Folding Mechanism of the SH3 Domain from Grb2

Francesca Troilo, Daniela Bonetti, Carlo Camilloni, Angelo Toto, Sonia Longhi, Maurizio Brunori, Stefano Gianni

The folding mechanism of the SH3 domain from Grb2

Journal:	*The Journal of Physical Chemistry*
Manuscript ID	Draft
Manuscript Type:	Special Issue Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Troilo, Francesca; Universita degli Studi di Roma La Sapienza Bonetti, Daniela; Universita degli Studi di Roma La Sapienza Camilloni, Carlo; Universita degli Studi di Milano, Deparment of Biosciences Toto, Angelo; Università di Roma, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli" and Istituto di Biologia e Patologia Molecolari del CNR Longhi, Sonia; CNRS and Universities Aix-Marseille I and II, AFMB, UMR 6098 Brunori, Maurizio; Università di Roma La Sapienza, Scienze Biochimiche Gianni, Stefano; Università di Roma, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli" and Istituto di Biologia e Patologia Molecolari del CNR

The folding mechanism of the SH3 domain from Grb2

Francesca Troilo[1], Daniela Bonetti[1], Carlo Camilloni[2], Angelo Toto[1], Sonia Longhi[3], Maurizio Brunori[1,*], and Stefano Gianni[1,*]

[1]Istituto Pasteur - Fondazione Cenci Bolognetti, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli" and Istituto di Biologia e Patologia Molecolari del CNR, Sapienza Università di Roma, 00185, Rome, Italy

[2] Dipartimento di Bioscienze, Università degli studi di Milano, 20133, Milan, Italy

[3] Aix-Marseille Univ, CNRS, Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, Marseille, France

*Corresponding authors: stefano.gianni@uniroma1.it;

maurizio.brunori@uniroma1.it

ABSTRACT

SH3 domains are small protein modules involved in the regulation of important cellular pathways. These domains mediate protein-protein interactions recognizing motifs rich in proline on the target protein. The SH3 domain from Grb2 (Grb2-SH3) presents the typical structure of an SH3 domain composed of two-three stranded antiparallel β–sheets orthogonally packed onto each other, to form a single hydrophobic core. Grb2 interacts, via SH3 domain, with Gab2, a scaffolding disordered protein, triggering some key metabolic pathways involved in cell death and differentiation. In this work we report a mutational analysis (Φ-value analysis) of the folding pathway of Grb2-SH3 that, coupled with molecular dynamic simulations, allows us to asses the structure of the transition state and the mechanism of folding of this domain. Data suggest that Grb2-SH3 folds via a native-like, diffused transition state with a concurrent formation of native-like secondary and tertiary structure (nucleation-condensation mechanism) and without the accumulation of folding intermediates. The comparison between our data and previous folding studies on SH3 domains belonging to other proteins, highlights that proteins of this class may fold via alternative pathways, stabilized by different nuclei leading or not to accumulation of folding intermediates. This comparative analysis suggests that the alternative folding pathways for this class of SH3 domains can be selectively regulated by the specific aminoacid sequences.

INTRODUCTION

One of the most informative approaches to address the folding mechanism of globular proteins is to compare experiments performed on homologous proteins [1-6]. In fact, by describing the folding of proteins sharing the same topology while displaying a different sequence, it is theoretically possible to draw some general rules on the basic principles governing folding. Comparative folding studies have been previously reported for example on the colicin immunity proteins Im7 and Im9 [2,4,7,8], on the immunoglobulin domains [3], on c-type cytochromes [9-11], on homeodomain-like proteins [5,12], on PDZ domains [1,13-15] and others. Whilst all these studies suggest that the overall general features of folding are by-and-large defined by protein topology, it appears that a closer look at the folding pathway of the different homologues appears to highlight some features specific for each globular protein.

In the context of comparative folding studies, the SH3 domain represents a debated system. In fact, whilst earlier comparison between the src and the spectrin SH3 domains suggested this class of proteins to fold via a robust two-state mechanism characterized by a polarized and highly conserved transition state [16-18], studies on the Sso7d domain revealed an additional complexity [19]. Indeed, while displaying less than 10% sequence identity, the Sso7d protein shares the typical topology of SH3 domains, except for the last β-strand that is a small α-helix in Sso7d. Interestingly, experimental and computational comparison of the folding of Sso7d with SH3 domains, revealed a substantial shift in the folding nucleus from the third to the second β-hairpin [19]. This finding highlighted that, even if protein topology plays a

major role in the selection of the folding pathways, the specific nature of the interactions stabilizing the protein is still critical to describe folding mechanisms. Furthermore, it is of interest to note that recent studies highlighted how the folding of SH3 may also occur via a multi-state scenario, with accumulation of intermediates characterized at equilibrium [20-22]. It appears therefore that even for a deeply investigated protein system, such as the SH3 domain, folding demands a careful study to be fully understood.

The SH3 domain from Grb2 (Grb2-SH3) corresponds to the typical structure of an SH3 domain composed of two three-stranded antiparallel β–sheets orthogonally packed onto each other, to form a single hydrophobic core [23,24]. Physiologically, the domain is involved in binding a proline rich stretch of amino acids of Gab2 (encompassing residues 503 to 524), with this interaction triggering some key metabolic pathways involved in cell death and differentiation. From the perspective of its primary structure, it is interesting to note that Grb2-SH3 displays a similar degree of sequence identity towards Sso7d and spectrin SH3 (Figure 1), posing this system as an interesting candidate to understand further the folding mechanism of this highly studied class of proteins.

Here we present the characterization of the folding of the Grb2-SH3 domain. By carrying out kinetic experiments on 23 site-directed variants in combination with restrained molecular dynamics simulations, we present the structure of the main folding transition state. The transition state is stabilized by contacts involving both the first β-hairpin and the N- and C-termini of the protein, a finding which appears different from what previously observed for src, spectrin and fyn SH3 [16,18,25,26].

Furthermore, we present evidence that this protein folds via a nucleation-condensation mechanism, with a diffused, rather than structurally polarized transition state. The data are discussed in the context of previous work on other SH3 domains.

