Cryptic disorder out of disorder: encounter between conditionally disordered CP12 and glyceraldehyde-3phosphate dehydrogenase

Hélène Launay, Patrick Barré, Carine Puppo, Yizhi Zhang, Stéphanie Maneville, Brigitte Gontero, Véronique Receveur-Bréchot

▶ **To cite this version:**

HAL Id: hal-03855995

https://cnrs.hal.science/hal-03855995

Submitted on 16 Nov 2022

Cryptic disorder out of disorder: encounter between conditionally disordered CP12 and glyceraldehyde-3-phosphate dehydrogenase

Helene Launay[1]*, Patrick Barre[2], Carine Puppo[1], Yizhi Zhang[1], Stéphanie Manneville[2], Brigitte Gontero[1]*, Véronique Receveur-Brechot[1,2]*.

1 Aix Marseille Univ, CNRS, BIP, UMR 7281, IMM, 13402 Marseille Cedex 20, France.
2 Aix Marseille Univ, CNRS, IMR, UPR 3243, IMM, 13402 Marseille Cedex 20, France.

ABSTRACT

Among intrinsically disordered proteins, conditionally disordered proteins undergo dramatic structural disorder rearrangements upon environmental changes and/or post-translational modifications that directly modulate their function. Quantifying the dynamics of these fluctuating proteins is extremely challenging but paramount to understanding the regulation of their function. The chloroplast protein CP12 is a model of such proteins, and acts as a redox switch by formation/disruption of its two disulfide bridges. It regulates the Calvin cycle by forming, in oxidized conditions, a supramolecular complex with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and then phosphoribulokinase. In this complex both enzymes are inactive. The highly dynamic nature of CP12 has so far hindered structural characterization explaining its mode of action. Thanks to a synergistic combination of small-angle X-ray scattering, nuclear magnetic resonance and circular dichroism that drove the molecular modeling of structural ensembles, we deciphered the structural behavior of *Chlamydomonas reinhardtii* oxidized CP12 (CP12$_{ox}$) alone and in the presence of GAPDH. Contrary to sequence-based structural predictions, the N-terminal region is unstable, oscillates at the ms timescale between helical and random conformations, and is connected through a disordered linker to its C-terminus, which forms a stable helical turn. Upon binding to GAPDH, CP12$_{ox}$ undergoes an induced unfolding of its N-terminus. This phenomenon called cryptic disorder contributes to decrease the entropy cost and explains CP12 unusual high affinity for its partners.

KEYWORDS: NMR, photosynthesis, protein dynamics, protein-protein interaction, SAXS.

CORRESPONDING AUTHORS:

Hélène Launay (helene.launay@inserm.fr)
Brigitte Gontero (bmeunier@imm.cnrs.fr)
Véronique Receveur-Bréchot (veronique.brechot@imm.cnrs.fr)[1]

---

[1] **ABBREVIATIONS:** CBB, Calvin-Benson-Bassham; CD, Circular Dichroism; CP12$_{ox}$ oxidized CP12; CP12$_{red}$ reduced CP12; Dmax, maximal internal distance ; DTT, Dithiothreitol; EDTA, Ethylene diamine tetraacetic acid; ESI-TWIM, ElectroSpray Ionization-Traveling Wave Ion Mobility; GAPDH : glyceraldehyde-3-phosphate dehydrogenase, het-NOE , heteronuclear nuclear Overhauser effect; HPLC, High Pressure Liquid

INTRODUCTION

Oxygenic photosynthetic organisms require a high level of regulation of their metabolic activities to cope with highly variable environmental conditions. Intrinsically disordered proteins (IDPs) have been described as essential regulators of cell cycle in eukaryotic species [1]. IDPs have the properties to sample rapidly a myriad of conformational states and many of them adopt a well-defined conformation upon interaction, a process referred to as "induced folding" [1,2]. Despite the increasing interest in these peculiar proteins, and their high abundance in eukaryotes, there are only few examples of IDPs characterized from photosynthetic organisms for which the disordered properties have been experimentally evidenced. Among such IDPs from Plantae are the dehydrin proteins [3–8], the GRAS proteins family [9], the NAC transcription factor [10], the chloroplast preprotein receptors [11] and the chloroplast protein CP12 [12–14].

CP12 is universally present in oxygenic photosynthetic organisms including cyanobacteria, algae, higher plants [15,16] as well as in cyanophages [17]. Also, a number of chloroplast proteins possess a C-terminal extension highly homologous with CP12, among which the B-type subunit of the A_2B_2 isoform of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in higher plants [18–21] as well as adenylate kinase 3 in the green alga *Chlamydomonas reinhardtii* [22]. CP12 controls CO_2 assimilation by regulating the Calvin-Benson-Bassham (CBB) cycle depending on its redox state [23,24]. Under light, CP12 is reduced (CP12$_{red}$) and is fully disordered [14]. Under dark, CP12 is oxidized (CP12$_{ox}$), partially ordered [12,13,25] and presents a high affinity for GAPDH, a CBB cycle key enzyme [26,27]. The GAPDH/CP12$_{ox}$ subcomplex in turn enables binding to a second CBB cycle enzyme, the phosphoribulokinase (PRK) [26]. In the ternary complex, the two enzymes are inactive. Despite the central role played by this regulatory protein, it has so far eluded a comprehensive structural characterization, thus hampering the mechanisms of inhibition of PRK and GAPDH from being deciphered.

In *C. reinhardtii*, the reduction/oxidation regulation of CP12 is mediated via its four cysteine residues, which are involved into two disulfide bridges under oxidized state (C23-C31 and C66-C75) [28]. The disorder to order transition induced by formation of these disulfide bridges endows CP12 with the same peculiar properties as other IDPs named conditionally disordered proteins. These proteins adopt different conformations allowing

different functions depending on environmental conditions [29]. CP12 has been proposed as a model for redox-conditionally disordered protein [30]. In *C. reinhardtii* and in the higher plant *Arabidopsis thaliana*, $CP12_{ox}$ has been described as being highly flexible [12,13] despite a significant number of α-helices that have been observed in circular dichroism (CD) experiments [12]. A sequence-based 3D-model of *C. reinhardtii* $CP12_{ox}$ suggested the presence of α-helical segments on each side of the N-terminal disulfide bridge (C23-C31), whereas the C-terminal region was expected to fold into a turn [31].

Very little experimental biophysical data could provide accurate insights on the conformational state of $CP12_{ox}$, and so far, information on only the 22 residues that are centered on the C-terminal disulfide bridge in *A. thaliana* CP12 is available, provided by Nuclear Magnetic Resonance (NMR) spectroscopy [13]. The structural properties of the rest of $CP12_{ox}$ have remained elusive, when studied either free or within the $GAPDH/CP12_{ox}$ subcomplex. In the crystal structure of the binary subcomplex from the higher plant *A. thaliana* [13], and in the cyanobacterium *Synechococcus elongatus* [32], only the same C-terminal moiety, bound to GAPDH, could be observed. The global shape of the *A. thaliana* subcomplex that was inferred from Small Angle X-ray Scattering (SAXS) data suggested that $CP12_{ox}$ becomes compact upon binding to GAPDH [33]. In contrast, Electronic Paramagnetic Resonance spectroscopy revealed that the *C. reinhardtii* $GAPDH/CP12_{ox}$ subcomplex is a fuzzy complex, with the N-terminal region of CP12 being very flexible [34]. These contradictory observations question the specific mode of action of CP12 and how it may depend on the organism.

All the obstacles to understanding the structural determinants of the function of IDPs such as CP12 arise from their inherent internal dynamics that lead to multitudinous conformations. However, this disorder and these highly dynamic properties are inseparable from their function. As highlighted recently [35], there is therefore an urgent need not only to observe the disordered nature of IDPs, but also to quantify these dynamics, notably by means of population ensembles [36]. Only this, will allow us to really understand how structure and dynamics are intertwined in these peculiar proteins, and how these features control their function. Such advances demand insightful and hybrid biophysical strategies [37].

Here, we synergistically combine several biophysical techniques (SAXS, NMR, CD and molecular modeling) that offer powerful means for the quantitative analysis of flexible systems, to assess the conformational sampling of *C. reinhardtii* free $CP12_{ox}$ and $CP12_{ox}$ in

association with GAPDH. We provide a quantitative molecular description of the structure and dynamics of CP12$_{ox}$ that sheds a new light on the thermodynamics and the mechanisms of the interaction between CP12$_{ox}$ and GAPDH.

RESULTS

CP12$_{ox}$ is intrinsically dynamic in solution, with a conformational exchange on the ms timescale

To gain insights into the conformation sampling of CP12$_{ox}$ at a residue-specific level we first used NMR spectroscopy. As was observed previously [12], the ^{1}H-^{15}N heteronuclear spectrum of CP12$_{ox}$ presents a series of 20 resonances that have a large ^{1}H frequency dispersion (7.1 to 8.9 ppm, black resonances in Figure 1a), typical of a globular fold. In addition, 33 resonances overlay the resonances that were observed on the ^{1}H-^{15}N heteronuclear spectrum of CP12$_{red}$ (Figure 1), suggesting that CP12$_{ox}$ also contains disordered segments. According to our previous findings [14], we have used in the following, the NMR resonances of CP12$_{red}$ as a reference of the fully disordered state of CP12.

The assignment of the resonances in the spectrum of CP12$_{ox}$ allowed us to compare, residue by residue, the NMR chemical shifts of the residues between CP12$_{ox}$ and CP12$_{red}$ (Figure 1b). The most drastic differences are in the C-terminal region L62-Y78 indicating that oxidation triggered the folding of this region (Figure 1). The Cα chemical shifts of this region indicate a helical formation (SI Figure 1). Robust dihedral angles are predicted for 13 residues in this L62-Y78 region (SI Figure 2). From these restraints, we built a structural model of this stable L62-Y78 region using CS-Rosetta [38]. Other residues of the sequence have chemical shift similar to those of CP12$_{red}$ and are predicted to be highly dynamic.

The N-terminal region of CP12$_{ox}$ exhibits unusual NMR characteristics. Most of the NMR chemical shifts assigned to this region are close to those in disordered CP12$_{red}$ (Figure 1b and SI Figure 1). Nonetheless, in the close vicinity of the N-terminal disulfide bridge, a few residues (for example W35) have chemical shifts highly displaced compared to CP12$_{red}$, which could be ascribed to the proximity of the disulfide bridge. In addition, many resonances in the ^{1}H-^{15}N heteronuclear spectrum could not be assigned due to significant broadening in the triple frequency experiments, and residues K17 to A24 and S28 to V33 were not assigned to any NMR resonance. These unassigned resonances have low proton

frequency dispersion and overlap (red in Figure 1a). A wide range of experimental conditions (temperature ranging from 10 to 27°C, pH ranging from 6 to 9, addition of Cu^{2+}) were tested to enhance the intensity of these broadened resonances, without success.

To understand the origin of the specific broadening of these N-terminal resonances beyond detection, ^{15}N relaxation experiments were performed on CP12$_{ox}$ and CP12$_{red}$ (Figure 2). As expected for CP12$_{red}$, the {^{1}H}-^{15}N heteronuclear NOE (het-NOE) values and ^{15}N R_2 relaxation rates are all typical of intrinsically disordered polypeptides ({^{1}H}-^{15}N het-NOE < 0.2 and ^{15}N R_2 < 5 s^{-1}, Figure 2). In the case of CP12$_{ox}$, the {^{1}H}-^{15}N het-NOE values are significantly higher than the reference CP12$_{red}$ values, confirming that CP12 is partially folded upon oxidation (Figure 2a). As expected for a folded region, the {^{1}H}-^{15}N het-NOE values for the L62-Y78 residues are higher than 0.6 (Figure 2a). The ratio of the two ^{15}N relaxation rates R_2/R_1 is a function of the rotational correlation time, and therefore of the hydrodynamic properties of the protein. For example, the expected values for CP12$_{ox}$ behaving as a rigid-body as modeled by Gardebien et al [31] is 8.1 ns according to hydropro [39] (SI Figure 3). Here, the experimental R_2/R_1 ratios measured for the L62-Y78 residues correspond to rotational correlation times of 4 ns (SI Figure 3). This correlation time is significantly shorter than what was expected. This excludes the possibility that CP12$_{ox}$ is fully globular. On the contrary, these relaxation parameters confirm the dynamical nature of CP12$_{ox}$, and indicate that the ordered L62-Y78 region is connected to the rest of the protein by residues with heterogeneous flexibilities. Indeed, upstream of this folded region, low {^{1}H}-^{15}N het-NOE values (< 0.4, Figure 2a), typical of a disordered linker between globular domains, are observed for residues A51-D60 and for the N-terminal extremity, up to residue D7. This last result suggests that these two segments of CP12 are random coil. As expected for disordered linker and tail, low ^{15}N R_2 relaxation rates are also measured for residues in these two small regions (Figure 2b).

