Orchestration of algal metabolism by protein disorder
Helene Launay, Véronique Receveur-Bréchot, Frédéric Carriere, Brigitte Gontero

To cite this version:
Helene Launay, Véronique Receveur-Bréchot, Frédéric Carriere, Brigitte Gontero. Orchestration of algal metabolism by protein disorder. Archives of Biochemistry and Biophysics, Elsevier, 2019, 672, pp.108070. 10.1016/j.abb.2019.108070. hal-02266032

HAL Id: hal-02266032
https://hal-amu.archives-ouvertes.fr/hal-02266032
Submitted on 6 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Orchestration of algal metabolism by protein disorder

Hélène Launay, Véronique Receveur-Bréchot, Frédéric Carrière, Brigitte Gontero*

Aix Marseille Univ, CNRS, BIP UMR 7281, 31 Chemin Joseph Aiguier, Marseille Cedex 20, 13402, France

*Correspondence to Brigitte Gontero

Dr Brigitte Gontero, Aix Marseille Univ, CNRS, BIP UMR 7281, 31 Chemin Joseph Aiguier, Marseille Cedex 20, 13402, France

Email: bmeunier@imm.cnrs.fr

Tel: + 33 4 91 16 45 49

Highlights (85 characters)

- IDPs help algae to cope with extreme conditions
- IDPs from algae are under-represented in disordered protein databases
- Only IDPs involved in CO₂ metabolism have been studied so far (EPYC1, CP12, RCA…)
- Disorder is involved in the regulation of enzymes belonging to CO₂ metabolism
- Disorder is also important in algal chaperones
Abstract

Intrinsically disordered proteins (IDPs) represent a family of proteins that provide many functional advantages in a large number of metabolic and signaling pathways. Because of their high flexibility that endows them with pressure-, heat- and acid-resistance, IDPs are valuable metabolic regulators that help algae to cope with extreme conditions of pH, temperature, pressure and light. They have, however, been overlooked in these organisms. In this review, we present some well-known algal IDPs, including the conditionally disordered CP12, a protein involved in the regulation of CO\textsubscript{2} assimilation, as probably the best known example, whose disorder content is strongly dependent on the redox conditions, and the essential pyrenoid component 1 that serves as a scaffold for ribulose-1, 5-bisphosphate carboxylase/oxygenase. We also describe how some enzymes are regulated by protein fragments, called intrinsically disordered regions (IDRs), such as ribulose-1, 5-bisphosphate carboxylase/oxygenase activase, the A\textsubscript{2}B\textsubscript{2} form of glyceraldehyde-3-phosphate dehydrogenase and the adenylate kinase. Several molecular chaperones, which are crucial for cell proteostasis, also display significant disorder propensities such as the algal heat shock proteins HSP33, HSP70 and HSP90. This review confirms the wide distribution of IDPs in algae but highlights that further studies are needed to uncover their full role in orchestrating algal metabolism.

**Keywords**: Calvin cycle, chaperone, CP12, EPYC, Intrinsically disordered protein, RuBisCO

**Abbreviations**

ADK: adenylate kinase; CBB: Calvin-Benson-Bassham; CCM: CO\textsubscript{2} concentrating mechanism; CD: circular dichroism; CP12: chloroplast protein of 12 kDa; EPYC1: Essential
Pyrenoid component 1; EST: expressed sequence tag; GAPDH: glyceraldehyde-3-phosphate dehydrogenase; HCA: hydrophobic cluster analysis; HSP: heat shock protein; IDP: intrinsically disordered protein; IDR: intrinsically disordered region; LSU: large subunit; NMR: nuclear magnetic resonance; PGA: phosphoglyceric acid; PRK: phosphoribulokinase; ROS: reactive oxygen species; PTM: post translational modification; RuBisCO: Ribulose-1, 5-bisphosphate carboxylase/oxygenase; RuBP: Ribulose-1, 5-bisphosphate; RCA: RuBisCO activase; SAXS: small angle X-ray scattering; SSU: small subunit
1. Introduction

Regulation of Algal Metabolism

Algae comprise a hugely diverse, polyphyletic group of photosynthetic organisms including prokaryotic cyanobacteria and eukaryotic algae from the Archaeplastida, Chromalveolata, Cryptophyta, Haptophyta and Dinoflagellata [1]. At least 72,500 algal species are known [2] but this is probably a large underestimate. Algae contribute about 46% of global primary production, mainly in the ocean [3] and also play an important role in global biogeochemical cycles such as oceanic carbon sequestration [4]. Algae are ubiquitous in aquatic and moist environments and some of them are ‘extremophiles’ tolerating, *inter alia*, desiccation [5], high or low temperatures, high or low pH, high salinity, high pressure [6] and high concentrations of heavy metals [7, 8]. Some survival strategies involve the production of novel biochemical compounds and there is a recent impetus towards the valorization of algae metabolic products, such as polyunsaturated fatty acids, fatty acid esters, polysaccharides, and the production of enzymes for various applications [9].

In eukaryotic algae, as in plants, chloroplasts have a central role in metabolism. They all derive from a cyanobacterium, but the number of membranes around a chloroplast varies depending on the number of endosymbioses involved in their evolutionary history [10]. Most of the chloroplastic enzymes involved in the metabolic primary steps are nuclear-encoded [11, 12] and their activity can be regulated at the transcription, translation, or protein level. Notably, at the protein level, post-translational modifications, protein-ligand or protein-protein interactions can regulate enzyme activity [13, 14].