EXPERIMENTAL AND THEORETICAL METHODS

Site-Directed Mutagenesis

C-SH3 domain of Grb2 was subcloned in a pET28b+ plasmid vector. The constructs encoding the site directed variants of SH3 were obtained using the gene encoding Grb2-SH3 *wt* as a template to perform site-directed mutagenesis using the QuickChange Lightning Site-Directed Mutagenesis kit (Agilent technologies) according to the manufactorer's instructions. All mutations are conservative All mutations were confirmed by DNA sequencing.

Protein expression and purification

The C-SH3 domain of Grb2 *wt* and all the site directed variants were expressed in *E. coli* cells BL21 (DE3). Bacterial cells were grown in LB medium, containing 30 μg/ml of kanamycin, at 37°C until OD_{600} = 0.7 - 0.8 and then protein expression was induced with 1mM IPTG. After induction cells were grown at 37°C over night and then collected by centrifugation.

To purify the protein, the bacterial pellet was resuspended in buffer 50 mM TrisHCl, 0.5 M NaCl, pH 7.5 with the addition of antiprotease tablet (Complete EDTA-free, Roche), then sonicated and centrifuged. The soluble fraction from bacterial lysate was loaded onto a nickel-charged HisTrap Chelating HP (GE Healthcare) column

equilibrated with 50 mM TrisHCl, 0.5 M NaCl, pH 7.5. The protein was then eluted with a gradient from 0 to 1 M imidazole by using an AKTA-prime system. Fractions containing the protein were collected and the buffer was exchanged to 25 mM Hepes pH 7.5 100 mM potassium acetate by using a HiTrap Desalting column (GE Healthcare). The purity of the protein was analyzed through SDS-page.

Protein concentration was estimated by measuring the absorbance of tryptophan residue at 280nm and calculated through the Lambert-Beer equation.

Equilibrium experiments

Equilibrium unfolding experiments were performed on a Fluoromax single photon counting spectrofluorometer (Jobin-Yvon, NJ, USA). C-SH3 protein and all the site directed variants, at a constant concentration of 3 μM, was excited at 280 nm and emission spectra were recorded between 300 and 400 nm, at increasing denaturant (urea) concentration. Experiments were performed at 25°C, using a quartz cuvette with a path length of 1 cm, in buffer 50 mM sodium phosphate buffer at pH 7.2.

Stopped-flow folding experiments

Unfolding and refolding kinetics experiments were carried out on a single-mixing SX-18 stopped-flow instrument (Applied Photophysics), monitoring the change of fluorescence emission. The experiments were performed at 25°C in buffer 50 mM sodium phosphate pH 7.2, by using urea as the denaturant. The excitation wavelength used was 280 nm and the fluorescence emission light was recorded by using a 320 nm cut-off glass filter. For each denaturant concentration usually 5 individual traces were averaged. The final concentration of Grb2-SH3 and all the variants was typically 1 μM. In all cases the fluorescence time courses obtained was satisfactorily fitted by using a single exponential equation.

Molecular Dynamics Simulations

Molecular dynamics simulations of SH3 were performed using the CHARMM22* force field [27] with the TIP3P water model [28]. All the simulations were run using GROMACS [29] and PLUMED2 [30]. A time step of 2 fs was used together with LINCS constraints [31]. Van der Waals and Coulomb interactions were implemented with a cut-off at 0.9 nm, and long-range electrostatic effects were treated with the particle mesh Ewald method on a grid with a mesh of 0.1 nm.
A standard 200 ns molecular dynamics simulation at 300 K was performed as a reference for the native state ensemble. The starting conformation was taken from an available X-Ray structure (PDB code 2VWF [23]) and solvated with 4531 water molecules and 4 sodium ions.

The transition state ensemble was determined following a standard procedure based on the interpretation of Φ value analysis in terms of fraction of native contacts. Briefly, given a set of experimental Φ values, a pseudo energy term has been added to the force field as the squared difference between experimental and simulated Φ values in order to maximize the agreement with the experimental value while keeping the simulation stable. Given two residues that are not nearest neighbours, the native contacts between them are defined as the number of heavy side-chain atoms located within 0.65 nm in the native structure. The Φ value for a residue *i* is calculated from the fraction of native contacts that it makes in a given conformation. With this approach only Φ values between 0 and 1 can be incorporated as structural restraints.

The transition state ensemble was generated using 1000 cycles of simulated annealing. Each cycle is 200 ps long, in which the temperature is varied between 300 K and 400 K. Only the structures sampled at the reference temperature are retained for further analysis, resulting in TSE of ~6000 conformations.

RESULTS

The kinetic folding mechanism of wild type Grb2-SH3

In order to characterize the folding mechanism of Grb2-SH3 we initially conducted experiments on the wild type protein. Urea-induced equilibrium denaturation of Grb2-SH3 measured at 25°C, pH 7.2 in 50 mM sodium phosphate buffer by decrease in Trp emission is reported in Figure 2. The observed transition is consistent with a simple two-state behavior, suggesting the absence of stable equilibrium intermediate(s) [32]. The unfolding free energy in water derived from two-state analysis is 3.1 kcal mol^{-1} displaying an $m_{D\text{-}N}$ value of 0.73 kcal mol^{-1} M^{-1}. This value, which is proportional to the change in accessible surface area upon unfolding, is consistent with what expected from a protein of 56 amino acids [33].

The folding and unfolding kinetics of Grb2-SH3 were measured by stopped-flow fluorimetry. As expected for a two-state folder, under all investigated conditions, folding and unfolding time courses were consistent with a single exponential decay. Furthermore, in analogy to what previously observed on other SH3 domains, the urea dependence of the observed rate constant (k_{obs}) on urea concentration conforms to a V-shaped chevron plot (Figure 3), a typical signature of two state folding [32].