Between these two disordered segments (i.e. residues L8 to D50), we observe, as mentioned above, unusual NMR relaxation properties. For the corresponding residues, the NMR resonances are of very low intensity because of very broad linewidths. Thus, many resonances could not be assigned and the dynamic data are sparse. The ^{15}N R_2 relaxation rates reflect the broadening of the resonances. Typically, six unassigned resonances exhibit very high ^{15}N R_2 rates (18 to 38 s^{-1}, Figure 2b). These resonances must correspond to residues within the regions A15 to K25 and S28 to L41 since these are the only residues with

missing resonances. High ^{15}N R_2 rates might arise from two phenomena [40]: (i) a very low rotational correlation time due to the formation of large species or (ii) a chemical exchange in the NMR intermediate exchange regime (100 µs - 100 ms timescale). To check the first hypothesis, we performed translational diffusion measurements on the observed CP12$_{ox}$ resonances. The diffusion coefficient of the observed species is 13±0.5 10^{-11} m^2s^{-1}, which corresponds to a Stokes radius of 19.1Å. This value is incompatible with the hypothesis of line broadening due to large species (SI Figure 4). This does not exclude the formation of larger aggregates but these would not be detected in our NMR data. On the other hand, the second hypothesis is supported by previous studies using ESI-TWIM (ElectroSpray Ionization-Traveling Wave Ion Mobility) mass spectrometry under non-denaturing conditions [25] which showed the co-existence of two monomeric conformers. Furthermore, the few resonances of this region that could be assigned indicate that these residues have chemical shifts identical to those in the disordered CP12$_{red}$ state (L8 to V16 and S42 to D50, Figure 1b). On the contrary, the sparse dynamical data that could be measured include high {^1H}-^{15}N het-NOE values (> 0.5 Figure 2a), indicating that these residues do not behave like a disordered linker. These apparent contradictory results are reconciled if we consider that this region is interconverting on a 100 µs - 100 ms timescale between a random-coil state and a more structured state.

Taken together, the NMR data on CP12$_{ox}$ show that the protein is composed of segments with distinct conformational states. The C-terminal end folds into a stable turn that is connected to the rest of the protein via a disordered linker. The N-terminal end including the histidine tag is also disordered. The central region surrounding the N-terminal disulfide bridge is in a conformational exchange on a 100 µs to 100 ms timescale, involving a disordered state and one or more structured state. The secondary structural propensities (SSP [41]) derived from NMR chemical shifts suggests that this more structured state is helical. Indeed, residues S42 to K48 had positive SSP values typical of IDPs with a tendency to form α-helices (Figure 1c). Similarly, our former NMR experiments on CP12$_{red}$ indicated a propensity, although very low, for the residues surrounding the N-terminal disulfide bridge to form α-helices [14], which was supported by *in silico* structure predictions [31].

Displacement of the conformation equilibrium towards a helical-state reveals a two-state exchange

The 100 µs to 100 ms exchange limits the number of NMR observable that could allow a residue specific characterization of each state in exchange. We therefore used CD as an alternative method to monitor the putative secondary structures of $CP12_{ox}$ on a global scale. The CD spectra were typical of a protein containing a significant amount of α-helices (Figure 3). Furthermore to check the hypothesis of a conformational exchange, we aimed at displacing the putative equilibrium using increasing concentrations of 2,2,2-trifluoroethanol (TFE), which is known to stabilize residual or transient structures. The CD spectra at five increasing TFE concentrations indicate an increase of the helical content as revealed by a lower value of ellipticity at 220 nm, as well as a decrease of disordered content as revealed by the higher value of ellipticity at 193 nm, and the shift of the minimum of ellipticity from 206 to 208 nm (Figure 3). Furthermore, an isodichroic point was observed at 203 nm, typical of a two-state equilibrium that is displaced by TFE (Figure 3).

In order to localize the helical secondary structure on $CP12_{ox}$, we performed ^1H-^{15}N heteronuclear experiments on $CP12_{ox}$ with increasing concentration of TFE (Figure 4). Several unassigned resonances, which were clustered on the central region of the ^1H NMR spectral width in 0% TFE, were significantly displaced with 20% TFE, indicating a shift of the equilibrium towards a more folded state. The broad linewidths of these resonances indicate that the conformational equilibrium is shifted, but not enough to escape the NMR intermediate exchange regime. Nevertheless, these resonances can be unambiguously assigned to the regions K17 to K25, S28 to V33 and D36 to L41, although we could not achieve the residue specific assignment because of broad linewidth. Besides, the recorded Cα chemical shifts of residues H47 to K53 are significantly displaced towards positive values in the presence of TFE (Figure 4), which confirms that TFE induces the formation of α-helices. On the contrary, the NMR resonances of the L62 -Y78 region shows limited chemical shift displacement with increasing TFE concentration (Figure 4 and SI Figure 5), as expected for a stable fold. These data confirm that the C-terminus of $CP12_{ox}$ is stable, and suggest that the region surrounding the N-terminal disulfide bridge has a propensity to form α-helices.

This is consistent with former CD experiments on two $CP12_{ox}$ mutants disrupting each disulfide bond[34]. These experiments showed that loss of the N-terminal disulfide bond leads to a lower helical content, while loss of the C-terminal disulfide bridge does not

destabilize the α-helices of the protein. These different results evidence that the strong helical propensity revealed by the TFE titration is localized within the N-terminal region.

So far, the following direct conclusions can be derived from our experimental data: the C-terminal region (L62-Y78) is stable, folded into a structure that we solved here. This domain is connected to the rest of the protein via a disordered linker. The region surrounding the N-terminal disulfide bound is fluctuating on a 100 µs – 100 ms timescale. Our data derived from NMR, CD and the TFE titration allowed us to raise the hypothesis that this region is interconverting between a disordered and a helical state.

Bimodal distribution of the CP12$_{ox}$ conformers between a disordered state and a helical hairpin

We then used the SAXS technique, which is particularly suitable for the structural analysis of disordered and/or dynamical systems [42], to assess the overall structural properties of the different CP12ox species present in solution, and to confront our experimental data to the hypothesis of a conformational exchange of CP12$_{ox}$ between a helical and a random-coil state (Figure 5).

We used size-exclusion chromatography in line with SAXS (SEC-SAXS) to separate the different species according to their molecular mass. Two well-separated peaks were observed (SI Figure 6), the first peak, which had low intensity, corresponding to large aggregates (unseen in the NMR experiments). Analysis of SAXS data from the second peak yielded a linear Guinier plot (SI Figure 7) and the molecular weight inferred from the I(0) scattering intensity was 12,650 Da, corresponding to monomeric CP12$_{ox}$ (11,060 Da for ^{15}N-CP12$_{ox}$ based on the sequence), discarding the hypothesis of several oligomeric species in the second elution peak. The radius of gyration, Rg, that was obtained using the Guinier analysis is 23 ± 0.1 Å (SI Figure 7), and the maximal internal distance Dmax is 100 ± 2 Å (SI Figure 8). These values can be compared to the values 28.2 ± 0.5 Å and 119.3 ± 5 Å respectively, that were measured for the fully disordered CP12$_{red}$ reference [14]. While the Rg is the result of an average over all the conformers, the inferred Dmax arises from the most extended conformers in the population [43]. A significantly lower Rg and an only slightly lower Dmax for CP12$_{ox}$ compared with the respective values in CP12$_{red}$ reveals the presence of several populations in CP12$_{ox}$, with at least extended and compact conformations present. The values of all these parameters definitely support the hypothesis of an equilibrium

between different conformational states versus the transient formation of oligomeric species.

SAXS *ab initio* shape determination can be extremely interesting when used to gain insights on specific features of the different conformations that are adopted by partially folded proteins. In the case of $CP12_{ox}$, excellent fits to the data are obtained when the overall envelope exhibits a compaction in the center of the protein and two extended protrusions (SI Figure 8). This envelope represents a population weighted average of the ensembles of conformers in exchange, and when a peculiar structural element is present in a majority of conformers, it can be observed in this global envelope. In the case of $CP12_{ox}$, a hairpin fold appears at the center of the polypeptide chain with another turn at one of the extremities. These features reflect the structures that are induced by the two disulfide bonds and suggest that the central region of $CP12_{ox}$ that corresponds to the 'NMR silent' region L8-D50, can be partially folded.

All the power of SAXS is expressed when one can combine complementary information from other biophysical techniques to build models and compare them with the results from the SAXS data. Taken together, the NMR, CD and *in silico* sequence analysis described above allowed us to build two series of models of $CP12_{ox}$, which may correspond to the two possible exchanging states (Figure 5c). Using RanCh [44], we generated a first pool of 10,000 structures. We chose to model the 'NMR silent' region in this pool, by two helices encompassing residues L8-A24 and S28-D50 and connected to a hairpin by the C23-C31 disulfide bridge, as in the molecular model proposed by Gardebien *et al* [31] since this model is consistent with all our data (secondary structures observed by CD and predicted by sequence homology, NMR secondary structural propensities, SAXS overall envelope, …).In these structures, the C-terminal residues L62-Y78 are folded into the structure that was determined using our NMR data and the rest of the protein is considered as random coil (Figure 5c). We generated a second pool of 10,000 structures to model the second exchanging state using the same framework except that residues L8-A24 and S28-D50 are in random conformations, as detailed above according to the NMR chemical shifts, and the two disulfide bridges are kept (Figure 5c).

From these two pools of conformations, an ensemble of conformers that best describes the scattering data was selected by GAJOE [45]. The best fit to the data (χ^2=1.5, Figure 5a) is obtained with an ensemble of conformers that followed a bimodal distribution

(Figure 5b and d). We could decompose this ensemble into two Gaussian distributions via a least-squares fitting. The conformers of the two ensembles have average Rg values of 19.6 ± 0.3 Å and 27 ± 1 Å, respectively. Strikingly, the two ensembles superimpose very well on the distributions generated by RanCh of the helical and of the random pools, with similar average Rg values and widths of the distributions for each pool (mean Rg values of 21 Å and 26.7 Å respectively, Figure 5b and d). Quantifying the relative populations of the exchanging $CP12_{ox}$ conformational states from this deconvolution leads to 55 to 60% of the conformers having the two helices L8-A24 and S28-D50 folded as modeled by Gardebien et al [31], and approximately 40 to 45% of the conformers having no stable conformation in this region (Figure 5c and d).