Algae comprise a large number of algal species with high biochemical and structural diversity, are ecologically important, and can survive and grow in a large range of different
environments. To understand the basic principles of how algae regulate their metabolism, and eventually support future biotechnological applications, we need to decipher the mechanisms of these processes at the molecular level. Among regulatory proteins, a peculiar class of proteins has emerged, the intrinsically disordered/ductile proteins (IDPs), or the proteins that contain intrinsically disordered regions (IDRs). In this review, we will discuss their role in the regulation of algal metabolisms.

Reassessment of the protein structure function paradigm

IDPs escape from the classic structure-function paradigm and are fully functional despite being fully or partially disordered. The principal feature of these IDPs or IDRs containing proteins is the absence of a unique stable conformation which results in a high structural flexibility [15, 16]. Nevertheless, IDPs can also present either small structural motifs, or have a significant propensity to fold into secondary structural elements [17, 18]. This propensity to fold is modulated by several factors such as post-translational modifications [19], protein or ligand binding [20-22], physico-chemical conditions such as redox-conditions [23-27], pH [28], temperature [29], binding to a membrane, adsorption at an interface [30], etc. Some IDPs are also named conditionally disordered proteins [31] when their disorder content and their function is directly modulated by different possible conditions, as redox-conditions. A large collection of names were found in the literature that describes these proteins, and the community eventually decided to keep the term IDP [32]. The term will be used in the rest of this review.

The signatures of IDPs are rather well-established, with a bias in their amino-acid composition and frequent sequences of low complexity [33, 34]. Based on the sequence peculiarities of IDPs, a plethora of disorder predictors has been developed during the last two decades, with increasing prediction accuracy, and are continually updated [35-37]. These
predictors have allowed large scale analyses of an ever growing number of fully sequenced and annotated genomes in all kingdoms of Life, which revealed the widespread and universal nature of IDPs. Several databases of IDPs have thus been implemented, which provide precomputed disorder predictions of completely-sequenced genomes (D²P²; http://d2p2.pro/), of all UniProt entries (MobiDB; http://mobidb.bio.unipd.it/), or which reference only experimentally verified IDPs (DisProt, http://www.disprot.org/; IDEAL, http://idp1.force.cs.is.nagoya-u.ac.jp/IDEAL/).

A range of biophysical approaches are required to characterise experimentally the disordered properties of IDPs, often advantageously in association with state-of-the-art modelling tools to comprehend the large ensemble of probable conformational states that coexist in a given condition [16, 38-41]. IDPs have thus been characterised by small angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR) spectroscopy, circular dichroism (CD) spectroscopy, protease sensitivity and many other techniques, to cite but a few (see Longhi, this issue, [16, 41-43], with each method having advantages and limitations.

The functions carried out by IDPs are diverse and complement those of ordered proteins. They play significant roles in many biological processes, such as control of the cell cycle, transcriptional activation, and signalling, and they frequently interact with or function as central hubs in protein interaction networks [44-48]. Moreover, correlation between disorder and post-translational modifications (PTMs) such as phosphorylation, methylation, acylation, glycosylation and ubiquitination has been reported [19, 49]. This is worth noting since PTMs can contribute to the diversification and functionality of proteomes by regulating different properties of proteins (stability, conformation, etc…). Hence IDPs via diverse PTMs participate in the regulation processes of the cell [19, 46, 50, 51].

**IDP phylogenetic distribution**
Analysis of the proteomes using predictors of disorder revealed that the proportion of intrinsic disorder in proteomes is variable in the evolution tree. There is a low proportion of disorder in bacteria and archaea, a higher one in viruses, and a much larger one in eukaryotes, especially in higher vertebrates, including mammals [52-54]. Recently, it was shown that IDPs play a role in the evolution of multi-cellularity and cell type specification [48, 55].

By contrast, the distribution of IDPs among the different kingdoms of Life, that have been experimentally investigated, does not correlate with the distribution of predicted IDPs. Notably, in the experimentally-verified database DisProt [56], IDPs from photosynthetic organisms, and especially from algae, are under-represented while IDPs from bacteria are over-represented, with 19% of all IDPs. In the other databases, this number ranges from 0 to 8.1% [56]. This discrepancy in the IDP databases might reflect, in a predominant manner, a biased interest in scientific research for pathogenic organisms that overlooks the wide distribution of disorder in the tree of Life. Another bias might also result from the fact that large scale projects on structural genomics have used *E. coli* as the main expression system for high throughput production and screening of proteins, which naturally limits the efficient and correct production of some eukaryotic proteins.

**IDP in plants and algae**

A specific feature of IDPs is that they are able to maintain their functional states in a wide range of physiological conditions, unlike well-folded proteins which can be extremely sensitive to variation in their environment. Indeed, IDPs do not aggregate under high temperature or low pH, conditions in which structured proteins usually undergo denaturation and loss-of-function [29, 57, 58]. As mentioned above, algae encounter extreme conditions. Indeed, some algae can sink down to 4000 m in the deep sea and survive despite the high hydrostatic pressure [6]. Similarly to plant cells, algae can maintain their cellular volume by
imposing a turgor pressure inside their cell wall [5, 59]. Variation of pressure can be also
detrimental for the conservation of functional cavities in folded enzymes and of
macromolecular scaffolds [60]. For instance, it has been shown that the activity of Calvin-
Benson-Bassham (CBB) cycle enzymes is strongly affected by high pressure [61]. IDPs,
which are devoid of internal cavities, are expected to be more stable upon pressure increase,
and indeed, α-synuclein [62] and the disordered region of titin [63] preserve their disordered
characteristics at high pressure. It is likely that IDPs in general are more stable in these
conditions but this needs further investigations.