Since Grb2-SH3 unfolding displays a low cooperativity, with an $m_{D\text{-}N}$ of 0.73 kcal kcal mol^{-1} M^{-1} an accurate determination of the folding parameters from each independent experiment is complicated. Therefore, to decrease the fitting error and, at the same time, to test the robustness of two-state folding of Grb2-SH3, equilibrium

and kinetic experiments were fitted globally to the following equations:

Equilibrium: $Y_{obs} = Y_N + Y_D \frac{e^{(m_{D-N}([urea]-[urea]_{1/2})}}{1+e^{(m_{D-N}([urea]-[urea]_{1/2})}}$

Kinetics: $k_{obs} = k_F e^{(-m_F[urea])} + k_U e^{(-m_U[urea])}$; $m_{D-N} = m_F + m_U$

with shared $m_{D\text{-}N}$ values. The fitting parameters calculated from the global analysis are reported in Table 1.

The structure of the folding transition state of Grb2-SH3

In order to characterize the transition state of folding of Grb2-SH3, we carried out a Φ value analysis [34,35], by producing 23 site directed variants. The Φ value is then calculated by dividing the effect of the substitution on the activation free energy by that of the stability of the native structure. The conservative variants were designed and the analysis carried out using the standard rules of Φ value analysis, as formalized previously [36].

Unfolding and folding of all the variants were measured both at equilibrium, by urea induced denaturation, and by kinetics, using the stopped-flow fluorimetry. In all cases, in analogy to what observed for wild type Grb2-SH3, folding and unfolding kinetics were consistent with a single exponential decay. Figure 4 shows the equilibrium and kinetic experiments carried out on each site directed variant. In all

cases, data were consistent with a two-state scenario, indicating that Grb2-SH3 folds via a robust mechanism, which does not involve any transient folding intermediates.

To determine the structure of the folding transition state of Grb2-SH3, we used the experimentally measured Φ values as restrains in molecular dynamics simulations. This method, which has been previously used and validated on several different protein systems [1,37-42], is based on the incorporation of the Φ values as biases on the fraction of formed native contacts in a molecular dynamics simulation trajectory (cf. Methods).

The structure of the folding transition state of Grb2-SH3, together with the associated contact map, is reported in Figure 5A. It is evident that the protein seems to fold via a native-like transition state that is characterized by the formation of the first β-hairpin, together with a consolidation of the interaction between the N- and C-termini of the protein. Structure gradually tapers off, with the region encompassing the β2-β3 interaction being the most disordered of the ensemble. The structural features of the transition state of folding of Grb2-SH3, in comparison to those previously depicted for other SH3 domains are analysed in the discussion section.

On the basis of the Φ value analysis of src and spectrin SH3 [16,18], it has been previously suggested the structure of the transition state of SH3 domains to be highly polarized. To test this hypothesis for Grb2-SH3, we analysed the Bronsted plot of this protein [43]. In fact, whilst a diffused native like structure is expected to return linear Bronsted plots, a polarized transition state is more likely to yield a scatter in the Bronsted plot, with only some positions playing a key role in stabilizing its structure

(characterized by high Φ value), with the others displaying low values of Φ [44]. As evident from Figure 5B, the transition state of Grb2-SH3 clearly displays a linear Bronsted plot, suggesting this protein to fold via a native-like diffused, rather than polarized, transition state. This finding appears consistent with a nucleation-condensation mechanism [45,46] for this SH3, in agreement with what proposed earlier by Shakhnovich and co-workers [26].

Robustness of two-state folding in Grb2-SH3

It has been proposed that some SH3 domains may retain some residual structure in their denatured state [47,48] and/or populate folding intermediates [20-22]. In order to test the robustness of the two-state folding in Grb2-SH3, we resorted to analyse the dependence of the folding parameters as a function of protein stability. In fact, comparing the parameters measured on different site-directed variants represents an efficient test to address the overall folding characteristics of transition and denatured state [49,50]. More specifically, since the dependence of activation and ground states free energies on the denaturant concentration (measured by the m_U, m_F and m_{D-N} values) are dependent from the changes in accessible surface area between the pertinent state [33], an analysis of their dependence may be reveal signatures of shifts of the transition and denatured states along the reaction coordinate, as well as the accumulation of folding intermediates. Figure 6 depicts the correlation between the m_{D-N}, m_U, and m_F and the $\Delta\Delta G_{D-N}$ for the different site-directed variants. It is evident that, in the case of Grb2-SH3, no detectable change in m_U, m_F and m_{D-N} values could be observed for the different variants, spanning a change in protein stability of about 3 kcal mol^{-1}. This observation suggests that, contrary to what observed in the case of fyn and PI3K SH3

domain [20-22], the folding mechanism of this protein is robust and consistent with two-state.

DISCUSSION

The first comparative Φ value analysis on globular proteins was presented in a two papers describing the folding of src and spectrin SH3 [16,17]. These studies suggested this protein family to fold via a conserved mechanism characterized by a structurally robust transition state. Furthermore, it was pointed out that the structure of the transition state was primarily stabilized by interactions taking place in the third β-hairpin of the protein, representing a polarized folding nucleus. Subsequently, also a Φ value analysis of fyn SH3 domain was reported, further supporting the robustness of the structure of the transition state [25] showing that even drastic non conservative mutations caused little structural rearrangements of the transition state [51]. A breakdown of such robustness could be observed in SSo7d, a protein sharing a similar topology with the other SH3 domains while displaying negligible sequence homology. In fact, in this case, a shift in the transition state nucleus from the third to the second β-hairpin was reported [19].

In the context of previous work on SH3 domains, it is therefore interesting to note how the structure of the transition state of folding of Grb2-SH3 is different from that of src, spectrin and fyn SH3. In fact, Grb2-SH3 displays an extended folding nucleus, which involves the β-sheet comprising the N- and C-termini of the protein together with the first β-hairpin. Since the structural architecture of the folding nucleus of

Grb2-SH3 appears to be distinct from that of Sso7d, it appears that this protein family may fold through a multitude of mechanisms comprising distinct regions the protein. Such pathways may then be selectively stabilized over others by the amino acid sequence, indicating that, whilst the overall features of folding are defined by protein topology, the nature of the interactions stabilising the native state are still critical to influence protein folding mechanisms, In this context, alternative pathways may emerge when the sequence is changed extensively.

