double-stranded RNA viruses) proceeds through a decapping enzyme carried by the Gag subunit of the capsid; this Gag subunit is responsible for covalently binding cap structures (7-methyl-GpppG (m^{7}GpppG), in which p is a phosphate group) of cellular mRNA[158]. The decapping activity of Gag aids in the expression of viral RNA, apparently by producing large amounts of cellular RNA decoys that inhibit the *S. cerevisiae* enzyme 5′–3′ exoribonuclease 1 (Xrn1)[159] and compete with degradation of viral RNA. How the viral mRNA is recruited by the eukaryotic translation initiation factor 4E (eIF4E) complex remains to be elucidated. In the case of the family *Poxviridae* (double-stranded DNA viruses), the decapping enzyme (D10) increases the turnover of host mRNAs and contributes to the shutdown of host protein expression[160]. Moreover, D10 seems to preferentially degrade m^{7}GpppGm$_{2'\text{-O}}$ rather than m^{7}GpppAm$_{2'\text{-O}}$ and thereby hydrolyses early-phase viral RNA carrying predominantly m^{7}GpppGm$_{2'\text{-O}}$ cap structures[161]. In other words, this viral pathway benefits from having mRNAs (produced by the capping apparatus) that will be recruited by the eIF4E complex, and simultaneously removes the potential competition from cellular mRNAs for ribosome binding. Viral 'cap-snatching' (see main text) also results in this imbalance, favouring expression of viral genes. Finally, several viruses have been reported to interfere with the cellular RNA-trafficking and decoy machinery. First, a single-stranded positive-sense RNA (ss(+)RNA) enterovirus was shown to inhibit the ability of cells to form stress granules by cleaving RAS·GAP-binding protein (G3BP) family members. The kinetics of poliovirus-induced processing (P)-body disruption correlates with production of viral proteases that induce the degradation of P-body proteins, such as DCP1A and the 3′-deadenylase complex component PAN3. Recently, the ss(+)RNA viruses hepatitis C virus and HIV were also reported to be connected to the cellular decapping machinery and to regulate it (reviewed in REF. 162).

Cell-based replicon assays
Assays that allow one to follow the replication of a 'minimal viral genome' encoding the viral replication complex but no structural or envelope proteins, which are usually replaced by reporter genes (such as luciferase or chloramphenicol acetyl transferase genes).

share the AdoMet-binding and active site, which comprises a conserved Lys-Asp-Lys-Glu catalytic tetrad, typical of 2′OMTases.

The RNA-capping pathway of the alphavirus-like togaviruses. Alphaviruses (ss(+)RNA viruses, such as Semliki Forest virus, sindbis virus and chikungunya virus) synthesize a cap-0 structure through a non-conventional mechanism (FIG. 3b). The Semliki Forest virus replicase protein nsP1 has N7MTase and GTase activities, the latter still incompletely characterized, and is presumably involved in the capping of viral mRNA after nsP2-mediated cleavage of the β–γ phosphate bond at the 5′ end[99,100]. Although the GTase activity of nsP1 has not been fully demonstrated, GTP leads to the formation of a covalent enzyme–GMP complex, albeit only in the presence of AdoMet[99,101]. Accordingly, mutagenesis and *in vitro* assays have also revealed that the MTase catalyses the transfer of a methyl group from AdoMet to the N7 position of GTP before the formation of the covalent m^{7}GMP–enzyme complex. The covalent link involves a conserved catalytic histidine residue that is required for the GTase reaction but not for MTase activity. Interestingly, brome mosaic virus replicase protein 1A, bamboo mosaic virus nsp[102], tobacco mosaic virus P126 (REF. 103) and hepatitis E virus p110 (REF. 104) have properties similar to the N7MTase and GTase of alphaviruses, pointing to a common evolutionary origin for these distantly related viruses of plants and animals. Crystal structures are not yet available for any of these enzymes.

Virus-mediated RNA cap snatching

Among the ss(−)RNA viruses, those of the families *Arenaviridae*, *Bunyaviridae* and *Orthomyxoviridae* have a segmented RNA genome and form the order tentatively named Multinegavirales (FIG. 2). These viruses do not have a cap-synthesizing machinery. However, they have evolved to steal caps from host mRNAs in order to prime their own viral replication, in a process known as cap snatching (REFS 105–107). Cap snatching involves three steps (FIG. 3c). First, the 5′-methylated cap-1 or cap-2 structure of a host mRNA is bound by a specific site in the viral RdRp (or possibly the N protein[108]). Then, endonucleolytic cleavage of the cellular mRNA occurs several nucleotides downstream from the cap structure. Finally, this short, capped RNA is used as a primer for the synthesis of viral mRNA by the RdRp. The sequence, length and structure of the mRNA 5′ end that comes with the cap varies from one virus to the other. Most sequences are 15–20 nucleotides long[106,109–112], but arenaviruses, nairoviruses and thogotoviruses use shorter primers[113–116]. Following endonucleolytic cleavage, the 'decapped' cellular mRNAs (BOX 2) are targeted to the degradation machinery, resulting in the downregulation of cellular mRNAs.

Enzymes from the cap-snatching pathway. Cap snatching was found initially in influenza viruses (ss(−)RNA viruses of the family *Orthomyxoviridae*), which serve as a model system for the other two viral families known to use snatching (ss(−)RNA viruses of the families *Bunyaviridae* and *Arenaviridae*), although differences in the proteins involved and the lengths of snatched sequences are expected, as shown for Thogoto virus[117]. The influenza virus polymerase is made of three subunits: PA, PB1 and PB2. A cap-binding domain was found in the central region of the PB2 subunit[118], and an endonuclease domain at the N terminus of the PA subunit[119,120]. The structures of both domains shed light on the molecular mechanisms leading to cap snatching (see below). The cap-binding domain has a novel fold, although the mode of m^{7}G binding by aromatic stacking is similar to that used in other cap-binding proteins. By contrast, the endonuclease domain of PA has a fold that is characteristic of the two-metal-dependent PD(D/E)XK nuclease superfamily but has the peculiarity of a metal-ligating histidine residue in the active site, conferring Mn^{2+} specificity[121] (FIG. 5).

In contrast to orthomyxoviruses, both arenaviruses and bunyaviruses have a single protein (L) carrying the polymerase and cap-snatching activities. Recent studies showed that a Mn^{2+}-dependent endonuclease that is homologous to that of orthomyxoviruses exists at the N terminus of arenaviral and bunyaviral L protein[122,123] (FIG. 5a). Mutational analysis and cell-based replicon assays demonstrated that viral nuclease activities are essential for cap-dependent transcription of viral mRNA[124]. These domains have a conserved architecture and mechanism, which suggests an evolutionary link between them despite their low sequence identity (FIG. 5b).

A preliminary electron microscopy study of L protein from Machupo virus (an arenavirus) showed a central

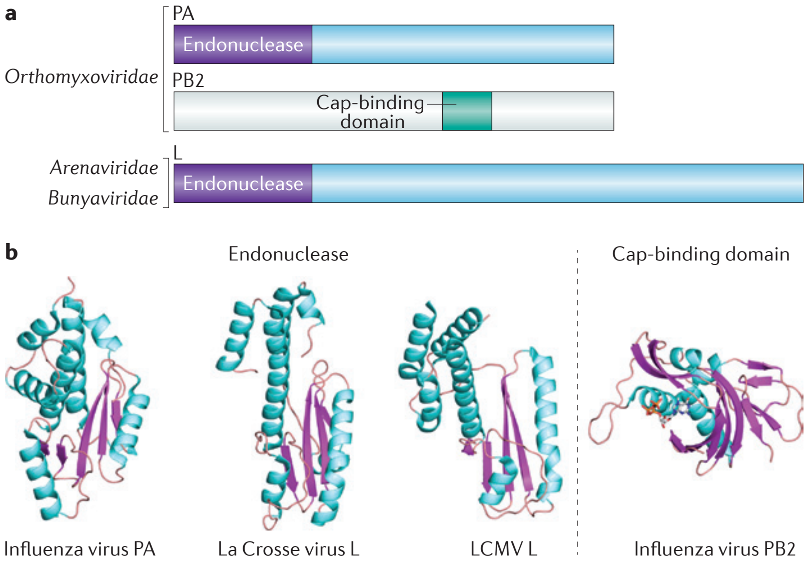


Figure 5 | **Unconventional capping machineries. Endonucleases and cap-binding PB2. a** | Domains involved in the 'cap-snatching' mechanism. The organization of cap-snatching domains of influenza viruses (single-stranded negative-sense RNA (ss(–) RNA) viruses of the family *Orthomyxoviridae*), and corresponding domains of distantly related viruses of the families *Arenaviridae* and *Bunyaviridae* (also ss(–)RNA viruses). Influenza virus polymerase is composed of three proteins of multiple domains: PA, PB1 and PB2. PA and PB2 are involved in cap snatching. PA carries the endonuclease domain in its amino terminus, whereas PB2 has an inner domain responsible for cap binding. Mapping of the domain organization for arenaviruses and bunyaviruses is less advanced; only the endonuclease domain is mapped to the amino terminus of L protein. The cap-binding domain is not clearly identified, as it is thought to be in either L protein or nucleocapsid (N or NP) protein, depending on the virus. **b** | Structures of the different endonuclease domains of viruses from the families *Orthomyxoviridae* (influenza viruses), *Arenaviridae* (lymphocytic choriomeningitis virus (LCMV)) and *Bunyaviridae* (La Crosse virus) (Protein Data Bank (PDB) accession codes 3HW4, 3JSB and 2XI5, respectively), and of the cap-binding domain from an influenza virus (PDB accession code 2VQZ). Despite having no sequence similarities, the folds of these endonuclease domains are conserved, suggesting a convergent evolution. Structures are coloured cyan for α-helices and pink for β-strands.

ring similar to the RdRp of dsRNA viruses but decorated with additional, non-conserved accessory domains. The authors speculate that these extra domains are involved in cap snatching, and that arenaviral and bunyaviral L proteins contain independently folded and functional domains, as does the influenza virus RdRp[125,126].

The recent nucleoprotein (NP) structure and enzymatic assay of Lassa virus (ss(–)RNA; *Arenaviridae*) revealed a second nuclease[108], which is probably involved in damping the interferon response[127], and a dTTP-binding site, which is proposed to be a cap-binding site involved in cap snatching. The idea that the cap-binding site for cap snatching may be in the nucleoprotein rather than the L protein had already been suggested for hantaviruses (ss(–)RNA; *Bunyaviridae*)[128]. In this model, L and N (as the NP protein is called in hantaviruses) cooperate in the process. Meanwhile, the structure of N from Rift Valley fever virus (also a bunyavirus)[129] does not present any of these features. These data call for caution before extending the assignation of such a function to related nucleoproteins.

Innate immunity and RNA capping

Mammalian cells have co-evolved with viruses and have developed several mechanisms to detect a viral infection (such as detecting uncapped or partially capped RNA, VPg–RNA, and so on) and induce an antiviral response in neighbouring cells[130–132]. This innate immunity is based on a small number of receptors called pattern recognition receptors (PRRs), which discriminate self from non-self components. Non-self detection depends on the recognition of a limited set of pathogen-associated molecular patterns (PAMPs), which are molecules or components that are characteristic of infectious agents, such as viral nucleic acids[133]. The presence of non-self nucleic acids is detected through sensors such as Toll-like receptors (TLRs), which recognize DNA or RNA in intracellular compartments that do not usually contain these molecules. Moreover, several PRRs sense the presence of foreign nucleic acids directly in the cytoplasm — namely, the NOD-like receptors (NLRs) and the retinoic acid-inducible gene (RIG)-like receptors (RLRs)[134]. As host cell RNA is present in the cytosol, PRRs sense uncommon RNA structures that are present in infected cells, such as dsRNA, RNA presenting a 5′-triphosphate, RNA with an incompletely methylated cap structure (cap-0, for mammalian standards) or RNA bearing a protein covalently attached to the 5′ end (such as VPg). The detection of PAMPs by PRRs triggers intracellular signalling events that mainly induce the production of type I interferon (IFN), interleukin-1 (IL-1) and pro-inflammatory cytokines, as well as the establishment of a cellular antiviral state in order to limit viral propagation[135,136] (FIG. 6).

Among the PRRs, the TLR family is the best studied innate immunity sensor family. TLRs are transmembrane proteins that are mainly expressed in immune cells, such as macrophages and dendritic cells. They are localized in endosomal compartments or at the cell surface. TLRs contain a leucine-rich repeat motif that recognizes PAMPs, and a Toll–IL-1 receptor (TIR) domain is present in the cytoplasmic part of the protein, ensuring signal transduction through TIR domain interaction with the TIR domains of cytoplasmic adaptor proteins such as MYD88 and TRIF (also known as TICAM1). TLR3, TLR7 and TLR8 are activated by different kinds of RNAs. TLR3 detects dsRNA, whereas TLR7 and TLR8 recognize ssRNA carrying a 5′-triphosphate or a cap-0 structure[137].

In contrast to TLRs, the NLRs and RLRs are localized in the cell cytoplasm[132] and detect the presence of intracellular invaders. Among the RLRs, RIG-I (also known as DDX58) and MDA5 (also known as IFIH1) seem to discriminate non-self RNA from self RNA on the basis of the RNA 5′ end[134]. Metazoan self RNA presents a 5′ cap-1 or cap-2 structure. RIG-I protein is specialized in the detection of 5′-triphosphate RNA, whereas MDA5 senses the presence of RNA with a cap-0 structure or linked to a protein such as VPg[138]. RIG-I and MDA5 consist of two N-terminal caspase-recruitment domains (CARDs), a central DEXH box-containing RNA helicase–ATPase domain and a C-terminal regulatory domain (CTD)[139]. It is likely that, in the absence

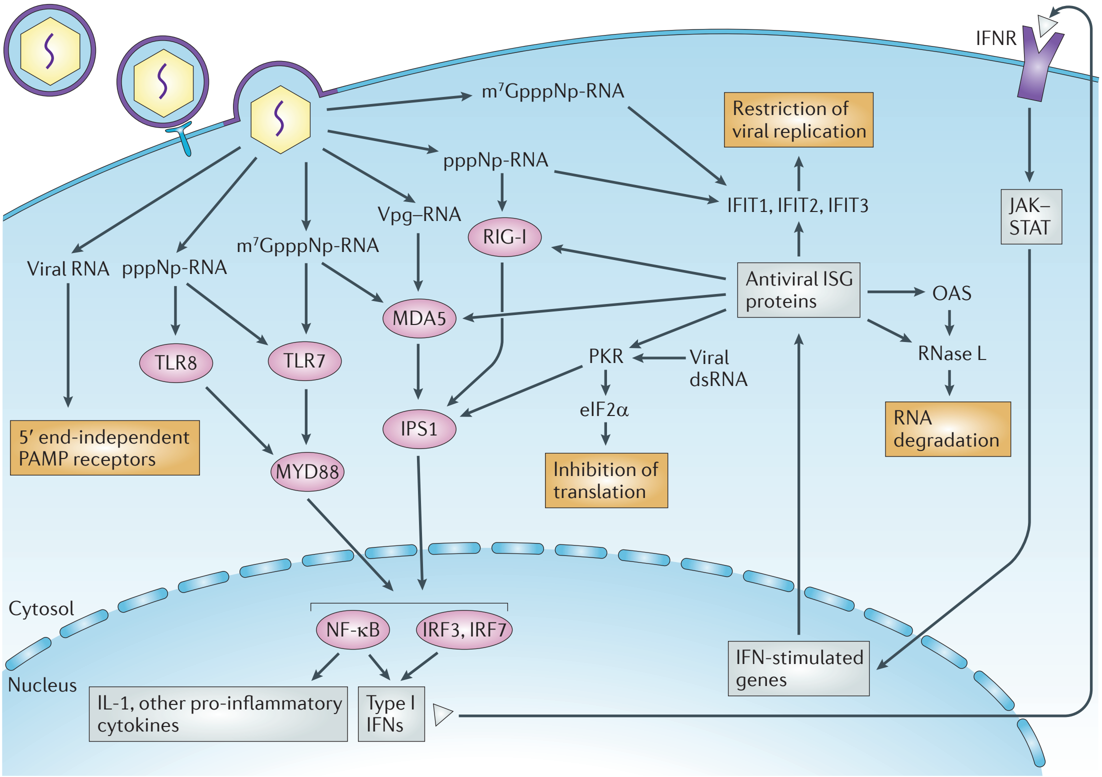


Figure 6 | **Sensing of viral RNA by the innate immune system.** A virus that infects a cell releases viral RNA into the host cytoplasm or endosomes in one of four forms: RNA protected by a 5′ cap-1 structure (7-methyl-Gppp-2′-*O*-methyl-NP-RNA, depicted here as m^{7}GpppNmp-RNA, in which N is the first transcribed nucleotide and p is a phosphate group), 5′-triphosphate RNA (pppNp-RNA), RNA linked to VPg or RNA carrying a 5′ cap-0 structure (m^{7}GpppNp-RNA). Cap-1 RNA is recognized by pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs), retinoic acid-inducible gene (RIG)-like receptors (RLRs) and NOD-like receptors (NLRs), which in this case recognize, for example, double-stranded RNA (dsRNA). In absence of a cap structure, the 5′-triphosphate of RNA is sensed by the RLR RIG-I, and the VPg–RNA and cap-0 RNA are recognized by another RLR, MDA5. Both RIG-I and MDA5 recruit the mitochondrial-anchored protein interferon-β (IFNβ) promoter stimulator 1 (IPS1), which in turn recruits appropriate inhibitor of NF-κB kinase (IKK) proteins to activate nuclear factor-κB (NF-κB) and IFN regulatory factors (IRFs). This results in the induction of type I IFNs and the production of pro-inflammatory cytokines such as interleukin-1 (IL-1). The endosomal receptor TLR7 can recognize cap-0 RNA and 5′-triphosphate RNA, whereas TLR8 recognizes only 5′-triphosphate RNA. TLR7 recruits the adaptor protein MYD88 through a Toll–IL-1 receptor (TIR)–TIR domain interaction, leading to the activation of NF-κB and IRF3 or IRF7; this results in the induction of type I IFNs and the production of IL-1 and other pro-inflammatory cytokines. Antiviral restriction factors are also stimulated by type I IFN receptor (IFNR) and are known to inhibit the replication of RNA viruses carrying non-capped genomes. Autocrine and paracrine IFN binds IFNR and initiates the JAK–STAT (Janus kinase–signal transducer and activator of transcription) signalling cascade. Among hundreds of proteins encoded by IFN-stimulated genes (ISGs), some antiviral proteins specifically target uncapped viral RNA. The IFIT (IFN-induced protein with tetratricopeptide repeats) proteins specifically sequester 5′-triphosphate RNA (IFIT1) or cap-0 RNA (IFIT2). Protein kinase, RNA activated (PKR) recognizes 5′-triphosphate RNA through an amino-terminal dsRNA-binding domain composed of two binding motifs. Activated PKR phosphorylates eukaryotic translation initiation factor 2α (eIF2α) through its kinase domain and blocks protein translation. RNase L is stimulated by 2′,5′oligo(A) oligonucleotides synthesized by oligoadenylyl synthases (OASs) and is also involved in the degradation of capped and non-capped viral RNA.

of ligands, the CARDs of RIG-I are auto-inhibited by other domains of the protein. For RIG-like proteins, nucleic acid binding to the RNA-binding site of the CTD induces a conformational change resulting in interaction of the CARD with the signalling adaptor molecule IFNβ promoter stimulator 1 (IPS1; also known as CARDIF, MAVS or VISA). IPS1 recruits a signalling complex in order to activate transcription factors such as interferon regulatory factor 3 (IRF3) and nuclear factor-κB (NF-κB), leading to the expression of IFNβ and other proteins that drive the antiviral response (FIG. 6).

The molecular basis for RNA recognition with and without 5′ end modification and/or overhanging nucleotides was analysed for RIG-I and MDA5. Initial studies indicated that RIG-I specifically recognizes 5′-triphosphate-containing ssRNA[140]. It was later found that RIG-I requires base-paired structures in conjunction with a 5′-triphosphate to trigger an antiviral response[141]. The molecular basis of the specific interaction between the RIG-I CTD and 5′-triphosphate dsRNA was deciphered by a crystallographical study of the CTD in complex with RNA[142]. This study revealed

Viral antigenomes
Viral RNAs that are complementary strands to the genome. The antigenome strand is used as a matrix for the synthesis of new viral genomes, and of viral mRNAs in the case of positive-sense RNA viruses.

that 5′-triphosphate dsRNA binds to the CTD of RIG-I more tightly than its single-stranded counterpart. The 5′-triphosphate is sequestered in a lysine-rich cleft of the CTD, with a phenylalanine residue stacked to the terminal base pair. Interestingly, 5′-triphosphate dsRNA methylated at the 2′-*O* position of its first or second nucleotide is expected to create a steric conflict with the CTD of RIG-I. Accordingly, it does not stimulate the RIG-I pathway.

In contrast to RIG-I, MDA5 is thought to recognize either a viral RNA 5′ end carrying structures that are distinct from a 5′-triphosphate or longer, structured (mesh) RNA that would be generated during the viral life cycle. Indeed, MDA5 was reported to sense both dsRNA and ssRNA bearing a 5′ cap-0 (REFS 138,143) or linked to VPg. Structural analysis of the MDA5 CTD indicated that its global fold is similar to that of the RIG-I CTD[144], with amino acid differences in the domains involved in the recognition of the RNA 5′ end.

The involvement of MDA5 and TLR7 in the detection of cap-0-containing RNA was recently characterized in the antiviral response observed for a coronavirus mutant lacking 2′OMTase activity[138]. The replication of this mutant virus was dramatically impaired in infected mice. However, the replication of this virus was restored in MDA5-, TLR7- or type I IFN receptor (IFNR)-deficient mice. It has also been suggested that other interferon-stimulated genes (ISGs) restrict the replication of 2′OMTase-deficient viruses[145]. Accordingly, IFIT1 (IFN-induced protein with tetratricopeptide repeats 1; also known as IFI56) and IFIT2 (also known as IFI54) were reported to limit the replication of West Nile virus (an ss(+)RNA virus of the genus Flavivirus), vaccinia virus (a dsDNA virus of the family *Poxviridae*) and murine hepatitis virus (an ss(+)RNA virus of the family *Coronaviridae*) lacking 2′OMTase activity[145]. IFIT1 was recently shown to bind and sequester 5′-triphosphate RNA into a multiprotein complex containing IFIT2 and IFIT3 (also known as IFI60) in order to exert its antiviral effect[146]. Therefore, it is likely that 2′-O-methylation of the RNA cap promotes escape from the host innate antiviral response through avoidance of IFIT-mediated suppression.

Conclusions

Since the discovery of 'blocked and methylated' mRNA ends nearly 40 years ago, viruses have played an essential part in deciphering the process of mRNA capping, as well as its relationship with various cellular processes such as transcription, translation and innate immunity. Viral RNA capping is a field that still has a lot of uncharted territory: whether the RNA 5′ ends are protected or not is still unknown for many neglected viral families, and the GTase resists identification even for some studied human pathogens (for example, the ss(+) RNA viruses of the order Nidovirales).

It is likely that, during their co-evolution with their hosts, viruses evolved different adaptation strategies to protect their RNA transcripts. The diversity of mechanisms expressed in nature to add a cap to an RNA 5′ end is larger than that described here. Future research will also aim to elucidate how the fine-tuning between host-mediated decapping of viral RNA, virus-mediated capping of viral RNA and host innate immunity is performed. Moreover, it has been shown that viral antigenomes are not capped[147], an observation that has now been extended to many different viruses. Thus, 'no capping' signals probably exist. In addition, the abundance of template RNAs must certainly need to be finely regulated for optimal viral replication. The corresponding signals and regulations are largely unknown.

Finally, owing to its spectacular mechanistic diversity, RNA capping is an attractive field for the design of antiviral drugs. Several molecules have been proposed to act directly or indirectly on viral RNA capping. Ribavirin is a broad-spectrum antiviral agent that is active against several viruses that add or snatch RNA caps, and its pleiotropic mechanism includes targeting the RNA-capping machinery[148,149]. So far, efforts to design MTase inhibitors have used the AdoMet and AdoHcy (*S*-adenosyl-L-homocysteine) backbone to synthesize analogues that are specific to viral enzymes[150]. The increasing knowledge about active-site differences between cellular and viral MTases is expected to provide antiviral selectivity. Last, inhibitors of cap-snatching endonuclease have long been known[119,151–153]. Recently published crystal structures[119,122,123] of their targets should inform antiviral-drug design projects.

1. Shatkin, A. Capping of eucaryotic mRNAs. *Cell* **9**, 645 (1976).
2. Darnell, J. E. Jr. Transcription units for mRNA production in eukaryotic cells and their DNA viruses. *Prog. Nucleic Acid Res. Mol. Biol.* **22**, 327–353 (1979).
3. Filipowicz, W. *et al.* A protein binding the methylated 5′-terminal sequence, m^{7}GpppN, of eukaryotic messenger RNA. *Proc. Natl Acad. Sci. USA* **73**, 1559–1563 (1976).
4. Schibler, U. & Perry, R. P. The 5′-termini of heterogeneous nuclear RNA: a comparison among molecules of different sizes and ages. *Nucleic Acids Res.* **4**, 4133–4149 (1977).
5. Liu, H. & Kiledjian, M. Decapping the message: a beginning or an end. *Biochem. Soc. Trans.* **34**, 35–38 (2006).
6. Nallagatla, S. R., Toroney, R. & Bevilacqua, P. C. A brilliant disguise for self RNA: 5′-end and internal modifications of primary transcripts suppress elements of innate immunity. *RNA Biol.* **5**, 140–144 (2008).
7. Rehwinkel, J. *et al.* RIG-I detects viral genomic RNA during negative-strand RNA virus infection. *Cell* **140**, 397–408 (2010).
8. Furuichi, Y. & Shatkin, A. J. Viral and cellular mRNA capping: past and prospects. *Adv. Vir. Res.* **55**, 135–184 (2000).
A historical and chronological perspective on the discovery of RNA capping and the structure of the RNA cap.
9. Furuichi, Y., Muthukrishnan, S. & Shatkin, A. J. 5′-Terminal m-^{7}G(5′)ppp(5′)G^{m}p *in vivo*: identification in reovirus genome RNA. *Proc. Natl Acad. Sci. USA* **72**, 742–745 (1975).
10. Shatkin, A. J. Methylated messenger RNA synthesis *in vitro* by purified reovirus. *Proc. Natl Acad. Sci. USA* **71**, 3204–3207 (1974).
11. Wei, C. M. & Moss, B. Methylated nucleotides block 5′-terminus of vaccinia virus messenger RNA. *Proc. Natl Acad. Sci. USA* **72**, 318–322 (1975).
12. Shatkin, A. J. & Both, G. W. Reovirus mRNA: transcription translation. *Cell* **7**, 305–313 (1976).
13. Furuichi, Y., Morgan, M., Muthukrishnan, S. & Shatkin, A. J. Reovirus messenger RNA contains a methylated, blocked 5′-terminal structure: m^{7}G(5′)ppp(5′)G^{m}pCp-. *Proc. Natl Acad. Sci. USA* **72**, 362–366 (1975).
14. Muthukrishnan, S., Both, G. W., Furuichi, Y. & Shatkin, A. J. 5′-Terminal 7-methylguanosine in eukaryotic mRNA is required for translation. *Nature* **255**, 33–37 (1975).
15. Marcotrigiano, J., Gingras, A. C., Sonenberg, N. & Burley, S. K. Cocrystal structure of the messenger RNA 5′ cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* **89**, 951–961 (1987).
16. Perry, K. L., Watkins, K. P. & Agabian, N. Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. *Proc. Natl Acad. Sci. USA* **84**, 8190–8194 (1987).
17. Beelman, C. A. & Parker, R. Degradation of mRNA in eukaryotes. *Cell* **81**, 179–183 (1995).
18. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763–776 (2009).

19. Otsuka, Y., Kedersha, N. L. & Schoenberg, D. R. Identification of a cytoplasmic complex that adds a cap onto 5′-monophosphate RNA. *Mol. Cell. Biol.* **29**, 2155–2167 (2009).

20. Schoenberg, D. R. & Maquat, L. E. Re-capping the message. *Trends Biochem. Sci.* **34**, 435–442 (2009).

21. Goodfellow, I. *et al.* Calicivirus translation initiation requires an interaction between VPg and eIF4E. *EMBO Rep.* **6**, 968–972 (2005).

22. Ambros, V., Pettersson, R. F. & Baltimore, D. An enzymatic activity in uninfected cells that cleaves the linkage between poliovirion RNA and the 5′ terminal protein. *Cell* **15**, 1439–1446 (1978).

23. Ambros, V. & Baltimore, D. Purification and properties of a HeLa cell enzyme able to remove the 5′-terminal protein from poliovirus RNA. *J. Biol. Chem.* **255**, 6739–6744 (1980).

24. Cong, P. & Shuman, S. Mutational analysis of mRNA capping enzyme identifies amino acids involved in GTP binding, enzyme-guanylate formation, and GMP transfer to RNA. *Mol. Cell. Biol.* **15**, 6222–6231 (1995).

25. De la Peña, M., Kyrieleis, O. J. P. & Cusack, S. Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap N7 methyl-transferase. *EMBO J.* **26**, 4913–4925 (2007).
Work showing activation of the MTase domain of D1 in complex with the D12 subunit at atomic resolution.

26. Mao, X. & Shuman, S. Intrinsic RNA (guanine-7) methyltransferase activity of the vaccinia virus capping enzyme D1 subunit is stimulated by the D12 subunit. Identification of amino acid residues in the D1 protein required for subunit association and methyl group transfer. *J. Biol. Chem.* **269**, 24472–24479 (1994).
The discovery and molecular basis of the activation of a viral MTase by a protein cofactor.

27. Schnierle, B. S., Gershon, P. D. & Moss, B. Cap-specific mRNA (nucleoside-$O^{2'}$-)-methyltransferase and poly(A) polymerase stimulatory activities of vaccinia virus are mediated by a single protein. *Proc. Natl Acad. Sci. USA* **89**, 2897–2901 (1992).

28. Benarroch, D. *et al.* The RNA helicase, nucleotide 5′-triphosphatase, and RNA 5′-triphosphatase activities of Dengue virus protein NS3 are Mg^{2+}-dependent and require a functional Walker B motif in the helicase catalytic core. *Virology* **328**, 208–218 (2004).

29. Myette, J. R. & Niles, E. G. Characterization of the vaccinia virus RNA 5′-triphosphatase and nucleotide triphosphate phosphohydrolase activities demonstrate that both activities are carried out at the same active site. *J. Biol. Chem.* **271**, 11945–11952 (1996).

30. Vasquez-Del Carpio, R., Gonzalez-Nilo, F. D., Riadi, G., Taraporewala, Z. F. & Patton, J. T. Histidine triad-like motif of the rotavirus NSP2 octamer mediates both RTPase and NTPase activities. *J. Mol. Biol.* **362**, 539–554 (2006).

31. Jayaram, H., Taraporewala, Z., Patton, J. T. & Prasad, B. V. V. Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold. *Nature* **417**, 311–315 (2002).
The first crystal structure of an RTPase with a HIT-like fold.

32. Taraporewala, Z., Chen, D. & Patton, J. T. Multimers formed by the rotavirus nonstructural protein NSP2 bind to RNA and have nucleoside triphosphatase activity. *J. Virol.* **73**, 9934–9943 (1999).

33. Taraporewala, Z. F. & Patton, J. T. Identification and characterization of the helix-destabilizing activity of rotavirus nonstructural protein NSP2. *J. Virol.* **75**, 4519–4527 (2001).

34. Benarroch, D., Smith, P. & Shuman, S. Characterization of a trifunctional mimivirus mRNA capping enzyme and crystal structure of the RNA triphosphatase domain. *Structure* **16**, 501–512 (2008).

35. Gu, M. & Lima, C. D. Processing the message: structural insights into capping and decapping mRNA. *Curr. Opin. Struc. Biol.* **15**, 99–106 (2005).

36. Lima, C. D., Wang, L. K. & Shuman, S. Structure and mechanism of yeast RNA triphosphatase: an essential component of the mRNA capping apparatus. *Cell* **99**, 533–543 (1999).
The first crystal structure of an RTPase involved in RNA capping.

37. Shuman, S. Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog. Nucleic Acid Res. Mol. Biol.* **66**, 1–40 (2001).

38. Lehman, K., Schwer, B., Ho, C. K., Rouzankina, I. & Shuman, S. A conserved domain of yeast RNA triphosphatase flanking the catalytic core regulates self-association and interaction with the guanylyltransferase component of the mRNA capping apparatus. *J. Biol. Chem.* **274**, 22668–22678 (1999).

39. Gorbalenya, A. E. & Koonin, E. V. One more conserved sequence motif in helicases. *Nucleic Acids Res.* **16**, 7734 (1988).

40. Gorbalenya, A. E. & Koonin, E. V. Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res.* **17**, 8413–8440 (1989).

41. Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17**, 4713–4730 (1989).

42. Luo, D. *et al.* Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. *EMBO J.* **27**, 3209–3219 (2008).

43. Gu, M. & Rice, C. M. Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl Acad. Sci. USA* **107**, 521–528 (2010).

44. Lescar, J. *et al.* Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. *Antiviral Res.* **80**, 94–101 (2008).

45. Luo, D. *et al.* Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications. *J Biol. Chem.* **285**, 18817–18827 (2010).

46. Luo, D. *et al.* Crystal structure of the NS3 protease-helicase from dengue virus. *J. Virol.* **82**, 173–183 (2008).

47. Sampath, A. *et al.* Structure-based mutational analysis of the NS3 helicase from dengue virus. *J. Virol.* **80**, 6686–6690 (2006).

48. Balistreri, G., Caldentey, J., Kääriäinen, L. & Ahola, T. Enzymatic defects of the nsP2 proteins of Semliki Forest virus temperature-sensitive mutants. *J Virol.* **81**, 2849–2860 (2007).

49. Changela, A., Ho, C. K., Martins, A., Shuman, S. & Mondragón, A. Structure and mechanism of the RNA triphosphatase component of mammalian mRNA capping enzyme. *EMBO J.* **20**, 2575–2586 (2001).

50. Takagi, T., Moore, C. R., Diehn, F. & Buratowski, S. An RNA 5′-triphosphatase related to the protein tyrosine phosphatases. *Cell* **89**, 867–873 (1997).

51. Shuman, S. & Lima, C. D. The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Curr. Opin. Struct. Biol.* **14**, 757–764 (2004).

52. Wang, S. P., Deng, L., Ho, C. K. & Shuman, S. Phylogeny of mRNA capping enzymes. *Proc. Natl Acad. Sci. USA* **94**, 9573–9578 (1997).

53. Cong, P. & Shuman, S. Covalent catalysis in nucleotidyl transfer. A KTDG motif essential for enzyme-GMP complex formation by mRNA capping enzyme is conserved at the active sites of RNA and DNA ligases. *J. Biol. Chem.* **268**, 7256–7260 (1993).

54. Niles, E. G. & Christen, L. Identification of the vaccinia virus mRNA guanyltransferase active site lysine. *J. Biol. Chem.* **268**, 24986–24989 (1993).

55. Shuman, S. & Hurwitz, J. Mechanism of mRNA capping by vaccinia virus guanylyltransferase: characterization of an enzyme–guanylate intermediate. *Proc. Natl Acad. Sci. USA* **78**, 187–191 (1981).

56. Schwer, B. & Shuman, S. Mutational analysis of yeast mRNA capping enzyme. *Proc. Natl Acad. Sci. USA* **91**, 4328–4332 (1994).

57. Lindahl, T. & Barnes, D. E. Mammalian DNA ligases. *Annu. Rev. Biochem.* **61**, 251–281 (1992).

58. Chu, C. *et al.* Structure of the guanylyltransferase domain of human mRNA capping enzyme. *Proc. Natl Acad. Sci. USA* **108**, 10104–10108 (2011).

59. Fabrega, C., Shen, V., Shuman, S. & Lima, C. D. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **11**, 1549–1561 (2003).

60. Håkansson, K., Doherty, A. J., Shuman, S. & Wigley, D. B. X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell* **89**, 545–553 (1997).
A study that deciphers the GTase reaction, with corresponding crystal structure snapshots.

61. Håkansson, K. & Wigley, D. B. Structure of a complex between a cap analogue and mRNA guanylyl transferase demonstrates the structural chemistry of RNA capping. *Proc. Natl Acad. Sci. USA* **95**, 1505–1510 (1998).

62. Reinisch, K. M., Nibert, M. L. & Harrison, S. C. Structure of the reovirus core at 3.6 A resolution. *Nature* **404**, 960–967 (2000).
A report detailing the complete orthoreoviral RNA cap assembly line at atomic resolution.

63. Sutton, G., Grimes, J. M., Stuart, D. I. & Roy, P. Bluetongue virus VP4 is an RNA-capping assembly line. *Nature Struct. Mol. Biol.* **14**, 449–451 (2007).

64. Cheng, L. *et al.* Atomic model of a cypovirus built from cryo-EM structure provides insight into the mechanism of mRNA capping. *Proc. Natl Acad. Sci. USA* **108**, 1373–1378 (2011).

65. Kim, J., Tao, Y., Reinisch, K. M., Harrison, S. C. & Nibert, M. L. Orthoreovirus and Aquareovirus core proteins: conserved enzymatic surfaces, but not protein–protein interfaces. *Virus Res.* **101**, 15–28 (2004).

66. Luongo, C. L., Reinisch, K. M., Harrison, S. C. & Nibert, M. L. Identification of the guanylyltransferase region and active site in reovirus mRNA capping protein λ2. *J. Biol. Chem.* **275**, 2804–2810 (2000).

67. Issur, M. *et al.* The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure. *RNA* **15**, 2340–2350 (2009).

68. Bisaillon, M. & Lemay, G. Viral and cellular enzymes involved in synthesis of mRNA cap structure. *Virology* **236**, 1–7 (1997).

69. Egloff, M. P., Benarroch, D., Selisko, B., Romette, J. L. & Canard, B. An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J.* **21**, 2757–2768 (2002).
Characterization of the first (+)RNA virus MTase at atomic resolution.

70. Martin, J. L. & McMillan, F. M. SAM (dependent) I AM: the *S*-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12**, 783–793 (2002).

71. Schubert, H. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* **28**, 329–335 (2003).

72. Decroly, E. *et al.* Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2′-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathog.* **7**, e1002059 (2011).

73. Egloff, M. P. *et al.* Structural and functional analysis of methylation and 5′-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. *J. Mol. Biol.* **372**, 723–736 (2007).

74. Dong, H., Ren, S., Li, H. & Shi, P. Y. Separate molecules of West Nile virus methyltransferase can independently catalyze the N7 and 2′-*O* methylations of viral RNA cap. *Virology* **377**, 1–6 (2008).

75. Ray, D. *et al.* West Nile virus 5′-cap structure is formed by sequential guanine N-7 and ribose 2′-O methylations by nonstructural protein 5. *J. Virol.* **80**, 8362–8370 (2006).

76. Fabrega, C., Hausmann, S., Shen, V., Shuman, S. & Lima, C. D. Structure and mechanism of mRNA cap (guanine-N7) methyltransferase. *Mol. Cell* **13**, 77–89 (2004).

77. Bujnicki, J. M. & Rychlewski, L. Reassignment of specificities of two cap methyltransferase domains in the reovirus λ2 protein. *Genome Biol.* **2**, RESEARCH0038.1–RESEARCH0038.6 (2001).

78. Ferron, F., Longhi, S., Henrissat, B. & Canard, B. Viral RNA-polymerases – a predicted 2′-O-ribose methyltransferase domain shared by all *Mononegavirales*. *Trends Biochem. Sci.* **27**, 222–224 (2002).

79. Hodel, A. E., Gershon, P. D., Shi, X. & Quiocho, F. A. The 1.85 Å structure of vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends. *Cell* **85**, 247–256 (1996).
The first crystal structure of a viral enzyme involved in RNA capping.

80. Lockless, S. W., Cheng, H. T., Hodel, A. E., Quiocho, F. A. & Gershon, P. D. Recognition of capped RNA substrates by VP39, the vaccinia virus-encoded mRNA cap-specific 2′-*O*-methyltransferase. *Biochemistry* **37**, 8564–8574 (1998).

81. Li, C., Xia, Y., Gao, X. & Gershon, P. D. Mechanism of RNA 2′-O-methylation: evidence that the catalytic lysine acts to steer rather than deprotonate the target nucleophile. *Biochemistry* **43**, 5680–5687 (2004).

82. Hodel, A. E., Gershon, P. D. & Quiocho, F. A. Structural basis for sequence-nonspecific recognition of 5′-capped mRNA by a cap-modifying enzyme. *Mol. Cell* **1**, 443–447 (1998).
An article discussing the molecular basis of N7-methylated-cap selectivity and detailing the first RNA cap protein crystal structure.

83. Bouvet, M. *et al. In vitro* reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog.* **6**, e1000863 (2010).

84. Chen, Y. *et al*. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc. Natl Acad. Sci USA* **106**, 3484–3489 (2009).

85. Decroly, E. *et al.* Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2′*O*)-methyltransferase activity. *J. Virol.* **82**, 8071–8084 (2008).

86. Ramadevi, N., Burroughs, N. J., Mertens, P. P., Jones, I. M. & Roy, P. Capping and methylation of mRNA by purified recombinant VP4 protein of bluetongue virus. *Proc. Natl Acad. Sci. USA* **95**, 13537–13542 (1998).

87. Ramadevi, N. & Roy, P. Bluetongue virus core protein VP4 has nucleoside triphosphate phosphohydrolase activity. *J. Gen. Virol.* **79**, 2475–2480 (1998).

88. Abraham, G., Rhodes, D. P. & Banerjee, A. K. The 5′ terminal structure of the methylated mRNA synthesized *in vitro* by vesicular stomatitis virus. *Cell* **5**, 51–58 (1975).

89. Abraham, G., Rhodes, D. P. & Banerjee, A. K. Novel initiation of RNA synthesis *in vitro* by vesicular stomatitis virus. *Nature* **255**, 37–40 (1975).

90. Gupta, K. C. & Roy, P. Alternate capping mechanisms for transcription of spring viremia of carp virus: evidence for independent mRNA initiation. *J. Virol.* **33**, 292–303 (1980).

91. Barik, S. The structure of the 5′ terminal cap of the respiratory syncytial virus mRNA. *J. Gen. Virol.* **74**, 485–490 (1993).

92. Ogino, T. & Banerjee, A. K. The HR motif in the RNA-dependent RNA polymerase L protein of Chandipura virus is required for unconventional mRNA-capping activity. *J. Gen. Virol.* **91**, 1311–1314 (2010).

93. Ogino, T. & Banerjee, A. K. Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol. Cell* **25**, 85–97 (2007).
The elucidation of the unconventional RNA-capping pathway of (−)RNA viruses.

94. Bujnicki, J. M. & Rychlewski, L. *In silico* identification, structure prediction and phylogenetic analysis of the 2′-*O*-ribose (cap 1) methyltransferase domain in the large structural protein of ssRNA negative-strand viruses. *Protein Eng.* **15**, 101–108 (2002).

95. Li, J., Fontaine-Rodriguez, E. C. & Whelan, S. P. Amino acid residues within conserved domain VI of the vesicular stomatitis virus large polymerase protein essential for mRNA cap methyltransferase activity. *J. Virol.* **79**, 13373–13384 (2005).

96. Rahmeh, A. A., Li, J., Kranzusch, P. J. & Whelan, S. P. J. Ribose 2′-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N-7 methylation by the large polymerase protein. *J. Virol.* **83**, 11043–11050 (2009).
A paper with significant implications for understanding the mechanisms of unconventional capping in VSV. The authors show the sequence requirements for methylation and the chain of events that characterize the mechanism.

97. Testa, D. & Banerjee, A. K. Two methyltransferase activities in the purified virions of vesicular stomatitis virus. *J. Virol.* **24**, 786–793 (1977).

98. Tekes, G., Rahmeh, A. A. & Whelan, S. P. A freeze frame view of vesicular stomatitis virus transcription defines a minimal length of RNA for 5′ processing. *PLoS Pathog.* **7**, e1002073 (2011).

99. Ahola, T. & Kääriäinen, L. Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc. Natl Acad. Sci. USA* **92**, 507–511 (1995).
The discovery of the unconventional RNA-capping pathway of alphaviruses, in which guanine-N7 methylation of GTP precedes m^{7}GMP transfer onto the putative GTase, nsP1.

100. Vasiljeva, L., Merits, A., Auvinen, P. & Kääriäinen, L. Identification of a novel function of the alphavirus capping apparatus. RNA 5′-triphosphatase activity of Nsp2. *J. Biol. Chem.* **275**, 17281–17287 (2000).

101. Ahola, T., Laakkonen, P., Vihinen, H. & Kääriäinen, L. Critical residues of Semliki Forest virus RNA capping enzyme involved in methyltransferase and guanylyltransferase-like activities. *J. Virol.* **71**, 392–397 (1997).

102. Li, Y. I., Chen, Y. J., Hsu, Y. H. & Meng, M. Characterization of the AdoMet-dependent guanylyltransferase activity that is associated with the N terminus of bamboo mosaic virus replicase. *J. Virol.* **75**, 782–788 (2001).

103. Merits, A. *et al.* Virus-specific capping of tobacco mosaic virus RNA: methylation of GTP prior to formation of covalent complex p126-^{m}7GMP. *FEBS Lett.* **455**, 45–48 (1999).

104. Magden, J. *et al.* Virus-specific mRNA capping enzyme encoded by hepatitis E virus. *J. Virol.* **75**, 6249–6255 (2001).

105. Bouloy, M., Plotch, S. J. & Krug, R. M. Globin mRNAs are primers for the transcription of influenza viral RNA *in vitro*. *Proc. Natl Acad. Sci. USA* **75**, 4886–4890 (1978).

106. Caton, A. J. & Robertson, S. Structure of the host-derived sequences present at the 5′ ends of influenza virus mRNA. *Nucleic Acids Res.* **8**, 2591–2603 (1980).

107. Plotch, S. J., Bouloy, M. & Krug, R. M. Transfer of 5′-terminal cap of globin mRNA to influenza viral complementary RNA during transcription *in vitro*. *Proc. Natl Acad. Sci. USA* **76**, 1618–1622 (1979).

108. Qi, X. *et al.* Cap binding and immune evasion revealed by Lassa nucleoprotein structure. *Nature* **468**, 779–783 (2010).

109. Bishop, D. H. M. Y., Gay, M. E. & Matsuoko, Y. Non Viral heterogeneous sequences are present at the 5′ ends of one species of snowshoe hare bunyavirus S. complementary RNA. *Nucleic Acids Res.* **11**, 6409–6418 (1983).

110. Bouloy, M., Pardigon, N., Vialat, P., Gerbaud, S. & Girard, M. Characterization of the 5′ and 3′ ends of viral messenger RNAs isolated from BHK21 cells infected with Germiston virus (Bunyavirus). *Virology* **175**, 50–58 (1990).

111. Eshita, Y., Ericson, B., Romanowski, V. & Bishop, D. H. Analyses of the mRNA transcription processes of snowshoe hare bunyavirus S and M RNA species. *J. Virol.* **55**, 681–689 (1985).

112. Patterson, J. L., Holloway, B. & Kolakofsky, D. La Crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependent endonuclease. *J. Virol.* **52**, 215–222 (1984).

113. Garcin, D. *et al.* The 5′ ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis. *J. Virol.* **69**, 5754–5762 (1995).

114. Jin, H. & Elliott, R. M. Non-viral sequences at the 5′ ends of Dugbe nairovirus S mRNAs. *J. Gen. Virol.* **74**, 2293–2297 (1993).

115. Raju, R. *et al.* Nontemplated bases at the 5′ends of Tacaribe virus mRNAs. *Virology* **174**, 53–59 (1990).

116. Weber, F., Haller, O. & Kochs, G. Nucleoprotein viral RNA and mRNA of Thogoto virus: a novel" cap-stealing" mechanism in tick-borne orthomyxoviruses? *J. Virol.* **70**, 8361–8367 (1996).

117. Leahy, M. B., Dessens, J. T. & Nuttall, P. A. *In vitro* polymerase activity of Thogoto virus: evidence for a unique cap-snatching mechanism in a tick-borne orthomyxovirus. *J. Virol.* **71**, 8347–8351 (1997).

118. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Mol. Biol.* **15**, 500–506 (2008).

119. Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458**, 914–918 (2009).

120. Yuan, P. *et al.* Crystal structure of an avian influenza polymerase PA$_N$ reveals an endonuclease active site. *Nature* **458**, 909–913 (2009).
Together with reference 119, this paper describes the structural and functional characterization of the cap-snatching endonuclease.

121. Crépin, T. *et al.* Mutational and metal binding analysis of the endonuclease domain of the influenza virus polymerase PA subunit. *J. Virol.* **84**, 9096–9104 (2010).

122. Reguera, J., Weber, F. & Cusack, S. Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. *PLoS Pathog.* **6**, e1001101 (2010).

123. Morin, B. *et al.* The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription. *PLoS Pathog.* **6**, e1001038 (2010).

124. Lelke, M., Brunotte, L., Busch, C. & Günther, S. An N-terminal region of Lassa virus L protein plays a critical role in transcription but not replication of the virus genome. *J. Virol.* **84**, 1934–1944 (2010).

125. Kranzusch, P. J. *et al.* Assembly of a functional Machupo virus polymerase complex. *Proc. Natl Acad. Sci. USA* **107**, 20069–20074 (2010).

126. Ruigrok, R. W. H., Crépin, T., Hart, D. J. & Cusack, S. Towards an atomic resolution understanding of the influenza virus replication machinery. *Curr. Opin. Struct. Biol.* **20**, 104–113 (2010).

127. Hastie, K. M., Kimberlin, C. R., Zandonatti, M. A., MacRae, I. J. & Saphire, E. O. Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3′ to 5′ exonuclease activity essential for immune suppression. *Proc. Natl Acad. Sci. USA* **108**, 2396–2401 (2011).

128. Mir, M. A., Duran, W. A., Hjelle, B. L., Ye, C. & Panganiban, A. T. Storage of cellular 5′ mRNA caps in P bodies for viral cap-snatching. *Proc. Natl Acad. Sci. USA* **105**, 19294–19299 (2008).
An article that casts light on how N protein binds preferentially to capped mRNAs, stores and protects these mRNAs in P-bodies, and potentially takes an active role in cap acquisition.

129. Ferron, F. *et al.* The hexamer structure of the rift valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. *PLoS Pathog.* **7**, e1002030 (2011).

130. Koyama, S., Ishii, K. J., Coban, C. & Akira, S. Innate immune response to viral infection. *Cytokine* **43**, 336–341 (2008).

131. Takeuchi, O. & Akira, S. Recognition of viruses by innate immunity. *Immunol. Rev.* **220**, 214–224, (2007).

132. Wilkins, C. & Gale, M., Jr Recognition of viruses by cytoplasmic sensors. *Curr. Opin. Immunol.* **22**, 41–47 (2010).

133. Yoneyama, M. & Fujita, T. Recognition of viral nucleic acids in innate immunity. *Rev. Med. Virol.* **20**, 4–22 (2010).

134. Brennan, K. & Bowie, A. G. Activation of host pattern recognition receptors by viruses. *Curr. Opin. Microbiol.* **13**, 503–507 (2010).

135. Hansen, J. D., Vojtech, L. N. & Laing, K. J. Sensing disease and danger: a survey of vertebrate PRRs and their origins. *Dev. Comp. Immunol.* **35**, 886–897 (2011).

136. Meylan, E., Tschopp, J. & Karin, M. Intracellular pattern recognition receptors in the host response. *Nature* **442**, 39–44 (2006).

137. Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S. & Reis e Sousa, C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* **303**, 1529–1531 (2004).

138. Züst, R. *et al.* Ribose 2′-*O*-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nature Immunol.* **12**, 137–143 (2011).
The authors show for the first time the role of 2′-*O*-methylation in the sensing of self RNA by the innate immune system through the RNA sensor MDA5.

139. Oshiumi, H., Sakai, K., Matsumoto, M. & Seya, T. DEAD/H BOX 3 (DDX3) helicase binds the RIG-I adaptor IPS-1 to up-regulate IFN-β-inducing potential. *Eur. J. Immunol.* **40**, 940–948 (2010).

140. Pichlmair, A. *et al.* RIG-I-mediated antiviral responses to single-stranded RNA bearing 5′-phosphates. *Science* **314**, 997–1001 (2006).
The 5′-triphosphate of viral RNA is identified as a major component in the RIG-I-mediated host innate immune response.

141. Schmidt, A. *et al.* 5′-triphosphate RNA requires base-paired structures to activate antiviral signaling via RIG-I. *Proc. Natl Acad. Sci. USA* **106**, 12067–12072 (2009).

142. Wang, Y. *et al.* Structural and functional insights into 5′-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nature Struct. Mol. Biol.* **17**, 781–787 (2010).

143. Luthra, P., Sun, D., Silverman, R. H. & He, B. Activation of IFN-β expression by a viral mRNA through RNase L and MDA5. *Proc. Natl Acad. Sci. USA* **108**, 2118–2123 (2011).

144. Li, X. *et al.* Structural basis of double-stranded RNA recognition by the RIG-I like receptor MDA5. *Arch. Biochem. Biophys.* **488**, 23–33 (2009).

145. Daffis, S. *et al.* 2′-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* **468**, 452–456 (2010).
The authors demonstrate that 2′-*O*-methylation of the viral RNA cap evades host innate antiviral responses through escape of IFIT-mediated suppression.

146. Pichlmair, A. *et al.* IFIT1 is an antiviral protein that recognizes 5′-triphosphate RNA. *Nature Immunol*. **12**, 624–630 (2011).
147. Garcin, D. & Kolakofsky, D. A novel mechanism for the initiation of Tacaribe arenavirus genome replication. *J. Virol.* **64**, 6196–6203 (1990).
148. Hong, Z. & Cameron, C. E. Pleiotropic mechanisms of ribavirin antiviral activities. *Prog. Drug Res.* **59**, 41–69 (2002).
149. Magden, J., Kääriäinen, L. & Ahola, T. Inhibitors of virus replication: recent developments and prospects. *Appl. Microbiol. Biotechnol.* **66**, 612–621 (2005).
150. Lim, S. P. *et al.* Small molecule inhibitors that selectively block dengue virus methyltransferase. *J. Biol. Chem.* **286**, 6233–6240 (2011).
151. Kuzuhara, T., Iwai, Y., Takahashi, H., Hatakeyama, D. & Echigo, N. Green tea catechins inhibit the endonuclease activity of influenza A virus RNA polymerase. *PLoS Curr.* **1**, RRN1052 (2009).
152. Parkes, K. E. B. *et al.* Use of a pharmacophore model to discover a new class of influenza endonuclease inhibitors. *J. Med. Chem.* **46**, 1153–1164 (2003).
153. Tomassini, J. *et al.* Inhibition of cap (m^{7}GpppXm)-dependent endonuclease of influenza virus by 4-substituted 2,4-dioxobutanoic acid compounds. *Antimicrob. Agents Chemother.* **38**, 2827–2837 (1994).
154. Balvay, L., Soto Rifo, R., Ricci, E. P., Decimo, D. & Ohlmann, T. Structural and functional diversity of viral IRESes. *Biochim. Biophys. Acta* **1789**, 542–557 (2009).
155. Guidotti, L. G. & Chisari, F. V. Noncytolytic control of viral infections by the innate and adaptive immune response. *Annu. Rev. Immunol.* **19**, 65–91 (2001).
156. Malmgaard, L. Induction and regulation of IFNs during viral infections. *J. Interferon Cytokine Res.* **24**, 439–454 (2004).
157. Garaigorta, U. & Chisari, F. V. Hepatitis C virus blocks interferon effector function by inducing protein kinase R phosphorylation. *Cell Host Microbe* **6**, 513–522 (2009).
158. Blanc, A., Goyer, C. & Sonenberg, N. The coat protein of the yeast double-stranded RNA virus L-A attaches covalently to the cap structure of eukaryotic mRNA. *Mol. Cell. Biol.* **12**, 3390–3398 (1992).
159. Naitow, H., Tang, J., Canady, M., Wickner, R. B. & Johnson, J. E. L-A virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism. *Nature Struct. Biol.* **9**, 725–728 (2002).
160. Parrish, S., Resch, W. & Moss, B. Vaccinia virus D10 protein has mRNA decapping activity, providing a mechanism for control of host and viral gene expression. *Proc. Natl Acad. Sci. USA* **104**, 2139–2144 (2007).
161. McLennan, A. G. Decapitation: poxvirus makes RNA lose its head. *Trends Biochem. Sci.* **32**, 297–299 (2007).
162. Gaglia, M. M. & Glaunsinger, B. A. Viruses and the cellular RNA decay machinery. *Wiley Interdiscip. Rev. RNA* **1**, 47–59 (2010).

Acknowledgements

The authors are in debt to all previous and current laboratory members, too numerous to name, for their contributions and involvement in the study of viral RNA replication and capping. Special thanks go to B. Selisko for her tireless dedication to and help with the scientific elaboration of this manuscript. This work was supported in part through funding by the Fondation pour la Recherche Médicale (Programme Aide aux équipes), the French Direction Générale de l'Armement (contrat 07co404), Infectiopôle-Sud and the European Union Seventh Framework Programme (FP7/2007–2013) through the project SILVER (Small inhibitor leads against emerging RNA viruses; grant agreement 260644).

Competing interests statement

The authors declare no competing financial interests.

DATABASES

Protein Data Bank: http://www.cancer.gov/drugdictionary
1AV6 | 1CKM | 1CKN | 1CKO | 1EJ6 | 1L9V | 1YN9 | 2BHR | 2JHC | 2P41 | 2QZE | 2VDW | 2VQZ | 2XI5 | 3HW4 | 3JSB

FURTHER INFORMATION

Bruno Canard's homepage:
http://www.afmb.univ-mrs.fr/Bruno-Canard

PyMOL: http://www.pymol.org

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

D. Copie des Publications

REVIEW

A Practical Overview of Protein Disorder Prediction Methods

François Ferron,[1,2] Sonia Longhi,[1*] Bruno Canard,[1] and David Karlin[3]
[1]*Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Universités Aix-Marseille I et II, Marseille, France*
[2]*Boston Biomedical Research Institute, Watertown, Massachusetts 02472*
[3]*Ecole de l'ADN, INMED, Marseille, France*

ABSTRACT **In the past few years there has been a growing awareness that a large number of proteins contain long disordered (unstructured) regions that often play a functional role. However, these disordered regions are still poorly detected. Recognition of disordered regions in a protein is important for two main reasons: reducing bias in sequence similarity analysis by avoiding alignment of disordered regions against ordered ones, and helping to delineate boundaries of protein domains to guide structural and functional studies. As none of the available method for disorder prediction can be taken as fully reliable on its own, we present an overview of the methods currently employed highlighting their advantages and drawbacks. We show a few practical examples of how they can be combined to avoid pitfalls and to achieve more reliable predictions. Proteins 2006;65:1–14.**
© 2006 Wiley-Liss, Inc.

Key words: protein disorder; prediction methods

INTRODUCTION

Intrinsically unstructured/disordered or natively unfolded proteins have a broad occurrence in living organisms. They are characterized by the lack of stable secondary and tertiary structure under physiological conditions and in the absence of a binding partner/ligand. Intrinsically disordered proteins fulfil essential functions, which are often linked with their disordered structural state.

A protein region is defined as disordered if it is devoid of stable secondary structure and if it has a large number of conformations as seen using methods such as X-ray crystallography (lack of electron density), nuclear magnetic resonance (NMR), circular dichroism (CD), small-angle X-ray scattering, and various hydrodynamic measurements.[1,2] However, this definition embraces several categories of disorder: molten globules, partially unstructured proteins (premolten globules), and random coils (by increasing mobility and decreasing residual secondary structure content (see Uversky[3]).

What is the practical interest of identifying disordered regions? Disorder prediction is an essential prerequisite to protein sequence analysis. Disordered regions often have a biased amino acid composition that can lead to spurious sequence similarity with unrelated proteins. The recognition of regions of disorder is thus crucial to avoid spurious sequence alignments with sequences of globular proteins (for examples, see Iyer et al.[4]). Moreover, the recognition of disordered regions facilitates the identification of Eukaryotic Linear Motifs (ELMs), which are short functional motifs occurring mainly (>70%) within disordered regions (e.g., SH3, PDZ, phosphorylation sites[5–7]).

Second, disordered regions often prevent crystallization of proteins, or the generation of interpretable NMR data. Therefore, structural biologists use disorder predictions to delineate compact domains to solve their 3D structure, or to dissect target sequences into a set of independently folded domains in order to facilitate tertiary structure and threading predictions.[8]

Although the identification of disordered regions of less than 20 residues in length is generally thought to be less accurate,[9] recent results suggest that progress has been made in predicting short disordered regions.[10,11] Accordingly, we herein consider also short (i.e., less than 20 residues) regions of disorder.

As in other areas of bioinformatics, the reliability of disorder prediction benefits from the use of several methods based on different concepts, different physicochemical parameters, or different implementations. Using a single disorder predictor is not sufficient to achieve predictions good enough to decipher the modular organization of a protein (for examples, see refs. 12–16). Herein, we briefly review the sequence features of disordered proteins. Disor-

*Correspondence to: Sonia Longhi, Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Universités Aix-Marseille I et II, Case 932, 163 Avenue de Luminy, 13288 Marseille Cedex 09, France. E-mail: longhi@afmb.cnrs-mrs.fr

Received 8 April 2005; Revised 22 December 2005; Accepted 4 April 2006

Published online 19 July 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21075

TABLE I. Main Features of Software Tools for Disorder Prediction

Predictor	What is predicted	Based on	Generates and uses multiple sequence alignment?
PONDR[23,68,69] http://www.pondr.com	All regions that are not rigid including random coils, partially unstructured regions, and molten globules	Local aa composition, flexibility, hydropathy, etc	No
SEG[24] http://mendel.imp.univie.ac.at/METHODS/seg.server.html http://www.ncbi.nlm.nih.gov/BLAST (simplified version with default settings) ftp://ftp.ncbi.nih.gov/pub/seg/seg (to download program)	Low-complexity segments that is, "simple sequences" or "compositionally biased regions"	Locally optimized low-complexity segments are produced at defined levels of stringency and then refined according to the equations of Wootton and Federhen[24]	No
Disopred2[11] http://bioinf.cs.ucl.ac.uk/disopred	Regions devoid of ordered regular secondary structure	Cascaded support vector machine classifiers trained on PSI-BLAST profiles	Yes
Globplot[18] http://globplot.embl.de	Regions with high propensity for globularity on the Russell/Linding scale (see next)	Russell/Linding scale of disorder (propensities for secondary structures and random coils)	No
Disembl[70] http://dis.embl.de	LOOPS (regions devoid of regular secondary structure); HOT LOOPS (highly mobile loops); REMARK465 (regions lacking electron density in crystal structure)	Neural networks trained on X-ray structure data	No
NORSp[71] http://cubic.bioc.columbia.edu/services/NORSp	Regions with No Ordered Regular Secondary Structure (NORS). Most, but not all, are highly flexible	Secondary structure and solvent accessibility	Yes
FoldIndex[72] http://bip.weizmann.ac.il/fldbin/findex	Regions that have a low hydrophobicity and high net charge (either loops or unstructured regions)	Charge/hydropathy analyzed locally using a sliding window	No
Charge/hydropathy method[58] http://www.pondr.com	Fully unstructured domains (random coils)	Global sequence composition (hydrophobicity versus net charge)	No
HCA (Hydrophobic Cluster Analysis)[73] http://smi.snv.jussieu.fr/hca/hca-seq.html	Hydrophobic clusters, which tend to form secondary structure elements	Helical visualization of amino acid sequence	No
PreLink[74] http://genomics.eu.org	Regions that are expected to be unstructured in all conditions, regardless of the presence of a binding partner.	Compositional bias and low hydrophobic cluster content.	No
IUPred[43] http://iupred.enzim.hu	Regions that lack a well-defined 3D structure under native conditions	Energy resulting from interresidue interactions, estimated from local amino acid composition	No
RONN[44] http://www.strubi.ox.ac.uk/RONN	Regions that lack a well-defined 3D structure under native conditions.	Bio-basis function neural network trained on disordered proteins	No

der prediction methods are described in Table I, with tips and caveats concerning each predictor listed in Table II. We present a general scheme for disorder prediction in Figure 1, while Figure 2 illustrates a possible pitfall in disorder prediction. We applied the general scheme described in Figure 1 to an in-depth analysis of two well-characterized proteins, the nucleoprotein of measles virus (Figs. 3 and 4) and the Ubiquitin-like protein domain of hPLIC-2 (Fig. 5), and show how prediction methods need to be combined to achieve accurate disorder predictions.