Besides, the chloroplast proteome undergoes variations of physico-chemical conditions
with redox switches [64]. The dark-light regulation of the photosynthetic pathway in the
chloroplast is mainly modulated via redox transitions [65-67]. Thanks to their propensity to
receive post-translation modifications that modify their conformational and functional states,
IDPs are particularly suited to sense these changes of physico-chemical conditions, thereby
regulating enzymes activity [68, 69]. For all these reasons, the properties of IDPs are valuable
assets for the regulation of enzymes and metabolic pathways, in the broad range of
environmental conditions that algae withstand.

Up to now, only few IDPs from photosynthetic organisms have been investigated, and
they mainly come from the higher plant Arabidopsis thaliana, which has one of the most
accurately annotated sequenced genome thanks to its small size. A recent analysis of 12 plant
genomes revealed that the occurrence of disorder in plants is similar to that in many other
eukaryotes [70, 71] though only few plant IDPs have been experimentally characterised (A.
Covarrubias, this issue). These IDPs or IDR containing proteins are, in the vast majority,
nuclear-encoded proteins, while chloroplast- and mitochondria-encoded proteins have a lower
disorder trend as in Archae and Bacteria, which is in agreement with their phylogenetic
origin.
In the model alga *Chlamydomonas reinhardtii*, a systematic search for disordered protein experimentally showed that this alga contains a large number of non-characterised proteins that have IDP-like properties [72]. Interestingly, the few IDPs that have been biophysically investigated so far are the ones involved in the regulation of their metabolic pathways. This bias arises from the keen interest in the metabolism of algae, which have a high potential in CO$_2$ assimilation and production of various added-value molecules.

In this review we shine a spotlight on the few known IDPs from algae that help these photosynthetic organisms to adjust their photosynthetic carbon metabolism to the prevailing conditions in their habitat. These IDPs are found in the chloroplast, where the primary step of photosynthesis occurs, and where CO$_2$ is assimilated and converted into building blocks to produce fatty acids, carbohydrates, proteins and various metabolites. Chaperones, which regulate the proteostasis, and therefore metabolism, through disordered regions, will also be discussed. Finally, we will extend the discussion to other algal predicted IDPs involved in metabolic regulation.
2. CO₂ acquisition and assimilation

During photosynthetic CO₂ assimilation, chemical energy, conveyed mainly by ATP and NAD(P)H, and supplied by electron transfer reactions through the photosystems I and II, is used by metabolic reactions of the CBB cycle. The efficiency of this assimilation relies on a fine tuning between energy demand and supply. Hence the CBB is highly regulated. This cycle functions in the light and is inactive in the dark. Some enzymes are active only under reducing conditions, corresponding to the light conditions, and are inactive under oxidising conditions corresponding to the dark [13, 65].

**RuBisCO activase (RCA)**

Ribulose-1, 5-bisphosphate carboxylase/oxygenase (RuBisCO) is the enzyme that initiates the CBB cycle by converting ribulose-1, 5-bisphosphate (RuBP) and carbon dioxide (CO₂) into two molecules of phosphoglyceric acid (PGA). In addition, RuBisCO catalyses a competitive reaction that converts RuBP in the presence of O₂, into one PGA and one toxic compound, phosphoglycolate [73]. RuBisCO is made up of eight large (LSU, RbcL) and eight small (SSU, RbcS) subunits, in algae and higher plants [74]. In order for RuBisCO to be active, it has long been suggested that a specific lysine residue on the LSU needs to be carbamylated [75]. Today, it is known that the mechanism for regulating RuBisCO is much more complex and involves not only changes in the carbamylation state of this lysine residue, but also conformational changes in the protein. These conformational changes are triggered by a protein called RuBisCO activase (RCA), which is a specific molecular chaperone and an AAA⁺ ATPase [66]. RuBisCO active sites can be locked by sugar phosphate in the close inactive state and RCA prevents this inhibition by promoting ATP-dependent conformational changes that open closed sites, thus facilitating dissociation of inhibitory sugar phosphate.
(Figure 1). Furthermore, the presence of RCA allows RuBisCO to be active even at suboptimal CO$_2$ concentrations that would not normally permit carboxylation \textit{in vivo} \cite{76, 77}.

Some algae in the charophyta phylum have multiple isoforms of RCA: two $\alpha$-isoform of $\sim$45-48 and 58 kDa, and a $\beta$-isoform of $\sim$41-43 kDa \cite{78} (Figure 2B). These isoforms are the product of alternative splicing and differ from each other by a larger and redox-regulated C-terminal extension on the $\alpha$-isoforms. This C-terminal extension contains three or two regulatory cysteine residues that allow the protein to be activated by light through the action of a small redox protein, the thioredoxin $f$ \cite{78, 79}. Interestingly, RCAs from another phylum of microalgae, the chlorophyta (including \textit{C. reinhardtii}), lack these two cysteine residues and therefore are unlikely to be redox-sensitive.

