Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2006) 186–195
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

# Construction of a 3D model of CP12, a protein linker

Fabrice Gardebien [a], Rajesh R. Thangudu [a], Brigitte Gontero [b], Bernard Offmann [a,*]

[a] *Laboratoire de Biochimie et Génétique Moléculaire, Université de La Réunion, 15 Avenue René Cassin, 97715 Saint-Denis Messag Cedex 09, La Réunion, France*
[b] *Laboratoire de Génétique et Membranes, Institut Jacques Monod, UMR 7592 CNRS, Universités Paris VI–VII, 2 Place Jussieu, 75 251 Paris Cedex 05, France*

Received 19 April 2005; received in revised form 10 November 2005; accepted 14 December 2005
Available online 19 January 2006

**Abstract**

The chloroplast protein CP12 is known to play a leading role in a complex formation with the enzymes GAPDH and PRK. As a preliminary step towards the understanding of the complex formation mechanism and the exact role of this protein linker, a comparative modelling of the CP12 protein of the green alga *Chlamydomonas reinhardtii* was performed. Because of the very few structural information and poor template similarities, the derivation of the model consisted in an iterative trial-and-error procedure using the comparative modelling program MODELLER, the following three structure validation programs PROCHECK, PROSA, and WHATIF, and molecular mechanics energy refinement of the model using the program CHARMM. The analysis of the final model reveals a scaffold of key residues that is believed to be essential in the folding mechanism and that coincides with the residues conserved throughout the CP12 family. Our results suggest that this protein is a typical disordered protein. Finally, the various mechanisms by which the CP12 protein can self-interact or binds to other enzymes are discussed in light of its modelled structure and characteristics.


*Keywords:* CP12; GAPDH; Comparative modelling; Multienzyme complex assembly; Unstructured proteins; Protein linker; Calvin cycle

## 1. Introduction

CP12 is a small nuclear-encoded 8.5 kDa chloroplast protein. This protein is also found across photosynthetic prokaryotic systems in cyanobacteria. For the green alga *Chlamydomonas reinhardtii*, it was shown that chloroplastic CP12 in its oxidized form acts as a protein linker between glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and phosphoribulokinase (PRK), thus providing evidence for a novel way for the light regulation of photosynthetic enzymes in the Calvin cycle pathway [1,2]. A model for the binding pathway between CP12 and its target enzymes has been experimentally derived [1,3]. Nevertheless, the structural basis for the mechanism underlying this interaction has yet to be demonstrated though low resolution of the whole complex was studied using cryoelectron microscopy [4]. The presence of CP12 in its oxidized form appears to be crucial in the complex formation. It should furthermore be pointed out that the CP12 sequence of this particular alga is very similar to the sequences reported for many prokaryotic and eukaryotic species.

To better grasp the role of CP12 in the complexation mechanism between GAPDH, PRK, and CP12, knowledge of the tertiary structure of this protein is essential. No experimental structure of any CP12 homologues has been reported so far and NOE experiments conducted for this protein in *C. reinhardtii* remain difficult to interpret [1]. However, experimental results clearly indicate the presence of two disulphide bridges when CP12 is in its oxidized state, one between the two N-ter cysteines and one between the two C-ter cysteines. These cysteines are highly conserved among prokaryotic and eukaryotic CP12 accessions. As reported previously, the formation of these disulphide bridges upon oxidation plays a key role in the folding of CP12 [1]. Information available from NMR and circular dichroism showed that oxidized CP12 is mainly composed of α-helices and coil segments and is very flexible, while reduced CP12 is mainly unstructured [1].

In this study, our aim was to derive a model of the tertiary structure of the oxidized form of the CP12 protein found in *C. reinhardtii* by using a comparative modelling approach. A drawback arises, however, from the fact that no homologues

* Corresponding author. Tel: +262 262 93 86 41; fax: +262 262 93 82 37.
*E-mail address:* Bernard.Offmann@univ-reunion.fr (B. Offmann).

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.12.003

was found for this 80 residues long mature peptide despite extensive investigation. To overcome this problem, our strategy was to consider an extensive iterative trial-and-error procedure that combines the following three steps: generation of models by comparative modelling, validation of a model by ad hoc structure validation programs, and energy refinement using molecular mechanics calculations.

The paper is organized as follows. In Section 2, all the details about the procedure we followed to derive the models are given. In Section 3, we first present the results of the comparative modelling calculation (Sections 3.1 and 3.2). Then, we discuss the mechanisms by which CP12 self-interacts (complexation can be observed) or binds to the enzymes GAPDH and PRK in light of its structure and characteristics (Section 3.3). Finally, a comparative discussion between the tertiary structure here reported and other disordered proteins is proposed (Section 3.4).