These values are consistent with the proportions of helical content that could be derived from the CD spectra of $CP12_{ox}$. Our structural models are indeed composed of 37 residues involved in interconverting helices in the N-terminal region, and 9 residues in the stable C-terminal helices, meaning that the helical models exhibit 46% helical content (46 residues involved in helices over a total of 100 residues for $CP12_{ox}$), while the random models exhibit only 9% helical content. A mixture of ~55-60% of the helical models and 40-45% of the models with random conformations in the N-terminal region would lead to a theoretical average of ~29-30% of helical content (0.55x46+0.45x9). This value is very close to the one derived from the mean residue molar ellipticity measured at 222 nm (Θ_{222}) for $CP12_{ox}$ in the absence of TFE, of 10,500 $deg.cm^2.dmol^{-1}$ (Figure 3), corresponding to a helical content of 27% with respect to the theoretical Θ_{222} of 38,500 $deg.cm^2.dmol^{-1}$ for a 100% helical $CP12_{ox}$. Similarly, in the presence of 40% TFE, the experimental Θ_{222} was 17,800 $deg.cm^2.dmol^{-1}$ (Figure 3), corresponding to 46% helical content, suggesting that at this TFE concentration, the equilibrium is totally or almost totally shifted towards the helical conformations. As most of the NMR resonances remained broad in the N-terminal region in the presence of 40% TFE, one cannot exclude that a small proportion of $CP12_{ox}$ is still exchanging conformation at the 100 µs - 100 ms timescale at high TFE concentration.

To sum up, all our experimental data and structure predictions on $CP12_{ox}$ indicate an equilibrium at the 100 µs – 100 ms timescale between two conformational ensembles: in a slight majority population, the region surrounding the N-terminal disulfide bridge is helical, while in the other population, it is unfolded.

Conformations of CP12$_{ox}$ bound to GAPDH

The stable C-terminal region of CP12$_{ox}$ that was characterized above has a high affinity for GAPDH [26]. Previous attempts using X-ray crystallography allowed to solve the structure of this C-terminal region in the GAPDH/CP12$_{ox}$ subcomplex from *A. thaliana* [13] and *S. elongatus* [32], but failed to gain information about the N-terminal region of CP12$_{ox}$ upon binding to GAPDH. We therefore used SAXS to gain structural information on the dynamical N-terminal extension of CP12$_{ox}$ within this high molecular weight complex from *C. reinhardtii* (Figure 7), using a similar strategy as Del Guidice *et al* on the *A. thaliana* complex [33].

On a global scale, the Rg that was determined from the scattering curve of the GAPDH/CP12$_{ox}$ subcomplex using the Guinier analysis is significantly larger than what we determined from the free GAPDH scattering curve (38.4 ± 0.2 Å versus 32.5 ± 0.5 Å, SI Figure 7). Similarly, the maximal internal distance in the GAPDH/CP12$_{ox}$ subcomplex (190 ± 5 Å) is drastically larger and more than twice that of free GAPDH (90 ± 0,5 Å; SI Figure 9). This result indicates that CP12$_{ox}$, when interacting with GAPDH, adopts an extended conformation. Determination of the molecular mass of the subcomplex using the online program SAXS MoW2 [46] yields a value of 166 kDa, compared with a sequence-based Mw of 10.9 kDa for His-tagged CP12$_{ox}$ and 147.5 for GAPDH tetramer. Similarly, the program GNOM [44] calculates, from the SAXS data, a Porod volume of 250,820 Å^3 for the subcomplex and 206,430 and 22,480 Å^3 for free GAPDH and free CP12$_{ox}$, respectively. These results consistently provide a stoichiometry of the subcomplex of two CP12$_{ox}$ monomers binding one GAPDH tetramer and corroborate former studies using ion mobility mass spectrometry under non-denaturing conditions [25].

To perform a molecular analysis of the SAXS data of the subcomplex, we needed an atomic model of GAPDH from *C. reinhardtii* beforehand. We thus built homology models based on the X-ray crystal structure of GAPDH from *A. thaliana* (PDB ID: **3K2B** [47]) and from spinach (PDB ID: **1NBO** [48]) and compared their theoretical scattering curve to the experimental scattering curve of free GAPDH from *C. reinhardtii*. The fits to the data were excellent and equivalent for the two models (χ^2 of 2.9 in both cases, see SI Figure 10). The same strategy was applied to build an atomic model of GAPDH in the complexed state when

bound to CP12$_{ox}$, using the crystal structure of *A. thaliana* GAPDH in presence of the last 20 C-terminal residues of CP12 (PDB ID: **3QV1** [13]).

Based on this latter GAPDH atomic model, we then built a series of models of the subcomplex GAPDH/CP12$_{ox}$. In these models, two molecules of CP12$_{ox}$ were bound to GAPDH tetramer, in agreement with the stoichiometry confirmed by SAXS. These molecules of CP12$_{ox}$ are attached to GAPDH through their C-terminal turn L62-Y78 in the conformation that was determined from our NMR data, since it is very close to the structure of the analogous *A. thaliana* CP12$_{ox}$ region bound to GAPDH (RMSD 0.91 Å, SI Figure 11; PDB ID: **3QV1** [13]).The N-terminal region of CP12$_{ox}$ proteins in the subcomplex was then modeled, using a similar strategy as for free CP12$_{ox}$ described above. Two families of 10,000 conformations each were built using Ranch: in the first ensemble, the dynamic N-terminal region of CP12$_{ox}$ is in the helical fold described above whereas in the second ensemble it is disordered. As for free CP12$_{ox}$, the N-terminal tail and the linker region (residues A51-D60) are modeled in random disordered conformations.

We ran the program GAJOE with each pool of conformations (the helical and the disordered one) individually as starting ensemble to select the population of conformers that best represent the SAXS data. Only the most extended conformations were selected, with Rg distributions significantly shifted towards longer Rg values (Figure 6). Accordingly, the disordered ensemble agrees better with the experimental data (χ^2=5.3, Figure 6c and d) than the helical ensemble (χ^2=29.7, Figure 6a and b), or even when the two ensembles are taken together as starting ensemble. This suggests that the N-terminal region adopts preferentially the most extended conformations, in agreement with the high observed Dmax values, and that the equilibrium between the helical and the disordered states of CP12$_{ox}$ is displaced towards the most unfolded conformations upon binding to GAPDH.

DISCUSSION

A hybrid experimental approach reveals the dynamic structural properties of CP12$_{ox}$

Because of the peculiar structural and dynamical properties of CP12, the molecular understanding of its regulation has been hindered so far. Here, we have quantified the structure and dynamics of this protein and have deciphered the wide range of timescales

that affects the large conformational dynamics of this protein. In particular, the region that encompasses the N-terminal disulfide bridge is involved in a conformational exchange that has eluded all previous structural characterizations. The same region of CP12$_{ox}$ was not observed by NMR in the homologous *A. thaliana* CP12$_{ox}$ (BMRB ID: **17926**, SI Figure 11) [13]. It was not observed either in the crystal structures of CP12$_{ox}$ bound to GAPDH from *A. thaliana* (PDB ID: **3QV1** [13]) and from *S. elongatus* (PDB ID: **3B1J** [32]). The study of such flexible states, described as 'dark states', presents challenges that require the synergy of multiple experimental and computational approaches as underlined by Bhowmick *et al* [37]. Using NMR, SAXS and CD, we have shown that the C-terminus of free CP12$_{ox}$ adopts a stable fold and that the seven N-terminal residues as well as residues within the A51-D60 region behave as disordered linkers, adopting a myriad of conformations in equilibrium at the ps-µs timescales. More strikingly, the rest of the protein (residues L8-D50, around the C23-C31 disulfide bridge), interconverts at the 100 µs - 100 ms timescale between two families of conformations, a folded helical state, and a disordered state.

CD spectroscopy allowed us to follow the displacement of this two-state conformational equilibrium, and, together with NMR spectroscopy, indicated that one of the families of conformers has a high helical content, consistent with the model proposed by Gardebien *et al* [31]. Our NMR data further identified that these two sets of conformers, one of which being disordered, arise from the structural interconversion of residues L8 to D50; we could infer from these data the timescale of this chemical exchange. These experimental observations allowed the construction of an ensemble of molecular models that was then compared with the results from the experimental SAXS data, with a proportion of extended (40%) and compact/helical (60%) conformations. Ion mobility mass spectrometry also observed the co-existence of these two conformers [25]. Thanks to that combined analysis, the conformational exchange of the N-terminal region of CP12$_{ox}$ in *C. reinhardtii*, could be finally deciphered.

The N-terminal region of free CP12$_{ox}$ is largely unstable despite confident sequence-based predictions for helices

Noticeably, the high scores of confidence of the secondary structure predictions [31] are not consistent with the observed highly dynamic properties of the N-terminal region of CP12$_{ox}$, all the more so since these predictors do not take into account the redox state of the

cysteine residues. The secondary and tertiary structure of CP12 is probably unstable because CP12 is too small a protein to form a sufficiently large hydrophobic core, and the predicted interatomic contacts involving conserved hydrophobic residues may not be sufficient to stabilize a 3D structure [31]. In the theoretical model, the absence of a disulfide constraint in this N-terminal region was predicted to destabilize the protein helices, in agreement with our data on CP12$_{red}$ [14]. Our experimental data here confirm that this helical hairpin is highly unstable. Nonetheless, one cannot exclude the possibility that the unstable folded conformer would be stabilized upon binding the corresponding partner of this region, PRK, which would thus provide an add-on hydrophobic moiety.

The low entropy cost of the binding allows a high affinity between GAPDH and CP12$_{ox}$

In contrast, the C-terminal end of CP12$_{ox}$ forms a very stable structural element. The structure solved here using chemical shift information that was obtained from the *C. reinhardtii* L62-Y78 region resembles that from *A. thaliana* that was solved by Fermani *et al* using chemical shifts and ^1H homonuclear NOEs (PDB ID: **2LJ9** [13], RMSD 1.82 Å). Interestingly, our structure from *C. reinhardtii* is closer to the conformation of CP12 bound to GAPDH either from *A. thaliana* (PDB ID: **3QV1** [13], RMSD 0.91 Å) or *S. elongatus* (PDB ID: **3B1J** [32], RMSD 0.91 Å), whereas in *A. thaliana,* free CP12$_{ox}$ is less superimposable to CP12$_{ox}$ bound to GAPDH (RMSD = 2.43 Å between **2LJ9** and **3QV1** PDB structures, SI Figure 11). This indicates that, unlike *A. thaliana* CP12$_{ox}$, free *C. reinhardtii* CP12$_{ox}$ is already in a conformation close to the one bound to GAPDH. Our dynamical data (^{15}N R$_2$, ^{15}N R$_1$ and {^1H}-^{15}N het-NOE) indicate that the L62-Y78 region is stable in *C. reinhardtii* CP12$_{ox}$ (Figure 7). On the contrary, for *A. thaliana* CP12$_{ox}$ and according to Fermani *et al* [13], the C-terminus of free protein undergoes a conformational exchange, which is in favor of a conformation selection model for GAPDH binding. While the entire binding site of the GAPDH enzyme is conserved between the enzymes of the higher plant and the alga (except for the T96S mutation, numbering according to *A. thaliana*), several mutations are observed between the two CP12$_{ox}$ homologs (SI Figure 11), with many aliphatic residues (A, L, V) in place of acidic (D, E, N) and glycine residues between the alga and the plant CP12. These mutations endow *C. reinhardtii* CP12 with a more solid hydrophobic core than in *A. thaliana* CP12, leading to a more stable structure. The higher affinity of CP12$_{ox}$ for GAPDH in the green alga (K$_D$ = 0.4 nM [26]) compared to the *A. thaliana* complex (K$_D$ = 0.18 µM [27]) is certainly a consequence of the

higher structural stability of the *C. reinhardtii* L62-Y78 region in a fold ready for GAPDH binding (Figure 7). In addition, the affinity of IDPs for their partner is generally medium to low ($K_D > 1$ µM), because they often undergo an induced folding upon binding, a process responsible for an important entropy cost [49]. The higher stability of the *C. reinhardtii* CP12 C-terminus certainly contributes to a lower entropy cost, as the binding fits well to the "lock and key" model, thereby increasing the affinity between the two partners compared to the analogous interaction in *A. thaliana*, which, by contrast follows the conformational selection model of binding [13] . In *A. thaliana*, of interest, the effect of CP12 on GAPDH activity was negligible, in contrast to the inhibition of GAPDH observed in *C. reinhardtii* [20,50,51].