An \textit{in silico} analysis \cite{80} showed that the C-terminal extension of RCA isoforms is intrinsically disordered (Figures 1B and 2A). Several small but significant differences were observed between the intrinsic disorder propensities in RCAs from two chlorophyta algae, \textit{Tetraselmis} sp. and \textit{Ostreococcus tauri}, probably reflecting a different mode of interaction between their RCAs and RuBisCO \cite{81}. The authors suggest that the difference in the disordered propensities can be ascribed to different evolution pressures on these two algae related to their motility. Indeed, \textit{Tetraselmis} sp. is motile and can actively migrate to optimal habitat, in contrast to \textit{O. tauri}, which is a non-motile microalga. For the latter, therefore, there is a higher evolution pressure on RCA to optimise photosynthesis under non-optimal conditions. However, one cannot exclude that other processes are involved. In other photosynthetic organisms such as in $\beta$ cyanobacteria, the RCA lacks the N terminal domain necessary for RuBisCO activation found in plants and green algae \cite{66, 78, 82}. In red lineage organisms, there is no RCA gene but it is substituted by another protein, called CbbX. In the purple bacterium \textit{Rhodobacter sphaeroides}, CbbX possesses a short disordered C-terminal
extension (Figures 1C and 2A), the function of this part though, remains unelucidated [83].

The homologues of CbbX from Rhodophyta lack this C-terminal extension (Figure 2C).

**Essential Pyrenoid component 1 (EPYC1)**

In many algae, RuBisCO is present within a chloroplastic membraneless compartment that concentrates locally proteins involved in carbon fixation, and called a pyrenoid. Pyrenoids are involved in CO$_2$ concentrating mechanisms (CCMs) that are essential for aquatic organisms, which face low CO$_2$ concentrations much below the optimal concentration for RuBisCO activity [84, 85]. These CCMs increase CO$_2$ concentration around RuBisCO and thereby decrease O$_2$ competition. The pyrenoid is located in the chloroplast, and is surrounded by a starch sheath and traversed by membrane tubules that are continuous with the thylakoid membranes [86]. Recently, it was shown that in the model alga *C. reinhardtii*, RuBisCO accumulation and localization within the pyrenoid is mediated by a disordered repeat nuclear-encoded protein, called Essential Pyrenoid component 1 (EPYC1), previously identified as a low-CO$_2$ inducible protein (LC15; Cre10.g436550) [87]. This protein is ubiquitous in pyrenoid-containing algae but is absent in pyrenoid-less algae. In addition to EPYC1, RuBisCO and its chaperone RCA, 135 to 190 putative CCMs proteins including some of their binding partner(s) were also found to be located in the membrane-less pyrenoid, by mass spectrometry analysis [88, 89]. Like many IDPs, EPYC1 is phosphorylated at low CO$_2$ concentration [90]. Furthermore EPYC1 was shown to interact with a predicted serine/threonine protein kinase (KIN4-2; Cre03.g202000) and two 14-3-3 kinases FTT1 and FTT2 [87]. These interactions with kinases probably regulate RuBisCO and EPYC1 interactions and activities.

EPYC1 protein displays characteristic features of IDPs. Its CD profile is typical of an IDP (Figure 3A), and it is resistant to denaturing conditions such as boiling for 15 min [91]. Its sequence contains four 60 amino-acids repeats, each domain being predicted to be
disordered (Figure 3C). The role for these tandem domain repetitions remains controversial. It was hypothesised that these low-complexity repeats allow EPYC1 to bind several RuBisCO (Figure 3B), which then drives formation of pyrenoid droplets with liquid-like properties [88]. However, it was later shown that a mutant of EPYC1 with a single repeat was still able to form membrane-less liquid-liquid phase separation droplets upon binding to RuBisCO in-vitro, suggesting that each repeat domain contains several binding sites to RuBisCO. In-vitro reconstitution of liquid-liquid phase separation droplets indicated that EPYC1 and active RuBisCO are necessary and sufficient for phase separation, and that the phase separation is governed by electrostatic interactions [91]. In line with this finding, it is known that proteins with IDRs and low complexity sequences tend to phase separate and to form membrane-less organelles [92]. The molecular nature of such membrane-less organelles is under intense scrutiny [93]. In the current model, the pyrenoid matrix is highly dynamic, in a liquid-phase that supports efficient molecular diffusion to increase CO$_2$ concentration and decrease the oxygenase activity of RuBisCO.

**CP12 and the regulation of the CBB cycle**

CP12 is a nuclear-encoded chloroplast protein of about 8.5 kDa that is present in many photosynthetic organisms, including cyanobacteria, algae, and plants [94, 95] as well as in cyanophages [96]. It has been especially well characterised in some model organisms such as the higher plant *A. thaliana*, the cyanobacterium *Synechococcus elongatus* and the green alga *C. reinhardtii* [97]. This small protein is a conditionally disordered protein that has been characterised by NMR, SAXS, CD, Förster resonance energy transfer, fluorescence correlation spectroscopy, mass spectrometry and molecular modelling [26, 27, 98-101]. CP12 is involved in the regulation of the CBB cycle but has multiple other faces, such as being a chaperone. CP12 from *C. reinhardtii* has two disulfide bridges, one at the N-terminus and one
at the C-terminus. While the C-terminal cysteine residues are conserved in all CP12 proteins, the N-terminal ones may be missing in some photosynthetic organisms, e.g. in rhodophytes, the glaucophyte *Cyanophora* and in *Synechococcus* [94, 102, 103]. The formation of the two disulfide bonds under dark or oxidizing conditions allows the algal CP12 to bind two enzymes from the CBB cycle, namely glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and phosphoribulokinase (PRK). The formation of this supramolecular ternary edifice results in the inactivation of these two enzymes and consequently leads to the arrest of the CBB cycle in the dark. In the light, CP12 is reduced and the breakage of its two disulfide bridges leads to the dissociation of the ternary complex, with the resulting dissociated forms of PRK and GAPDH being fully active. CP12 therefore participates in the regulation and in the “work on demand” of the CBB cycle that is active under light and inactive under dark.