A number of studies have shown that proteins may fold with or without folding intermediates, depending on solvent conditions and changes in sequence composition [2,7,8,12,15,38]. Accordingly, whilst the folding of SH3 domains has been classically descripted with a two-state mechanism, Dokholyan and co-workers predicted [52], by analysing different molecular dynamics simulations, that this class of protein may populate stable intermediates as a consequence of the local stabilization of individual structural elements. This finding was later supported by NMR and by pulse hydrogen exchange mass spectrometry, that revealed that presence of at least one folding intermediate in the case of Fyn [20,21] and PI3K SH3 [22] respectively. In both cases, the stabilization of the intermediate appears to arise from the stabilization of non-native hydrophobic interactions, leading to a polarized structure formation upon folding. The analysis of the Bronsted plot of Grb2-SH3 suggests this protein to fold via a transition state with diffused native-like structure. In this case, therefore, the protein seems consistent with a nucleation-condensation mechanism, characterized by a concurrent formation of native secondary and tertiary structure [45,46]. On the light of this finding, it is not surprising to observe that, contrary to the SH3 domains of Fyn and PI3K, Grb2-SH3 seems to conform to two state folding, even when challenged with

different site-directed variants, as illustrated by the robustness of the measured m_F, m_U and m_{D-N} values, which are essentially independent of protein stability.

Taken together, our analysis of the folding pathway of Grb2-SH3 supports a view whereby this class of proteins may fold via alternative pathways, stabilized by different nuclei, that can be selectively balanced by sequence composition. In agreement with previous finding on other protein systems, local stabilization of such alternative nuclei may lead to the accumulation of intermediates, switching two-state to multi-state folding.

ACKNOWLEDGEMENTS

Work partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (Progetto di Interesse 'Invecchiamento' to S.G.), Sapienza University of Rome (C26A155S48, B52F16003410005 and RP11715C34AEAC9B to S.G), the Associazione Italiana per la Ricerca sul Cancro (Individual Grant - MFAG 2016, 18701 to S.G.). FT is a recipient of a PhD fellowship from the Italo-French University.

REFERENCES

1 Calosci, N., Chi, C. N., Richter, B., Camilloni, C., Engstrom, A., Eklund, L., Travaglini-Allocatelli, C., Gianni, S., Vendruscolo, M. & Jemth, P. Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proc. Natl. Acad. Sci. U S A* **2008** 105, 19241-19246,

2 Capaldi, A. P., Shastry, M. C., Kleanthous, C., Roder, H. & Radford, S. E. Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat Struct Biol* **2001** 8, 68-72,

3 Clarke, J., Cota, E., Fowler, S. B. & Hamill, S. J. Folding studies of Ig-like beta-sandwich proteins suggest they share a common folding pathway. *Structure* **1999** 7, 1145-1153,

4 Friel, C. T., Capaldi, A. P. & Radford, S. E. Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. *J. Mol. Biol.* **2003**, 293-305,

5 Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W., DeMarco, M. L., Daggett, V. & Fersht, A. R. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. U S A* **2003** 100, 13286-13291,

6 Zarrine-Afsar, A., Larson, S. M. & Davidson, A. R. The family feud: do proteins with similar structures fold via the same pathway? *Curr Opin Struct Biol* **2005** 15, 42-49,

7 Capaldi, A. P., Kleanthous, C. & Radford, S. E. Im7 folding mechanism: misfolding on a path to the native state. *Nature Structural Biology* **2002** 9, 209-216,

8 Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. & Radford, S. E. Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* **1999** 286, 1597-1608,

9 Gianni, S., Travaglini_Allocatelli, C., Cutruzzola, F., Brunori, M., Shastry, M. C. & Roder, H. Parallel pathways in cytochrome c(551) folding. *J Mol Biol* **2003** 330, 1145-1152,

10 Travaglini-Allocatelli, C., Gianni, S. & Brunori, M. A common folding mechanism in the cytochrome c family. *Trends Biochem. Sci.* **2004** 29, 535-541,

11 Travaglini-Allocatelli, C., Gianni, S., Morea, V., Tramontano, A., Soulimane, T. & Brunori, M. Exploring the cytochrome c folding mechanism: cytochrome c552 from thermus thermophilus folds through an on-pathway intermediate. *J. Biol. Chem.* **2003** 278, 41136-41140,

12 White, G. W., Gianni, S., Grossmann, J. G., Jemth, P., Fersht, A. R. & Daggett, V. Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J. Mol. Biol.* **2005** 350, 757-775,

13 Chi, C. N., Gianni, S., Calosci, N., Travaglini-Allocatelli, C., Engstrom, Å. & Jemth, P. A conserved folding mechanism for PDZ domains. *FEBS Lett.* **2007**, Feb 15; [Epub ahead of print],

14 Gianni, S., Calosci, N., Aelen, J. M., Vuister, G. W., Brunori, M. & Travaglini-Allocatelli, C. Kinetic folding mechanism of PDZ2 from PTP-BL. *Prot. Eng. Des. Sel.* **2005** 18, 389-395,

15 Ivarsson, Y., Travaglini-Allocatelli, C., Jemth, P., Malatesta, F., Brunori, M. & Gianni, S. An on-pathway intermediate in the folding of a PDZ domain. *J. Biol. Chem.* **2007** 282, 8568-8572,

16 Grantcharova, V. P., Riddle, D. S., Santiago, J. V. & Baker, D. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol* **1998** 5, 714-720,

17 Martinez, J. C., Pisabarro, M. T. & Serrano, L. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat Struct Biol* **1998** 6, 721-729,

18 Martínez, J. C. & Serrano, L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* **1999** 6, 1010-1016,

19 Guerois, R. & Serrano, L. The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* **2000** 304, 967-982,

20 Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **2004** 430, 586-590,

21 Ollerenshaw, J. E., Kaya, H., Chan, H. S. & Kay, L. E. Sparsely populated folding intermediates of the Fyn SH3 domain: matching native-centric essential dynamics and experiment. *Proc. Natl. Acad. Sci. U S A* **2004** 101, 14748-14753,