Another decrease in entropy cost in *C. reinhardtii* GAPDH/CP12$_{ox}$ complexation may arise from the N-terminal region of CP12$_{ox}$, which appears to undergo a remote induced unfolding upon its C-terminus binding to GAPDH. Indeed according to our SAXS data, CP12$_{ox}$ when bound to GAPDH does not behave as when it is free, and adopts preferentially the most unfolded conformations (Figure 6). Preliminary NMR experiments on the subcomplex also showed that the N-terminal region of CP12$_{ox}$ bound to GAPDH is highly dynamic (SI Figure 12). Such a phenomenon of induced unfolding is termed cryptic disorder [29]. Thus, the gain in entropy of this region probably also contributes to the increased affinity between *C. reinhardtii* CP12$_{ox}$ and GAPDH, in contrast to the *A. thaliana* complex, in which the CP12$_{ox}$ N-terminus becomes more compact than in the free protein [33] (Figure 7). The unfolding of the N-terminus domain of the algal CP12$_{ox}$ may also be the consequence of the loss of possible stabilizing contacts between this region and the C-terminal domain that were described by Gardebien *et al* [31], as these contacts reinforce the small hydrophobic core of free CP12$_{ox}$. Once the C-terminal subdomain is bound to GAPDH, it would no longer be available to stabilize the already unstable helical N-terminal domain, which results in its unfolding.

Consequences of GAPDH binding for PRK recruitment

These dynamical residues that surround the N-terminal disulfide bridge are involved in the recruitment of PRK by CP12$_{ox}$ to form the ternary complex PRK/GAPDH/CP12$_{ox}$ [52] but the mechanisms of this recognition remains obscure. This flexibility is certainly essential for the regulation of the interaction. As shown here for *C. reinhardtii* CP12, the conformational sampling of this region differs significantly in the free protein or in the GAPDH/CP12$_{ox}$ subcomplex. In the free protein, it oscillates between a disordered state and a more

compact hairpin fold, whereas in the GAPDH/CP12$_{ox}$ subcomplex, it preferentially adopts extended conformations. Noteworthy, PRK exhibits a higher affinity for the subcomplex GAPDH/CP12$_{ox}$ than for free CP12$_{ox}$ (K_D=60 nM versus 1.3 μM [26]). The extended conformations of CP12$_{ox}$ in the complex may provide a larger capture radius, which would probably help encounters with PRK by widely opening arms to the target. On the other hand, the PRK binding site of CP12$_{ox}$ may be more exposed and accessible when CP12$_{ox}$ adopts extended conformations, than when it is buried in its folded conformation. Interestingly, the two N-terminal cysteine residues of CP12, which allow the partial folding of this region upon disulfide bridge formation, are not conserved in the cyanobacterial proteins. This lack of disulfide does not impair interaction with PRK [32,53], suggesting that formation of these helices are not mandatory for PRK binding and that the extended disordered conformations might be those that preferably interact with PRK. Only higher resolution structural data would help answer these questions and decrypt the mechanisms of recognition of PRK by CP12$_{ox}$, whether already bound to GAPDH or not.

The heterogeneous conformational sampling relates to the moonlighting activities of CP12

CP12 is a jack of all trades and the master of the CBB cycle in photosynthetic organisms [54] and is able to bind to multiple partners, with variable affinities [55]. It has also moonlighting activities that includes metal ion binding [56,57] and chaperone activities [58]. Indeed, CP12 function cannot be reduced to inactivating the CBB cycle during the night [59,60]. The dynamic properties of CP12 elucidated here provide a clue on how CP12 carries these variable functions. These functions are probably regulated via the multiple structural transitions that occur in this conditionally disordered protein. A first transition between CP12$_{red}$ and CP12$_{ox}$ drastically modifies the conformational sampling of the protein. CP12$_{ox}$ also undergoes another drastic structural transition between the free and the GAPDH-bound protein, and thereby partly recovers the disordered properties of the reduced state.

Finally, our results provide a striking example of a recognition process that involves an IDP and that is not strictly conserved evolutionarily, with different structural behavior and affinities in the complex in different species. The example of CP12 illustrates the extraordinary complexity and diversity of IDPs and the difficulty to relate, especially in conditionally disordered proteins, a sequence to predicted structures, dynamics, thermodynamics, and recognition mechanisms, even by comparing a protein from one

species to another. CP12 exemplifies the invalid dichotomy between order and disorder in protein structures, and the wide range of sometimes complex situations that exist in the continuum between these two extreme states.

MATERIAL AND METHODS

Protein production:

The protocol for the expression and purification of $CP12_{ox}$ was described in Graciet et al [12]. The 100% oxidation level of the sample was controlled using SDS-PAGE analysis and thiol titration, using Ellman's reagent [61]. Correct cysteine pairing was further checked by mass spectrometry [28]. Furthermore, the NMR Cβ chemical shift of C66 and C73 confirm the oxidation states of $CP12_{ox}$ during NMR experiments (SI Figure 1).

Nuclear Magnetic Resonance:

NMR experiments were performed on oxidized CP12 at a concentration of 200 µM, in 50 mM phosphate buffer with 50 mM NaCl and 20 mM oxidized DTT, using 600 MHz, 900 MHz and 800 MHz Bruker spectrometers that were equipped with a cryogenic probe. A range of protein concentration (from 200 µM to 1 mM), pH (from 6 to 9), as well as temperature (from 283 to 300K) were tested, and the conditions for which the highest number of distinct resonances was observed was chosen (pH 6.5 and temperature 300K). Sequential assignment of the observed resonances was performed using the standard triple resonance experiments HNCO, HN(Ca)CO, HNCaCb, HN(CO)CaCb [62], and the data were processed using NMRPipe [63] and analyzed using CcpNMR [64] and NMRFAM-Sparky [65]. 2D heteronuclear ^1H-^{15}N HSQC spectra were recorded using the fast-HSQC pulse sequence, with an acquisition time of 121ms in the ^1H dimension and 84ms in the ^{15}N dimension [66].

Dihedral angle restraints were generated from Cα, Cβ, CO, N and HN chemical shifts using TALOS+ [67] on the e-NMR platform [68]. CS-Rosetta [38] was used to generate a model for the L62-Y78 region based on the robust dihedral angle restraints. The SSP software was used on the Cα, Cβ, CO, N and HN to measure the secondary structure propensity [41].

Backbone dynamic experiments were recorded using standard ^{15}N-R_1, ^{15}N-R_2 and {^1H}-^{15}N heteronuclear NOE experiments [69]. A 4 s saturation time was used in the {^1H}-^{15}N heteronuclear NOE experiments. Nine T_2 relaxation delays, ranging from 7.6 ms to 113.7 ms

and including one replicate, were used in the ^{15}N-R$_2$ experiments. Eight T$_1$ relaxation delays, ranging from 20 ms to 900 ms and including one replicate, were used in the ^{15}N-R$_1$ experiments. The data were processed and analyzed using NMRPipe [63].

Translational diffusion was measured using a bipolar gradient stimulated ^{1}H echo experiment [70]. The diffusion delay (Δ) was set to 100 ms, the z-gradient pulse was a squared pulse of 2 ms (total gradient length, δ, of 4ms), using 10 gradients strengths varying from 2% to 98% of the maximal gradient strength. The data were processed in NMRPipe [63] and analyzed in Octave. The diffusion coefficient is determined from the logarithm of the signal intensity that follows the equation:

$$\ln(I) = D \left[(\delta^2 . \gamma_H^2 . (\Delta - \frac{\delta}{6}) \right]$$

where γ_H is the ^{1}H gyromagnetic ratio. The stokes radius is determined from the diffusion coefficient using the Stokes-Einstein equation:

$$r_H = \frac{k_B \, T}{D.6.\eta\pi}$$

^{1}H-^{15}N HSQC spectrum of the GAPDH/CP12ox subcomplex was recorded on a 600 MHz spectrometer, at 283K. An excess of unlabeled GAPDH (50 µM of tetramer) was added to 60 µM of ^{15}N labelled CP12$_{ox}$ in 20mM HEPES pH 6.8, 0.1mM NAD, 1mM Cys. ^{1}H-^{15}N HSQC spectra were recorded using the fast-HSQC pulse sequence, with an acquisition time of 30ms in the ^{1}H dimension and 21ms in the ^{15}N dimension [66]. Dissociation of the complex was achieved by adding 20 mM DTT and 0.02 mM 1,3-bisphophoglycerate to the sample.

Small Angle X-ray scattering:

The same CP12$_{ox}$ sample that was used for NMR experiments was analyzed using SAXS on the SWING beamline at the SOLEIL synchrotron. Scattering vectors (q=4 π/λ sinθ, where 2θ is the scattering angle and λ the wavelength, equal to 1.033 Å) ranging from 0.018 to 0.47 Å^{-1}, were reached with a sample-to-detector (CCD Aviex) distance of 1790 mm. Upstream of the measurement capillary, the sample was submitted to size-exclusion chromatography using a BioSEC-3, 150 Å Agilent HPLC column (SI Figure 6). Prior to the elution of CP12$_{ox}$, 90 frames were collected to measure the blank scattering of the buffer. During the elution of CP12$_{ox}$, 250 frames of 1.5 s with a 0.5 s dead-time between two individual frames, were accumulated. The frames recorded during the second elution peak were carefully compared

with each other, and data corresponding to identical scattering profiles were averaged to increase the signal-to-noise ratio.

Data analysis was performed using the ATSAS suite of software [44]. The Rg and forward scattering intensity I(0) were obtained via PRIMUS using the Guinier approximation, and the distance distribution function P(r) was obtained via GNOM. The molecular mass M of $CP12_{ox}$ was determined using the forward scattering intensity I(0) of the frame corresponding to the top of the peak, at the highest protein concentration, and according to the following equation[42]:

$$I(0) = c \, M \, [(\rho_p - \rho_s) \, v_p]^2 / \mathcal{N} \; ,$$

where c is the protein concentration, ρ_p and ρ_s the scattering length density of the protein and of the buffer respectively, v_p the partial specific volume of the protein and $\mathcal{N}$ Avogadro's constant. Ab initio models of 3D envelopes corresponding to the scattering curves were constructed using DAMMIF [44] and GASBOR [44]. Two pools of 10,000 conformers each were generated using RanCh as detailed in the main text. For each conformer, backbones were reconstituted using PD2 ca2main [71] and side chains were added using scwrl4 [72]. The subensemble of conformers that was compatible with the experimental scattering curve was then selected using GAJOE [45].

SAXS experiments on GAPDH and on the $GAPDH/CP12_{ox}$ subcomplex were performed on the same beamline, with sample-to-detector distances of 1817 mm and 1832 mm respectively, and with acquisition times of 2 s and 1 s and dead-times of 1 s and 0.5 s respectively. The proteins were diluted in 30 mM Tris at pH 7.9, 4 mM EDTA, 0.1 mM NAD with 5 mM free cysteine, with a protein concentration of 23 mg/ml for the GAPDH sample, and 16 mg/ml in total for the complex. One equivalent of $CP12_{ox}$ (monomer) was added to four equivalents of GAPDH (monomer) to form the $GAPDH/CP12_{ox}$ subcomplex. A single elution peak was observed on the HPLC chromatography of GAPDH. Two peaks were observed in the HPLC chromatogram of the $GAPDH/CP12_{ox}$ subcomplex. Only the first and dominant peak, which corresponded to the complex, was used in the analysis; the second peak corresponded to free GAPDH.

The SAXS data analysis of the GAPDH and $GAPDH/CP12_{ox}$ data was performed using the same tools as for the $CP12_{ox}$ data. *C. reinhardtii* tetrameric GAPDH models were built using homology modeling with Modeller on the Chimera interface [73], from the GAPDH crystal structures with PDB IDs: (i) **1NBO**, free GAPDH from spinach, 80% sequence identity

[48]), (ii) **3K2B**, free GAPDH from *A. thaliana*, 79% sequence identity[47] and (iii) **3QV1**, GAPDH from *A. thaliana* in complex with CP12 C-terminus [13]. The loops at the subunits interfaces were refined using Sybil. These atomic models were confronted to the experimental SAXS data using CRYSOL. Pools of conformers of the complex were constructed in RanCh by adding two molecules of $CP12_{ox}$ (each in 10,000 possible conformations) to the GAPDH atomic model, which was built from the structure with PDB ID **3QV1**.