## 2. Materials and methods

### 2.1. Selection of templates and input preparation

Prediction of the secondary structure for the 80 residues long mature peptide of CP12 from *C. reinhardtii* was performed through the JPRED server (http://www.compbio.dundee.ac.uk/www-jpred) [5]. The results indicate two long helices between residues 8 and 23 and between 29 and 51; a shorter helix is also predicted between residues 61 and 65. Besides, as derived experimentally [2], all thiol groups from the four cysteines are expected to be in the oxidized form, thereby resulting in two S–S bonds between CYS-23 and CYS-31 and between CYS-66 and CYS-75. Strong structural constraints therefore arise from the above expected features. Besides the N- and C-terminal sequence structures, three coiled regions, i.e. two cysteine–cysteine loops and the coil between residues 52 and 60, determine the relative orientation between the predicted helical regions.

Because of the absence of detectable homologues of CP12 in structural databases using classical local similarity search algorithm and of very low similarity with other known protein structures, all the input information about CP12 structure were derived from both experimental and prediction data. Hence, two kinds of restraints were used in our comparative modelling approach. First, we imposed two disulphide bridges between cysteine residues 23 and 31 and 66 and 75, respectively, to comply with experimental data. Second, on the basis of experimental observation from circular dichroism, partial NOE [1] and theoretical predictions (see Section 2), two helices were considered between residues 8 and 23 and 29 and 51, respectively.

The comparative modelling program MODELLER (version 7v7) [6] in conjunction with the following validation programs PROCHECK [7], PROSA [8], and WHATIF [9] were used throughout this study to derive a model for the CP12 protein. The challenge we had to face was to derive a structure for this protein that does not have any known homologues in the PDB. Only protein structures sharing overall low similarities ($< 25\%$) were used in our modelling approach as templates. Our basic idea was to select templates based on the following criteria: (i) the sequence must share sufficient similarity (at least 15% with respect to the 80 residue-long sequence of CP12); (ii) once the CP12 and template respective fragment sequences aligned, the secondary structure of the template must match the predicted secondary structure of the target sequence (if not, the alignment is reconsidered); (iii) if present in the template, the cysteine residues must be aligned or very close to those in the target sequence; (iv) the residues aligned with the bridged cysteines must have their $C^{\alpha}$ atoms distant by at most $\sim 7$ Å. A template was rejected whenever one of the above criteria was not fulfilled.

At first, several approaches were considered to search for templates sharing global similarities with the CP12 sequences. They were obtained from the Modeller (version 7v7) internal structure database [6], by enquiring DSDBase for similar disulphide connectivity as described by Thangudu et al. [10] and by threading CP12 sequence onto known structures by using the 3D-PSSM server [11]. Despite the various databases used, very few good candidates were found. We finally selected one template structure from the 3D-PSSM server based on its threading scores and the fulfilment of the above-described requirements: the portion of sequence of this selected template shares 15% global similarity with CP12, a good agreement is found between the observed and predicted secondary structures of the template and CP12, and finally, a correct alignment between respective disulphide bridges. This template was chosen as first template in the final alignment (Fig. 1) and defines a general framework for the CP12 fold. Then, in order to bring more quality and precision in the models derived from the comparative modelling program, we enriched the template pool by choosing short template sequences that share high similarity with the CP12 sequence: the target sequence was split into three parts (from N-ter to residue 31, from residue 32 to residue 75, and from residue 66 to C-ter, respectively) and template–target alignments was performed for each target subunit sequence. Although a large number of potential templates could be obtained from this approach, only a very few of them satisfied the above-described requirements on the structures. Nineteen templates were finally selected in the comparative modelling calculations. Appendix A presents the fragments used in the modelling approach along with the final model for CP12.

A further step consisted in performing a multiple alignment of all the templates and target sequences using the program Clustalw with the PAM substitution matrix. The alignment was then manually modified such as to contain the main features of all the individual template–target alignment. The alignment obtained served as a basis but was further constantly manually modified throughout the modelling process (vide infra).

### 2.2. Modelling strategy

Several sets of models were generated by using the program MODELLER. In this program, the models are generated by satisfaction of spatial restraints. The restraints include distance and dihedral angles for the backbone and sidechains that are

```
>1KW2-F1  PSGTFEQVSQLVKEVVSLTEACCAEGADPDCYDTRTSALSAKSCESNSPFP-----VHPGTAECCTKEGLERKLC-----
>1KW2-F2  -PTADLEDVLPLAEDITNILSKCCESASEDCMAK-------------------------------------------------
>1QHG     SEQLLAHLNKEQQEAVR-------------------------------------------------------------------
>1APY     ------------KNATEAAWRALASGGSALDAV---------------------------------------------------
>1D2D     -----MVYDKIAAQGEVVRKLKAEKAPKAKVTEAVECLLSLKAEYKEKT-----------------------------------
>1D8H     ---PDDDLTKSVQDWV-------------------------------------------------------------------
>1PZQ     AASPAVDIGDRLDELEKALEALSAED---------------------------------------------------------
>1A1T     ------------------------------------------------------------------CGKEGHIAKNCRA---
>1XYF     ---------------------------------------------------------------------------DAGELRVYGD
>1A3Q     ---------------------------------------------------------------------------DDIEVRFYED
>1BG6     ------------------------------------------------AVKDADVIL-----------------------------
>1DX5     ------------------------------------------------------------CHNLPGTFECICGPD
>1R12     ------------------------------------------------------------CSDNPNARECR----
>1N7O     ---------------------------------TYKDISLMQSLLSD-----------------------------
>1CU1     ------------------------------------------------AVQNEVTLTHPITKY---------------
>1MXM     ---------------------------------LIIARMISNAVNRLMIS-----------------------------
>1HCX     ------------------------------------------------KPDGTLAD--------------------
>1CMX     ------------------------------------------------KSDPTATD--------------------
>1K6D     ------------------------------------------------LIALAADITLVEP-------------------
>cp12     SGQPAVDLNKKVQDAVKEAEDACAKGTSADCAVAWDTVEELSAAVSHKKDAVKADVTLTDPLEAFCKDAPDADECRVYED
```