SEQUENCE FEATURES OF DISORDERED PROTEINS

Sequence Composition

Intrinsically disordered proteins generally have a biased amino acid composition. A consensus of two independent studies, focusing respectively on the amino acids preferred at the surface of globular proteins or on those found less frequently in secondary structures[17,18] established the following empirical rule: G, S, and P are disorder-promoting amino acids, W, F, I, Y, V, and L are order-

TABLE II. Tips and Caveats of Disorder Prediction Methods

	Tips	Caveats
PONDR	PONDR comes in several versions: VL-XT may be used to look for binding sites, indicated by sharp drops in the middle of long disordered regions. VSL1 performs better to identify short regions of disorder. VL3 should be preferred to delineate domains as it gives smoother predictions.	
SEG	The stringency of the search for low-complexity segments is determined by three user-defined parameters: trigger window length [W], trigger complexity [K(1)] and extension complexity [K(2)] Parameters for disorder prediction: for long nonglobular domains, use long window lengths, typically: seg sequence 45 3.4 3.75 for shorter nonglobular domains, typically use: seg sequence 25 3.0 3.3	
Disopred2		Prediction accuracy is lower if there are few homologues
Globplot	Gives easy overview of modular organization of large proteins thanks to user-friendly, built-in SMART, PFAM, and low-complexity predictions. Changes of slope often correspond to domain boundaries	
Disembl	Gives predictions of low complexity regions and of aggregation propensity	Use the Loop predictor only as a filter to remove false disorder predictions of Hot Loops and Remark465, i.e., if Loop predicts that a region is ordered whereas Hot Loops or Remark 465 predict the opposite, Loop should be trusted
NORSp		Beware: some NORS are rigid, whereas some highly mobile regions have predicted secondary structure Prediction accuracy is lower if there are few homologs available
FoldIndex	Highlights some regions that are probably short loops better than a simple hydrophobicity plot	Foldindex does not provide results on the N- and C-termini. Therefore, it is not convenient to use it on small proteins (<100 aa).
Charge/hydropathy method	This method has not been trained on disordered proteins. Therefore, it is expected to recognize types of disordered proteins that are underrepresented in the disordered protein databases.	Requires prior knowledge of modular organization of protein. Applicable only to domains without disulfide bonds and without metal-binding regions.
HCA	Highlights coiled coils and regions with a biased composition Highlights regions with potential for induced folding Highlights very short potential globular domains Allows meaningful comparison with related proteins Allows a better definition of the boundaries of disordered regions	User's interpretation required
Prelink		Prelink generally predicts as ordered unstructured regions that have the potential to be ordered in the presence of a partner (i.e., to undergo induced folding)
IUPred	IUPRED uses a novel algorithm that was not trained on disordered proteins. Therefore, it is expected to recognize types of disordered proteins that are underrepresented in the disordered protein databases.	Applicable only to proteins without disulfide bonds and without metal-binding regions.
RONN		If RONN is being used to search for short regions of disorder it is advisable to inspect the plot for regions close to, but below, the threshold.

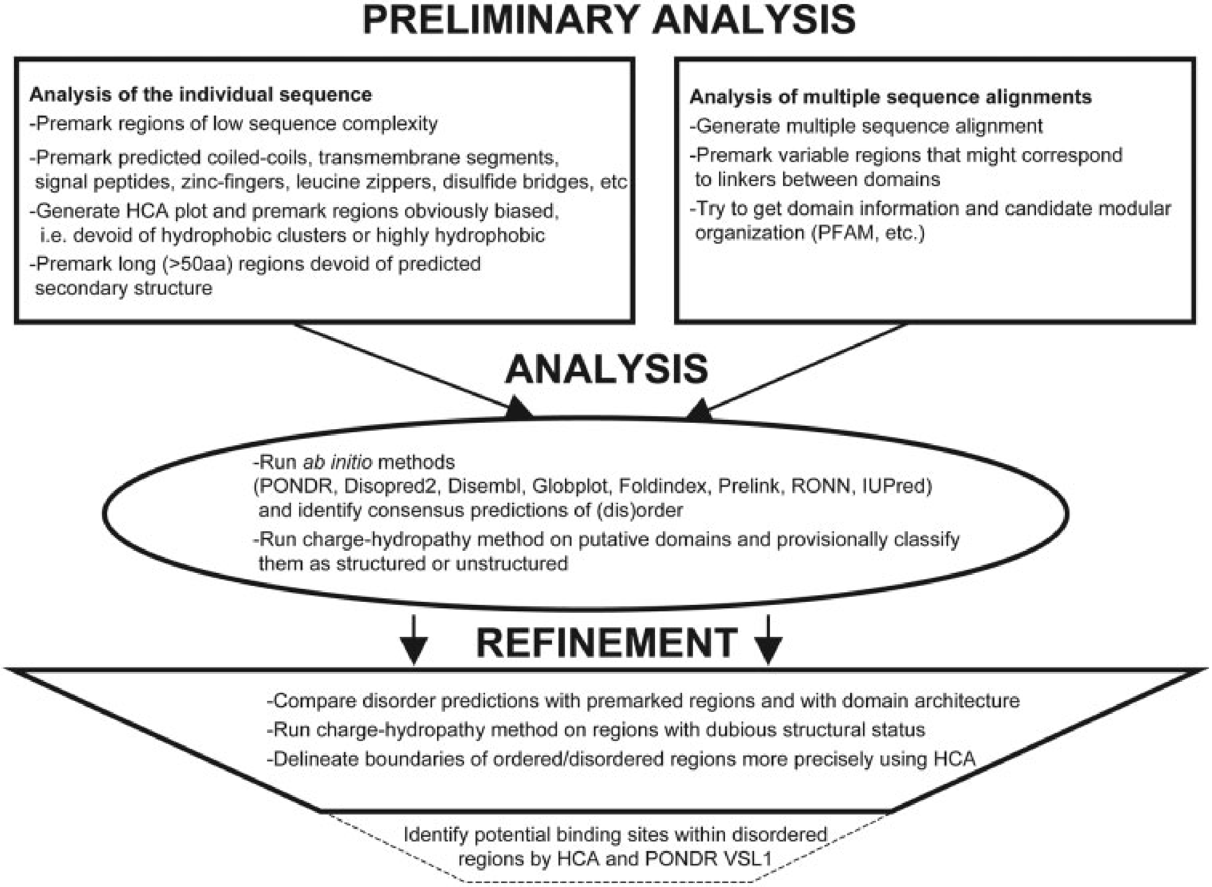

PRELIMINARY ANALYSIS
Analysis of the individual sequence
-Premark regions of low sequence complexity
-Premark predicted coiled-coils, transmembrane segments, signal peptides, zinc-fingers, leucine zippers, disulfide bridges, etc
-Generate HCA plot and premark regions obviously biased, i.e. devoid of hydrophobic clusters or highly hydrophobic
-Premark long (>50aa) regions devoid of predicted secondary structure
Analysis of multiple sequence alignments
-Generate multiple sequence alignment
-Premark variable regions that might correspond to linkers between domains
-Try to get domain information and candidate modular organization (PFAM, etc.)
ANALYSIS
-Run ab initio methods (PONDR, Dispred2, Disembl, Globplot, Foldindex, Prelink, RONN, IUPred) and identify consensus predictions of (dis)order
-Run charge-hydropathy method on putative domains and provisionally classify them as structured or unstructured
REFINEMENT
-Compare disorder predictions with premarked regions and with domain architecture
-Run charge-hydropathy method on regions with dubious structural status
-Delineate boundaries of ordered/disordered regions more precisely using HCA
Identify potential binding sites within disordered regions by HCA and PONDR VSL1

Fig. 1. General scheme for prediction of disordered and ordered regions in a protein.

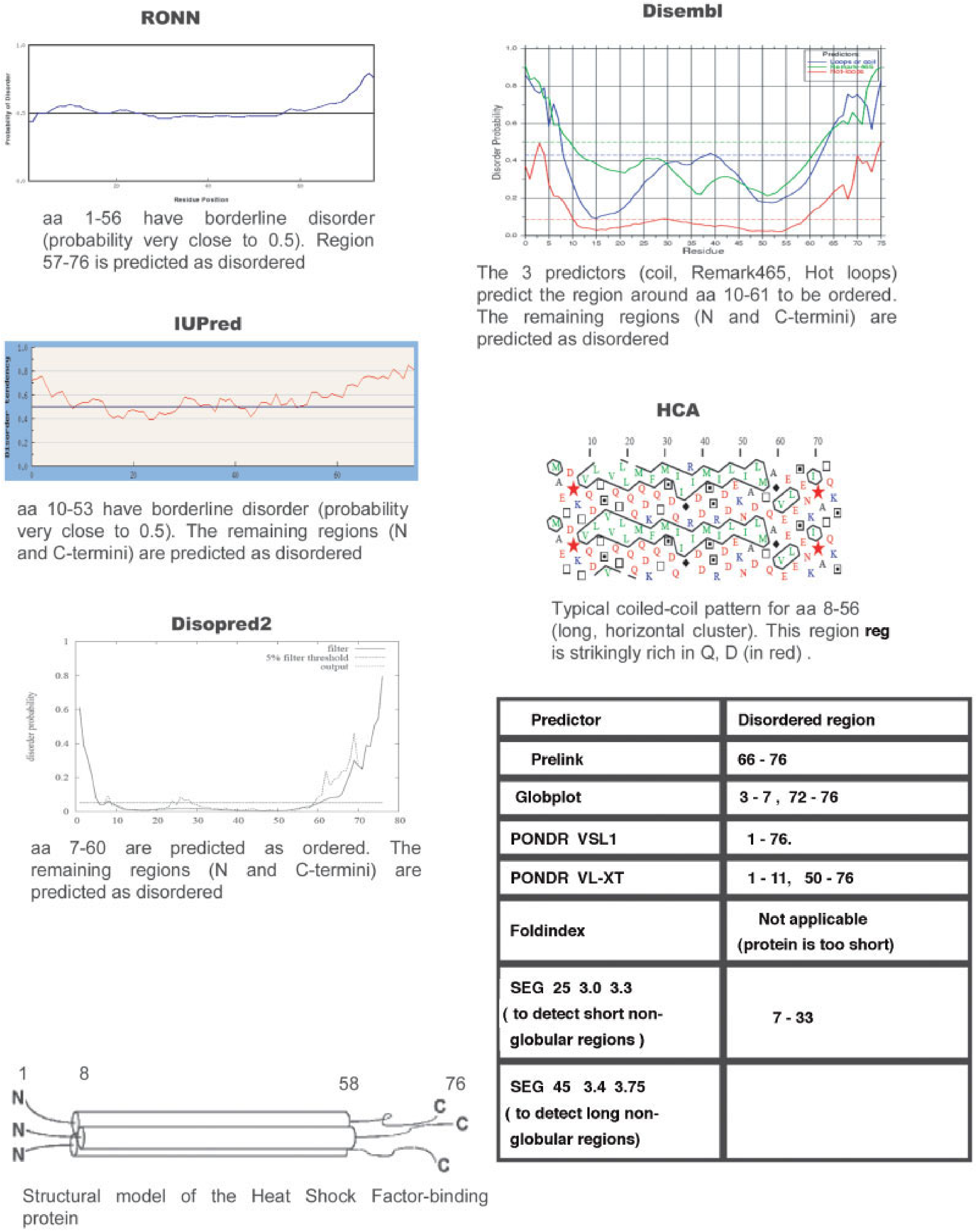
promoting amino acids, while H and T are considered neutral with respect to disorder. Using sequence composition as the sole predictive parameter of disorder is not reliable. For instance, the RNA cap 2′-O-methyltransferase domain of dengue virus polymerase is structured[19] and yet is heavily depleted in some order-promoting residues and markedly enriched in some disorder-promoting residues (data not shown). However, Weathres et al.[20] recently reported that amino acid composition alone could allow recognition of intrinsically disordered proteins with a good accuracy. In any case, it is recommended to always analyze the sequence composition of proteins prior to further sequence analysis.[21]

Low Predicted Secondary Structure Content

Secondary structure prediction is based on the propensity of each amino acid to belong to each type of secondary structure element, computed along sliding windows. Long (>70 aa) regions devoid of predicted secondary structure elements (as judged by using a combination of methods) are generally disordered. There are a few exceptions, called "loopy proteins," which have no regular secondary structure and yet are ordered, like the Kringle domain, a triple-looped, disulphide-linked domain, found in some serine proteases and in some plasma proteins.[22]

Low-Sequence Complexity

Low Complexity Regions (LCRs) are regions with a biased composition (homopolymeric runs, short-period repeats, and more subtle overrepresentation of a few residues), making use of fewer types of amino acids. Intrinsically disordered proteins tend to have a low sequence complexity, although it is not a general rule.[23,24] It has been shown recently that the more low-complexity regions an eukaryotic protein has, the less it is likely to be solubly expressed in bacteria. This might be related to the fact that low-complexity regions, which are more frequent in eukaryotic proteins than in bacterial proteins, are more sensitive to proteolytic degradation.[25] Given the tendency of IDPs to have low complexity, one could expect that they are less soluble than globular proteins. However, this is not a general rule. For instance, the intrinsically disordered N-terminal domain of the measles virus phosphoprotein is even more soluble than the structured C-terminal domain (see Karlin et al.[26] and Longhi et al.[27]). Likewise, the intrinsically disordered C-terminal domain of the measles virus nucleoprotein has a solubility comparable to that of its structured domain (see Longhi et al.[27]). Furthermore, intrinsically disordered proteins are less prone to aggregation compared to globular proteins,[28,29] which facilitates their purification and conservation.



RONN

aa 1-56 have borderline disorder (probability very close to 0.5). Region 57-76 is predicted as disordered

Disembl

The 3 predictors (coil, Remark465, Hot loops) predict the region around aa 10-61 to be ordered. The remaining regions (N and C-termini) are predicted as disordered

IUPred

aa 10-53 have borderline disorder (probability very close to 0.5). The remaining regions (N and C-termini) are predicted as disordered

HCA

Typical coiled-coil pattern for aa 8-56 (long, horizontal cluster). This region reg is strikingly rich in Q, D (in red) .

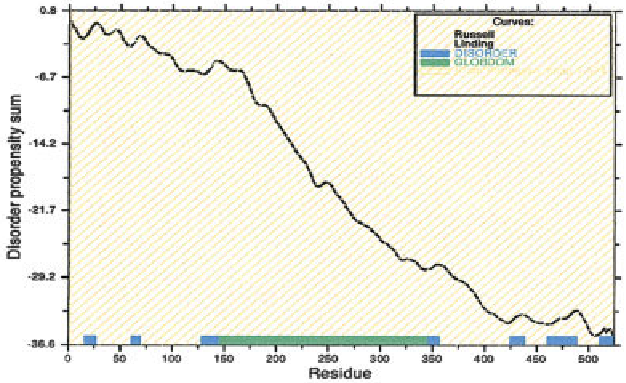
Disopred2

aa 7-60 are predicted as ordered. The remaining regions (N and C-termini) are predicted as disordered

Predictor	Disordered region
Prelink	66 - 76
Globplot	3 - 7 , 72 - 76
PONDR VSL1	1 - 76.
PONDR VL-XT	1 - 11, 50 - 76
Foldindex	Not applicable (protein is too short)
SEG 25 3.0 3.3 (to detect short non-globular regions)	7 - 33
SEG 45 3.4 3.75 (to detect long non-globular regions)	

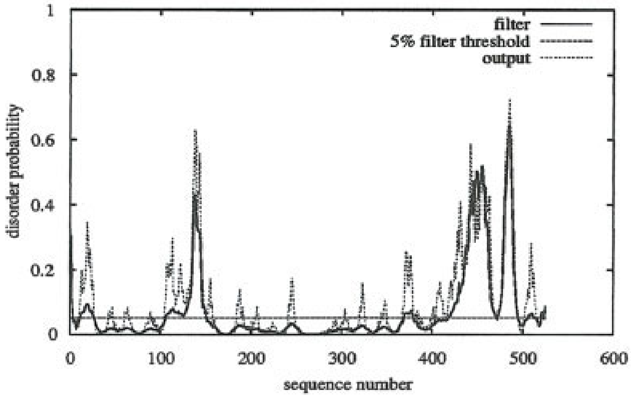
Structural model of the Heat Shock Factor-binding protein

Fig. 2. Analysis of the human heat shock factor-binding protein 1 (Genbank accession number: AF068754) using different predictors. The graphical output of each method and the corresponding interpretation is shown. The precise boundaries of ordered and disordered regions were derived from the corresponding text output (not shown). Bottom, structural model of the protein reproduced from ref. 59 with permission of Richard Morimoto and the American Society for Biochemistry and Molecular Biology. The N- and C-termini (thin lines) are disordered, whereas the central region forms a triple coiled-coil (cylinders). Numbers correspond to the amino acid boundaries of these regions. SEG parameters are explained in Table II.



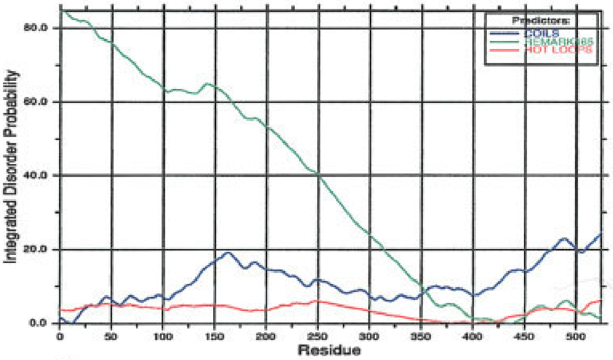
Globplot of MV Nucleoprotein

Conclusions: A globular domain spanning residues 145-344 (▬) is predicted. Other regions are not reliably predicted.



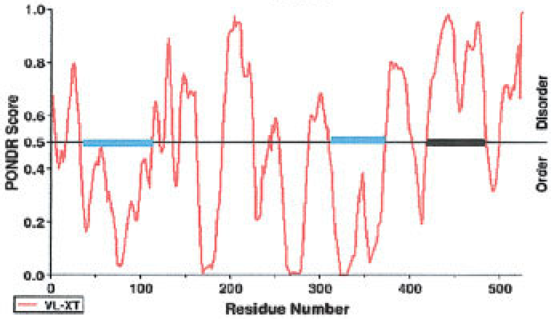
Disopred of MV Nucleoprotein

Conclusions: Disordered regions spanning residues 131-149 and 426-494 are predicted. Other regions are predicted as globular.



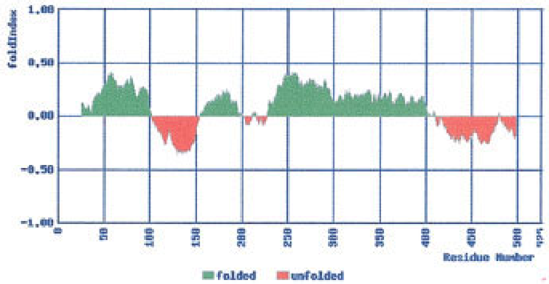
Disembl of MV Nucleoprotein

Conclusions: Disordered regions spanning residues 131-144, 437-466, and 478-490 are predicted. Other regions are predicted as globular. That interpretation is based on the downward slope of remark 465.



PONDR® of MV Nucleoprotein

Conclusions: A single disordered region (aa 419-484) is predicted (thick black line). Long (>40 aa) ordered regions span residues 33-113 and 308-371 (cyan bar).



FoldIndex of Nucleoprotein MV

Conclusions: Disordered regions (▬) spanning residues 100-150 and 420-494 are predicted. Other regions are predicted as globular (▬).

Fig. 3. Measles virus (MV) nucleoprotein (N) (accession number: P35972) analyzed with different predictors. The graphical output of each method and the corresponding interpretation is shown. The precise boundaries of ordered and disordered regions were derived from the corresponding text output (not shown). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

C

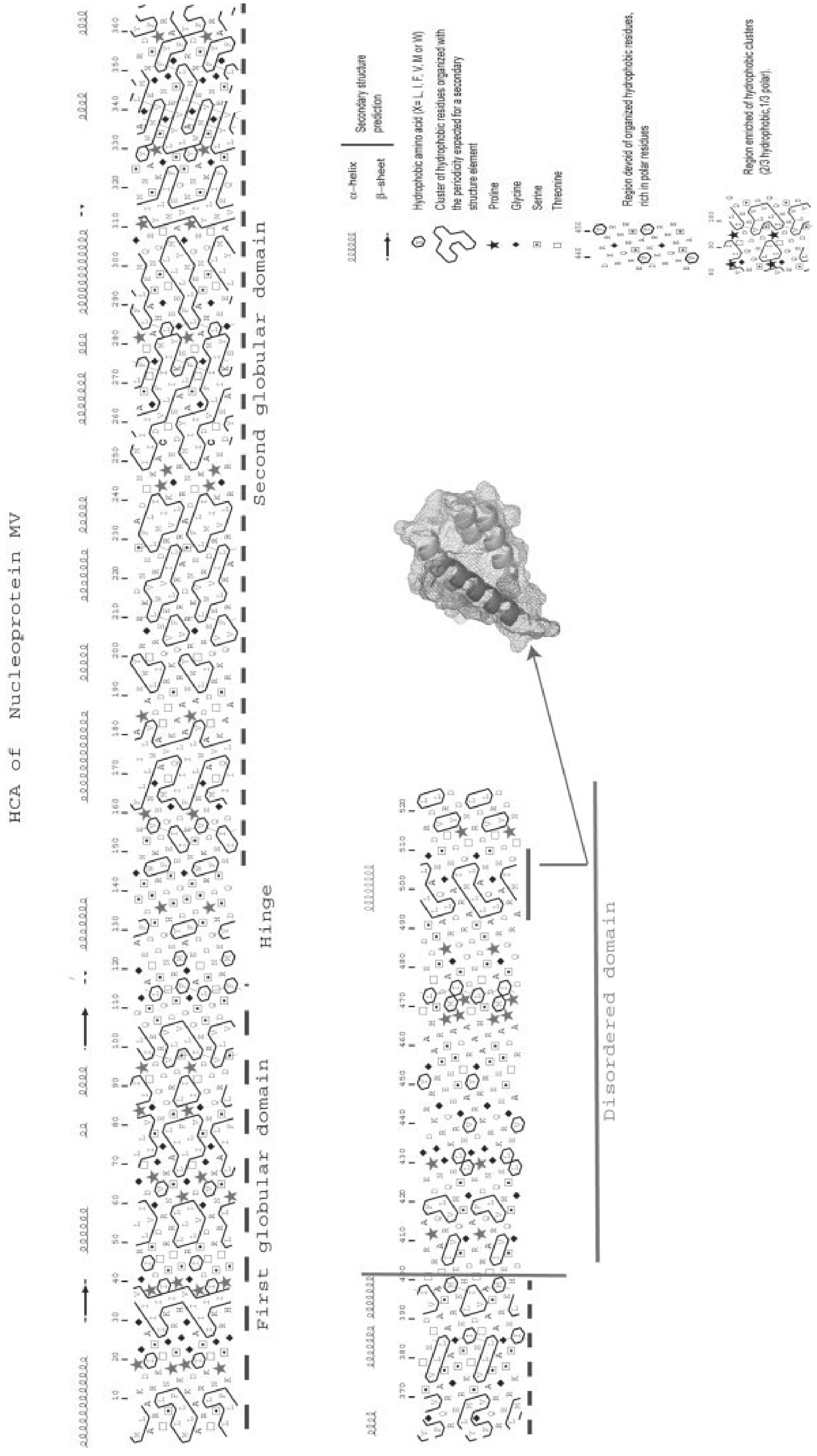
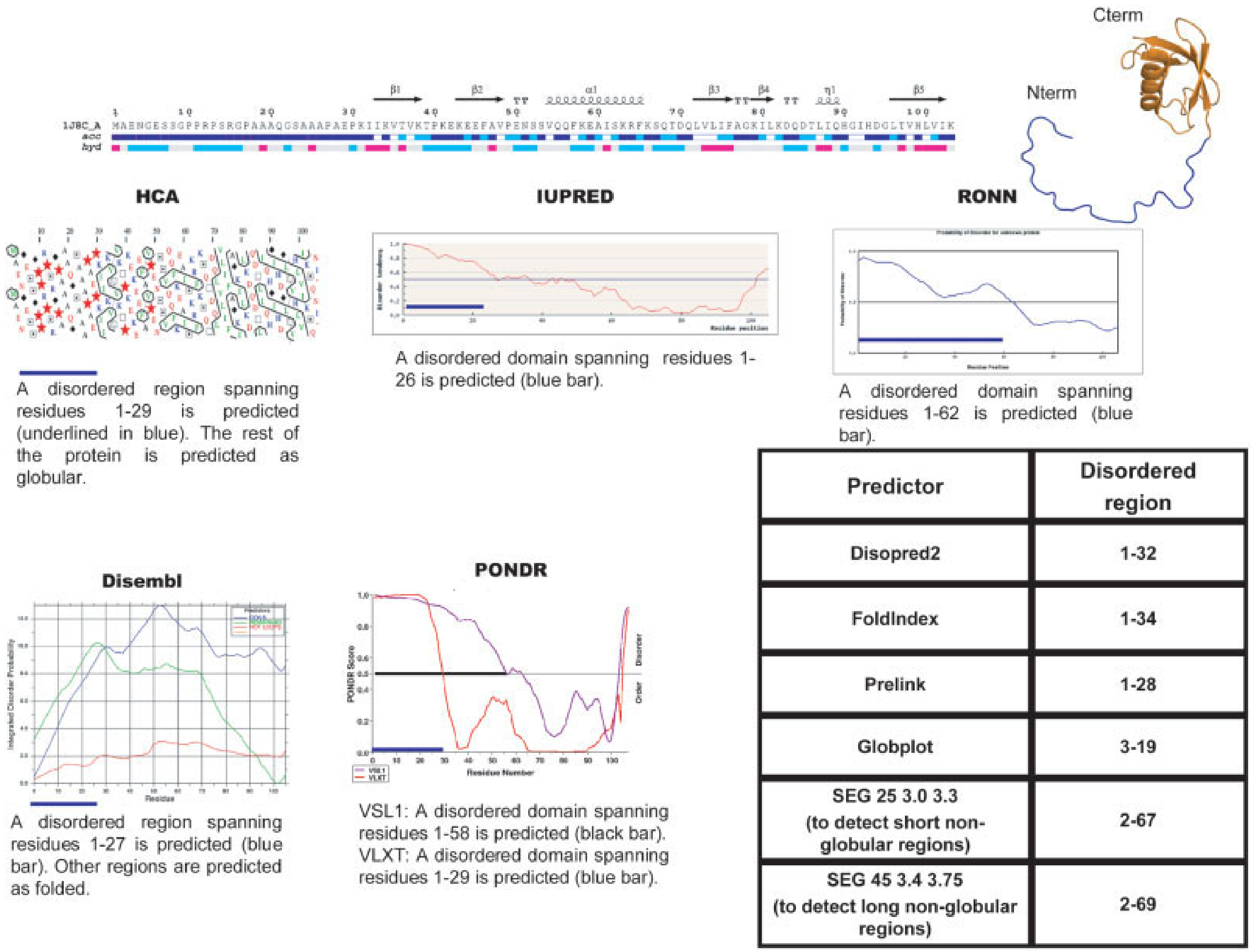

HCA of Nucleoprotein MV
First globular domain
Hinge
Second globular domain
Disordered domain
α–helix
β–sheet
Secondary structure prediction
Hydrophobic amino acid (X= L, I, F, V, M or W)
Cluster of hydrophobic residues organized with the periodicity expected for a secondary structure element
Proline
Glycine
Serine
Threonine
Region devoid of organized hydrophobic residues, rich in polar residues
Region enriched of hydrophobic clusters (2/3 hydrophobic, 1/3 polar).

Fig. 4. HCA plot of measles virus nucleoprotein. Conventions are explicited in the caption. Globular regions (framed) are characterized by a thick distribution of hydrophobic clusters, while unstructured regions are poor or devoid of hydrophobic clusters. Long disordered regions and predicted secondary structure elements are shown. There is no low-complexity region in measles virus N. The induced folding region is underlined, and the corresponding structure (dark gray α-helix) is presented in complex with the C-terminal domain of the measles virus P[75] (PDB code: 1T6O). The picture of the measles virus P was obtained using Pymol.



A disordered region spanning residues 1-29 is predicted (underlined in blue). The rest of the protein is predicted as globular.

A disordered domain spanning residues 1-26 is predicted (blue bar).

A disordered domain spanning residues 1-62 is predicted (blue bar).

A disordered region spanning residues 1-27 is predicted (blue bar). Other regions are predicted as folded.

VSL1: A disordered domain spanning residues 1-58 is predicted (black bar). VLXT: A disordered domain spanning residues 1-29 is predicted (blue bar).

Predictor	Disordered region
Disopred2	1-32
FoldIndex	1-34
Prelink	1-28
Globplot	3-19
SEG 25 3.0 3.3 (to detect short non-globular regions)	2-67
SEG 45 3.4 3.75 (to detect long non-globular regions)	2-69

Fig. 5. Analysis of the Ubiquitin-like protein domain of hPLIC-2 (accession number: Q9UHD9) using different predictors. The protein sequence, with the secondary structure elements derived from the structure (PDB code 1J8C), is shown. The solvent accessibility of each residue is plotted below the sequence. The ribbon representation of the structure is shown; the globular domain is colored in gold while the disordered region is in blue. The picture was obtained using Pymol. The graphical output of various prediction methods and the corresponding interpretation are shown. The precise boundaries of ordered and disordered regions were derived from the corresponding text output (not shown). A blue bar highlights the disordered region for each graphical output, except for VSL1 for which the disordered region is highlighted with a black line. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Some special cases of low-complexity sequences are found in proteins with a certain amino acid periodicity (such as coiled-coils) and other nonglobular, yet ordered proteins (collagen, for example). It is recommended to always look for LCRs, coiled-coils, and repeats in a protein prior to further sequence analysis (using programs such as Paircoil[30] and Multicoil[31]). More subtle parameters to discriminate between globular and nonglobular proteins using the program SEG are discussed in refs. 21 and 24 (see also Table I).

High-Sequence Variability

Disordered regions are on average much more variable than ordered ones.[32] The reason why they evolve faster is not clear at present. The relationship between sequence variability and flexibility is well known by crystallographers: when a protein does not crystallize despite repeated attempts, crystallographers are used to removing hypervariable regions, presumed to be flexible linkers. High-sequence variability is not by itself an evidence of disorder, but only an indicator. A simple method to appreciate sequence variability is visual inspection of a multiple sequence alignment. However, it can sometimes be misleading. Programs that rely on nucleotide substitution rate (as described by Brown et al.[32]) can be very informative and should be used for a more rigorous analysis.[33]

PREDICTORS OF DISORDER

Several programs have been developed to predict disordered regions using the sequence features reviewed above. They are presented with their philosophy in Table I, and their salient points are briefly discussed below. A detailed description of each predictor is outside the scope of this review, and the reader interested in more details on a specific predictor is invited to refer to the relevant article, indicated in Table I.

Predictors of Disorder

Different predictors rely on different physicochemical parameters. Therefore, a given predictor can be more performant in detecting a given feature of a disordered protein. Thus, predictors are complementary, a point illustrated in the section focused on practical examples (see below). As there is no consensus on what disorder means, it is necessary to know precisely what is predicted by each method. For instance, "long disordered regions" predicted by PONDR correspond to regions that are not rigid including random coils, partially unstructured regions, and molten globules (Table I). On the contrary, if a protein is predicted to be unstructured by the charge/hydropathy method, it means that it is probably fully unstructured (random coil) (Table I). This issue has been stressed recently by a systematic comparison between these two prediction methods.[34]

Most predictors rely on training against a dataset of disordered protein regions. These datasets are either entirely built up by the authors or represent improvements of existing datasets. Despite continuous efforts, these datasets retain some inconsistencies and are necessarily biased, because large regions of disorder can prevent crystallization. Furthermore, these datasets contain relatively few disordered proteins. Indeed, Disprot (http://www.disprot.org/),[35] which is the largest publicly available database of disordered proteins whose disorder has been experimentally assessed, contains only about 400 entries. For these reasons, it is useful to distinguish two kinds of predictors: those that have been trained on datasets of disordered proteins (PONDR, Globplot, Disembl, Disopred2, RONN, PreLink), and those that have not, namely the charge/hydropathy method (and its derivative Foldindex), NORSp, and IUPred. The latter avoid the shortcomings and biases associated to the disordered datasets. Therefore, they are expected to perform better than the former methods on disordered proteins presently under-represented in training datasets.

PONDR, a neural network based on local amino acid composition, flexibility, and other sequence features, was the first predictor to be developed (Table I). It is now available in various versions, each having its own specificities (e.g., VL-XT allows highlighting of potential protein-binding regions), and Table II suggests which version should be chosen according to the user's goal. Noteworthy, PONDR has found an application in the study of structured proteins. Indeed, there is a strong inverse correlation between the VL-XT score within *ordered* regions and the presence of dehydrons, which are underwrapped backbone hydrogen bonds, recently identified as a major determinant of protein–protein interactions.[36,37]

Globplot uses a new scale called "Russell/Linding," specially developed to express the propensity for a given amino acid to be in "random coil" or in "regular secondary structure" (Table I). In Globplot, changes of slope often correspond to domain boundaries (Table II).

Disembl consists of three separate predictors, trained on separate datasets, that respectively comprise residues within "loops/coils" (as defined by DSSP[38]), "hot loops" (loops with high B-factors, i.e., very mobile from X-ray crystal structure), or that are missing from the PDB X-ray structures (called "Remark 465") (Table I). The partitioning of residues into different flexibility groups is very useful depending upon the user's goals (i.e., "hot loop" may be used to correlate certain functional aspects of proteins with mobile loops, while "Remark 465" may be used to detect linkers likely to affect crystallization).

Disopred2 (Table I) is also based on a neural network, but incorporates information from multiple sequence alignments because its inputs are derived from sequence profiles generated by PSI-BLAST.

RONN (Table I) uses a novel approach, a bio-basis function neural network. It relies on the calculation of "distances," as determined by sequence alignment, from well-characterized prototype sequences (ordered, disordered, or a mixture of both). Its key feature is that amino acid side chain properties are not considered at any stage.

Prelink (Table I) relies on amino acid composition and on low hydrophobic cluster content. In this respect, it is a derivative of HCA, a powerful approach that is discussed below. PreLink is the first predictor that statistically proved the ability of HCA to detect linkers, an ability that had long been noticed before but never previously demonstrated.

The charge/hydropathy analysis is based on the elegant reasoning that folding of a protein is governed by a balance between attractive forces (of hydrophobic nature) and repulsive forces (electrostatic, between similarly charged residues), Thus, globular proteins can be distinguished from unstructured ones based on the ratio of their net charge versus their hydropathy (Table I). A drawback of this approach is that it gives only a global indication, not valid if the protein is composed of both ordered and disordered regions (Table II). A derivative of that method, Foldindex, solves this problem by computing the charge/hydropathy ratio along the protein (Table I).

NORSp (Table I) relies on the principle that long regions predicted to be devoid of secondary structure and accessible to the solvent are generally unstructured. However, this is not always true, as in the case of the Kringle domain mentioned above.

IUPred uses a novel algorithm that evaluates the energy resulting from interresidues interactions. Although it was derived from the analysis of the sequences of globular proteins only, it allows the recognition of disordered proteins based on their lower interaction energy. This provides a new way to look at the lack of a well-defined structure, which can be viewed as a consequence of a significantly lower capacity to form favorable contacts, correlating results of another study.[39]

The program SEG (Table I), which computes sequence complexity, has not been developed to detect disordered regions but has been used successfully in that aim by the group of Koonin.[21] Typical SEG parameters for disorder prediction are found in Table II.

Table I also includes a nonautomated method that is very useful for unveiling unstructured regions: Hydrophobic Cluster Analysis (HCA).[40] HCA makes use of a twodimensional helical representation of protein sequences in which hydrophobic clusters are plotted along the sequence (the reader is invited to refer to the excellent review by Callebaut et al.[40]). HCA stands aside from other predictors, because they only give insights on the extent of disorder/order, but do not correlate this information with the sequence by itself. Furthermore, there is little one can actually *learn* from comparing the output of these predictors for homologous proteins. In contrast, HCA provides a representation of the short-range environment of each amino acid, thus giving information not only on order/disorder but also on the folding potential (see paragraph on induced folding below). Although HCA does not provide a quantitative prediction of disorder and rather presently requires human interpretation, it provides additional, qualitative information compared to automated predictors, a point illustrated in the section focused on practical examples (see below).

Error Rate of Predictors

A general error rate is difficult to evaluate, because it depends of the definition of disorder used, on the evaluation set, and on the criteria of evaluation. These points are well illustrated by the evaluation of disorder predictors within the recent Critical Assessment of Protein Structure Prediction, CASP6, where very different rankings were obtained as a function of the criteria used.[41] Moreover, the accuracy of a given predictor can be limited when predicting a type of disorder different from that against which it was trained.

Another reason that prevents the meaningful calculation of a precise error rate is the fact that a protein can be disordered by itself, and yet adopt a structure either in a cellular context (when binding to a partner, a phenomenon called "induced folding") or because of artefacts [crystal contacts during crystallization, for instance, or structure solved in a nonaqueous medium such as trifluoroethanol (TFE)]. Many prediction "errors" fall, in fact, in these categories, as discussed in two recent articles by the groups of Poupon and of Dunker (see references therein cited, and Figs. 4–5 in reference[42] and Figure 6 in ref. 34).

Because of all the reasons stated above, error rates are not given in Table I. In general, predictors are more reliable in predicting order than in predicting disorder, as (1) ordered sequences comprise only a very narrow portion of sequence space, that is, their sequence properties are much more recognizable; and (2) because of the limited number of disordered protein sequences available for predictor training. From the authors' personal communications, a conservative accuracy for ab initio methods, such as Disopred2, Disembl, and PONDR, is around 60–70% for predicting disorder, and about 80% for predicting order. Reportedly, the charge/hydropathy method has the best overall accuracy (83%[39]). However, this method requires prior knowledge of the domain boundaries (see Table I). Despite the inherent difficulty of estimating meaningful error rates, recent studies have pointed out that disorder predictors can been grossly classified into three categories.[43,44] These categories are by no means absolute, and are presented only for convenience, to allow a better interpretation of the results provided by the predictors and to optimize their use in combination.

Some predictors perform better on short disordered regions in the context of globally ordered proteins: Disopred2, Prelink, and Disembl (Remark465). Actually, they were specifically developed with that aim. These predictors also have a good specificity (i.e., they predict relatively few ordered residues to be disordered), but a moderate sensitivity (i.e., they miss a significant number of disordered residues). IUPred performs comparatively well for predicting long disordered segments, and has a good sensitivity. Finally, although no method has both a very high specificity and a very high sensitivity, some predictors are "polyvalent" (RONN, PONDR VSL1, Disembl "hot loops," FoldIndex and Globplot).

However, we would like to point out once again that these observations are only aimed at guiding the user, and in no way they are intended to provide a quantitative comparison. The section focused on practical examples (see below) will also give a qualitative idea of the respective sensitivities and specificities of these methods.

PREDICTING INDUCED FOLDING

The analysis of hydrophobic clusters and of secondary structures is of major interest for studying induced folding, because burial of hydrophobic residues provides the major driving force in protein folding. This force is, in turn, regulated by secondary structures that play a role in guiding the folding pathway. In some cases, hydrophobic clusters are found within secondary structure elements that are unstable in the native protein, but can stably fold upon binding a partner. Therefore, HCA can be very informative in highlighting potential induced folding. As an example, we suspected that the isolated hydrophobic cluster with a predicted α-helix within the disordered N_{TAIL} domain (Fig. 4) could correspond to a binding region for one of the partners of *N*, the viral phosphoprotein *P*, and that *N* would undergo induced folding upon binding *P*. Later, this hypothesis was proven experimentally[27,45–47] (see region highlighted by a bar in Fig. 4).

Molecular Recognition Elements (MoREs) are regions within an intrinsically disordered protein that have a propensity to bind to a partner and thereby to undergo induced folding. It has long been noticed that PONDR can highlight potential MoREs[48] (Table II). For instance, a fine analysis of PONDR plots led to the identification of segments of increased structural propensity (i.e., prone to induced folding) in the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E.[49] The group of Dunker recently developed a program to identify α-helix-forming MoREs (called α-MoREs) from the amino acid sequence.[50] For instance, the above-mentioned region within N_{TAIL} that undergoes an α-helical transition upon binding to *P*, was successfully identified. Linding also showed how neural network predictions of disorder can indicate the propensity of ELMs to undergo induced folding.[6]

PRACTICAL EXAMPLES SHOWING HOW COMBINING DIFFERENT METHODS IMPROVES DISORDER PREDICTION

Figure 1 illustrates a general sequence analysis scheme that integrates the peculiarities of each method to predict globular and disordered regions. As a first step, one should perform an analysis of sequence composition[51] and complexity,[24] a search for signal peptides, transmembrane regions,[52] leucine zippers,[53] and coiled-coil regions,[54–56] to premark regions of biased composition. This step is crucial in that it can avoid pitfalls that can lead to miss-predictions, as exemplified in the next section.

It is also recommended to use DIpro[57] to identify possible disulfide bridges and to search for possible metal-binding regions by looking for conserved Cys_3–His or Cys_2–His_2 motifs in multiple sequence alignments. Indeed, the presence of conserved cysteines and/or of metal-binding motifs prevents meaningful local predictions of disorder within these regions, as they may display features typifying disorder while gaining structure upon disulfide formation or upon binding to metal ions.[58]

Then, ab initio methods, such as Globplot, Disembl, PONDR, Disopred2, IUPred, RONN, Prelink, Foldindex, and NORSp (Table I) can be combined to define a consensus on both globular and unstructured regions. Of course, any supplemental information, as for instance sequence similarity of a protein region to multidomain proteins, are precious in terms of domain boundary definition. Once a gross domain architecture for the protein of interest is established, the case of domains whose structural state is uncertain can be settled using the charge/hydropathy method, which has a quite low error rate (see above).

An Example of a Pitfall: A Coiled-Coil

To illustrate a possible pitfall in disorder prediction, we have chosen the Heat-Shock Factor binding Protein 1. Biophysical and biochemical analyses have shown that it consists of a long, trimeric coiled-coil, with the N- and C-termini (respectively aa 1–8 and 58–76) being disordered.[59] Foldindex was not used, as it could not give reliable predictions due the small size of the protein (Table II). PONDR VSL1 predicts the whole protein as disordered (Fig. 2). IUPred and RONN predict borderline disorder for most of the protein (Fig. 2) and the C-terminus as disordered. SEG does not detect any long, nonglobular region (Fig. 2) (using the parameters shown in Table II). However when using more sensitive parameters (Table II), it detects a medium-length region of biased composition (aa 7–33). All other automated predictors predict the central region as ordered and the N- and C-termini as disordered (Fig. 2). A preliminary analysis using Multicoil[31] and HCA would have solved these discrepancies. Multicoil gives a high probability of coiled-coil over aa 30–60 (not shown), while the HCA plot is typical of a coiled-coil (a long and horizontally extended hydrophobic cluster encompassing aa 8–56) (Fig. 2). Furthermore, it is quite obvious from the plot that the protein has a biased composition, being rich in Q and D residues (noticeable thanks to their red color; see Fig. 2). In particular, the Q-rich region roughly corresponds to the low-complexity region detected by SEG (aa 7–33).

Thus, performing the preliminary analysis shown in Figure 1 would have allowed detection of a coiled-coil (which fooled some predictors into giving a wrong prediction of borderline disorder) and would have overcome this pitfall, while giving precious information on the protein (biased composition). Once the structural status of the region 8–60 has been established as a coiled-coil, the comparative analysis of the ensemble of the results gives a more accurate prediction. Indeed, almost all predictors correctly predict disordered N- and C-termini with reasonably accurate boundaries (Fig. 2). This example also illustrates the advantage of using predictors that rely on different principles: for instance, because Prelink is based on HCA, it is expected to correctly predicted coiled-coils as ordered. As another example, PONDR VL-XT, gives a correct prediction, whereas another version of PONDR, VSL1, optimized to detect short disordered regions (Table II), is completely fooled by the coiled-coil, and predicts it as disordered.

Domain Identification

Figures 3 and 4 illustrate the approach used to study the domain organization of the nucleoprotein (N) of measles virus, a protein that encapsidates the viral RNA. Experimental data available indicate that N is organized into two regions, N_{CORE} (aa 1–399) and N_{TAIL} (aa 400–525), respectively ordered[60] and disordered.[46,61] As shown in Figure 3, most ab initio methods converge to show the presence of a disordered region at its C-terminus (consensus is aa 437–484), and of a globular core (aa 145–344). Interestingly, Foldindex (Table I) highlights a very hydrophilic region (aa 100–150) that is also visible as a short plateau (aa 131–144) in the output of Disembl Remark 465 predictor and that is predicted by Disopred2 too (aa 131–149). Moreover, this region is hypervariable in sequence among *Morbillivirus* members (not shown). Finally, because changes in slope of Globplot often correspond to domain boundaries (see Table I), from this analysis one would suspect the following domain organization: a first domain or subdomain encompassing residues 1–130, that is not confidently predicted but might be ordered (cf. the negative slope of Globplot together with PONDR prediction); an exposed loop spanning aa 131–149; a second, more compact domain (aa 150–400, cf. steep negative slope), and a disordered domain encompassing aa 401–525. Finally, the charge/hydropathy method predicts that both suspected subdomains are ordered and confirms that the C-terminal domain is disordered (not shown).

HCA helps to refine these predictions. As shown in Figure 4, the density of hydrophobic clusters indicates without ambiguity that both subdomains identified by the combination of previous methods are ordered, and the lack of hydrophobic clusters within the 422–525 region indicates that it cannot be ordered by itself (the hydrophobic clusters in the 494–525 region are not long enough to lead to the formation of a compact domain).

Thus, no single method, nor even a combination of two predictors, could successfully unveil the organization of measles virus N, whereas the combined use of all predictors proved to be much more powerful in terms of domain boundary recognition. The hypervariable region (aa 131–149) is indeed accessible to antibodies and thus exposed to the solvent.[62] However, a wealth of mutational data (see Karlin et al.[60] and references therein) indicates that N_{CORE} cannot be divided into independent modules, but rather that the subdomains indicated above (aa 1–130 and aa 145–400) probably fold cooperatively. Thus, the exposed region (aa 131–149) is probably a loop and not a linker that would connect two mobile domains. Whether it is disordered or not is not known. However, as it is not sensitive to proteolysis[60] it is probably at least partially ordered. These unsolved issues nicely illustrate the present limits of disorder prediction.

Manual Refinement of Domain Boundaries using HCA

Figure 5 illustrates a frequently encountered case in disorder prediction, namely the occurrence of an extended region of intermediate length (20 > aa < 40) at one extremity of the protein followed by a globular domain (~70 aa). We have used the example of the Ubiquitin-like domain of hPLIC-2, whose structure has been solved by NMR.[63] As shown in Figure 5 (top), the region encompassing residues 1–31 is devoid of regular secondary structure elements and is extended in solution. All predictors detect a disordered region at the amino terminus (Fig. 5); however, the predicted boundaries of the disordered region vary from one predictor to another, with a predicted length ranging from 19 to 69 residues (see Fig. 5). Globplot predicts disorder for the 1–20 region, four predictors (Prelink, DisEMBL, VL-XT, and IUPred) define the C-terminal boundary of the disordered region around residue 28, two predictors (Disopred and Foldindex) predict a disordered region spanning residue 1–32, two predictors (VSL1 and RONN) extend the prediction of disorder to the region encompassing residues 1–60 (see Fig. 5), and SEG predicts a potential nonglobular region within aa 2–69.

Based on these results, the user can be confident than the N-terminal moiety of the protein is disordered, although the exact C-terminal boundary of the unstructured region remains uncertain (predictions vary from residue 19 to residue 69!) Use of HCA helps to reduce this uncertainty. The HCA plot clearly allows the identification of the 1–30 region as disordered, based on its almost total depletion in hydrophobic residues, and of the 53–103 region as ordered, given its high density in hydrophobic clusters. Thus, one can confidently predict that the protein is organized into two moieties, with HCA having narrowed the boundary between these two regions down to the 30–52 region. In the absence of functional or biochemical clues (such as limited proteolysis studies), the production of various truncated versions of each moiety is recommended in view of functional or structural studies. Such constructs should start or end at incremental positions between residues 30 and 52 (i.e., at the ends of predicted secondary structure elements; not shown). Indeed, the experience that we gained in the context of past and present structural genomics projects developed in our laboratory (see SPINE and VIZIER projects at http://www.afmb.univ-mrs.fr/-The-Spine-Program- and http://www.afmb.univ-mrs.fr/-VIZIER-) has shown that a critical factor in obtaining good-quality protein crystals is the number of constructs generated around the predicted boundary of the domain under study.

CONTRIBUTION OF DISORDER PREDICTIONS TO BIOINFORMATICS ANALYSES

As already mentioned, the identification of disorder can also avoid gross mistakes in protein sequence analysis. For instance, Iyer et al.[4] recently reported two examples in which the SEG program (Table I), in combination with multiple alignment and secondary structure prediction, invalidates previous functional assignments for two proteins, ATF-2 and PIF3, made on the basis of distant sequence similarity to two domains (respectively histone acetyltransferase (HAT) and PAS domain). The HAT and PAS domains are globular, whereas the similar regions of ATF-2 and PIF3 are confidently predicted to be unstructured, casting a strong doubt on their suspected homology.

Once regions of disorder have been identified, then further bioinformatics analyses, aimed at identifying related proteins, can be carried out avoiding spurious sequence similarity. For instance, Rabitsch et al. illustrated how to perform search for homologs of proteins composed mostly of unstructured regions (Sgo1 and Sgo2). They searched for candidate proteins having short stretches of sequence or structural similarity (i.e., presence of a coiled-coil) to Sgo1 and Sgo2, and distributed in a similar fashion as compared to the candidate protein sequence.[64]

Disorder prediction can also greatly help to identify short modules. For instance, it was instrumental in identifying five novel groups of Lsm domain proteins.[65] The architecture of these proteins, which guided the research for sequence similarities, was elucidated using the consensus of Globplot, Disembl, NORSp, and PONDR analyses. The authors identified conserved motifs described as “stable islands in a large sea of intrinsically unstructured sequence regions.” This is probably true of many large human proteins for which very short conserved motifs in the middle of long disordered regions remain to be discovered.[7] Another method that can be of great use in identifying short (50–70 aa), globular domains located within long, disordered regions (e.g., chromodomains) is HCA.[40]

CONCLUSION

As we have seen from two detailed examples, no single predictor can reveal the structural organization of a protein. However, in combination they provide relatively accurate results. Thus, there is obvious room for improvement of predictors by combining features of several programs (i.e., by including information on predicted secondary structure elements, or information derived from multiple sequence alignments). It would also be of major interest to check whether known regions of induced folding correlate well with isolated hydrophobic clusters, corresponding to predicted α-helices, within disordered regions.[45–47] Other improvements may arise from a better understanding of the different types, or "flavors" of disorder.[66]

As a last, optimistic note, one should never give up carrying out a crystallization experiment because of an order/disorder prediction: recently, Mavrakis et al. submitted the phosphoprotein of rabies virus to crystallization trials, despite the fact that the N-terminal moiety, which accounts for more than half of the protein, was predicted to be disordered. Crystals formed in the drop. In fact the N-terminus had been cleaved off by contaminating proteases and the crystals were made of the C-terminal part . . . whose structure was readily solved![67]

ACKNOWLEDGMENTS

We thank K. Dunker, P. Romero, J. Ward, R. Linding, V. Uversky, J. Wootton, I. Callebaut, A. Poupon, R. Esnouf, and Z. Dosztányi for their useful comments on their respective predictors. We are also grateful to V. Receveur-Bréchot and Sung Haeng Lee for critical reading of the manuscript.

REFERENCES

1. Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–533.
2. Receveur-Bréchot V, Bourhis JM, Uversky VN, Canard B, Longhi S. Assessing protein disorder and induced folding. Proteins Struct Funct Bioinformat 2006;62:24–45.
3. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002;11:739–756.
4. Iyer LM, Aravind L, Bork P, Hofmann K, Mushegian AR, Zhulin IB, Koonin EV. Quoderat demonstrandum? The mystery of experimental validation of apparently erroneous computational analyses of protein sequences. Genome Biol 2001;2:RESEARCH0051.
5. Puntervoll P, Linding R, Gemund C, Chabanis-Davidson S, Mattingsdal M, Cameron S, Martin DM, Ausiello G, Brannetti B, Costantini A, Ferre F, Maselli V, Via A, Cesareni G, Diella F, Superti-Furga G, Wyrwicz L, Ramu C, McGuigan C, Gudavalli R, Letunic I, Bork P, Rychlewski L, Kuster B, Helmer-Citterich M, Hunter WN, Aasland R, Gibson TJ. ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. Nucleic Acids Res 2003;31:3625–3630.
6. Linding R. Linear functional modules. Implication for protein function. PhD Thesis, University of Heidelberg; 2004.
7. Neduva V, Linding R, Su-Angrand I, Stark A, Masi FD, Gibson TJ, Lewis J, Serrano L, Russell RB. Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS Biol 2005;3:e405.
8. Friedberg I, Jaroszewski L, Ye Y, Godzik A. The interplay of fold recognition and experimental structure determination in structural genomics. Curr Opin Struct Biol 2004;14:307–312.
9. Melamud E, Moult J. Evaluation of disorder predictions in CASP5. Proteins 2003;53(Suppl 6):561–565.
10. Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 2005;61:166–182.
11. Dosztanyi Z, Csizmok V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol 2005;347:827–839.
12. Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in Paramyxovirinae N and P. J Gen Virol 2003;84(Pt 12):3239–3252.
13. Ferron F, Rancurel C, Longhi S, Cambillau C, Henrissat B, Canard B. VaZyMolO: a tool to define and classify modularity in viral proteins. J Gen Virol 2005;86(Pt 3):743–749.
14. Severson W, Xu X, Kuhn M, Senutovitch N, Thokala M, Ferron F, Longhi S, Canard B, Jonsson CB. Essential amino acids of the hantaan virus N protein in its interaction with RNA. J Virol 2005;79:10032–10039.
15. Ferron FP. Approches bioinformatiques et structurales des réplicase virales. Marseille: Aix-Marseille II; 2005.
16. Llorente MT, Barreno-Garcia B, Calero M, Camafeita E, Lopez JA, Longhi S, Ferron F, Varela PF, Melero JA. Structural analysis of the human respiratory syncitial virus phosphoprotein: characterization of an a-helical domain involved in oligomerization. J Gen Virol 2006;87:159–169.
17. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. J Mol Graph Model 2001;19:26–59.
18. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 2003;31:3701–3708.
19. Egloff MP, Benarroch D, Selisko B, Romette JL, Canard B. An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. EMBO J 2002;21:2757–2768.
20. Weathers EA, Paulaitis ME, Woolf TB, Hoh JH. Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein. FEBS Lett 2004;576:348–352.
21. Koonin E V, Galperin M. Sequence–evolution–function: computational approaches in comparative genomics. New York: Kluwer Academic Publishers; 2003.
22. Liu J, Tan H, Rost B. Loopy proteins appear conserved in evolution. J Mol Biol 2002;322:53–64.
23. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42:38–48.
24. Wootton JC. Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput Chem 1994;18:269–285.
25. Dyson MR, Shadbolt SP, Vincent KJ, Perera RL, McCafferty J. Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. BMC Biotechnol 2004;4:32.
26. Karlin D, Longhi S, Receveur V, Canard B. The N-terminal domain of the phosphoprotein of morbilliviruses belongs to the natively unfolded class of proteins. Virology 2002;296:251–262.
27. Longhi S, Receveur-Brechot V, Karlin D, Johansson K, Darbon H, Bhella D, Yeo R, Finet S, Canard B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J Biol Chem 2003;278:18638–18648.
28. Linding R, Schymkowitz J, Rousseau F, Diella F, Serrano L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. J Mol Biol 2004;342:345–353.
29. Tartaglia GG, Pellarin R, Cavalli A, Caflisch A. Organism complexity anti-correlates with proteomic beta-aggregation propensity. Protein Sci 2005;14:2735–2740.
30. Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS. Predicting coiled coils by use of pairwise residue correlations. Proc Natl Acad Sci USA 1995;92:8259–8263.
31. Wolf E, Kim PS, Berger B. MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Sci 1997;6:1179–1189.
32. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Keith Dunker A. Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol 2002;55:104–110.
33. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 2002;18:486.
34. Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK. Comparing and combining predictors of mostly disordered proteins. Biochemistry 2005;44:1989–2000.
35. Vucetic S, Obradovic Z, Vacic V, Radivojac P, Peng K, Iakoucheva LM, Cortese MS, Lawson JD, Brown CJ, Sikes JG, Newton CD, Dunker AK. DisProt: a database of protein disorder. Bioinformatics 2005;21:137–140.
36. Fernandez A, Berry RS. Molecular dimension explored in evolution to promote proteomic complexity. Proc Natl Acad Sci USA 2004;101:13460–13465.

37. Fernandez A, Scott R, Berry RS. The nonconserved wrapping of conserved protein folds reveals a trend toward increasing connectivity in proteomic networks. Proc Natl Acad Sci USA 2004;101: 2823–2827.
38. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
39. Garbuzynskiy SO, Lobanov MY, Galzitskaya OV. To be folded or to be unfolded? Protein Sci 2004;13:2871–2877.
40. Callebaut I, Courvalin JC, Worman HJ, Mornon JP. Hydrophobic cluster analysis reveals a third chromodomain in the Tetrahymena Pdd1p protein of the chromo superfamily. Biochem Biophys Res Commun 1997;235:103–107.
41. Jin Y, Dunbrack RL Jr. Assessment of disorder predictions in CASP6. Proteins 2005;61:167–175.
42. Coeytaux K, Poupon A. Prediction of unfolded segments in a protein sequence based on amino acid composition. Bioinformatics 2005;21:1891–1900.
43. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 2005;21:3433–3434.
44. Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 2005;21: 3369–3376.
45. Johansson K, Bourhis JM, Campanacci V, Cambillau C, Canard B, Longhi S. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. J Biol Chem 2003;278:44567–44573.
46. Bourhis JM, Johansson K, Receveur-Brechot V, Oldfield CJ, Dunker KA, Canard B, Longhi S. The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. Virus Res 2004;99:157–167.
47. Kingston RL, Baase WA, Gay LS. Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins. J Virol 2004;78:8630–8640.
48. Garner E, Romero P, Dunker AK, Brown C, Obradovic Z. Predicting binding regions within disordered proteins. Genome Inform Ser Workshop Genome Inform 1999;10:41–50.
49. Callaghan AJ, Aurikko JP, Ilag LL, Gunter Grossmann J, Chandran V, Kuhnel K, Poljak L, Carpousis AJ, Robinson CV, Symmons MF, Luisi BF. Studies of the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E. J Mol Biol 2004;340:965–979.
50. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 2005;44: 12454–12470.
51. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 1999;112:531–552.
52. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 2003;31:3784–3788.
53. Bornberg-Bauer E, Rivals E, Vingron M. Computational approaches to identify leucine zippers. Nucleic Acids Res 1998;26: 2740–2746.
54. Lupas A. Prediction and analysis of coiled-coil structures. Methods Enzymol 1996;266:513–525.
55. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science 1991;252:1162–1164.
56. Lupas A. Predicting coiled-coil regions in proteins. Curr Opin Struct Biol 1997;7:388–393.
57. Baldi P, Cheng J, Vullo A. Large-scale prediction of disulphide bond connectivity. Adv Neural Inf Process Syst 2004;17:97–104.
58. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.
59. Tai LJ, McFall SM, Huang K, Demeler B, Fox SG, Brubaker K, Radhakrishnan I, Morimoto RI. Structure–function analysis of the heat shock factor-binding protein reveals a protein composed solely of a highly conserved and dynamic coiled-coil trimerization domain. J Biol Chem 2002;277:735–745.
60. Karlin D, Longhi S, Canard B. Substitution of two residues in the measles virus nucleoprotein results in an impaired self-association. Virology 2002;302:420–432.
61. Longhi S, Receveur-Brechot V, Karlin D, Johansson K, Darbon H, Bhella D, Yeo R, Finet S, Canard B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J Biol Chem 2003;278:18638–18648.
62. Giraudon P, Jacquier MF, Wild TF. Antigenic analysis of African measles virus field isolates: identification and localisation of one conserved and two variable epitope sites on the NP protein. Virus Res 1988;10:137–152.
63. Walters KJ, Kleijnen MF, Goh AM, Wagner G, Howley PM. Structural studies of the interaction between ubiquitin family proteins and proteasome subunit S5a. Biochemistry 2002;41:1767–1777.
64. Rabitsch KP, Gregan J, Schleiffer A, Javerzat JP, Eisenhaber F, Nasmyth K. Two fission yeast homologs of *Drosophila* Mei-S332 are required for chromosome segregation during meiosis I and II. Curr Biol 2004;14:287–301.
65. Albrecht M, Lengauer T. Novel Sm-like proteins with long C-terminal tails and associated methyltransferases. FEBS Lett 2004;569:18–26.
66. Vucetic S, Brown C, Dunker K, Obradovic Z. Flavors of protein disorder. Proteins 2003;52:573–584.
67. Mavrakis M, McCarthy AA, Roche S, Blondel D, Ruigrok RW. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. J Mol Biol 2004;343:819–831.
68. Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Dunker AK. Identifying disordered regions in proteins from amino acid sequences. Proceedings of the IEEE International Conference on Neural Networks. 1997. p 90–95.
69. Li X, Romero P, Rani M. Dunker AK, Obradovic AZ. Predicting protein disorder for N-, C- and internal regions. Genome Informatics 1999;10:30–40.
70. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. Structure (Camb) 2003;11:1453–1459.
71. Liu J, Rost B. NORSp: predictions of long regions without regular secondary structure. Nucleic Acids Res 2003;31:3833–3835.
72. Zeev-Ben-Mordehai T, Rydberg EH, Solomon A, Toker L, Auld VJ, Silman I, Botti S, Sussman JL. The intracellular domain of the *Drosophila* cholinesterase-like neural adhesion protein, gliotactin, is natively unfolded. Proteins 2003;53:758–767.
73. Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon JP. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci 1997;53:621–645.
74. Coeytaux K, Poupon A. Prediction of unfolded segments in a protein sequence based on amino acid composition. Bioinformatics 2005;21:1891–1900.
75. Kingston RL, Hamel DJ, Gay LS, Dahlquist FW, Matthews BW. Structural basis for the attachment of a paramyxoviral polymerase to its template. Proc Natl Acad Sci USA 2004;101:8301–8306.

Available online at www.sciencedirect.com

ScienceDirect

Journal of Structural Biology 155 (2006) 230–238

Journal of Structural Biology

www.elsevier.com/locate/yjsbi

Crystal structure of the actin-binding domain of α-actinin 1: Evaluating two competing actin-binding models ☆

Emma Borrego-Diaz, Frederic Kerff, Sung Haeng Lee, François Ferron, Yu Li, Roberto Dominguez *

Boston Biomedical Research Institute, 64 Grove Street, Watertown, MA 02472, USA

Received 4 December 2005; accepted 29 January 2006
Available online 25 April 2006

Abstract

α-Actinin belongs to the spectrin family of actin crosslinking and bundling proteins that function as key regulators of cell motility, morphology and adhesion. The actin-binding domain (ABD) of these proteins consists of two consecutive calponin homology (CH) domains. Electron microscopy studies on ABDs appear to support two competing actin-binding models, extended and compact, whereas the crystal structures typically display a compact conformation. We have determined the 1.7 Å resolution structure of the ABD of α-actinin 1, a ubiquitously expressed isoform. The structure displays the classical compact conformation. We evaluated the two binding models by surface conservation analysis. The results show a conserved surface that spans both domains and corresponds to two previously identified actin-binding sites (ABS2 and ABS3). A third, and probably less important site, ABS1, is mostly buried in the compact conformation. However, a thorough examination of existing structures suggests a weak and semi-polar binding interface between the two CHs, leaving open the possibility of domain reorientation or opening. Our results are consistent with a two-step binding mechanism in which the ABD interacts first in the compact form observed in the structures, and then transitions toward a higher affinity state, possibly through minor rearrangement of the domains.