While in the oxidised state, CP12 has some α-helices, in the reduced state it becomes fully disordered (Figure 4A and 4B) [27, 99, 104, 105]. However, in both reduced and oxidised states, CP12 is highly mobile. NMR, CD and SAXS showed that the algal oxidised CP12 is composed of three regions with distinct conformational sampling [27]: its C-terminus folds in a stable turn, ready for GAPDH binding, it is linked to the N-terminal domain through a randomly disordered linker, while the N-terminus oscillates on the millisecond timescale between helical and random states (Figure 4C). The existence of a conformational equilibrium between two states of the algal CP12 was also shown previously by ion mobility mass spectrometry [98]. The combined analysis of NMR and SAXS data enabled the quantification of the relative population of each state (40% random-coil state and 60% helical). Upon binding to GAPDH, this equilibrium shifts towards more extended conformations (Figure 4D). In other words CP12 undergoes an induced unfolding. This phenomenon, called cryptic disorder, which contributes to decreasing the entropy cost of binding, might explain the very high affinity of CP12 for its partners, with a dissociation constant in the nM range [106]. This
GAPDH/CP12 sub-complex in turn enables the binding of PRK [107]. The idea that two proteins could form a high-affinity complex while remaining highly dynamic has long been “nearly heretical”. However, a recent study has brought to light the case of two fully disordered proteins forming a dynamic complex of very high affinity (pM range), without any gain of structure in the two proteins [108]. This last example shows that disordered proteins might interact strongly without anchor sites, nor disorder-to-order transitions, opening a new concept and avenue in the field of IDPs.

CP12 can bind metal ions, such as Cu$^{2+}$, Ni$^{2+}$ [109] and Ca$^{2+}$ [110], as well as enzymes such as the FBP aldolase [111]. Of interest also, CP12 can play the role of a specific chaperone for GAPDH preventing its aggregation and inactivation by heat treatment [112]. Hence, CP12 is a versatile protein considered as a jack-of-all-trades but master of the CBB cycle [113], that indirectly regulates the CO$_2$ assimilation by interacting with PRK and GAPDH. This regulation by CP12 is conserved in different species, but the molecular mechanism of the interactions and the strength of binding between the PRK/GAPDH/CP12 complex differ among species. For example, in *A. thaliana* the dissociation constant for GAPDH/CP12 is in the µM range [114], and CP12 undergoes an induced folding upon binding to GAPDH, in contrast to *C. reinhardtii* [27]. The broad adaptation to multiple environmental conditions might explain these divergent molecular mechanisms.

**GAPDH-B and ADK3, two enzymes with C-terminal extensions homologous to CP12**

There are two main forms of chloroplast GAPDH. The first one, GapA, is present in the majority of photosynthetic organisms [102, 115] and forms A$_4$ homotetramers. The second isoform, GapB, is found in higher plants and in the alga *Ostreococcus*. It is extremely similar to GapA, but has a C-terminal extension (about 30 amino acids) homologous to CP12 [116, 117]. The GapB subunit forms functional A$_2$B$_2$ heterotetramers with the GapA subunit, as
well as inactive A₈B₈ oligomers, although the relevance and the role of these complexes remain disputed. However, it is well established that the disordered "tail" or the C-terminal extension on GapB endows GAPDH with autonomous redox regulation [115].

In *C. reinhardtii*, where only GapA exists, this autonomous regulation is impossible. However, as the first steps of glycolysis and the oxidative pathway of phosphate pentoses take place in the chloroplast [118], there is an obligatory need for a fine regulation of the activity of this A₄ isoform. It has been shown that the redox regulation of this chloroplast A₄ GAPDH, depends on its interaction with CP12. Therefore, the regulation of GAPDH relies, as for RuBisCO, on IDR{s} either on the enzyme itself or on its partner, CP12. Since GAPDH is a moonlighting protein, with many cellular functions [119], fusion with this CP12-like tail may endows it with further other regulatory functions yet unknown and to be explored.

Other important enzymes are the adenylate kinases (ADK{s}). These enzymes maintain the energetic balance in cellular compartments including chloroplasts and mitochondria, by catalyzing the interconversion of AMP, ATP, and ADP (AMP+ATP→2ADP). In *C. reinhardtii*, ten genes encoding ADK{s} are present. Within the chloroplast stroma of photosynthetic organisms, where the ATP:ADP ratio has to be kept at an optimal level to allow a sustainable rate of CO₂ fixation, ADK₃ plays a major role in the regulation of concentrations of adenylates [120].