Fig. 1. Final alignment that enables the derivation of the model for CP12. Nineteen template fragments were used in our approach with a minimal number of gaps. The identification code is taken from the RCSB Protein Data Bank; the codes 1KW2-F1 and 1KW2-F2 refer to two fragments used for the protein of PDB identification code 1KW2.

calculated from the alignment with the templates, bonds and angles values taken from the CHARMM-22 force field, internal statistical preferences for dihedral angles and non-bonded interatomic distances, as well as user-defined restraints (for example on the secondary structure packing). These restraints are expressed in terms of probability density functions, which are combined into an objective function. The optimization of the latter function allows the generation of several models. For deriving our models, in addition to the restraints generated from an initial alignment, two other types of restraints were used: the first to maintain the residue segment 8–23 and 29–51 as helical and the second to define the disulphide bond between residues 23 and 31 and 66 and 75, respectively. Several series of comparative modelling calculation were considered for many input alignments. Twenty models were generated in each series and only a few representative models were then selected on the basis of the lowest MODELLER objective function. Among these models, only those that showed satisfactory PROSA and WHATIF profiles are considered in a subsequent minimization step. The minimization is performed using the CHARMM force field using the corresponding program within the InsightII modelling environment [12]. The minimized structures were further evaluated using PROSA and WHATIF and, in case of unsatisfactory results, the whole modelling procedure is restarted by generating a new set of models with a modified alignment and/or modified extent of secondary structure.

To check the stability of the final CP12 model, we performed molecular dynamics (MD) simulation of the hydrated protein at ambient temperature: the protein was immersed into a spherical droplet of 3162 water molecules. To minimize boundary effects, stochastic boundary conditions were considered by using the Langevin dynamics for the outer water shell (this method consists in splitting the spherical droplet into an inner core treated in the NVE ensemble and an outer region where the newtonian equations are augmented by a friction and a random forces to mimic the collisions with solvent molecules). The radius of the water droplet chosen allows the complete solvation of the protein. A time step of 2 fs was used (all bonds involving hydrogen atoms were constrained by the SHAKE algorithm [13]). For the non-bonded interactions, the Lennard–Jones potential was driven smoothly to zero for distances between 8 and 10 Å.

## 3. Results and Discussion

### 3.1. Comparative modelling

The modelling approach can be summarized as follows. From an initial alignment, it consists in four steps: (i) the derivation of sets of models by MODELLER, (ii) the validation tests (PROSA and WHATIF) made for a few models that are chosen from their relative objective function values, (iii) an energy minimization run is subsequently considered for the structures that have passed these tests, and (iv) a final validation step (as in (ii)). The alignment was manually modified and process restarted from step (i) as long as all the quality requirements for the minimized structures were not fulfilled.

From the above methodology, it follows that the final alignment, presented in Fig. 1, is the result of many intermediate revised versions. The model that satisfied all the validation criteria as assessed by the following programs PROCHECK, PROSA, and WHATIF is presented in Fig. 2. A good overall stereochemistry is obtained for the model with 95.9% of the residue $\Psi/\Phi$ angles falling in the most favoured regions and 4.1% in the allowed regions. The Ramachandran plot is shown in Fig. 3. The interaction energy per residue was also calculated by the program PROSA. Fig. 4 displays the PROSA profiles calculated for the CP12 model along with