Keywords: Spectrin family; Surface conservation analysis; X-ray crystallography; Protein–protein interaction

1. Introduction

The dynamic assembly/disassembly of the actin cytoskeleton is regulated by a large number of actin-binding proteins (ABPs). These proteins interact with monomeric (G-) and filamentous (F-) actin through structurally conserved motifs. One of the most abundant actin-binding motifs is the calponin homology (CH) domain (~110 residues), which can occur in single or multiple copies (Broderick and Winder, 2005; Gimona et al., 2002). CH domains are present in all members of the spectrin family such as β-spectrin, dystrophin, utrophin, and α-actinin, which play key roles in the regulation of cell motility, morphology and adhesion. The actin-binding domain (ABD) of these proteins consists of two N-terminal CH domains, CH1 and CH2, connected by a flexible linker. Although CH1 accounts for most of the F-actin-binding affinity, CH2 (which does not bind actin by itself) contributes somewhat to the overall binding affinity of the ABD (Way et al., 1992).

In the cell α-actinin crosslinks actin filaments and serves as a connection between the cytoskeleton and signaling and membrane proteins (Broderick and Winder, 2005; Otey and Carpen, 2004). α-Actinins are a family of four closely related isoforms. Isoforms 2 and 3 are expressed exclusively in muscle, where they crosslink actin filaments in the Z disc (Beggs et al., 1992). Of the two non-muscle isoforms, α-actinin 1 is ubiquitously expressed and located primarily in focal adhesions (Pavalko and Burridge, 1991), whereas isoform 4 is present in membrane ruffles (Honda et al., 1998). α-Actinin functions as

☆ PDB IDS: 1EYI, 1EYN.

* Corresponding author. Fax: +1 617 972 1753.
E-mail address: rdominguez@bbri.org (R. Dominguez).

1047-8477/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2006.01.013

an antiparallel homodimer. Each monomer consists of an N-terminal ABD, followed by four spectrin repeats, and a C-terminal calmodulin-like (CaM) domain. The function of α-actinin appears to be regulated by the binding of phosphatidylinositol (4,5)-bisphosphate (PIP_2) to the ABD (Fukami et al., 1992, 1996) and the binding of Ca^{2+} to the CaM domain in non-muscle isoforms.

The crystal structures of several ABDs have been determined, including those of human, *A. thaliana* and *S. pombe* fimbrin (Goldsmith et al., 1997; Klein et al., 2004), utrophin (Keep et al., 1999b), dystrophin (Norwood et al., 2000), human and mouse plectin (Garcia-Alvarez et al., 2003; Sevcik et al., 2004) and α-actinin 3 (Franzot et al., 2005). With the exception of the first ABD of *A. thaliana* fimbrin (Klein et al., 2004), which contains a unique mutation at the CH1–CH2 interface (discussed below), all the ABD structures display a similar compact conformation, characterized by extensive interactions between the two CH domains. In two of the structures, those of utrophin (Keep et al., 1999b) and dystrophin, the compact conformation results from domain swapping between two molecules in the asymmetric unit, a relatively common artefact of X-ray crystallography (Liu and Eisenberg, 2002). However, in solution the ABD of utrophin is monomeric (Winder, 1996) and that of dystrophin exists in a monomer–dimer equilibrium (Norwood et al., 2000). Moreover, both proteins appear to function as monomers in the cell (Broderick and Winder, 2005). There is therefore little doubt that most ABDs assume a compact conformation in the unbound state, similar to that observed in the structures.

By contrast, the conformation of actin-bound ABDs has yet to be clarified (Galkin et al., 2003; Lehman et al., 2004). EM studies of ABD-decorated actin filaments have suggested two different modes of binding, compact (Hanein et al., 1998; McGough et al., 1994; Sutherland-Smith et al., 2003) and extended (Galkin et al., 2002; Moores et al., 2000). The compact model holds that the ABD binds actin with only minor changes relative to the crystal structures, whereas the extended model maintains that the two CHs become fully separated upon binding. The latter model implies that CH1 and CH2 interact semi-independently with actin. These results appear to suggest that different ABDs, while sharing a common fold, bind actin in completely different conformations. Furthermore, based on studies on utrophin (Galkin et al., 2002) and plectin (Cherepanova et al., 2005), it has been proposed that a single ABD can bind actin in multiple conformations (polymorphic binding). The above scenarios are uncommon, and possibly unlikely, since proteins of the same family tend to share a similar mode of interaction.

We have determined the crystal structure of the ABD of α-actinin 1, a ubiquitously expressed isoform. The structure displays a compact conformation similar to that observed in most ABDs. To gain further insights into the nature of the ABD-actin interaction, we have mapped a conserved surface area on the compact structure, which corresponds approximately to two previously identified actin-binding sites (ABS2 and ABS3). This analysis was then extended to the separated CH domains, so that the merits of the two binding models could be evaluated. An examination of the existing structures reveals a moderately conserved and semi-polar CH1–CH2 interface. This interface is unlikely to become fully exposed but could accommodate limited structural changes. The results are consistent with a two-step binding mechanism in which the ABD interacts first in the compact form observed in most of the structures, and then transitions toward a higher affinity state, possibly through minor rearrangement of the two domains.

2. Experimental

2.1. Cloning, expression, and purification of ABD

The cDNA encoding the ABD of human α-actinin 1 (GenBank ID BC003576) was purchased from ATCC, amplified by PCR, and inserted between the *Nde*I and *Eco*RI sites of the vector pTYB12 (New England BioLabs). A fusion protein, consisting of chitin-binding domain, intein and the ABD, was expressed in the *E. coli* strain BL21Star(DE3) (Invitrogen), and purified on a chitin affinity column. The final protein contains, in addition to α-actinin 1 residues 30–253, an N-terminal (Ala-Gly-His-Met) and a C-terminal (Glu-Phe-Leu-Glu-Pro-Gly) extension derived from the expression plasmid.

2.2. Crystallization and data collection

Crystals were obtained in two closely related forms and under similar conditions at 4 °C, using the hanging-drop diffusion method (Table 1). Crystals of form I were grown in 50 mM NaCl, 1 mM EDTA, 100 mM imidazole, pH 6.7, and 24% PEG 5000 MME, and crystals of form II in 50 mM NaCl, 1 mM EDTA, 200 mM K_2HPO_4, 20% w/v polyethylene glycol 3350, pH 9.2. Datasets were collected at the IMCA-CAT facilities of the Advanced Photon Source (APS, Argonne, IL). The diffraction data were indexed and scaled with the program HKL2000 (HKL Research, Inc.).

2.3. Structure determination

A molecular replacement solution for crystal form I was obtained with the program AMoRE (Navaza and Saludjian, 1997), using the first CH domain of plectin (Garcia-Alvarez et al., 2003) and the second CH domain of β-spectrin (Carugo et al., 1997) as search models and data from 9.0 to 3.3 Å resolution. The correlation coefficient for this solution was 47.4% (R_{factor}, 0.458). After several rounds of manual rebuilding with the program O (Jones et al., 1991) and refinement with the program CNS (Brunger et al., 1998), the final R_{factor} was 18.3%

Table 1
Data collection and refinement statistics

	Crystal form I	Crystal form II
Diffraction data statistics		
Space group	$P2_12_12$	$P2_12_12$
Unit cell parameters		
a, *b*, *c* (Å)	56.76, 110.89, 34.91	53.01, 111.07, 31.55
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	50 – 1.7 (1.76–1.70)	50 – 1.8 (1.85–1.80)
Measured reflections	257,342	98,540
Unique reflections	23,591	17,130
I/σ	53.5 (6.6)	32.7 (3.5)
Completeness (%)	94.1 (69.0)	99.7 (61.6)
R_{merge}	0.077 (0.292)	0.067 (0.283)
Refinement statistics		
Resolution range (Å)	40 – 1.7 (1.75–1.70)	40 – 1.8 (1.87–1.80)
No. of reflections	23,546	16,457
σ-Cutoff	None	None
R_{factor}	0.183 (0.232)	0.184 (0.214)
R_{free}	0.205 (0.273)	0.216 (0.246)
No. of atoms		
Protein	1866	1814
Solvent	322	
RMS deviations	0.009	0.007
Bonds (Å)	0.009	0.007
Angles (°)	1.19	1.28
PDB accession code	(2EYI)	(2EYN)

Values in parentheses correspond to highest resolution shell. $R_{\text{merge}} = \sum_{hkl} |I - <I>|/\sum I$, where *I* and *<I>* are the observed and mean intensity values of reflection *hkl*. $R_{\text{factor}} = \sum_{hkl} ||F_o| - |F_c||/\sum |F_o|$, where F_o and F_c are the observed and calculated structure factors amplitudes of reflection *hkl*. R_{free}, R_{factor} calculated for a randomly selected subset of the reflections (5%) that were omitted during refinement.

($R_{\text{free}} = 20.5\%$) for all the data to 1.7 Å resolution (Table 1). This model was then used to determine the structure of the more compact crystal form II (final $R_{\text{factor}} = 18.4\%$ and $R_{\text{free}} = 21.6\%$ at 1.8 Å resolution).

2.4. Surface conservation analysis

A total of 18 representative ABD sequences were selected for this study (Fig. 1) and aligned with the program Muscle (Edgar, 2004). The alignment was manually improved with the program SeaView (Galtier et al., 1996). Per-residue conservation indices were then calculated with the program JalView (Clamp et al., 2004). Residues of the α-actinin 1 ABD (crystal form I) were gradient colored according to the conservation indices, using the program ProtSkin (http://www.mcgnmr.ca/ProtSkin/). Conserved surface areas were then visualized with the program PyMol (http://pymol.sourceforge.net/).

2.5. Surface electrostatic potential and CH1–CH2 contacts

The electrostatic potential was mapped onto the solvent exposed surfaces of the ABD and the individual CH domains (Fig. 3) using the program GRASP (Nicholls et al., 1991). Atom-to-atom contacts between the two CH domains were analyzed with CCP4 program CONTACT (CCP4, 1994). The distance cut-offs for van der Waal interactions, hydrogen bonds, and salt bridges were set, respectively, to 4.2, 3.2, and 3.2 Å (Fig. 4).

3. Results and discussions

3.1. Structure of the ABD of human α-actinin 1

We have determined the crystal structure of the actin-binding domain (ABD) of human α-actinin 1, a ubiquitously expressed isoform (Fig. 2A). Data have been collected for two crystal forms (I and II, see Section 2 and Table 1). Crystal form I diffracted to a higher resolution (1.7 Å), and the atomic structure built from this form contains all the residues of the ABD. By contrast, in crystal form II (determined to 1.8 Å resolution), several residues (81–83, 108–109) are disordered and thus not included in the final model. Therefore, subsequent structural analyses were performed on the structure resulting from crystal form I.

The ABD of α-actinin 1 displays a compact conformation with extensive contact between the two CH domains. This conformation has been observed in almost all the ABD structures thus far available (Goldsmith et al., 1997; Keep et al., 1999b; Klein et al., 2004; Norwood et al., 2000), (Franzot et al., 2005; Garcia-Alvarez et al., 2003; Sevcik et al., 2004). Strong similarities are found between the current structure and the recently published structure of the ABD of α-actinin 3 (Franzot et al., 2005), with an RMS deviation of only 0.67 Å for all the Cα atoms. Unlike α-actinin 1 whose actin-binding activity is regulated by Ca^{2+}, the binding of the muscle-specific α-actinin 3 is insensitive to Ca^{2+} (Landon et al., 1985). Such disparity can be attributed to a dysfunctional CaM domain in α-actinin 1. The ABDs of these two isoforms are rather similar (with 89.3% sequence identity) and do not appear to contribute to the functional differences mentioned above. However, there is one clear difference between the two ABD structures in the conformation of the loop connecting helices E and F of CH1 (residues 102–110). In α-actinin 1, this loop is located closer to helix G, shielding the latter more effectively from the solvent. This difference probably results from the intrinsic flexibility of the E–F loop, since the two isoforms share identical sequence in this area. Several studies have identified helix G of CH1 as a major actin-binding site (ABS2, see discussions below). Therefore, flexibility of the E–F loop might be important in allowing the solvent accessibility of this site to change readily.

3.2. Surface conservation and actin binding

The conformation of an actin-bound ABD would most likely depend on the nature of its binding interface: if this interface consists of a continuous area, spanning both CH domains, a compact conformation would be more likely. On the other hand, two independent sites (one on each CH domain) could result in an extended binding conformation. Thus, a precise characterization of the actin-binding

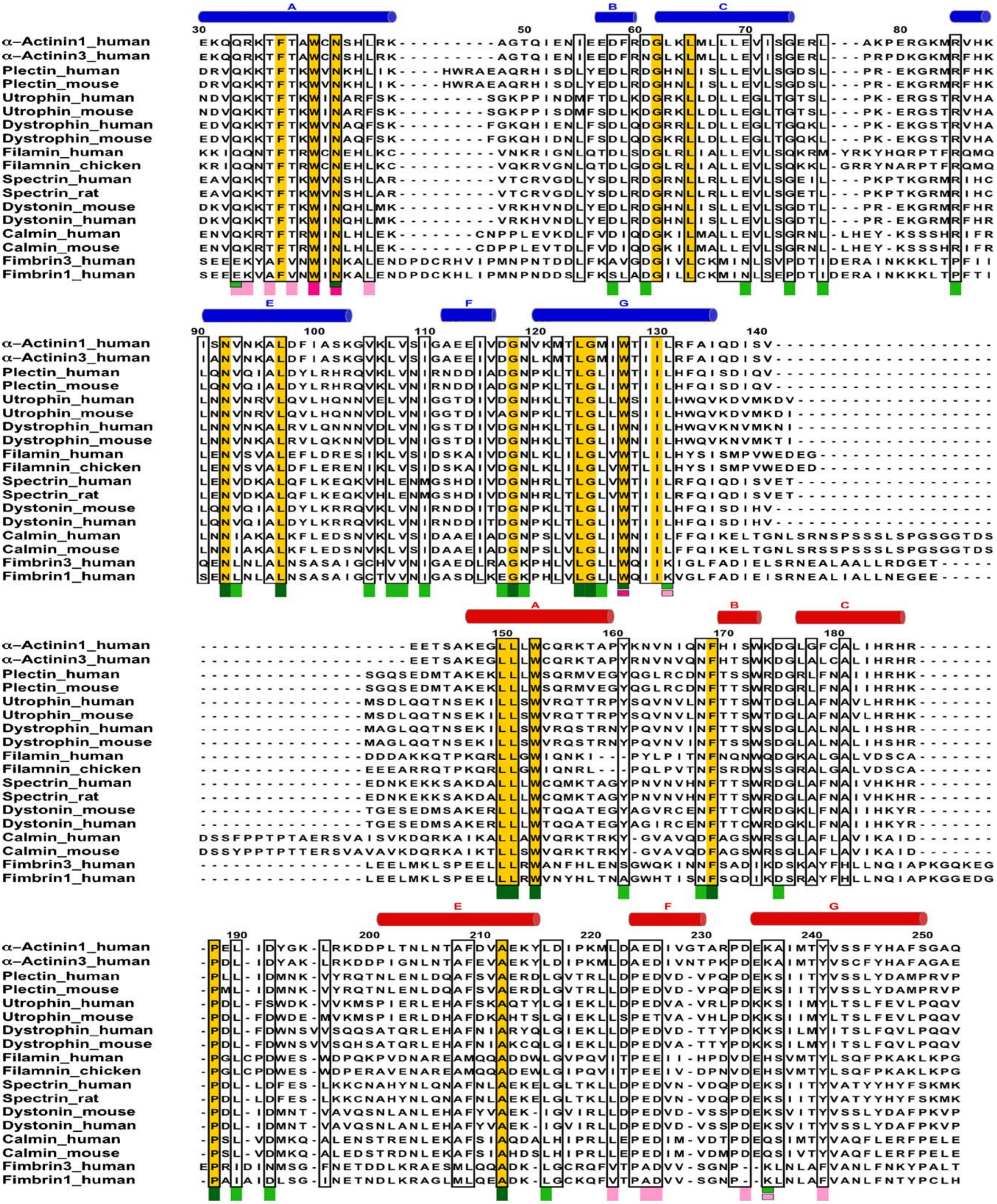


Fig. 1. Alignment of 18 representative sequences of ABD domains. Accession codes are (from top to bottom): human α-actinin 1, P12814; human α-actinin 3, Q08043; human plectin, G02520; mouse plectin, Q9QXS1; human utrophin, P46939; mouse utrophin (NP_035812), human dystrophin, P11532; mouse dystrophin, P11531; human filamin, O75369; chicken filamin, NP_989905; human β-spectrin, O15020; rat β-spectrin, Q9QWN8; mouse dystonin, NP_598594; human dystonin, Q03001; human calmin, Q96JQ2; mouse calmin, Q8C5W0; human fimbrin (plastin 3, ABD1), BAD96521; and human fimbrin (plastin 1, ABD1), AAH31083. Invariant residues are highlighted in yellow, while residues with greater than 70% conservation are boxed. The helices of CH1 and CH2, defined according to the structure of the ABD of α-actinin 1, are represented as blue and red cylinders. Residues numbers correspond to human α-actinin 1. Conserved and exposed amino acids are indicated below the alignment by dark-green (100% conservation) and light-green (>70% conservation) rectangles. Similarly, conserved amino acids at the interface of the two CHs are indicated by purple (>75% conservation) and magenta (>50% conservation) rectangles.

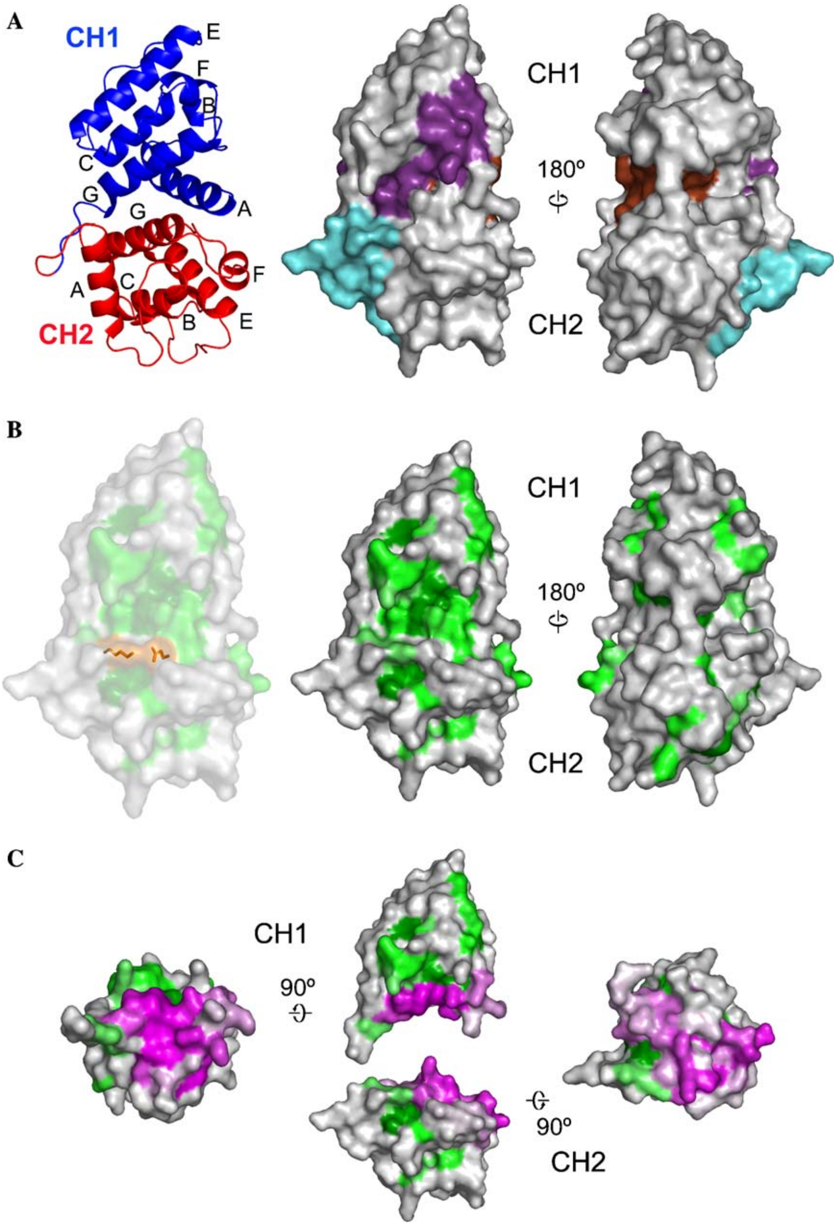


Fig. 2. Potential actin-binding interface of the ABD domain. (A) Ribbon representation of the structure of the ABD of human α-actinin 1 (CH1, blue; CH2, red) and surface representation showing the location of three identified actin-binding sites (ABS1, brown; ABS2, purple; and ABS3, cyan). (B) Surface representation of the ABD of α-actinin 1 showing the potential actin-binding region suggested by sequence conservation analysis (corresponding to amino acids with >70% conservation, gradient colored in green). Notice the general agreement between this region and ABS2 and ABS3 shown in (A). This region consists of a larger area on CH1 and a smaller area on CH2, which are connected by a salt bridge between Lys137 and Glu235 (orange, left panel). (C) Surface conservation on the separated domains of the α-actinin ABD. Conserved residues at the CH1–CH2 interface are gradient colored in pink according to their conservation indices.

interface, based on structural and sequence analysis, might give us key insights as to how ABDs interact with F-actin.

A simple and frequently attempted approach is to visualize known actin-binding regions on the surfaces of high-resolution structures. Previous mutagenesis and NMR studies have suggested three major actin-binding sites (ABSs) on various ABDs (Bresnick et al., 1991; Corrado et al., 1994; Fabbrizio et al., 1993; Kuhlman et al., 1992; Levine et al., 1990, 1992). These three sites, ABS1 (residues 34–43 of human α-actinin 1), ABS2 (114–133), and ABS3 (141–160), correspond respectively to the N- and C-terminal helices of CH1 (helices A and G), and to the interdomain linker and the N-terminal helix A of CH2. When the above information was applied to the structure of α-actinin 1, ABS1 appears as an isolated site, whereas ABS2 and ABS3 are contiguous (Fig. 2A). Similar results have been reported in a recent study on the ABD of α-actinin 3 (Franzot et al., 2005). The outcomes of such analysis, however, could be ambiguous for a number of reasons: first, each of these ABSs is likely to contain extra residues in addition to the region minimally required for actin binding; second, only a portion of each ABS is exposed, and thus would be in direct contact with F-actin; finally, the designation of ABS segments, particularly that of ABS1, is only tentative, since it was made before any structural information was available (Broderick and Winder, 2005).

Actin is one of the most conserved proteins in nature. Therefore, we reasoned that the corresponding actin-binding interface of any given family of actin-binding proteins (ABPs) would consist of highly conserved areas on otherwise variable surfaces. (Based on the assumption that the interaction is conserved among all the member of the family.) To identify such binding sites on ABDs, we first calculated the per-residue conservation indices for 18 representative ABDs with the program JalView (Clamp et al., 2004), based on a careful alignment of their sequences (Fig. 1). In the next step, all residues of the ABD of α-actinin 1 were gradient colored according the conservation indices, and surface residues were visualized using the program PyMol (see Section 2). The above analysis reveals two highly conserved regions on the surface of the ABD, consisting of a larger area centered around Leu124, Gly125 and Trp128 of ABS2 (helix G, CH1), and the E–F loop of CH1, and a smaller area centered around Leu151 of ABS3 (helix A, CH2) and Phe169 (A-B linker, CH2) (Fig. 2B). The two conserved areas are connected by a salt bridge between Lys147 (ABS3, helix A, CH2) and Glu235 (helix G, CH2), which shields the side chain of Met239 (helix G, CH2) from the solvent. A similar analysis has been performed on the crystal structure of the first ABD of human fimbrin (Banuelos et al., 1998). Interestingly, this ABD lacks such a salt bridge due to the replacement of the Glu residue with a proline. As a result, the conserved surface area spans both domains, from Gly229 and Trp232 on CH1 to Leu367 and Leu272 on CH2 (corresponding to residues Gly125, Trp128, Met239, and Leu151 of α-actinin 3). However, fimbrin ABD1 appears to be an exception since the above-mentioned ionic pair is present in the crystal structures of β-spectrin (Banuelos et al., 1998), α-actinin 3 (Franzot et al., 2005), mouse plectin (Garcia-Alvarez et al., 2003; Sevcik et al., 2004), and human plectin (Garcia-Alvarez et al., 2003; Sevcik et al., 2004). Moreover, based on sequence comparison, we further suggest that filamin and dystonin also contain this salt bridge (Fig. 1). Could Lys147 and Glu235 be part of a continuous binding surface in these ABDs? Given the conservation of this interaction, such a scenario is likely. However, the polarity of this interaction is reversed in both utrophin (Keep et al., 1999a) and dystrophin (Norwood et al., 2000), with Lys on helix A and Glu on helix G. Although such reversals may have little effect on protein stability, it is unclear if they would be tolerated in a specific binding interface.

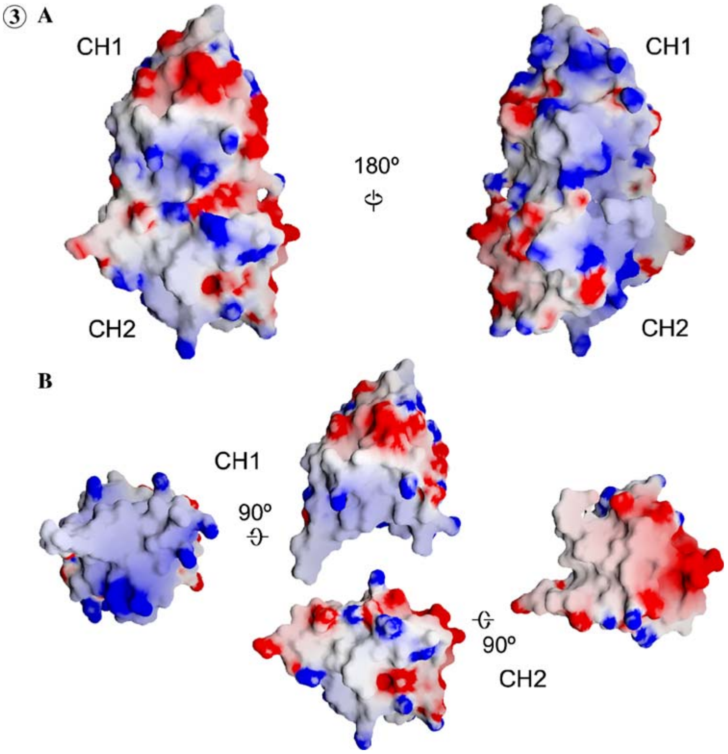
Surprisingly, in a compact ABD, no conserved surface area of considerable size is formed by the residues of ABS1, which is partially buried at the CH1–CH2 interface. A reasonable explanation is that ABS1 may become accessible to F-actin upon separation or at least reorientation of the two CH domains. Indeed, when the same surface conservation analysis was performed on individual domains (corresponding to CH domains in the extended conformation), we identified one more highly conserved area on the surface of CH1, dominated by ABS1 and ABS2 (helices A and G) (Fig. 2C). By contrast, the additionally exposed area on the separated CH2 is only moderately conserved (Fig. 2C).

3.3. CH1–CH2 interface and ABD conformation

Do the two CH domains separate upon F-actin binding? Some important hints can be garnered from a thorough analysis of the interface between CH1 and CH2. Indeed, strong and highly conserved contacts between the two domains would support a compact model, whereas weak or less conserved interactions would be consistent with an extended model. To better understand the nature of the CH1–CH2 interaction, we performed and exhaustive analysis of all the existing crystal structures of ABDs (corresponding to 9 proteins, 11 ABDs, 12 crystal forms, 19 crystallographically independent molecules). As part of this analysis, we tabulated and categorized all the atom-to-atom contacts between CH1 and CH2 within the distance of 4.2 Å (a generous distance cut-off for van der Waal contacts). The most important results of our study are presented below.

The first criterion examined was the hydrophobicity of the CH1–CH2 interface. As observed in this and previous studies, a substantial surface area on each domain (700–800 Å^2) becomes buried by the CH1–CH2 interaction. However, the entropic gain from this interaction does not depend on the size of the buried surface alone, but also on the general hydrophobicity of the buried residues. Our analysis of various ABDs indicates that there are relatively few incidences in which two non-polar side chains from the two CHs contact each other (Fig. 4A). For instance, at the CH1–CH2 interface of α-actinin 1 we found only two such contacts; between Trp128 and Met239, and between Ile136 and Met239. At least one salt bridge and several hydrogen bonds are usually observed at the interface between domains. However, their occurrence appears to be random as their positions are poorly conserved (Fig. 4A). In addition, the interfaces of various ABDs display a number of relatively less favorable contacts, including those between polar and non-polar side chains, and between hydrophobic side chains and backbone atoms. Electrostatic mapping of the ABD and the individual CH domains appears to confirm the above analysis, with a partial positive charge on CH1 and a partial negative charge on CH2 (Fig. 3). In summary, the interface between CH1 and CH2 can be best described as semi-polar.

As noted in previous studies, the general architecture of the CH1–CH2 interface remains the same for all ABDs, and involves the packing of helices A and G of CH1 against the C-terminal region of CH2 (helices F and G and the E–F linker) (Figs. 2A and 4A). For a given pair of residues involved in interdomain contacts in one ABD, an equivalent pair can usually be found in another ABD. However, the type of interaction is rarely conserved. For example, the


③ A
CH1
CH1
180º
CH2
CH2
B
CH1
90º
90º
CH2

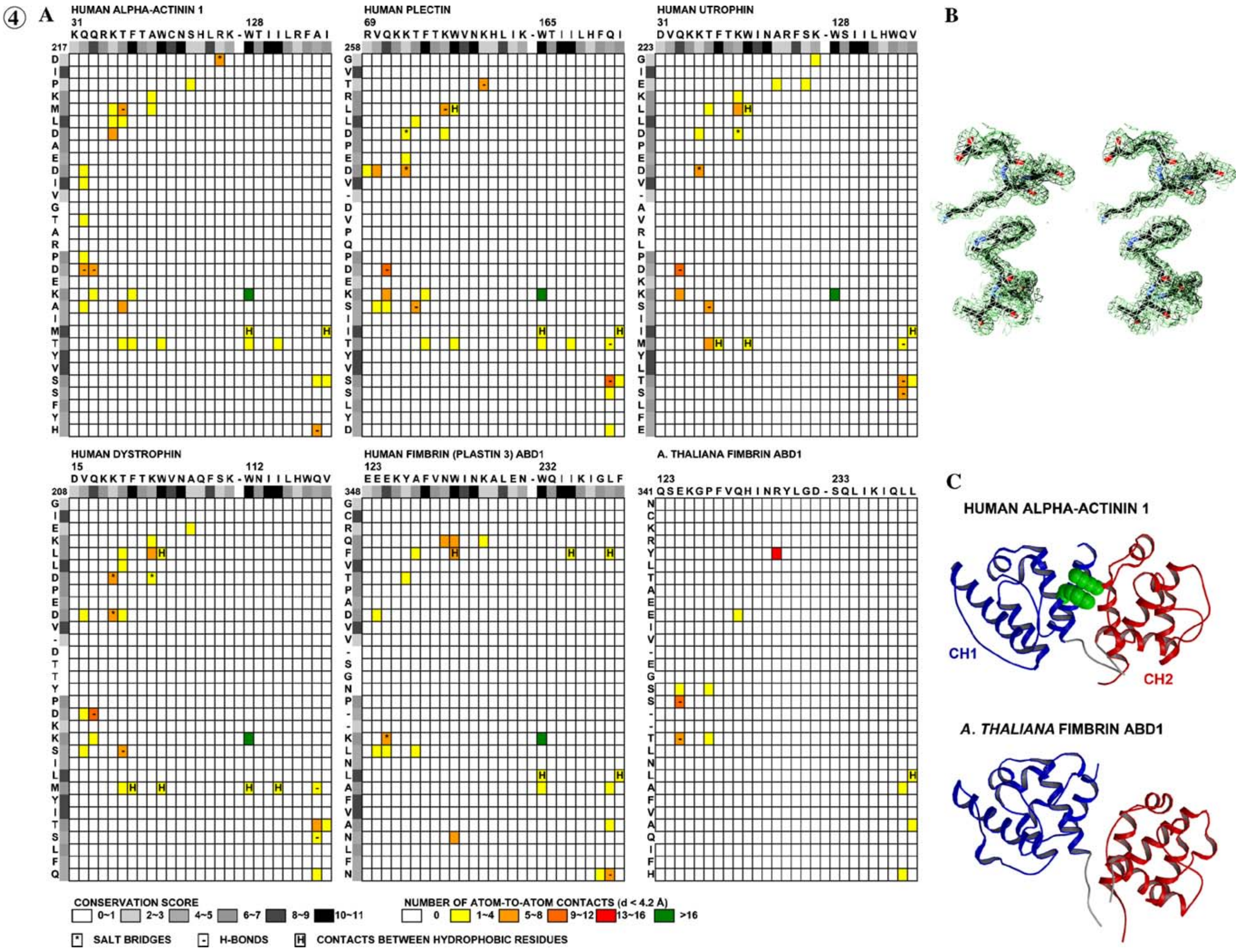

④ A
HUMAN ALPHA-ACTININ 1
31 KQQRKTFTAWCNSHLRK-WTIILRFAI 128
217
HUMAN PLECTIN
69 RVQKKTFTKWVNKHLIK-WTIILHFQI 165
258
HUMAN UTROPHIN
31 DVQKKTFTKWINARFSK-WSIILHWQV 128
223
HUMAN DYSTROPHIN
15 DVQKKTFTKWVNAQFSK-WNIILHWQV 112
208
HUMAN FIMBRIN (PLASTIN 3) ABD1
123 EEEKYAFVNWINKALEN-WQIIKIGLF 232
348
A. THALIANA FIMBRIN ABD1
123 QSEKGPFVQHINRYLGD-SQLIKIQLL 233
341
CONSERVATION SCORE
0~1
2~3
4~5
6~7
8~9
10~11
NUMBER OF ATOM-TO-ATOM CONTACTS (d < 4.2 Å)
0
1~4
5~8
9~12
13~16
>16
SALT BRIDGES
H-BONDS
CONTACTS BETWEEN HYDROPHOBIC RESIDUES
B
C
HUMAN ALPHA-ACTININ 1
CH1
CH2
A. THALIANA FIMBRIN ABD1

Fig. 3. Electrostatic surface representation of α-actinin's ABD. (A) Electrostatic potential mapped onto the solvent exposed surface of the ABD of α-actinin 1 calculated with the program GRASP (Nicholls et al., 1991, p. 646). Blue corresponds to positive and red to negative. (B) Electrostatic potential calculated for the individual CH domains of α-actinin 1. Notice the partially charged CH1–CH2 interface.

Fig. 4. Interdomain contacts in the structures of ABDs. (A) Schematic representations showing interdomain contacts in the crystal structures of 6 different ABDs (human α-actinin 1, human plectin, human utrophin, human dystrophin, human fimbrin ABD1 and *A. thaliana* fimbrin ABD1. Residues involved in interdomain contacts include those on helices A and G of CH1 (shown horizontally) and those in the C-terminal region of CH2 (helices F and G and the E–F and F–G linkers; shown vertically). The conservation of each residue is indicated by a gray scale. The numbers of atom-to-atom contacts are color-coded. (B) A stereo view representation of the conserved contact between Trp128 and Lys236, which involves both hydrophobic and cation-π interactions. A $2F_o - F_c$ electron density map contoured at 1.1σ is also shown. (C) Comparison of the α-actinin 1 ABD (top) and the first ABD (ABD1) of *A. thaliana* fimbrin (bottom). In the former, the strong interaction between Trp128 and Lys236 (green spheres) appears to hold the two domains together in a compact conformation. The ABD1 of *A. thaliana* fimbrin lacks such a contact, and in one of the molecules of the asymmetric unit (molecule A), the two CH domains are nearly separated.

interacting pair Ala39–Lys220 in α-actinin 1 is replaced with Asn131–Gln151 in human fimbrin. This observation is not totally unexpected, in particular since the C-terminal region of CH2 is only moderately conserved (Fig. 4A). Furthermore, although there are several conserved residues in helices A and G of CH1, they tend to forge intradomain instead of interdomain contacts. One contact, however, between the side chains of Trp128 (helix G, CH1 of α-actinin 1) and Lys236 (helix G, CH2), is highly conserved. This contact, which involves extensive hydrophobic interaction and occasionally a π-cation bond (Crowley and Golovin, 2005) appears to dominate the interdomain interface (Fig. 4B). Most members of the spectrin family display a similar pair, with the occasional replacement of the Lys residue by Arg or Gln (Fig. 1). An exception is found in the first ABD of *A. thaliana*, in which the corresponding positions are occupied by Ser and Thr. Interestingly, the crystal structure of this ABD (Klein et al., 2004) displays two conformations within the asymmetric unit: one resembles the compact form seen in other ABDs (albeit with altered interdomain interactions); the other displays a significantly different conformation characterized by few interactions between CHs (Fig. 4). Therefore, we suggest that the Trp-Lys pair functions as a hinge, linking CH1 and CH2 while allowing for some plasticity at the interface.

3.4. Implications for F-actin binding

Our analysis shows a highly conserved surface area spanning both CH domains of the compact ABD. We suggest that this area represents the initial actin recognition site in the compact conformation. The existence of a compact initial binding state is supported by a study on the ABD of plectin (Garcia-Alvarez et al., 2003). In that work, actin binding was demonstrated for a mutant ABD trapped in a compact conformation by an inter-domain disulfide bond. The above-mentioned site corresponds largely to ABS2, and to a minor extent, ABS3. By contrast, the conserved portion of ABS1 is not exposed in the compact ABD, possibly because this site contributes minimally to actin binding. Indeed, only ABS2 is absolutely required for binding, as shown in a study of N- and C-terminally truncated α-actinin constructs (Hemmings et al., 1992; Kuhlman et al., 1992). Alternatively, ABS1 may become exposed as a result of an actin-induced conformational change. Consistent with this view, our analysis of the separated CH1 shows an additional conserved area (centered on ABS1), which is otherwise buried in the compact ABD. The conservation of this area may indicate a role in actin binding since the separated CH2 lacks such an area.

A second binding step may result in minor rearrangement of the two domains. Indeed, although our analysis shows a poorly conserved and semi-polar interface between the two domains, in most ABDs there exists a strongly interacting Lys-Trp pair that seems to keep the domains together. Therefore, relative rotation of the two domains around the Lys-Trp hinge is a likely scenario. However, based on currently available data, the total separation of domains suggested by the extended model cannot be ruled out. Further research is required to distinguish between the two binding models.

Acknowledgments

Supported by NIH Grant HL075404 and American Heart Association Established Investigator Award 0240089N. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Use of IMCA-CAT beamline 17-ID was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with the University of Chicago.

References

Banuelos, S., Saraste, M., Carugo, K.D., 1998. Structural comparisons of calponin homology domains: implications for actin binding. Structure 6, 1419–1431.

Beggs, A.H., Byers, T.J., Knoll, J.H., Boyce, F.M., Bruns, G.A., Kunkel, L.M., 1992. Cloning and characterization of two human skeletal muscle alpha-actinin genes located on chromosomes 1 and 11. J. Biol. Chem. 267, 9281–9288.

Bresnick, A.R., Janmey, P.A., Condeelis, J., 1991. Evidence that a 27-residue sequence is the actin-binding site of ABP-120. J. Biol. Chem. 266, 12989–12993.

Broderick, M.J., Winder, S.J., 2005. Spectrin, alpha-actinin, and dystrophin. Adv. Protein Chem. 70, 203–246.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L., 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 54, 905–921.

Carugo, K.D., Banuelos, S., Saraste, M., 1997. Crystal structure of a calponin homology domain. Nat. Struct. Biol. 4, 175–179.

CCP4, 1994. The CCP4 suite: programs for protein crystallography. Acta Cryst. D50, 760–763.

Cherepanova, O.A., Orlova, A., Galkin, V.E., Kostan, J., Wiche, G., Egelman, E.H., 2005. The interaction of the plectin ABD with actin. 49th Annual Meeting of the Biophysical Society, Long Beach, California, Abstract No. 2424.

Clamp, M., Cuff, J., Searle, S.M., Barton, G.J., 2004. The Jalview Java alignment editor. Bioinformatics 20, 426–427.

Corrado, K., Mills, P.L., Chamberlain, J.S., 1994. Deletion analysis of the dystrophin-actin binding domain. FEBS Lett. 344, 255–260.

Crowley, P.B., Golovin, A., 2005. Cation-pi interactions in protein–protein interfaces. Proteins 59, 231–239.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Fabbrizio, E., Bonet-Kerrache, A., Leger, J.J., Mornet, D., 1993. Actin-dystrophin interface. Biochemistry 32, 10457–10463.

Franzot, G., Sjoblom, B., Gautel, M., Djinovic Carugo, K., 2005. The crystal structure of the actin binding domain from alpha-actinin in its closed conformation: structural insight into phospholipid regulation of alpha-actinin. J. Mol. Biol. 348, 151–165.

Fukami, K., Furuhashi, K., Inagaki, M., Endo, T., Hatano, S., Takenawa, T., 1992. Requirement of phosphatidylinositol 4,5-bisphosphate for alpha-actinin function. Nature 359, 150–152.

Fukami, K., Sawada, N., Endo, T., Takenawa, T., 1996. Identification of a phosphatidylinositol 4,5-bisphosphate-binding site in chicken skeletal muscle alpha-actinin. J. Biol. Chem. 271, 2646–2650.

Galkin, V.E., Orlova, A., VanLoock, M.S., Egelman, E.H., 2003. Do the utrophin tandem calponin homology domains bind F-actin in a compact or extended conformation? J. Mol. Biol. 331, 967–972.

Galkin, V.E., Orlova, A., VanLoock, M.S., Rybakova, I.N., Ervasti, J.M., Egelman, E.H., 2002. The utrophin actin-binding domain binds F-actin in two different modes: implications for the spectrin superfamily of proteins. J. Cell Biol. 157, 243–251.

Galtier, N., Gouy, M., Gautier, C., 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. 12, 543–548.

Garcia-Alvarez, B., Bobkov, A., Sonnenberg, A., de Pereda, J.M., 2003. Structural and functional analysis of the actin binding domain of plectin suggests alternative mechanisms for binding to F-actin and integrin beta4. Structure (Camb.) 11, 615–625.

Gimona, M., Djinovic-Carugo, K., Kranewitter, W.J., Winder, S.J., 2002. Functional plasticity of CH domains. FEBS Lett. 513, 98–106.

Goldsmith, S.C., Pokala, N., Shen, W., Fedorov, A.A., Matsudaira, P., Almo, S.C., 1997. The structure of an actin-crosslinking domain from human fimbrin. Nat. Struct. Biol. 4, 708–712.

Hanein, D., Volkmann, N., Goldsmith, S., Michon, A.M., Lehman, W., Craig, R., DeRosier, D., Almo, S., Matsudaira, P., 1998. An atomic model of fimbrin binding to F-actin and its implications for filament crosslinking and regulation. Nat. Struct. Biol. 5, 787–792.

Hemmings, L., Kuhlman, P.A., Critchley, D.R., 1992. Analysis of the actin-binding domain of alpha-actinin by mutagenesis and demonstration that dystrophin contains a functionally homologous domain. J. Cell Biol. 116, 1369–1380.

Honda, K., Yamada, T., Endo, R., Ino, Y., Gotoh, M., Tsuda, H., Yamada, Y., Chiba, H., Hirohashi, S., 1998. Actinin-4, a novel actin-bundling protein associated with cell motility and cancer invasion. J. Cell Biol. 140, 1383–1393.

Jones, T.A., Zou, J.-Y., Cowan, S.W., Kjeldgaard, M., 1991. Improved methods for building protein models in electron density maps and location of errors in these models. Acta Crystallogr. Sect. A 47, 110–119.

Keep, N.H., Norwood, F.L., Moores, C.A., Winder, S.J., Kendrick-Jones, J., 1999a. The 2.0 A structure of the second calponin homology domain from the actin-binding region of the dystrophin homologue utrophin. J. Mol. Biol. 285, 1257–1264.

Keep, N.H., Winder, S.J., Moores, C.A., Walke, S., Norwood, F.L., Kendrick-Jones, J., 1999b. Crystal structure of the actin-binding region of utrophin reveals a head-to-tail dimer. Structure Fold Des. 7, 1539–1546.

Klein, M.G., Shi, W., Ramagopal, U., Tseng, Y., Wirtz, D., Kovar, D.R., Staiger, C.J., Almo, S.C., 2004. Structure of the actin crosslinking core of fimbrin. Structure (Camb.) 12, 999–1013.

Kuhlman, P.A., Hemmings, L., Critchley, D.R., 1992. The identification and characterisation of an actin-binding site in alpha-actinin by mutagenesis. FEBS Lett. 304, 201–206.

Landon, F., Gache, Y., Touitou, H., Olomucki, A., 1985. Properties of two isoforms of human blood platelet alpha-actinin. Eur. J. Biochem. 153, 231–237.

Lehman, W., Craig, R., Kendrick-Jones, J., Sutherland-Smith, A.J., 2004. An open or closed case for the conformation of calponin homology domains on F-actin? J. Muscle Res. Cell Motil. 25, 351–358.

Levine, B.A., Moir, A.J., Patchell, V.B., Perry, S.V., 1990. The interaction of actin with dystrophin. FEBS Lett. 263, 159–162.

Levine, B.A., Moir, A.J., Patchell, V.B., Perry, S.V., 1992. Binding sites involved in the interaction of actin with the N-terminal region of dystrophin. FEBS Lett. 298, 44–48.

Liu, Y., Eisenberg, D., 2002. 3D domain swapping: as domains continue to swap. Protein Sci. 11, 1285–1299.

McGough, A., Way, M., DeRosier, D., 1994. Determination of the alpha-actinin-binding site on actin filaments by cryoelectron microscopy and image analysis. J. Cell Biol. 126, 433–443.

Moores, C.A., Keep, N.H., Kendrick-Jones, J., 2000. Structure of the utrophin actin-binding domain bound to F-actin reveals binding by an induced fit mechanism. J. Mol. Biol. 297, 465–480.

Navaza, J., Saludjian, P., 1997. AMoRe: an automated molecular replacement program package. Methods Enzymol. 276, 581–594.

Nicholls, A., Sharp, K.A., Honig, B., 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 11, 281–296.

Norwood, F.L., Sutherland-Smith, A.J., Keep, N.H., Kendrick-Jones, J., 2000. The structure of the N-terminal actin-binding domain of human dystrophin and how mutations in this domain may cause Duchenne or Becker muscular dystrophy. Structure Fold Des. 8, 481–491.

Otey, C.A., Carpen, O., 2004. Alpha-actinin revisited: a fresh look at an old player. Cell Motil. Cytoskeleton 58, 104–111.

Pavalko, F.M., Burridge, K., 1991. Disruption of the actin cytoskeleton after microinjection of proteolytic fragments of alpha-actinin. J. Cell Biol. 114, 481–491.

Sevcik, J., Urbanikova, L., Kost'an, J., Janda, L., Wiche, G., 2004. Actin-binding domain of mouse plectin. Crystal structure and binding to vimentin. Eur. J. Biochem. 271, 1873–1884.

Sutherland-Smith, A.J., Moores, C.A., Norwood, F.L., Hatch, V., Craig, R., Kendrick-Jones, J., Lehman, W., 2003. An atomic model for actin binding by the CH do mains and spectrin-repeat modules of utrophin and dystrophin. J. Mol. Biol. 329, 15–33.

Way, M., Pope, B., Weeds, A.G., 1992. Evidence for functional homology in the F-actin binding domains of gelsolin and alpha-actinin: implications for the requirements of severing and capping. J. Cell Biol. 119, 835–842.

Winder, S.J., 1996. Structure-function relationships in dystrophin and utrophin. Biochem. Soc. Trans. 24, 497–501.

Modulation of actin structure and function by phosphorylation of Tyr-53 and profilin binding

Kyuwon Baek*†, Xiong Liu†‡, François Ferron*, Shi Shu‡, Edward D. Korn‡§, and Roberto Dominguez*§

*Department of Physiology, 3700 Hamilton Walk, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6085; and ‡Laboratory of Cell Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892

Contributed by Edward D. Korn, June 17, 2008 (sent for review May 22, 2008)

On starvation, *Dictyostelium* cells aggregate to form multicellular fruiting bodies containing spores that germinate when transferred to nutrient-rich medium. This developmental cycle correlates with the extent of actin phosphorylation at Tyr-53 (pY53-actin), which is low in vegetative cells but high in viable mature spores. Here we describe high-resolution crystal structures of pY53-actin and unphosphorylated actin in complexes with gelsolin segment 1 and profilin. In the structure of pY53-actin, the phosphate group on Tyr-53 makes hydrogen-bonding interactions with residues of the DNase I-binding loop (D-loop) of actin, resulting in a more stable conformation of the D-loop than in the unphosphorylated structures. A more rigidly folded D-loop may explain some of the previously described properties of pY53-actin, including its increased critical concentration for polymerization, reduced rates of nucleation and pointed end elongation, and weak affinity for DNase I. We show here that phosphorylation of Tyr-53 inhibits subtilisin cleavage of the D-loop and reduces the rate of nucleotide exchange on actin. The structure of profilin–*Dictyostelium*-actin is strikingly similar to previously determined structures of profilin–β-actin and profilin–α-actin. By comparing this representative set of profilin–actin structures with other structures of actin, we highlight the effects of profilin on the actin conformation. In the profilin–actin complexes, subdomains 1 and 3 of actin close around profilin, producing a 4.7° rotation of the two major domains of actin relative to each other. As a result, the nucleotide cleft becomes moderately more open in the profilin–actin complex, probably explaining the stimulation of nucleotide exchange on actin by profilin.

actin phosphorylation | profilin–actin structure | pY53-actin structure | *Dictyostelium discoideum* actin | gelsolin–actin structure

Multiple cellular functions, including cell motility, cell division, and endocytosis, involve dynamic remodeling of the actin cytoskeleton (1, 2). Various studies have established a connection between actin phosphorylation and cytoskeleton remodeling. For example, fibroblast stimulation with epidermal growth factors induces actin phosphorylation on serine residues and the formation of membrane ruffles (3). Actin is also one of the proteins found to be tyrosine phosphorylated in fibroblasts expressing constitutively active Src tyrosine kinase and displaying significant cytoskeleton rearrangement (4). In *Mimosa pudica*, a plant that closes its leaves and droops its petioles when touched, actin is heavily tyrosine phosphorylated, and a decrease in actin phosphorylation correlates with petiole bending (5). In none of these examples, however, is the precise connection between actin phosphorylation and cytoskeleton remodeling well understood. Such a connection is better established in *Dictyostelium* cells, in which the developmental cycle correlates closely with the extent of actin tyrosine phosphorylation (6–11).

Dictyostelium cells grow and divide as amoebae in nutrient medium, but they aggregate and differentiate into multicellular organisms on starvation, ultimately forming fruiting bodies that contain spores that germinate when conditions become favorable for growth (12). The extent of actin phosphorylation, which is very low in growing vegetative cells, begins to increase 12–24 h into the developmental cycle, reaching ≈50% of the total actin at ≈36 h (9–11). At this high level of actin phosphorylation, the spores of the mature fruiting bodies remain viable for ≈20 days, at which time viability and actin phosphorylation levels both decrease, disappearing entirely by ≈30 days. Increases in actin phosphorylation also occur when vegetative cells are exposed to heat stress, sodium azide, and the phosphotyrosine phosphatase inhibitor phenylarsine oxide (6, 8, 11, 13).

The phosphorylation site on *Dictyostelium* actin has been mapped to residue Tyr-53 (8, 11). This site is near the DNase I-binding loop (D-loop) of actin (residues 40–50), which is implicated in intersubunit contacts in the filament (14–17). The D-loop is disordered in most crystal structures of actin, including in the two structures of unphosphorylated *Dictyostelium* actin described here in complexes with profilin and gelsolin segment 1 (G1). In contrast, the D-loop is partially stabilized by hydrogen-bonding contacts with the phosphate group on Tyr-53 in the structure of Tyr-53-phosphorylated actin (pY53-actin), which is also described here in complex with G1. The stabilization of the D-loop is further supported by biochemical characterization of pY53-actin in solution.

Profilin stimulates nucleotide exchange on actin (18, 19). However, the structural bases for this activity are not well understood. The crystal structure of profilin–*Dictyostelium*-actin described here is strikingly similar to previously determined structures of profilin–β-actin (20) and profilin–α-actin (21). Based on this representative group of profilin–actin structures, we analyze the effects of profilin on the conformation of actin and its role in nucleotide exchange.

Results and Discussion

Structures of pY53-Actin and Unphosphorylated Actin Complexed with G1. The different biochemical properties of pY53-actin and unphosphorylated actin suggested that their structures may be different (11). To test this possibility, we set out to crystallize pY53-actin and unphosphorylated *Dictyostelium* actin under identical conditions, so that their structures could be compared directly. Actin's natural tendency to polymerize constitutes an obstacle to crystallization. Different approaches have been used to overcome this problem, including the crystallization of complexes of actin with actin-binding proteins (ABPs) (20, 22–26) and toxins (27), and blocking actin polymerization by mutagenesis (28) or chemical cross-linking (29). We chose to attempt

Author contributions: E.D.K. and R.D. designed research; K.B., X.L., F.F., and S.S. performed research; K.B., X.L., E.D.K., and R.D. analyzed data; and K.B., E.D.K., and R.D. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3CHW, 3CIP, 3CI5).

[†] K.B. and X.L. contributed equally to this work.

[§] To whom correspondence may be addressed. E-mail: edk@nih.gov or droberto@mail.med.upenn.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0805852105/DCSupplemental.

Table 1. Crystallographic data and refinement statistics

	Profilin–actin–VASP$_{202–244}$	Gelsolin–actin	Gelsolin–pY53–actin
Space group	$P2_12_12_1$	$C2$	$C2$
Unit cell $a/b/c$, Å	38.04/76.18/180.82	178.88/69.17/56.56	178.42/69.05/56.54
Unit cell $\alpha/\beta/\gamma$, °	90/90/90	90/104.36/90	90/104.2/90
Resolution, Å	2.3–50 (2.3–2.38)	1.6–50 (1.6–1.66)	1.7–50 (1.7–1.76)
Completeness, %	89.7 (80.9)	96.8 (78.4)	94.9(57.9)
Multiplicity	11.0 (11.1)	7.3 (3.8)	8.0 (2.1)
R_{sym}, %*	11.7 (34.9)	7.1 (52.1)	8.4 (40.7)
I/σ	20.3 (15.8)	36.6 (2.1)	23.1 (2.2)
R_{factor}, %†	15.7	15.0	14.1
R_{free}, %‡	22.7	19.7	19.1
rms bonds, Å	0.015	0.011	0.012
rms angles, °	1.504	1.326	1.365
B-factor actin/ABP, Å^2	22.54/20.40	24.88/23.21	29.40/30.32
B-factor solvent, Å^2	22.89	38.60	44.71
Number of aa/waters	513/318	491/461	494/460
PDB code	3CHW	3CIP	3CI5

Values in parentheses correspond to highest resolution shell.

*$R_{\mathrm{sym}} = \Sigma(I - \langle I \rangle)/\Sigma sI$; I and $\langle I \rangle$, intensity and mean intensity of a reflection.

†$R_{\mathrm{factor}} = \Sigma|F_o - F_c|/\Sigma|F_o|$; F_o and F_c, observed and calculated structure factors.

‡R_{free}; R_{factor} of 5% of the reflections that were not used in refinement.

crystallization with profilin, gelsolin, vitamin D-binding protein (DBP), and toxofilin, all proteins that bind to actin on the opposite side from subdomain 2, which is where Tyr-53 is located and phophorylation-dependent structural changes are more likely to occur. Unphosphorylated *Dictyostelium* actin had already been crystallized in complex with G1 (30), and we obtained crystals under similar conditions of both the unphosphorylated and phosphorylated forms (see *Materials and Methods*). In addition, we obtained crystals of unphosphorylated actin with profilin and the vasodilator-stimulated phosphoprotein (VASP) polyproline peptide ^{198}GAGGGPPPAPPLAAQ^{213}, which binds to profilin on the opposite side from actin (21). pY53-actin failed to crystallize with profilin under similar conditions (see *Conclusions*). Toxofilin (26) did not crystallize with either form of actin, and although DBP-actin yielded large crystals, they diffracted x-rays poorly.

The structures of complexes of G1 with unphosphorylated actin and pY53-actin were determined to resolutions of 1.6 Å and 1.7 Å, respectively (Table 1). The two structures are strikingly similar to each other and to prior structures of mammalian skeletal α-actin with G1 (23) and gelsolin segments 1 to 3 (31), and *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and *Dictyostelium discoideum* actin with G1 (30). In particular, the conformation of actin is nearly identical in all these complexes, with the notable exception of the D-loop (supporting information (SI) Fig. S1). The D-loop is one of the most flexible parts of the actin molecule, and it is disordered in most of the structures. Because of its intrinsic flexibility, we cannot compare the conformation of the D-loop among different structures. Factors such as crystal packing, crystallization conditions, actin isoform, and the identity of the ABP used in the crystallization may influence the conformation of the D-loop. In contrast, in the current study we crystallized unphosphorylated actin and pY53-actin under identical conditions, allowing for a direct comparison of the effect of phosphorylation on the conformation of the D-loop.

The conformation of the side chain of Tyr-53 is essentially the same in pY53-actin and unphosphorylated actin (Fig. 1 and Movie S1) and similar to other structures of actin, except for the phosphate group on Tyr-53, which is clearly defined in the electron density map of pY53-actin, but absent in unphosphorylated actin. To further validate this observation, we analyzed crystals of the two forms of actin using tandem mass spectrometry (Fig. S2). After separation by metal affinity chromatography of phospho-peptides from tryptic digestions of protein samples from the crystals, peptides containing phosphorylated Tyr-53 were detected only in crystals of pY53-actin. This result confirmed that the crystals of unphosphorylated actin were free of the phosphorylated form. Although we cannot exclude the presence of trace amounts of unphosphorylated actin in crystals of pY53-actin, crystallographic occupancies of the phosphate atoms in this high-resolution structure indicate that phosphorylation is ≈100%.

The phosphate oxygens on Tyr-53 make hydrogen-bonding interactions with the side-chain nitrogen atoms of Lys-61 and Gln-49 and with the main-chain nitrogen atoms of Gln-49 and Gly-48 in the D-loop (Fig. 1*B*). As a result, the D-loop, which is fully disordered in the unphosphorylated structure (residues 42–49 were not visualized), becomes partially ordered in pY53-actin, where four additional residues of the loop were resolved in the electron density map (Gly-42, Met-47, Gly-48, Gln-49). Although additional density was present, residues 43–46 could still not be unambiguously traced. The average temperature factor for the four additional residues observed in the D-loop is 79.4 Å^2, compared to ≈30 Å^2 for the rest of the structure (Table 1), suggesting that although more constrained, the D-loop is still quite dynamic in this structure. Other than the changes in the D-loop, the structural differences due to phosphorylation appear minor (Movie S1). However, a symmetry-related molecule in the crystal is located near the D-loop (Fig. S3), and it is possible that the proximity of this crystal contact limited the full extent of the conformational change. Another factor that could have limited the magnitude of the conformational change is the presence of G1 in the structure.

Probing the Structures of pY53-Actin and Unphosphorylated Actin in Solution. The crystal structures suggest that phosphorylation of Tyr-53 affects the conformation of the D-loop. We used subtilisin cleavage of the D-loop to test this observation in solution. The susceptibility of the D-loop to subtilisin cleavage between residues Met-47 and Gly-48 has been commonly used to monitor conformational changes in the D-loop resulting from factors such as the type of nucleotide and divalent cation bound to actin (32, 33), and the binding of actin-depolymerizing factor (ADF)/

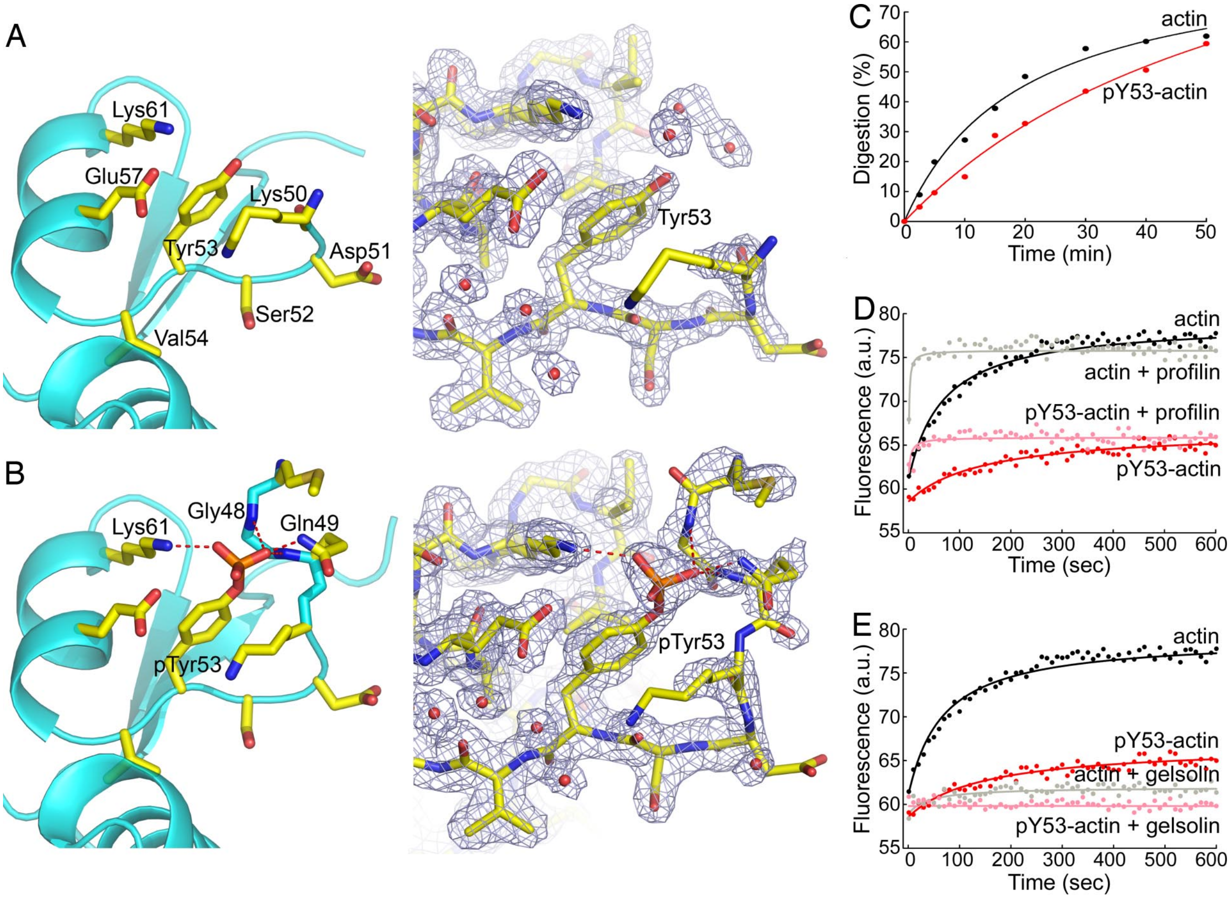


Fig. 1. Conformational change in actin subdomain 2 on phosphorylation of Tyr-53. (*A* and *B*) Close views of subdomain 2 in the structures of unphosphorylated actin and pY53-actin, showing omit electron density maps (contoured at 1 σ) around Tyr-53 (see Figs. S3 and S5 for a full view of the G1-actin structure). The D-loop was not visualized in the unphosphorylated structure. Hydrogen-bonding contacts (red dashed lines) between the oxygen atoms of the phosphate group on Tyr-53 and residues of the D-loop stabilize the conformation of the D-loop in the structure of pY53-actin. This and other figures of the paper were generated with the program PyMOL (http://pymol.sourceforge.net/). (*C*) Phosphorylation protects the D-loop from subtilisin cleavage, as shown by the ≈50% decrease in the initial rate of digestion. (*D* and *E*) Based on the increase in fluorescence as etheno-ATP replaces actin-bound ATP, phosphorylation reduces the rate of nucleotide exchange from 0.011 s^{-1} for unphosphorylated actin to 0.006 s^{-1} for pY53-actin, but profilin accelerates and gelsolin inhibits nucleotide exchange to the same extents for both forms of actin. The increase in fluorescence at equilibrium for pY53-actin is only 50% of the increase for unphosphorylated actin. Data were recorded every 10 s.

cofilin to the filament (34). We found that phosphorylation of Tyr-53 protects the D-loop from subtilisin cleavage, as shown by a ≈50% reduction in the initial rate of cleavage of the D-loop in pY53-actin compared to unphosphorylated actin (Fig. 1*C*). The protection of the D-loop from subtilisin cleavage is consistent with the more stably folded conformation of the loop observed in the structure of pY53-actin.