ADK₃ from *C. reinhardtii* comprises two domains: a canonical domain, homologous to the *Escherichia coli* ADK (44% identity), and a "tail" of 25 residues, very close to the C-terminal portion of CP12 (68% identity) [121]. The kinetic parameters for the wild-type and truncated ADK₃ (removing the C terminus, namely ADK₃-ΔCP12) are very similar, indicating that the C-terminal extension of ADK₃ does not affect the activity and kinetic properties of the enzyme, nor its global secondary structure. Since CP12-like extension
confers GAPDH redox regulation, it was hypothesised that it might be the same with ADK3. This redox regulation was not observed, but the CP12-like extension confers ADK3 with a high affinity for the chloroplast GAPDH [122]. Via this disordered region, ADK3 down-regulates the NADPH-dependent GAPDH activity, therefore stopping the consumption of NADPH by the CBB cycle. The C-terminal extension on ADK is also found in the sequence of ADK from other algae and might play the same role.

In addition, the C terminus of ADK3 has a strong effect on the thermal stability of the enzyme, and this effect is probably the result of its predicted disorder propensity, in agreement with IDPs heat-resistance [29, 57].

**Protection against oxidative damage**

The metabolism of microalgae is tightly redox regulated. On the other hand, the transient accumulation of reactive oxygen species (ROS) causes subsequent oxidative damage, protein unfolding and aggregation. Chloroplasts are organelles where molecular oxygen is produced during photosynthesis and where ROS production occurs, therefore protein-thiols protective mechanisms are active in these organelles. Under conditions causing oxidative stress, glutathionylation and nitrosylation play roles in enzyme regulation and signalling [123]. A proteomic study on *C. reinhardtii* has identified 225 proteins that are potential targets of glutathionylation, including 10 out of 11 enzymes of the CBB cycle [124], suggesting a central role of such modification in the global regulation of the carbon assimilation pathway. The role of CP12 in oxygen photoautotrophs beyond light/dark regulation, includes protection of PRK and GAPDH from oxidative stress [125] and a role in the glutathionylation of ADK3 that remains to be defined. Indeed, while ADK3 can be glutathionylated, the truncated ADK3-ΔCP12 cannot, even though the site of glutathionylation is not on the cysteine residues present in the C-terminal extension on ADK3 [121].
3. Molecular chaperones

Protein homeostasis, or proteostasis, is controlled by the so-called chaperones, in response to stress, and in normal conditions [126]. These proteins are thus essential for the “orchestration” of metabolism. Several molecular chaperones have now been identified and classified as conditionally disordered proteins [127]. The functions and underlying mechanisms of molecular chaperones are very well studied in bacteria and in the cytosol or endoplasmic reticulum of eukaryotic cells, but their role in the metabolic processes of the chloroplast have been overlooked.

An experimental search for IDPs in *C. reinhardtii* showed that chaperones and ribosomal proteins were the most represented IDPs [72]. Twenty proteins related to chaperone-function were identified, such as the heat-shock proteins HSP33, HSP40, HSP70s, FKBP, HSP90 co-chaperone, …. The disordered propensities of HSP70, HSP90 and HSP90 co-chaperone (Figure 5) have been shown experimentally in other organisms than algae [128], except for HSP33 that was characterised in *C. reinhardtii* [69].

HSP33 has been well studied in *E. coli* and is a redox chaperone that has a flexible C terminus that contains a conserved cysteine-rich center, Cys-X-Cys-Xn-Cys-X-X-Cys [23, 25]. These cysteine residues bind Zn$^{2+}$ under reducing conditions, thus maintaining the protein in its inactive form. In the oxidised form, the C-terminus unfolds after Zn$^{2+}$ release and the protein becomes active. To be active, this bacterial chaperone therefore needs to become disordered and this internal order-to-disorder transition is required to control substrate binding and release [129]. Interestingly, HSP33 is present in only two clades of eukaryotes: (i) Kinetoplastids and parasitic oomycetes and (ii) photosynthetic organisms, but only in algae and mosses, not in higher land plants [69]. In *C. reinhardtii*, and in general in the green algal lineage, the first cysteine residue of the conserved cysteine-rich center has been lost and
replaced by a glycine residue. As a result, the C-terminus of HSP33 has lost its ability to bind Zn in reducing conditions and is disordered in both oxidised and reduced forms, unlike the bacterial protein. Only few structural changes of HSP33 from *C. reinhardtii* have been observed experimentally between the oxidised and the reduced forms, as predicted (Figure 5A), and this was linked to a lower affinity for Zn$^{2+}$ [69]. In *C. reinhardtii*, HSP33 expression increases in response to light-induced oxidative stress and appears to participate in the rapid response to exposure of the algal cells to oxidizing conditions. This protein is therefore essential to algal proteostasis and metabolism.

In the cytosol of eukaryotic cells, the function of a plethora of proteins involved in signal transduction, like hormone receptors and kinases, depends on their folding by the HSP90 and HSP70 machineries [130, 131]. The HSP70s family comprises one subset of the best-studied chaperones in *C. reinhardtii*. These proteins play a role in all major subcellular compartments of the cell, and the HSP70B, HSP70C and HSP70A are targeted to the chloroplast, mitochondrion and cytosol, respectively [132]. The chloroplast HSP70B is induced either by light, or by heat or cold stress [132, 133]. Interestingly, the algal HSP70B that is highly homologous to other HSP70B from different species (see Table 1 in [132]) exhibits long disorder regions of more than 30 amino acid residues, and in particular a 71-amino acid disordered region at its C terminus (Figure 5B). Its flexibility allows binding and release of clients, in good agreement with its function. In *C. reinhardtii*, HSP70B and HSP90C can form a multimeric complex [134], although this complex has never been successfully reconstituted *in vitro* from the two recombinant proteins [135]. The algal HSP90C also bears a disordered C-terminus of about 50 aminoacid residues (Figure 5C). Both HSP70B and HSP90C were found in the Chlamydomonas phosphoproteome [136] indicating that they are phosphorylated as many IDPs. Despite their importance in the cell, these heat-
shock proteins, HSP70B, HSP90C and HSP33 remain understudied in *C. reinhardtii* and other algae compared to other organisms [137, 138].