Circular dichroism

Protein $CP12_{ox}$ was diluted in filtered 50 mM Na_2HPO_4 pH7 at final protein concentration of about 15 µM. CD spectra were recorded at 25°C from 260 to 185 nm on a Jasco J-1100 spectropolarimeter using 2 mm thick quartz cells, and were averaged over 8 successive scans Mean residue molar ellipticity (Θ, $deg.cm^2.dmol^{-1}$) was calculated as $\Theta = E \times 10^6/[lc(n\text{-}1)]$, where E is the raw ellipticity (mdeg), l is the cell pathlength (mm), c is the protein concentration (µM), (n-1) is the number of peptide bonds and n the number of residues (n=100 for $CP12_{ox}$ including the His-Tag). The helical content of $CP12_{ox}$ wild type and mutants was estimated by the ratio between the experimental mean residue molar ellipticity at 222 nm (Θ_{222}) and the theoretical Θ_{222} of CP12 with 100% helical conformation, given by the formula $39,500 \times (1\text{-}2.57/n)^{74}$.

Structural data deposition

Chemical shift assignment of oxidized CP12 has been deposited in the Biological Magnetic Resonance Data Bank under accession code **26013**.

Small angle X-ray scattering data have been deposited at SASBDB (https://www.sasbdb.org/) under accession codes SASDCY9, SASDCZ9 and SASDD22 for oxidized CP12, GAPDH, and the subcomplex GAPDH/$CP12_{ox}$ respectively.

Author Contributions

Conceptualization, H.L., V.R.B and B.G.; Methodology, H.L., P.B., V.R.B and B.G.; Investigation, H.L., P.B., C.P.,Y. Z., S.M. and V.R.B; Resources, C.P., Y. Z.; Validation, H.L.,

V.R.B and B.G.; Writing – Original Draft, H.L., V.R.B and B.G.; Writing – Review & Editing, H.L., V.R.B and B.G.; Funding Acquisition, V.R.B and B.G.

REFERENCES

1. Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6,** 197–208

2. Mollica, L. *et al.* (2016). Binding mechanisms of intrinsically disordered proteins: theory, simulation, and experiment. *Front. Mol. Biosci.* **3,** 52

3. Mouillon, J.-M., Gustafsson, P. & Harryson, P. (2006). Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. *Plant Physiol.* **141,** 638–650

4. Findlater, E. E. & Graether, S. P. (2009). NMR assignments of the intrinsically disordered K2 and YSK2 dehydrins. *Biomol. NMR. Assign.* **3,** 273–275

5. Rahman, L. N. *et al.* (2010). Interactions of intrinsically disordered Thellungiella salsuginea dehydrins TsDHN-1 and TsDHN-2 with membranes - synergistic effects of lipid composition and temperature on secondary structure. *Biochem. Cell Biol.* **88,** 791–807

6. Szalainé Ágoston, B., Kovács, D., Tompa, P. & Perczel, A. (2011). Full backbone assignment and dynamics of the intrinsically disordered dehydrin ERD14. *Biomol. NMR Assign.* **5,** 189–193

7. Clarke, M. W. *et al.* (2015). Structural and functional insights into the cryoprotection of membranes by the intrinsically disordered dehydrins. *J. Biol. Chem.* **290,** 26900–26913

8. Atkinson, J., Clarke, M. W., Warnica, J. M., Boddington, K. F. & Graether, S. P. (2016). Structure of an intrinsically disordered stress protein alone and bound to a membrane surface. *Biophys. J.* **111,** 480–491

9. Sun, X. *et al.* (2011). A functionally required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Mol. Biol.* **77,** 205–223

10. Jensen, M. K. *et al.* (2010). The Arabidopsis thaliana NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling. *Biochem. J.* **426,** 183–196

11. Richardson, L. G., Jelokhani-Niaraki, M. & Smith, M. D. (2009). The acidic domains of the Toc159 chloroplast preprotein receptor family are intrinsically disordered protein domains. *BMC Biochemistry* **10,** 35

12. Graciet, E. *et al.* (2003). The small protein CP12: a protein linker for supramolecular complex assembly. *Biochemistry* **42,** 8163–8170

13. Fermani, S. *et al.* (2012). Conformational selection and folding-upon-binding of intrinsically disordered protein CP12 regulate photosynthetic enzymes assembly. *J. Biol. Chem.* **287,** 21372–21383

14. Launay, H. *et al.* (2016). Absence of residual structure in the intrinsically disordered regulatory protein CP12 in its reduced state. *Biochem. Biophys. Res. Commun.* **477,** 20–26

15. Groben, R. *et al.* (2010). Comparative sequence analysis of CP12, a small protein involved in the formation of a Calvin cycle complex in photosynthetic organisms. *Photosynth. Res.* **103,** 183–194

16. Stanley, D. N., Raines, C. A. & Kerfeld, C. A. (2013). Comparative analysis of 126 cyanobacterial genomes reveals evidence of functional diversity among homologs of the redox-regulated CP12 protein. *Plant Physiol.* **161,** 824–835

17. Thompson, L. R. *et al.* (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *PNAS* **108,** E757–E764

18. Scheibe, R., Wedel, N., Vetter, S., Emmerlich, V. & Sauermann, S.-M. (2002). Co-existence of two regulatory NADP-glyceraldehyde 3-P dehydrogenase complexes in higher plant chloroplasts. *Eur. J. Biochem.* **269,** 5617–5624

19. Sparla, F., Pupillo, P. & Trost, P. (2002). The C-terminal extension of glyceraldehyde-3-phosphate dehydrogenase subunit B acts as an autoinhibitory domain regulated by thioredoxins and nicotinamide adenine dinucleotide. *J. Biol. Chem.* **277,** 44946–44952

20. Marri, L. *et al.* (2009). Prompt and easy activation by specific thioredoxins of Calvin cycle enzymes of Arabidopsis thaliana associated in the GAPDH/CP12/PRK supramolecular complex. *Mol. Plant* **2,** 259–269

21. Thieulin-Pardo, G., Avilan, L., Kojadinovic, M. & Gontero, B. (2015). Fairy 'tails': flexibility and function of intrinsically disordered extensions in the photosynthetic world. *Front. Mol. Biosci.* **2,** 23

22. Thieulin-Pardo, G. *et al.* (2016). The intriguing CP12-like tail of adenylate kinase 3 from Chlamydomonas reinhardtii. *FEBS J.* **283,** 3389–3407

23. Howard, T. P., Metodiev, M., Lloyd, J. C. & Raines, C. A. (2008). Thioredoxin-mediated reversible dissociation of a stromal multiprotein complex in response to changes in light availability. *PNAS* **105,** 4056–4061

24. López-Calcagno, P. E., Howard, T. P. & Raines, C. A. (2014). The CP12 protein family: a thioredoxin-mediated metabolic switch? *Front Plant Sci* **5,** 9

25. Kaaki, W., Woudstra, M., Gontero, B. & Halgand, F. (2013). Exploration of CP12 conformational changes and of quaternary structural properties using electrospray ionization traveling wave ion mobility mass spectrometry. *Rapid Commun. Mass Spectrom.* **27,** 179–186

26. Graciet, E. *et al.* (2003). The small protein CP12: a protein linker for supramolecular complex assembly. *Biochemistry* **42,** 8163–8170

27. Marri, L. *et al.* (2008). Spontaneous assembly of photosynthetic supramolecular complexes as mediated by the intrinsically unstructured protein CP12. *J. Biol. Chem.* **283,** 1831–1838

28. Delobel, A. *et al.* (2005). Mass spectrometric analysis of the interactions between CP12, a chloroplast protein, and metal ions: a possible regulatory role within a PRK/GAPDH/CP12 complex. *Rapid Commun. Mass Spectrom.* **19,** 3379–3388

29. Jakob, U., Kriwacki, R. & Uversky, V. N. (2014). Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem. Rev.* **114,** 6779–6805

30. Reichmann, D. & Jakob, U. (2013). The roles of conditional disorder in redox proteins. *Curr. Opin. Struct. Biol.* **23,** 436–442

31. Gardebien, F., Thangudu, R. R., Gontero, B. & Offmann, B. (2006). Construction of a 3D model of CP12, a protein linker. *J. Mol. Graph. Model.* **25,** 186–195

32. Matsumura, H. *et al.* (2011). Structure basis for the regulation of glyceraldehyde-3-phosphate dehydrogenase activity via the intrinsically disordered protein CP12. *Structure* **19,** 1846–1854

33. Del Giudice, A. *et al.* (2015). Unravelling the shape and structural assembly of the photosynthetic GAPDH-CP12-PRK complex from Arabidopsis thaliana by small-angle X-ray scattering analysis. *Acta Crystallogr. D Biol. Crystallogr.* **71,** 2372–2385

34. Mileo, E. *et al.* (2013). Dynamics of the intrinsically disordered protein CP12 in its association with GAPDH in the green alga Chlamydomonas reinhardtii: a fuzzy complex. *Mol. BioSyst.* **9,** 2869–2876

35. Sormanni, P. *et al.* (2017). Simultaneous quantification of protein order and disorder. *Nat. Chem. Biol.* **13,** 339–342

36. Bonomi, M., Heller, G. T., Camilloni, C. & Vendruscolo, M. (2017). Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* **42,** 106–116

37. Bhowmick, A. *et al.* (2016). Finding our way in the dark proteome. *J. Am. Chem. Soc.* **138,** 9730

38. Shen, Y. *et al.* (2010). De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein Sci.* **19,** 349–356

39. Ortega, a, Amorós, D. & García de la Torre, J. (2011). Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **101,** 892–8

40. Peng, J. W. & Wagner, G. (1994). Investigation of protein motions via relaxation measurements. *Meth. Enzymol.* **239,** 563–596

41. Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. (2006). Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci.* **15,** 2795–2804

42. Receveur-Brechot, V. & Durand, D. (2012). How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* **13,** 55–75

43. Receveur, V., Czjzek, M., Schülein, M., Panine, P. & Henrissat, B. (2002). Dimension, Shape, and Conformational Flexibility of a Two Domain Fungal Cellulase in Solution Probed by Small Angle X-ray Scattering. *J. Biol. Chem.* **277,** 40887–40892