Because the exchange of ATP for ADP on actin is thought to alter the conformation of the D-loop (32), we speculated that changes in the conformation of the loop might reciprocally affect nucleotide exchange on actin. Consistent with this prediction, we found that phosphorylation of Tyr-53 reduces the initial rate of nucleotide exchange on actin by ≈45% (as measured by the increase in fluorescence as etheno-ATP replaces bound ATP), doubles the half-time to reach equilibrium, and reduces the fluorescence of etheno-ATP at equilibrium by ≈50% (Fig. 1 *D* and *E*). Because the conformational change in the D-loop does not seem to extend to the nucleotide cleft (Movie S1), the structural reasons for the differences in the rate of nucleotide exchange and etheno-ATP equilibrium fluorescence are unclear. Probably, pY53-actin and unphosphorylated actin have different affinities for ATP and etheno-ATP, and the fluorescence of etheno-ATP bound in the nucleotide cleft of pY53-actin may be masked by the conformational change in the D-loop. Finally, it is known that profilin accelerates (18, 19) and G1 inhibits (35) nucleotide exchange on actin. Consistent with these reports, we found that profilin stimulates (Fig. 1*D*) and G1 inhibits (Fig. 1*E*) nucleotide exchange on *Dictyostelium* actin. The effects of these two ABPs are very similar for phosphorylated and unphosphorylated actin.

Structure of Unphosphorylated *Dictyostelium* Actin Complexed with Profilin. The structural bases for the stimulation of nucleotide exchange by profilin (18, 19) (Fig. 1*D*) have remained unclear. Although the original structure of profilin–β-actin revealed a moderately open nucleotide cleft in actin (20), a subsequent structural determination suggested a far more open structure (36). Some biochemical observations have also been interpreted as evidence of a more open cleft in actin than suggested by the majority of the crystal structures (37). However, a wide-open cleft appears to be structurally unstable (38). More important, the wide-open structure of profilin–β-actin (36) was obtained in an unconventional way, by transferring the original crystals (20) into a high-phosphate solution. Thus, opening of the cleft was obtained by crystal manipulation rather than a physiologically relevant factor.

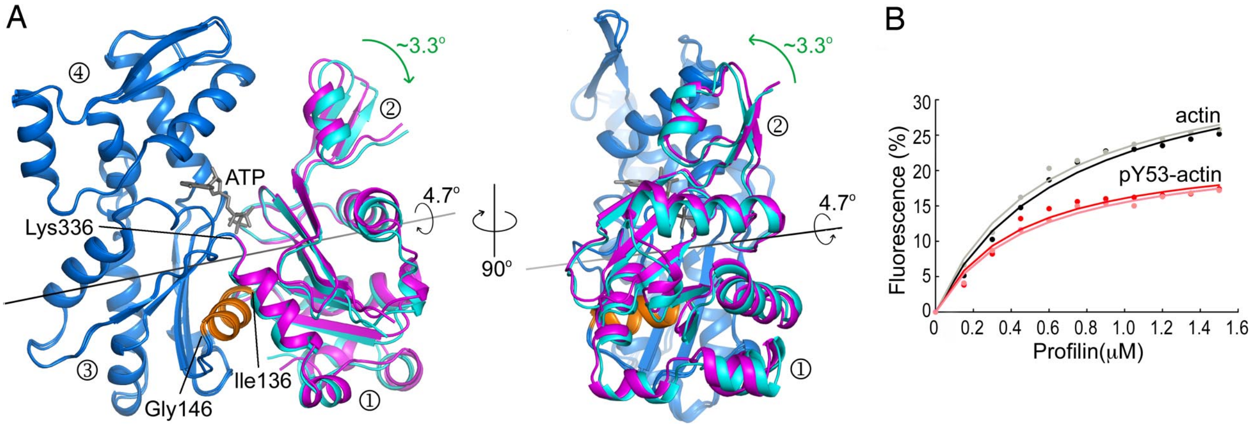


Fig. 2. Profilin binding causes a moderate opening of the nucleotide cleft in actin. (*A*) Superimposition of the structures of profilin–*Dictyostelium*-actin (blue and cyan) and uncomplexed monomeric actin (28) (blue and magenta). Two orientations are shown, rotated by 90°. The latter structure was obtained by mutagenesis in subdomain 4 and is thought to be free of perturbations resulting from the binding of an ABP or chemical cross-linking. For clarity, profilin is not shown in this figure (see Figs. S5 and S6 for a full view of the profilin–actin structure). Subdomains 3 and 4 of the structures were superimposed (blue) to highlight the relative movement of subdomains 1 and 2 (magenta or cyan). Using the classical view of actin as a reference (left view), the 4.7° rotation (calculated with the program DynDom, http://www.sys.uea.ac.uk/dyndom/) between the two major domains of actin can be visualized as two perpendicular rotations of ≈3.3°. The center of this rotation approximately coincides with the junctions between domains, consisting of residue Lys-336 and the helix between residues Ile-136 and Gly-146. Comparison of the profilin–actin structures with any other structure of actin, except for the wide-open structure of profilin–β-actin (36), results in a similar motion of the two major domains (see also Movies S2 and S3). This movement appears less dramatic than previously anticipated (36, 37), but it is probably sufficient to explain the stimulation of nucleotide exchange by profilin. (*B*) Quenching of tryptophan fluorescence on profilin binding (the results of two identical experiments, with different preparations of both actins, are shown). Profilin binds pY53-actin and unphosphorylated actin with similar affinities (K_d = 0.090 and 0.057 μM, respectively), but the quenching of tryptophan fluorescence is significantly less for profilin–pY53-actin.

We recently reported two structures of profilin–α-actin (21), determined to resolutions of 1.8 Å and 1.5 Å in the presence of two different fragments of human VASP. Here we describe the structure of profilin–*Dictyostelium*-actin to 2.3 Å resolution (Fig. 2*A* and Table 1). The two structures of profilin–α-actin and that of profilin–*Dictyostelium*-actin are strikingly similar to one another and to the original 2.55 Å structure of profilin–β-actin (20) (Fig. S4*A*), but different from the wide-open structure of profilin–β-actin (36) (Fig. S4*B*). Therefore, we now have possession of a representative set of four profilin–actin structures, determined under different crystallization conditions and crystal packing environments, and corresponding to different isoforms and sources of actin.

Can we identify structural features unique to the profilin–actin complex that would explain the acceleration of nucleotide exchange? To answer this question, we compared the profilin–actin structures to all of the actin structures in the protein data bank. The principal conclusion from this analysis is that compared to any other structure of actin, the nucleotide cleft is moderately more open in the profilin–actin complex (Fig. 2*A*). The conformational change leading to cleft opening begins with subdomains 1 and 3 of actin closing around the profilin molecule. As a result, the two major domains of actin on each side of the nucleotide cleft rotate by 4.7° relative to each other (Fig. 2*A*). By using the classical view of actin as a reference (Fig. 1*A Left*), this rotation can be described as two perpendicular rotations of ≈3.3° (Fig. 2*A*), roughly corresponding to the propeller-like twisting and scissor-like opening of the domains suggested by normal-mode analysis (39). An important distinction, however, is that rather than a scissor-like motion, we observe a clamp-like motion, i.e., as the target-binding cleft of actin "clamps" on the profilin molecule, the nucleotide cleft opens at the opposite side of the molecule. To better illustrate this movement, we generated a movie of the conformational change by linear interpolation between the atomic coordinates of the profilin–*Dictyostelium*-actin structure and the ATP-bound structure of uncomplexed monomeric actin (28) (Movie S2). Note that the latter structure was chosen for this comparison because it was obtained by mutagenesis of two residues in subdomain 4 (which prevents polymerization), and it is therefore free of perturbations resulting from the binding of an ABP or chemical cross-linking. To further show that cleft opening is a general feature of the profilin–actin complex, we generated a second interpolation, using as a reference the structure of unphosphorylated *Dictyostelium* actin in complex with gelsolin (Movie S3).

Although a rotation of the actin domains similar to that described above had already been pointed out when the structure of profilin–β-actin was first reported (20), this effect was attributed to differences between α- and β-actin. In contrast, we find that the profilin-induced rotation is independent of the source of the actin. Moreover, the earlier comparison was made to the only other actin structure available at the time, that of the complex with DNase I (22), which binds atop actin subdomains 2 and 4 and probably limits nucleotide cleft opening.

As mentioned above, we could not obtain crystals of pY53-actin with profilin, which led us to suspect that Tyr-53 phosphorylation affects the interaction with profilin. The binding of profilin to actin is accompanied by a large decrease in the intensity of tryptophan fluorescence, which is thought to result from shielding of actin Trp-356 by profilin in the complex (40). Therefore, we measured the quenching of tryptophan fluorescence to estimate the binding affinities of profilin for pY53-actin and unphosphorylated actin. We found that profilin binds pY53-actin and unphosphorylated actin with similar affinities, although the quenching of tryptophan fluorescence is significantly less for profilin–pY53-actin (Fig. 2*B*). Because the structure of profilin–pY53-actin could not be obtained, it is unclear what structural differences are responsible for this observation. Interestingly, in the two gelsolin structures, the region around Trp-356 is flexible, with residues Phe-352 to Trp-356 displaying alternative side-chain rotamers (Figs. S1 and S5). In all other structures of actin, Trp-356 presents a single and unique side-chain orientation (Fig. S1), which is also the orientation of Trp-356 in the structure of profilin–*Dictyostelium*-actin (Fig. S5). If the region around Trp-356 is intrinsically flexible, as suggested by the two gelsolin structures, quenching of tryptophan fluorescence during profilin binding might result from locking Trp-356 into a single conformation that is less exposed to solvent, rather than direct masking of Trp-356 by profilin. Finally, because the binding affinity of profilin is not significantly affected by actin phosphorylation (Fig. 2*B*), it is likely that our inability to obtain crystals of profilin–pY53-actin resulted from a conflict between the conformational change in the D-loop and crystal contacts. Indeed, analysis of crystal packing contacts reveals a symmetry-related molecule near the D-loop (Fig. S6).

Conclusions

The extent of actin phosphorylation at Tyr-53 varies dramatically during the different stages of the developmental cycle of *Dictyostelium* cells (9–11). Biochemically, actin Tyr-53 phosphorylation had been shown to increase the critical concentration for polymerization, reduce the rates of nucleation and pointed end elongation, and decrease the affinity for DNase I (11). The proximity of Tyr-53 to the D-loop, which is implicated in intersubunit contacts in the filament (14–17) and mediates the interaction with DNase I (22), was believed to account for most of the biochemical properties of pY53-actin (8, 11). The structures of pY53-actin and unphosphorylated actin seem to confirm this hypothesis. The D-loop is flexible and unresolved in most actin structures. However, in the structure of pY53-actin, hydrogen-bonding contacts between the phosphate oxygens on Tyr-53, and residues of the D-loop help stabilize its conformation (Fig. 1*B*). Phosphorylation of Tyr-53 protects the D-loop from subtilisin cleavage (Fig. 1*C*), providing additional evidence for a more stable conformation of the loop in solution. A more stably folded D-loop would probably interfere with the binding of DNase I, and with intersubunit contact in the filament.

Our biochemical analyses suggest that phosphorylation of Tyr-53 also affects other regions of the actin molecule, including the nucleotide cleft and the target-binding cleft between subdomains 1 and 3, where both profilin and gelsolin bind (Fig. S5). Thus, phosphorylation reduces the quenching of tryptophan fluorescence that accompanies the binding of profilin (Fig. 2*B*), lowers the rate of nucleotide exchange on actin, and reduces the fluorescence of etheno-ATP bound to actin (Fig. 1 *D* and *E*). Probably connected with these observations, the hydrolysis of ATP on polymerization was previously shown to be much slower for pY53-actin than for unphosphorylated actin (11). The causes of these biochemical differences are not apparent from analysis of the structures, which show little difference between pY53-actin and unphosphorylated actin other than the conformational change in subdomain 2 (Movie S1). However, there are multiple examples of long-range allosteric interactions in actin, including intramolecular coupling between the target-binding cleft and subdomain 2, which occur both in the monomer and in the filament (41). The causes of these allosteric effects are not always clear from comparisons of the structures. Phosphorylation of Tyr-53 appears to be yet another example of a modification in subdomain 2 that alters the dynamic equilibrium between different regions of the actin molecule, from the D-loop, through the nucleotide cleft, and down to the target-binding cleft.

We find that profilin–actin complexes present a moderately more open nucleotide cleft than other actin structures. Opening of the nucleotide cleft results from a relatively small rotation of 4.7° of the two major domains of actin relative to each other (Fig. 2*A* and Movies S2 and S3). This movement is less dramatic than previously anticipated (36, 37), but it is probably sufficient to explain the stimulation of nucleotide exchange produced by profilin. Indeed, the most important factor in determining nucleotide exchange might not necessarily be the degree of cleft opening, but rather changes in the intricate network of hydrogen bonding interactions that coordinate the nucleotide in the catalytic cleft. Opening of the cleft is accompanied by slight changes in this network and a small twist of the nucleotide (Movies S2 and S3). Whereas the moderately open cleft stabilized by profilin may explain the stimulation of nucleotide exchange on actin, the short-lived state when actin releases its nucleotide may be characterized by an even more open cleft, probably analogous to that of the wide-open structure of profilin–β-actin (36) (Fig. S4*B*). Finally, in the actin filament, the nucleotide cleft appears to be more closed than in any of the existing structures of the monomer (17), suggesting that addition of profilin–actin at the barbed end of growing filaments results in closure of the nucleotide cleft, which might then lower the affinity for profilin and stimulate its rapid release.

The conformational changes associated with Tyr-53 phosphorylation and profilin binding described here are less extensive than previously anticipated, but are associated with long-range allosteric effects throughout the actin molecule. Large motions such as the acto-myosin power stroke are visually arresting, but conformational changes in proteins are usually less dramatic and result from changes in the equilibrium between different states in a landscape of nearly isoenergetic conformations (42).

Materials and Methods

Preparation of Proteins and Peptide. Phosphorylated and unphosphorylated *Dictyostelium* actin were purified as described (11). Human profilin-I and the VASP peptide 198–213 (corresponding to the last polyPro region of human VASP) were prepared as described (21). The cDNA encoding for human gelsolin was purchased from ATCC (catalog number MGC-39262). Gelsolin segment 1 or G1 (residues Met-52-Phe-176) was amplified by PCR and cloned between the NdeI and XhoI sites of vector pTYB12 (New England BioLabs). This vector comprises a chitin-binding domain (for affinity purification) and an intein (for self-cleavage after purification). BL21(DE3) competent cells (Invitrogen) were transformed with this constructs and grown in LB medium at 37°C until the OD_{600} reached a value of 0.5. Expression was induced by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and carried out overnight at 20°C. Cells were harvested by centrifugation, resuspended in chitin-affinity-column equilibration buffer (20 mM Tris·HCl, pH 7.5; 0.5 M NaCl; 0.1% Triton X-100; 1 mM EDTA) and lysed using a microfluidizer apparatus (MicroFluidics). Affinity purification on the chitin column was done according to the manufacturer's protocol (New England Biolabs). G1 was eluted from the column after self-cleavage of the intein, which was induced with 50 mM DTT for 2 days at 4°C. The protein was then dialyzed against 10 mM Tris·HCl, pH 8.0; 40 mM NaCl; 0.2 mM EDTA; and 1 mM NaN_3 and further purified on a superose12 column (Amersham Pharmacia).

Crystallization. Phosphorylated and unphosphorylated actin at ≈20 μM concentration in G-buffer (2 mM Tris, pH 7.5; 0.2 mM $CaCl_2$; 0.2 mM ATP; 1 mM NaN_3) were mixed with profilin and G1 at a 1:1.2 molar ratio. The complexes were purified on a superose12 column preequilibrated with 2 mM Tris·HCl, pH 8.0; 2 mM ATP; 4 mM $MgCl_2$; 0.5 mM EDTA; and 1 mM NaN_3. The profilin–actin and G1–actin complexes were concentrated to ≈5 mg/ml and ≈13 mg/ml, respectively by using Vivaspin centrifugal devices (Sartorius). Crystals of G1–actin were obtained under similar conditions for the unphosphorylated and phosphorylated forms: 100 mM Hepes, pH 7.5; 1.7 M Li_2SO_4; 2 mM ATP; 1 mM EDTA; and 10% glycerol at 20°C in 4-μl hanging drops. Except for the use of glycerol in this crystallization, these conditions are similar to those published before (30). The crystals were flash-frozen in liquid nitrogen by using 20% glycerol as a cryoprotectant and stored for data collection. Crystals of the profilin–actin complex were obtained only for the unphosphorylated form. This complex was crystallized with a VASP polyproline peptide ($VASP_{198-213}$), which binds profilin and facilitates crystallization. The crystallization conditions were similar to those described by us for profilin–α-actin (21): 0.1 M Bis-Tris, pH 6.5; 25% (wt/vol) PEG 3350. The crystals were flash-frozen in liquid nitrogen for data collection with Paratone-N as a cryoprotectant.

Data Collection and Structure Determination. An x-ray dataset was collected to the resolution of 2.3 Å from a profilin–actin crystal by using the beamline F2 of the Cornell High Energy Synchrotron Source (Ithaca, NY). Datasets were collected from crystals of G1 complexes with unphosphorylated and phosphorylated actin to the resolutions of 1.6 Å and 1.7 Å, respectively, using the 17-BM beamline of the IMCA-CAT facility at the Advance Photon Source (Argonne, IL). All of the datasets were indexed and scaled with the program HKL2000 (HKL Research). The structures were determined by molecular replacement using the CCP4 (43) program Phaser and the structures of profilin–α-actin (2PAV) or G1-actin (1NM1) as search models. Model building and refinement were performed with the CCP4 programs Coot and Refmac (Table 1).

Subtilisin Cleavage of the D-Loop. Actin, 13 μM in G-buffer, was cleaved with subtilisin at a molar ratio of actin:subtilisin of 8,000:1. Similar results were obtained at a ratio of 5,000:1. The percentage digestion for each reaction was determined as the ratio of the intensities of the digested fraction (lower band, ≈38,000 Da, on SDS PAGE) to the sum of the intensities of the digested and undigested actin fractions.

Etheno-ATP Exchange. The fluorescence of etheno-ATP increases when it replaces the ATP bound to actin. ATP-exchange experiments were carried out in 4 mM Tris, pH 7.5; 1 mM DTT; 0.1 mM $CaCl_2$; 0.01% NaN_3; and 10 μM ATP, with addition of 50 μM etheno-ATP at time 0. The experiments were performed with 1 μM actin in the presence of either 0.5 μM profilin, which acts catalytically (19), or 1.2 μM gelsolin.

Quenching of Tryptophan Fluorescence. Quenching of tryptophan fluorescence by profilin was done with 0.6 μM actin in 4 mM Tris pH 7.5, 1 mM DTT, 0.1 mM $CaCl_2$, 0.01% NaN_3 and 0.2 mM ATP. Kaleida graph software was used to fit the data and determine K_d values.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grant GM073791 and the Division of Intramural Research, National Heart, Lung, and Blood Institute. Data collection at the Cornell High Energy Synchrotron Source and Macromolecular Diffraction facilities was supported by National Science Foundation Grant DMR-0225180 and National Institutes of Health Grant RR-01646. Use of IMCA-CAT beamline 17-BM was supported by the Industrial Macromolecular Crystallography Association through a contract with the University of Chicago. The Advanced Photon Source was supported by Department of Energy Contract W-31-109-Eng-38.

1. Pollard TD, Borisy GG (2003) Cellular motility driven by assembly and disassembly of actin filaments. *Cell* 112:453–465.
2. Engqvist-Goldstein AE, Drubin DG (2003) Actin assembly and endocytosis: From yeast to mammals. *Annu Rev Cell Dev Biol* 19:287–332.
3. van Delft S, Verkleij AJ, Boonstra J, van Bergen en Henegouwen PM (1995) Epidermal growth factor induces serine phosphorylation of actin. *FEBS Lett* 357:251–254.
4. Rush J, *et al.* (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* 23:94–101.
5. Kameyama K, *et al.* (2000) Tyrosine phosphorylation in plant bending. *Nature* 407:37.
6. Schweiger A, Mihalache O, Ecke M, Gerisch G (1992) Stage-specific tyrosine phosphorylation of actin in *Dictyostelium discoideum* cells. *J Cell Sci* 102:601–609.
7. Howard PK, Sefton BM, Firtel RA (1993) Tyrosine phosphorylation of actin in *Dictyostelium* associated with cell-shape changes. *Science* 259:241–244.
8. Jungbluth A, *et al.* (1995) Stress-induced tyrosine phosphorylation of actin in *Dictyostelium* cells and localization of the phosphorylation site to tyrosine-53 adjacent to the DNase I binding loop. *FEBS Lett* 375:87–90.
9. Gauthier ML, Lydan MA, O'Day D, Cotter AD (1997) Endogenous autoinhibitors regulate changes in actin tyrosine phosphorylation during *Dictyostelium* spore germination. *Cell Signal* 9:79–83.
10. Kishi Y, Clements C, Mahadeo DC, Cotter DA, Sameshima M (1998) High levels of actin tyrosine phosphorylation: Correlation with the dormant state of *Dictyostelium* spores. *J Cell Sci* 111:2923–2932.
11. Liu X, Shu S, Hong MS, Levine RL, Korn ED (2006) Phosphorylation of actin Tyr-53 inhibits filament nucleation and elongation and destabilizes filaments. *Proc Natl Acad Sci USA* 103:13694–13699.
12. Williams HP, Harwood AJ (2003) Cell polarity and *Dictyostelium* development. *Curr Opin Microbiol* 6:621–627.
13. Jungbluth A, *et al.* (1994) Strong increase in the tyrosine phosphorylation of actin upon inhibition of oxidative phosphorylation: correlation with reversible rearrangements in the actin skeleton of *Dictyostelium* cells. *J Cell Sci* 107:117–125.
14. Holmes KC, Popp D, Gebhard W, Kabsch W (1990) Atomic model of the actin filament. *Nature* 347:44–49.
15. Hegyi G, *et al.* (1998) Intrastrand cross-linked actin between Gln-41 and Cys-374. I. Mapping of sites cross-linked in F-actin by N-(4-azido-2-nitrophenyl) putrescine. *Biochemistry* 37:17784–17792.
16. Khaitlina SY, Strzelecka-Golaszewska H (2002) Role of the DNase-I-binding loop in dynamic properties of actin filament. *Biophys J* 82:321–334.
17. Oda T, Stegmann H, Schroder RR, Namba K, Maeda Y (2007) Modeling of the F-actin structure. *Adv Exp Med Biol* 592:385–401.
18. Mockrin SC, Korn ED (1980) Acanthamoeba profilin interacts with G-actin to increase the rate of exchange of actin-bound adenosine 5′-triphosphate. *Biochemistry* 19:5359–5362.
19. Goldschmidt-Clermont PJ, Machesky LM, Doberstein SK, Pollard TD (1991) Mechanism of the interaction of human platelet profilin with actin. *J Cell Biol* 113:1081–1089.
20. Schutt CE, Myslik JC, Rozycki MD, Goonesekere NC, Lindberg U (1993) The structure of crystalline profilin-beta-actin. *Nature* 365:810–816.
21. Ferron F, Rebowski G, Lee SH, Dominguez R (2007) Structural basis for the recruitment of profilin–actin complexes during filament elongation by Ena/VASP. *EMBO J* 26:4597–4606.
22. Kabsch W, Mannherz HG, Suck D, Pai EF, Holmes KC (1990) Atomic structure of the actin:DNase I complex. *Nature* 347:37–44.
23. McLaughlin PJ, Gooch JT, Mannherz HG, Weeds AG (1993) Structure of gelsolin segment 1-actin complex and the mechanism of filament severing. *Nature* 364:685–692.
24. Otterbein LR, Cosio C, Graceffa P, Dominguez R (2002) Crystal structures of the vitamin D-binding protein and its complex with actin: Structural basis of the actin-scavenger system. *Proc Natl Acad Sci USA* 99:8003–8008.
25. Chereau D, *et al.* (2005) Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. *Proc Natl Acad Sci USA* 102:16644–16649.
26. Lee SH, Hayes DB, Rebowski G, Tardieux I, Dominguez R (2007) Toxofilin from *Toxoplasma gondii* forms a ternary complex with an antiparallel actin dimer. *Proc Natl Acad Sci USA* 104:16122–16127.
27. Klenchin VA, *et al.* (2003) Trisoxazole macrolide toxins mimic the binding of actin-capping proteins to actin. *Nat Struct Biol* 10:1058–1063.
28. Rould MA, Wan Q, Joel PB, Lowey S, Trybus KM (2006) Crystal structures of expressed non-polymerizable monomeric actin in the ADP and ATP states. *J Biol Chem* 281:31909–31919.
29. Otterbein LR, Graceffa P, Dominguez R (2001) The crystal structure of uncomplexed actin in the ADP state. *Science* 293:708–711.
30. Vorobiev S, *et al.* (2003) The structure of nonvertebrate actin: Implications for the ATP hydrolytic mechanism. *Proc Natl Acad Sci USA* 100:5760–5765.
31. Burtnick LD, Urosev D, Irobi E, Narayan K, Robinson RC (2004) Structure of the N-terminal half of gelsolin bound to actin: Roles in severing, apoptosis and FAF. *EMBO J* 23:2713–2722.
32. Strzelecka-Golaszewska H, Moraczewska J, Khaitlina SY, Mossakowska M (1993) Localization of the tightly bound divalent-cation-dependent and nucleotide-dependent conformation changes in G-actin using limited proteolytic digestion. *Eur J Biochem* 211:731–742.
33. Strzelecka-Golaszewska H, Wozniak A, Hult T, Lindberg U (1996) Effects of the type of divalent cation, Ca2+ or Mg2+, bound at the high-affinity site and of the ionic composition of the solution on the structure of F-actin. *Biochem J* 316:713–721.
34. Muhlrad A, *et al.* (2004) Cofilin induced conformational changes in F-actin expose subdomain 2 to proteolysis. *J Mol Biol* 342:1559–1567.
35. Bryan J (1988) Gelsolin has three actin-binding sites. *J Cell Biol* 106:1553–1562.
36. Chik JK, Lindberg U, Schutt CE (1996) The structure of an open state of beta-actin at 2.65 A resolution. *J Mol Biol* 263:607–623.
37. Schuler H (2001) ATPase activity and conformational changes in the regulation of actin. *Biochim Biophys Acta* 1549:137–147.
38. Minehardt TJ, Kollman PA, Cooke R, Pate E (2006) The open nucleotide pocket of the profilin/actin x-ray structure is unstable and closes in the absence of profilin. *Biophys J* 90:2445–2449.
39. Tirion MM, ben-Avraham D (1993) Normal mode analysis of G-actin. *J Mol Biol* 230:186–195.
40. Perelroizen I, Marchand JB, Blanchoin L, Didry D, Carlier MF (1994) Interaction of profilin with G-actin and poly(L-proline). *Biochemistry* 33:8472–8478.
41. Egelman EH (2001) Actin allostery again? *Nat Struct Biol* 8:735–736.
42. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* 254:1598–1603.
43. CCP4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
44. Robinson RC, *et al.* (2001) Crystal structure of Arp2/3 complex. *Science* 294:1679–1684.

JOURNAL OF VIROLOGY, July 2009, p. 6534–6545 Vol. 83, No. 13
0022-538X/09/$08.00+0 doi:10.1128/JVI.00189-09

The Crystal Structures of Chikungunya and Venezuelan Equine Encephalitis Virus nsP3 Macro Domains Define a Conserved Adenosine Binding Pocket[▽]

Hélène Malet,[1]† Bruno Coutard,[1] Saïd Jamal,[1] Hélène Dutartre,[1]‡ Nicolas Papageorgiou,[1] Maarit Neuvonen,[2] Tero Ahola,[2] Naomi Forrester,[3]§ Ernest A. Gould,[3] Daniel Lafitte,[4] Francois Ferron,[1] Julien Lescar,[1] Alexander E. Gorbalenya,[5] Xavier de Lamballerie,[6] and Bruno Canard[1]*

Architecture et Fonction des Macromolécules Biologiques, CNRS and Universités d'Aix-Marseille I et II, UMR 6098, ESIL Case 925, 13288 Marseille, France[1]; Institute of Biotechnology, University of Helsinki, P.O. Box 56, Viikinkaari 9, 00014 Helsinki, Finland[2]; CEH Oxford, Mansfield Road, Oxford OX1 3SR, United Kingdom[3]; Marseille Protéomique, INSERM UMR 911 CRO2, Aix-Marseille Université, Faculté de Pharmacie, 27 Bd. Jean Moulin, 13285 Marseille cedex 05, France[4]; Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands[5]; and UMR190, Emergence des Pathologies Virales, Institut de Recherche pour le Développement—Université de la Méditerranée, Faculté de Médecine de Marseille, 27 Bd. Jean Moulin, 13005 Marseille cedex 05, France[6]

Received 27 January 2009/Accepted 14 April 2009

Macro domains (also called "X domains") constitute a protein module family present in all kingdoms of life, including viruses of the *Coronaviridae* and *Togaviridae* families. Crystal structures of the macro domain from the Chikungunya virus (an "Old World" alphavirus) and the Venezuelan equine encephalitis virus (a "New World" alphavirus) were determined at resolutions of 1.65 and 2.30 Å, respectively. These domains are active as adenosine di-phosphoribose 1″-phosphate phosphatases. Both the Chikungunya and the Venezuelan equine encephalitis virus macro domains are ADP-ribose binding modules, as revealed by structural and functional analysis. A single aspartic acid conserved through all macro domains is responsible for the specific binding of the adenine base. Sequence-unspecific binding to long, negatively charged polymers such as poly(ADP-ribose), DNA, and RNA is observed and attributed to positively charged patches outside of the active site pocket, as judged by mutagenesis and binding studies. The crystal structure of the Chikungunya virus macro domain with an RNA trimer shows a binding mode utilizing the same adenine-binding pocket as ADP-ribose, but avoiding the ADP-ribose 1″-phosphate phosphatase active site. This leaves the AMP binding site as the sole common feature in all macro domains.

The *Togaviridae* virus family comprises two positive-sense RNA virus genera, viz., *Alphavirus* and *Rubivirus* (52). The genus *Alphavirus* contains at least 28 viruses and has a worldwide distribution, even though each virus has a local distribution. Some alphaviruses are not known to cause illness, but others can cause severe disease in higher eukaryotes, in particular, humans and horses. Alphaviruses present in the "Old World" principally cause arthritis and skin rashes, whereas alphaviruses of the "New World" may cause severe and even fatal encephalitis. Most have vertebrate hosts, principally mammals, birds, or fish, and are transmitted by mosquitoes, although other hematophagous arthropods such as lice or mites can be vectors of some alphaviruses (37).

Although known for decades as a virus that produces sporadic human outbreaks in Africa, India, Southeast Asia, and the Philippines, Chikungunya virus (CHIKV) unexpectedly emerged as a major arbovirus pathogen in many islands of the Indian Ocean in 2005, and, more recently (2007) in Singapore and Australia. The virus was also introduced into northern Italy in 2007, where it established localized outbreaks. In urban areas, the virus is mainly transmitted by the mosquito *Aedes aegypti* but an adaptive mutation facilitated transmission by another urban mosquito, *Aedes albopictus*, in La Réunion island in 2005 as well as other places (7). Epidemics are sporadic, but a large part of the population can be infected, as was the case in La Réunion island in 2006, where one-third of the population (~250,000 people) was infected, leading to 237 deaths. CHIKV symptoms in humans include fever, rash, and severe arthritis, which usually disappear after 1 week, but, in some cases, arthritis can persist for more than 1 year (18). Venezuelan equine encephalitis virus (VEEV) is a New World alphavirus, present in the United States and Central and South America. Different subtypes of VEEV exist, some of which cause epidemics while others are zoonotic. In a few cases, VEEV causes encephalitis in humans, particularly children (40). This virus is also pathogenic for horses, with an observed case fatality rate of around 80 to 90%. No vaccine or drugs are licensed for treatment against alphaviruses.

* Corresponding author. Mailing address: Architecture et Fonction des Macromolécules Biologiques, CNRS and Universités d'Aix-Marseille I et II, UMR 6098, ESIL Case 925, 13288 Marseille, France. Phone: 33 491 82 86 44. Fax: 33 491 82 86 46. E-mail: bruno.canard@afmb.univ-mrs.fr.

† Present address: School of Crystallography, Birkbeck College, University of London, Malet St., London WC1E 7HX, United Kingdom.

‡ Present address: Baylor Institute for Immunological Research, INSERM U899, 3434 Live Oak St., Dallas, TX 75204.

§ Present address: Department of Pathology, University of Texas Medical Branch, Galveston, TX 77551.

▽ Published ahead of print on 22 April 2009.

The 5′ region of the positive-stranded RNA [(+) RNA] genome encodes four nonstructural proteins (nsP1, nsP2, nsP3, nsP4), and the 3′ region encodes the three major structural proteins (the capsid and two envelope proteins). The nsP1 protein is a membrane-associated protein that bears methyltransferase and guanylyltransferase activities. It is involved in the capping of the (+) RNA genome (1). The nsP2 protein is made of three domains, the first containing helicase, RNA triphosphatase, and nucleoside triphosphatase activities (39, 53), whereas the second and third domains are a papaine-like protease and a nonfunctional methyltransferase, respectively (47). Moreover, nsP2 contains a nuclear localization sequence which allows 50% of the translated nsP2 to be translocated into the nucleus (35). The nsP4 protein contains the RNA-dependent RNA polymerase, involved in genome replication and transcription (42). The functions, roles, and activities of the nsP3 protein are less well understood. Although it is involved in the transcription process at an early stage of the infection (51), no precise function or activity has been attributed to this protein. It is made of two domains, the first one being a unique macro domain (described below) located in the conserved N-terminal region. The C-terminal region is less conserved and is phosphorylated in up to 16 positions on serines and threonines (22, 25, 50). The role of phosphorylation is not well documented, but deletion of the phosphorylated residues decreases the RNA synthesis level (49). Moreover, the absence of phosphorylation on nsP3 in variants of the alphavirus Semliki Forest virus (SFV) decreases viral pathogenicity, and the absence of the C terminus of nsP3 alters SFV neurovirulence (48). The C terminus of nsP3 is thus thought to have a nonessential regulatory role.

The first 160 residues of the N-terminal region of nsP3, known as the macro domain, have been initially identified based on sequence similarity analysis between alphavirus, rubivirus, and coronavirus (15). The domain was named the “X domain,” referring to a domain with an unknown function conserved in these viruses. Subsequent sequence analysis revealed that this domain is remarkably conserved in all kingdoms of life (2): 1,081 domains showing similarity to X domains are currently indexed in the SMART database (24). They may represent a protein or a single domain in a larger protein, and the domain can also exist in duplicate or triplicate in the same protein (21). In particular, this domain is present in a variant of histone H2A. This variant is called macroH2A, and its difference from the conventional histone H2A is the presence of an additional domain called “macro,” which shows similarity to X domains. Consequently, all X domains have also been called macro domains.

In viruses, macro domains exist in alphaviruses and in viruses related to the genus *Alphavirus*, such as rubella virus (genus *Rubivirus*) and hepatitis E virus (HEV) (genus *Hepevirus*). *Coronavirus* and *Torovirus*, which belong to the *Coronaviridae* family, are the only other viral genera containing a macro domain, located in their large multifunctional nsp3 protein, which is otherwise unrelated to alphavirus nsP3.

The crystal structures of several macro domains have been determined in archaebacteria (*Archaeoglobus fulgidus*) (2), eubacteria (*Escherichia coli* and *Thermus thermophilus*), and eukaryotes (*Saccharomyces cerevisiae*, *Rattus norvegicus*, and *Homo sapiens*) (4, 19, 20). The crystal structure of the macro domain of the coronavirus responsible for severe acute respiratory syndrome (SARS-CoV) has also been determined (10, 27, 43). The structural conservation between these structures is remarkable and they have been defined as a family in the SCOP database (30). This structural conservation suggests an important role in the biology of their host organism, but this role remains elusive so far.

An enzymatic activity was first discovered for the yeast macro domain, which acts as an ADP-ribose 1″-phosphate phosphatase with a somewhat low turnover constant of 1.7 min^{-1} (28). This activity is involved in the downstream processing of ADP-ribose 1″-phosphate, a side product of cellular pre-tRNA splicing, thus controlling the metabolism of ADP-ribose 1″-phosphate or other ADP-ribose derivatives with known regulatory functions in the cell. This activity has also been reported for *A. fulgidus* (17), SARS-CoV (10, 43), and human CoV (HCoV) (38), but was at the limit of detection for the alphavirus SFV (10). The affinity of ADP-ribose for the SARS-CoV enzyme was low (dissociation constant [K_d] of 52.7 μM) (43), and the pathway in which ADP-ribose 1″-phosphate phosphatase is putatively involved remains elusive. A recent study showed that ADP-ribose binding is not a common property to all macro domains. Indeed, the group 3 CoV macro domain does not bind ADP-ribose, whereas a group 1 CoV macro domain binds it with a dissociation constant K_d of 29 μM (36).

In the case of *A. fulgidus*, poly(ADP-ribose) (PAR) binding has been reported (17). Such a binding property has also been identified in CoVs (10), showing that some macro domains, in particular viral macro domains, are able to bind long, negatively charged polymers. The role of PAR binding could be related to a cellular response to viral infection. Indeed, in response to inflammation or stress, the nuclear enzyme PAR polymerase-1 (PARP-1) promotes PAR synthesis (5). Alphavirus infection can induce PARP-1 activation (31) which leads to the depletion of ATP and NAD$^+$ present in the cell and apoptosis-inducing factor (AIF)-induced apoptosis.

In this article, we present the biochemical, enzymatic, and structural analysis of two alphavirus macro domains, one from CHIKV, a representative of the Old World alphaviruses, the other from VEEV, a New World alphavirus present in the Americas. The crystal structures of the CHIKV and VEEV macro domains in complex with ADP-ribose show essentially the same positioning of the ligand as that seen in the SARS-CoV macro domain structure. The crystal structure of the CHIKV macro domain in complex with adenosine-containing short RNAs shows that the adenosine binding site is generally conserved in macro domains. Moreover, binding of the short RNA is enhanced by the presence of positively charged patches present at the surface of the protein. The latter patches, located in the immediate vicinity of the ADP-ribose binding crevice, are significantly more charged than that of the SARS-CoV macro domain.

MATERIALS AND METHODS

Oligonucleotides. RNA (AAAAAAAAAGCUACC, AAA, UUU, GGGGGG, UCGGGGGCUGGC) and DNA (AAAGCCAAAAA) oligonucleotides were purchased from Dharmacon.

Expression and purification of alphavirus macro domains. The cDNAs corresponding to two alphavirus macro domains from the nsP3 protein of CHIKV (strain Ross, amino acids [aa] 1 to 160 of nsP3) and VEEV (strain P676, aa 1 to 160) were cloned into the pDest14 plasmid using the “Gateway” cloning procedure (Invitrogen). A hexa-histidine sequence tag was fused at either the N-

TABLE 1. Crystallization analysis data for CHIKV and VEEV

Parameter	Result for[d]:					
	CHIKV				VEEV	
	Native	Selenomethionylated protein	Soaked with ADP-ribose	Soaked with RNA	Native	Cocrystallized with ADP-ribose
Beamline	ESRF ID14-EH2	ESRF ID23-1	ESRF ID23-1	ESRF ID14-EH2	ESRF ID23-1	SOLEIL PROXIMA1
Space group	$P3_1$	$P3_1$	$P3_1$	$P3_1$	I4	$P2_12_12_1$
Cell dimensions (Å)	a = b = 87.06, c = 84.49	a = b = 86.81, c = 84.70	a = b = 87.96, c = 84.17	a = b = 86.82, c = 81.32	a = b = 129.63, c = 42.49	a = 74.00, b = 87.0, c = 105.00
Wavelength (Å)	0.9330	0.9792	0.8856	0.9330	0.974	0.980
Resolution range (Å)	35.00–1.65 (1.74–1.65)	35.00–1.80 (1.90–1.80)	35.00–1.90 (2.00–1.90)	30.00–2.00 (2.11–2.00)	35.00–2.30 (2.42–2.30)	30.00–2.60 (2.70–2.60)
Total no. of reflections	493,231 (71,649)	2,000,862 (237,557)	310,123 (38,681)	213,703 (30,275)	108,812 (16,349)	109,136 (10,819)
No. of unique reflections	86,209 (12,610)	64,424 (8,446)	57,403 (8,387)	46,206 (6,793)	15,851 (2,317)	20,549 (2,014)
Completeness (%)	100 (100)	97.3 (87.3)	99.9 (99.9)	99.6 (100.0)	99.2 (100.0)	99.2 (89.2)
$I/\Sigma(I)$	18.1 (3.2)	37.4 (6.1)	15.7 (2.4)	17.4 (2.8)	9.0 (5.4)	15.74 (4.72)
R_{sym} (%)[a]	5.2 (46.8)	8.8 (50.2)	7.2 (55.4)	5.2 (52.5)	13.9 (26.4)	10.4 (49.7)
Multiplicity	5.7 (5.7)	31.1 (28.1)	5.4 (4.6)	4.6 (4.5)	6.9 (7.1)	5.3 (5.4)
Anomalous completeness (%)		97.1 (86.1)				
Anomalous multiplicity		15.4 (14.0)				
R (%)[b]	16.6		17.0	21.5	20.3	19.7
R_{free} (%)[c]	19.6		20.1	26.0	26.1	27.5
rmsd bond length (Å)	0.012		0.013	0.014	0.013	0.010
rmsd angle (°)	1.347		1.447	1.398	1.348	1.433
Protein Data Bank no.	3GPG		3GPO	3GPQ	3GQE	3GQO

[a] $R_{\mathrm{sym}} = \Sigma\,|I - \langle I\rangle|\,/\Sigma\,I$, where I is the observed intensity and $\langle I\rangle$ is the average intensity. Values in parentheses refer to the highest-resolution shell.
[b] $R = \Sigma\,\|\mathrm{Fo}| - |\mathrm{Fc}\|/\Sigma|\mathrm{Fo}|$.
[c] R_{free} is calculated as R, but on 5% of all reflections that are never used in crystallographic refinement.
[d] Values in parentheses refer to the highest-resolution shell. ESRF, European Synchrotron Radiation Facility.

terminal end (CHIKV) or the C-terminal-end (VEEV). An incomplete factorial expression screening in *E. coli* was performed for each construct in order to design the best expression conditions required for scale-up, as previously described (3). The proteins (wild type and mutants) were then produced under the following conditions: (i) with *E. coli* Rosetta (DE3)(pLysS) (Novagen) cells at 25°C in SB medium (Athena Enzymes) overnight after induction with 500 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) for the CHIKV macro domain and (ii) with *E. coli* Rosetta (DE3)(pLysS) cells at 37°C in 2YT medium for 4 h after induction with 500 μM IPTG for the VEEV macro domain. Cells were harvested by centrifugation at 2,800 × *g*. Cell pellets were then resuspended in 50 mM Tris buffer, 300 mM NaCl, 10 mM imidazole, 0.1% Triton, and 5% glycerol (pH 8.0). Lysozyme (0.25 mg/ml), phenylmethylsulfonyl fluoride (1 mM), DNase I (2 μg/ml), and EDTA-free protease cocktail (Roche) were added before performing a sonication step. The lysates were centrifuged at 12,000 × *g* for 45 min, and the supernatant was recovered. The recombinant proteins were purified using the Akta Xpress fast protein liquid chromatography system (GE Healthcare) as follows. The first purification step (immobilized metal affinity chromatography) was performed on a 5-ml His prep column (GE Healthcare). The proteins were eluted with 50 mM Tris buffer (pH 8.0) containing 300 mM NaCl and 500 mM imidazole. The purification was then refined by a size exclusion chromatography step on a preparative Superdex 200 column (GE Healthcare) preequilibrated in a buffer designed for its ability to keep the protein soluble and stable (14): 10 mM Tris (pH 7.5) plus 300 mM NaCl for the CHIKV macro domain and 10 mM Bicine (pH 8.5) plus 50 mM NaCl for the VEEV macro domain. Proteins were then concentrated up to 14.5 mg/ml and 14 mg/ml using a Vivaspin 10-kDa molecular-mass-cutoff centrifugal concentrator (Vivascience) for the CHIKV and VEEV macro domains, respectively. The Eastern equine encephalitis virus (EEEV) macro domain was obtained from EEEV cDNA by the same procedure described above and purified to homogeneity following the same protocol as that of the CHIKV macro domain. It was used in phosphatase assays (see Fig. 4). Its crystal structure will be reported elsewhere.

A seleno-methionine-substituted protein was used to determine the CHIKV macro domain structure. The protein was expressed according to standard conditions of methionine biosynthesis pathway inhibition (9) and purified following the same procedure as that of the native protein.

Crystallization. Initial crystallization trials were set up for CHIKV and VEEV macro domains with a nano-drop dispenser (Honeybee; Genomic Solutions) in 96-well sitting drop plates (Greiner Bio One) using three commercial crystallization kits: Structure Screen combination, Stura footprints (Molecular Dimensions Limited), and Nextal SM1 (Qiagen).

The initial crystallization conditions of the CHIKV protein were further optimized by the hanging drop vapor diffusion method in Linbro plates by mixing 3 μl of protein solution with 2 μl of reservoir solution. Crystals grew in a mixture of 46% polyethylene glycol 600 (PEG 600) and 100 mM HEPES (pH 7.4). For the VEEV macro domain, initial hits were optimized by mixing 100 nl of protein and 100 nl of reservoir solution using the sitting drop vapor diffusion method. Optimized crystallization conditions of VEEV macro domain were 18% PEG 3350 plus 11 mM sodium acetate. Diffraction intensities were recorded on different beamlines (Table 1) at the European Synchrotron Radiation Facility (Grenoble, France) and at SOLEIL Synchrotron (Gif-sur-Yvette, France). Integration of the different datasets was performed using MOSFLM (23) or XDS (16). Scaling and merging of the intensities were carried out using programs from the CCP4 suite (6) or XSCALE (16), depending on the data sets. Statistics are provided in Table 1.

Structure determination. The structure of the CHIKV macro domain was determined using the single-wavelength anomalous dispersion method (SAD) on a 1.80-Å data set collected at the peak of the selenium absorption edge from a seleno-methionine derivative crystal. Location of all the 16 selenium atoms (4 selenium atoms in each of the 4 molecules of the asymmetric unit) was performed using SHELXD (44). Phases and figures of merit were calculated using SHELXE (45). The excellent quality of the map allowed the program ARPWARP to build residues 2 to 160 for molecules A and B and to partially build molecules C and D. Refinement was performed against a 1.65-Å resolution data set using REFMAC (6), Buster-TNT (41), and COOT (11). Refinement statistics are listed in Table 1. The VEEV macro domain was determined by molecular replacement with the program PHASER (29) using the CHIKV macro domain as a model. Crystals diffracted to a 2.30-Å resolution for the VEEV protein. Refinement was performed using REFMAC, Buster-TNT, and COOT.

Crystal soaking and cocrystallization experiments. CHIKV and VEEV macro domain crystals were soaked for 16 h with ADP-ribose (4 mM) (Sigma). The

CHIKV macro domain was soaked for 16 h with PAR (1.3 mM in the ADP-ribose monomer) (Biomol International LP.), short RNA oligonucleotides AAA (2 to 10 mM) (Dharmacon) and UUUUUU (2 to 10 mM) (Dharmacon), and the DNA oligonucleotide AAAGCCAAAAA (10 mM) (Dharmacon). Crystals were transferred in drops containing the mother liquor solution and the ligand. Co-crystallization of the VEEV macro domain with ADP-ribose (5 mM) was performed in 0.2 M ammonium sulfate, 0.1 M sodium cacodylate (pH 6.5), and 30% PEG 8000.

Site-directed mutagenesis. Based on the CHIKV macro domain structure, four point mutations were made using Stratagene's Quickchange site-directed mutagenesis. Single-amino-acid substitutions to alanine were performed in the CHIKV macro domain for D10, N24, H67, and Y114.

Thermal denaturation shift assay. Several potential ligands were tested at 2 mM each (ADP-ribose, ADP-glucose, ATP, ADP, AMP, *S*-adenosyl-homocysteine, manganese chloride [$MnCl_2$], magnesium chloride [$MgCl_2$], NAD^+, and GDP using the quantitative PCR machine ICycler IQ (Bio-Rad) according to the previously described thermal shift assay (12). Briefly, the CHIKV macro domain (final concentration of 3 mg/ml in the size exclusion chromatography buffer described above) was mixed with the ligand and a SYPRO orange solution in concentrations recommended by the manufacturer (Invitrogen) in a final volume of 25 μl. Incremental steps of temperature from 20 to 90°C were applied to the samples. The denaturation of the protein was monitored by following the increase of the fluorescence emitted by the probe that binds the exposed hydrophobic regions of the denatured protein. A melting temperature (T_m) can be calculated as the mid-log of the transition phase from the native to the denatured protein using a Boltzmann model. ADP-ribose titration from 0 to 2.5 mM was then assessed on the CHIKV (wild type and D10A single mutant), VEEV and SARS-CoV macro domains at 1 mg/ml in 10 mM HEPES plus 150 mM NaCl (pH 7.5).

ITC. For isothermal titration calorimetry (ITC), ADP-ribose binding to CHIKV and VEEV macro domains was measured using a Microcal MCS isothermal titration calorimeter. Experiments were carried out at 25°C in a buffer containing 10 mM HEPES plus 150 mM NaCl (pH 7.5). The protein concentration in the cell was 20 μM, whereas the ADP-ribose concentration in the syringe was 300 μM. Heats of dilution were measured by injecting the ligand into the protein solution. The recorded heat curve was subtracted from the experimental curves prior to data analysis. Titration curves were fitted using MicroCal Origin software, assuming one set of sites and enthalpy changes (ΔH), equilibrium constants (K_d), and stoichiometry were extracted.

PAR binding assay. The PAR (1‴-2″ branched polymer of ADP-ribose linked by 1″-2′ glycosidic bonds), was synthesized by auto-poly-ADP-ribosylation of PARP-1 in a reaction volume of 400 μl using 4 U of human PARP-1 (Sigma), 150 μM NAD^+, and 40 μCi of [^{32}P]NAD^+ (GE Healthcare). After 2 h of incubation at 24°C, the reaction was stopped by dilution (30-fold) with the slot blot buffer (10 mM Tris [pH 7.5], 300 mM NaCl, 0.05% Tween 20). PAR binding on the macro domains of several alphaviruses (CHIKV, VEEV, Sindbis virus, and SFV) as well as SARS-CoV was tested. Various amounts of each protein (2,000, 1,000, 500, 250, 125, 62.5, 31.25, 15.63, 7.813, 3.906, 1.953, and 0.9766 pmol) were blotted on a nitrocellulose membrane (Schleicher & Schuell) using a slot blot apparatus (Bio-Rad). Bovine serum albumin (BSA) was also included as a negative control. The membrane was then incubated for 1 h in the PAR preparation and then washed extensively in slot blot buffer (five times in 100 ml), and the membrane-bound fraction of PAR was analyzed using photostimulated plates and an FLA3000 fluorescent image analyzer (Fuji).

RNA binding assay. Fifty microliters of RNA (AAAAAAAAAGCUACC; 100 μM) was labeled using 25 U of T4 polynucleotide kinase (New England Biolabs), 100 μM ATP, and 50 μCi [γ-^{32}P]ATP (GE Healthcare). After 30 min at 37°C, the reaction was stopped by incubating the mixture for 10 min at 70°C. RNA binding to the CHIKV macro domain native and mutants was tested as follows: 2,000, 1000, 500, 250, 125, 62.5, 31.25, 15.63, 7.813, 3.906, 1.953, and 0.9766 pmol of protein were blotted onto a nitrocellulose membrane (Schleicher & Schuell) using a slot blot apparatus (Bio-Rad). BSA and hepatitis C virus RNA-dependent RNA polymerase (HCVpol) were also included as negative and positive controls, respectively. The radiolabeled RNA was diluted (300×) in blotting buffer (10 mM Tris [pH 7.5], 150 mM NaCl, 0.05% Tween 20). The membrane supporting the blotted proteins was incubated for 1 h in this preparation at room temperature. The membrane was washed five times in 100 ml of blotting buffer, and the membrane-bound fraction of RNA was analyzed using photostimulated plates and an FLA3000 fluorescent image analyzer (Fuji).

ADP-ribose 1″-phosphate phosphatase assay. ADP-ribose-1″-phosphate phosphatase activity was detected with thin-layer chromatography (TLC) assays for the CHIKV (wild type and mutants), EEEV, and VEEV macro domains. The yeast protein Poa1p was used as a positive control. First, ADP-ribose 1″-phosphate was produced by incubating 7 mM ADP-ribose 1″-2″-cyclic phosphate with cyclic phosphodiesterase (15 ng/ml) for 3 h at 28°C. Then, 1 μl of the reaction was mixed with 1 μl of 45 μM of the macro domain tested and 1 μM of 60 mM MES (morpholineethanesulfonic acid) (pH 5.0). The reaction mixture was incubated for 1 h at 28°C. Two microliters of the reaction mixture was then spotted onto a polyethyleneimine-F (PEI-F) cellulose TLC plate, and the plate was developed at room temperature in 150 mM NaCl and 150 mM sodium formate (pH 3.0). The spots were detected from the fluorescent background under a UV lamp emitting at a 254-nm wavelength (10).

Protein structure accession number. The coordinates of CHIKV macro domain structure, apo form, in complex with ADP-ribose and with RNA AAA have been deposited in the Protein Data Bank under accession no. 3GPG, 3GPO, and 3GPQ, respectively. Those of the VEEV macro domain structure apo form and in complex with ADP-ribose have been deposited in the Protein Data Bank under accession no. 3GQE and 3GQO.

RESULTS

Protein production and crystallization. A cDNA construct encompassing aa 1 to 160 of the alphavirus nsP3 macro domain was designed based on available macro domain crystal structures (2, 4, 19, 20, 43), secondary structures, and disorder predictions. CHIKV and VEEV nsP3 macro domains were selected in order to compare macro domains from New and Old World alphaviruses. These constructs were expressed in *E. coli* and purified, and crystals were obtained for these two proteins. Molecular replacement techniques using the closest structural homologue (*E. coli* macro domain, identity of 29.9% in 1 to 160 aa for the CHIKV macro domain) were unsuccessful. The crystal structure of the CHIKV macro domain was determined using the SAD technique in conjunction with seleno-methionylated protein crystals that diffracted to 1.80 Å. It was then refined against a native data set at a 1.65-Å resolution. The CHIKV crystals belong to the space group $P3_1$, and four molecules are present in the asymmetric unit. VEEV macro domain structure in an apo form was then determined at a 2.30-Å resolution using molecular replacement and the CHIKV macro domain as a template. The VEEV crystals of the apo form belong to the space group I4, with two molecules in the asymmetric unit.

Overall structure of alphavirus macro domains and comparison with other macro domains. The structures of the alphavirus macro domain consist of a central twisted six-stranded β sheet surrounded by three helices on one side and one on the other (Fig. 1A). The core β sheet is well conserved within the existing structures of macro domains from all origins also. Positions of α helices are also well conserved, even if a deletion is present in alphavirus macro domains near residue 48 (in CHIKV macro domain; the numbering of CHIKV will be used throughout), leading thus to the absence of one helix present in other macro domains (Fig. 1E, α3 in SARS-CoV). Moreover, the α helix 2 (aa 78 to 100) is longer in alphavirus macro domains than in other macro domains (Fig. 1B and E). The loops connecting the secondary structure elements are variable in sequence and structure when compared to non-alphavirus macro domain structures available, in particular the loops between the β3-β4, β4-α2, and β5-α3 elements. The root mean square deviations (rmsds) between the CHIKV macro domain and other reported macro domain structures are comprised between 2.01 Å and 2.97 Å on 160 aa (sequence identity between 21.9 and 30.0%). Remarkably, the closest structure is the *E. coli* macro domain and not the SARS-CoV macro do-

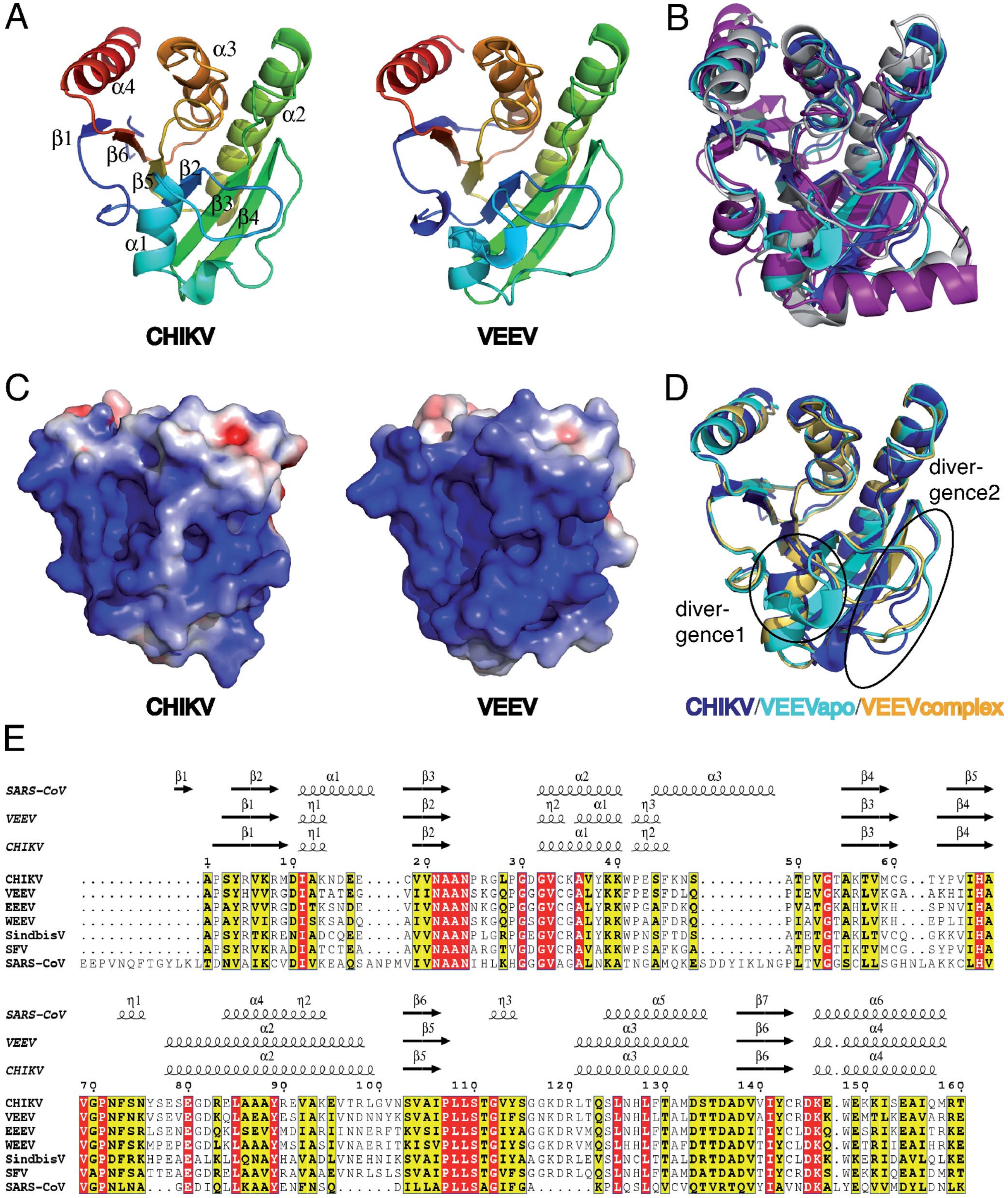

A
α3
α4
α2
β1
β6
β5
β2
β3
β4
α1
CHIKV
VEEV
B
C
CHIKV
VEEV
D
diver-
gence2
diver-
gence1
CHIKV/VEEVapo/VEEVcomplex
E
SARS-CoV
VEEV
CHIKV
CHIKV .............APSYRVKRMDIAKNDEE....CVVNAANPRGLPGDGVCKAVYKKWPESFKNS.........ATPVGTAKTVMCG..TYPVIHA
VEEV .............APSYHVVRGDIATATEG....VIINAANSKGQPGGGVCGALYKKFPESFDLQ.........PIEVGKARLVKGA..AKHIIHA
EEEV .............APAYRVVRGDITKSNDE....VIVNAANNKGQPGGGVCGALYRKWPGAFDKQ.........PVATGKAHLVKH...SPNVIHA
WEEV .............APAYRVIRGDISKSADQ....AIVNAANSKGQPGSGVCGALYRKWPAAFDRQ.........PIAVGTARLVKH...EPLIIHA
SindbisV .............APSYRTKRENIADCQEE....AVVNAANPLGRPGEGVCRAIYKRWPNSFTDS.........ATETGTAKLTVCQ..GKKVIHA
SFV .............APSYRVKRADIATCTEA....AVVNAANARGTVGDGVCRAVAKKWPSAFKGA.........ATPVGTIKTVMCG..SYPVIHA
SARS-CoV EEPVNQFTGYLKLTDNVAIKCVDIVKEAQSANPMVIVNAANIHLKHGGGVAGALNKATNGAMQKESDDYIKLNGPLTVGGSCLLSGHNLAKKCLHV
CHIKV VGPNFSNYSESEGDRELAAAYREVAKEVTRLGVNSVAIPLLSTGVYSGGKDRLTQSLNHLFTAMDSTDADVVIYCRDKE.WEKKISEAIQMRT
VEEV VGPNFNKVSEVEGDKQLAEAYESIAKIVNDNNYKSVAIPLLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKK.WEMTLKEAVARRE
EEEV VGPNFSRLSENEGDQKLSEVYMDIARIINNERFTKVSIPLLSTGIYAGGKDRVMQSLNHLFTAMDTTDADITIYCLDKQ.WESRIKEAITRKE
WEEV VGPNFSKMPEPEGDLKLAAAYMSIASIVNAERITKISVPLLSTGIYSGGKDRVMQSLHHLFTAFDTTDADVTIYCLDKQ.WETRIIEAIHRKE
SindbisV VGPDFRKHPEAEALKLLQNAYHAVADLVNEHNIKSVAIPLLSTGIYAAGKDRLEVSLNCLTTALDRTDADVTIYCLDKK.WKERIDAVLQLKE
SFV VAPNFSATTEAEGDRELAAVYRAVAAEVNRLSLSSVAIPLLSTGVFSGGRDRLQQSLNHLFTAMDATDADVTIYCRDKS.WEKKIQEAIDMRT
SARS-CoV VGPNLNA...GEDIQLLKAAYENFNSQ......DILLAPLLSAGIFGA......KPLQSLQVCVQTVRTQVYIAVNDKALYEQVVMDYLDNLK

FIG. 1. Structures of the macro domains from the alphaviruses CHIKV and VEEV. (A) Representation of CHIKV and VEEV macro domains in a purple-to-red gradient (from N terminus to C terminus). Secondary structure elements are labeled on the CHIKV macro domain. (B) Superposition of representations of the CHIKV, VEEV, SARS-CoV, and *E. coli* macro domains, colored, respectively, in dark blue, cyan, purple, and white. (C) Electrostatic surface potential presented between −4 and 4 kT/e for the CHIKV and VEEV macro domains. The potential was generated using the PDB2PQR server (8) and Adaptative Poisson-Boltzmann Solver (APBS) plug-in in Pymol (http://www.pymol.org). (D) Representation of the superposition between the macro domains from CHIKV (apo form in dark blue), VEEV (apo form in cyan and complexed with ADP-ribose in light orange). The two main divergences between the three structures are indicated and circled. (E) Sequence alignment of the macro domains studied. They belong to the genus *Alphavirus*, except for SARS-CoV. Residues in red boxes are strictly conserved, while those in yellow boxes are conserved by at least four out of seven viruses aligned. Secondary structure elements from the viral macro domain crystal structures obtained are represented above the alignment.

main present in the nsp3 protein product of the SARS-CoV orf1a replicase polyprotein (10, 46). The rmsds between the Cα of SARS-CoV and alphavirus macro domains are between 2.80 and 2.97 Å on 160 aa. Structural differences between SARS-CoV and the alphavirus macro domain are consistent with the differences observed in all of the known macro domain crystal structures. Moreover, the SARS-CoV macro domain is longer at the N terminus and the additional 13 residues allow the formation of a seventh β strand in the central β sheet. Therefore, from a structural point of view, there is no obvious cluster of related viral macro domain structures.