22 Dasgupta, A. & Udgaonkar, J. B. Four-state folding of a SH3 domain: salt-induced modulation of the stabilities of the intermediates and native state. *Biochemistry* **2012** 51, 4723-4734,

23 Harkiolaki, M., Tsirka, T., Lewitzky, M., Simister, P. C., Joshi, D., Bird, L. E., Jones, E. Y., O'Reilly, N. & Feller, S. M. Distinct Binding Modes of Two Epitopes in Gab2 that Interact with the Sh3C Domain of Grb2. *Structure* **2009** 17, 809-822,

24 Toto, A., Bonetti, D., De Simone, A. & Gianni, S. Understanding the mechanism of binding between Gab2 and the C terminal SH3 domain from Grb2. *Oncotarget* **2017** 8, 82344-82351,

25 Northeyk, J. G., Di Nardo, A. A. & Davidson, A. R. Hydrophobic core packing in the SH3 domain folding transition state. *Nat Struct Biol* **2002** 9, 126-130,

26 Hubner, I. A., Edmonds, K. A. & Shakhnovich, E. I. Nucleation and the transition state of the SH3 domain. *J. Mol. Biol.* **2005** 349, 424-434,

27 Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? . *Biophys. J.* **2011** 100, 47-49,

28 Jorgensen, W. L. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. . *J. Am. Chem. Soc.* **1981** 103, 335-340,

29 Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. & Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. . *SoftwareX* **2015** 1-2, 19-25,

30 Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED2: New feathers for an old bird. *Comput. Phys. Commun.* **2014** 185, 604-613,

31 Hess, B. P-lincs: A parallel linear constraint solver for molecular simulation. *J. Chem. Theor. Inf.* **2008** 4, 116-122,

32 Jackson, S. E., Fersht, A.R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **1991** 30, 10428-10435,

33 Myers, J. K., Pace, C.N., Scholtz, J.M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995** 4, 2138-2148,

34 Fersht, A. R., Matouschek, A. & Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **1992** 224, 771-782,

35 Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **1989** 340, 122-126,

36 Fersht, A. R. & Sato, S. Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. U S A* **2004** 101, 7976-7981,

37 Geierhaas, C. D., Salvatella, X., Clarke, J. & Vendruscolo, M. Characterisation of transition state structures for protein folding using 'high', 'medium' and 'low' {Phi}-values. *Protein Eng. Des. Sel.* **2008** 21, 215-222,

38 Gianni, S., Ivarsson, Y., De Simone, A., Travaglini-Allocatelli, C., Brunori, M. & Vendruscolo, M. Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nat Struct Mol Biol* **2010** 17, 1431-1437,

39 Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **2001** 409, 641-645,

40 Camilloni, C., Bonetti, D., Morrone, A., Giri, R., Dobson, C. M., Brunori, M., Gianni, S. & Vendruscolo, M. Towards a structural biology of the hydrophobic effect in protein folding. *Sci. Rep.* **2016** 6, 28285,

41 Gianni, S., Camilloni, C., Giri, R., Toto, A., Bonetti, D., Morrone, A., Sormanni, P., Brunori, M. & Vendruscolo, M. Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. *Proc. Natl. Acad. Sci. U S A* **2014** 111, 14141-14146,

42 Gsponer, J., Hopearuoho, H., Whittaker, S. B., Spence, G. R., Moore, G. R., Paci, E., Radford, S. E. & Vendruscolo, M. Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl. Acad. Sci. U S A* **2006** 103, 99-104,

43 Leffler, J. E. Parameters for the description of transition states. *Science* **1953** 117, 340-341,

44 Fersht, A. R. Relationship of Leffler (Bronsted) alpha values and protein folding Phi values to position of transition-state structures on reaction coordinates. *Proc. Natl. Acad. Sci. U. S. A.* **2004** 101, 14338-14342,

45 Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* **1994** 33, 10026-10036,

46 Fersht, A. R. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U. S. A.* **1995** 21, 10869-10873,

47 Crowhurst, K. A., Tollinger, M. & Forman-Kay, J. D. Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain. *J. Mol. Biol.* **2002** 322, 163-178,

48 Kortemme, T., Kelly, M. J., Kay, L. E., Forman-Kay, J. & Serrano, L. Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol.* **2000** 297, 1217-1229,

49 Sanchez, I. E. & Kiefhaber, T. Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.* **2003** 327, 867-884.,
50 Scaloni, F., Gianni, S., Federici, L., Falini, B. & Brunori, M. Folding mechanism of the C-terminal domain of nucleophosmin: residual structure in the denatured state and its pathophysiological significance. *FASEB J.* **2009** 23, 2360-2365,
51 Northey, J. G., Maxwell, K. L. & Davidson, A. R. Protein folding kinetics beyond the phi value: using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state. *J. Mol. Biol.* **2002** 320, 389-402,
52 Borreguero, J. M., Ding, F., Buldyrev, S. V., Stanley, H. E. & Dokholyan, N. V. Multiple folding pathways of the SH3 domain. *Biophys. J.* **2004** 87,

FIGURE LEGENDS

Figure 1: Cartoon rapresentation of Grb2-SH3 (A), Spectrin-SH3 (B) and Sso7d-SH3 (C) structures and sequences alignments. As discussed in the text, Grb2-SH3 displays a comparable sequence identity to both Sso7d (18,8%) and spectrin SH3 (19.5%).

Figure 2: Equilibrium denaturation experiment of the Grb2-SH3 domain carried out in buffer 50 mM sodium phosphate pH 7.2 at 25°C. The change of the intrinsic fluorescence of the tryptophan residue versus urea concentrations was fitted with a two-state equation (see text for details).

Figure 3: Chevron plot of the Grb2-SH3 domain obtained in buffer 50 mM sodium phosphate pH 7.2 at 25°C.

Figure 4: Equilibrium denaturations and chevron plots of Grb2-SH3 and its site directed mutants. All experiments were carried out at 25 °C and pH 7.2 in 50 mM sodium phosphate buffer. Each mutant was globally fitted to a two state mechanism by assuming the m_{D-N} value at equilibrium to be equivalent to the sum between the kinetic m_F and m_U values. In all cases, data were consistent with a two-state scenario [32], indicating the absence of transient folding intermediates.