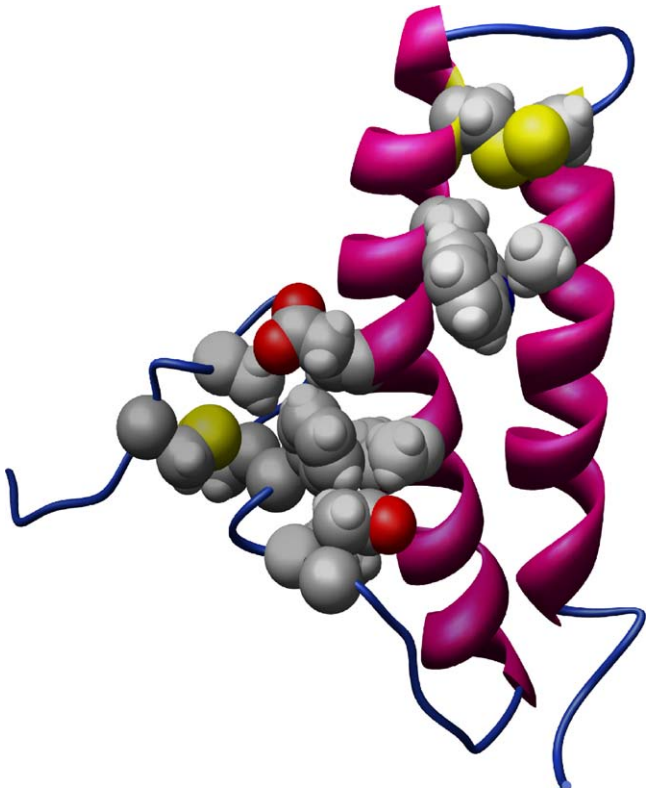

Fig. 2. Model derived for the CP12 protein. The two helices are displayed in magenta. Sulfur, oxygen, carbon, and hydrogen atoms are displayed in yellow, red, grey, and white, respectively. The respective atomic environment of TRP-35 and PHE-65 is highlighted by the use of spherical representation, see discussion in Section 3.2.

those for the two fragments of one of the main template 1KW2-F1 and 1KW2-F2 (two fragments used of the human Vitamin D-binding protein, 1KW2). Caution should be exercised in our case when interpreting the potentials of mean force as

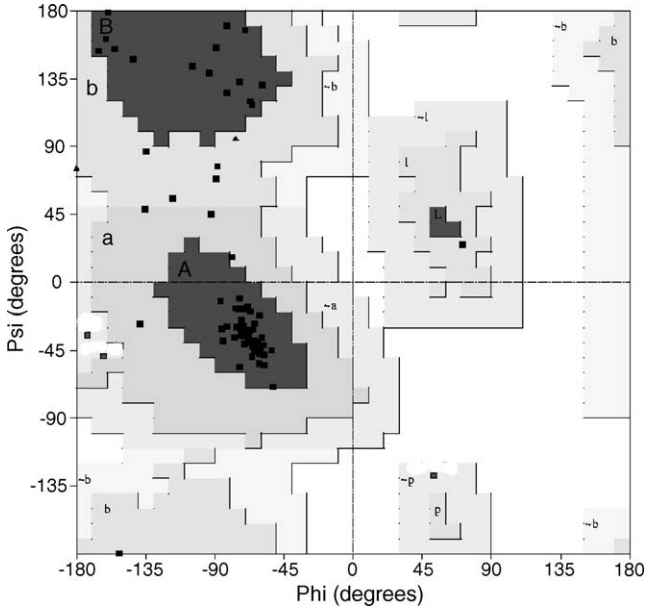


Fig. 3. Ramachandran plot of the $\Psi$/ $\Phi$ distribution of the CP12 model as obtained by PROCHECK: 95.9% residues are in most favoured regions and 4.1% are in additional allowed regions.

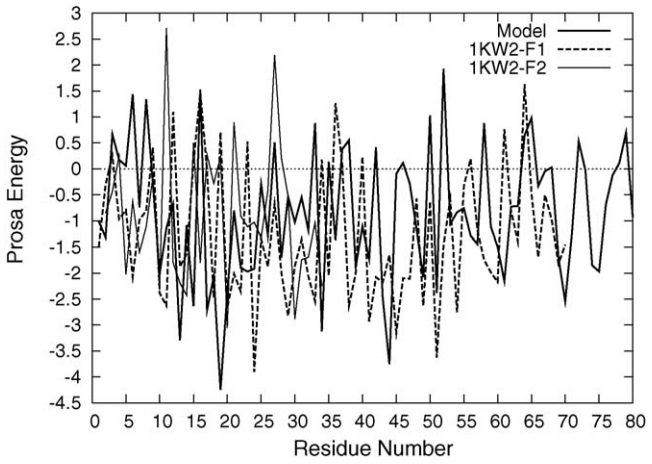


Fig. 4. PROSA energy plot for the CP12 protein model. For comparison, the PROSA energy of two fragments of the main template (labelled 1KW2-F1 and 1KW2-F2) is also shown.

calculated for the residues by PROSA. Indeed, as they are based on statistics obtained for large globular proteins, they are not appropriately fitted for small disulphide-rich proteins. Thus, the profile derived for the model should not be evaluated on the basis of absolute energy criteria but rather using energy criteria relative to characterized disulphide-rich proteins as for the protein 1KW2 (PDB identification code). Although it is difficult to compare individual residue energy between the model and the two fragments of the disulphide-rich template 1KW2 (as the structure of the model is actually derived using other template fragments), it appears that similar overall average quality is obtained for the model and this template. A final test is the packing quality of each residue as assessed by the WHATIF program. Fig. 5 presents the profile obtained with respect to the residues. As the evaluation criterion corresponds in this case to a threshold of −5, one can see that all residues show satisfactory packing values except for three residues for which the score is slightly below the threshold: −5.5 for LYS-25 (in the first CYS–CYS loop), and −5.2 and −5.8 for TYR-78 and GLU-79, respectively, in the C-terminal region. It should be noted that in solution the last two to three residues are expected