Electrostatic surface analysis of alphavirus macro domains reveals a highly positively charged patch located both in the crevice, previously defined as an ADP-ribose 1″-phosphate phosphatase active site (17) and at its periphery. The other face of the protein, located far from the active site, is negatively charged (not shown). This bimodal charge distribution is much more pronounced in alphavirus macro domains than in other macro domain structures.

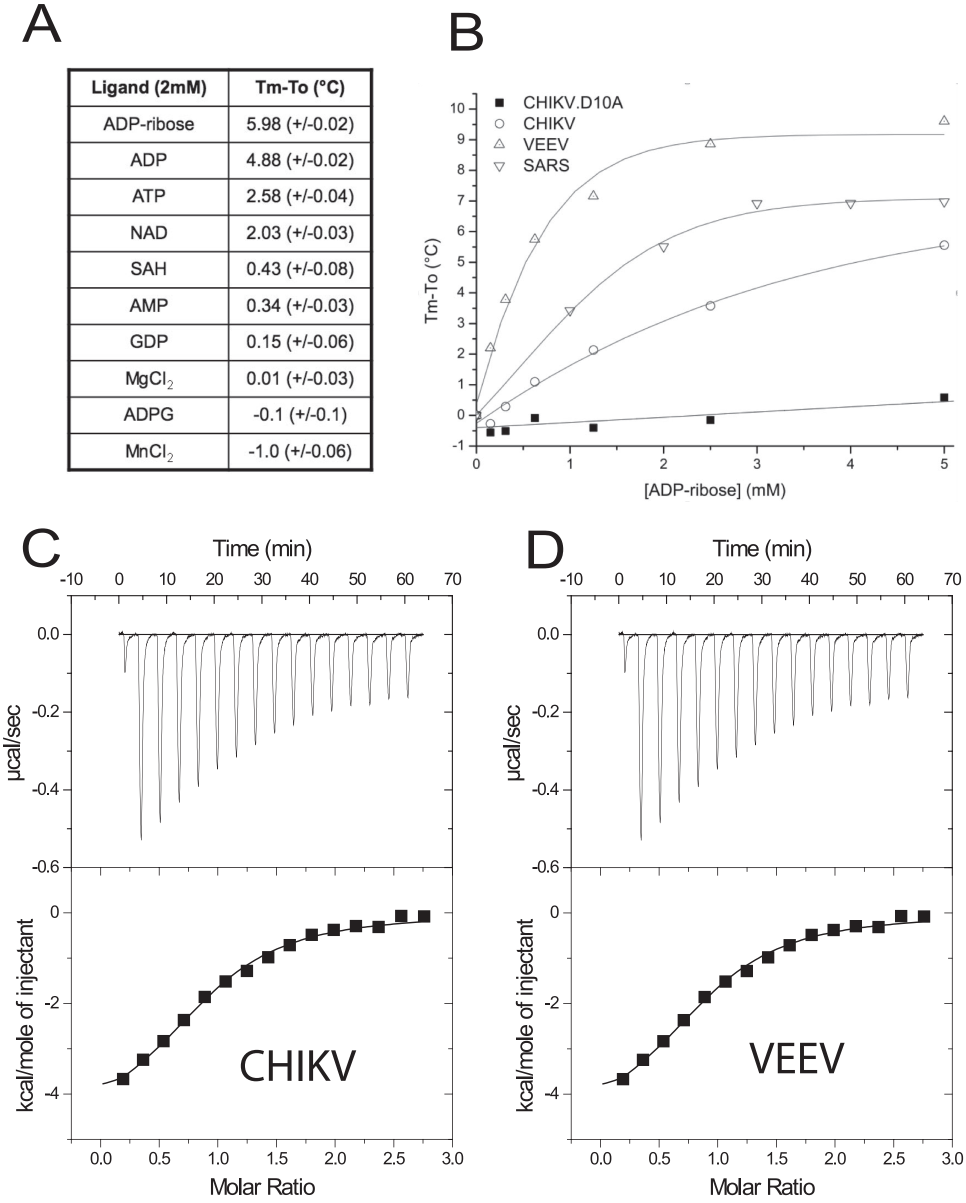
Between the two alphavirus macro domain structures, the sequence identity is 57% and the crystal structures are also very similar, the rmsd between CHIKV and VEEV macro domains being 0.91 Å on 160 aa (Fig. 1A and D). Interestingly, the main structural divergence appears between residues 30 to 37 when comparing CHIKV and VEEV macro domains in the apo form (Fig. 1D, divergence 1). Some of these residues, located close to the active site, are involved in RNA and ADP-ribose binding (see below). This structural difference could thus impact substrate binding differentially within the alphaviruses (see Discussion). Interestingly, the corresponding loop in other available macro domain structures is very well conserved and corresponds to the position of the CHIKV macro domain loop. In addition, two loops located far from the active site, from positions 48 to 52 and 62 to 65, are different between the CHIKV and VEEV macro domains (Fig. 1D, divergence 2). These loops are located far from the active site. The divergence of these loops does not originate from crystal packing. Indeed, a crystal form obtained for the VEEV macro domain in complex with ADP-ribose (see below) shows a similar position of these loops despite a different packing.

Biochemical and structural basis of ADP-ribose binding. Since several macro domains were reported as ADP-ribose binding proteins (17), the binding of ADP-ribose and other putative substrates was tested on the CHIKV macro domain using a thermal shift assay following the strategy developed previously (33). The T_m of the protein is measured using SYPRO orange fluorescence, which increases when this dye binds to hydrophobic core of the protein upon thermal denaturation. The melting temperature with the ligand (T_m) is compared to the T_m of the protein alone (T_o) in order to assess ligand binding. The ligands tested on the CHIKV macro domain can be divided into two groups (Fig. 2A). First of all, ADP-ribose, ADP, ATP, and NAD^+ shift significantly the T_m of the protein to higher values (between 2 and 6°C). This first group contains an ADP moiety. A ribose at the distal position contributes to stabilizing the macro domain, better than either nothing (for ADP), a phosphate (for ATP), a nicotinamide group (for NAD^+), or a glucose (for ADP-glucose). The other ligands have either a poor effect or no effect on the thermal melting resistance, suggesting that they do not bind the protein. Altogether, these results indicate that the CHIKV macro domain has a specificity for (i) an adenine rather than a guanine, (ii) two phosphate groups rather than either one or three, and (iii) a ribose at the distal position.

Both CHIKV and VEEV macro domains bind ADP-ribose, although they exhibit a quite different thermal denaturation shift profile (Fig. 2B). The large T_m-T_o observed for the VEEV macro domain will find a structural explanation (see Fig. 1D and Fig. 3B for details). When the D10 position is changed to alanine, the CHIKV macro domain loses its ability to bind ADP-ribose (Fig. 2B; see the structural rationale for the selection of this mutant below).

The ADP-ribose binding efficacy was then assessed by ITC. The K_ds for CHIKV (Fig. 2C) and VEEV (Fig. 2D) macro domains are 5 ± 0.4 μM and 3.9 ± 0.65 μM, respectively. Compared to the existing data, the binding affinity of both CHIKV and VEEV macro domains for ADP-ribose is almost the same as that of the yeast protein Poa1p (32) and 5 and 10 times stronger than that of SARS-CoV and HEV (10). In contrast, the macro domain of SFV, the only other virus analyzed in this way, does not appear to bind ADP-ribose, although protein instability might have impaired precise measurement (32).

The structural basis of ADP-ribose specificity was revealed by soaking and cocrystallization with ADP-ribose for CHIKV and VEEV macro domains, respectively. ADP-ribose binds to a crevice, previously defined as the ADP-ribose 1″-phosphate phosphatase active site (17, 28), in a similar manner to the two alphaviruses' macro domains. The crevice is located at the top of the β strands 2, 4, and 5 and surrounded by the loops connecting β2-α1 and β5-α3 (Fig. 1A and 3A). ADP-ribose lies in a slightly bent conformation comparable to that seen in other known macro domain–ADP-ribose complexes. The specificity for the adenosine base detected by the thermal shift assay (Fig. 2) is at least in part provided by the D10 residue. The adenine moiety is selectively hydrogen bonded via its N6 nitrogen to the D10 side chain (Fig. 3B). Interestingly, this residue is conserved in most macro domain sequences but not Sindbis virus (Fig. 1E). However, in the latter virus, the N10 residue could engage the carbonyl group of its side chain amide into a similar hydrogen bond with a N6 adenine. The D residue at this position was shown to be responsible for the adenosine specificity in *A. fulgidus* (17). In the CHIKV macro domain, the N6 nitrogen is also hydrogen bonded to R144. In addition, G32 contributes to the binding of adenine. The 3′-OH of the ADP-ribose proximal ribose (the adenosine ribose) provides a hydrogen bond to T111 (Fig. 3B). Several water molecules also interact with the proximal ribose. The phosphate binding site involves the main chain NH groups of the residues G112, V113, and Y114 for the CHIKV macro domain and G112, I113, and F114 for the VEEV macro domain (Fig. 3B). These residues define a positively charged pocket which is likely to bind bulky negatively charged groups. The distal ribose of ADP-ribose is coordinated by the Y/F114 side chain and also makes hydrogen bonds with residues N24 and D/G31 in the main chain in CHIKV and VEEV, respectively. These residues interacting with ADP-ribose are moderately conserved in alphavirus macro domains (Fig. 1E). A noticeable amino acid polymorphism in the ADP-ribose binding pocket is the isoleucine and phenylalanine in some alpha-



Ligand (2mM)	Tm-To (°C)
ADP-ribose	5.98 (+/-0.02)
ADP	4.88 (+/-0.02)
ATP	2.58 (+/-0.04)
NAD	2.03 (+/-0.03)
SAH	0.43 (+/-0.08)
AMP	0.34 (+/-0.03)
GDP	0.15 (+/-0.06)
$MgCl_2$	0.01 (+/-0.03)
ADPG	-0.1 (+/-0.1)
$MnCl_2$	-1.0 (+/-0.06)

FIG. 2. ADP-ribose binding. (A) Thermal denaturation shift at a given ligand concentration (2 mM) for the CHIKV macro domain. T_m is the melting temperature of the protein in the presence of the ligand, and T_o is the melting temperature of the protein alone. (B) ADP-ribose titration using the thermal denaturation shift assay on CHIKV (wild type and D10A mutant), VEEV, and SARS-CoV macro domains. (C) ITC assay of ADP-ribose binding to the CHIKV macro domain. (Upper panel) ITC raw data of ADP-ribose binding to the CHIKV and VEEV macro domains. (Lower panel) ADP-ribose binding isotherm derived from raw data. (D) Same experiment as in panel C but using the VEEV macro domain.

viruses, such as VEEV, in lieu of V113 and Y114 as in CHIKV (Fig. 1E). Even if these changes do not modify the conformation of the ADP-ribose bound in the CHIKV and VEEV macro domain (Fig. 3B), they are likely to promote differences in ADP-ribose binding affinity between them (see Discussion). This binary complex structure allowed the selection of mutants in order to probe ligand-protein interactions. The CHIKV macro domain mutant D10A expectedly showed a reduced thermal shift compared to the same assay with the native protein (Fig. 2B).

No noticeable conformational change occurs in the CHIKV macro domain between the apo and ADP-ribose bound forms. On the contrary, in VEEV macro domain, the loop containing residues 30 to 37 changes its conformation when it binds to ADP-ribose, and adopts a conformation similar to the corresponding loop the CHIKV macro domain (Fig. 1D and 3B). Additionally, residue R114 adopts a different rotamer. These rearrangements are necessary in order to allow coordination of the adenine and the proximal ribose of the ADP-ribose. Furthermore, the residue N24 adopts two different rotamers, depending on the asymmetric unit of the molecule. These modifications are probably related to the differences in thermal denaturation shifts yet lead to similar measured equilibrium constants around 4 to 5 μM. Moreover, these observed rearrangements could explain the absence of ADP-ribose binding in VEEV crystals soaked with ADP-ribose despite the affinity of this domain for ADP-ribose as seen using ITC. Furthermore, in the VEEV macro domain crystal, one molecule of the asymmetric unit contains a Bicine molecule (Fig. 3B), likely originating from the gel filtration buffer. This Bicine molecule could have prevented ADP-ribose binding in the soaking experiment but corroborates the proposition that this pocket is able to accommodate negatively charged groups.

ADP-ribose 1″-phosphate phosphatase activity. ADP-ribose 1″-phosphate phosphatase was the first activity detected for the yeast Poa1p macro domain based on a genome-wide search of this specific activity (28). In viruses, this activity has already been detected for SARS-CoV, SFV, and HEV macro domains (10) but was associated with a poor turnover constant. In the case of alphaviruses, the SFV ADP-ribose 1″-phosphate phosphatase activity is at the limit of detection (10). This observation, together with our ADP-ribose complexes presented above, prompted us to test this activity for the alphavirus macro domains studied here. CHIKV and VEEV macro domains showed an activity comparable to that of yeast Poa1p macro domain (Fig. 4). To get insight into the characterization of the residues involved in catalytic activities, we tested whether the residues involved in ADP-ribose binding in the crystal structure were indeed able to alter the ADP-ribose 1″-phosphate phosphatase activity. The CHIKV macro domain N24A and Y114A mutants were inactive, while the D10A mutant showed only a decreased activity under the conditions used (Fig. 4). As pinpointed by our structural model, N24 is vicinal to the electrophilic center, i.e., the 1″ phophosphorus of the ADP-ribose 1″ phosphate. The decrease but not annihilation, of activity promoted by D10A is indeed consistent with a role in binding, and not catalysis, as D10 is hydrogen bonded to the ADP-ribose adenine (Fig. 3B), far away from the 1″ phophosphorus of the ADP-ribose 1″ phosphate. In contrast, N24 could well be involved in the phosphatase reaction, as in the case of the SARS-CoV macro domain (10). N24 is

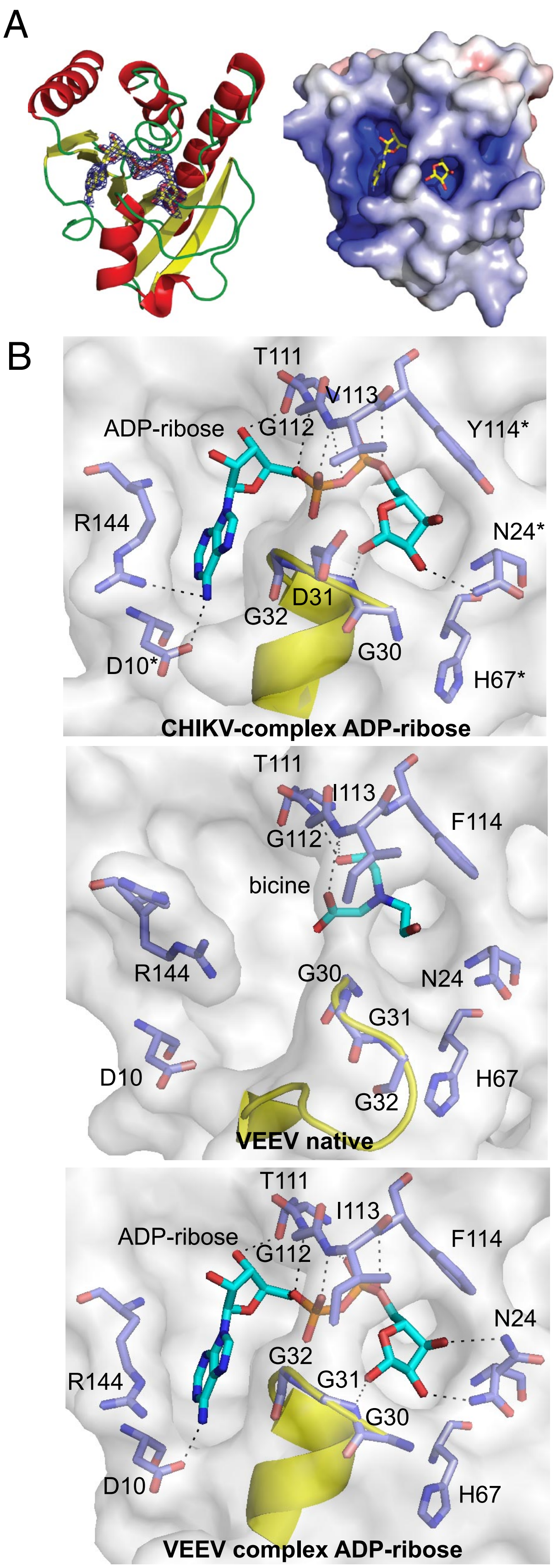


FIG. 3. Structural basis for ADP-ribose binding. (A) ADP-ribose binding site of the CHIKV macro domain. On the left side is a representation of the CHIKV macro domain with helices, strands, and loops colored, respectively, in red, yellow, and green. ADP-ribose is displayed in sticks with carbons in yellow, oxygens in red, nitrogens in blue, and phosphorus in orange. The Fo − Fc difference map, contoured at 3σ, was calculated at a 1.80-Å resolution from a model in which the ligand was omitted. On the right side is an electrostatic surface representation of the CHIKV macro domain in complex with ADP-ribose. The electrostatic potential is shown between −8 and 8 kT/e and has been generated as in Fig. 1C. The ADP-ribose molecule is shown as on the left side of panel A. (B) The ADP-ribose binding site presented with the same orientation as in panel A. The CHIKV macro domain in complex with ADP-ribose is shown in the upper part, the VEEV macro domain in complex with a Bicine molecule originating from the buffer is presented in the middle part, and the VEEV macro domain in complex with ADP-ribose is shown in the lower part. ADP-ribose and Bicine are shown in cyan. Residues interacting with ADP-ribose in the CHIKV macro domain are shown in blue, and mutated residues are indicated with an asterisk. Corresponding residues in VEEV are colored in blue. Hydrogen bonds between the ligand and the protein are shown in black dotted lines. The loops containing residues 30 to 37 are shown in yellow in CHIKV and VEEV macro domains: their conformations diverged particularly between residues 32 and 35.

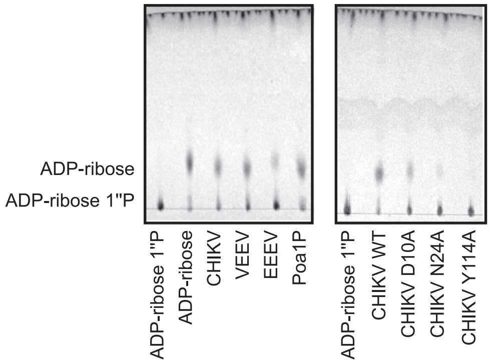


FIG. 4. Detection by TLC of ADP-ribose 1″-P, the product of the ADP-ribose phosphatase activity. Control lanes (ADP-ribose 1″-phosphate and ADP-ribose) and reaction lanes with corresponding enzymes and D10A, N24A, and Y114A mutants are indicated. The EEEV lane corresponds to the purified EEEV macro domain designed and produced in the same manner as the CHIKV macro domain. The yeast macro domain protein Poa1p serves as a positive control.

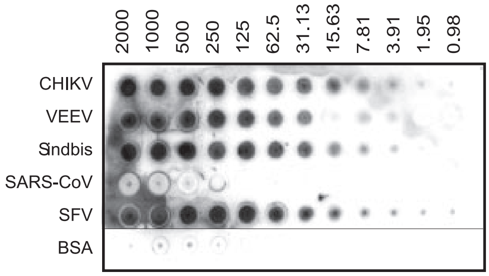


FIG. 5. Slot blot PAR binding assay of macro domains from CHIKV, VEEV, Sindbis virus, SARS-CoV, and SFV. BSA was used as a negative control. PAR was synthezised using PARP-1 and $[^{32}P]NAD^{+}$, diluted, and used as a probe to label immobilized proteins (2,000 to 0.98 pmol) blotted onto a nitrocellulose membrane as described in Materials and Methods.

correctly positioned to activate a water molecule, which in turn could promote a nucleophilic attack on the phosphorus of the ADP-ribose 1″ phosphate. The distance of 6 Å between N24 and the 1″ position of the ribose could be consistent with such a reaction, leaving enough space for both a water molecule and the phosphate.

PAR binding. Conserved positive patches on the electrostatic surface of the protein (Fig. 1C) suggest possible binding of longer negatively charged chains such as PAR or RNA. Macro domain binding to PAR has already been shown for *A. fulgidus* and SARS-CoV macro domains (10, 17). The difficulty of obtaining homogeneous PAR in sufficient quantities inhibited our attempts to obtain a crystal structure of an alphavirus macro domain in complex with PAR. However, we tested PAR binding for several alphavirus macro domains in comparison with SARS-CoV PAR binding using a slot-blot assay. All alphavirus macro domains tested (CHIKV, VEEV, Sindbis virus, and SFV) showed stronger PAR binding than SARS-CoV macro domain (Fig. 5) (10). The alanine single-amino-acid mutants described above were assayed for PAR binding. The mutants did not induce a significant decrease in PAR binding either (not shown). This indicates that any such single mutation is not sufficient to destabilize the interaction with a PAR chain significantly.

RNA binding. Since alphavirus are single-stranded RNA viruses and macro domains are able to bind long negatively charged chains, it had been hypothesized that a viral macro domain could also bind RNA (21, 34). Neuvonen and Ahola (32) have recently shown that several viral macro domains are able to bind PAR as well as RNA. The interaction between RNA and the CHIKV macro domain was studied using a slot blot assay in conjunction with the HCV RNA-dependent RNA polymerase serving as a positive control. Single-stranded RNA binding was indeed detected using a wide variety of RNA oligonucleotides of unrelated sequences. It was detected not only for the wild-type CHIKV macro domain but also for the D10A, N24A, and Y114A mutants (Fig. 6A). No clear difference in binding affinity appeared when compared to wild type. This means that as in the case of PAR, either the RNA binding site does not correspond to the ADP-ribose binding site or a single-amino-acid change has no drastic impact on RNA binding, as previously hypothesized in the case of PAR binding.

Structural basis of RNA binding. The structural basis of RNA binding was then assessed by soaking crystals of the CHIKV macro domain with several small RNAs (see Materials and Methods) and one DNA (AAAGCCAAAAA). Only small adenine-containing RNA or DNA gave a significant extra density upon analysis of the soaked crystals. In these cases, a density was observed within the hydrophobic crevice that binds ADP-ribose. The AMP was virtually superimposable to the adenosine moiety observed in the binary complexes with ADP-ribose. However, the visible density differs between the molecules of the asymmetric unit (data not shown). In contrast, in the case of the AAA RNA, two molecules of the asymmetric unit show a clear density corresponding to the soaked ligand. This density corresponds to a bent RNA that is partially bound in the crevice corresponding to the ADP-ribose binding site (Fig. 6B and C).

Interestingly, an adenine base of the RNA is bound in a similar fashion to that of the ADP-ribose adenine. The RNA is partly disordered at its extremities. Due to steric constraints, a 3′ phosphate cannot be positioned near the loop linking β5 to α3 (containing residues 111 to 113; Fig. 6B). Thus, the first nucleotide (nt 1 [at the 5′ end]) of the RNA is located near this latter loop (Fig. 6B). Only the 3′ phosphate of nt 1 is visible, together with the entire nt 2 and nt 3.

The binding of nt 1 is promoted by coordination of its phosphate with the main chain nitrogen group of residues G32, G112, and V113 (Fig. 6C). The 2′-O position of the nt 2 ribose is coordinated by W148 and C143. As in the case of ADP-ribose, specificity for adenosine of nt 2 is mediated by the same hydrogen bond between D10 and the adenine N6. The adenine of nt 2 is also coordinated by residues G32 and R144. The phosphate of nt 3 makes hydrogen bonds with nitrogen groups of residues R144 and D145, whereas the ribose of nt 3 interacts with D145 and W148. The adenine of nt 3 does not make any interaction with the protein. Therefore, we conclude that the CHIKV macro domain exhibits an RNA binding pocket encompassing an AMP binding pocket identical to that used for ADP-ribose binding.

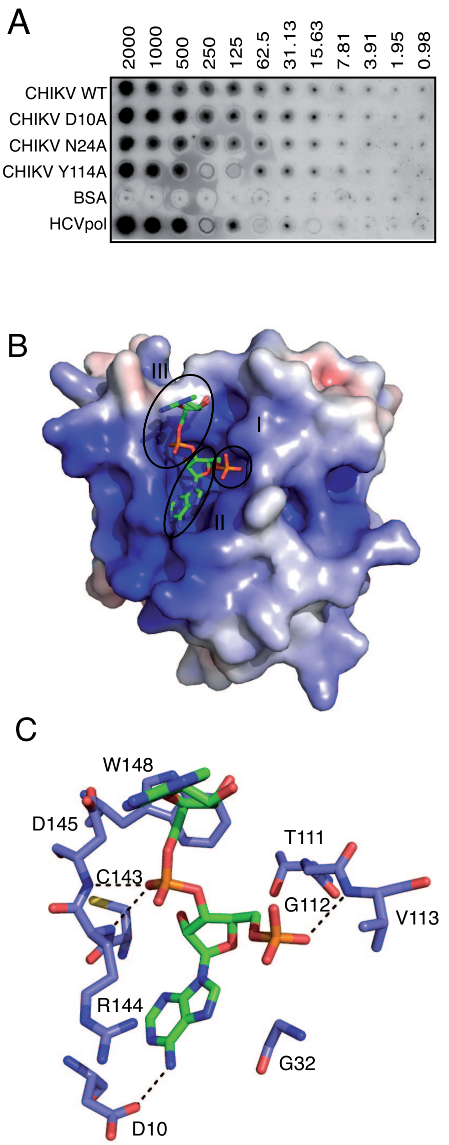


FIG. 6. RNA binding assays. (A) Slot blot RNA binding assay with wild-type and mutant CHIKV macro domains using the 5′-^{32}P-labeled RNA AAAAAAAAAGCUACC as a probe to label immobilized proteins. The amount of proteins blotted onto the membrane ranges from 2,000 to 0.98 pmol, as indicated. (B) Surface representation of the CHIKV macro domain in complex with the RNA AAA. Proposed nt 1, 2, and 3 (I, II, and III, respectively) are circled. The electrostatic potential is shown between −6 and 6 kT/e and has been generated as in Fig. 1C. (C) RNA binding site. RNA is shown in green and interacting residues in blue, with oxygens in red. H-bonds are indicated with dotted lines.

DISCUSSION

We have determined two novel alphavirus macro domain crystal structures and shown that these domains can bind ADP-ribose, RNA, and PAR. The crystal structures of the alphavirus macro domains show a high degree of conservation with other available macro domain structures (2, 4, 10, 19, 20, 27, 43). Surprisingly, the macro domain that shows the highest sequence and structural similarity to alphavirus macro domain is that of *E. coli* and not coronavirus. A direct reflection of this observation is that alphaviruses exhibit a much higher affinity for both ADP-ribose and PAR than coronaviruses do. This might suggest that the macro domain module has been acquired twice during evolution for different purposes within the viral world or that the difference of macro domain function between alphavirus and coronavirus has promoted divergence. This point attracts attention to the fact that one should not automatically assume that these macro domains perform the same role, still as elusive, in coronaviruses and alphaviruses. Along these lines, Neuvonen and Ahola (32) have reported a comparative study of viral and cellular macro domains and concluded that the function of a given macro domain might not directly illuminate the function of any other related macro domain.

ADP-ribose binding and ADP-ribose 1″-phosphate phosphatase activity, shown in most of the macro domains studied, are also detected in alphaviruses (10, 17, 28, 43). The level of activity is comparable to that of other tested viral macro domains (Fig. 4) (10, 32), except for that of SFV, which appears to be lower, assuming that protein stability is not interfering with activity measurements (32).

Using mutagenesis experiments, we have confirmed the binding determinants of the ADP-ribose molecule. The D10 residue conserved in all but one macro domain plays a central role in adenine specificity, but not in ADP-ribose 1″-phosphate phosphatase activity that is dependent on N24 and Y114 residues. We show here that the CHIKV macro domain also binds RNA. The crystal structure of the CHIKV macro domain in complex with the AAA trimer RNA allows the identification of residues involved in substrate binding. This D10 residue is also involved in the specific recognition of adenines, in the same manner as in the case of ADP-ribose. The RNA is only partially bound in the ADP-ribose binding site. However, even if the D10A substitution is able to suppress ADP-ribose binding, the CHIKV D10A macro domain is still able to bind longer adenine-containing polymers, such as RNA or PAR. This indicates that either binding of PAR/RNA involves more than a single AMP unit or RNA/PAR binds to another site rather than the ADP-ribose binding cleft.

A notable structural characteristic of macro domains is the presence of positively charged residues close to the ADP-ribose binding site, these basic patches being particularly pronounced in the alphavirus macro domain. Since macro domains from different species bind PAR and RNA (10, 17, 32), these characteristics suggest that residues of these patches could be involved in the binding of such negatively charged substrates longer than ADP-ribose. In agreement with this localized enhanced basicity present in alphavirus macro domains, our results reveal a higher binding affinity to these substrates in alphaviruses than in coronaviruses. Also, since random sequence RNAs bind quite well to the macro domains of different alphaviruses, the basic surface (Fig. 1C) might represent another unspecific, negatively charged polymer binding site distinct from the AMP binding crevice.

Although many RNA oligonucleotides of unrelated sequences bind to the macro domain, it is remarkable that a single AMP remains well defined at all electronic densities, remaining the sole ligand moiety common to all macro domains so far. Together with the possible RNA or PAR binding site outside the ADP-ribose binding cleft on the basic patches, it is tempting to speculate that the viral macro domains work with another yet-to-be defined AMP-containing substrate. Recent results have also proposed this possibility when CoV macro domains were compared and found to bind ADP-ribose very differently: although structurally very close, a group 3 CoV macro domain does not bind ADP-ribose, whereas a group 1 CoV macro domain binds it with a K_d of 29 μM (36).

There are a number of investigational avenues regarding the role of the viral macro domains. First, alphavirus infection can induce synthesis of a large quantity of PAR upon PARP-1 activation (31). This leads to depletion of ATP and NAD^+ present in the cell and induces activation of the AIF, which in turn promotes apoptosis. The role of the viral macro domain could thus be related to the binding of PAR and to the modification of the cellular response to viral infection. It remains to be investigated if and how PAR, which is synthesized by the PARP-1 in the nucleus, could interact directly with alphavirus macro domain, the latter having been detected only in the cytoplasm.

Second, the binding of RNA certainly requires further biochemical studies using the replication complex of these viruses. Indeed, the macro domain, being part of a large replicase, could serve either as a non-sequence-specific RNA recruitment factor or adenine-containing RNA recruitment factor in order to provide the RNA template to the neighboring nonstructural proteins.

Third, there are other areas of investigation that have been suggested to depend on the function of the macro domain, such as inflammation (13). It is also possible that functions described in the second and third points above might be combined since the coronavirus macro domain is located in the same nsp3 as the papain-like proteinase domain, which has been shown to possess deubiquitinylating and interferon antagonistic activity (26).

Our work provides a structural basis with which to begin to address these questions in alphaviruses, for which replicons and infectious recombinant clones exist and which might prove less demanding than equivalent studies using much larger CoV genomes.

ACKNOWLEDGMENTS

We thank Violaine Lantez, Karen Dalle, and Nadège Brisbarre for technical support, as well as the European and SOLEIL Synchrotron Radiation Facility staffs for help with crystallographic data collection. A.E.G. thanks Dmitriy Samborsky and Alexander Kravchenko for administrating the Viralis software.

This work was supported in part by the VIZIER integrated project (LSHG-CT-2004-511960) of the European Union 6th Framework, the Direction Générale de l'Armement, and the Agence Nationale de la Recherche (CHIKVIRULENCE grant).

REFERENCES

1. **Ahola, T., and L. Kaariainen.** 1995. Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. Proc. Natl. Acad. Sci. USA **92:**507–511.
2. **Allen, M. D., A. M. Buckle, S. C. Cordell, J. Lowe, and M. Bycroft.** 2003. The crystal structure of AF1521 a protein from Archaeoglobus fulgidus with homology to the non-histone domain of macroH2A. J. Mol. Biol. **330:**503–511.
3. **Berrow, N. S., K. Bussow, B. Coutard, J. Diprose, M. Ekberg, G. E. Folkers, N. Levy, V. Lieu, R. J. Owens, Y. Peleg, C. Pinaglia, S. Quevillon-Cheruel, L. Salim, C. Scheich, R. Vincentelli, and D. Busso.** 2006. Recombinant protein expression and solubility screening in Escherichia coli: a comparative study. Acta Crystallogr. D Biol. Crystallogr. **62:**1218–1226.
4. **Chakravarthy, S., S. K. Y. Gundimella, C. Caron, P.-Y. Perche, J. R. Pehrson, S. Khochbin, and K. Luger.** 2005. Structural characterization of the histone variant macroH2A. Mol. Cell. Biol. **25:**7616–7624.
5. **Chiarugi, A., and M. A. Moskowitz.** 2002. Cell biology. PARP-1—a perpetrator of apoptotic cell death? Science **297:**200–201.
6. **Collaborative Computational Project, No. 4.** 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol. Crystallogr. **50:** 760–763.
7. **de Lamballerie, X., E. Leroy, R. N. Charrel, K. Ttsetsarkin, S. Higgs, and E. A. Gould.** 2008. Chikungunya virus adapts to tiger mosquito via evolutionary convergence: a sign of things to come? Virol. J. **5:**33.
8. **Dolinsky, T. J., P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker.** 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. **35:**W522–W525.
9. **Doublie, S.** 1997. Preparation of selenomethionyl proteins for phase determination. Methods Enzymol. **276:**523–530.
10. **Egloff, M.-P., H. Malet, Á. Putics, M. Heinonen, H. Dutartre, A. Frangeul, A. Gruez, V. Campanacci, C. Cambillau, J. Ziebuhr, T. Ahola, and B. Canard.** 2006. Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains. J. Virol. **80:**8493–8502.
11. **Emsley, P., and K. Cowtan.** 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60:**2126–2132.
12. **Ericsson, U. B., B. M. Hallberg, G. T. Detitta, N. Dekker, and P. Nordlund.** 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal. Biochem. **357:**289–298.
13. **Eriksson, K. K., L. Cervantes-Barragán, B. Ludewig, and V. Thiel.** 2008. Mouse hepatitis virus liver pathology is dependent on ADP-ribose-1″-phosphatase, a viral function conserved in the alpha-like supergroup. J. Virol. **82:**12325–12334.
14. **Geerlof, A., J. Brown, B. Coutard, M. P. Egloff, F. J. Enguita, M. J. Fogg, R. J. Gilbert, M. R. Groves, A. Haouz, J. E. Nettleship, P. Nordlund, R. J. Owens, M. Ruff, S. Sainsbury, D. I. Svergun, and M. Wilmanns.** 2006. The impact of protein characterization in structural proteomics. Acta Crystallogr. D Biol. Crystallogr. **62:**1125–1136.
15. **Gorbalenya, A. E., E. V. Koonin, and M. M. Lai.** 1991. Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. FEBS Lett. **288:**201–205.
16. **Kabsch, W.** 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. **26:**795–800.
17. **Karras, G. I., G. Kustatscher, H. R. Buhecha, M. D. Allen, C. Pugieux, F. Sait, M. Bycroft, and A. G. Ladurner.** 2005. The macro domain is an ADP-ribose binding module. EMBO J. **24:**1911–1920.
18. **Kennedy, A. C., J. Fleming, and L. Solomon.** 1980. Chikungunya viral arthropathy: a clinical description. J. Rheumatol. **7:**231–236.
19. **Kumaran, D., S. Eswaramoorthy, F. W. Studier, and S. Swaminathan.** 2005. Structure and mechanism of ADP-ribose-1″-monophosphatase (Appr-1″-pase), a ubiquitous cellular processing enzyme. Protein Sci. **14:**719–726.
20. **Kustatscher, G., M. Hothorn, C. Pugieux, K. Scheffzek, and A. G. Ladurner.** 2005. Splicing regulates NAD metabolite binding to histone macroH2A. Nat. Struct. Mol. Biol. **12:**624–625.
21. **Ladurner, A. G.** 2003. Inactivating chromosomes: a macro domain that minimizes transcription. Mol. Cell **12:**1–3.
22. **Lastarza, M. W., A. Grakoui, and C. M. Rice.** 1994. Deletion and duplication mutations in the C-terminal nonconserved region of Sindbis virus nsP3: effects on phosphorylation and on virus replication in vertebrate and invertebrate cells. Virology **202:**224–232.
23. **Leslie, A. G. W.** 1992. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 and ESF-EACMB Newsl. Protein Crystallogr. **26.**
24. **Letunic, I., R. R. Copley, B. Pils, S. Pinkert, J. Schultz, and P. Bork.** 2006. SMART 5: domains in the context of genomes and networks. Nucleic Acids Res. **34:**D257–D260.
25. **Li, G. P., M. W. La Starza, W. R. Hardy, J. H. Strauss, and C. M. Rice.** 1990. Phosphorylation of Sindbis virus nsP3 in vivo and in vitro. Virology **179:**416–427.
26. **Lindner, H. A., N. Fotouhi-Ardakani, V. Lytvyn, P. Lachance, T. Sulea, and R. Ménard.** 2005. The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. J. Virol. **79:**15199–15208.
27. **Malet, H., K. Dalle, N. Bremond, F. Tocque, S. Blangy, V. Campanacci, B. Coutard, S. Grisel, J. Lichiere, V. Lantez, C. Cambillau, B. Canard, and M. P. Egloff.** 2006. Expression, purification and crystallization of the SARS-CoV macro domain. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. **62:**405–408.
28. **Martzen, M. R., S. M. McCraith, S. L. Spinelli, F. M. Torres, S. Fields, E. J. Grayhack, and E. M. Phizicky.** 1999. A biochemical genomics approach for identifying genes by the activity of their products. Science **286:**1153–1155.

29. **McCoy, A. J.** 2007. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. D Biol. Crystallogr. **63:**32–41.

30. **Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia.** 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. **247:**536–540.

31. **Nargi-Aizenman, J. L., C. M. Simbulan-Rosenthal, T. A. Kelly, M. E. Smulson, and D. E. Griffin.** 2002. Rapid activation of poly(ADP-ribose) polymerase contributes to Sindbis virus and staurosporine-induced apoptotic cell death. Virology **293:**164–171.

32. **Neuvonen, M., and T. Ahola.** 2009. Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites. J. Mol. Biol. **385:**212–225.

33. **Pantoliano, M. W., E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane, and F. R. Salemme.** 2001. High-density miniaturized thermal shift assays as a general strategy for drug discovery. J. Biomol. Screen. **6:**429–440.

34. **Pehrson, J. R., and R. N. Fuji.** 1998. Evolutionary conservation of histone macroH2A subtypes and domains. Nucleic Acids Res. **26:**2837–2842.

35. **Peränen, J., M. Rikkonen, P. Liljeström, and L. Kääriäinen.** 1990. Nuclear localization of Semliki Forest virus-specific nonstructural protein nsP2. J. Virol. **64:**1888–1896.

36. **Piotrowski, Y., G. Hansen, A. L. Boomars-van der Zanden, E. J. Snijder, A. E. Gorbalenya, and R. Hilgenfeld.** 2009. Crystal structures of the X-domains of a group-1 and a group-3 coronavirus reveal that ADP-ribose binding may not be a conserved property. Protein Sci. **18:**6–16.

37. **Powers, A. M., A. C. Brault, Y. Shirako, E. G. Strauss, W. Kang, J. H. Strauss, and S. C. Weaver.** 2001. Evolutionary relationships and systematics of the alphaviruses. J. Virol. **75:**10118–10131.

38. **Putics, Á., W. Filipowicz, J. Hall, A. E. Gorbalenya, and J. Ziebuhr.** 2005. ADP-ribose-1″-monophosphatase: a conserved coronavirus enzyme that is dispensable for viral replication in tissue culture. J. Virol. **79:**12721–12731.

39. **Rikkonen, M., J. Peränen, and L. Kääriäinen.** 1994. ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2. J. Virol. **68:**5804–5810.

40. **Rivas, F., L. A. Diaz, V. M. Cardenas, E. Daza, L. Bruzon, A. Alcala, O. De la Hoz, F. M. Caceres, G. Aristizabal, J. W. Martinez, D. Revelo, F. De la Hoz, J. Boshell, T. Camacho, L. Calderon, V. A. Olano, L. I. Villarreal, D. Roselli, G. Alvarez, G. Ludwig, and T. Tsai.** 1997. Epidemic Venezuelan equine encephalitis in La Guajira, Colombia, 1995. J. Infect. Dis. **175:**828–832.

41. **Roversi, P., E. Blanc, C. Vonrhein, G. Evans, and G. Bricogne.** 2000. Modelling prior distributions of atoms for macromolecular refinement and completion. Acta Crystallogr. D Biol. Crystallogr. **56:**1316–1323.

42. **Rubach, J. K., B. R. Wasik, J. C. Rupp, R. J. Kuhn, R. W. Hardy, and J. L. Smith.** 2009. Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. Virology **384:**201–208.

43. **Saikatendu, K. S., J. S. Joseph, V. Subramanian, T. Clayton, M. Griffith, K. Moy, J. Velasquez, B. W. Neuman, M. J. Buchmeier, R. C. Stevens, and P. Kuhn.** 2005. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1″-phosphate dephosphorylation by a conserved domain of nsP3. Structure **13:**1665–1675.

44. **Schneider, T. R., and G. M. Sheldrick.** 2002. Substructure solution with SHELXD. Acta Crystallogr. D Biol. Crystallogr. **58:**1772–1779.

45. **Sheldrick, G. M.** 2002. Macromolecular phasing with SHELXE. Z. Kristallogr. **217:**644–650.

46. **Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. Spaan, and A. E. Gorbalenya.** 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. **331:**991–1004.

47. **Strauss, E. G., R. J. De Groot, R. Levinson, and J. H. Strauss.** 1992. Identification of the active site residues in the nsP2 proteinase of Sindbis virus. Virology **191:**932–940.

48. **Tuittila, M. T., M. G. Santagati, M. Röyttä, J. A. Määttä, and A. E. Hinkkanen.** 2000. Replicase complex genes of Semliki Forest virus confer lethal neurovirulence. J. Virol. **74:**4579–4589.

49. **Vihinen, H., T. Ahola, M. Tuittila, A. Merits, and L. Kaariainen.** 2001. Elimination of phosphorylation sites of Semliki Forest virus replicase protein nsP3. J. Biol. Chem. **276:**5745–5752.

50. **Vihinen, H., and J. Saarinen.** 2000. Phosphorylation site analysis of Semliki forest virus nonstructural protein 3. J. Biol. Chem. **275:**27775–27783.

51. **Wang, Y.-F., S. G. Sawicki, and D. L. Sawicki.** 1994. Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA. J. Virol. **68:**6466–6475.

52. **Weaver, S. C., T. K. Frey, H. V. Huang, R. M. Kinney, C. M. Rice, J. T. Roehrig, R. E. Shope, and E. G. Strauss.** 2005. Togaviridae, p. 999–1008. *In* C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball (ed.), Virus taxonomy. VIIIth Report of the ICTV. Elsevier/Academic Press, London, United Kingdom.

53. **Wengler, G.** 1993. The NS 3 nonstructural protein of flaviviruses contains an RNA triphosphatase activity. Virology **197:**265–273.

BMC Research Notes

BioMed Central

Short Report

Crystallographic structure of ubiquitin in complex with cadmium ions

Insaf A Qureshi[1], Francois Ferron[2], Cheah Chen Seh[1], Peter Cheung[1] and Julien Lescar*[1,2]

Address: [1]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551, Singapore and [2]AFMB UMR 6098 CNRS, Marseille, France

Email: Insaf A Qureshi - iaqureshi@ntu.edu.sg; Francois Ferron - francois.ferron@afmb.univ-mrs.fr; Cheah Chen Seh - ccseh@ntu.edu.sg; Peter Cheung - pcfcheung@ntu.edu.sg; Julien Lescar* - julien@ntu.edu.sg

* Corresponding author

Published: 15 December 2009

BMC Research Notes 2009, **2**:251 doi:10.1186/1756-0500-2-251

Received: 12 October 2009
Accepted: 15 December 2009

This article is available from: http://www.biomedcentral.com/1756-0500/2/251

Abstract

Background: Ubiquitination plays a critical role in regulating many cellular processes, from DNA repair and gene transcription to cell cycle and apoptosis. It is catalyzed by a specific enzymatic cascade ultimately leading to the conjugation of ubiquitin to lysine residues of the target protein that can be the ubiquitin molecule itself and to the formation of poly-ubiquitin chains.

Findings: We present the crystal structure at 3.0 Å resolution of bovine ubiquitin crystallized in presence of cadmium ions. Two molecules of ubiquitin are present in the asymmetric unit. Interestingly this non-covalent dimeric arrangement brings Lys-6 and Lys-63 of each crystallographically-independent monomer in close contact with the C-terminal ends of the other monomer. Residues Leu-8, Ile-44 and Val-70 that form a hydrophobic patch at the surface of the Ub monomer are trapped at the dimer interface.

Conclusions: The structural basis for signalling by poly-Ub chains relies on a visualization of conformations of alternatively linked poly-Ub chains. This arrangement of ubiquitin could illustrate how linkages involving Lys-6 or Lys-63 of ubiquitin are produced in the cell. It also details how ubiquitin molecules can specifically chelate cadmium ions.

Background

Ubiquitin (Ub) is an evolutionary conserved protein of 76 amino-acids found in all eukaryotes. Ub participates in the regulation of various cellular processes through conjugation to other cellular proteins, which is catalyzed by an enzymatic cascade that involves the sequential actions of three enzymes known as Ub-activating (E1), Ub-conjugating (E2) and Ub-protein ligase (E3) enzymes. As a result, the C terminal glycine residue of Ub is ligated to the target protein through the formation of an isopeptide bond to one of its lysine residues, by a complex formed by the E2 and E3 enzymes. Covalent attachment of the C-terminal end of Ub to lysine residues of a protein substrate, a process known as ubiquitination, targets the substrate for a range of possible fates including protein degradation. Conjugation of a single Ub molecule regulates transcription, endocytosis and membrane trafficking. Since the ubiquitin protein itself contains seven lysine residues: Lys-6, Lys-11, Lys-27, Lys-29, Lys-33, Lys-48 and Lys-63, additional Ub molecules can be ligated to each of these seven

Table 1: Data collection and refinement statistics

Data collection statistics	
Space group	P4(3)32
Unit cell a, b, c (Å)	105.23
No. of molecules per asymmetric unit	2
Resolution (Å)	30.0-3.0
No. of measured reflections	92197
No. of unique reflections	4367
Completeness	99.9 (100.0)
Redundancy	21.1 (22.9)
R_{sym} (%)	6.9 (56.7)
I/ (I)	35.7 (5.8)
Refinement statistics	
No. of reflections (working set/test set)	3613/172
R factor (R_{work}/R_{free})	0.222/0.257
No. of atoms	
Protein	1196
Cadmium	5
Water	42
B mean (Å^2)	87.88
B (Wilson Plot, Å^2)	101.29
B Protein atoms (Å^2)	72.37
B cadmium atoms (Å^2)	69.76
B water atoms (Å^2)	59.02
R.M.S. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.44
Ramachandran statistics (%)	
Most favored	81.8
Additionally allowed	17.4
Generously allowed	0.8
Disallowed	0.0

*Values in parentheses refer to the corresponding values of the highest resolution shell (3.00-3.12)

sites, leading to the formation of poly-Ub chains with distinct linkages between individual Ub molecules. The most common poly-Ub chain is linked through Lys-48 and serves as a signal for rapid degradation of substrates by the proteasome-dependent proteolysis pathway [1]. Poly-Ub chains linked through Lys-63 trigger non-proteolytic signals, such as for DNA repair, ribosomal protein synthesis, inflammatory signalling, endocytosis and vesicular trafficking. However, a recent report suggested that *in vivo*, Lys-63-linked ubiquitin chain may also serve as a targeting signal for the 26S proteaseome [2]. Lys-6-modified ubiquitin is a potent and specific inhibitor of ubiquitin-mediated protein degradation [3]. Lys-11- and Lys-29-linked chains also target the substrate for proteasome-dependent protein degradation. Lys-29- or Lys-33-linked ubiquitin chains regulate the enzymatic activity of kinases [4]. Thus, different poly-Ub chains, despite being assembled from identical Ub units are recognized as distinct signals by the cell machinery. Therefore, a visualization of conformations of alternatively linked poly-Ub chains should help deciphering the molecular and structural basis for their signalling activities.

In the present study, we crystallized bovine ubiquitin in the presence of a high concentration of cadmium salt in a form that comprises two independent molecules per asymmetric unit and report this structure at 3Å resolution.

Methods

Crystallization

Bovine ubiquitin (Sigma) was subjected to gel filtration chromatography on a HiLoad 16/60 Superdex 75 column (GE Healthcare) using buffer A (50 mM HEPES pH 7.5, 150 mM NaCl and 5 mM dithiothreitol). Fractions were pooled and concentrated to 10 mg/ml in buffer A using an Amicon Ultra-5 filter (Millipore). Screening for crystallization conditions was performed using sitting-drop vapour diffusion in 96-well plates by mixing a volume of 0.2 μl of protein solution with 0.2 μl of precipitant. Drops were equilibrated against 0.1 ml of precipitant solution at 291 K. Crystallization screens tested included the sparse-matrix Crystal Screens 1 and 2, Crystal screen Lite & Cryo, Index, SaltRx, MembFac, PEG/Ions, grid screens for sodium malonate and ammonium sulfate (Hampton Research) using a CyBio crystallization robot. Crystals were observed using a reservoir solution containing 100 mM HEPES pH 7.5, 50 mM cadmium sulfate, 1 M sodium citrate.

Data collection and structure determination

Crystals suitable for diffraction experiments grew within one week to maximum dimensions of approximately 0.20 × 0.25 × 0.25 mm and diffracted to 3.0 Å resolution at third generation synchrotron beamlines. The data set was collected at beamline ID14-4, ESRF (Grenoble, France) using crystals frozen in the reservoir solution supplemented with 30% glycerol. Reflection data were indexed, integrated and scaled using MOSFLM and the CCP4 program suite (Table 1). The structure was determined by molecular replacement with the program MOLREP [5] using the monomeric Ub structure (PDB ID code 1UBQ) as a search probe. The model was refined using the program CNS [6] and rebuilt with program O [7]. The geometry of the final model was checked with PROCHECK [8] and figures for molecular representation were prepared using PyMOL http://www.pymol.org. Statistics for structure refinement and validation are given in Table 1. The

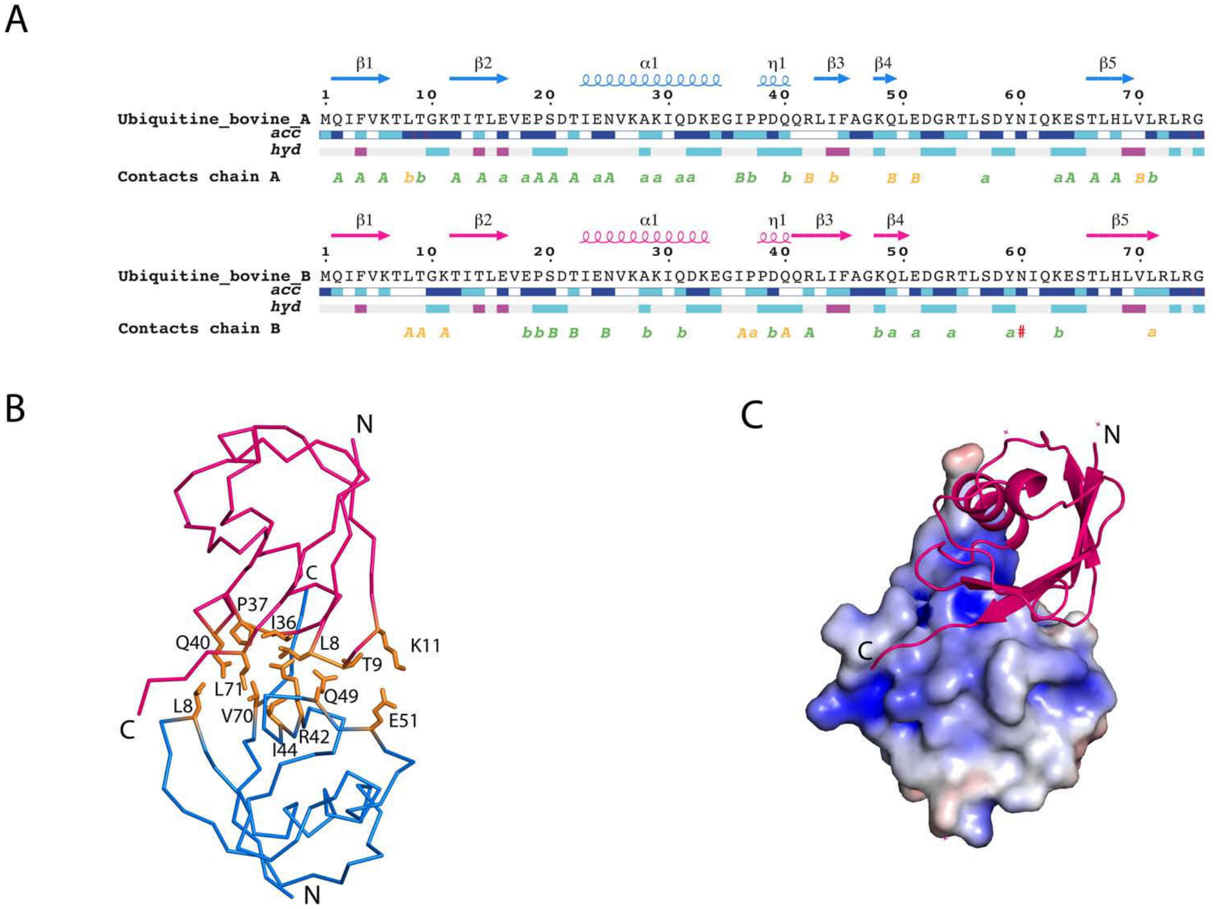

A
β1
β2
α1
η1
β3
β4
β5
1
10
20
30
40
50
60
70
Ubiquitine_bovine_A MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRG
acc
hyd
Contacts chain A A A A bb A A a aAA A aA aa aa Bb b B b B B a aA A A Bb
Ubiquitine_bovine_B MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRG
acc
hyd
Contacts chain B AA A bbB B B b b Aa bA A ba a a a# b a
B
N
C
P37
I36
Q40
L8
K11
T9
L71
Q49
L8
V70
E51
R42
I44
C
N
C
N
C

Figure 1

Overall Structure. A. Schematic representation of the intermolecular contacts established by Ub monomer A (blue) and B (pink). Non-crystallographic (yellow) or crystallographic (green) contacts are marked under the corresponding residue. A capital letter marks an interaction <3.2Å and a small letter a contact <4Å. B. Secondary structure elements are indicated above the respective sequence. The relative accessibility (acc) is calculated by DSSP and rendered as blue-colored boxes (from dark-blue: fully accessible, to white: buried). The hydropathic character of a sequence is calculated according to the algorithm of Kyte and Doolittle with a window of 3 and is rendered as red (hydrophobic) to blue (hydrophilic) boxes. C. The Ub non-covalent crystallographic dimer using the color scheme in A). Side-chains of residues involved in the association are represented as yellow sticks and labeled. N- and C-terminal ends of each Ub chain are indicated.

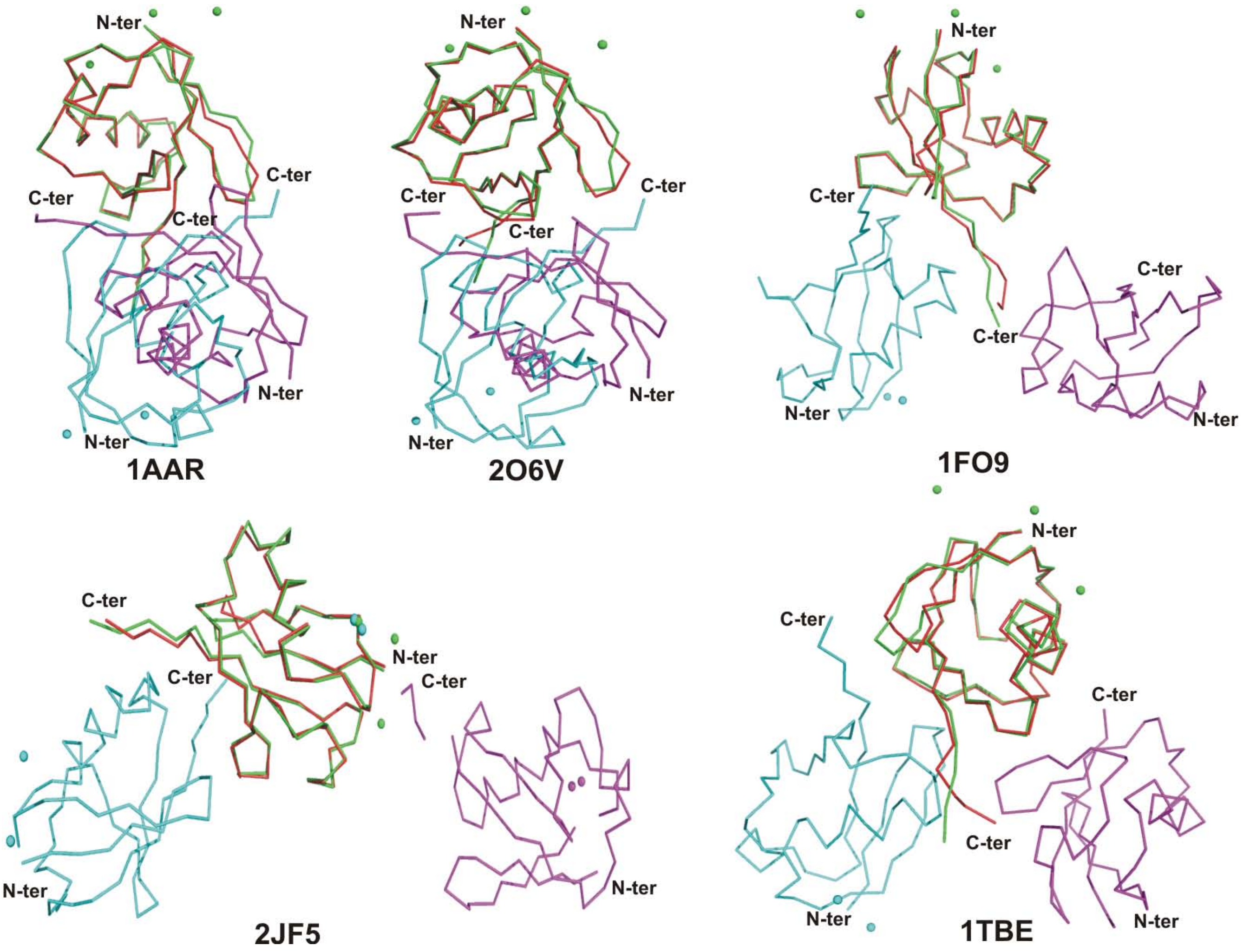


Figure 2

Comparison of Ub crystallographic dimers. Monomer A of Ub (in green, this work) was superimposed with one monomer of either 1AAR, 2O6V, 1F09, 2JF5, 1TBE. Only the α-carbon traces are displayed. The monomer B of Ub is in cyan, while the A and B monomers of other ubiquitins are colored in red and magenta. When present, cadmium ions from the crystallization media are shown as spheres.

coordinates are available from the Protein Data Bank with accession number 3H1U.

Results and Discussion

Overall Structure

The asymmetric unit contains two non-covalently linked Ub monomers (Figure. 1) hereafter named A and B that can be superimposed with an r.m.s. deviation of 0.89 Å for 75 equivalent α-carbon atoms. The absence of covalent bond formation between the two Ub monomers was confirmed both by SDS PAGE and mass spectrometry using dissolved Ub crystals that only revealed the presence of monomeric species. Structural differences with other crystallized ubiquitins [9-11] are located in one loop spanning residues Thr-7-Gly-10 and at the C-terminal ends that are engaged in different crystal packing interactions (Figure. 2). Compared to the Lys-63-linked diubiquitin that was also crystallized in presence of cadmium ions [[11], PDB code 2JF5], significant variations in conformation are observed for monomer A with a r.m.s.d. of 0.88 and a r.m.s.d of 1.34 Å for the B monomer: Of note, the α-carbon atoms of residues Leu-8 and Thr-9 are displaced by 3.6 and 3.3 Å respectively compared to the Lys-63-linked diubiquitin structure. Residue Leu-8 is an important recognition element for the binding of Lys-48-linked tetraubiquitin to the 26S proteasome and the Leu-8Ala mutation causes the largest defect in substrate degradation as compared to other mutations targeting the hydrophobic patch (e.g. Ile-44Ala and Val-70Ala) [12,13]. The ubiquitin molecule displays extensive acidic and basic

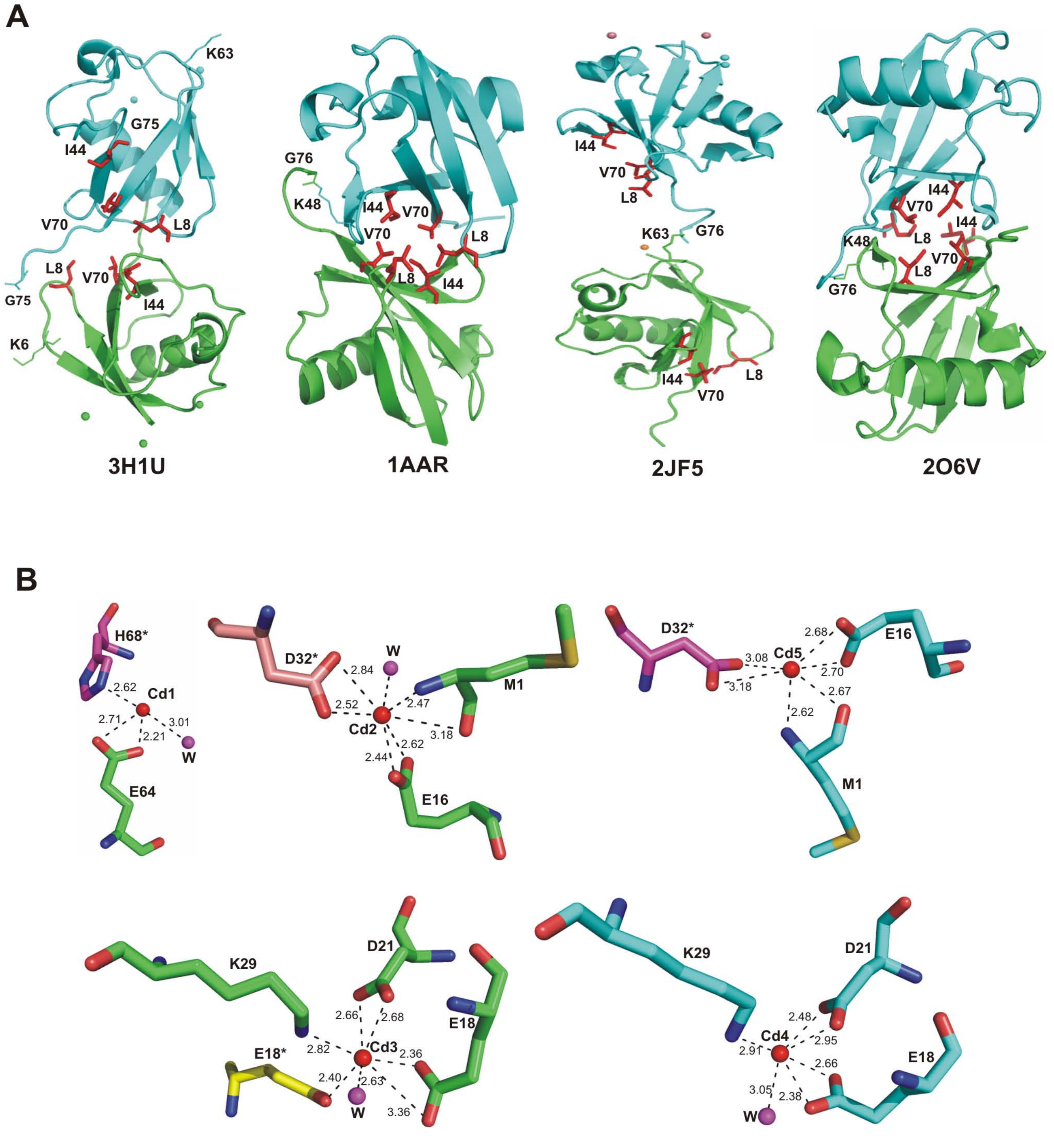


Figure 3

The hydrophobic patch and binding of cadmium ions. A. Mapping the locations of the residues that form the hydrophobic patch (Leu-8, Ile-44, and Val-70) with respect to the crystallographic dimer interface in four Ub crystal structures. Chain A are shown in green and chain B in cyan. Residues from the hydrophobic patch are shown as red sticks and labelled. When present, cadmium ions are shown as either green or cyan spheres, according to the Ub monomer they bind to. B. The five cadmium binding sites observed. Residues of chain A are in green while symmetry related molecules are shown in different colours. Contacts between cadmium ions (red spheres) and water molecules (magenta spheres) are indicated by dashed lines, with the corresponding distances labelled in Å. Oxygen atoms are coloured in red and nitrogen atoms in blue.