**Algal IDPs in other cellular processes**

The experimental search for putative IDPs in *C. reinhardtii* also revealed few enzymes with long intrinsically disordered segments, involved in the algal metabolism (Figure 6) that are nuclear-encoded and targeted to either the mitochondrion or the chloroplast [72]. These IDR-containing enzymes have not been studied so far to our knowledge. Yet, one can hypothesise that the IDRs of these enzymes may also play important roles in the metabolism.

To cite but a few, the magnesium chelatase complex, involved in the light capture in the chloroplast, contain an IDR. This enzyme synthesises a precursor of chlorophyll. Downstream of the CBB, the synthetic pathway of the storage compound starch also involves an IDR-containing protein, the soluble starch synthase III (Figure 6B). Besides, four enzymes with long IDRs targeted to the mitochondrion were identified by Zhang *et al.*: the peptide methionine sulfoxide reductase, the dihydrolipoamide succinyltransferase oxoglutarate dehydrogenase E2 component (DLST-2OGDH) (Figure 6B), the F1F0 ATP-synthetase ε-subunit and the carbonic anhydrase CA-8, all involved in the respiration process.

**5. Conclusions**

The autonomous chloroplastic and mitochondrial genomes encode only a small subset of proteins and the majority of the proteins present in these organelles are nuclear-encoded and imported after their synthesis in the cytosol. This import process allows the optimization of metabolic pathways by IDPs, although the chloroplast and the mitochondrion originate from bacteria, which have a lower content in IDPs. Furthermore, IDPs, with their ductile properties, are very robust and well-suited to cope with the drastic environmental changes encountered
by algae (extreme changes in temperature, pressure, pH, light etc.) and tune the metabolic pathways according to these changes. The data obtained from the experimental study on *C. reinhardtii*, although not exhaustive [72], and the predictions from algal genomes [71] demonstrate the wide distribution of protein disorder in algae. Remarkably a study comparing the distribution of disorder in different cell types, from mammals (including *Homo sapiens*), plants, algae, and bacteria showed that the proteome of *C. reinhardtii* contains the highest number of residues with a disorder score [55].

Whereas very few studies have been undertaken hitherto, the growing interest in microalgae and their metabolic properties will undoubtedly trigger original studies on the role of IDPs in the regulation of cellular and metabolic processes in these microorganisms.

**Acknowledgements**

The authors are grateful to Pr Stephen Maberly who read our manuscript and made helpful comments.

**Funding**

Financial support is provided by CNRS, Aix Marseille Université, and the ANR SIGNAUX BioNRJ (ANR-15-CE05-0021-03).

**Figure captions**

**Figure 1: Mechanism of action of RCA.** A: Model of the interaction between RuBisCO and RCA and of the mechanism of action of RCA activation, figure extracted from [139]. The disordered C-terminal extension of RCA is indicated by the letter C. B: Model structure of RCA from the higher plant tobacco (*Nicotiana tabacum*), with C-terminal extensions not resolved in the electron microscopy images, represented by black lines (RCSB entry
3ZW6.pdb, [140]). C: Model structure of the red-type AAA+ rubisco activase CbbX from the purple bacteria *R. sphaeroides*. The C-terminal extensions not resolved in the negative stain images are represented by a black line (RCSB entry 3ZUH.pdb [83]).

**Figure 2: Disorder propensity of the C-terminal extension of RuBisCO activase and CbbX.**

A. Hydrophobic Cluster Analysis (HCA, [141]) plots of the C-terminal extension of rubisco activase and CbbX from the following organisms, from top to bottom: *C. reinhardtii* (accession number XP_001692244.1); *Chlorokybus atmophyticus*, isoform β (EST HO20293, [78]); *Klebsormidium flaccidum*, isoforms α2 and α1 (EST HO431775.1, [78]); the purple bacterium *Rhodobacter sphaeroides* (accession number AAC44827.1). B: Schematic alignment of RCA from cyanobacteria, chlorophyta and charophyta highlighting their C-terminal extension. Figure extracted from [78]. C: Schematic alignment of CbbX from purple bacteria, rhodophyta and heterokonta. Interestingly, and contrary to the green-type viridiplantae, the rhodophyta seem to have lost the C-terminal disordered extensions.