Figure 5: Structure of the folding transition state of Grb2-SH3, together with the associated contact map (Panel A). The top left of the contact map refers to the contacts between amino acids in the native state; whereas the bottom right to the

contacts in the transition state. As explained in the text, the protein seems to fold via a native-like transition state characterized by the formation of the first β-hairpin, together with a consolidation of the interaction between the N- and C-termini of the protein.

Panel B: Bronsted plot. As explained in the Results, the linearity of the Bronsted plot suggests that this protein fold via a native-like diffused, transition state. This finding appears consistent with a nucleation-condensation mechanism [45,46].

Figure 6: Correlation between the m_{D-N} (open circle), m_U (rhombus), and m_F (squares) and the $\Delta\Delta G_{D-N}$ for the different site-directed variants. As discussed in the text, no detectable dependence of m_U, m_F and m_{D-N} values can be observed for the different variants suggesting that the folding mechanism of this protein is robust and consistent with two-state.

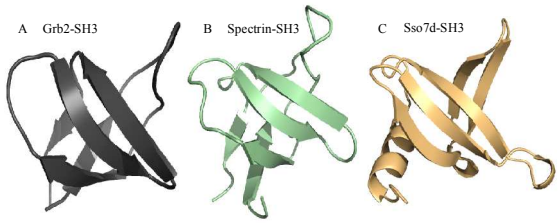
Table 1. Kinetic folding parameters of Grb2-SH3 and its site-directed variants.

MUTANT	k_F (s^{-1})	k_U (s^{-1})	m_F (kcal/M mol)	m_U (kcal/M mol)	m_{D-N} (kcal/M mol)	$[urea]_{1/2}$ (M)	**ϕ**
WT	16.0±1.4	0.14±0.02	0.59±0.02	0.14±0.06	0.73±0.01	4.2±1.0	
T1S	13.0±1.3	0.21±0.02	0.59±0.06	0.12±0.12	0.71±0.10	3.1±0.2	0.28±0.26
Y2A	8.6±0.9	0.69±0.07	0.59±0.10	0.07±0.10	0.73±0.04	1.8±0.1	0.26±0.07
V3A	4.6±0.6	0.33±0.05	0.75±0.07	0.08±0.09	0.83±0.06	2.5±0.3	0.58±0.11
A5G	1.7±0.7	0.36±0.04	0.63±0.05	0.12±0.08	0.73±0.06	0.2±0.1	0.70±0.17
L6A	8.0±0.6	0.67±0.05	0.68±0.04	0.13±0.06	0.81±0.04	2.3±1.1	0.30±0.06
F7A	6.4±0.64	0.21±0.02	0.66±0.07	0.11±0.10	0.77±0.07	3.2±0.1	0.68±0.15
F19A	2.6±0.3	1.23±0.05	0.57±0.09	0.07±0.13	0.73±0.09	1.6±0.2	0.45±0.04
F24A	3.7±0.3	0.70±0.03	0.70±0.04	0.12±0.06	0.81±0.04	1.8±0.2	0.47±0.05
I25V	6.1±0.5	0.64±0.03	0.70±0.04	0.08±0.06	0.78±0.04	2.4±1.7	0.38±0.06
H26A	33.0±8.0	0.23±0.05	0.62±0.05	0.15±0.06	0.77±0.03	3.6±1.5	*
S31A	11.0±1.0	0.42±0.04	0.67±0.03	0.08±0.04	0.75±0.03	2.8±0.4	0.24±0.10
A39G	16.0±2.0	0.92±0.07	0.71±0.04	0.06±0.06	0.77±0.04	2.4±1.3	-0.03±0.07
H41A	13.0±1.3	0.56±0.06	0.65±0.07	0.07±0.07	0.72±0.02	3.0±0.1	0.11±0.09
T44S	14.0±1.0	0.53±0.05	0.69±0.07	0.05±0.07	0.74±0.02	2.8±0.1	0.07±0.09
Y51A	11.0±0.5	0.43±0.02	0.57±0.02	0.14±0.02	0.71±0.01	2.6±0.4	0.23±0.07
T53S	11.0±3.0	0.23±0.05	0.53±0.06	0.12±0.09	0.65±0.07	2.7±0.3	0.41±0.37
A54G	16.2±1.6	0.62±0.06	0.55±0.03	0.12±0.04	0.67±0.02	2.4±1.1	-0.05±0.09

V55A	15.2±1.5	0.45±0.04	0.47±0.05	0.23±0.09	0.70±0.07	3.4±0.3	-0.01±0.11

Table 1: The mutants F9A, L17A, V27A, F47A and V52A expressed poorly and could not be characterized.

*This mutant shows $\Delta\Delta G_{D-N}$ < 0.4 kcal mol^{-1}, preventing reliable calculation of the Φ-value [36].




```
Grb2-SH3 !---GPTYVQALFDFDPQEDGELGFRRGDFIHVMDNSDPNWWKGACHGQTGMFPRNYVTAVNIQPPPVNRNLKPDRK !
Spectrin-SH3 !DFQGKELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGYVPAAYVKKLD---------- !
  Identity: 19.5% !

Grb2-SH3 !PTYVQALFDF-DPQEDGELG-------FRGDFI-HVNDNSDPNWWKGACHGQTGMFPRNYVTAVNIQPPPVNRNLKPDRK !
Sso7d-SH3 !----MATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGG---------GKTG---RGAVSEKD-APKELLQMLEKQKK !
  Identity: 18.8% !
```

Figure 1

!