Table 2: Non-covalent interactions between two ubiquitin monomers

Molecule A		Molecule B		
Residue	Atom	Residue	Atom	Distance (Å)
Leu8	CD1	Gln40	NE2	3.23
Leu8	CD2	Gln40	OE1	3.99
Gln40	N	Leu8	O	2.70
Gln40	OE1	Thr9	CG2	3.73
Gln40	NE2	Thr9	CG2	2.91
Arg42	NH1	Glu34	O	2.69
Arg42	NH1	Ile36	N	3.17
Arg42	NH2	Pro37	CD	3.38
Ile44	CG2	Leu71	CD2	3.44
Glu51	OE2	Lys11	NZ	3.19
Val70	CG1	Gln40	OE1	3.20
Val70	CG2	Gln40	NE2	3.69

patches at opposite faces of the monomer. The hydrophobic patch at the surface of each ubiquitin molecule includes Leu-8, Ile-44 and Val-70 and mediates interaction between mono- and polyubiquitin and a variety of ubiquitin-binding domains [14]. In the present crystal structure these residues are trapped at the interface between monomers A and B (Figure 3A). A list of the interactions between the two crystallographically-independent Ub monomers is given in Table 2. Interestingly, the relative arrangement of the two Ub monomers brings Lys-6 from monomer A in close proximity to the C-terminal end of monomer B. Likewise, Lys-B63 is next to the C-terminus of a symmetry-related molecule of chain A (Figure 3A). Recently, it was shown that formation of Lys6-linked polyubiquitin chains was catalysed by a heterodimeric RING E3 ligase complex consisting of the breast cancer-susceptibility protein BRCA1 and BARD1 [15]. As BRCA1/BARD1 are localized at sites of DNA damage through binding of their adaptor protein RAP80 to Lys6- and Lys63-linked ubiquitin chains [16], both of these chain types might be involved in DNA repair.

The buried surface area between the two monomers present in the asymmetric unit is 1082 Å2 indicating that this dimeric arrangement will not be maintained in solution. This was confirmed by gel filtration experiments (not shown). Interestingly, the mode of association in the crystal between molecules A and B (Figure 3A) is grossly related to the Lys-48-linked diubiquitin structure observed by Cook et al [9] which also buries the hydrophobic residues (Leu-8, Ile-44 and Val-70) at the interface but has a much more extensive dimerization interface of 1507 Å2 ([9] PDB code 1AAR]. However, no clear correlation seems to exist between isopeptide bond formation, the oligomerization state and the surface area buried at the molecular interface: tetra-ubiquitin chains have been observed in various crystal structures either with a surface of 1492 Å2 for the Lys-48-linked tetra-ubiquitin ([17] PDB code 2O6V] or with much smaller buried interface of 659 Å2 ([10] PDB code 1TBE], 435 Å2 ([18] PDB code 1F9J], or 209 Å2 ([11] PDB code 2JF5]. The Lys-63-linked diubiquitin ([11] PDB code 2JF5] adopts an extended conformation with no clear interface formation between Ub domains and fully exposes both hydrophobic patch residues of each Ub to the solvent. Thus, in the present structure, the overall conformation of ubiquitin molecules differs from those seen in previously reported poly-Ub structures (Figure 3A), which probably accounts for distinct roles in cell signalling. The analysis of crystal packing shows that both molecules within the asymmetric unit come in close contact with symmetry-related molecules (Additional File 1: Fig. S1). An electron density omit map after removing Gly-75 of both chains from the phasing model, indicates that these residues are well ordered in the crystal (Additional File 2: Fig. S2). Although all known Ub structures are closely related but their modes of association differs. Thus the identification of a new interface may explain the formation of Lys-6 or Lys-63 linked diubiquitin.

Metal binding

Molecule A interacts with three cadmium ions and molecule B chelates two cadmium ions. Cadmium ions contribute directly to lattice formation by bridging residues from four neighbouring ubiquitin molecules. The first cadmium ion (Cd1) completes its tetrahedral coordination sphere by binding the carboxylic group of Glu64, atom NE2 of His68* (the star denotes a residue from a symmetry-related protein) and one water molecule (Figure 3B). Cadmium ions Cd2 and Cd5 are bound to the carboxylic group of Glu16, main chain amide and carbonyl groups of residue Met1 and the carboxylic group of Asp32* and a water molecule. Cd2 binds to a water molecule but Cd5 does not interact with any water molecule (Figure 3B). Cadmium ions Cd3 and Cd4 are chelated by the carboxylic groups of Glu18 and Asp21 and atom NZ of Lys29. One water molecule completes their coordination shells (Figure 3B). Interestingly, residues that bind cadmium ions, especially Glu16, Glu18 and Glu64 show a remarkable displacement from their positions in native ubiquitin; detailed views of the various cadmium binding sites are shown in Additional File 3: Fig. S3. A comparison of the crystal structure reported here with other ubiquitin complexed to metal ions ([11] PDB code 2JF5) is shown in Figure 2. In addition, three out of five cadmium ions bind to ubiquitin in different ways: Positions of Cd3 and Cd4 are analogous to that observed in structure 2JF5, but positions of Cd2 and Cd5 are occupied by magnesium and cobalt ions, respectively. No ligand is present in structure 2JF5 corresponding to the position occupied by Cd1. These results support the findings by Dokmanic and co-workers [19] stating that the most abundant amino-acid residues in the tetrahedral or octahedral cadmium coordination spheres are Cys followed by Glu and Asp.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

IAQ crystallized the protein and took part in refinement. FF collected diffraction data at ESRF. JL solved and refined the structure. All authors contributed reagents and participated in writing the paper.

Additional material

Additional file 1

Figure S1: Packing of molecules in crystal of Ub*. A) The non-crystallographic dimer using a ribbon representation. B) Nearest symmetry molecules (at 4Å) in the same orientation as A). C, D, E) rotation of B) by 90° around the horizontal axis, at 90° along Y, at 180° along X respectively. F) rotation of C) at 180° along Y.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1756-0500-2-251-S1.PNG]

Additional file 2

Figure S2: Omit electron density map around the C-terminal of Ub*. Omit electron density map around the C-terminal of chain A (panel A) and chain B (panel B) after removing residue Gly75 from the phase calculation. The* F_o-F_c *difference electron density map contoured at 3.0 level is coloured in green and* $2F_o$-F_c *map contoured at 1.5 level (in blue). Density accounting for Gly-75 is visible for both chains.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1756-0500-2-251-S2.PNG]

Additional file 3

Figure S3: Comparison between structures of native and cadmium bound Ub*. Superposition of the structures of native ubiquitin (magenta) and bovine ubiquitin (green) in the region of the binding sites of cadmium (red). The interactions between cadmium ions and donor atoms are indicated by dotted lines.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1756-0500-2-251-S3.PNG]

Acknowledgements

This work was supported by a BMRC grant 05/1/22/19/405 and an ATIP grant from CNRS to the laboratory of JL. We thank the ESRF, Grenoble for generously granting beam time and for expert help during data collection.

References

1. Pickart CM, Fushman D: **Polyubiquitin chains: polymeric protein signals.** *Curr Opin Chem Biol* 2004, **8:**610-616.
2. Saeki Y, Kudo T, Sone T, Kikuchi Y, Yokosawa H, Toh-e A, Tanaka K: **Lysine 63-linked polyubiquitin chain may serve as a targeting signal for the 26S proteasome.** *EMBO J* 2009, **28(4):**359-371.
3. Shang F, Deng G, Liu Q, Guo W, Haas AL, Crosas B, Finley D, Taylor A: **Lys6-modified Ubiquitin Inhibits Ubiquitin-dependent Protein Degradation.** *J Biol Chem* 2005, **280:**20365-20374.
4. Al-Hakim AK, Zagorska A, Chapman L, Deak M, Peggie M, Alessi DR: **Control of AMPK-related kinases by USP9X and atypical Lys(29)/Lys(33)-linked polyubiquitin chains.** *Biochem J* 2008, **411(2):**249-260.
5. Vagin A, Teplyakov A: **MOLREP: an automated program for molecular replacement.** *J Appl Crystallogr* 1997, **30:**1022-1025.
6. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL: **Crystallography & NMR system: A new software suite for macromolecular structure determination.** *Acta Crystallogr D* 1998, **54:**905-921.
7. Jones TA, Zou JY, Cowan SW, Kjeldgaard M: **Improved methods for building protein models in electron density maps and the location of errors in these models.** *Acta Crystallogr A* 1991, **47(Pt 2):**110-119.
8. Laskowski RA, MacArthur MW, Moss DS, Thornton JM: **PROCHECK: a program to check the stereochemical quality of protein structures.** *J Appl Cryst* 1993, **26:**283-291.
9. Cook WJ, Jeffrey LC, Carson M, Chen Z, Pickart CM: **Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2).** *J Biol Chem* 1992, **267(23):**16467-16471.
10. Cook WJ, Jeffrey LC, Kasperek E, Pickart CM: **Structure of tetraubiquitin shows how multiubiquitin chains can be formed.** *J Mol Biol* 1994, **236(2):**601-609.
11. Komander D, Reyes-Turcu F, Licchesi JD, Odenwaelder P, Wilkinson KD, Barford D: **Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains.** *EMBO Rep* 2009, **10:**466-473.
12. Beal R, Deveraux Q, Xia G, Rechsteiner M, Pickart C: **Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting.** *Proc Natl Acad Sci USA* 1996, **93(2):**861-866.
13. Beal RE, Toscano-Cantaffa D, Young P, Rechsteiner M, Pickart CM: **The hydrophobic effect contributes to polyubiquitin chain recognition.** *Biochemistry* 1998, **37(9):**2925-34.
14. Hicke L, Schubert HL, Hill CP: **Ubiquitin-binding domains.** *Nat Rev Mol Cell Biol* 2005, **6(8):**610-21.
15. Nishikawa H, Ooka S, Sato K, Arima K, Okamoto J, Klevit RE, Fukuda M, Ohta T: **Mass spectrometric and mutational analyses reveal Lys-6-linked polyubiquitin chains catalyzed by BRCA1-BARD1 ubiquitin ligase.** *J Biol Chem* 2004, **279:**3916-3924.
16. Sobhian B, Shao G, Lilli DR, Culhane AC, Moreau LA, Xia B, Livingston DM, Greenberg RA: **RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites.** *Science* 2007, **316:**1198-1202.
17. Eddins MJ, Varadan R, Fushman D, Pickart CM, Wolberger C: **Crystal Structure and solution NMR studies of Lys48-linked tetraubiquitin at neutral pH.** *J Mol Biol* 2007, **367:**204-211.
18. Phillips CL, Thrower J, Pickart CM, Hill CP: **Structure of a new crystal form of tetraubiquitin.** *Acta Crystallogr D* 2001, **57(Pt 2):**341-344.
19. Dokmanić I, Sikić M, Tomić S: **Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination.** *Acta Crystallogr D* 2008, **64(Pt 3):**257-263.

Acta Crystallographica Section F
**Structural Biology
and Crystallization
Communications**

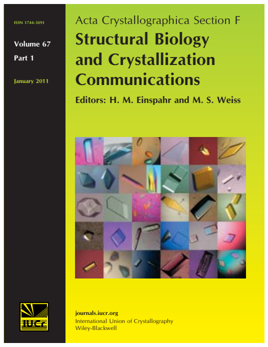
ISSN 1744-3091

Editors: **H. M. Einspahr and M. S. Weiss**

Crystallization and diffraction analysis of the SARS coronavirus nsp10–nsp16 complex

Claire Debarnot, Isabelle Imbert, François Ferron, Laure Gluais, Isabelle Varlet, Nicolas Papageorgiou, Mickaël Bouvet, Julien Lescar, Etienne Decroly and Bruno Canard

Acta Cryst. (2011). F**67**, 404–408



Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Crystallization and diffraction analysis of the SARS coronavirus nsp10–nsp16 complex

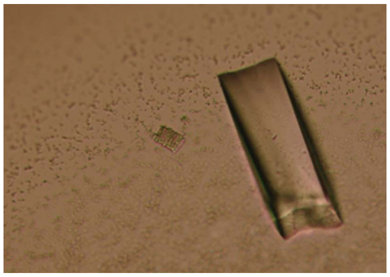
Claire Debarnot,[a] Isabelle Imbert,[a] François Ferron,[a] Laure Gluais,[a] Isabelle Varlet,[a] Nicolas Papageorgiou,[a] Mickaël Bouvet,[a] Julien Lescar,[a,b] Etienne Decroly[a]* and Bruno Canard[a]*

[a]Département de Virologie Structurale, Architecture et Fonction des Macromolécules Biologiques, UMR 6098, 163 Avenue de Luminy, 13288 Marseille CEDEX 09, France, and [b]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore

Correspondence e-mail:
etienne.decroly@afmb.univ-mrs.fr,
bruno.canard@afmb.univ-mrs.fr

Received 19 November 2010
Accepted 21 January 2011

To date, the SARS coronavirus is the only known highly pathogenic human coronavirus. In 2003, it was responsible for a large outbreak associated with a 10% fatality rate. This positive RNA virus encodes a large replicase polyprotein made up of 16 gene products (nsp1–16), amongst which two methyltransferases, nsp14 and nsp16, are involved in viral mRNA cap formation. The crystal structure of nsp16 is unknown. Nsp16 is an RNA-cap AdoMet-dependent (nucleoside-2′-*O*-)-methyltransferase that is only active in the presence of nsp10. In this paper, the expression, purification and crystallization of nsp10 in complex with nsp16 are reported. The crystals diffracted to a resolution of 1.9 Å resolution and crystal structure determination is in progress.



1. Introduction

In 2003, an outbreak of a novel virus occurred in China and spread through several countries. The identified agent, a previously unknown coronavirus, was named severe acute respiratory syndrome coronavirus (SARS-CoV). Much like other coronaviruses, SARS-CoV has an ~30 kb RNA genome with a 5′ cap and a 3′ poly(A) tail (Snijder *et al.*, 2003). The genome of SARS-CoV codes for two large polyproteins 1a and 1ab, which are autoproteolytically processed by at least two viral proteases, producing 16 nonstructural proteins (nsps). These nsps are thought to form a huge protein complex that is responsible for viral RNA replication and transcription. In addition, a set of subgenomic mRNAs encoding structural and accessory proteins is produced by a sophisticated mechanism involving nested mRNA synthesis using the full-length genomic RNA as template. These mRNAs are thought to be capped and polyadenylated (Lai & Stohlman, 1981; van Vliet *et al.*, 2002).

Since 2003, interest in SARS-CoV has greatly promoted structural and functional studies of its nsps. In recent years, many three-dimensional structures of replication proteins have appeared in the literature, including those of nsp1 (Almeida *et al.*, 2007); several domains of nsp3 [including (i) an N-terminal Glu-rich acidic domain (AD; Serrano *et al.*, 2007), (ii) an X domain (XD; Egloff *et al.*, 2006), (iii) the SUD domain (SARS-CoV unique domain; Chatterjee *et al.*, 2009) and (iv) the papain-like protease PLP2 (Ratia *et al.*, 2006)]; nsp5 (Anand *et al.*, 2002); a complex consisting of nsp7–nsp8 (Zhai *et al.*, 2005); the RNA-binding protein nsp9 (Egloff *et al.*, 2004); the zinc-binding protein nsp10 (Joseph *et al.*, 2006, Su *et al.*, 2006); and the hexameric RNA endonuclease nsp15 (Ricagno *et al.*, 2006). In many cases, the crystal structures allowed scientists to suggest or to ascertain a biochemical function for the nsps, such as nsp7–nsp8, which acts as an RNA-dependent RNA primase (Imbert *et al.*, 2006), and the nsp15 endonuclease, the active site of which has been found to share structural homology to the active site of RNAse A (Ricagno *et al.*, 2006; Joseph *et al.*, 2007). The process of RNA capping, however, has not yet benefitted from this structural work. Structural data is lacking for the enzymes that are putatively involved in mRNA capping.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

No. of crystals	1
Beamline	PROXIMA 1 (SOLEIL synchrotron)
Wavelength (Å)	0.979
Detector	ADSC Q315r
Crystal-to-detector distance (mm)	256.35
Rotation range per image (°)	1
Total rotation range (°)	90
Exposure time per image (s)	1
Resolution range (Å)	37.52–2.00 (2.11–2.00)
Space group	$C222_1$
Unit-cell parameters (Å, °)	a = 68.1, b = 184.6, c = 128.8, $\alpha = \beta = \gamma$ = 90.00
Mosaicity (°)	0.097
Total no. of measured intensities	273149
Unique reflections	54947 (7963)
Multiplicity	3.7 (3.7)
Mean $I/\sigma(I)$	7.5 (3.2)
Completeness (%)	99.7 (100)
$R_{\rm merge}$†	0.114 (0.430)
$R_{\rm meas}$	0.136 (0.500)
Overall B factor from Wilson plot (Å^2)	37.3

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations of reflection hkl.

In eukaryotes, mRNA capping results from a series of three to four canonical reactions acting on the 5′-end of the mRNA. The nascent mRNA transcript endures a limited dephosphorylation in which the last 5′-phosphate is removed by a 5′-RNA triphosphatase. A guanylyltransferase (also named capping enzyme) attaches a GMP molecule onto the 5′-diphosphate RNA in a 5′-to-5′ orientation. The capped RNA is then methylated at the N7 position of the cap guanine nucleotide by an N7-guanine methyltransferase. This yields a cap-0 structure ($^{\rm m7}$GpppNN . . .) as found mainly in yeasts and lower eukaryotes. In higher eukaryotes and plants, a second methyltransferase acts on the first transcribed nucleotide at its 2′-O position to yield a cap-1 structure ($^{\rm m7}$GpppN$_{2'{\rm Om}}$N . . .), such as that found on coronavirus mRNAs (Lai & Stohlman, 1981; van Vliet *et al.*, 2002).

We have recently identified the RNA cap 2′-*O*-methyltransferase of the SARS-CoV (Bouvet *et al.*, 2010; Lugari *et al.*, 2010). This activity is harboured by nsp16 as predicted by signature-sequence analysis; however, the enzyme is only active in the presence of nsp10. The latter has been shown to interact strongly with nsp16 using the yeast two-hydrid method, indicating that the enzymes associate into a complex that is stable enough to withstand purification (Imbert *et al.*, 2008).

Here, we report the cloning of nsp10 and nsp16 in a prokaryotic expression vector. Upon expression of both genes, a stable complex consisting of nsp10 and nsp16 (nsp10–nsp16) can be purified and crystallized. We present X-ray diffraction data from these SARS-CoV nsp10–nsp16 crystals.

2. Materials and methods

2.1. Tandem cloning of the nsp10 and nsp16 SARS-CoV genes

The Frankfurt 1 isolate of SARS-CoV (GenBank accession No. AY291315; Thiel *et al.*, 2003) was amplified in Vero-E6 cells and used for production of the nsp10–nsp16 complex as follows: genes coding for SARS-CoV nsp10 (139 amino acids, 14.84 kDa) and nsp16 (298 amino acids, 33.5 kDa) were cloned into the pmCOX *Escherichia coli* dual expression plasmid kindly provided by Dr Bruno Coutard (AFMB, France) containing two separate promoters. In this backbone, SARS-CoV nsp10 is expressed under a *tetA* promoter and encodes a protein fused with an N-terminal *Strep*-Tag (eight amino acids; WSHPQFEK) and nsp16 is expressed under a T7*lac* promoter and encodes a protein fused with an N-terminal hexahistidine tag (Bouvet *et al.*, 2010). This system allows independent regulation of the expression levels of the two genes by the addition of different concentrations of tetracycline and IPTG (isopropyl β-D-1-thiogalactopyranoside).

2.2. Expression and purification of the nsp10–nsp16 complex

SARS-CoV nsp10–nsp16 co-expression was performed in *E. coli* strain C41 (DE3) (Avidis SA, France) harbouring the pLysS plasmid (Novagen). Cultures were grown at 310 K until the OD$_{600\,{\rm nm}}$ reached 0.6. Expression was induced by adding 50 µ*M* IPTG and 200 µg l^{-1} anhydrotetracycline and the cells were incubated for 16 h at 297 K. Bacterial cell pellets were collected by centrifugation at 6000*g*, frozen and resuspended in lysis buffer (50 m*M* HEPES pH 7.5, 300 m*M* NaCl, 5 m*M* MgSO$_4$) supplemented with 10 µg ml^{-1} DNase I. After cell disruption at 100 MPa and 277 K (Constant Cell, UK) and clarification by centrifugation at 20 000*g* for 30 min, the soluble protein fraction was incubated with *Strep*-Tactin Sepharose (IBA BioTAGnology). After three washes in buffer (50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 1 m*M* TCEP, 5 m*M* MgCl$_2$), bound proteins were eluted in wash buffer supplemented with 2.5 m*M* D-desthiobiotin.

2.3. Crystallization and diffraction analysis

Stura Footprint Screen, JCSG+ Screen and Structure Screens I and II (Molecular Dimensions) were tried in CrystalQuick plates with three wells per reservoir (Greiner Bio-One) using 400, 300 and 200 nl drops. The drops contained increasing volumes (100, 200 and 300 nl) of protein solution at a concentration of 5.4 mg ml^{-1} in elution buffer (50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 1 m*M* TCEP, 5 m*M* MgCl$_2$, 2.5 m*M* D-desthiobiotin) and constant volumes of precipitant. A hit condition was observed in 0.1 *M* bicine pH 9, 2 *M* MgCl$_2$ after one week. Crystallization optimization was performed by vapour diffusion using the hanging-drop method at 293 K in Linbro plates (Hampton Research). 2 µl drops (consisting of 1.5 µl protein solution and 0.5 µl precipitant solution) were equilibrated against 1 ml reservoir solution. Optimization of crystal-growth conditions led to the following condition: 0.1 *M* CHES pH 9, 1.52 *M* MgCl$_2$. Crystals appeared after 24 h and visible growth stopped after 48–72 h. Before cooling the crystals to 100 K in a nitrogen-gas stream (Oxford Cryosystems), the crystals were briefly soaked in a cryoprotectant consisting of 15%(*v*/*v*) glycerol added to the mother liquor. Native diffraction data were recorded using either an ADSC Quantum 315 3 × 3 array detector on beamline PROXIMA 1 at the SOLEIL Synchrotron Radiation Facility or an ADSC Quantum 210 2 × 2 array detector on beamline ID14-EH1 at the European Synchrotron Radiation Facility. Data were processed using *XDS* (Kabsch, 2010) and scaled with *SCALA* (Collaborative Computational Project, Number 4, 1994). Crystal parameters and data-collection statistics are summarized in Table 1.

2.4. Cross-linking of the purified nsp10–nsp16 complex

Purified nsp10–nsp16 complex (4 µg) in 50 m*M* HEPES pH 7.5, 150 m*M* NaCl, 5 m*M* MgCl$_2$, 1 m*M* TCEP and 5% glycerol was incubated overnight at 277 K with a solution of suberic acid bis(*N*-hydroxysuccinimide ester) (SAB; Sigma) at a concentration of 0.005%. Samples were then denatured with an equal volume of 2× dissociation buffer [100 m*M* Tris–HCl pH 6.8, 20% glycerol and 200 m*M* DTT, 4% sodium dodecyl sulfate (SDS) and 0.2% bromophenol blue]. After 5 min of heating at 368 K, the proteins were separated and analyzed on an SDS NuPAGE 4–12% gel (Invitrogen).

3. Results and discussion

3.1. Expression, purification and crystallization of a stable nsp10–nsp16 complex

Many attempts in the laboratory to crystallize nsp16 on its own remained unsuccessful. Although significant amounts (on a milligram scale) of nsp16 can be obtained when expressed alone, the protein is unstable in a variety of buffers and precipitates under various storage conditions. Moreover, although signature-sequence analyses unambiguously identified a SAM-dependent methyltransferase fold (von Grotthuss *et al.*, 2003; Decroly *et al.*, 2008), SARS-CoV nsp16 alone does not exhibit this enzymatic activity. In contrast, nsp10 is expressed to high levels in *E. coli* and can readily be purified and crystallized. Two structures of nsp10 crystal forms were published in 2006: one revealed monomers and dimers (Joseph *et al.*, 2006) at 1.8 Å resolution (PDB entry 2fyg), whereas a complex dodecameric structure at 2.1 Å resolution (PDB entry 2g9t) was observed when nsp10 was expressed and crystallized as a fusion with nsp11 (Su *et al.*, 2006). Nevertheless, no function has been either predicted or demonstrated for nsp10, which is a zinc-binding protein.

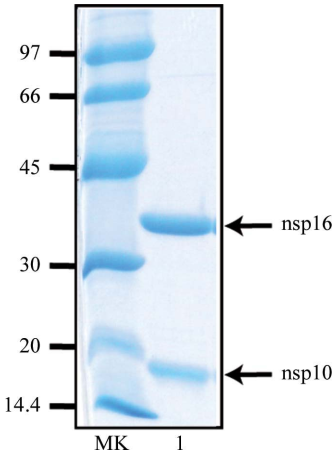


Figure 1
Purification of SARS-CoV nsp10 in complex with nsp16. The purified SARS-CoV nsp10–nsp16 complex was analyzed by 12% SDS–PAGE and stained using Coomassie Blue. Lane MK, molecular-weight markers; lane 1, 2 µg nsp10–nsp16 protein complex eluted from the *Strep*-Tactin column.

We have reported that nsp10 and nsp16 interact *in vitro* when co-expressed in yeast (Imbert *et al.*, 2008) or bacteria (Bouvet *et al.*, 2010). This observation prompted us to co-express the two proteins in the prokaryotic expression vector and to attempt purification of the complex. Nsp10 was cloned under the control of the *tetA* promoter in fusion with a *Strep*-Tag peptide at its N-terminus and nsp16 tagged with a hexahistidine tag at its N-terminus was inserted into the same plasmid under the *T7lac* promoter. After transformation into bacterial strain C41, protein expression was induced by the addition of IPTG and tetracycline for 16 h at 297 K. The bacterial lysate was clarified and nsp10 was adsorbed onto *Strep*-Tactin Sepharose beads. After several washes, the proteins bound to *Strep*-Tactin were eluted with 2.5 m*M* of the biotin analogue desthiobiotin. Upon SDS–PAGE analysis, we detected the presence of *Strep*-nsp10 and His$_6$-nsp16 proteins migrating as expected at around 15 and 35 kDa, respectively, indicating that the two proteins had co-purified (Fig. 1). The overall yield of the purified complex was typically 1 mg per litre of *E. coli* culture. Estimation of protein concentration and normalization with regard to molecular mass indicated an approximate 1:1 ratio of the proteins.

3.2. Characterization of the nsp10–nsp16 complex

The nsp10–nsp16 complex was further characterized before attempting to identify crystallization conditions. We first confirmed the identity of each recombinant protein by matrix-assisted laser-

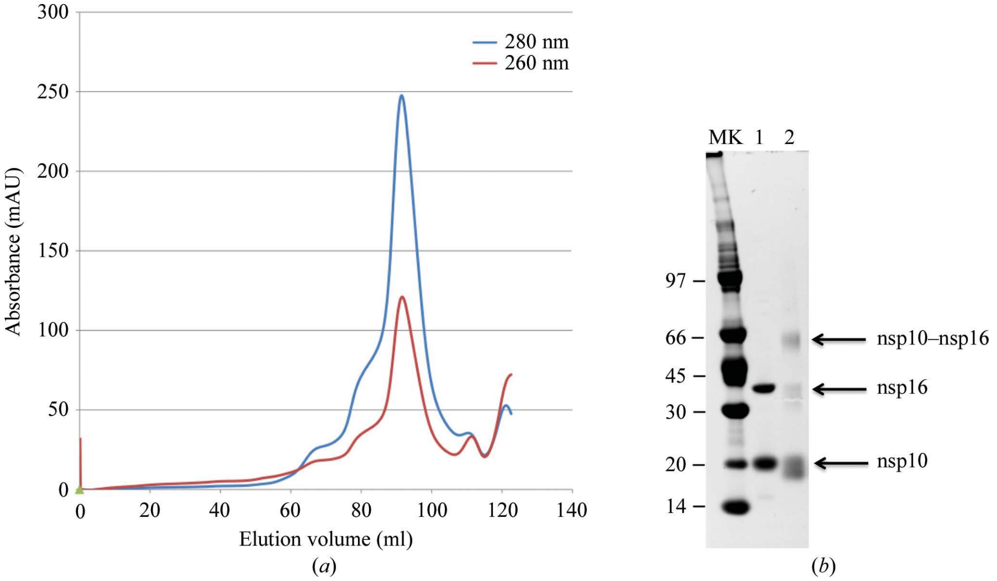


Figure 2
Characterization of SARS-CoV nsp10 in complex with nsp16. (*a*) Gel-filtration chromatogram of the SARS-CoV nsp10–nsp16 complex. The nsp10–nsp16 complex eluted from the *Strep*-Tactin column was analyzed on a 16/60 S200 gel-filtration column and the elution of protein and nucleic acid was followed by measuring the absorption at 280 nm (blue) and 260 nm (orange), respectively. The main peak eluting after 90 ml corresponds to elution of a 50 kDa protein. (*b*) Cross-linking experiment. The purified SARS-CoV nsp10–nsp16 complex was loaded onto a 4–12% NuPAGE gel and stained using Coomassie Blue. Lane MK, molecular-weight markers; lane 1, 4 µg non-cross-linked nsp10–nsp16 complex, lane 2, 4 µg of the nsp10–nsp16 complex incubated overnight at 277 K with 0.005% SAB cross-linker.

desorption ionization time-of-flight mass spectrometry after in-gel trypsin digestion (data not shown). We also analyzed the proteins eluted from the *Strep*-Tactin column by size-exclusion chromatography. Fig. 2(*a*) shows the gel-filtration elution profile. We observed that the stable nsp10–nsp16 complex elutes as a single peak corresponding to 50 kDa. The interaction between nsp10 and nsp16 was also confirmed by a cross-linking experiment. For this purpose, the nsp10–nsp16 complex was incubated with 0.005% SAB, a cross-linking reagent that specifically reacts with lysine residues. After SDS–PAGE separation, we detected the bands corresponding to nsp10 and nsp16 monomers as well as an additional band corresponding to a nsp10–nsp16 complex migrating around 50 kDa (lane Xtal in Fig. 3*b*). The presence of nsp16 in the complex was also demonstrated by a methyltransferase assay (Selisko *et al.*, 2010) which showed that the purified fraction containing the nsp10–nsp16 complex exhibited 2′-*O*-methyltransferase activity on short N7-methylated capped RNA substrates (Bouvet *et al.*, 2010). Together, these results indicate that we have developed a method allowing the production and purification of an active nsp10–nsp16 complex by a one-step procedure. This protein preparation was used without additional treatment for crystal growth.

3.3. Crystal growth and data collection

The complex was initially crystallized using 0.1 *M* bicine pH 9, 2 *M* $MgCl_2$ (condition E2 fom Structure Screens I and II). Crystal growth was then optimized and the best condition was finally chosen as 0.1 *M* CHES pH 9, 1.52 *M* $MgCl_2$. Crystals typically appeared overnight and visible growth stopped after 48–72 h (Fig. 3*a*). To confirm that the crystal contained the nsp10–nsp16 complex, we collected ten crystals,

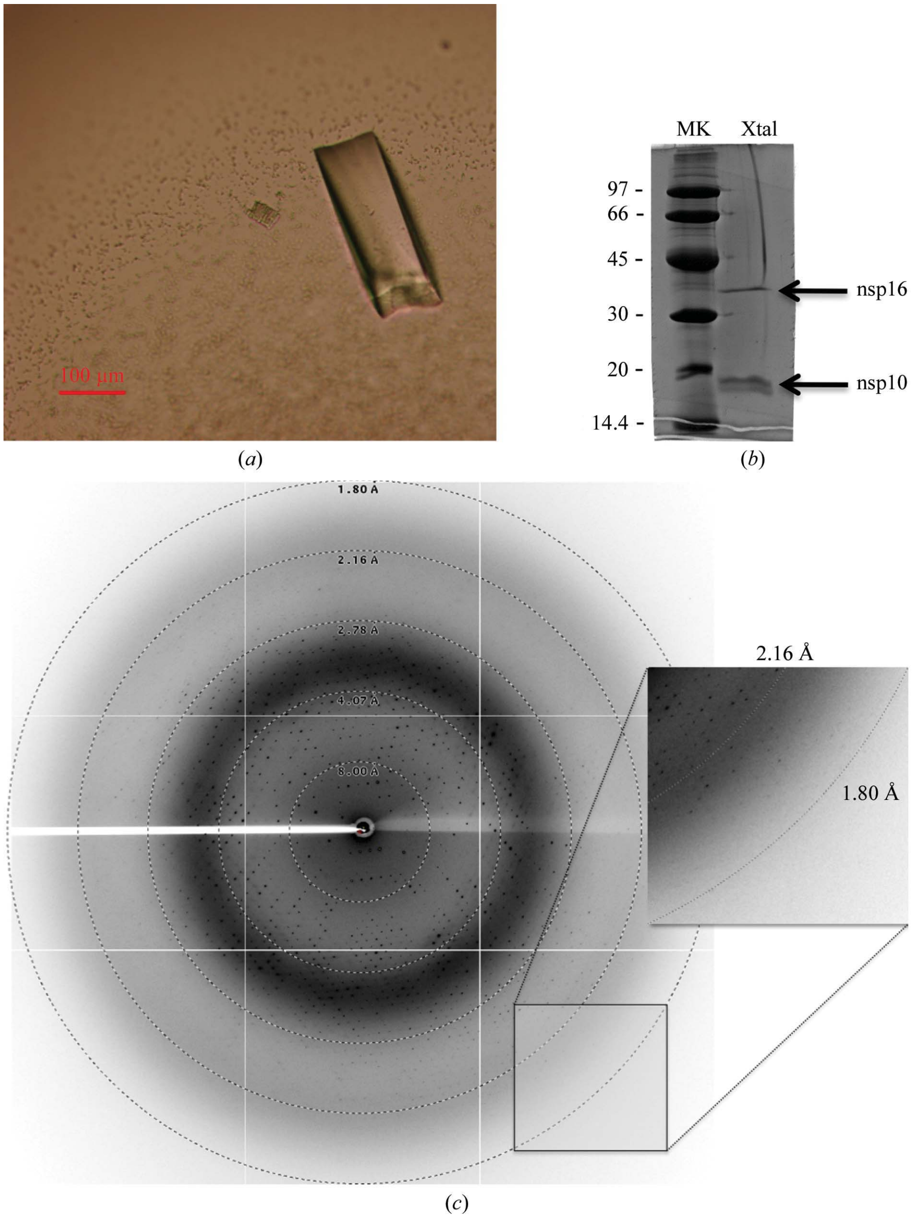


Figure 3
(*a*) Optimized crystal of the SARS-CoV nsp10–nsp16 complex. The scale bar is 100 µm in length. (*b*) NuPAGE analysis of the nsp10–nsp16 complex from optimized crystals. Ten optimized crystals were loaded onto a 4–12% NuPAGE gel and stained using Coomassie blue. Lane MK, molecular-weight markers; lane Xtal, nsp10–nsp16 complex. (*c*) X-ray diffraction pattern from a crystal of the SARS-CoV nsp16–nsp10 protein complex. Resolution arcs are shown. Reflections are observed to below 2 Å (an enlargement is shown in the inset).

which were analyzed by SDS–PAGE after two washes. Fig. 2(*b*) shows that both proteins, nsp10 and nsp16, were detected on Coomassie Blue staining. This confirms that the crystals consist of the nsp10–nsp16 complex with a similar stoichiometry as observed prior to crystallization.

Crystals were flash-cooled in the same buffer supplemented with 15%(*v*/*v*) glycerol before exposure to synchrotron X-ray radiation. A typical diffraction image is shown in Fig. 3(*c*), in which reflections are visible to 2 Å resolution. Data integration and reduction indicated that the crystals belonged to space group $C222_1$, with one complex per asymmetric unit (Table 1). The crystals (space group $C222_1$, unit-cell parameters $a = 68.1$, $b = 184.6$, $c = 128.8$ Å) contained one nsp10–nsp16 complex per asymmetric unit, with a solvent content of 70% and a $V_{\rm M}$ value of 4.17 Å^3 Da^{-1}.

4. Conclusions

We have crystallized a complex of the SARS-CoV nsp10 and nsp16 proteins. The presence of nsp10 in the complex confers 2′-*O*-methyltransferase activity on nsp16. The crystal structure of nsp10 is known, whereas that of nsp16 is not. Given the quality of the crystals described here and the fact that nsp10 is a zinc-binding protein, it should be possible to determine the crystal structure using molecular-replacement techniques merged with phases obtained by SAD studies, taking advantage of the phasing power of the Zn atoms.

This work was supported in its initial phase by the VIZIER integrated project (LSHG-CT-2004-511960) of the European Union Sixth Framework and then by the French National Research agency under reference ANR-08-MIEN-032, the Fondation pour la Recherche Médicale (Programme équipe FRM) to BC and the Direction Générale de l'Armement (contract 07co404). MB has a fellowship from the Direction Générale de l'Armement.

References

Almeida, M. S., Johnson, M. A., Herrmann, T., Geralt, M. & Wüthrich, K. (2007). *J. Virol.* **81**, 3151–3161.

Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). *EMBO J.* **21**, 3213–3224.

Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E. J., Canard, B. & Decroly, E. (2010). *PLoS Pathog.* **6**, e1000863.

Chatterjee, A., Johnson, M. A., Serrano, P., Pedrini, B., Joseph, J. S., Neuman, B. W., Saikatendu, K., Buchmeier, M. J., Kuhn, P. & Wüthrich, K. (2009). *J. Virol.* **83**, 1823–1836.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Decroly, E., Imbert, I., Coutard, B., Bouvet, M., Selisko, B., Alvarez, K., Gorbalenya, A. E., Snijder, E. J. & Canard, B. (2008). *J. Virol.* **82**, 8071–8084.

Egloff, M. P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E. J., Gorbalenya, A. E., Cambillau, C. & Canard, B. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 3792–3796.

Egloff, M. P., Malet, H., Putics, A., Heinonen, M., Dutartre, H., Frangeul, A., Gruez, A., Campanacci, V., Cambillau, C., Ziebuhr, J., Ahola, T. & Canard, B. (2006). *J. Virol.* **80**, 8493–8502.

Grotthuss, M. von, Wyrwicz, L. S. & Rychlewski, L. (2003). *Cell*, **113**, 701–702.

Imbert, I., Guillemot, J. C., Bourhis, J. M., Bussetta, C., Coutard, B., Egloff, M. P., Ferron, F., Gorbalenya, A. E. & Canard, B. (2006). *EMBO J.* **25**, 4933–4942.

Imbert, I., Snijder, E. J., Dimitrova, M., Guillemot, J. C., Lécine, P. & Canard, B. (2008). *Virus Res.* **133**, 136–148.

Joseph, J. S., Saikatendu, K. S., Subramanian, V., Neuman, B. W., Brooun, A., Griffith, M., Moy, K., Yadav, M. K., Velasquez, J., Buchmeier, M. J., Stevens, R. C. & Kuhn, P. (2006). *J. Virol.* **80**, 7894–7901.

Joseph, J. S., Saikatendu, K. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C. & Kuhn, P. (2007). *J. Virol.* **81**, 6700–6708.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Lai, M. M. & Stohlman, S. A. (1981). *J. Virol.* **38**, 661–670.

Lugari, A., Betzi, S., Decroly, E., Bonnaud, E., Hermant, A., Guillemot, J. C., Debarnot, C., Borg, J. P., Bouvet, M., Canard, B., Morelli, X. & Lécine, P. (2010). *J. Biol. Chem.* **285**, 33230–33241.

Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, N., Baker, S. C., Stevens, R. C. & Mesecar, A. D. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 5717–5722.

Ricagno, S., Egloff, M. P., Ulferts, R., Coutard, B., Nurizzo, D., Campanacci, V., Cambillau, C., Ziebuhr, J. & Canard, B. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 11892–11897.

Selisko, B., Peyrane, F. F., Canard, B., Alvarez, K. & Decroly, E. (2010). *J. Gen. Virol.* **91**, 112–121.

Serrano, P., Johnson, M. A., Almeida, M. S., Horst, R., Herrmann, T., Joseph, J. S., Neuman, B. W., Subramanian, V., Saikatendu, K. S., Buchmeier, M. J., Stevens, R. C., Kuhn, P. & Wüthrich, K. (2007). *J. Virol.* **81**, 12049–12060.

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J. & Gorbalenya, A. E. (2003). *J. Mol. Biol.* **331**, 991–1004.

Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R., Joachimiak, A., Zhang, X. C., Bartlam, M. & Rao, Z. (2006). *J. Virol.* **80**, 7902–7908.

Thiel, V., Ivanov, K. A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E. & Ziebuhr, J. (2003). *J. Gen. Virol.* **84**, 2305–2315.

Vliet, A. L. van, Smits, S. L., Rottier, P. J. & De Groot, R. J. (2002). *EMBO J.* **21**, 6571–6580.

Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M. & Rao, Z. (2005). *Nature Struct. Mol. Biol.* **12**, 980–986.

Antiviral Research 86 (2010) 296–305

Contents lists available at ScienceDirect

Antiviral Research

journal homepage: www.elsevier.com/locate/antiviral

Dengue virus replicons: Production of an interserotypic chimera and cell lines from different species, and establishment of a cell-based fluorescent assay to screen inhibitors, validated by the evaluation of ribavirin's activity

Nicolas Massé[a], Andrew Davidson[b], François Ferron[a], Karine Alvarez[a], Mike Jacobs[c], Jean-Louis Romette[a], Bruno Canard[a], Jean-Claude Guillemot[a,*]

[a] Architecture et Fonction des Macromolécules Biologiques, CNRS and Universités d'Aix-Marseille I et II, UMR 6098, ESIL Case 925, 13288 Marseille, France
[b] Department of Cellular and Molecular Medicine, School of Medical Sciences, University of Bristol, Bristol, UK
[c] Department of Infection, University College London Medical School, London, UK

ARTICLE INFO

Article history:
Received 9 September 2009
Received in revised form 8 March 2010
Accepted 16 March 2010

Keywords:
Dengue virus
Replicon
HTS
NS5
Antiviral

ABSTRACT

The prevention and treatment of flavivirus infections are public health priorities. Dengue fever is the most prevalent mosquito-borne viral disease of humans, affecting more than 50 million people annually. Despite the urgent need to control dengue infections, neither specific antiviral therapies nor licensed vaccines exist and the molecular basis of dengue pathogenesis is not well understood. In this study we produced a novel dengue virus type 2 (DV2) subgenomic replicon that expresses a fusion protein comprised of Enhanced Green Fluorescent Protein (EGFP) and Puromycin N-Acetyltransferase (PAC). We successfully established BHK, COS and Huh7 cell lines that stably expressed the DV2 replicon. Using EGFP as a reporter of DV replication complex activity, we set up a new HTS assay. The assay was validated using the inhibitor ribavirin, confirmed by flow cytometry analysis and the analysis of NS5 expression by Western-blot analysis. In order to develop a system to test antivirals against the NS5 proteins of all four DV serotypes in a similar cellular environment, the replicon was further modified, to allow easy exchange of the NS5 gene between DV serotypes. As proof of principle, a chimeric replicon in which the DV2 NS5 gene was substituted with that of DV type 3 was stably expressed in BHK cells and used in ribavirin inhibition studies. The assays described in this study will greatly facilitate DV drug discovery by serving as primary or complementary screening. The approach should be applicable to the development of fluorescent cell-based HTS assays for other flaviviruses, and useful for the study of many aspects of DV, including viral replication and pathogenesis.

1. Introduction

Dengue fever is the most prevalent mosquito-borne viral disease of humans, affecting more than 50 million people annually. Dengue virus (DV) is a member of the Flaviviridae family together with other important pathogens such as yellow fever, West Nile, and Japanese encephalitis viruses. The four genetically related but serologically distinct serotypes of DV (types 1–4) cause dengue fever (DF), dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS). Despite the urgent need to control dengue infections, neither specific antiviral therapies nor licensed vaccines exist, and the molecular basis of dengue pathogenesis is not well understood. DV has a capped positive-sense single stranded RNA genome containing a single open reading frame (ORF), flanked by 5′ and 3′ untranslated regions (UTR). The ORF encodes a polyprotein precursor, which is subsequently cleaved by cellular and viral proteases into three structural proteins (capsid (C), premembrane (prM) and envelope (E)) as well as seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5). The non-structural proteins assemble with cellular proteins to form a replication complex, where the viral RNA is synthesized (Lindenbach et al., 2007). Two of the 10 viral proteins have key enzymatic activities and are thus potential targets of inhibitors. NS3 acts as a serine protease (with NS2B as a cofactor), 5′RNA triphosphatase, nucleotide phosphatase, and helicase (Lescar et al., 2008). NS5 is a bifunctional protein: its amino-terminal domain carries a methyltransferase (MTase) involved in methylation of the 5′ cap structure of genomic RNA (Dong et al., 2008), whereas its carboxy-terminal domain is the RNA-dependent RNA polymerase (RdRp) responsible for the synthesis of viral RNA (Malet et al., 2008).

Abbreviations: DV, dengue virus; DV2, DV serotype 2; DV3, DV serotype 3; EGFP, Enhanced Green Fluorescent Protein; HTS, high throughput screening; IRES, internal ribosome entry site; MTase, methyltransferase; NS5, non-structural protein 5; PAC, Puromycin N-Acetyltransferase; RdRp, RNA-dependent RNA polymerase.

[*] Corresponding author. Tel.: +33 4 9182 8630; fax: +33 4 9182 8646.
E-mail address: Jean-Claude.Guillemot@afmb.univ-mrs.fr (J.-C. Guillemot).

0166-3542/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.antiviral.2010.03.010

To control the spread of DV, new tools are required, both to improve our understanding of the replication complex and provide a cost-effective means of screening antiviral compounds. Subgenomic replicon constructs contain all of the genetic elements needed for genome amplification in permissive cells but lack the major part of the genes encoding the structural proteins. The RNAs consequently replicate but are not packaged into viral particles. Subgenomic replicon can also encode drug selectable markers and/or reporter proteins, that may substitute for the structural protein coding regions. As they are non-infectious, subgenomic replicons are important tools for the study of viral replication, the screening of antiviral compounds and the testing of their inhibitory potential (Khromykh et al., 2001; Lo et al., 2003a,b; Tilgner and Shi, 2004; Alvarez et al., 2005, 2008; Liu et al., 2005; Rossi et al., 2005, 2007; Tilgner et al., 2005; Filomatori et al., 2006; Gu et al., 2006; Ng et al., 2007; Noueiry et al., 2007). They avoid problems associated with the use of infectious virus and are particularly important for the study of BL-3 pathogens, like DV.

In this study, we established stable cell lines expressing a novel DV type 2 (DV2) replicon. In place of the DV structural genes, the replicon contained a gene construct encoding a fusion protein (EGFF–PAC) comprised of the Enhanced Green Fluorescent Protein (EGFP) and Puromycin N-Acetyltransferase (PAC). Cells containing the replicon could be selected by puromycin treatment, whilst EGFP could be used as a reporter to follow the replication level of the subgenomic replicon. First, BHK cells expressing the DV2 replicon were established. The replicon systems were then extended for the first time to the COS monkey cell line and a human hepatic cell line (Huh7). Extension of the replicon system to COS cells is important as these cells are not easily naturally infected (unpublished observations (Rodrigo et al., 2006)). The DV2 replicon Huh7 stable cell line constitutes a suitable tool for studying the interactions between viral proteins and hepatic proteins, some of which may be responsible for the infection-induced liver damage (Ekkapongpisit et al., 2007; Pattanakitsakul et al., 2007; Higa et al., 2008).

Second, we set up a screening assay of inhibitors, based on the use of EGFP as a readout. To validate the assay, we used ribavirin, a broad spectrum antiviral molecule, which inhibits DV replication (Koff et al., 1982; Crance et al., 2003; Takhampunya et al., 2006). The inhibitory effect of ribavirin observed in the DV2 replicon cell lines, was similar to that described using infected cells, establishing a new fluorescent HTS assay in a BL-2 environment. Moreover, the establishment of multiple cell lines carrying the DV replicon enables evaluation of the inhibitory potential of selected compounds in cell lines from different organs and species.

Finally, with the aim to test antivirals against the NS5 proteins of the four DV serotypes and potential NS5 mutants in a similar cellular environment, we designed an original method. We modified the DV2 replicon to allow easy exchange of a NS5 gene cassette. We validated the construct by introducing the DV type 3 (DV3) NS5 gene into the DV2 replicon. The chimeric replicon construct was shown to be viable as it was possible to establish a stable cell line expressing the DV2/3-NS5 replicon. In addition to demonstrating that a replication complex containing the NS1 to NS4B proteins from DV2 and the NS5 protein from DV3 is functional, the chimeric replicon can be used as a tool to test the effectiveness of antivirals specifically against the MTase and the RdRp of the different DV serotypes in a cellular context.

2. Materials and methods

2.1. Cells, media and reagents

Cells were grown in Dulbecco's modified eagle medium (DMEM) (PAA) supplemented with penicillin, streptomycin and 5% (BHK cells) or 10% (COS and Huh7 cells) foetal calf serum (FCS). For BHK-, COS- and Huh7-DV EGFP-replicon containing cells, puromycin (Sigma) was added at 3.5, 2 and 1 μg/ml, respectively. Replicon replication inhibition tests were done using a medium without phenol red (Invitrogen), supplemented with penicillin, streptomycin, 2 mM L-glutamine, 1 mM sodium pyruvate and 5% FCS, without puromycin. Ribavirin (Sigma) was resuspended in 100% DMSO at 20 mM and stored at −20 °C.

2.2. Replicon plasmid construction

2.2.1. DV2 (new Guinea C strain) EGFP-PAC replicon plasmid

A DV2 replicon expressing an EGFP-PAC fusion protein was constructed using the replicon pDENΔCprME-PAC2A (Jones et al., 2005) that was previously derived from the genome length DV2 (new Guinea C strain) cDNA clone pDVWS601 (Gualano et al., 1998; Pryor et al., 2001). Initially, the EGFP gene was amplified from pEGFP (Clontech) by PCR, using primers P1 and P2 (all primers are listed in Table 1) and cloned into pCR-Blunt-II-TOPO (Invitrogen) resulting in the plasmid pCR-EGFP. The PAC-1D2A gene cassette was then amplified from pDENΔCprME-PAC2A by PCR as two overlapping fragments (in order to replace an internal XbaI site with a BglII site) using primers P3, P4, P5 and P6 respectively. The PAC and 1D2A gene fragments were then joined by overlap PCR using primers P3 and P6 and the resulting fragment cloned into pCR-Blunt-II-TOPO to produce the plasmid pCR-PAC-1D2A. The PAC-1D2A gene fragment was excised from pCR-PAC-1D2A with NdeI and BamHI and cloned into the corresponding sites of pCR-EGFP resulting in the plasmid pCR-EGFP-PAC-1D2A. The EGFP and PAC genes are separated by a 21 nt linker containing the restriction enzyme sites BglII, SphI and NdeI. The EGFP-

Table 1
List and sequences of the primers used in this study.

Primer	Complete name	Sequence 5′-3′
P1	P1_Sbf_5-GFP	TCGACTGTACAAACTCCTGCAGGTACGCGTACCATGGTGAGCAAGGGCGAGGAGCTG
P2	P2 GFP-3 2r	GTCCATATGGCATGCAGATCTTCGGTACAGCTCG
P3	P3_PAC/G-3	GCATGCCATATGGACATGACCGAGTACAAGCCCACGGTGCGCCTC
P4	P4 PAC/B-2r	CAGATCTGGCACCGGGCTTGCGGGTCATGCACCAG
P5	P5 1D2A/PAC	CCGCAAGCCCGGTGCCAGATCTGTCACCGAGTTGCTTTACCGG
P6	P6_1D2A/Sna/Pme	GTTTAAACGCTACGTACGGGCCCAGGGTTGGACTCGACGTC
P7	P7 F attBl DV2 largeNS5	GGGGACAAGTTTGTACAAAAAAGCAGGCTGCAAGTATAGCGGCTAGAGG
P8	P8_R_attB2_DV2_largeNS5	GGGGACCACTTTGTACAAGAAAGCTGGGTGGTTGAAGGCTCTCAAGGGCATCG
P9	P9 F pDON-DV2-largeNS5+AgeI	CGAGAAGGGGAACCGGTAACATAGGAGAGACG
P10	P10_R_pDON-DV2-largeNS5+AgeI	CGTCTCTCCTATGTTACCGGTTCCCCTTCTCG
P11	P11_ F_pDON-DV2-largeNS5+ScaI	GGAAGAGGCAGGAGTaCTGTGGTAGAAGGC
P12	P12_R_pDON-DV2-largeNS5+ScaI	GCCTTCTACCACAGtACTCCTGCCTCTTCC
P13	P13 F attBl DV3 NS5	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGGAACCGGTTCACAAGGTGAAACCTTAGGAGA
P14	P14_R_attB2_DV3_NS5	GGGGACCACTTTGTACAAGAAAGCTGGGTCCACAGTACTCCCTCTGACTCCTCCTCCTTCCTGA
P15	P15 F_pDON-DV2/3-largeNS5 VLtoAI	CAGAGGGAGCAATTTGGTAAAAGGCAAAACTAACATG
P16	P16 R_pDON-DV2/3-largeNS5 VLtoAI	CATGTTAGTTTTGCCTTTTACCAAATTGCTCCCTCTG

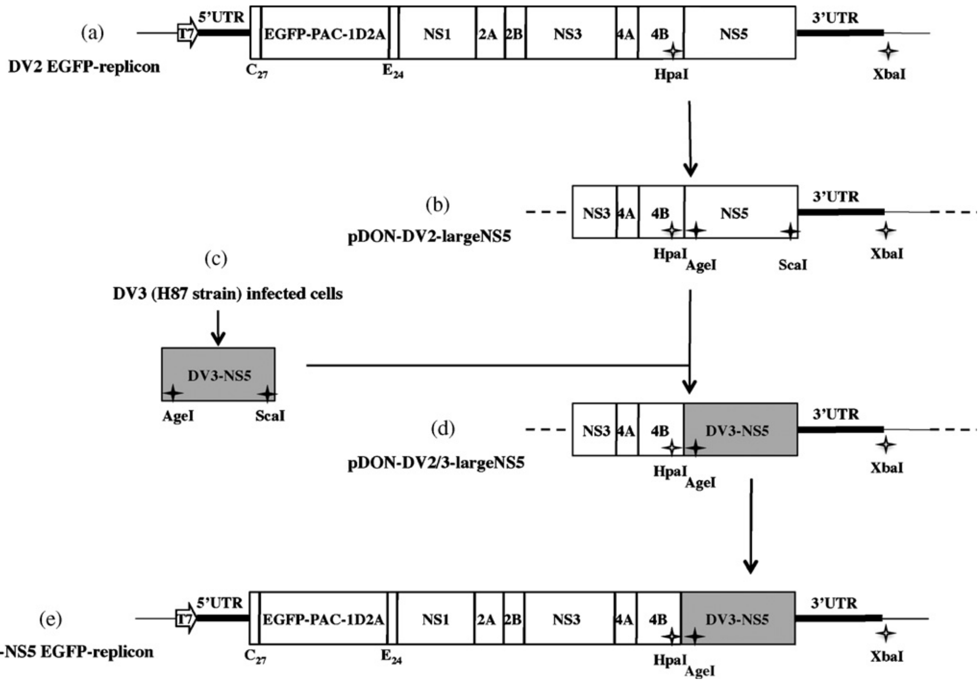


Fig. 1. Schematic representation of the subgenomic DV2 and DV2/3-NS5 EGFP-replicon plasmids. (a) From the full length DV2 NGC infectious cDNA clone pDVWS601, a large part of the C-prM-E coding sequence was deleted, leaving sequence encoding the N-terminal 27 amino acids of the C protein (C_{27}) and the C-terminal 24 amino acids of the E protein (E_{24}). The deleted sequence was replaced by an EGFP-PAC-1D2A gene fusion. This replicon is under the control of a T7 promoter. (b) The pDON-DV2-largeNS5 plasmid was obtained by recombination of the pDON201 vector with a PCR product obtained from the DV2 EGFP replicon plasmid. (c) The DV3 (H87 strain) NS5 sequence was obtained by RT-PCR. (d) The pDON-DV2/3-largeNS5 plasmid was obtained by exchange of the NS5-containing AgeI/ScaI fragments between the DV3 PCR product and the pDON-DV2-largeNS5. (e) The DV2/3-NS5 EGFP-replicon was obtained by exchange of the NS5-containing HpaI/XbaI fragments between the pDON-DV2/3-largeNS5 and the DV2 EGFP replicon. For all plasmids, white and black stars indicate restriction enzyme sites initially present or added by PCR, respectively.

PAC-1D2A gene fusion was excised from pCR-EGFP-PAC-1D2A by digestion with BsrGI and SnaBI and cloned into the corresponding sites of pDENΔCprME-PAC2A resulting in the production of the replicon plasmid pDENΔCprME-EGFP-PAC-1D2A (hereafter termed DV2 EGFP-replicon). The replicon plasmid contains a T7 promoter upstream of the DV sequence (such that transcription adds a single G nt at the 5′ end of the DV 5′UTR) and retains DV2 sequences coding for the first 27 amino acids of the C protein (including the first methionine) and the last 24 amino acids of the E protein. The remainder of the sequence coding for the viral structural proteins was replaced by the EGFP-PAC-1D2A gene fusion (Fig. 1a).

2.3. Construction of a DV2/3-NS5 EGFP-replicon plasmid

Using the DV2 EGFP-replicon plasmid as a template, a sequence including the NS5 gene (extending from the middle of the NS3 gene to the plasmid sequence downstream of the viral 3′UTR) and containing two pre-existing unique enzyme restriction sites near the 5′ and 3′ termini of NS5 (HpaI and XbaI, respectively) was amplified by PCR with primers P7 and P8. They contained attB1 and attB2 sequences, allowing recombination with the pDON201 vector using the Gateway® (Invitrogen) cloning technology. Unique enzyme restriction sites were then added by site-directed mutagenesis at the exact 5′ (AgeI site, with primers P9 and P10) and exact 3′ (ScaI site, with primers P11 and P12) termini of the NS5 gene resulting in the production of the plasmid pDON-DV2-largeNS5 plasmid (Fig. 1b).

The NS5 gene of DV3 (H87 strain) was amplified by RT-PCR using RNA from DV3-infected cells, with primers P13 and P14 introducing the AgeI and ScaI sites at the exact 5′ and exact 3′ termini (Fig. 1c). Primers P13 and P4 also contained the attB1 and attB2 sequences, allowing recombination with a pDON201 vector. By restriction (with AgeI and ScaI) and ligation (with T4 ligase), the DV2 NS5 sequence was removed from the pDON-DV2-largeNS5 plasmid and replaced with the DV3 NS5 sequence, producing the plasmid pDON-DV2/3-largeNS5 (Fig. 1d). The amino acid substitutions created by the introduction of the ScaI site into the DV3 NS5 sequence was restored (Val.Leu to Ala.Ile) by site-directed mutagenesis (primers P15 and P16).

The HpaI-XbaI fragment from either the pDON-DV2-largeNS5 plasmid (control) or the pDON-DV2/3-largeNS5 plasmid were then used to replace the equivalent fragment in the DV2 EGFP-replicon plasmid. The new EGFP-replicon plasmids were named DV2-NS5 and DV2/3-NS5, respectively (Fig. 1e). All plasmids were sequenced. pDON201-derived plasmids and replicon plasmids were amplified in *E. coli* DH5α at 37 °C and Stbl2 at 30 °C, respectively.

2.4. In vitro RNA transcription and transfection

DV2, DV2-NS5 and DV2/3-NS5 EGFP-replicon plasmids were linearized by XbaI, extracted with phenol/chloroform and precipitated with ethanol. The linearized plasmids were used as templates (2.5 μg in a total reaction volume of 50 μl) for the production of replicon RNAs in the presence of 6 m ^{m7}GpppA cap analog (NEBiolabs), with the T7-dependent MEGAscript® kit from Ambion. Replicon RNA was purified with the RNeasy® kit (Qiagen). Replicon RNA (2 μg) was transfected with the TransMessenger® kit (Qiagen) into BHK and COS cells previously plated in a 6-well plate. Huh7 cells (8×10^6 cells resuspended in 800 μl of cytomix buffer (Liang et al., 2005) were electroporated with 3 μg of replicon RNA (1 pulse, 950 μF, 270 V). After 48 h, transfected and electroporated cells were selected for 7 days with puromycin, at the concentrations described above. Cells expressing high levels of EGFP were sorted with a Becton-Dickinson FACSVantage^{TM} and then propagated in the presence of puromycin.

2.5. Inhibition of the replication of the replicon by ribavirin

Two methods were developed to assess the effects of compounds on the replication of the DV replicon based on either 24 or 96 well formats. Control cells, DV2 and DV2/3-NS5 EGFP-replicon cells were seeded either in 24-well plates or in black 96-well plates at densities of 5×10^4 or 1×10^4 cells per well, respectively, in complete uncolored medium supplemented with 0.5% DMSO (v/v). After 24 h, the medium was removed and cells were incubated with 400 μl (24-well plate) or 100 μl (96-well plate) of the same medium containing a range of ribavirin concentrations (1 μM to 100 μM). In each 96-well plate, incubations were done in triplicate and 6 control wells were also included. The media were renewed after 24 h, as cells were incubated with ribavirin for a total of 48 h.

For the 96-well plate assay, the inhibitory effect of ribavirin was defined according to its cytotoxic effect as follows. In addition to the ribavirin treated and control cells, a range of cell numbers, from 1×10^4 to 8×10^4 were freshly seeded, in triplicate, into each 96 well plate and incubated for 5 to 6 h. The media was then removed from all wells and replaced by 100 μl of PBS. The EGFP fluorescence from each well was read with a Tecan SafireII® fluorimeter at 490 nm (excitation) and 510 nm (emission). Then 20 μl of Celltiterblue® reagent (Promega) was added per well, the plates incubated for a further 75 min at 37 °C and 5% CO_2 and the fluorescence read at 560 nm (excitation) and 590 nm (emission). A 590 nm fluorescence curve produced using the range of cell numbers was then used to define an equation to calculate the number of cells present in each well. The 510 nm fluorescence values were divided by the calculated cell number and mean values for each point of concentration were reported as a percentage of the mean control value.

For the 24 well based assay, the EGFP-fluorescence intensity of the cells in each well was analyzed by flow cytometry as described below. Values for each ribavirin concentration were reported as a percentage of the control value.

2.6. Protein expression analysis

2.6.1. Flow cytometry

Cells were washed with PBS and trypsinized. Pelleted cells were washed twice with PBS containing 2% FCS and the EGFP-fluorescence intensity was analyzed by flow cytometry with a Becton-Dickinson FACScan™.