44. Petoukhov, M. V. *et al.* (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350

45. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2,** 207–217
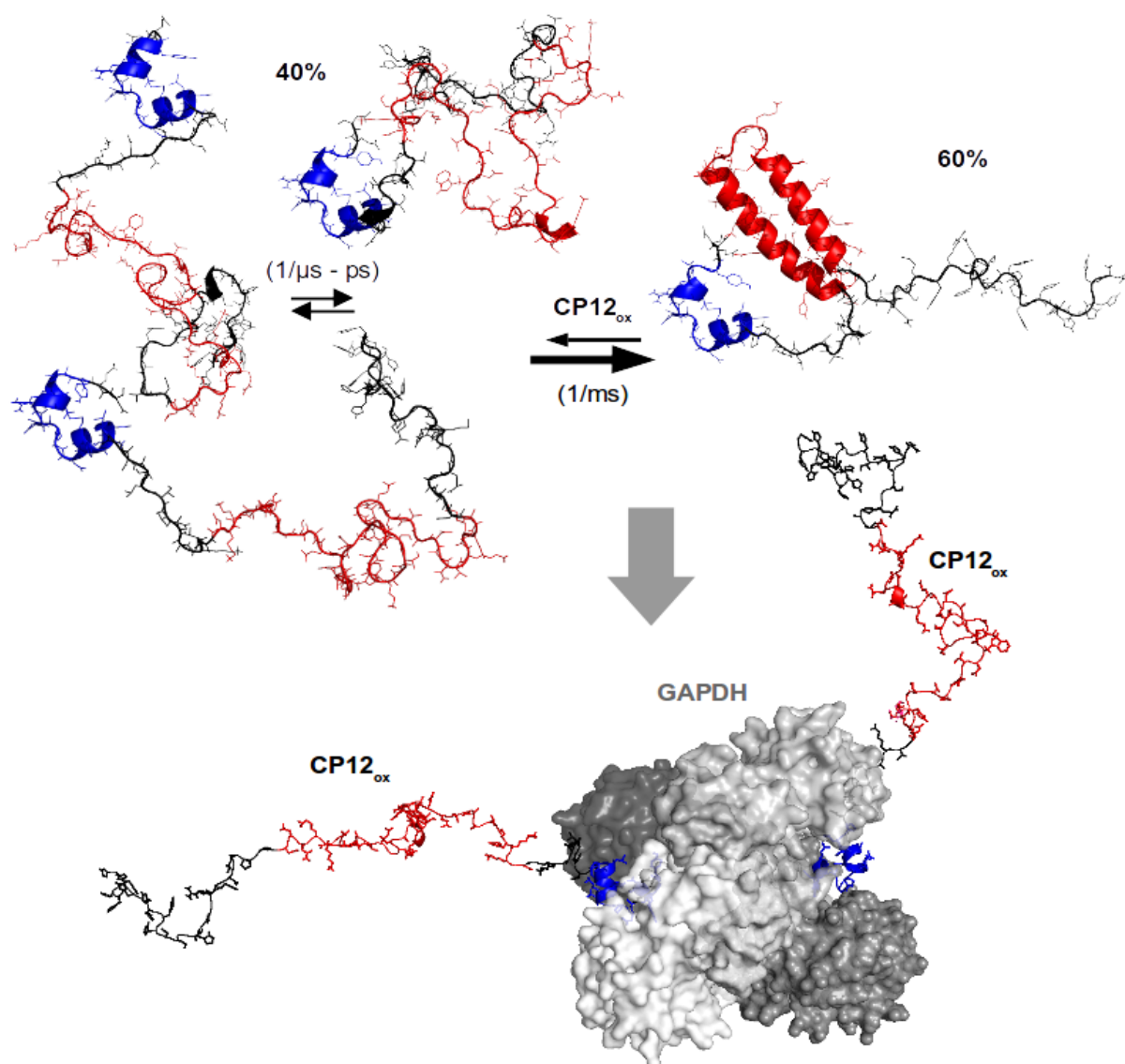
46. Fischer, H., De Oliveira Neto, M., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. (2010). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Cryst.* **43,** 101–109

47. Fermani, S. *et al.* (2010). Structure of photosynthetic glyceraldehyde-3-phosphate dehydrogenase (isoform A4) from Arabidopsis thaliana in complex with NAD. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66,** 621–626

48. Falini, G. *et al.* (2003). Dual coenzyme specificity of photosynthetic glyceraldehyde-3-phosphate dehydrogenase interpreted by the crystal structure of A4 isoform complexed with NAD. *Biochemistry* **42,** 4631–4639

49. Flock, T., Weatheritt, R. J., Latysheva, N. S. & Babu, M. M. (2014). Controlling entropy to tune the functions of intrinsically disordered regions. *Curr. Opin. Struct. Biol.* **26,** 62–72

50. Lebreton, S., Graciet, E. & Gontero, B. (2003). Modulation, via protein-protein interactions, of glyceraldehyde-3-phosphate dehydrogenase activity through redox phosphoribulokinase regulation. *J. Biol. Chem.* **278,** 12078–12084

51. Sparla, F. *et al.* (2005). Regulation of Photosynthetic GAPDH Dissected by Mutants. *Plant Physiol* **138,** 2210–2219

52. Avilan, L. *et al.* (2012). CP12 residues involved in the formation and regulation of the glyceraldehyde-3-phosphate dehydrogenase-CP12-phosphoribulokinase complex in Chlamydomonas reinhardtii. *Mol Biosyst* **8,** 2994–3002

53. Tamoi, M., Miyazaki, T., Fukamizo, T. & Shigeoka, S. (2005). The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions. *Plant J.* **42,** 504–513

54. Gontero, B. & Maberly, S. C. (2012). An intrinsically disordered protein, CP12: jack of all trades and master of the Calvin cycle. *Biochem. Soc. Trans.* **40,** 995–999

55. Erales, J., Avilan, L., Lebreton, S. & Gontero, B. (2008). Exploring CP12 binding proteins revealed aldolase as a new partner for the phosphoribulokinase/glyceraldehyde 3-phosphate

dehydrogenase/CP12 complex – purification and kinetic characterization of this enzyme from Chlamydomonas reinhardtii. *FEBS J.* **275,** 1248–1259

56. Erales, J., Gontero, B., Whitelegge, J. & Halgand, F. (2009). Mapping of a copper-binding site on the small CP12 chloroplastic protein of Chlamydomonas reinhardtii using top-down mass spectrometry and site-directed mutagenesis. *Biochem. J.* **419,** 75–82, 4 p following 82

57. Rocha, A. G. & Vothknecht, U. C. (2013). Identification of CP12 as a novel calcium-binding protein in chloroplasts. *Plants (Basel)* **2,** 530–540

58. Erales, J., Lignon, S. & Gontero, B. (2009). CP12 from Chlamydomonas reinhardtii, a permanent specific 'chaperone-like' protein of glyceraldehyde-3-phosphate dehydrogenase. *J. Biol. Chem.* **284,** 12735–12744

59. Howard, T. P. *et al.* (2011). Antisense suppression of the small chloroplast protein CP12 in tobacco alters carbon partitioning and severely restricts growth. *Plant Physiol.* **157,** 620–631

60. Elena López-Calcagno, P., Omar Abuzaid, A., Lawson, T. & Anne Raines, C. (2017). Arabidopsis CP12 mutants have reduced levels of phosphoribulokinase and impaired function of the Calvin–Benson cycle. *J Exp Bot* **68,** 2285–2298

61. Ellman, G. L. (1959). Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **82,** 70–77

62. Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. B* **103,** 203–216

63. Delaglio, F. *et al.* (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–93

64. Vranken, W. F. *et al.* (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59,** 687–696

65. Lee, W., Tonelli, M. & Markley, J. L. (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31,** 1325–1327

66. Mori, S., Abeygunawardana, C., Johnson, M. O. & van Zijl, P. C. (1995). Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. B* **108,** 94–98

67. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–223

68. Bonvin, A. M. J. J., Rosato, A. & Wassenaar, T. A. (2010). The eNMR platform for structural biology. *J. Struct. Funct. Genomics* **11,** 1–8

69. Skelton, N. J. *et al.* (1993). Practical aspects of two-dimensional proton-detected 15N spin relaxation measurements. *J. Magn. Res. B* **102,** 253–264

70. Stejskal, E. O. & Tanner, J. E. (1965). Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics* **42,** 288

71. Moore, B. L., Kelley, L. A., Barber, J., Murray, J. W. & MacDonald, J. T. (2013). High-quality protein backbone reconstruction from alpha carbons using Gaussian mixture models. *J. Comput. Chem.* **34,** 1881–1889

72. Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77,** 778–795

73. Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815

74. Chen, Y. H., Yang, J. T. & Chau, K. H. (1974). Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13,** 3350–3359
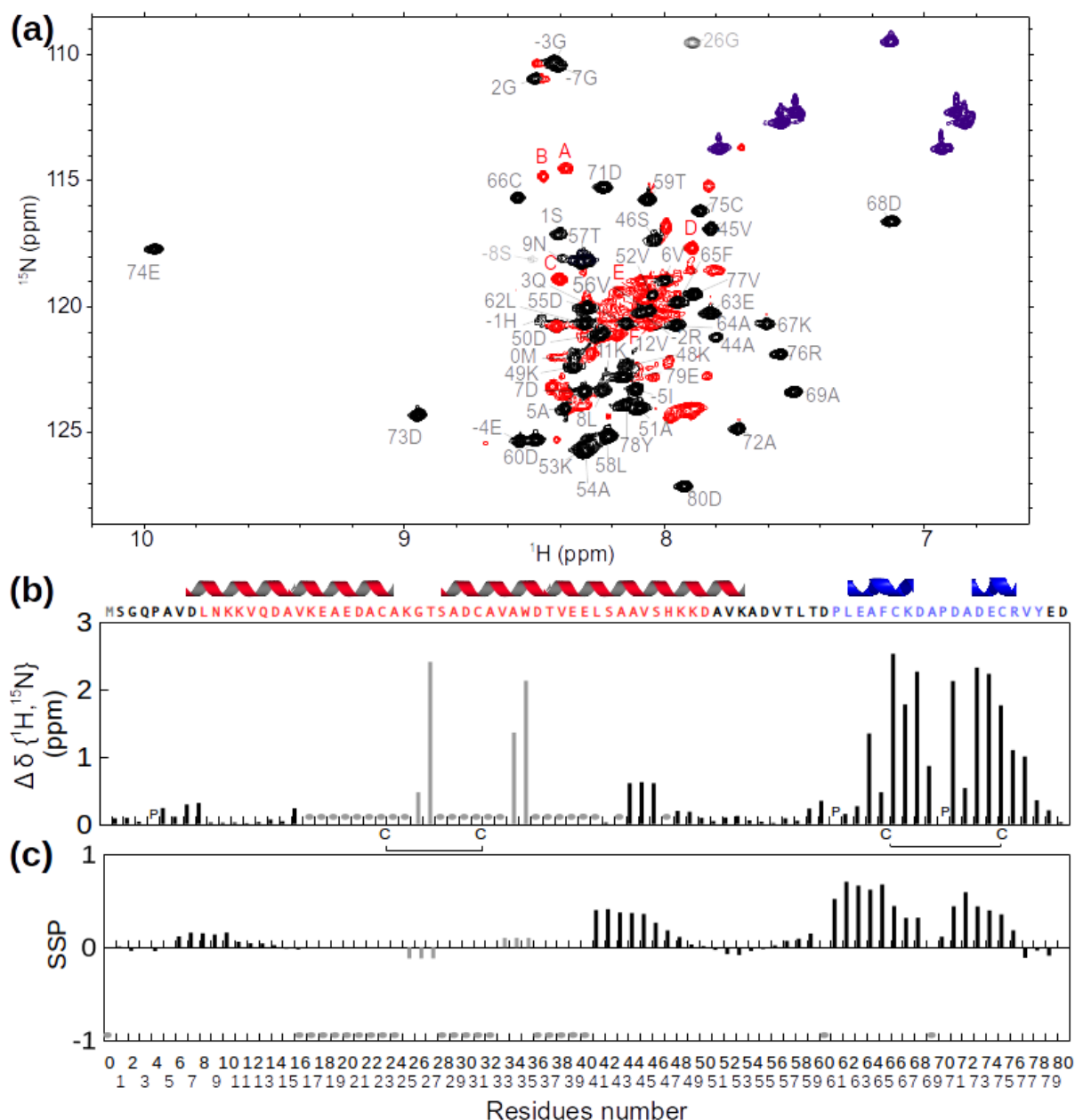
FIGURES