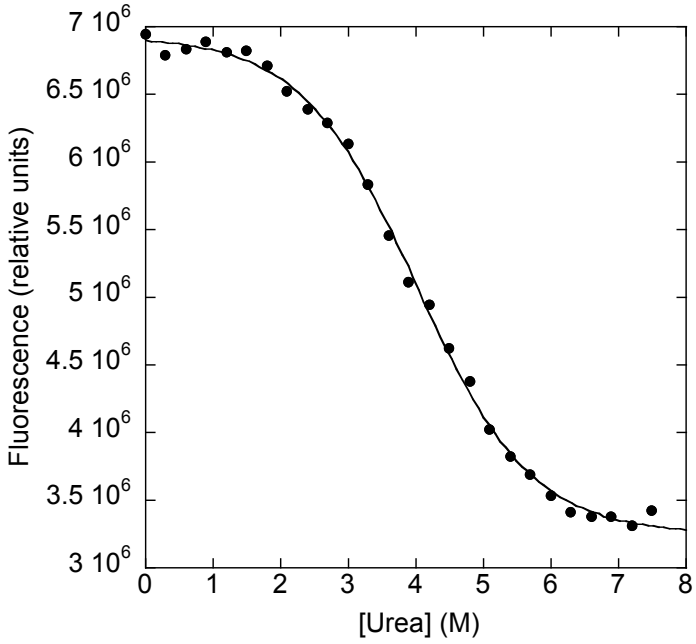

7 10^6
6.5 10^6
6 10^6
5.5 10^6
5 10^6
4.5 10^6
4 10^6
3.5 10^6
3 10^6
Fluorescence (relative units)
0
1
2
3
4
5
6
7
8
[Urea] (M)

Figure 2

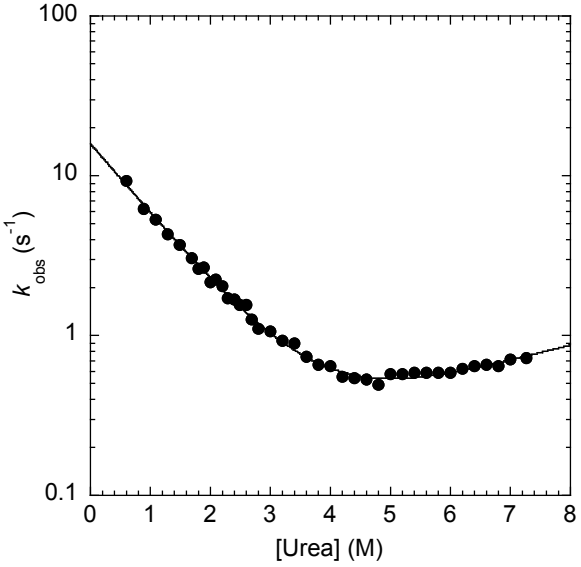


Figure 3

!

25

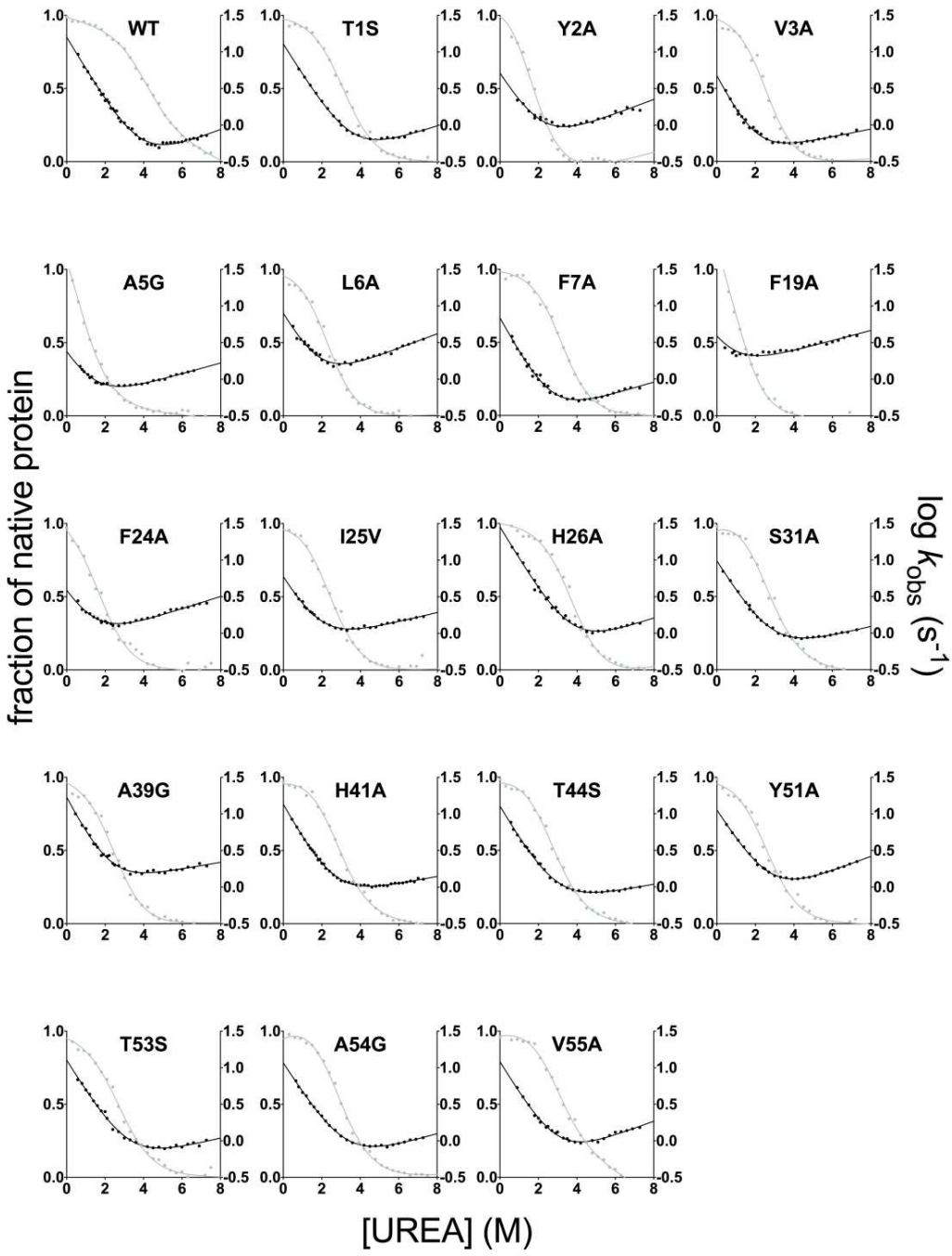

WT
T1S
Y2A
V3A
A5G
L6A
F7A
F19A
F24A
I25V
H26A
S31A
A39G
H41A
T44S
Y51A
T53S
A54G
V55A
fraction of native protein
log k_{obs} (s^{-1})
[UREA] (M)