**Figure 3: Disorder in the essential pyrenoid component EPYC1 forming membraneless pyrenoid with RuBisCO.**

A: The circular dichroism spectrum of EPYC1 indicates random-coil structural properties, in contrast to the spectra obtained for RuBisCO and lysozyme. Figure extracted from supplementary figure 1 of [91]. B: Schematic of the composition of the pyrenoid, with a ratio of EPYC1: RuBisCO from one to four. Among the 135-190 proteins that are present in the pyrenoid, we chose to represent RCA that is discussed in this review. C: Schematic of the EPYC1 protein sequence with its four repeating domains. The HCA plot of the first repeat is shown (*C. reinhardtii* sequence, accession number PNW77439.1).
**Figure 4: Redox conformational transitions of CP12 governing the light-dark CBB-cycle regulation.** A: Schematic of the redox transitions of CP12 and GAPDH-CP12 formation. The two pairs of cysteine residues of the green-lineage CP12 are redox regulated by the thioredoxins (Trx) that are reduced (under light) or oxidised (under dark). B: Under reducing conditions, CP12 is fully disordered and highly flexible. C: Under oxidizing conditions, the two pairs of cysteine residues form two disulfide bridges, and CP12 is partially disordered. The C-terminal bridge is in a stable structure, depicted in blue, while the N-terminal bridge is in a region that interconverts between two states on a millisecond timescale, depicted in red: a random-coil state and a helical hairpin. D: This latter region unfolds when the C-terminal region is bound to GAPDH.

**Figure 5: Three chloroplastic intrinsically disordered chaperons of C. reinhardti.** HSP33, HSP70b and HSP90C are predicted as containing disordered regions using Pondr-VLXT (A score higher than 0.5 indicates a propensity for structural disorder)[142]. A: Pondr-VLXT score of C. reinhardti HSP33 (accession number, XP_001700446). The prediction of the reduced state of HSP33 is obtained using the mutant C274S,C309S,C312S. B: Pondr-VLXT score of C. reinhardti HSP70b (accession number XP_001696432). C: Pondr-VLXT score of C. reinhardti HSP90 (accession number XP_001702984.1).

**Figure 5: Examples of C. reinhardti metabolic enzymes with long intrinsically disordered regions.** A: Schematic of the metabolic pathways which possess enzymes with long intrinsically disordered regions and that are heat-resistant. These enzymes have been identified by Zhang et al. [72]. Some enzymes are chloroplastic: the magnesium chelatase complex (Mg$^{2+}$ CC, accession number: XP_001700902.1), the soluble starch synthase III (SSS III, accession number: AAY42381.1); while others are targeted to the mitochondrion:
the peptide methionine sulfoxide reductase (PMSR, accession number: XP_001689879.1), the
carbonic anhydrase 8 (CA-8, accession number: XP_001697606.1), the dihydrolipoamide
succinyltransferase, the oxoglutarate dehydrogenase E2 component (DLST-2OGDH,
accession number: XP_001692539.1) and the F1F0 ATP synthetase ε subunit (F1F0 ATP ε,
accession number: XP_001702609.1). Other IDPs described in this review are also included,
in particular those regulating the CBB cycle, and chaperones. B: The IUPred2A [143]
disorder prediction for the mitochondrial DLST-2OGDH and the chloroplastic soluble starch
synthase III are given as examples. A score higher than 0.5 indicates a propensity for
structural disorder.

References

1-526.
Duarte, Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological


[75] G.H. Lorimer, H.M. Miziorko, Carbamate formation on the epsilon-amino group of a lysyl residue as the basis for the activation of ribulose bisphosphate carboxylase by CO$_2$ and Mg$^{2+}$, Biochemistry 19(23) (1980) 5321-5328.


Rubisco

A

RCA

Rubisco

L_8S_8

RbcL

RbcS

End state

Release of the inhibitory sugar phosphate

B

Nicotiana tabacum

RCA

C

Rhodobacter sphaeroides

CbbX
A Chlorophyta RCA C-ter
(C. reinhardtii)

Charophyta β C-ter
(Chlorokybus atmophyticus)

Charophyta α2 C-ter (Klebsormidium flaccidum)

Charophyta RCA α1 C-ter (K. flaccidum)

Purple bacterium CbbX C-ter
(Rhodobacter sphaeroides)

B

C

purple bacterium
Rhodophyta
Heterokonta

Cyanobacteria
Chlorophyta
Charophyta

α1

α2

β

Rubisco recognition C
CTE

A

Chlorophyta
Charophyta
Cyanobacteria

B

C

Charophyta

α

β

C-terminal

ATPase

N

cTP
(1/µs – ps)

A

B

C

D

GAPDH

CP12_{red}

Trx

CP12_{ox}

S-S

S-S

GAPDH

60%

40%

k_{ex} (1/ms)

CP12_{ox}

GAPDH

(1/µs – ps)
A) Pondr-VLXT score

HSP33 reduced
HSP33 oxidized

Residue number

B) HSP70B

Pondr-VLXT score

Residue number

C) HSP90C

Pondr-VLXT score

Residue number
**Chloroplast**

- CO$_2$
- EPYC1
- CP12
- HSP33, HSP70B, HSP90C and HSP90 co-chap.
- HSP70C, HSP70E, FKBP, HSP40

**Mitochondrion**

- Methionine redox
  - PMSR
- ATP synthesis
  - F$_1$F$_0$ ATP $\varepsilon$
- Krebs cycle
  - DLST-2OGDH
- ADK-3
  - ATP
  - NADPH
- HSP70 co-chap.

**Chaperone**

- HSP33, HSP70, HSP90C and HSP90 co-chap.

**Carbohydrate synthesis**

- SSS III

**ATP synthesis**

- F$_1$F$_0$ ATP $\varepsilon$

**RuBisCO**

- thylakoid

**Chlorophyll synthesis**

- Mg$^{2+}$ CC pigments

**Krebs cycle**

- DLST-2OGDH

**Methionine redox**

- PMSR

**Krebs cycle**

- DLST-2OGDH

**IUPred 2A score**

**Residue number**

**IUPred 2A score**

**Residue number**