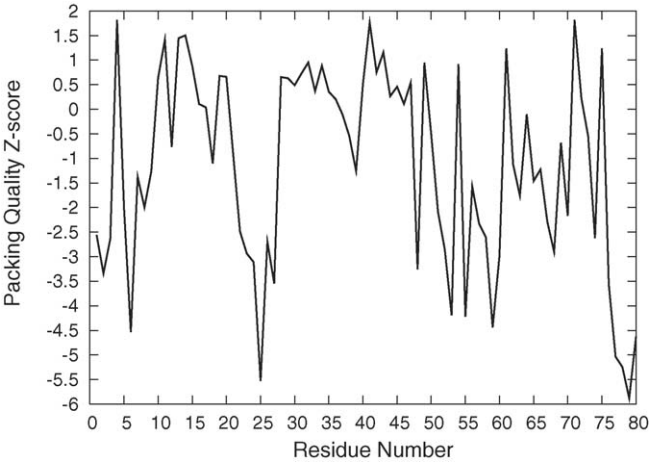


Fig. 5. WHATIF quality control values calculated for the CP12 protein model.

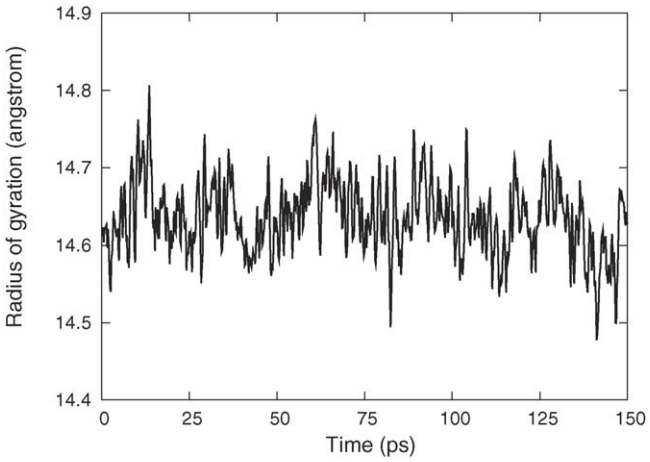


Fig. 6. Radius of gyration of the CP12 model during 150 ps of hydrated MD simulation.

to be rather mobile due to: (i) polar or charged character of these residues (favourable interactions with water solvent molecules) and (ii) the well-known chain-end effect in polymeric systems (highest mobility associated to the terminal parts of a polymer chain).

As a further means of validating the model, MD calculations were performed using explicit water molecules that surround the protein. Such simulations provide a test of the stability of the structure under ambient temperature in water. Indeed, the native fold of the protein will remain unchanged while a non-native structure will denaturate. It can be seen from Fig. 6 that the radius of gyration fluctuates around a mean value during 150 ps of production step. As mentioned above, the two chain ends that consist in polar residues are exposed to the solvent and proved to be rather mobile. However, the mean radius of gyration is constant and this result supports the stability of the model.

### 3.2. Analysis of the model

A multiple alignment was performed for 13 CP12 protein sequences of both eukaryotic and prokaryotic species using the program Clustalw [14]. Further analysis of the alignment using the ConSeq server [15] shows a scaffold of residues that are expected to be essential for the function of the protein and/or in the folding mechanism and the overall stability of the system.

From this multiple sequence alignment (data not shown), we defined as ‘conserved’ in the text and figures all residues in the *C. reinhardtii* sequence that are strictly conserved and those that retain their physico-chemical properties.

Fig. 7 displays the contact map that derives from the interatomic distance calculation for the CP12 model as obtained from the WHATIF server (http://www.swift.cmbi.kun.nl/WIWWWI). A contact is displayed in the map whenever at least one interatomic distance between two residues is within 1 Å range based on their overlapping Van der Waals radii.

In the left-hand column and lowest header row is displayed the sequence of the protein. The conserved residues among both prokaryotes and eukaryotes are highlighted in yellow and in green for those that are conserved in plants only. Contacts are represented according to the following color code: (i) in orange between two conserved residues, (ii) in yellow between pairs that comprise at least one conserved residue, (iii) in green, contacts involving at least one conserved residue among eukaryotic accessions and (iv) in black those involving non-conserved residues.

Note that the conserved residues are found only in the two helices and in the C-terminal part (beyond PHE-65). It is striking that off-diagonal pair interactions (i.e. for amino-acids separated by at least three residues) always involve at least one residue (yellow and orange squares) that is conserved with only few exceptions that mainly concern the interactions between the N-terminal part (sequence GQPA) and residues 53 up to 56 in the loop region. The number of pair interactions involving two conserved residues (orange squares) amounts to 22 for a total of 24 conserved residues. Four of these residues in the C-terminal part (PRO-70, GLU-74, TYR-78 and GLU-79) are not making contacts with any of the other conserved residues. Thus, the number of average contacts established per conserved residues is then about 1.