Figure 4

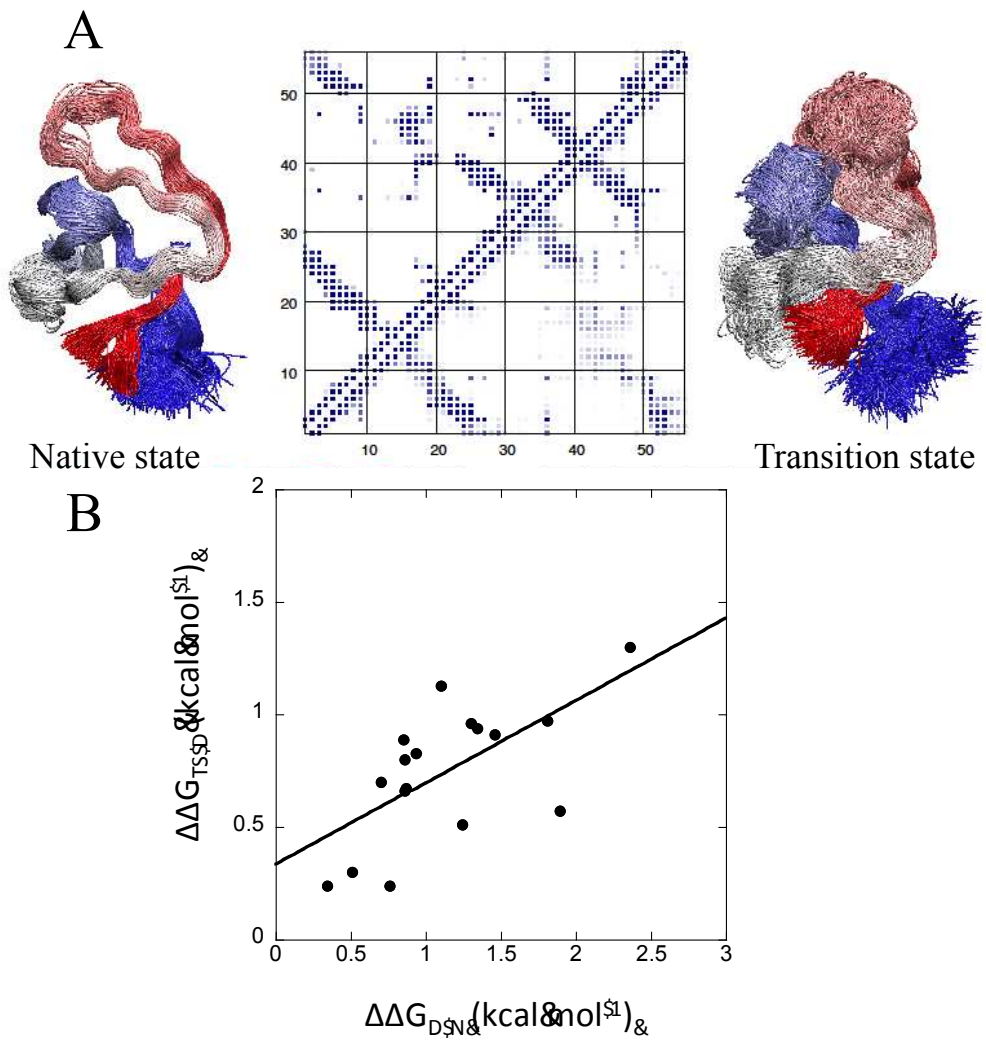

A
Native state
Transition state
B
ΔΔG_TS,D (kcal mol^-1)
ΔΔG_D,N (kcal mol^-1)

Figure 5

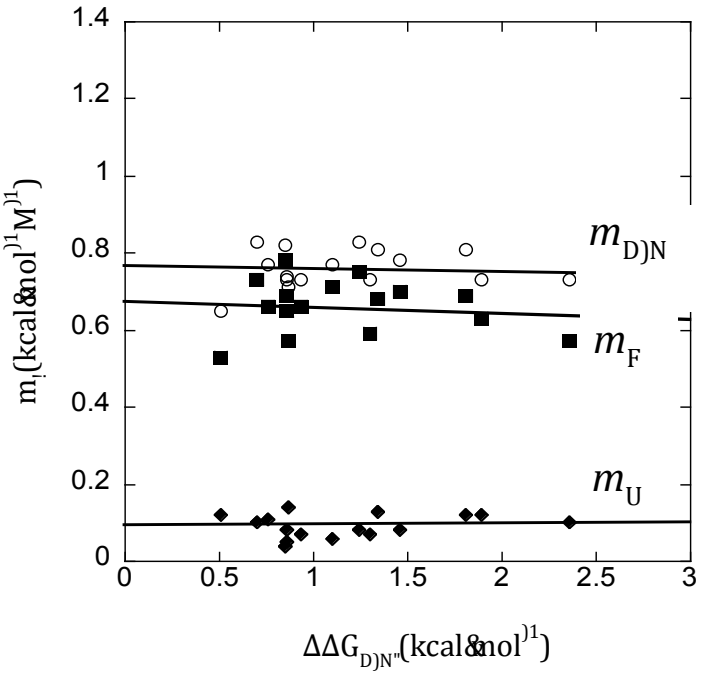

1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2
2.5
3
$m_{D)N}$
m_F
m_U
m_i(kcal&mol$^{)1}$M$^{)1}$)
$\Delta\Delta G_{D)N"}$(kcal&mol$^{)1}$)

Figure 6

TABLE OF CONTENT GRAPHICS