Abstract

Figure 1. NMR analysis of CP12ox.compared to CP12red (a) ^1H-^{15}N HSQC spectrum of CP12ox. Unassigned resonances are shown in red. Six of them which had sufficient intensity to allow relaxation analysis are labelled with a letter. Ambiguous assignments are in light grey. Purple resonances are Arg, Glu and Asn side-chains resonances. **(b)** Difference in chemical shift between the CP12ox and CP12red resonances. The differences are expressed as $\sqrt{(0.2(\delta^{15}N_{red}-\delta^{15}N_{ox})^2+(\delta^1H_{red}-\delta^1H_{ox})^2)}$. Residues that do not have an assigned NMR resonance are indicated with grey dots. Ambiguously assigned residues are in grey. **(c)** Secondary structure propensity calculated using amine proton and nitrogen, Cα and Cβ chemical shift [41] of CP12ox resonances. Positive values indicate helical propensity and negative values suggest strand propensity. Above the graphs is the sequence of CP12. Black residues have NMR parameters typical of disordered segments. Blue residues have NMR parameters of a folded structure. Red residues have broadened resonances or are unassigned. Above the sequence, the N-terminal helices predicted in the model of Gardebien and al [31] are shown in red, and the two helices we solved in the C-terminal turn are shown in blue.
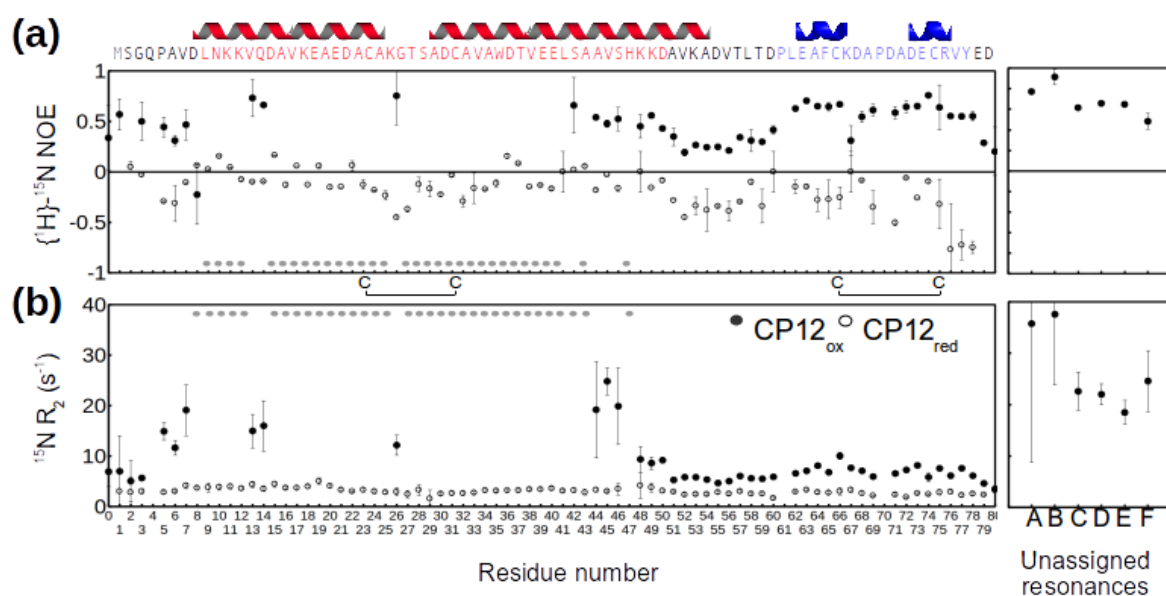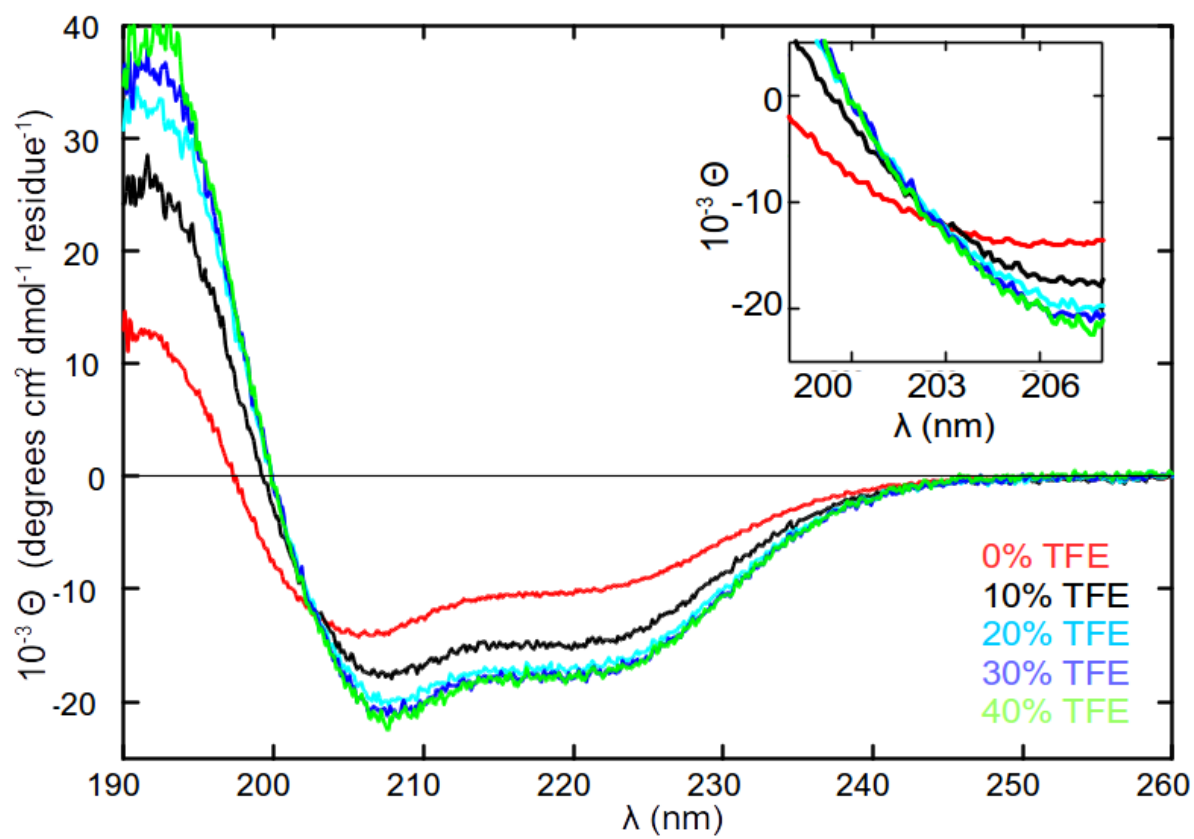
Figure 2. Backbone dynamics of CP12$_{red}$ and CP12$_{ox}$. (a) {^{1}H}-^{15}N het-NOE and **(b)** ^{15}N-R$_2$ measured for resonances of CP12$_{red}$ (opened circles) and CP12$_{ox}$ (filled circles). Residues that do not have an assigned NMR resonance in CP12$_{ox}$ are indicated with grey dots. The right panels show the values obtained for six unassigned resonances for CP12$_{ox}$ labelled in Figure 1. Top of the graph: sequence of CP12 with the same color coding as in Figure 1. Above the sequence, the N-terminal helices predicted in the model of Gardebien and al [31] are shown in red, and the two helices we solved in the C-terminal turn are shown in blue.

Figure 3. CD spectra of CP12$_{ox}$ in the presence of increasing concentration of TFE. The isodichroic point at 203 nm is shown in the inset.
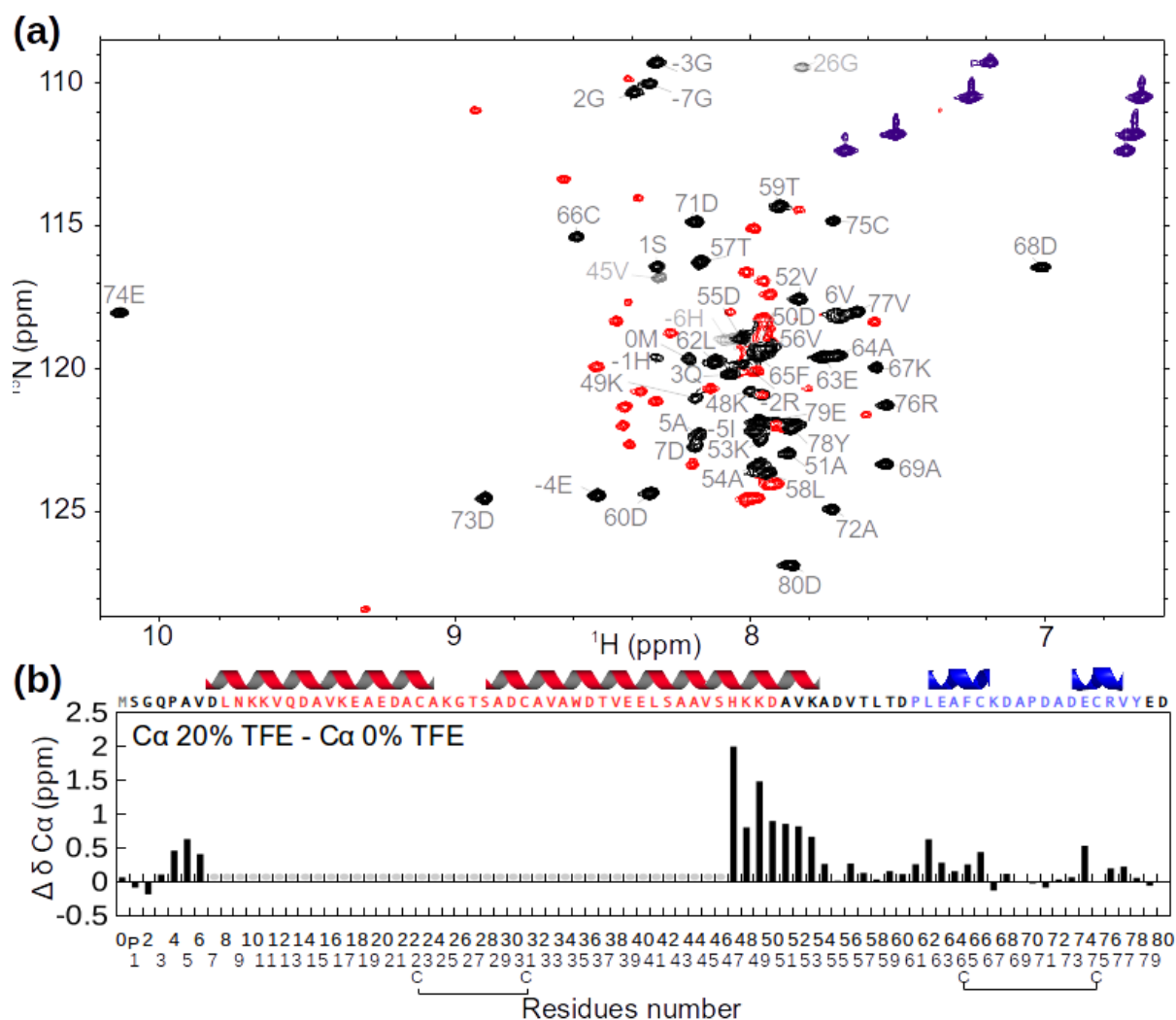
(a)

(b)