In the following, the discussion will refer to interactions between conserved residues only (orange squares). In the first helix, two interactions are observed between the four conserved residues; the latter residues are also responsible for six interactions with four conserved residues in the second helix. Moreover, in the second helix, seven pairs of conserved residues are interacting with each other and two conserved residues are also making three contacts with two conserved residues of the C-terminal region. Finally, in the latter region, nine conserved residues are found for a total of four contacts.

Among the above-discussed contacts, two striking examples of pair residue interactions concern the residue TRP-35 on the one hand, and the residue PHE-65 on the other hand. In Fig. 2, the local environment of these two residues is emphasized by the use of spherical representation of the respective neighbouring residue atoms. TRP-35 shows a strong interaction with ALA-19 (owing to a hyperconjugation-type interaction). Hence, besides the CYS-23–CYS-31 bridge, this interaction appears to be another major linkage that holds the two helices close in space. These two closely packed antiparallel α-helices actually correspond to the αα motif (supersecondary structure). In the vicinity of the other main aromatic residue PHE-65, there is a ‘hydrophobic’ environment as can be clearly seen from Fig. 2. Two strong interactions (of hyperconjugation-type) are worth noting between PHE-65 and the conserved residues ALA-43 and ALA-44, respectively (the latter being conserved only in plants).

We have also performed solvent accessibility calculation on the model. Interestingly, it can be noted from Fig. 8 that lower solvent accessibility is associated with higher residue conservation score among aligned CP12 accessions. The buried residues (LEU-8, ALA-15, CYS-31, SER-42, ALA-43, ALA-44, SER-46, HIS-47, ALA-69, ALA-72, CYS-75) are highly conserved among CP12 accessions. These are mainly non-polar residues which, as expectedly, specially in the case of small polypeptide system, would tend to avoid solvent exposition by establishing non-polar residue–residue contacts. This is in fact confirmed upon observation of the contact map of