2.7. Western-blot

Total proteins were extracted with cell lysis buffer (50 mM Tris Cl pH 7.4, 150 mM NaCl, 0.5% Na deoxycholate, 1% NP40, 100 μg/ml PMSF, protease inhibitor cocktail from Sigma®) and loaded on a SDS-PAGE (10%) gel. Purified recombinant DV2 NS5 protein (Selisko et al., 2006) was also loaded as a control. Proteins were transferred to a PVDF membrane (Millipore), which was then blocked with PBS/0.1% Tween20/5% milk, and then incubated with primary and secondary antibodies. The following antibodies were used; a mouse monoclonal anti-GAPDH coupled to peroxidase (clone GAPDH-71.1, from Sigma®, diluted at 1/25,000), a mouse monoclonal anti-EGFP (clone JL-8, from Clontech, diluted at 1/4000), a rat monoclonal anti-NS3 helicase domain (culture supernatant of the clone 31F9.11, diluted at 1/10, produced in house), a rat monoclonal anti-NS5 RdRp domain (clone 19A8.2, diluted at 1/500, produced in-house), a goat anti-rat IgG (Jackson Immunoresearch, diluted at 1/10,000) and a goat anti-mouse IgG (from Sigma, diluted at 1/4000), both coupled with horse-radish peroxidase. Western-blots were detected by chemiluminescence with ECL (Amersham™) and Biomax Light Films (Kodak).

3. Results

3.1. Construction of the DV2- and chimeric DV2/3-NS5 EGFP-replicon plasmids

As an initial step towards the development of a cell-based fluorescent screening assay for DV inhibitors, a DV2 (New Guinea C strain) replicon expressing the drug selectable marker puromycin N-acetyl-transferase (PAC) gene (Jones et al., 2005) was modified to express an EGFP-PAC gene fusion (Fig. 1a). Previous DV replicons that express both a reporter gene and a drug selectable marker gene have used internal ribosome entry sites (IRES) to initiate translation of the DV non-structural proteins (Puig-Basagoiti et al., 2006; Ng et al., 2007). However, translation from IRES sequences can be much less efficient than cap-dependent translation (Mizuguchi et al., 2000; Ibrahimi et al., 2009) and influenced by positional (Hennecke et al., 2001) and cell type specific effects (Hellen and Sarnow, 2001). Therefore, we expressed the Enhanced Green Fluorescent reporter Protein (EGFP) as a fusion with the PAC gene product. The 1D2A peptide derived from foot-and-mouth disease virus was engineered after the PAC gene. The 1D2A peptide induces a translation ‘skip’ (Doronina et al., 2008) which served to separate the EGFP-PAC-1D2A fusion protein from the remainder of the DV polyprotein. The EGFP-PAC-1D2A gene fusion replaced the major part of the DV2 structural genes. The sequence coding for the first 27 amino acids of the C protein was retained, as it contains the 5′ cyclization sequence (5′CS) important for replication (You and Padmanabhan, 1999; Khromykh et al., 2001; Alvarez et al., 2005). In addition, the sequence coding for the last 24 amino acids of the E protein were retained because the C-terminal uncharged hydrophobic sequence of the E protein acts as a signal sequence for translocation of NS1 across the endoplasmic reticulum (Rice et al., 1985; Falgout et al., 1989).

With the aim to test inhibitors against the NS5 protein of the different DV serotypes in a cellular environment, we established a DV2 replicon carrying an NS5 gene cassette. To produce chimeric DV EGFP replicon plasmids, we optimised a method designed to avoid the generation of recombination events, frequently observed when working with long viral sequences. We produced a plasmid (pDON-DV2-largeNS5) that can be used to reliably exchange the NS5 sequence in the DV2 EGFP-replicon plasmid in two cloning steps. The DV3 (H87 strain) NS5 sequence was first amplified by RT-PCR from total RNA extracted from infected cells. The NS5 gene cassette in the DV2 replicon was replaced with the DV3 NS5 sequence, producing the DV2/3-NS5-EGFP-replicon plasmid. Details of these constructions are described in Section 2 and in Fig. 1.

3.2. Production of hamster, simian and human cell lines supporting the DV2 EGFP-replicon

The DV2 EGFP-replicon plasmid was used as a template to produce a ^{m7}GpppA-capped replicon RNA that was introduced into hamster BHK, simian COS, and human Huh7 cells. The replicon RNA would be translated as a unique polyprotein by the cellular machinery. The fusion protein EGFP-PAC-1D2A would be separated from the remainder of the polyprotein by the ribosome skip induced by the 1D2A peptide. The DV non-structural proteins would be cleaved either by cellular proteases or by the NS2B/NS3 protease complex (Lindenbach et al., 2007) and assemble into a replication complex. The puromycin concentration used for selection of replicon containing cells was set according to the differential sensitivity level of each cell line: BHK cells were less sensitive than COS cells, which were less sensitive than Huh7 cells. BHK cells transfected with the replicon RNA were resistant to puromycin at 3.5 μg/ml, whereas non-transfected BHK cells were highly sensitive and died in 2 days at 1 μg/ml of puromycin. COS- and Huh7- DV2 cells trans-

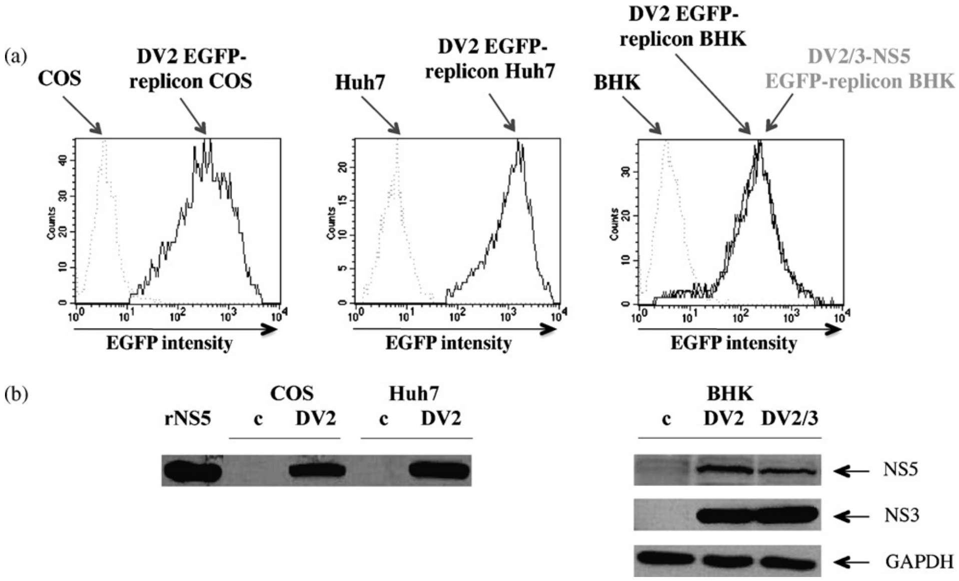


Fig. 2. Analysis of the expression of EGFP, NS3, NS5 and GAPDH proteins in replicon containing cells. (a) Analysis by flow cytometry of the fluorescence level of nontransfected (dotted line) or stable established EGFP-replicon containing (solid line) cells. (b) Analysis by Western-blot of the NS5 (higher panel), NS3 (middle panel) and GAPDH (lower panel) expression in non-transfected cells (lanes 'c') and in EGFP-replicon containing cells. Recombinant NS5 (rNS5) was loaded as a control. DV2: DV2 EGFP-replicon; DV2/3: DV2/3-NS5 EGFP-replicon.

fected with the DV EGFP-replicon were selected using 2 μg/ml and 1 μg/ml of puromycin, respectively.

Transfected puromycin-resistant cells, but not non-transfected cells, also expressed EGFP, as verified by fluorescent microscopy and flow cytometry (data not shown). Cells expressing high levels of EGFP were sorted and amplified. Replicon RNA was extracted and sequenced, and no mutations observed. EGFP expression was maintained at the same level for at least a period of 2 months in the presence of the respective puromycin concentration. BHK-, COS- and Huh7-DV2 EGFP-replicon containing cells were shown to express EGFP (Fig. 2a) and the NS5 protein (Fig. 2b, higher panels). EGFP expression was monitored by flow cytometry, whilst NS5 protein expression was revealed by Western-blot with an in-house produced and well-characterized monoclonal antibody directed to the DV2 NS5 RdRp domain.

3.3. *Production of a BHK cell line supporting a chimeric DV2 EGFP-replicon with full length NS5 from DV3*

The DV2-NS5- and the DV2/3-NS5-EGFP-replicon plasmids were linearized and *in vitro* transcribed. BHK cells, transfected with one or the other replicon RNA, were resistant to puromycin at 3.5 μg/ml and expressed EGFP. Cells expressing high levels of EGFP were sorted and amplified. In order to detect any adaptive mutations, replicon RNA was once again extracted and sequenced. No mutations were observed. Besides the EGFP protein (Fig. 2a), BHK DV2/3-NS5 EGFP-replicon containing cells expressed the NS5 protein (Fig. 2b, higher right panel). The NS5 signal, as revealed with an anti-NS5 monoclonal antibody directed to the DV2 NS5 RdRp domain, was greater in BHK DV2 cells than in the BHK DV2/3-NS5 cells. However, the levels of expression of the viral NS3 protein

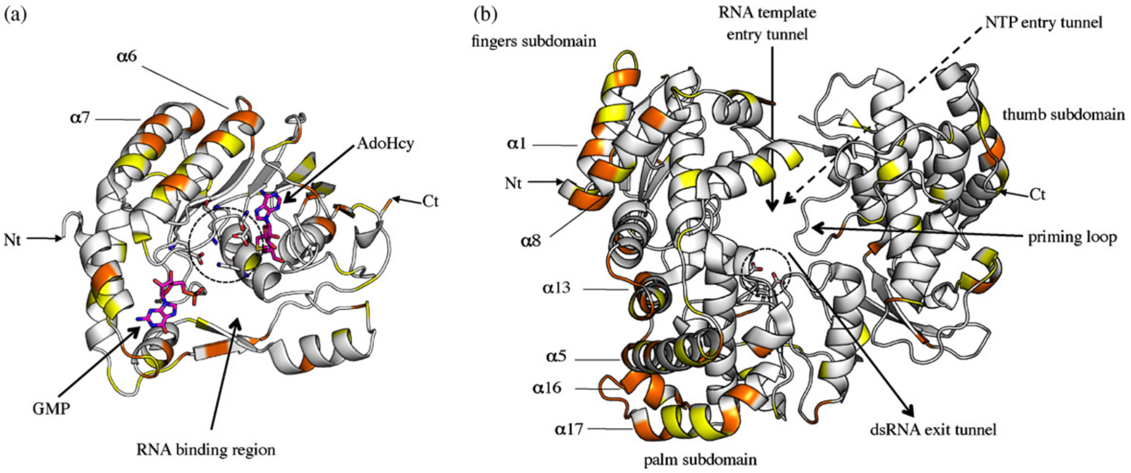


Fig. 3. Localization of solvent-accessible aminoacids differing between DV serotypes 2 and 3 on the NS5 MTase and RdRp domains. On the 3D ribbon views, non-conserved, similar and conserved amino acids are colored in orange, yellow and white, respectively. (a) MTase (left; PDB code: 2P41): the catalytic tetrad (K61-D146-K182-E218), the cofactor S-adenosyl-homocysteine (AdoHcy), the RNA binding region, and the GMP in the cap-binding site are indicated. (b) RdRp (PDB code: 2J7U): fingers, palm, thumb subdomains, tunnels of NTP entry, RNA entry, RNA exit, the catalytic residues (D533-D663), and the priming loop, are indicated. (a) and (b) Dotted circles represent active sites. Amino acids side chains are shown as sticks with the following color code: grey: C, blue: N, red: O. AdoHcy and GMP are shown as sticks with the following color code: magenta: C, blue: N, red: O, yellow: S, orange: P. Nt: amino-terminal, Ct: carboxy-terminal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(middle right panel) and the cellular protein GAPDH (Fig. 2b, lower right panel) were similar between the BHK DV2 and BHK DV2/3-NS5 EGFP-replicon containing cells, indicating that both replicons replicated at similar levels. The lower NS5 signal in the BHK DV2/3-NS5 EGFP-replicon containing cells may result from a lower affinity of the monoclonal antibody for the DV3-NS5 protein than the DV2-NS5 protein.

The exchange of the DV2 NS5 gene with that of DV3, allowed the assembly of a functional replication complex, as evidenced by the long term maintenance of EGFP and NS5 expression, indicative of replicon replication. The DV2 (NGC strain) and DV3 (H87 strain) NS5 proteins have 79% identity and 91% homology. The 3D structure of the NS5 MTase (Egloff et al., 2002) and RdRp (Yap et al., 2007) domains are shown in Fig. 3. The catalytic tetrade Lys-Asp-Lys-Glu, the Adenosyl-Homocysteine (AdoHcy) cofactor, the cap binding site and the RNA binding region are indicated on the MTase 3D view (Fig. 3a). The palm, thumb and fingers subdomains, the tunnels of NTP entry, RNA entry, RNA exit, the catalytic Asp residues, and the priming loop, are indicated on the RdRp 3D view (Fig. 3b). The amino acids potentially involved in protein-protein interactions are solvent-accessible. Based on an alignment of the DV2 NS5 and DV3 NS5 amino acid sequences (see Supplementary Fig. I), we distinguished three types of solvent-accessible amino acids: identical residues are labeled in white, similar residues in yellow and non-conserved residues in orange. Surface residues that differ between the DV2 and DV3 NS5 MTase domain are mainly observed on the α6 and α7 helices. In the case of the RdRp domain, the main differences are localized on six α-helices: α1 and α8 on the thumb subdomain, and α5, α13, α16 and α17 on the palm subdomain. As shown in Fig. 3, these helices are outside of the catalytic sites, and this suggests that they are not the major determinants of interactions with viral or cellular proteins crucial for a functional DV replication complex.

3.4. Set-up and validation of a DV inhibitor EGFP-based screening assay

In stable cell lines supporting autonomously replicating subgenomic replicons, RNA replication involves multiple viral and cellular proteins which form the replication complex. These cells provide multiple targets whose functions could be inhibited or interactions disrupted by small molecule inhibitors of viral replication. We therefore examined whether the replicon cell lines described above could be used as the basis of an antiviral screening assay. The EGFP protein is cotranslated from the replicon RNA with the DV non-structural proteins, its level of expression is dependent on a functional replication complex. Thus, the inhibition of the replication and/or the translation of the replicon RNA would lead to a decrease in the level of EGFP expression, that would reflect a similar decrease in the expression of the viral non-structural proteins.

The decrease in EGFP expression is easily measurable in a high throughput assay using a plate-carrying fluorimeter. Identification of inhibitors requires quantification of an EGFP expression decrease in the steady-state level of RNA replicon. The cytotoxicity and the inhibitory potential of the compounds were tested on the same plate. To do so, cells were maintained with the molecules for 48 h, the solution being renewed after 24 h. The initial cell number per well and the duration of the assay are thus critical since the cell density could affect the replication efficiency of the RNA replicon. The cell monolayer may be close to the subconfluent state at the end of the assay, 1×10^4 cells per well in 96-well plates were initially seeded and the readout measurements were done 3 days later. Since phenol red could interfere with fluorescence measurements, to varying degrees depending on the cell density, we decided to use an uncolored culture medium without phenol red during the experiment, and to replace media by PBS before measuring EGFP fluorescence. As chemical compounds have to be dissolved in DMSO, the effect of DMSO concentration on cell viability was examined. Replicon containing cells were fully viable at 0.5% DMSO, but stopped dividing at 1% DMSO and died in less than 24 h above 1% DMSO. The inhibition assays were therefore carried out in 96-well plates, with media containing a final concentration of 0.5% DMSO. Under these conditions, the signal-to-noise ratio (EGFP fluorescence level from replicon-containing cells divided by the background signal from naïve BHK cells) was around 20 to 25. At the end of a 2-day treatment with potential inhibitors, the Celltiterblue® reagent was used to calculate the cell number in each well. Total EGFP fluorescence was then correlated to this cell number. With only these two fluorescence readouts, we could quantify inhibitory activity and toxicity and measure the EC_{50} (the effective concentration that led to 50% of the control fluorescence per replicon cell) and the CC_{50} (the concentration that led to 50% of the control cell number) of selected compounds.

To validate our screening assay, we used ribavirin for two reasons: its antiviral activities and its moderate cellular toxicity (Graci and Cameron, 2006), which are the two characteristics quantified with our assay. Ribavirin, a structural analog of GTP, is an inhibitor of the replication of a broad spectrum of viruses (Sidwell et al., 1972; Streeter et al., 1973). Ribavirin has been shown to inhibit the replication of DV in infection assays (Koff et al., 1982; Crance et al., 2003; Takhampunya et al., 2006) and in a DV subgenomic replicon assay (Ng et al., 2007). Five distinct mechanisms have been proposed to explain the antiviral properties of ribavirin (for a review, see Parker (2005), Graci and Cameron (2006)). These include both indirect mechanisms (inosine monophosphate dehydrogenase inhibition, immunomodulatory effects) and direct mechanisms (interference with RNA capping, polymerase inhibition, lethal mutagenesis). At least one direct mechanism has been reported for a DV protein: the *in vitro* inhibition of the 2′-O MTase activity of the DV NS5 protein (Benarroch et al., 2004).

Experiments carried out in the 96-well plate format, with BHK DV2 EGFP-replicon containing cells, allowed us to calculate a reproducible EC_{50} value of 15.5 ± 1.1 μM for ribavirin (Fig. 4a, solid line). This value is representative of 4 experiments and is consistent with values published previously using other cellular systems (Koff et al., 1982; Crance et al., 2003; Takhampunya et al., 2006; Ng et al., 2007). We calculated a Z' factor, according to the mean values and the standard deviations of the EGFP fluorescence levels of BHK DV2 EGFP-replicon containing cells, either not treated or treated with 100 μM or 50 μM of ribavirin. The Z' factor was 0.76 and 0.73, respectively, indicating that the assay is reliable (Zhang et al., 1999). Similar reproducible EC_{50} values were calculated with the other DV2 EGFP-replicon cell lines (Table 2): 12.7 ± 1.7 μM for the COS DV2 EGFP-replicon containing cells and 11.9 ± 0.7 μM for the Huh7 DV2 EGFP-replicon containing cells. As for the BHK DV2 EGFP-replicon containing cells, we tested the effect of ribavirin on BHK DV2/3-NS5 EGFP-replicon cells and measured a reproducible EC_{50} value of 12 ± 1.2 μM (Table 2).

An alternate assay was developed to directly quantify the fluorescence of EGFP in individual cells treated with ribavirin, independently of the cell count number. Replicon containing cells seeded in 24-well plates were treated with ribavirin following the same protocol used for 96-well plates but analysed by flow cytometry. Fluorescence of individual cells was measured for each ribavirin concentration and the mean values of the fluorescence intensities were directly compared to the control value. Using this assay, the EC_{50} value calculated for BHK DV2 EGFP-replicon containing cells was 18.4 ± 1.8 μM (Fig. 4a, dotted line, representative of 3 individual experiments), which is similar to the EC_{50} value obtained with the high throughput 96-well assay. With the 24-well format method, EC_{50} values were 5.8 ± 0.6 μM, 19.2 ± 1.8 μM and 20 ± 1.5 μM, for COS- and Huh7- DV2-, and BHK DV2/3-NS5-

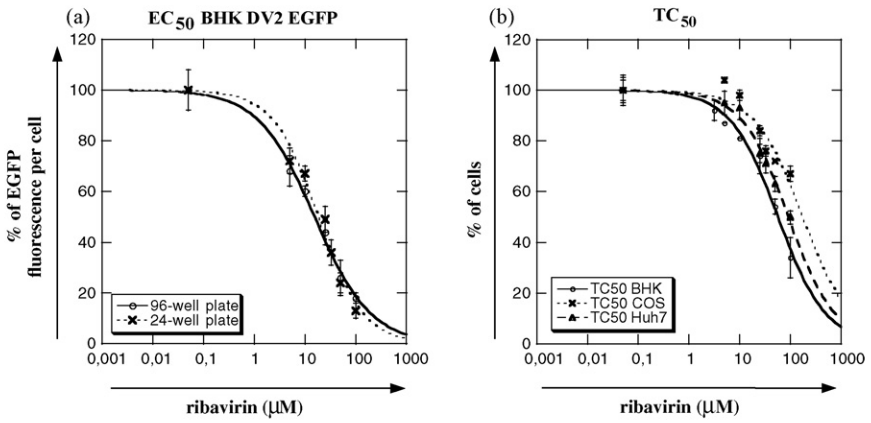


Fig. 4. Measurement of CC_{50} and EC_{50} values of ribavirin on DV EGFP-replicon cells. (a) Ribavirin EC_{50} values were calculated on BHK DV2 EGFP-replicon containing cells, from plate-carrying fluorimeter measurements (o, solid line) or obtained from flow cytometry measurements (×, dotted line). (b) From plate-carrying fluorimeter measurements, ribavirin CC_{50} values were calculated on BHK cells (solid line), COS cells (large dotted line) and Huh7 cells (thin dotted line).

Table 2
EC50, CC50 and SI values of ribavirin on DV EGFP-replicon containing cell lines.

	EC_{50} (µM)		CC_{50} (µM)[c]	SI[d]
BHK DV2 EGFP	15.5 ± 1.1[a]	18.4 ± 1.8[b]	57.2 ± 5.6	3.7
BHK DV2/3-NS5 EGFP	12 ± 1.2[a]	20 ± 1.5[b]		4.7
COS DV2 EGFP	12.7 ± 1.7[a]	5.8 ± 0.6[b]	>100	>7.9
Huh7 DV2 EGFP	11.9 ± 0.7[a]	19.2 ± 1.8[b]	91.9 ± 7.6	7.7

[a] EC50 values determined with the 96-wells plate assay.
[b] EC50 values determined with the 24-well plate assay.
[c] CC50 values were calculated with non-transfected cells.
[d] SI: selectivity index.

EGFP-replicon containing cells, respectively (Table 2). The high throughput fluorescence assay was validated as EC_{50} values are similar to those obtained with the flow cytometry assay. These results demonstrated unequivocally that combining the total EGFP fluorescence intensity measurement and cell counting using a fluorescent reagent is an appropriate method to rapidly calculate inhibitory activity and EC_{50} values of selected compounds.

Moreover, with the same assay, we could measure CC_{50} values and calculate the selectivity index ($SI = CC_{50}/EC_{50}$). For BHK cells, the ribavirin CC_{50} was 57.2 ± 5.6 µM (Fig. 4b, solid line). This gave SI values of 3.7 and 4.7, for BHK DV2- and BHK DV2/3-NS5- EGFP-replicon containing cells, respectively. Ribavirin CC_{50} values were higher than 100 µM for COS cells (Fig. 4b, large dotted line) and 91.9 ± 7.6 µM for Huh7 cells (Fig. 4b, thin dotted line), giving SI values above 7.7. Even though the EC_{50}/CC_{50} and SI values of ribavirin showed only minor variation between different cell types (Table 2), the availability of a number of different DV replicon containing cell lines should prove useful for screening inhibitors that may have different cell type specific characteristics.

3.5. The decrease in ribavirin-induced EGFP expression correlates with replicon NS5 protein expression

To confirm that the decrease in EGFP expression reflected a decrease in the expression of the viral non-structural proteins, we analyzed the amounts of DV NS5 protein following ribavirin treatment. Equal quantities of total protein extracted from ribavirin-treated cells were loaded on a SDS-PAGE gel and transferred to a PVDF membrane. The level of expression of the cellular protein GAPDH was not significantly affected by ribavirin treatment (Fig. 5, lower panel). As expected, the EGFP protein expression level decreased in BHK DV2 EGFP-replicon containing cells treated with ribavirin (Fig. 5, middle panel). Furthermore, the DV2 NS5 protein expression level decreased with increasing ribavirin concentrations (Fig. 5, upper panel). These results demonstrated that ribavirin inhibited the DV replication complex in the replicon containing cell lines and that its antiviral effect can be determined by monitoring the EGFP expression level correlated to the cell number. In conclusion, our Western-blot analysis confirmed that the high-throughput fluorescence measurements on the DV EGFP replicon-containing cells could be used to identify new inhibitory compounds.

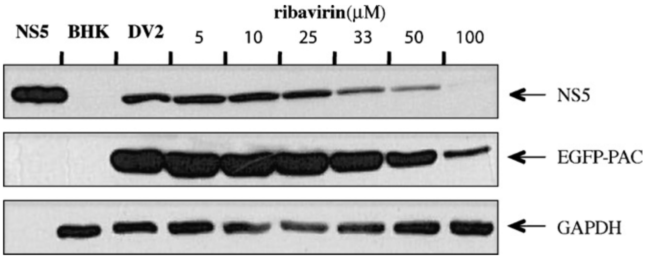


Fig. 5. Analysis by Western-blot of the expression of DV-NS5, EGFP and GAPDH proteins upon ribavirin treatment. Upper panel: NS5, middle panel: EGFP, lower panel: GAPDH. From left to right: purified recombinant NS5 protein (NS5), BHK cells extract (BHK), BHK DV2 EGFP-replicon containing cells extracts, untreated (DV2) or treated with ribavirin, 5, 10, 25, 33, 50 and 100 µM.

4. Discussion

The goal of this study was to extend current genetic systems for the development of HTS assays for anti-flavivirus drug discovery. Using DV2 as a model, we describe the construction of subgenomic DV EGFP-replicon plasmids and the production of several subgenomic replicon containing cell lines.

Self-replicating, non-infectious flavivirus subgenomic replicons have been widely used in studies of adaptive mutations (Liu et al., 2005; Rossi et al., 2005, 2007), viral assembly and packaging

(Harvey et al., 2004; Fayzulin et al., 2006; Leung et al., 2008), coupling between replication and packaging (Khromykh et al., 2001), viral RNA sequence and/or structure requirements for translation or replication (Lo et al., 2003a; Tilgner and Shi, 2004; Alvarez et al., 2005, 2008; Tilgner et al., 2005; Filomatori et al., 2006) and the antiviral immune response (Jones et al., 2005; Hershkovitz et al., 2008). Among these replicons, WNV-derived replicons have also been used for the screening of inhibitors, utilizing the luciferase reporter protein (Lo et al., 2003b; Rossi et al., 2005; Gu et al., 2006; Ng et al., 2007; Noueiry et al., 2007), resulting in several screening successes (Puig-Basagoiti et al., 2005; Goodell et al., 2006; Gu et al., 2006; Noueiry et al., 2007; Puig-Basagoiti et al., 2009). Two reports have described an HTS assay based on a DV replicon, both in BHK cells, with a luciferase reporter detection (Puig-Basagoiti et al., 2006; Ng et al., 2007). The assays described in this study demonstrate the establishment of a new DV replicon system based on an EGFP readout.

The establishment of replicon containing cell lines that can be used for HTS assays requires the expression of genes coding for a drug selectable marker and a reporter protein. A number of approaches have been used to produce the two individual proteins, including the use of self cleaving protein sequences to process polyproteins containing the two proteins and the use of an IRES element to either initiate translation of the viral non-structural proteins or a heterologous protein encoded by a gene cassette inserted in the viral 3′UTR. In the DV2 EGFP replicon, the EGFP and PAC proteins were produced as fusion protein that was cleaved from the DV2 non-structural proteins by the 1D2A peptide sequence from foot-and-mouth-disease virus. Previous studies have shown that translation initiated by an IRES sequence downstream from a cap-dependent translation start site is often inefficient (Mizuguchi et al., 2000; Ibrahimi et al., 2009) and can be subject to cell type specific effects. By contrast, the separation of heterologous gene product/s from the viral non-structural proteins by the 1D2A peptide essentially results in equimolar amounts of products (de Felipe et al., 2006). As the EGFP fluorescence from DV2 EGFP replicon containing cells was taken as a measure of non-structural protein expression, it may be beneficial to have the EGFP-PAC fusion protein separated from the non-structural proteins through the use of the 1D2A sequence rather than an IRES sequence. Both the EGFP and PAC proteins were active in the context of a fusion protein with the EGFP-PAC protein giving similar resistance to puromycin as the PAC gene product alone in previous studies (Jones et al., 2005). Although a gene construct containing an IRES element and a selectable marker gene have been successfully introduced into the 3′ UTR of a number of flavivirus replicons (Khromykh et al., 2001; Shi et al., 2002; Rossi et al., 2005; Gu et al., 2006), including DV type 1 (Suzuki et al., 2007), we and others have found it difficult to reliably insert similar gene cassettes into the DV2 genome (unpublished data (Ng et al., 2007)). The ability to insert a unique gene cassette encoding an EGFP-PAC fusion protein in place of the structural protein genes circumvents this potential difficulty in DV replicon production.

The DV replicon containing cell lines produced in this study will serve as useful tools for high throughput screening of DV replication inhibitors. Our assay is robust as the Z' factor is greater than 0.7 and the signal-to-noise ratio is 20 to 25, values well suited for a HTS assay. The EC_{50} and CC_{50} values determined for ribavirin with the new DV replicon containing cell lines are similar to the values described using virus infected cell lines (Koff et al., 1982; Crance et al., 2003; Takhampunya et al., 2006; Ng et al., 2007). The effectiveness of the inhibition of the DV replication complex, based on the EGFP readout, was confirmed by examining the DV NS5 expression level with a specific monoclonal antibody. Our results demonstrate unequivocally that the determination of replicon replication levels using a fluorescence-based method that correlates the total EGFP fluorescence with total cell numbers is appropriate for the measuring inhibitory activity and for calculating the EC_{50} and CC_{50} values of selected compounds. In a HTS, in 96- or 384-wells plates, the process would be divided in two parts. A single EGFP readout will select compounds with both antiviral activity and/or cytotoxic effect. The selected compounds, decreasing the EGFP level, will be tested for their cytotoxicity, thus allowing to keep the non-cytotoxic compounds with antiviral activity.

Two types of approaches are routinely used for antiviral assay development. One approach is biochemistry-based, in which the enzymatic activity of purified viral protein is assayed. For flaviviruses, enzymatic HTS assays have been developed for the multifunctional NS3 (protease, helicase, nucleoside triphosphatase, and 5′-RNA triphosphatase) and NS5 (MTase and RdRp) proteins (Johnston et al., 2007; Lim et al., 2008). The principal advantage of the biochemistry-based assay is that the targets of the identified inhibitors are known. The other approach is replicon-based, a biosafe alternative to using infectious virus and involves multiple targets in the viral life cycle. The assays described in this study provide additional means for DV drug discovery. For example, inhibitors obtained from anti-MTase and anti-RdRp based screens can now be tested using the DV2 EGFP replicon-containing cells. In addition, with the use of the DV2 or DV2/3-NS5 replicon-containing cells it will be possible to determine the potential of the MTase/RdRp inhibitors in the context of different DV serotypes. Moreover, with the replicon carrying the NS5 gene cassette, it will be possible to characterize NS5 mutants derived from biochemical/structural analysis or resistance studies, in a cellular environment.

A range of cell types were used in this study to stably express the DV replicons. As in a number of previous studies we initially established the DV2 EGFP replicon in BHK cells which were then used to optimize the assays for HTS. The DV inter-serotypic chimeric replicon, expressing the NS1 to NS4 genes from DV2, and the NS5 gene from DV3 was also established in BHK cells. In addition to its use for drug discovery, the chimeric replicon system provides an interesting tool to study the determinants of interactions which occur in the viral replication complex.

Once established in BHK cells, the replicon systems were then extended to a human hepatic cell line (Huh7) and the COS monkey cell line. DV can infect many cell types and causes diverse clinical and pathological effects. Although the main target cells in humans are believed to reside in the reticuloendothelial system, both clinical and experimental observations suggest that there is also liver damage during DV infection. Clinical evidence of liver involvement in DV infections includes the presence of hepatomegaly and increased levels of serum liver enzymes (Seneviratne et al., 2006). DV is able to replicate in both hepatocytes and Kupffer cells (Huerre et al., 2001). Thus, the DV2 replicon Huh7 stable cell line constitutes a suitable tool for studying the interactions between viral proteins and hepatic proteins, some of which may be responsible for the infection-induced liver damage (Ekkapongpisit et al., 2007; Pattanakitsakul et al., 2007; Higa et al., 2008). Extension of the replicon system to COS cells demonstrated that even cells that are poorly infected with DV *in vitro* (unpublished observations (Rodrigo et al., 2006)) can be used for maintaining DV replicons if they are permissive for genome replication. Monkey kidney-derived cells have been previously used for DV protein interaction studies *in vitro* (Kapoor et al., 1995). We therefore produced a COS cell line containing the DV replicon, allowing studies of the DV replication complex with antibodies directed to viral or cellular proteins. Compared to BHK cells, COS cells have the advantage that they are less likely to be susceptible to cross reactivity with mouse or rat monoclonal antibodies, commonly used for immunoprecipitation studies. Overall the production of a range of DV replicon cell lines provides new opportunities for the study of DV replication in a cellular context. At the same time, the establishment of multiple cell lines carrying the DV replicon demonstrates the feasibility of directly evaluating the potential of selected compounds in cell lines from different organs and species.

In summary, the assays described in this study will greatly facilitate DV drug discovery by serving as primary or complementary screening assays. The approach should be applicable to the development of cell-based HTS assay for other flaviviruses, and should also be useful for the study of DV replication and pathogenesis.

Acknowledgments

Authors thank Xavier de Lamballerie, Rémi Charrel and Hugues Tolou for their contribution and Cécile Baronti for the DV3-NS5 PCR product, Marc Barad for the cell-sorting, and Etienne Decroly for the rNS5 protein. Authors also appreciate the work of Hélène Codoul, Audrey Blanchard, Florian Madura (Internship ESIL) and Marion Ponserre (InternshipBTSMarieCurie). Authors are grateful to Barbara Selisko for carefully reading the manuscript. Authors thank the Région PACA, the Association de Recherche contre le Cancer, the Infectiopole, the Université de la Méditerranée (BQR), the Direction Généralede l'Armement (DGA #09cà402), and the Agence Nationale de la Recherche (ANR-07-BLAN 0285 Dengue D-D), for their support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.antiviral.2010.03.010.

References

Alvarez, D.E., Filomatori, C.V., Gamarnik, A.V., 2008. Functional analysis of dengue virus cyclization sequences located at the 5' and 3'UTRs. Virology 375, 223–235.

Alvarez, D.E., Lodeiro, M.F., Ludueña, S.J., Pietrasanta, L.I., Gamarnik, A.V., 2005. Long-range RNA-RNA interactions circularize the dengue virus genome. J. Virol. 79, 6631–6643.

Benarroch, D., Egloff, M.P., Mulard, L., Guerreiro, C., Romette, J.L., Canard, B., 2004. A structural basis for the inhibition of the NS5 dengue virus mRNA 2'-O-methyltransferase domain by ribavirin 5'-triphosphate. J. Biol. Chem. 279, 35638–35643.

Crance, J.M., Scaramozzino, N., Jouan, A., Garin, D., 2003. Interferon, ribavirin, 6-azauridine and glycyrrhizin: antiviral compounds active against pathogenic F flaviviruses. Antiviral Res. 58, 73–79.

de Felipe, P., Luke, G.A., Hughes, L.E., Gani, D., Halpin, C., Ryan, M.D., 2006. E unum pluribus: multiple proteins from a self-processing polyprotein. Trends Biotechnol. 24, 68–75.

Dong, H., Zhang, B., Shi, P.Y., 2008. Flavivirus methyltransferase: A novel antiviral target. Antiviral Res. 80, 1–10.

Doronina, V.A., Wu, C., de Felipe, P., Sachs, M.S., Ryan, M.D., Brown, J.D., 2008. Site-specific release of nascent chains from ribosomes at a sense codon. Mol. Cell. Biol. 28, 4227–4239.

Egloff, M.P., Benarroch, D., Selisko, B., Romette, J.L., Canard, B., 2002. An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. EMBO J. 21, 2757–2768.

Ekkapongpisit, M., Wannatung, T., Susantad, T., Triwitayakorn, K., Smith, D.R., 2007. cDNA-AFLP analysis of differential gene expression in human hepatoma cells (HepG2) upon dengue virus infection. J. Med. Virol. 79, 552–561.

Falgout, B., Chanock, R., Lai, C.J., 1989. Proper processing of dengue virus nonstructural glycoprotein NS1 requires the N-terminal hydrophobic signal sequence and the downstream nonstructural protein NS2a. J. Virol. 63, 1852–1860.

Fayzulin, R., Scholle, F., Petrakova, O., Frolov, I., Mason, P.W., 2006. Evaluation of replicative capacity and genetic stability of West Nile virus replicons using highly efficient packaging cell lines. Virology 351, 196–209.

Filomatori, C.V., Lodeiro, M.F., Alvarez, D.E., Samsa, M.M., Pietrasanta, L., Gamarnik, A.V., 2006. A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. Genes Dev. 20, 2238–2249.

Goodell, J.R., Puig-Basagoiti, F., Forshey, B.M., Shi, P.Y., Ferguson, D.M., 2006. Identification of compounds with anti-West Nile Virus activity. J. Med. Chem. 49, 2127–2137.

Graci, J.D., Cameron, C.E., 2006. Mechanisms of action of ribavirin against distinct viruses. Rev. Med. Virol. 16, 37–48.

Gu, B., Ouzunov, S., Wang, L., Mason, P., Bourne, N., Cuconati, A., Block, T.M., 2006. Discovery of small molecule inhibitors of West Nile virus using a high-throughput sub-genomic replicon screen. Antiviral Res. 70, 39–50.

Gualano, R.C., Pryor, M.J., Cauchi, M.R., Wright, P.J., Davidson, A.D., 1998. Identification of a major determinant of mouse neurovirulence of dengue virus type 2 using stably cloned genomic-length cDNA. J. Gen. Virol. 79, 437–446.

Harvey, T.J., Liu, W.J., Wang, X.J., Linedale, R., Jacobs, M., Davidson, A., Le, T.T., Anraku, I., Suhrbier, A., Shi, P.Y., Khromykh, A.A., 2004. Tetracycline-inducible packaging cell line for production of flavivirus replicon particles. J. Virol. 78, 531–538.

Hellen, C.U., Sarnow, P., 2001. Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev. 15, 1593–1612.

Hennecke, M., Kwissa, M., Metzger, K., Oumard, A., Kröger, A., Schirmbeck, R., Reimann, J., Hauser, H., 2001. Composition and arrangement of genes define the strength of IRES-driven translation in bicistronic mRNAs. Nucleic Acids Res. 29, 3327–3334.

Hershkovitz, O., Zilka, A., Bar-Ilan, A., Abutbul, S., Davidson, A., Mazzon, M., Kümmerer, B.M., Monsoengo, A., Jacobs, M., Porgador, A., 2008. Dengue virus replicon expressing the nonstructural proteins suffices to enhance membrane expression of HLA class I and inhibit lysis by human NK cells. J. Virol. 82, 7666–7676.

Higa, L.M., Caruso, M.B., Canellas, F., Soares, M.R., Oliveira-Carvalho, A.L., Chapeaurouge, D.A., Almeida, P.M., Perales, J., Zingali, R.B., Da Poian, A.T., 2008. Secretome of HepG2 cells infected with dengue virus: implications for pathogenesis. Biochim. Biophys. Acta 1784, 1607–1616.

Huerre, M.R., Lan, N.T., Marianneau, P., Hue, N.B., Khun, H., Hung, N.T., Khen, N.T., Drouet, M.T., Huong, V.T., Ha, D.Q., Buisson, Y., Deubel, V., 2001. Liver histopathology and biological correlates in five cases of fatal dengue fever in Vietnamese children. Virchows Arch. 438, 107–115.

Ibrahimi, A., Vande Velde, G., Reumers, V., Toelen, J., Thiry, I., Vandeputte, C., Vets, S., Deroose, C., Bormans, G., Baekelandt, V., Debyser, Z., Gijsbers, R., 2009. Highly efficient multicistronic lentiviral vectors with peptide 2A sequences. Hum. Gene Ther. 20, 845–860.

Johnston, P.A., Phillips, J., Shun, T.Y., Shinde, S., Lazo, J.S., Huryn, D.M., Myers, M.C., Ratnikov, B.I., Smith, J.W., Su, Y., Dahl, R., Cosford, N.D., Shiryaev, S.A., Strongin, A.Y., 2007. HTS identifies novel and specific uncompetitive inhibitors of the two-component NS2B-NS3 proteinase of West Nile virus. Assay Drug Dev. Technol. 5, 737–750.

Jones, M., Davidson, A., Hibbert, L., Gruenwald, P., Schlaak, J., Ball, S., Foster, G.R., Jacobs, M., 2005. Dengue virus inhibits alpha interferon signaling by reducing STAT2 expression. J. Virol. 79, 5414–5420.

Kapoor, M., Zhang, L., Ramachandra, M., Kusukawa, J., Ebner, K.E., Padmanabhan, R., 1995. Association between NS3 and NS5 proteins of dengue virus type 2 in the putative RNA replicase is linked to differential phosphorylation of NS5. J. Biol. Chem. 270, 19100–19106.

Khromykh, A.A., Meka, H., Guyatt, K.J., Westaway, E.G., 2001. Essential role of cyclization sequences in flavivirus RNA replication. J. Virol. 75, 6719–6728.

Koff, W.C., Elm Jr., J.L., Halstead, S.B., 1982. Antiviral effects if ribavirin and 6-mercapto-9-tetrahydro-2-furylpurine against dengue viruses in vitro. Antiviral Res. 2, 69–79.

Lescar, J., Luo, D., Xu, T., Sampath, A., Lim, S.P., Canard, B., Vasudevan, S.G., 2008. Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. Antiviral Res. 80, 94–101.

Leung, J.Y., Pijlman, G.P., Kondratieva, N., Hyde, J., Mackenzie, J.M., Khromykh, A.A., 2008. Role of nonstructural protein NS2A in flavivirus assembly. J. Virol. 82, 4731–4741.

Liang, C., Rieder, E., Hahm, B., Jang, S.K., Paul, A., Wimmer, E., 2005. Replication of a novel subgenomic HCV genotype 1a replicon expressing a puromycin resistance gene in Huh-7 cells. Virology 333, 41–53.

Lim, S.P., Wen, D., Yap, T.L., Yan, C.K., Lescar, J., Vasudevan, S.G., 2008. A scintillation proximity assay for dengue virus NS5 2'-O-methyltransferase-kinetic and inhibition analyses. Antiviral Res. 80, 360–369.

Lindenbach, B.D., Thiel, H.J., Rice, C.M., 2007. Flaviviridae: the viruses and their replication. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology, Vol. 1. Lippincott Williams & Wilkins, pp. 1101–1152.

Liu, W.J., Wang, X.J., Mokhonov, V.V., Shi, P.Y., Randall, R., Khromykh, A.A., 2005. Inhibition of interferon signaling by the New York 99 strain and Kunjin subtype of West Nile virus involves blockage of STAT1 and STAT2 activation by nonstructural proteins. J. Virol. 79, 1934–1942.

Lo, M.K., Tilgner, M., Bernard, K.A., Shi, P.Y., 2003a. Functional analysis of mosquito-borne flavivirus conserved sequence elements within 3' untranslated region of West Nile virus by use of a reporting replicon that differentiates between viral translation and RNA replication. J. Virol. 77, 10004–10014.

Lo, M.K., Tilgner, M., Shi, P.Y., 2003b. Potential high-throughput assay for screening inhibitors of West Nile virus replication. J. Virol. 77, 12901–12906.

Malet, H., Massé, N., Selisko, B., Romette, J.L., Alvarez, K., Guillemot, J.C., Tolou, H., Yap, T.L., Vasudevan, S., Lescar, J., Canard, B., 2008. The flavivirus polymerase as a target for drug discovery. Antiviral Res. 80, 23–35.

Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E., Hayakawa, T., 2000. IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. Mol. Ther. 1, 376–382.

Ng, C.Y., Gu, F., Phong, W.Y., Chen, Y.L., Lim, S.P., Davidson, A., Vasudevan, S.G., 2007. Construction and characterization of a stable subgenomic dengue virus type 2 replicon system for antiviral compound and siRNA testing. Antiviral Res. 76, 222–231.

Noueiry, A.O., Olivo, P.D., Slomczynska, U., Zhou, Y., Buscher, B., Geiss, B., Engle, M., Roth, R.M., Chung, K.M., Samuel, M., Diamond, M.S., 2007. Identification of novel small-molecule inhibitors of West Nile virus infection. J. Virol. 81, 11992–12004.

Parker, W.B., 2005. Metabolism and antiviral activity of ribavirin. Virus Res. 107, 165–171.

Pattanakitsakul, S.N., Rungrojcharoenkit, K., Kanlaya, R., Sinchaikul, S., Noisakran, S., Chen, S.T., Malasit, P., Thongboonkerd, V., 2007. Proteomic analysis of host responses in HepG2 cells during dengue virus infection. J. Proteome Res. 6, 4592–4600.

Pryor, M.J., Carr, J.M., Hocking, H., Davidson, A.D., Li, P., Wright, P.J., 2001. Replication of dengue virus type 2 in human monocyte-derived macrophages: comparisons of isolates and recombinant viruses with substitutions at amino acid 390 in the envelope glycoprotein. Am. J. Trop. Med. Hyg. 65, 427–434.

Puig-Basagoiti, F., Deas, T.S., Ren, P., Tilgner, M., Ferguson, D.M., Shi, P.Y., 2005. High-throughput assays using a luciferase-expressing replicon, virus-like particles, and full-length virus for West Nile virus drug discovery. Antimicrob. Agents Chemother. 49, 4980–4988.

Puig-Basagoiti, F., Qing, M., Dong, H., Zhang, B., Zou, G., Yuan, Z., Shi, P.Y., 2009. Identification and characterization of inhibitors of West Nile virus. Antiviral Res. 83, 71–79.

Puig-Basagoiti, F., Tilgner, M., Forshey, B.M., Philpott, S.M., Espina, N.G., Wentworth, D.E., Goebel, S.J., Masters, P.S., Falgout, B., Ren, P., Ferguson, D.M., Shi, P.Y., 2006. Triaryl pyrazoline compound inhibits flavivirus RNA replication. Antimicrob. Agents Chemother. 50, 1320–1329.

Rice, C.M., Lenches, E.M., Eddy, S.R., Shin, S.J., Sheets, R.L., Strauss, J.H., 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science 229, 726–733.

Rodrigo, W.W., Jin, X., Blackley, S.D., Rose, R.C., Schlesinger, J.J., 2006. Differential enhancement of dengue virus immune complex infectivity mediated by signaling-competent and signaling-incompetent human Fcgamma RIA (CD64) or FcgammaRIIA (CD32). J. Virol. 80, 10128–10138.

Rossi, S.L., Fayzulin, R., Dewsbury, N., Bourne, N., Mason, P.W., 2007. Mutations in West Nile virus nonstructural proteins that facilitate replicon persistence in vitro attenuate virus replication in vitro and in vivo. Virology 364, 184–195.

Rossi, S.L., Zhao, Q., O'Donnell, V.K., Mason, P.W., 2005. Adaptation of West Nile virus replicons to cells in culture and use of replicon-bearing cells to probe antiviral action. Virology 331, 457–470.

Selisko, B., Dutartre, H., Guillemot, J.C., Debarnot, C., Benarroch, D., Khromykh, A., Desprès, P., Egloff, M.P., Canard, B., 2006. Comparative mechanistic studies of de novo RNA synthesis by flavivirus RNA-dependent RNA polymerases. Virology 351, 145–158.

Seneviratne, S.L., Malavige, G.N., de Silva, H.J., 2006. Pathogenesis of liver involvement during dengue viral infections. Trans. R. Soc. Trop. Med. Hyg. 100, 608–614.

Shi, P.Y., Tilgner, M., Lo, M.K., 2002. Construction and characterization of subgenomic replicons of New York strain of West Nile virus. Virology 296, 219–233.

Sidwell, R.W., Huffman, J.H., Khare, G.P., Allen, L.B., Witkowski, J.T., Robins, R.K., 1972. Broad-spectrum antiviral activity of Virazole: 1-beta-D-ribofuranosyl-1,2,4-triazole-3-carboxamide. Science 177, 705–706.

Streeter, D.G., Witkowski, J.T., Khare, G.P., Sidwell, R.W., Bauer, R.J., Robins, R.K., Simon, L.N., 1973. Mechanism of action of 1- -D-ribofuranosyl-1,2,4-triazole-3-carboxamide (Virazole), a new broad-spectrum antiviral agent. Proc. Natl. Acad. Sci. U.S.A. 70, 1174–1178.

Suzuki, R., de Borba, L., Duarte dos Santos, C.N., Mason, P.W., 2007. Construction of an infectious cDNA clone for a Brazilian prototype strain of dengue virus type 1: characterization of a temperature-sensitive mutation in NS1. Virology 362, 374–383.

Takhampunya, R., Ubol, S., Houng, H.S., Cameron, C.E., Padmanabhan, R., 2006. Inhibition of dengue virus replication by mycophenolic acid and ribavirin. J. Gen. Virol. 87, 1947–1952.

Tilgner, M., Deas, T.S., Shi, P.Y., 2005. The flavivirus-conserved penta-nucleotide in the 3' stem-loop of the West Nile virus genome requires a specific sequence and structure for RNA synthesis, but not for viral translation. Virology 331, 375–386.

Tilgner, M., Shi, P.Y., 2004. Structure and function of the 3' terminal six nucleotides of the west nile virus genome in viral replication. J. Virol. 78, 8159–8171.

Yap, T.L., Xu, T., Chen, Y.L., Malet, H., Egloff, M.P., Canard, B., Vasudevan, S.G., Lescar, J., 2007. Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. J. Virol. 81, 4753–4765.

You, S., Padmanabhan, R., 1999. A novel in vitro replication system for Dengue virus. Initiation of RNA synthesis at the 3'-end of exogenous viral RNA templates requires 5'- and 3'-terminal complementary sequence motifs of the viral RNA. J. Biol. Chem. 274, 33714–33722.

Zhang, J.H., Chung, T.D., Oldenburg, K.R., 1999. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J. Biomol. Screen. 4, 67–73.

Antiviral Research 96 (2012) 21–31

Contents lists available at SciVerse ScienceDirect

Antiviral Research

journal homepage: www.elsevier.com/locate/antiviral

Review

The viral RNA capping machinery as a target for antiviral drugs

François Ferron, Etienne Decroly, Barbara Selisko, Bruno Canard [*]

Centre National de la Recherche Scientifique and Aix-Marseille Université, UMR 7257, Architecture et Fonction des Macromolécules Biologiques, 163 Avenue de Luminy, 13288 Marseille Cedex 09, France

ARTICLE INFO

Article history:
Received 21 May 2012
Revised 9 July 2012
Accepted 13 July 2012
Available online 26 July 2012

Keywords:
5′-Triphosphatase
Guanylyltransferase
Methyltransferase
Endonuclease
Cap snatching
Mechanism
Structure
Inhibitor

ABSTRACT

Most viruses modify their genomic and mRNA 5′-ends with the addition of an RNA cap, allowing efficient mRNA translation, limiting degradation by cellular 5′–3′ exonucleases, and avoiding its recognition as foreign RNA by the host cell. Viral RNA caps can be synthesized or acquired through the use of a capping machinery which exhibits a significant diversity in organization, structure and mechanism relative to that of their cellular host. Therefore, viral RNA capping has emerged as an interesting field for antiviral drug design. Here, we review the different pathways and mechanisms used to produce viral mRNA 5′-caps, and present current structures, mechanisms, and inhibitors known to act on viral RNA capping.

Contents

[*] Corresponding author. Tel.: +33 491 82 86 44; fax: +33 491 82 86 46.
E-mail address: bruno.canard@afmb.univ-mrs.fr (B. Canard).

0166-3542/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.antiviral.2012.07.007

1. Introduction

In the eukaryotic cell, RNA capping is a co-transcriptional event consisting of a chemical modification of the nascent mRNA 5′-end. Since the early 70s, viruses have played a pivotal role in the discovery and structural characterization of the RNA cap (Fig. 1), as well as in the mechanistic elucidation of the RNA capping pathway. Incidentally, virus families for which these discoveries were made (*Reoviridae* and *Poxviridae*) use an RNA capping pathway that turned out to be the same as that of their eukaryotic hosts. Soon after these discoveries, other virus families were found to deviate substantially from this 'conventional' pathway (see below).

In the conventional pathway, a cap structure is added to the nascent 5′-triphosphate mRNA in a series of reactions (Fig. 2). The 5′-triphosphate is first hydrolysed by an RNA 5′-triphosphatase (RTPase). A guanylyltransferase (GTase), also called "capping enzyme", adds the cap structure under the form of a guanosine 5′-monophosphate in a 5′–5′ orientation. The cap is then methylated onto the N-7 position of its guanine by an RNA cap guanine N-7-methyltransferase (N-7 MTase). This generates the minimal cap-0 (m7GpppN...), found in metazoan and lower eukaryotes. In higher eukaryotes, further methylation by ribose 2′-O-methyltransferases (2′-O MTases) occurs at the 2′-position of the riboses of the original transcript to yield mainly cap-1 (m7GpppNmN...) but also cap-2 (m7GpppNmNmN...) structures.

The presence of the cap ensures stability of the transcript against a variety of cellular 5′–3′ exonucleases and recognition of the mRNA by the ribosomal protein eIF4E for efficient translation. The capping was also shown to be involved in other cellular processes such as RNA splicing and export (Darnell, 1979; Filipowicz et al., 1976; Schibler and Perry, 1977).

Because of its coupling to RNA transcription, RNA capping is mainly a nuclear process, although some RNA re-capping events are suspected to occur in the cytoplasm (reviewed in Schoenberg and Maquat (2009)). Viruses generally replicate in the cytoplasm, causing there to be a time-window during which viral RNAs are synthesized but not yet capped. Virus and cell co-evolution has generated a number of cellular pathways and proteins involved in sensing the presence of viral RNAs. The absence of RNA cap as well as the presence of double stranded RNA are strong tokens for a viral infection. These RNA species, alone or together, are detected as "non-self" RNA by cellular sensors triggering an innate cellular immunity response (Koyama et al., 2008; Takeuchi and Akira, 2007; Wilkins and Gale, 2010). Viruses have evolved numerous strategies to escape detection, including rapid and efficient viral RNA capping.

Few virus families and genera do not rely on RNA capping. *Picornaviridae* use a protein as an RNA synthesis primer and this protein replaces the RNA cap in its role for transcription promotion and RNA protection. Viruses from *Pestivirus* and *Hepacivirus* genera use unprotected 5′-triphosphate RNA ends and other strategies to defend their RNA from the cell immunity systems (Garaigorta and Chisari, 2009; Guidotti and Chisari, 2001; Malmgaard, 2004).

However, the vast majority of viruses use RNA capping. With the ongoing deciphering of viral RNA capping machineries, a true diversity of mechanisms, partners, and pathway organizations which invariably leads to the same RNA structure are progressively being uncovered. This diversity and its differences from the cellular RNA capping machineries are drawing a lot of attention for antiviral drug design.

2. Is RNA capping an appropriate target for antiviral research?

There are factors to bear in mind before considering RNA capping as an interesting drug design target:

Specificity: a most often put forward requirement is the uniqueness of the viral target, i.e., the non-existence of a similar cellular target that could also be hit by any antiviral drug and cause serious side-effects. Interestingly, even when viral enzymes remain close in structure and mechanism to their cellular counterpart, there remain structural and functional differences potentially useful to achieve differential inhibition, i.e., selectivity for the viral target. In most cases, viral enzymes are profoundly original in folding, organization, and mechanisms, providing a large chemical space for drug design and drug selectivity.

Potency: another important parameter is the expected outcome of viral target inhibition. The question of whether the *in vitro* inhibition effectively leads to a significant block of viral growth must be answered. One has to consider the two major mechanisms of action of the viral target. It can be an enzyme, and inhibition of its enzyme

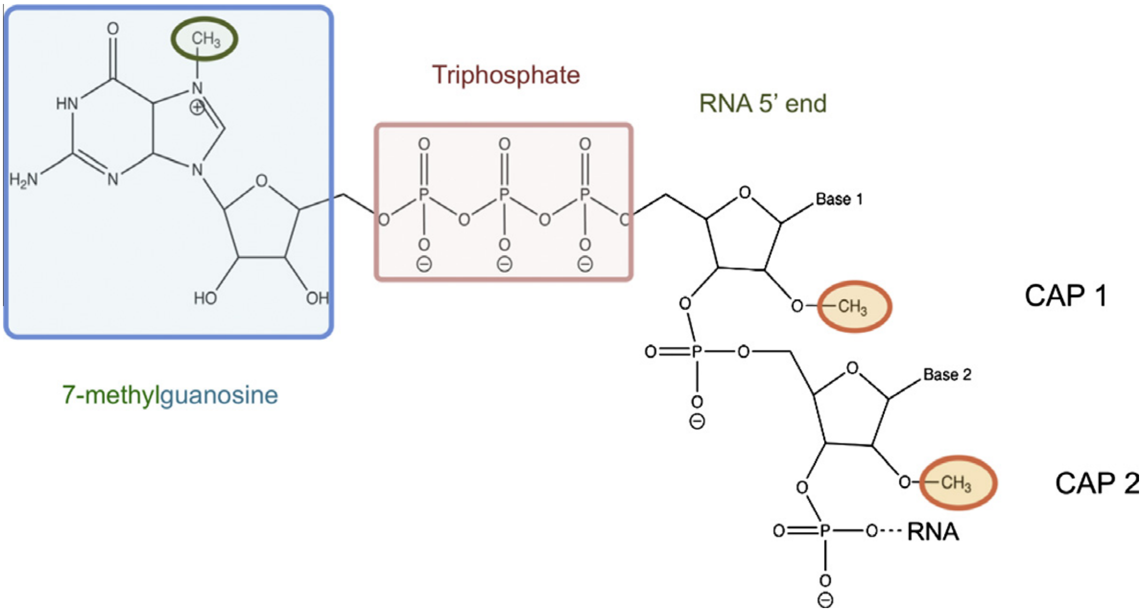


Fig. 1. *The RNA cap structure:* the cap consists of a 7-methylguanosine (blue box) linked to the 5′ nucleoside of the messenger RNA chain through a 5′–5′ triphosphate bridge (pink box). The methyl group of the guanosine at its N-7 position is surrounded in green, and the 2′-O methyl group of the first and second nucleotide residue forming the cap-1 and the cap2 structures, respectively, are surrounded in light orange.

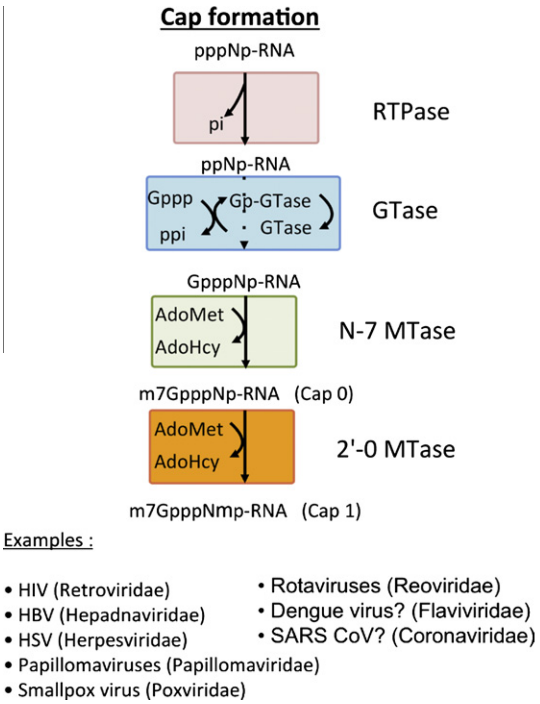


Fig. 2. *Canonical RNA capping mechanisms.* The cap-0 structure is formed on nascent RNA chains by the sequential action of three enzymes. (1) The RNA triphosphatase (RTPase, pink) hydrolyses the γ phosphate of the nascent RNA (pppN-RNA, where N denotes the first transcribed nucleotide) to yield a diphosphate RNA (ppN-RNA) and inorganic phosphate (P_i). (2) RNA guanylyltransferase (GTase, light blue) reacts with the α phosphate of GTP releasing pyrophosphate (PP_i) and forms a covalent enzyme–guanylate intermediate (Gp-GTase). The GTase then transfers the GMP molecule (Gp) to the 5′ diphosphate RNA to create GpppN-RNA. (3) RNA (guanine-N-7)-methyltransferase (N-7 MTase, green), recruited by the GTase, transfers the methyl group from *S*-adenosyl-L-methionine (AdoMet) to the cap guanine to form the cap-0 structure (^{m7}GpppN) and releases *S*-adenosyl-L-homocysteine (AdoHcy) as a by-product. The capping reaction is then completed by the methylation of the ribose-2′-O position of the first nucleotide by the AdoMet-dependant (nucleoside-2′-O)-methyltransferase (2′-O MTase, orange), generating cap-1 structure (^{m7}GpppN$_{m2'O}$). Below the schematic are listed examples of viruses that acquire their cap structures using the cellular capping machinery (first four examples) or encode their own viral capping machinery that adopt the canonic pathway. Question marks indicate viruses likely to follow this conventional pathway.

activity may have exponential consequences. On the other hand, the virus may develop alternative pathways in order to resist antiviral molecules. For example, *Flavivirus* capping inhibitors had to be validated since dengue virus was also shown to perform cap-independent translation of its RNA genome (Edgil et al., 2006). Another possibility is that the target protein can be a binding partner devoid of catalytic activity and part of a protein–protein binding equilibrium. Its inhibition would shift the equilibrium according to the law of mass action. The protein dosage must be fine tuned to exert a powerful antiviral effect. In both cases, a significant effect on viral growth must be observed to validate the target as appropriate.

Amplifying mechanism: lastly, one has to consider the number of events that the protein or target is involved in during the virus lifecycle. For example, RNA-dependent RNA polymerases (RdRp) incorporate several thousand nucleotides to produce a single RNA genome. Inhibition of the viral RdRp at each nucleotide incorporation should exhibit a powerful antiviral effect. In contrast, RNA capping events can vary from a single capping event (e.g., the genome of ss(+)RNA viruses), to many RNA capping events (e.g., *Nidovirales*, *Mononegavirales*, etc.). Nevertheless, the connection between RNA capping and antiviral innate immunity pathways is expected to amplify any drug action on RNA capping. Immature or incomplete RNA caps may not only exhibit an altered expression profile, but may also be detected by a variety of innate immunity sensors that add another direct negative effect on viral growth. In this context, it is noteworthy that the efficacy of RNA capping inhibition has been demonstrated. The inhibition of *Flavivirus* MTase, which catalyzes two methylations per synthesis of an entire RNA genome, has been reported to be able to completely suppress viral replication (Dong et al., 2008a).

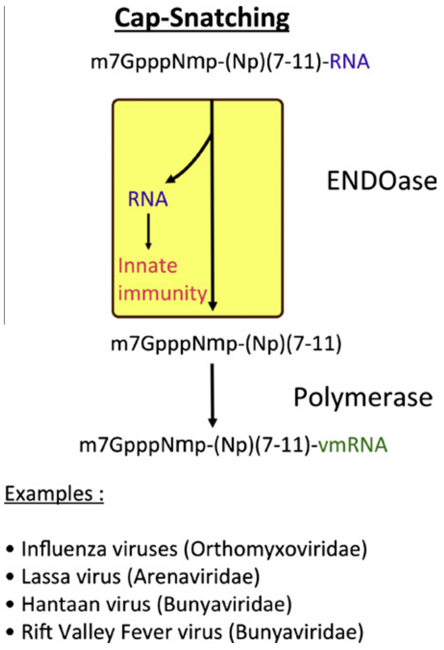

Cap-Snatching
m7GpppNmp-(Np)(7-11)-RNA
RNA
Innate immunity
ENDOase
m7GpppNmp-(Np)(7-11)
Polymerase
m7GpppNmp-(Np)(7-11)-vmRNA
Examples :
• Influenza viruses (Orthomyxoviridae)
• Lassa virus (Arenaviridae)
• Hantaan virus (Bunyaviridae)
• Rift Valley Fever virus (Bunyaviridae)

Fig. 3. *Cap-snatching mechanisms.* The endonuclease captures the cellular mRNA from the cell (blue) and snatches (cleaves and transfers) the cap structure followed by seven to eleven bases to the polymerase to start the viral mRNA (green). The residual cellular RNA serves as decoy for the innate immunity. The RNAs capped by viral enzymes are undistinguishable from cellular mRNA and can thus be translated into proteins by the host-cell ribosomal machinery. Below the schematic are listed examples of viruses that acquire their cap structures using the cap snatching mechanism.