Figure 4: Displacement of the conformational equilibrium of CP12$_{ox}$ in the presence of TFE. (a) ^{1}H-^{15}N HSQC spectrum of CP12$_{ox}$ in 20% TFE. Unassigned resonances are shown in red. Ambiguous assignments are in light grey. Purple resonances are Arg, Glu and Asn side-chains resonances.**(b)** Differences of Cα chemical shift values between 20% and 0% TFE. Top of the graph: sequence of CP12 with the same color coding as in Figure 1. Above the sequence, the N-terminal helices predicted in the model of Gardebien and al [31] are shown in red, and the two helices we solved in the C-terminal turn are shown in blue.
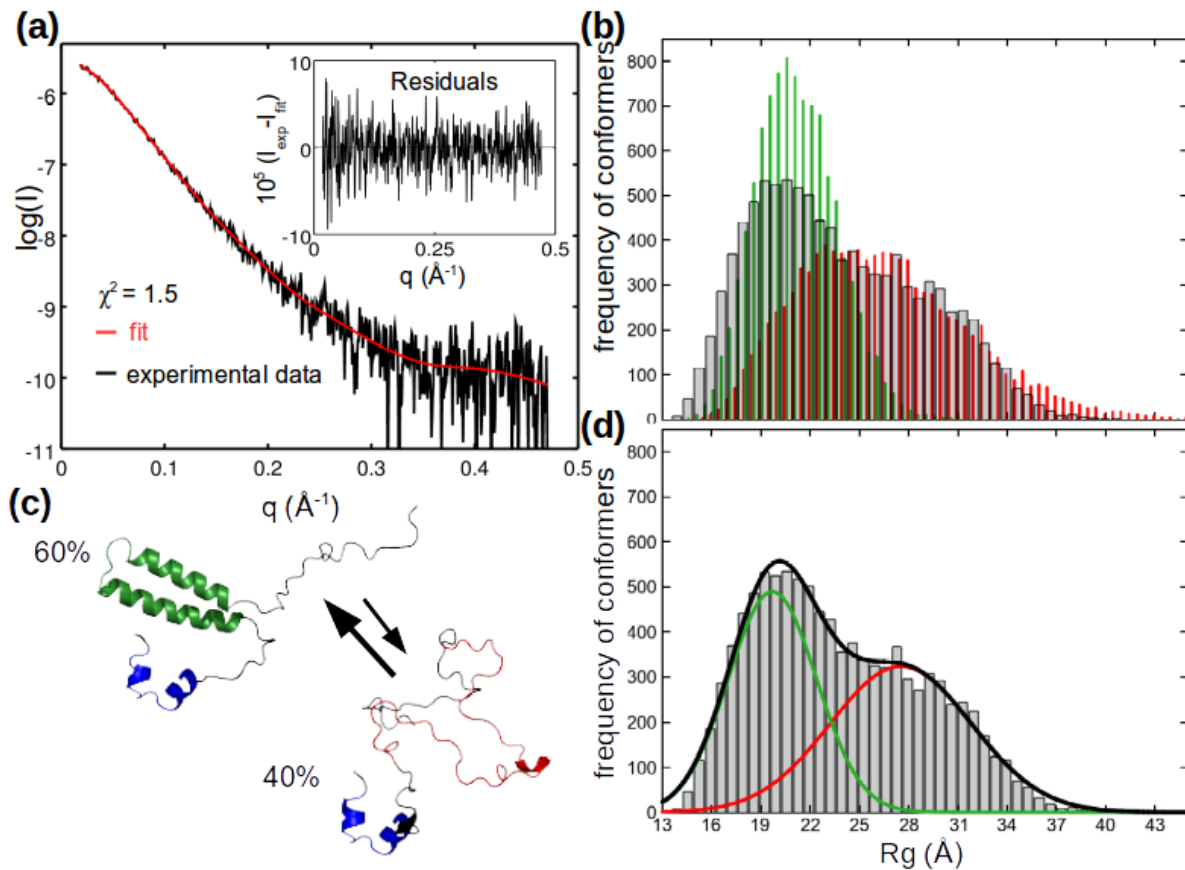
Figure 5. Selection of conformer ensembles compatible with the SAXS data. (a) Experimental SAXS curve (black) and calculated scattering curve (red) of the selected conformers. The residuals are displayed in the inset. **(b)** Distribution of Rg of conformers. The distribution of Rg of all the models generated with the N-terminal region in a random coil state is in red. The distribution of Rg of all the models generated with the N-terminal region in hairpin of two helices is in green. The grey bars are the Rg distribution of the selected conformers using GAJOE [44] when the red and green ensembles are set as input together. **(d)** That latter Rg distribution is fit by the sum of two-Gaussian curves using a least-squares fitting (black line). The fitting results in a first Gaussian (green) that is centered at Rg°=°19.7°±°0.3°Å, which is very close to the mean Rg for the CP12$_{ox}$ models with an N-terminal helix (Rg = 21 Å). The second Gaussian (red) is centered at Rg°=°27°±°0.3°Å, which corresponds to the mean Rg for the CP12$_{ox}$ models with a random-coil N-terminal (Rg = 26.7 Å). **(c)** Two typical models from each population schematizing the conformational exchange between the two states used in the analysis. The relative populations are determined from the deconvolution that is shown in **(d)**. The color coding of the residues is as follow: the blue residues correspond to the C-terminal domain folded in the structure we solved by NMR, the black residues are disordered ; the green residues correspond to the N-terminal region that is folded as in the molecular model of Gardebien et al [31] ; the red residues correspond to this same region in the disordered conformation.
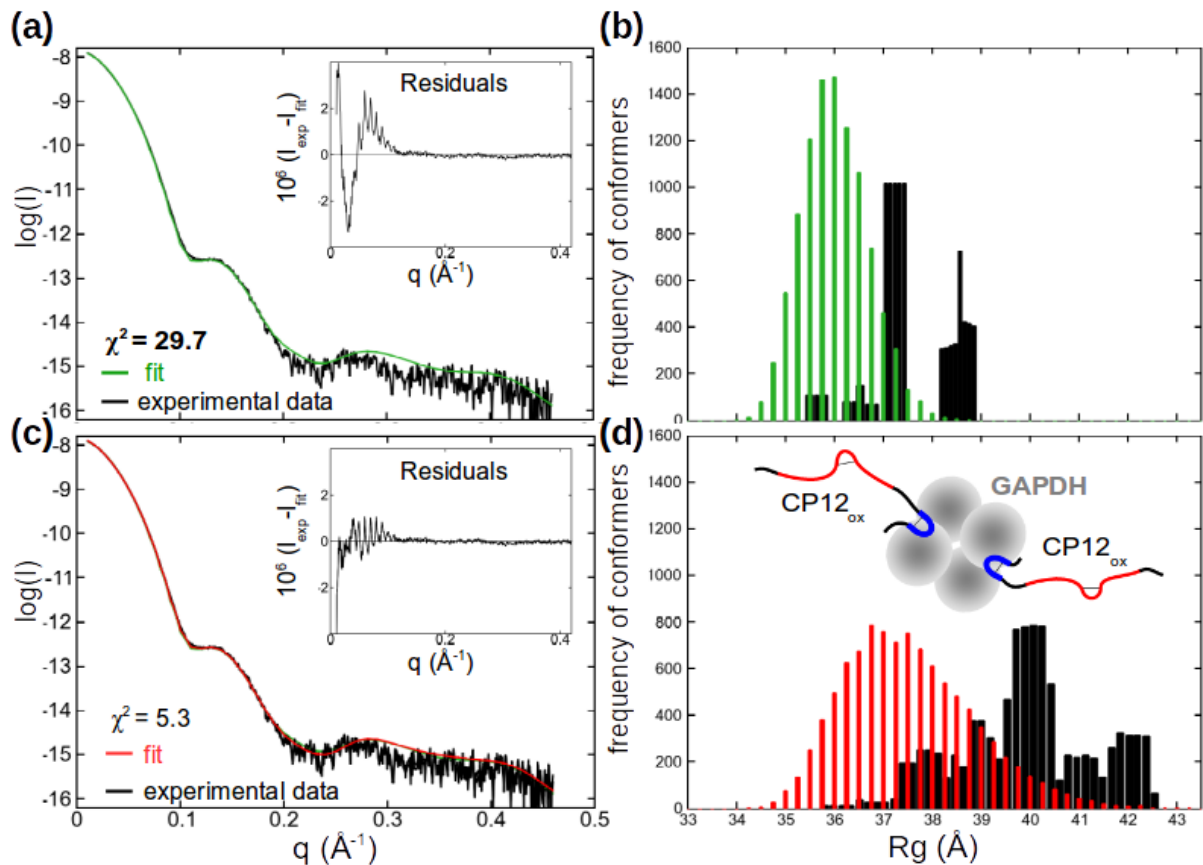
Figure 6. Selected ensemble of conformers for the GAPDH/CP12$_{ox}$ subcomplex. Two series of models for the N-terminal end of CP12 in the GAPDH/CP12$_{ox}$ subcomplex were tested separately. The green color corresponds to the ensemble of models generated with the helical N-terminal region **(a & b).** The red color corresponds to the ensemble generated with the disordered N-terminal region (d & e). **(a & c)** Experimental SAXS curve (black) and fit using GAJOE [44] (green and red repectively). Residuals are displayed as inserts. The model with two N-terminal helices (green) does not fit the experimental data as highlighted by the high χ^2. **(c & d)** Distribution of Rg of the selected conformers using GAJOE [44] (black) and of the starting pools of conformers (green and red). The distributions of the selected conformers are displaced towards high Rg compared to the starting pools. **(d)** In the distribution that fits the data, i.e. selected from the input ensemble with the N-terminal region disordered (red) the schematic of the GAPDH/CP12$_{ox}$ subcomplex is shown. The color coding of CP12 is the same as the sequence in Figure 1.
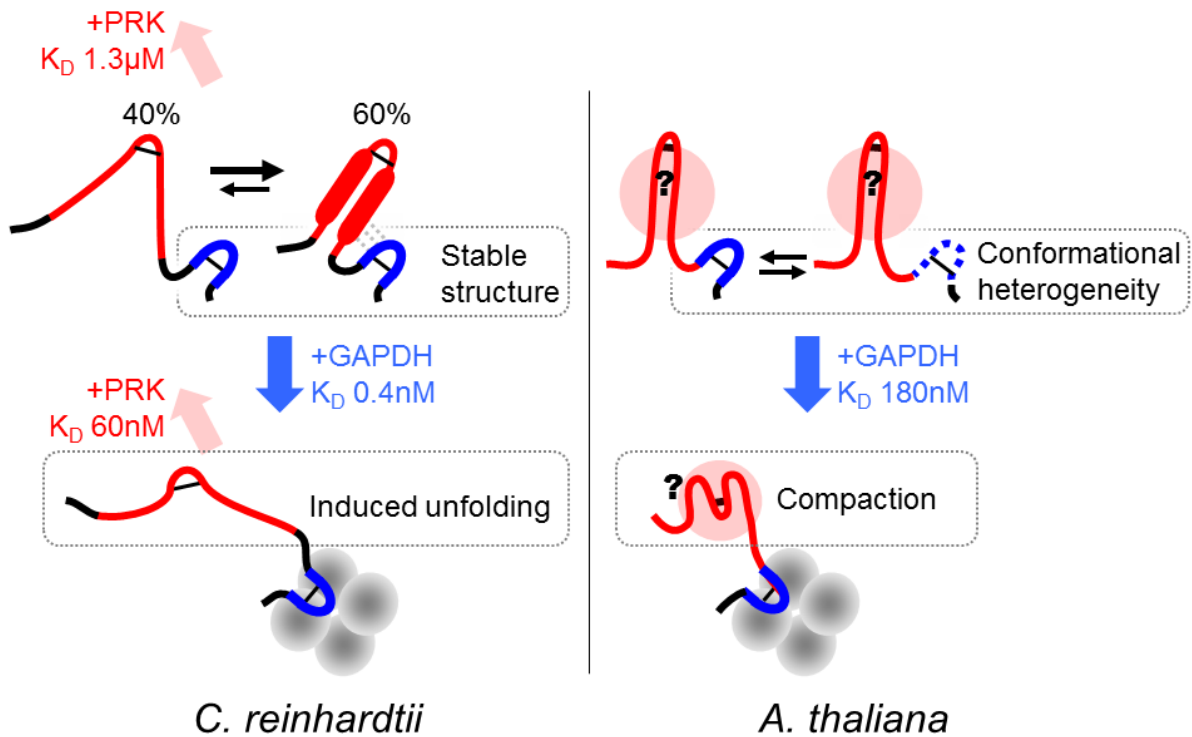
Figure 7. Schematic representation of the structural behavior involved in the CP12$_{ox}$ binding to GAPDH. This scheme highlights the differences between CP12$_{ox}$ of the two organisms *C. reinhardtii* and *A. thaliana.* At the C-terminal end of CP12$_{ox}$, a conformational selection governs GAPDH binding in *A. thaliana*, contrary to the case in *C. reinhardtii* .The N-terminal moiety of CP12$_{ox}$ undergoes an induced unfolding upon complex formation in *C. reinhardtii*, contrary to the *A. thaliana* protein. These two phenomena bring out the role of a lower entropy cost in the interaction that leads to a much higher affinity in the case of the *C. reinhardtii* interaction.