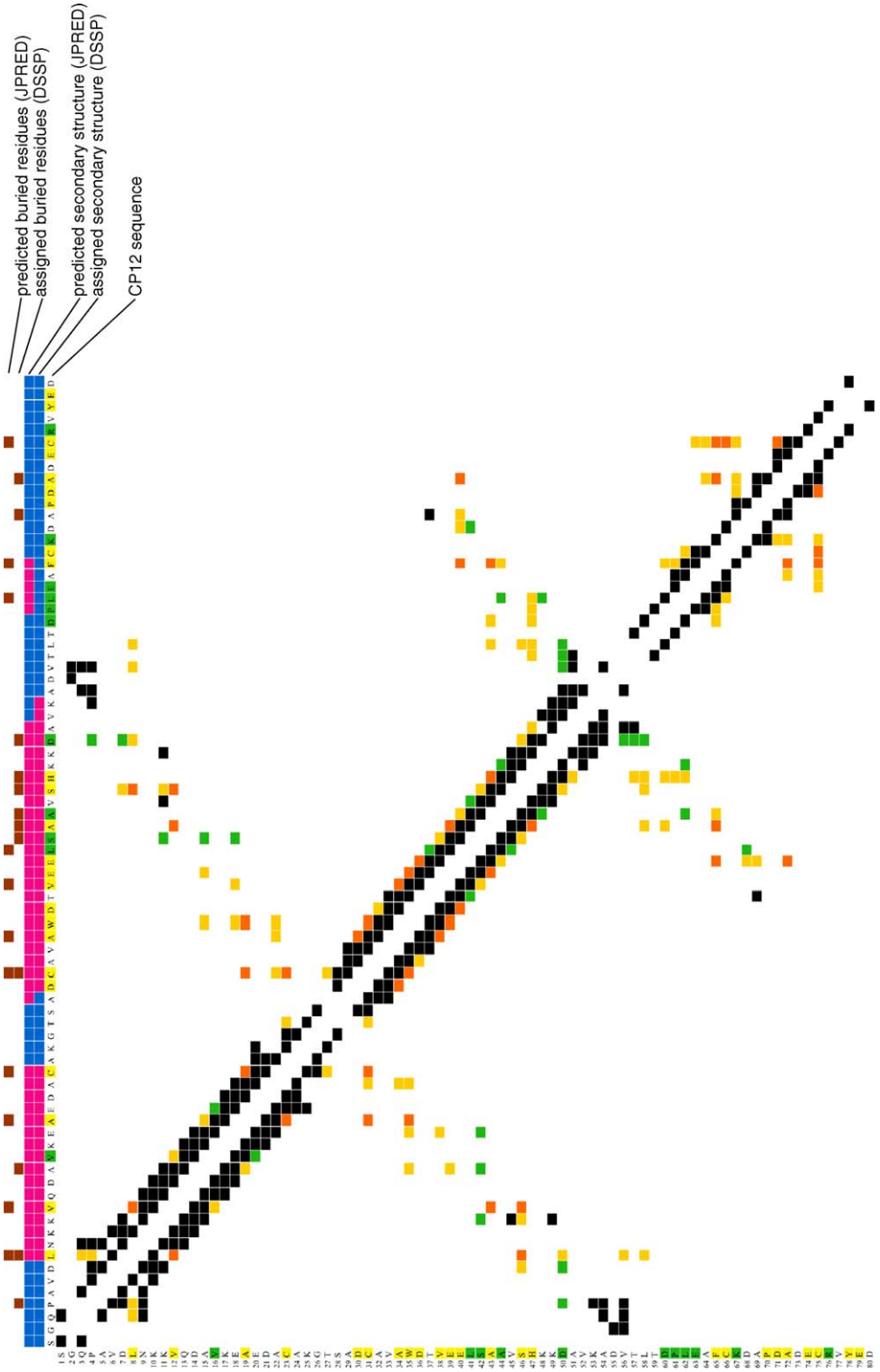


Fig. 7. Contact map calculated for the CP12 model as obtained from the WHATIF server. A contact is defined as two atoms for which the distance between the Van der Waals surfaces is $<$ 1.0 Å (Van der Waals radii for the atoms C, N, O, and S are 1.8, 1.7, 1.4, and 2.0 Å, respectively). Contacts are represented as follows: (i) in orange between two conserved residues, (ii) in yellow between pairs with at least one conserved residue, (iii) in green, those with at least one conserved residue among eukaryotic accessions and (iv) in black, those involving two non-conserved residues.

the model (Fig. 8). The off-diagonal residue–residue contacts are mainly established between non-polar conserved residues. Folding of CP12 would hence be driven by (i) the formation of the two non-interlocked disulphide bridges (CYS-23–CYS-31 and CYS-66–CYS-75), (ii) the hydrophobic contacts between conserved non-polar residues resulting in their partial shielding from solvent and (iii) the two strongly interacting helices (TRP-35–ALA-19).

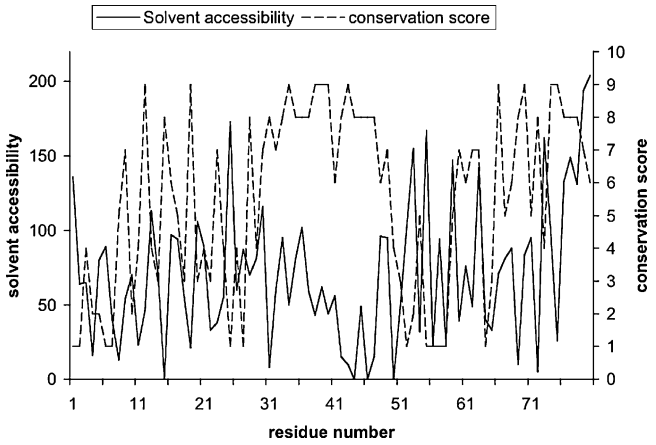


Fig. 8. Analysis of the buried residues in modelled *C. reinhardtii* CP12 and residue conservation among 90 aligned eukaryotic and prokaryotic CP12 accessions. Solvent accessibility was calculated using DSSP program [28]. Residue conservation scores (1, low conservation; 9, high conservation) were calculated from ConSurf server (http://www.consurf.tau.ac.il) [15].

From this analysis and under the hypothesis that the interactions between the conserved residues indeed contribute in driving the amino-acid polymer towards this fold, then one may expect that the other 12 members of the family mentioned present the same folding pattern. This will be the subject of future investigation.

CP12 acts as a protein linker and it was previously shown that its presence was necessary for the formation of a complex between PRK and GAPDH [1]. On the other hand, one may expect from any kind of linker an ambivalent nature, that is, it must show characteristics that are compatible with either of the parties. Hence, in some respects, CP12 should possess this dual and complementary nature. In Fig. 9, the surface of the CP12 model has been represented with negative and positive charges displayed in red and blue, respectively; the light blue represents the non-charged polar atoms while the remaining apolar atoms are shown in grey. Fig. 9 a, c, and b correspond to the front views and side view of the model, respectively, with in Fig. 9a the N-terminal part in the bottom-right, the first CYS–CYS bridge on the top and the C-terminal part on the left.

First, it is noteworthy that the protein is rather flat as seen in Fig. 9 b. Second, the two surfaces of each face exhibit different distribution of charges. A large grey surface beside light blue colored areas predominates in the helices part of the protein (Fig. 9 a) while a large deep blue (positively charged) area along with smaller grey portion is seen on the other side (Fig. 9 c). On both sides of the two helices, one can also note that the proportion of red area (negatively charged) is rather similar but yet with different spatial distribution (vide infra). In view of the dissimilar electrostatic features of the two faces (mostly polar and apolar surface versus highly charged surface), we conjecture that both faces are playing part in the binding process between the two enzymes PRK and GAPDH. In the following, we name these two faces CP12*hp* and CP12*pn*: *hp* stands for hydrophobic and polar character and *pn* for positively and negatively charged character.