3. Conventional and unconventional viral RNA capping mechanisms

Alongside the discovery of conventional cellular and viral RNA capping (Fig. 2), few viruses provided remarkable exceptions to the seemingly ubiquitous presence of capped mRNAs in Eukarya (as mentioned above for picorna-, hepaci- and pestiviruses). It was soon discovered that other viruses also used ‘unconventional’ RNA capping reactions deviating from the conventional RNA capping scheme (example in Fig. 3 and for complete review (Decroly et al., 2012)). Nevertheless, it is truly remarkable that several of these widely different RNA capping reactions and pathways converge to the same RNA cap structure (Fig. 4). This observation indicates that the evolutionary pressure to keep this structure protecting RNA must be significant. Viral RNA capping is thus an interesting target for drug design. In the following sections we summarize our current knowledge on viral enzymes involved in conventional and unconventional RNA capping (main reference structures in Fig. 5) and reported inhibitors (Fig. 6).

4. Conventional viral RNA cap-synthesizing enzymes

4.1. RNA triphosphatases

RTPases (Fig. 5, RTPase) are the enzymes responsible for the first step of cap formation, hydrolyzing the γ–β phosphodiester bond of

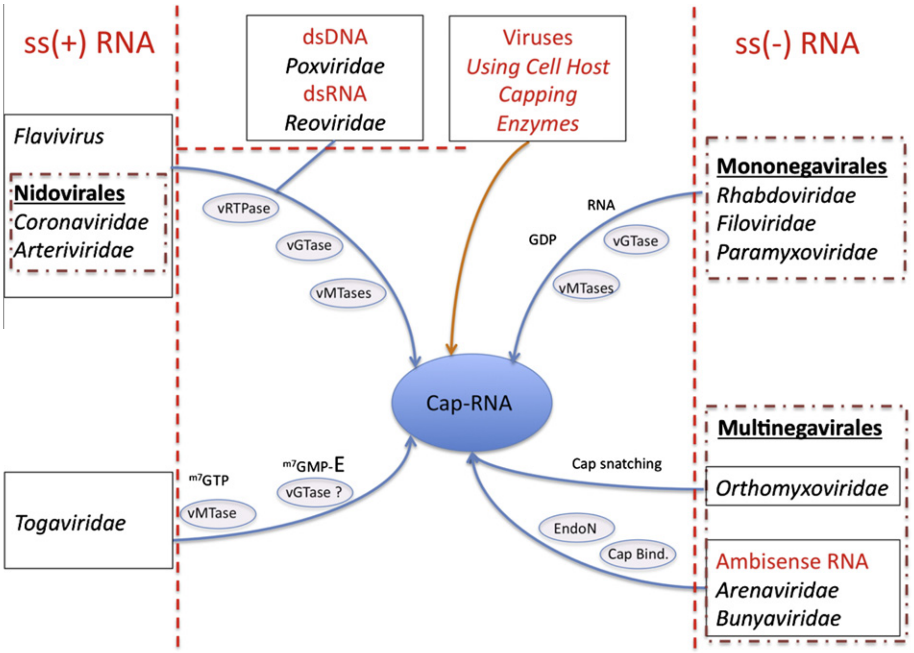


Fig. 4. *Viral RNA capping pathways.* Mammalian viruses, with the exception of ssRNA(+) pestiviruses and ssRNA(+) hepaciviruses, use strategies to chemically modify their mRNA 5′-ends through either covalent attachment of a protein (Viral Protein genome-linked, VPg, for ssRNA(+) -*Picornaviridae*, -*Caliciviridae*, and -*Astroviridae*), or covalent attachment of an RNA cap structure (all other viruses). When possible the viruses were divided by red dashed lines according to the nature of their viral genome (single stranded, double stranded, DNA, or RNA), and virus families and genera are written in italics. Three virus orders (ssRNA(+) *Nidovirales* and ssRNA(−) *Mononegavirales* and *Multinegavirales* (ssRNA(−) and ambisense RNA) are shown in bold and underlined. Enzymes and mechanisms are indicated along arrows leading to Cap-RNA. Small letters "v" ahead of the enzyme name indicate that the enzyme is of viral origin. The list of presented families is non exhaustive and used as an example.

5′-triphosphorylated mRNA. There is no enzyme conservation amongst different kingdoms of life, meaning that the activity can be found within a domain or a single protein presenting structural and mechanistic diversities (for detailed review Decroly et al. (2012)). Theses diversities make the RTPase an attractive target for the development of specific and selective inhibitors. RTPases can indeed be separated into two classes sorted by mechanistic criteria, i.e., either with a metal-independent and -dependent mechanism (Gu and Lima, 2005; Shuman, 2002). This separation also divides higher eucaryote RTPases (including human) from the vast majority of viral RTPases.

The metal-independent RTPase is so far a lone structural family, including human RTPase, higher eukaryotes (plants) and viruses like baculovirus. The substrate binding and catalytic site of these enzymes is located in a P-loop enriched with histidines and a cysteine that mediates a two-step reaction with a covalent phosphoenzyme intermediate (Changela et al., 2001; Denu and Dixon, 1998).

Alternatively, for viral metal-dependent RTPases a structural and mechanistic variety was found over the years. Metal-dependent RTPases can be organized into three superfamilies: the HIT-like fold (α–β complex) (Jayaram et al., 2002), the 'triphosphate tunnel metalloenzyme' (TTM) (Benarroch et al., 2008; Gu and Lima, 2005; Lima et al., 1999; Shuman, 2001), and the viral RNA helicase-like fold that carries one or more of the so-called Walker motifs (Benarroch et al., 2004b).

The HIT-like fold (Fig. 5, RTPase III) found in *Rotavirus* NSP2 protein has a magnesium dependent NTPase/RTPase hydrolysis activity capable of removing the γ-phosphate of either NTP or RNA (Jayaram et al., 2002). Both activities share the same catalytic site and mechanism. Also, the switch from one activity to the other is substrate dependent (NTP or pppRNA). NSP2 is a protein that self-assembles into a doughnut-shape octamer, which binds to single-stranded RNA after destabilizing RNA–RNA duplexes. Each NSP2 monomer has two subdomains separated by a deep electropositive cleft containing histidine residues involved in the binding and the hydrolysis of NTP as well as RNA (Vasquez-Del Carpio et al., 2006). The cleft is oriented on the surface of the doughnut. The N-terminal subdomain is mostly α-helical while the C-terminal domain adopts an α/β fold with a central twisted and anti-parallel β-sheet made of 5 β-strands, which is flanked by 5 α-helices. The protein is both a structural component of the RNA packaging and part of the capping machinery. Due to its unique structural features, it should be considered as a target of choice for the development of antiviral compounds.

Enzymes from the TTM superfamily hydrolyze NTPs to NDP + Pi in the presence of manganese or cobalt and are found in fungi, protozoa, and most of DNA viruses that encode a RTPase (*Orthopoxvirus*, Chlorovirus, *Baculoviridae* and *Mimivirus*). The RTPase of *Saccharomyces cerevisiae* Cet1 serves as a model (Fig. 5, RTPase I) and presents a structural tunnel composed of eight antiparallel β strands with a motif encompassing two glutamates (Lehman et al., 1999; Lima et al., 1999). The tunnel harbors several charged and hydrophilic side chains coordinating manganese and sulfate ions. The sulfate is thought to indicate the position of the γ-phosphate of the newly synthesized mRNA. In a docking/modeling-based study, a series of nucleoside analogues (6-chloropurine-riboside-5′-triphosphate, 6-methylthioguanosine-5′-triphosphate, or 8-iodo-guanosine-5′-triphosphate) was identified with high affinity for binding and resistant to hydrolysis (Despins et al., 2010; Issur et al., 2009a).

The last identified superfamily is the viral RNA helicase-like fold (Fig. 5, RTPase II), which carries NTPase-helicase activity. These RTPases are found in the RNA virus genera or families *Flavivirus* (NS3), *Coronaviridae*, *Orthoreovirus*, *Alphavirus* and *Potexvirus.* Structurally, they can belong to either SF1 or SF2 helicase superfamily. No viral helicase crystal structure belonging to the SF1 superfamily has been determined to date, yet the fold is thought to be divergent from the canonical SF1 (Cordin et al., 2006). On the other hand, the SF2 fold is formed by two RecA-like sub-do-

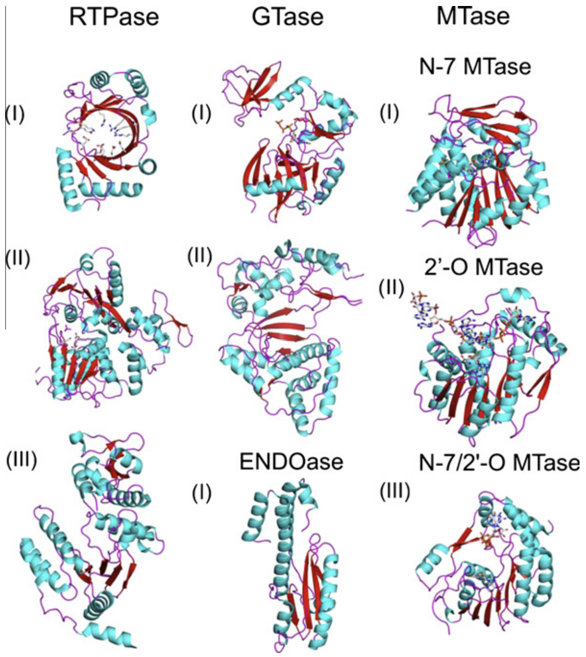


Fig. 5. *Structural models of viral enzymes involved in RNA capping. RTPase folds:* (I) the triphosphate tunnel metalloenzyme (TTM) fold is examplified by the structure of the RTPase (PDB code 2QZE) from the genus *Mimivirus*, which consists of double-stranded DNA (dsDNA) viruses; (II) the helicase fold of the RTPase from the single-stranded positive-sense RNA (ss(+)RNA) dengue virus (PDB code 2BHR); (III) the histidine triad (HIT)-like fold of the RTPase from dsRNA rotaviruses (PDB code 1L9V). *GTase folds:* (I) Structure of the dsDNA virus *Paramecium bursaria* chlorella virus 1 GTase, showing the loading reaction: GTP onto the active site, with the GTase in an open conformation (PDB code 1CKO); (II) GTase domain of the Cypovirus (PDB code 3IZ3, residues 1–366). *MTases folds*: (I) N-7-MTase: the (guanine-N-7)-methyltransferase (N-7 MTase) domain of protein D1 from the dsDNA virus vaccinia virus (PDB accession code 2VDW) in complex with a molecule of *S*-adenosyl-L-homocysteine (AdoHcy). II) 2′-O MTase: VP39, the 2′-O MTase of the dsDNA virus vaccinia virus, in complex with a capped RNA and AdoHcy (PDB code 1AV6). III) N-7/2′-O methyltransferase: the N-7 MTase-(nucleoside-2′-O)-methyltransferase (2’-O MTase) domain of NS5 from dengue virus (ss(+)RNA) in complex with the cap analogue 7-methyl-G (m7G)-pppGm2′-O and AdoHCY (PDB code 2P41). *Endonuclease fold*: (I) endonuclease (N-terminal) domain of L protein of lymphocytic choriomeningitis virus (PDB code 3JSB). Structures are colored cyan for α-helices and red for β-strands. All figures were prepared using PyMOL.

mains, separated by a cleft that can accommodate either a nucleotide or 5′-triphosphate RNA substrate. The two sub-domains carry the Walker motifs: the A motif forms the P-loop stabilizing the terminal phosphate moiety of the substrate, and the B motif is composed of acidic residues (D-E-x-D) coordinating the divalent ion needed for substrate hydrolysis.

All the catalytic sites described above, or clefts accommodating the RNA substrate or nucleotides could be used as a target for drug design. However, more efforts are needed to better understand the fine mechanistics that would lead to rational drug design.

4.1.1. Inhibitors

Amongst the different families of viral RTPases, very few inhibition studies have identified potentially interesting compounds (Fig. 6). In that respect, the least neglected RTPase is cvRTP1, from the DNA *Paramecium bursaria Chlorella* virus 1. Not surprisingly, phosphate analogues (see Fig. 6, RTPase inhibitor I and II, also vanadate) are weak inhibitors, with no “drug-like” properties (Takagi et al., 2003). The family of *Flavivirus* NS3-like enzymes may have greater value, since in addition to the RTPase activity they habour two other activities (helicase and NTPase). The RTPase active site is super-imposable to the NTPase active site, which provides energy to unwind dsRNA. Suppression of the RTPase/NTPase activity abrogates helicase activity, therefore, it is likely that bi-functional inhibitors will be discovered in the near future (Lescar et al., 2008). One example for this approach has been a study of the inhibitory potential of purine analogues (one structure in Fig. 6, RTPase inhibitors III) that are expected to inhibit all three activities (Despins et al., 2010).

4.2. Guanylyltransferases

GTases (Fig. 5, GTase) are responsible for attaching a GMP molecule onto the pre-mRNA. Contrary to RTPases, fold and mechanism seem to be very well conserved amongst all kingdoms of life, making *de facto* the GTase a difficult target for specific inhibitor development. Yet there is a possibility that in the viral kingdom, GTases exhibit other mechanisms and new folds that would make GTases good candidates for drug development. This assumption comes from both the difficulty in identifying GTases in the viral world and the presence of GTase domains embedded within larger proteins such as protein L in Mononegavirales or the variant GTase domain found in cypovirus (Cheng et al., 2011) (Fig. 5, GTase II). Known GTase proteins so far adopt a modular organization containing an N-terminal nucleotide transferase domain with an ATP-grasp fold and a C-terminal domain with an oligonucleotide/oligosaccharide binding fold (OB-fold) that is positioned as a lid over the base subdomain of the N-terminal domain (Fig. 5, GTase I). The GTP-binding site is located between the base and the hinge, and is highly conserved in GTases from dsDNA viruses to humans. The catalytic site can be defined by the presence of six conserved motifs, for which Motif I contains a lysine that covalently links GMP by hydrolysis of ppi from GTP, before its transfer onto RNA. Other conserved motifs form the binding pocket for the nucleotide. GTases from the ds RNA viruses of the *Orthoreovirus* genus form part of a multidomain protein (or assembly line) carrying all RNA capping functions. The GTase domain probably exists transiently and no fold can be clearly defined, whereas the methyltransferase domains are identifiable. Nevertheless, the catalytic residue is also a conserved lysine (Sutton et al., 2007).

4.2.1. Inhibitors

Very few compounds have been described that could provide a useful start for drug design. In the case of DNA viruses, the enzyme shares structural and functional similarities to the ligase protein superfamily, raising potential difficulties in achieving selectivity inside the host cell. The pyrophosphate analogue foscarnet (a phosphonic acid derivative (Fig. 6, GTase inhibitor I)) has been shown to inhibit the reaction, not surprisingly since the reaction produces pyrophosphate (Soulière et al., 2008). In the case of several viral families the GTase is still unknown (coronaviruses, flaviviruses). In the flaviviruses, the GTase activity might be expressed by the N-terminal domain of protein NS5 carrying also other related RNA capping activities (MTases, Issur et al., 2009b). Therefore, structural and mechanistic studies may provide efficient drug leads targeting several activities at once as it is the case for Ribavirin, a guanine nucleotide analogue, targeting GTase and MTase (Benarroch et al., 2004a; Bougie and Bisaillon, 2004).

4.3. Methyltransferases

The last step of the cap formation consists in the methylation of the cap by RNA cap methyltransferases (Fig. 5, MTase). The N-7 position of guanine is methylated by the N-7 MTase and the first nucleotide of the RNA transcript is further methylated at the ribose 2′-OH position by 2′-O MTase (see Fig. 1). *S*-adenosyl-L-methionine (SAM) is the methyl donor for both the N-7 and 2′-O methylations, generating *S*-adenosyl-L-homocysteine (SAH) as a by-product. The two methylations are either done by specific proteins (or domains)

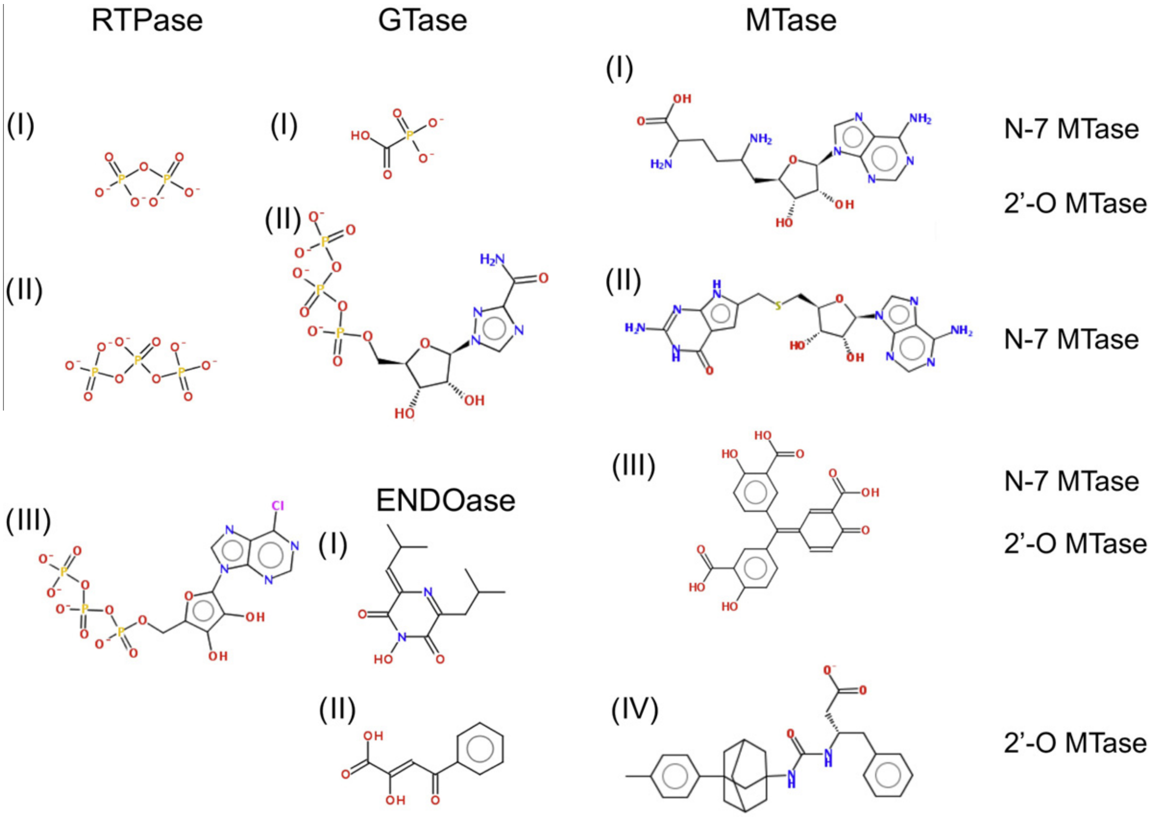


Fig. 6. *Capping enzymes inhibitors.* Inhibitors structure examples. *RTPase inhibitors*: (I) PPi (II) PPPi (III) 6-chloropurine-ribose-5′-triphosphate. *GTase inhibitors*: (I) foscarnet (II) ribavirin triphosphate. *MTase inhibitors*: (I) sinefungin (II) dAPPMA 2′-O MTase inhibitors: (III) aurintricaboxylic acid (IV) inhibitor 7. *Endonuclease inhibitors*: (I) DPBA (II) Flutimide.

(*Poxviridae, Coronaviridae, Reoviridae*) or by one single protein (or domain) (*Flavivirus*).

The canonical SAM-dependent MTase fold presents alternating β strands and α helices, that form a seven-stranded β sheet with at least three parallel α helices on each side (α/β fold). The structure is reminiscent of the Rossmann fold also found in the dinucleotide-binding domains of dinucleotide-binding proteins. The SAM-binding region is located at the N-terminal part of the β sheet and is formed in part by residues from loops following the central strands of the sheet. The substrate-binding region that is responsible for selectively binding nucleic acids is located in the C-terminal part of the β sheet. Depending on nature of the methylation (N-7 or 2′-O or both) additional structural features can be found to help accommodate or stabilize the substrate.

The N-7 MTase of vaccinia virus (VV) serves as structural and reference model of N-7 MTases. The activity is carried by the complex of the C-terminal domain of protein D1 and protein D12. Both the domain of D1 and D12 present the characteristic core α/β fold of the MTase family (Fig. 5, N-7 MTase I), but the D12 protein lacks a proper SAM binding site and does not show any activity on its own. In contrast, it stimulates the MTase activity of D1 by 30- to 50-fold. D12 does not affect the extent of substrate binding to the catalytic subunit. Rather, it has an allosteric role increasing the affinity for ligands as well as increasing the overall stability of the complex. Recently, similar finding have shown that the SARS 2′-O MTase activity harbored in protein nsp16 is stimulated by a small regulatory protein nsp10 (Bouvet et al., 2010; Decroly et al., 2011).

VP39 protein of VV serves as reference for 2′-O MTase structures. It is a single-domain protein with a typical α/β fold (Fig. 5 2′-O MTase II). This structure has all the features of the canonical core domain. The catalytic residues of 2′-O MTases are defined by a conserved tetrad motif K-D-K-E located at the interface of SAM-binding cavity and the substrate-binding cleft (Fig. 5, 2′-O MTase II). VP39 exhibits a hydrophobic pocket, at the end of the RNA binding cleft, which presents aromatic side chain residues involved in stacking the capped RNA.

Single domain MTases are exemplified by the *Flavivirus* MTase domain of protein NS5. The NS5 protein of flaviviruses is divided into two domains, an MTase residing in the N-terminal (one-third of NS5) and an RNA-dependent RNA polymerase domain in the C-terminal (two-thirds of NS5). The NS5 MTase domain was first unambiguously characterized as a 2′-O MTase (Egloff et al., 2002) and later it was shown that the same SAM-binding site is used for N-7 MTase activity (Ray et al., 2006). The core structure of the domain consists of the 7-stranded β-sheet surrounded by four α-helices (Fig. 5, N-7/2′-O MTase III) and its catalytic site is identified by the universal tetrad K-D-K-E (Fig. 5, N-7/2’-O MTase III). The core of the structure is not supposed to rearrange, yet optimal N-7 and 2′-O methylations of the flavivirus cap require distinct biochemical conditions, suggesting that the two methylations occur through different mechanisms. Mutagenesis of the tetrad has shown that only the D of the K-D-K-E catalytic tetrad is essential for the N-7 methylation, and structural studies are still investigating the determination of the complete catalytic site for the N-7 MTase. Ray and colleagues (Ray et al., 2006) have shown that the two cap methylation events are sequential and that the reaction is controlled by a steric constraint for the substrates. Indeed the N-7 methylation requires wild-type nucleotides at the 2nd (G) and 3rd (U) positions, with a 5′-stem–loop structure, whilst the subsequent 2′-O methylation requires wild-type nucleotides at the 1st (A) and 2nd (G) positions and a minimum of 20 nucleotides of viral RNA.

4.3.1. Inhibitors

The exploration of MTases constitutes the most actively growing field amongst viral RNA capping targets. The last 10 years have seen viral RNA cap MTases coming increasingly into focus as targets for the discovery and development of antivirals (Dong et al., 2008b; Issur et al., 2011; Liu et al., 2010). The inhibition of N-7 MTase activity is expected to exert a deleterious effect on viral replication because it blocks viral RNA translation. The essentiality of the N-7-methylation for viral replication has clearly been demonstrated for several viral families (Almazan et al., 2006; Dong et al., 2008a; Ray et al., 2006; Zhou et al., 2007). In contrast, mutational analysis indicates that 2′-O MTase inhibition has weak effects on virus replication in cell culture (Ray et al., 2006; Zhou et al., 2007). Nevertheless, 2′-O MTases are now also considered as potential targets due to the demonstration that RNA, lacking 2’-O-methylation on their cap structure, induces antiviral response through the MDA5 sensor (Züst et al., 2011). Viruses mutated in their 2′-O MTase gene accordingly show strongly reduced replication capacity in animal models (Daffis et al., 2010). This has led to a re-evaluation of the importance of 2′-O-methylation (reviewed in (Decroly et al., 2012)).

The discovery of inhibitors of viral MTases started and the elucidation of RNA capping started in the dsDNA *Poxviridae* field. A SAM analogue, sinefungin (Fig. 6, MTase inhibitor I), was reported as the first potent inhibitor of the VV N-7 MTase activity (Pugh et al., 1978) with a potent antiviral effect shown in a plaque formation assay. Another early study described inhibition of N-7 MTase and 2′-O MTase activities of VV by 2′,5′-linked oligo(adenylic acid) triphosphates (Sharma and Goswami, 1981). Interestingly, these molecules are “natural” metabolites produced by the host cell as part of the innate immune response against viral infection (reviewed in Decroly et al. (2012)). Later, multi substrate adducts with inhibitory activity against VV N-7 MTase were reported (Fig. 6, MTase inhibitor II) (Benghiat et al., 1986). SAH analogues were shown to inhibit VV and herpes simplex virus (HSV) and their respective MTases were proposed as potential molecular targets (Balzarini et al., 1992).

Concerning RNA virus MTases being part of the conventional capping machinery, *Flavivirus* and *Coronavirus* MTases have been particularly explored. The dengue virus 2′-O MTase activity was shown to be inhibited by the GTP-analogue ribavirin triphosphate (Benarroch et al., 2004a), sinefungin, and SAH (Selisko et al., 2010). Based on known complex structures of the dengue virus NS5 MTase domain (Benarroch et al., 2004a; Egloff et al., 2007, 2002) which define the RNA cap binding sites during 2′-O-methylation and N-7-methylation (the latter being the actual active site) as well as the SAM-binding site, several groups have employed virtual screening and docking procedures to identify potential ligands and thus inhibitors (Lim et al., 2011; Luzhkov et al., 2007; Milani et al., 2009; Podvinec et al., 2010). Some molecules were found to present inhibitory activity in the micromolar range in *in vitro* enzymatic assays (Luzhkov et al., 2007; Milani et al., 2009; Podvinec et al., 2010), see example in Fig. 6, MTase inhibitor IV). A new conserved hydrophobic binding pocket next to the SAM-binding site was identified on *Flavivirus* MTases, which is critical for viral replication and cap methylations (Dong et al., 2010). This invites to the design of SAM analogues that also interact with this adjacent hydrophobic site and should be thus more potent and specific in comparison to SAM analogues as sinefungin. High-throughput screens using MTase activity assays and GTP-binding assays were set up and used to identify inhibitor molecules (Geiss et al., 2011; Lim et al., 2008). Concerning *Coronavirus* MTases the search for antivirals has been less active. Some SAM analogues have been tested and found to inhibit the N-7 MTase activity of nsp14 and the 2′-O MTase activity of nsp16 in the low micromolar range (Bouvet et al., 2010). One of the identified molecules, aurintricarboxylic acid (ATA, (Fig. 6, MTase inhibitor III)), also inhibits SARS-CoV replication in infected cells (He et al., 2004). An indirect approach to inhibit viral MTases is being considered for *Coronavirus* 2′-O MTase nsp16, the activity of which depends on its interaction with the nsp10 protein (Bouvet et al., 2010; Decroly et al., 2011; Lugari et al., 2010). Molecules that disrupt the interaction are expected to inhibit the MTase activity of nsp16 and thus viral replication. This approach can also be considered for the N-7 MTase activity of protein D1 of VV, which depends on the activation by the stimulatory protein D12 (De la Peña et al., 2007).

4.4. Cap assembly lines

Cap-assembly lines are proteins found in several dsRNA viruses (*Reoviridae*). They are multi-domain and function proteins that are packaged with the genome in the viral particle and are able to perform the four reactions needed to synthesize a cap-1 structure. Domains of the cap assembly line must be tightly regulated and coordinated to accomplish the sequential steps of the cap synthesis pathway as they can exhibit activity independent from or in synergy with the capping process. The individual domains corresponding to GTase and MTase have been characterized and have similar folds to those previously described, although it is unclear how and when RTPase activity occurs. A complete pathway has been proposed in which guanylyl transfer is followed by N-7-methylation and 2′-O-methylation of the mRNA.

Each domain/function could be targeted individually. The entire protein is also to be considered, as there is room for specific development to sterically impair the necessary dynamic movements of the respective domains. So far, though, no specific inhibitors of cap assembly lines or either of their associated enzyme activities have been described.

5. Unconventional cap synthesis pathways

The first indication for the existence of deviations from the conventional viral RNA capping pathway came in the early 1970s, around the time of the discovery of the RNA cap structure. Since then, it has been demonstrated that vesicular stomatitis virus (VSV, Mononegavirales) and alphaviruses (*Togaviridae*) can synthesize a viral RNA cap that is identical to a cellular RNA cap, albeit through two completely different mechanisms. *Togaviridae* do not proceed further than synthesizing a cap-0 structure, which remains an interesting enigma since many members of this family productively infect both insects and mammals.

5.1. The Mononegavirales RNA capping pathway

The *Mononegavirales* order, also referred to as negative-strand (−) non-segmented (NNS) RNA viruses comprises major human pathogens such as rabies virus (*Rhabdoviridae*), measles virus (*Paramyxovirinae*), bornavirus (*Bornaviridae*) and Ebola/Marburg viruses (*Filoviridae*). RNA capping is achieved by the multifunctional L protein, which carries both RNA-dependent RNA polymerase (RdRp) and RNA cap synthesis activities. Shortly after the discovery of the RNA cap and conventional capping pathway, it was observed that the VSV L protein (Abraham et al., 1975a, b) transferred GDP rather than GMP onto the nascent transcript. Subsequently, L proteins from spring viremia of carp virus (Gupta and Roy, 1980), human respiratory syncytial virus (Barik, 1993), and chandipura virus (Ogino and Banerjee, 2010) were shown to perform a similar reaction and presumably follow the same RNA capping pathway.

This discovery adds a novel enzyme to the process as the pathway involves a unique L-encoded GDP polyribonucleotidyltransferase activity (PRNTase), which forms a covalent enzyme-p-RNA intermediate involving a phosphoamidate bond. A conserved catalytic histidine is present in a "HR" motif instead of lysine used by conventional GTases (Ogino and Banerjee, 2010). GDP generated from GTP (Ogino and Banerjee, 2007) by a yet-unknown NTPase is then transferred to the 5′-monophophorylated viral mRNA 5′-end covalently attached to the PRNTase. Interestingly, the methylation sequence also is different from the conventional pathway sequence. The RNA cap is first methylated at the ribose 2′-O position of the first nucleotide followed by the guanine N-7 position, generating an RNA cap-1 structure. No crystal structure is available yet to guide drug design against these unique enzymes. The PRNTase should be considered as a prime target, perhaps more amenable to antiviral design than the more common MTase fold.

5.1.1. Inhibitors

Most effort for antiviral design has been made for RSV, for which a sizeable market exists. Interestingly, two compound families having submicromolar activities (IC_{50} = 21–200 nM) have been described (Liuzzi et al., 2005; Sudo et al., 2005) that target the L protein. The most potent compounds of respective series have been used to elicit resistance and the resistance mutation map to an L region or domain consistent with the presence of the PRNTase. Although the PRNTase was discovered after the publication of these compounds, mechanistic studies clearly indicated that the guanylylation step was the actual target of the compounds. For RSV, only two products are approved for antiviral therapy or prevention: a monoclonal antibody (Palivizumab, Synagis) and ribavirin (Empey et al., 2010; Vigant and Lee, 2011). The latter served as a control compound with moderate antiviral effect in the discovery of PRNTase inhibitors, but the precise mechanism of action of ribavirin (i.e., capping, error catastrophe, IMP dehydrogenase, or else (Leyssen et al., 2006; Severson et al., 2003; Zhou et al., 2003)) against *Mononegavirales* remains to be investigated.

5.2. The ssRNA(+) Togaviridae (alpha-like) RNA capping pathway

The *Togaviridae* (ss(+)RNA viruses) as exemplified by some alphaviruses (Semliki Forest virus, sindbis virus and chikungunya virus) synthesize their cap-0 structure through another non-conventional mechanism (Fig. 4). The enzymes presumably involved in the capping pathway are poorly characterized and their crystal structures are not yet available. Nevertheless, it is likely that capping begins with N-7-guanine methylation of a GTP molecule by the nsp1 N-7 MTase. This methylation is seemingly followed by the formation of a covalent m^7GMP–enzyme complex involving a conserved histidine catalytic residue (Ahola et al., 1997). The N-7-methylation of GTP seems to precede the formation of the enzyme–GMP complex, since the guanylate intermediate is not observed in the absence of AdoMet (Ahola and Kääriäinen, 1995; Ahola et al., 1997). Whereas the m^7GMP transfer onto the 5′ end of viral RNA was not yet formally demonstrated, it is expected that the 5′ end of viral RNA bears a diphosphate. The hydrolysis of the β–γ phosphate bond at the 5′ end is mediated by the nsp2 RTPase domain (Vasiljeva et al., 2000). Brome mosaic virus replicase protein 1A (Ahola and Ahlquist, 1999), bamboo mosaic virus ORF1 protein (Li et al., 2001), tobacco mosaic virus P126 (Merits et al., 1999) and hepatitis E virus p110 (Magden et al., 2001) were reported to share properties with alphavirus N-7 MTases and GTases. Interestingly, Sindbis virus and Semliki Forest virus were reported more than 30 year ago to contain additional methyl groups attached to the exocyclic N2 of the cap structure (HsuChen and Dubin, 1976; van Duijn et al., 1986). The role of this methylation in virus replication is still unknown, but it is reminiscent to 2,2,7-trimethylguanosine (TMG) cap that is found on non-coding eukaryal RNAs such as small nuclear (sn), small nucleolar (sno) RNAs, and telomerase RNA (Busch et al., 1982; Seto et al., 1999).

5.2.1. Inhibitors

Ribavirin shows broad spectrum *in vitro* inhibitory activity against RNA viruses, through presumably different modes of action (Leyssen et al., 2006; Severson et al., 2003; Zhou et al., 2003). In the case of alphaviruses, resistance mutants to ribavirin were mapped into the GTase domain of the nsP1 protein (Scheidel et al., 1987; Scheidel and Stollar, 1991) suggesting ribavirin interferes with the GTase activity. Nevertheless, the ribavirin action mode remains controversial. Ribavirin also reduces the intracellular concentration of GTP (Leyssen et al., 2005) and/or stimulates the expression of interferon-stimulated genes such as that of protein MDA5 (Thomas et al., 2011), which is known to sense viral RNA devoid of 2’-O-methylation (Decroly et al., 2012; Züst et al., 2011).

In addition some efforts have focused on the identification of specific inhibitors targeting specifically the nsP1 capping activity. Several GTP analogues have been reported to inhibit Semliki forest virus MTase and GTase activities with K_i values below 100 μM (Lampio et al., 1999).

6. Virus-mediated RNA cap ‘snatching’

6.1. Enzymes from the cap-snatching pathway

The cap snatching mechanism is a common and alternative viral strategy for cap formation. Instead of making its own RNA cap, the virus has a machinery to remove and transfer (snatch) the cap from host mRNA onto its own pre-mRNA (Fig. 3). Three major human pathogen families, *Arenaviridae*, *Bunyaviridae* and *Orthomyxoviridae* (Fig. 4), all of them ss(−)RNA viruses, have developed this strategy (Bouloy et al., 1978; Caton and Robertson, 1980; Plotch et al., 1979). The *Orthomyxoviridae* serves as a paradigm for both mechanistic and structural studies. In the case of the influenza virus, the replication complex is made of three proteins forming the polymerase complex (PA, PB1, PB2). Host cell mRNAs present in the cytoplasm are targeted and recruited by PB2 that presents the cap from the mRNA to the endonuclease domain (PA) to snip the cap off. Then the short capped RNA is used as primer to polymerize the viral RNA. The first structural endonuclease domain potentially involved in cap snatching was recently identified in both *Arenaviridae*, and *Bunyaviridae* L protein (Morin et al., 2010; Reguera et al., 2010) (ENDOase, Fig. 5). The endonuclease fold features four mixed β-strands forming a twisted sheet surrounded by seven α-helices (Fig. 5). The β sheet forms the bottom of a negatively charged cavity creating a binding site for divalent cations. Above it, a C-terminal helix with a positive patch creates another pocket to accommodate the RNA where the typical conserved PD...(D/E)XK nuclease motif defines the catalytic site.

A fourth example of cap snatching has recently been described in *Totiviridae* for a fungal virus (Fujimura and Esteban, 2011). However, in this case the Totivirus L-A transfers only m7GMP snatched from cellular mRNA. Hence, this mechanism lies somewhere between alphaviruses, which transfer m7GMP acquired from GTP, and "conventional" cap-snatching viruses described above, which transfer longer capped RNA oligonucleotides. As an additional similarity to alphaviruses, the Totivirus L-A cap-snatching enzyme makes a covalent m7GMP-enzyme intermediate through a histidine residue (Fujimura and Esteban, 2011; Ahola et al., 1997).

6.1.1. Inhibitors

Ribavirin again is an effective inhibitor of *Arenaviridae* (Olschlager et al., 2011) and *Bunyaviridae* (Livonesi et al., 2006; Severson et al., 2003), but it has pleïotropic effects as discussed above and thus there is no direct evidence that the RNA cap snatching process is an actual target.

The first discovery of inhibitors of the cap snatching pathway was made in 1994, when Tomassini et al. reported that 2,4-dioxobutanoic acid compounds were able to inhibit Influenza A and B viruses in infected cell cultures (Tomassini et al., 1994). Remarkably, the compounds exhibited IC_{50} values in the micromolar range, comparable to IC_{50} values observed when purified polymerase cores were used in an *in vitro* endonuclease assay. The following chemical diversification of the parent compound could achieve a high specificity for influenza. Indeed, the L-735,882 compound reached an IC_{50} value of 1.1 μM for influenza in an *in vitro* endonuclease assay and an IC_{50} value of 2 μM in a viral titer assay using influenza virus infected cells. The La Crosse virus, a *Bunyaviridae* for which RNA capping and RNA polymerase priming rely also on cap snatching, remained unaffected both at the enzyme and virus level. Interestingly, the La Crosse virus replicates in the cytoplasm exclusively, unlike the influenza virus. Therefore, RNA caps are likely snatched in different cellular compartments or environments by these two viruses and this should be considerated at early steps of drug design for large spectrum inhibitors.

In 2009 and 2010, a structural basis for the inhibition mechanisms was provided by the crystal structure of the La Crosse virus endonuclease domains in complex with DPBA (Reguera et al., 2010) (Fig. 6, ENDOase inhibitor I)). The structural homology between the three available crystal structures indicates that a common pharmacophore including Mn^{2+} chelating functional groups might be active against all three viral families. Based on these crystal structures, several compound families have been described and their antiviral activities characterized using enzyme-based assays, viral growth assays and protein structure data. The first compounds benefiting from structure-based activity analysis are catechins isolated from green tea. Although they seem to target both neuraminidase (Song et al., 2005) and endonuclease, one active pharmacophore was identified as 3,4-dihydroxyphenyl. The same group extended the family of interesting compounds with thalidomide derivatives, active through their 3,4-dihydroxyphenethyl moieties, and later with marchantin-like phytochemicals produced in very high yield in liverworts (Iwai et al., 2011). Together with enzyme and virus-infected cell assays, docking the latter compounds into the PA endonuclease active site gave very interesting perspectives for effective drugs targeting influenza RNA capping.

7. Conclusions

The viral RNA capping machinery is remarkably diverse in organization, structure, and mechanisms used by its enzymes. All RNA viruses capping their own viral RNAs have evolved enzymes profoundly different from those of their host cell. As the power of antiviral compound screening and design increases, there are no doubts that viral RNA capping enzymes will be the target of novel highly efficient and selective drugs. Despite few exceptions, the identification and atomic resolution structure of target enzymes from major families has made significant progress in the last ten years. The recent connection of viral RNA capping to the host cell innate immunity mechanisms is more than ever a promising antiviral research field.

Acknowledgements

We are indebt to all previous and current lab members, too numerous to name, from the virology team and the UMR7257 for their contribution and help in the study of viral RNA replication and capping.

This work was supported in part through funding by the Fondation pour la Recherche Médicale (Programme Aide aux équipes), the Direction Générale de l'Armement (Contrat 07co404), the Infectiopôle-Sud, and the European Union Seventh Framework Programme (FP7/2007-2013) through the Project SILVER (Small Inhibitor Leads against Emerging RNA viruses, Grant Agreement No. 260644).

References

Abraham, G., Rhodes, D.P., Banerjee, A.K., 1975a. The 5′ terminal structure of the methylated mRNA synthesized in vitro by vesicular stomatitis virus. Cell 5, 51–58.

Abraham, G., Rhodes, D.P., Banerjee, A.K., 1975b. Novel initiation of RNA synthesis in vitro by vesicular stomatitis virus. Nature 255, 37–40.

Ahola, T., Ahlquist, P., 1999. Putative RNA capping activities encoded by brome mosaic virus: methylation and covalent binding of guanylate by replicase protein 1a. J. Virol. 73, 10061–10069.

Ahola, T., Kääriäinen, L., 1995. Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. Proc. Nat. Acad. Sci. USA 92, 507–511.

Ahola, T., Laakkonen, P., Vihinen, H., Kääriäinen, L., 1997. Critical residues of Semliki Forest virus RNA capping enzyme involved in methyltransferase and guanylyltransferase-like activities. J. Virol. 71, 392–397.

Almazan, F., Dediego, M.L., Galan, C., Escors, D., Alvarez, E., Ortego, J., Sola, I., Zuniga, S., Alonso, S., Moreno, J.L., Nogales, A., Capiscol, C., Enjuanes, L., 2006. Construction of a severe acute respiratory syndrome coronavirus infectious cDNA clone and a replicon to study coronavirus RNA synthesis. J. Virol. 80, 10900–10906.

Balzarini, J., De Clercq, E., Serafinowski, P., Dorland, E., Harrap, K.R., 1992. Synthesis and antiviral activity of some new *S*-adenosyl-L-homocysteine derivatives. J. Med. Chem. 35, 4576–4583.

Barik, S., 1993. The structure of the 5′ terminal cap of the respiratory syncytial virus mRNA. J. General Virol. 74 (Pt 3), 485–490.

Benarroch, D., Egloff, M.P., Mulard, L., Guerreiro, C., Romette, J.L., Canard, B., 2004a. A structural basis for the inhibition of the NS5 dengue virus mRNA 2′-O-methyltransferase domain by ribavirin 5′-triphosphate. J. Biol. Chem. 279, 35638–35643.

Benarroch, D., Selisko, B., Locatelli, G.A., Maga, G., Romette, J.-L., Canard, B., 2004b. The RNA helicase, nucleotide 5′-triphosphatase, and RNA 5′-triphosphatase activities of Dengue virus protein NS3 are Mg2+-dependent and require a functional Walker B motif in the helicase catalytic core. Virology 328, 208–218.

Benarroch, D., Smith, P., Shuman, S., 2008. Characterization of a trifunctional mimivirus mRNA capping enzyme and crystal structure of the RNA triphosphatase domain. Structure (London, England: 1993) 16, 501–512.

Benghiat, E., Crooks, P.A., Goodwin, R., Rottman, F., 1986. Inhibition of vaccinia RNA guanine 7-methyltransferase by compounds designed as multisubstrate adducts. J. Pharm. Sci. 75, 142–145.

Bougie, I., Bisaillon, M., 2004. The broad spectrum antiviral nucleoside ribavirin as a substrate for a viral RNA capping enzyme. J. Biol. Chem. 279, 22124–22130.

Bouloy, M., Plotch, S.J., Krug, R.M., 1978. Globin mRNAs are primers for the transcription of influenza viral RNA in vitro. Proc. Natl. Acad. Sci. USA 75, 4886–4890.

Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E.J., Canard, B., Decroly, E., 2010. In vitro reconstitution of SARS-coronavirus mRNA cap methylation. PLoS Pathog. 6, e1000863.

Busch, H., Reddy, R., Rothblum, L., Choi, Y.C., 1982. SnRNAs, SnRNPs, and RNA processing. Annu. Rev. Biochem. 51, 617–654.

Caton, A.J., Robertson, S., 1980. Structure of the host-derived sequences present at the 5′ ends of influenza virus mRNA. Nucleic Acids Res. 8, 2591.

Changela, A., Ho, C.K., Martins, A., Shuman, S., Mondragón, A., 2001. Structure and mechanism of the RNA triphosphatase component of mammalian mRNA capping enzyme. EMBO J. 20, 2575–2586.

Cheng, L., Sun, J., Zhang, K., Mou, Z., Huang, X., Ji, G., Sun, F., Zhang, J., Zhu, P., 2011. Atomic model of a cypovirus built from cryo-EM structure provides insight into the mechanism of mRNA capping. Proc. Nat. Acad. Sci. USA 108, 1373–1378.

Cordin, O., Banroques, J., Tanner, N.K., Linder, P., 2006. The DEAD-box protein family of RNA helicases. Gene 367, 17–37.

Daffis, S., Szretter, K.J., Schriewer, J., Li, J., Youn, S., Errett, J., Lin, T.-Y., Schneller, S., Zust, R., Dong, H., Thiel, V., Sen, G.C., Fensterl, V., Klimstra, W.B., Pierson, T.C., Buller, R.M., Gale, M., Shi, P.-Y., Diamond, M.S., 2010. 2′-O Methylation of the viral mRNA cap evades host restriction by IFIT family members. Nature 468, 452–456.

Darnell Jr., J.E., 1979. Transcription units for mRNA production in eukaryotic cells and their DNA viruses. Prog. Nucleic Acid Res. Mol. Biol. 22, 327–353.

De la Peña, M., Kyrieleis, O.J.P., Cusack, S., 2007. Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap N7 methyl-transferase. EMBO J. 26, 4913–4925.

Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., Gluais, L., Papageorgiou, N., Sharff, A., Bricogne, G., Ortiz-Lombardia, M., Lescar, J., Canard, B., 2011. Crystal structure and functional analysis of the SARS-coronavirus RNA Cap 2′-O-methyltransferase nsp10/nsp16 complex. PLoS Pathog. 7, e1002059.

Decroly, E., Ferron, F., Lescar, J., Canard, B., 2012. Conventional and unconventional mechanisms for capping viral mRNA. Nat. Rev. Microbiol. 10, 51–65.

Denu, J.M., Dixon, J.E., 1998. Protein tyrosine phosphatases: mechanisms of catalysis and regulation. Curr. Opin. Chem. Biol. 2, 633–641.

Despins, S., Issur, M., Bougie, I., Bisaillon, M., 2010. Deciphering the molecular basis for nucleotide selection by the West Nile virus RNA helicase. Nucleic Acids Res. 38, 5493–5506.

Dong, H., Liu, L., Zou, G., Zhao, Y., Li, Z., Lim, S.P., Shi, P.Y., Li, H., 2010. Structural and functional analyses of a conserved hydrophobic pocket of flavivirus methyltransferase. J. Biol. Chem. 285, 32586–32595.

Dong, H., Ren, S., Zhang, B., Zhou, Y., Puig-Basagoiti, F., Li, H., Shi, P.Y., 2008a. West Nile virus methyltransferase catalyzes two methylations of the viral RNA cap through a substrate-repositioning mechanism. J. Virol. 82, 4295–4307.

Dong, H., Zhang, B., Shi, P.Y., 2008b. Flavivirus methyltransferase: a novel antiviral target. Antiviral Res. 80, 1–10.

Edgil, D., Polacek, C., Harris, E., 2006. Dengue virus utilizes a novel strategy for translation initiation when cap-dependent translation is inhibited. J. Virol. 80, 2976–2986.

Egloff, M.-P., Decroly, E., Malet, H., Selisko, B., Benarroch, D., Ferron, F., Canard, B., 2007. Structural and functional analysis of methylation and 5′-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. J. Mol. Biol. 372, 723–736.

Egloff, M.P., Benarroch, D., Selisko, B., Romette, J.L., Canard, B., 2002. An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. EMBO J. 21, 2757–2768.

Empey, K.M., Peebles Jr., R.S., Kolls, J.K., 2010. Pharmacologic advances in the treatment and prevention of respiratory syncytial virus. Clin. Infect. Dis. 50, 1258–1267.

Filipowicz, W., Furuichi, Y., Sierra, J.M., Muthukrishnan, S., Shatkin, A.J., Ochoa, S., 1976. A protein binding the methylated 5′-terminal sequence, m7G pppN, of eukaryotic messenger RNA. Proc. Nat. Acad. Sci. USA 73, 1559–1563.

Fujimura, T., Esteban, R., 2011. Cap-snatching mechanism in yeast L-A double-stranded RNA virus. Proc. Nat. Acad. Sci. USA 108, 17667–17671.

Garaigorta, U., Chisari, F.V., 2009. Hepatitis C virus blocks interferon effector function by inducing protein kinase R phosphorylation. Cell Host Microbe 6, 513–522.

Geiss, B.J., Stahla-Beek, H.J., Hannah, A.M., Gari, H.H., Henderson, B.R., Saeedi, B.J., Keenan, S.M., 2011. A high-throughput screening assay for the identification of flavivirus NS5 capping enzyme GTP-binding inhibitors: implications for antiviral drug development. J. Biomol. Screen. 16, 852–861.

Gu, M., Lima, C.D., 2005. Processing the message: structural insights into capping and decapping mRNA. Curr. Opin. Struct. Biol. 15, 99–106.

Guidotti, L.G., Chisari, F.V., 2001. Noncytolytic control of viral infections by the innate and adaptive immune response. Annu. Rev. Immunol. 19, 65–91.

Gupta, K.C., Roy, P., 1980. Alternate capping mechanisms for transcription of spring viremia of carp virus: evidence for independent mRNA initiation. J. Virol. 33, 292–303.

He, R., Adonov, A., Traykova-Adonova, M., Cao, J., Cutts, T., Grudesky, E., Deschambaul, Y., Berry, J., Drebot, M., Li, X., 2004. Potent and selective inhibition of SARS coronavirus replication by aurintricarboxylic acid. Biochem. Biophys. Res. Commun. 320, 1199–1203.

HsuChen, C.C., Dubin, D.T., 1976. Di-and trimethylated congeners of 7-methylguanine in Sindbis virus mRNA. Nature 264, 190–191.

Issur, M., Despins, S., Bougie, I., Bisaillon, M., 2009a. Nucleotide analogs and molecular modeling studies reveal key interactions involved in substrate recognition by the yeast RNA triphosphatase. Nucleic Acids Res. 37, 3714–3722.

Issur, M., Geiss, B.J., Bougie, I., Picard-Jean, F., Despins, S., Mayette, J., Hobdey, S.E., Bisaillon, M., 2009b. The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure. RNA (New York, NY) 15, 2340–2350.

Issur, M., Picard-Jean, F., Bisaillon, M., 2011. The RNA capping machinery as an anti-infective target. Wiley Interdisciplinary Reviews: RNA 2, 184–192.

Iwai, Y., Murakami, K., Gomi, Y., Hashimoto, T., Asakawa, Y., Okuno, Y., Ishikawa, T., Hatakeyama, D., Echigo, N., Kuzuhara, T., 2011. Anti-influenza activity of marchantins, macrocyclic bisbibenzyls contained in liverworts. PLoS ONE 6, e19825.

Jayaram, H., Taraporewala, Z., Patton, J.T., Prasad, B.V.V., 2002. Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold. Nature 417, 311–315.

Koyama, S., Ishii, K.J., Coban, C., Akira, S., 2008. Innate immune response to viral infection. Cytokine 43, 336–341.

Lampio, A., Ahola, T., Darzynkiewicz, E., Stepinski, J., Jankowska-Anyszka, M., Kaariainen, L., 1999. Guanosine nucleotide analogs as inhibitors of alphavirus mRNA capping enzyme. Antiviral Res. 42, 35–46.

Lehman, K., Schwer, B., Ho, C.K., Rouzankina, I., Shuman, S., 1999. A conserved domain of yeast RNA triphosphatase flanking the catalytic core regulates self-association and interaction with the guanylyltransferase component of the mRNA capping apparatus. J. Biol. Chem. 274, 22668–22678.

Lescar, J., Luo, D., Xu, T., Sampath, A., Lim, S.P., Canard, B., Vasudevan, S.G., 2008. Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. Antiviral Res. 80, 94–101.

Leyssen, P., Balzarini, J., De Clercq, E., Neyts, J., 2005. The predominant mechanism by which ribavirin exerts its antiviral activity in vitro against flaviviruses and paramyxoviruses is mediated by inhibition of IMP dehydrogenase. J. Virol. 79, 1943–1947.

Leyssen, P., De Clercq, E., Neyts, J., 2006. The anti-yellow fever virus activity of ribavirin is independent of error-prone replication. Mol Pharmacol 69, 1461–1467.

Li, Y.I., Chen, Y.J., Hsu, Y.H., Meng, M., 2001. Characterization of the AdoMet-dependent guanylyltransferase activity that is associated with the N terminus of bamboo mosaic virus replicase. J. Virol. 75, 782–788.

Lim, S.P., Wen, D., Yap, T.L., Yan, C.K., Lescar, J., Vasudevan, S.G., 2008. A scintillation proximity assay for dengue virus NS5 2′-O-methyltransferase-kinetic and inhibition analyses. Antiviral Res. 80, 360–369.

Lim, S.V., Rahman, M.B., Tejo, B.A., 2011. Structure-based and ligand-based virtual screening of novel methyltransferase inhibitors of the dengue virus. BMC Bioinformatics 12 (Suppl. 13), S24.

Lima, C.D., Wang, L.K., Shuman, S., 1999. Structure and mechanism of yeast RNA triphosphatase: an essential component of the mRNA capping apparatus. Cell 99, 533–543.

Liu, L., Dong, H., Chen, H., Zhang, J., Ling, H., Li, Z., Shi, P.-Y., Li, H., 2010. Flavivirus RNA cap methyltransferase: structure, function, and inhibition. Front. Biol. 5, 286–303.

Liuzzi, M., Mason, S.W., Cartier, M., Lawetz, C., McCollum, R.S., Dansereau, N., Bolger, G., Lapeyre, N., Gaudette, Y., Lagace, L., Massariol, M.J., Do, F., Whitehead, P., Lamarre, L., Scouten, E., Bordeleau, J., Landry, S., Rancourt, J., Fazal, G., Simoneau, B., 2005. Inhibitors of respiratory syncytial virus replication target cotranscriptional mRNA guanylylation by viral RNA-dependent RNA polymerase. J. Virol. 79, 13105–13115.

Livonesi, M.C., De Sousa, R.L., Badra, S.J., Figueiredo, L.T., 2006. In vitro and in vivo studies of ribavirin action on Brazilian Orthobunyavirus. Am. J. Trop. Med. Hyg. 75, 1011–1016.

Lugari, A., Betzi, S., Decroly, E., Bonnaud, E., Hermant, A., Guillemot, J.-C., Debarnot, C., Borg, J.-P., Bouvet, M., Canard, B., Morelli, X., Lecine, P., 2010. Molecular mapping of the RNA Cap 2[prime]-O-methyltransferase activation interface between SARS coronavirus nsp10 and nsp16. J. Biol. Chem. 285, 33230–33241.

Luzhkov, V.B., Selisko, B., Nordqvist, A., Peyrane, F., Decroly, E., Alvarez, K., Karlen, A., Canard, B., Qvist, J., 2007. Virtual screening and bioassay study of novel inhibitors for dengue virus mRNA cap (nucleoside-2′O)-methyltransferase. Bioorg. Med. Chem. 15, 7795–7802.

Magden, J., Takeda, N., Li, T., Auvinen, P., Ahola, T., Miyamura, T., Merits, A., Kaariainen, L., 2001. Virus-specific mRNA capping enzyme encoded by hepatitis E virus. J. Virol. 75, 6249–6255.

Malmgaard, L., 2004. Induction and regulation of IFNs during viral infections. J. Interferon Cytokine Res. 24, 439–454.

Merits, A., Kettunen, R., Makinen, K., Lampio, A., Auvinen, P., Kaariainen, L., Ahola, T., 1999. Virus-specific capping of tobacco mosaic virus RNA: methylation of GTP prior to formation of covalent complex p126–m7GMP. FEBS Lett. 455, 45–48.

Milani, M., Mastrangelo, E., Bollati, M., Selisko, B., Decroly, E., Bouvet, M., Canard, B., Bolognesi, M., 2009. Flaviviral methyltransferase/RNA interaction: structural basis for enzyme inhibition. Antiviral Res. 83, 28–34.

Morin, B., Coutard, B., Lelke, M., Ferron, F., Kerber, R., Jamal, S., Frangeul, A., Baronti, C., Charrel, R., de Lamballerie, X., Vonrhein, C., Lescar, J., Bricogne, G., Günther, S., Canard, B., 2010. The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription. PLoS Pathog. 6 (9), e1001038.

Ogino, T., Banerjee, A.K., 2007. Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. Mol. Cell 25, 85–97.

Ogino, T., Banerjee, A.K., 2010. The HR motif in the RNA-dependent RNA polymerase L protein of Chandipura virus is required for unconventional mRNA-capping activity. J. General Virol. 91, 1311–1314.

Olschlager, S., Neyts, J., Gunther, S., 2011. Depletion of GTP pool is not the predominant mechanism by which ribavirin exerts its antiviral effect on Lassa virus. Antiviral Res. 91, 89–93.

Plotch, S.J., Bouloy, M., Krug, R.M., 1979. Transfer of 5′-terminal cap of globin mRNA to influenza viral complementary RNA during transcription in vitro. Proc. Natl. Acad. Sci. USA 76, 1618–1622.

Podvinec, M., Lim, S.P., Schmidt, T., Scarsi, M., Wen, D., Sonntag, L.S., Sanschagrin, P., Shenkin, P.S., Schwede, T., 2010. Novel inhibitors of dengue virus methyltransferase: discovery by in vitro-driven virtual screening on a desktop computer grid. J. Med. Chem. 53, 1483–1495.

Pugh, C.S., Borchardt, R.T., Stone, H.O., 1978. Sinefungin, a potent inhibitor of virion mRNA(guanine-7-)-methyltransferase, mRNA(nucleoside-2′-)-methyltransferase, and viral multiplication. J. Biol. Chem. 253, 4075–4077.

Ray, D., Shah, A., Tilgner, M., Guo, Y., Zhao, Y., Dong, H., Deas, T.S., Zhou, Y., Li, H., Shi, P.Y., 2006. West Nile virus 5′-cap structure is formed by sequential guanine N-7 and ribose 2′-O methylations by nonstructural protein 5. J. Virol. 80, 8362–8370.

Reguera, J., Weber, F., Cusack, S., 2010. Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. PLoS Pathog. 6 (9), e1001101.

Scheidel, L.M., Durbin, R.K., Stollar, V., 1987. Sindbis virus mutants resistant to mycophenolic acid and ribavirin. Virology 158, 1–7.

Scheidel, L.M., Stollar, V., 1991. Mutations that confer resistance to mycophenolic acid and ribavirin on Sindbis virus map to the nonstructural protein nsP1. Virology 181, 490–499.

Schibler, U., Perry, R.P., 1977. The 5′-termini of heterogeneous nuclear RNA: a comparison among molecules of different sizes and ages. Nucleic Acids Res. 4, 4133–4149.

Schoenberg, D.R., Maquat, L.E., 2009. Re-capping the message. Trends Biochem. Sci. 34, 435–442.

Selisko, B., Peyrane, F.F., Canard, B., Alvarez, K., Decroly, E., 2010. Biochemical characterization of the (nucleoside-2′O)-methyltransferase activity of dengue virus protein NS5 using purified capped RNA oligonucleotides (7Me)GpppAC(n) and GpppAC(n). J. General Virol. 91, 112–121.

Seto, A.G., Zaug, A.J., Sobel, S.G., Wolin, S.L., Cech, T.R., 1999. *Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle. Nature 401, 177–180.

Severson, W.E., Schmaljohn, C.S., Javadian, A., Jonsson, C.B., 2003. Ribavirin causes error catastrophe during Hantaan virus replication. J. Virol. 77, 481–488.

Sharma, O.K., Goswami, B.B., 1981. Inhibition of vaccinia mRNA methylation by 2′,5′-linked oligo(adenylic acid) triphosphate. Proc. Nat. Acad. Sci. USA 78, 2221–2224.

Shuman, S., 2001. Structure, mechanism, and evolution of the mRNA capping apparatus. Prog. Nucleic Acid Res. Mol. Biol. 66, 1–40.

Shuman, S., 2002. What messenger RNA capping tells us about eukaryotic evolution. Nat. Rev. Mol. Cell Biol. 3, 619–625.

Song, J.M., Lee, K.H., Seong, B.L., 2005. Antiviral effect of catechins in green tea on influenza virus. Antiviral Res. 68, 66–74.

Soulière, M.F., Perreault, J.-P., Bisaillon, M., 2008. Kinetic and thermodynamic characterization of the RNA guanylyltransferase reaction. Biochemistry 47, 3863–3874.

Sudo, K., Miyazaki, Y., Kojima, N., Kobayashi, M., Suzuki, H., Shintani, M., Shimizu, Y., 2005. YM-53403, a unique anti-respiratory syncytial virus agent with a novel mechanism of action. Antiviral Res. 65, 125–131.

Sutton, G., Grimes, J.M., Stuart, D.I., Roy, P., 2007. Bluetongue virus VP4 is an RNA-capping assembly line. Nat. Struct. Mol. Biol. 14, 449–451.

Takagi, T., Walker, A.K., Sawa, C., Diehn, F., Takase, Y., Blackwell, T.K., Buratowski, S., 2003. The Caenorhabditis elegans mRNA 5′-capping enzyme. In vitro and in vivo characterization. J. Biol. Chem. 278, 14174–14184.

Takeuchi, O., Akira, S., 2007. Recognition of viruses by innate immunity. Immunol. Rev. 220, 214–224.

Thomas, E., Feld, J.J., Li, Q., Hu, Z., Fried, M.W., Liang, T.J., 2011. Ribavirin potentiates interferon action by augmenting interferon-stimulated gene induction in hepatitis C virus cell culture models. Hepatology 53, 32–41.

Tomassini, J., Selnick, H., Davies, M.E., Armstrong, M.E., Baldwin, J., Bourgeois, M., Hastings, J., Hazuda, D., Lewis, J., McClements, W., 1994. Inhibition of cap (m7G pppXm)-dependent endonuclease of influenza virus by 4-substituted 2,4-dioxobutanoic acid compounds. Antimicrob. Agents Chemother. 38, 2827–2837.

van Duijn, L.P., Kasperaitis, M., Ameling, C., Voorma, H.O., 1986. Additional methylation at the N(2)-position of the cap of 26S Semliki Forest virus late mRNA and initiation of translation. Virus Res. 5, 61–66.

Vasiljeva, L., Merits, A., Auvinen, P., Kääriäinen, L., 2000. Identification of a novel function of the alphavirus capping apparatus. RNA 5′-triphosphatase activity of Nsp2. J. Biol. Chem. 275, 17281–17287.

Vasquez-Del Carpio, R., Gonzalez-Nilo, F.D., Riadi, G., Taraporewala, Z.F., Patton, J.T., 2006. Histidine triad-like motif of the rotavirus NSP2 octamer mediates both RTPase and NTPase activities. J. Mol. Biol. 362, 539–554.

Vigant, F., Lee, B., 2011. Hendra and nipah infection: pathology, models and potential therapies. Infect. Disord. Drug Targets 11, 315–336.

Wilkins, C., Gale Jr., M., 2010. Recognition of viruses by cytoplasmic sensors. Curr. Opin. Immunol. 22, 41–47.

Zhou, S., Liu, R., Baroudy, B.M., Malcolm, B.A., Reyes, G.R., 2003. The effect of ribavirin and IMPDH inhibitors on hepatitis C virus subgenomic replicon RNA. Virology 310, 333–342.

Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z., Guo, Y., Bernard, K.A., Shi, P.Y., Li, H., 2007. Structure and function of flavivirus NS5 methyltransferase. J. Virol. 81, 3891–3903.

Züst, R., Cervantes-Barragan, L., Habjan, M., Maier, R., Neuman, B.W., Ziebuhr, J., Szretter, K.J., Baker, S.C., Barchet, W., Diamond, M.S., Siddell, S.G., Ludewig, B., Thiel, V., 2011. Ribose 2′-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. Nat. Immunol. 12, 137–143.