### 3.3. Preliminary insights into the biology of CP12

CP12 in the oxidized form was shown to be involved in the regulation of the Calvin cycle by modulating the activity of both photosynthetic GAPDH and PRK in the green unicellular alga *C. reinhardtii*[1,2]. The enzymes are in fact inactivated when they form a stable complex with oxidized CP12. Activation of the enzymes is triggered upon thioredoxin light-mediated reduction of both disulphide bridges in CP12 whereby the complex dissociates [16]. The stoechiometry of the complex in *C. reinhardtii* is two homotetramers $A_4$ of GAPDH, two dimers of PRK and two monomers of CP12 and is obtained through the following assembly pathway as described previously [1]: because GAPDH has higher affinity to CP12 than PRK, a homotetramer of the former enzyme first associates with CP12 to form a very stable complex as shown experimentally. This association induces GAPDH tetramer conformation to change thus acquiring higher affinity for PRK. Subsequently, a dimer of PRK is able to bind to the GAPDH–CP12 complex. The final complex is obtained by dimerization whereby interactions involve CP12–CP12 contacts and additional enzyme–enzyme interactions.

It has been argued that the CP12 binds to GAPDH and PRK through its C- and N-terminus cysteine–cysteine corresponding loops [2]. Our model shows that C-terminus of CP12 is stuck to the second helix owing to interactions with PHE-65 as described above. This C-terminus, which is predicted to be disordered (data not shown), displays a long coil region from

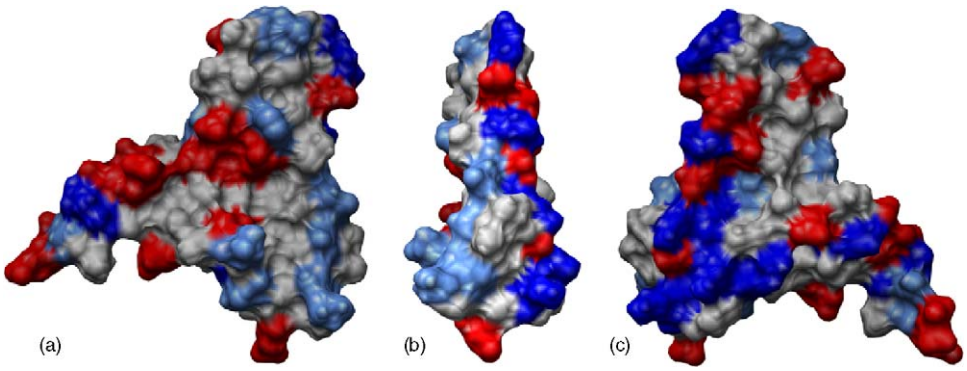


Fig. 9. Surface representation of the CP12 model: two front views (a) and (c) and one side view (b). Apolar, polar, positively and negatively charged atoms are shown in grey, light blue, deep blue, and red, respectively.

ASP-55 to ASP-80. It is noteworthy that a small region spanning from LEU-62 to CYS-66 displays a helix-like conformation similar to molten-globules which coincides with the short helix predicted by JPRED. The flexible property of this region and its structural characteristic tend to support the hypothesis that the C-terminus of CP12 plays a major role in binding to GAPDH but the interaction would also involve other parts of the molecule. The constraints imposed on the backbone in this region by the disulphide bridge between CYS-66 and CYS-75 may be determinant for the binding properties of CP12 with respect to GAPDH (see Wedel et al. [2]). Indeed, these authors claimed, based on experimental evidences, that the interaction between CP12 and GAPDH occurs at the C-terminal loop of CP12. The absence of disulphide bridge in this region would in fact result in a totally unconstrained structure, which can severely affect the binding properties of the protein [1].

Similary, binding of PRK to CP12 has been suggested to occur in the loop delimited between CYS-23 and CYS-31 [2]. Absence of this disulphide has been shown to prevent binding of PRK to CP12. Here, it can be argued from our model that this loop may not directly bind to PRK. In fact, absence of disulphide constraint in this region would trigger the two helices apart and probably unfold this whole region. This may explain the absence of binding of CP12 in the corresponding CYS-SER mutants studied by Wedel et al. [2]. This situation is quite different in the C-terminus disulphide connected loop where the whole region remains flexible and could be considered independently from the rest of the molecule.

As described above, the two faces in the CP12 model termed CP12*hp* and CP12*pn* exhibit distinct electrostatic properties. On the CP12*pn* face, a particular distribution of charges is observed along the first helix where four positive charges alternate with three negative charges (Fig. 9 c). As quoted by Wedel et al. [2], dimerization of two CP12 molecules might be achieved via electrostatic interactions between a cluster of positively and negatively charged residues, able to form salt bridges, when a second N-terminus is placed in an antiparallel position, hence leading to a four-helix bundle-like structure. We are currently investigating the dimerization of CP12 using molecular simulations in order to test this hypothesis.

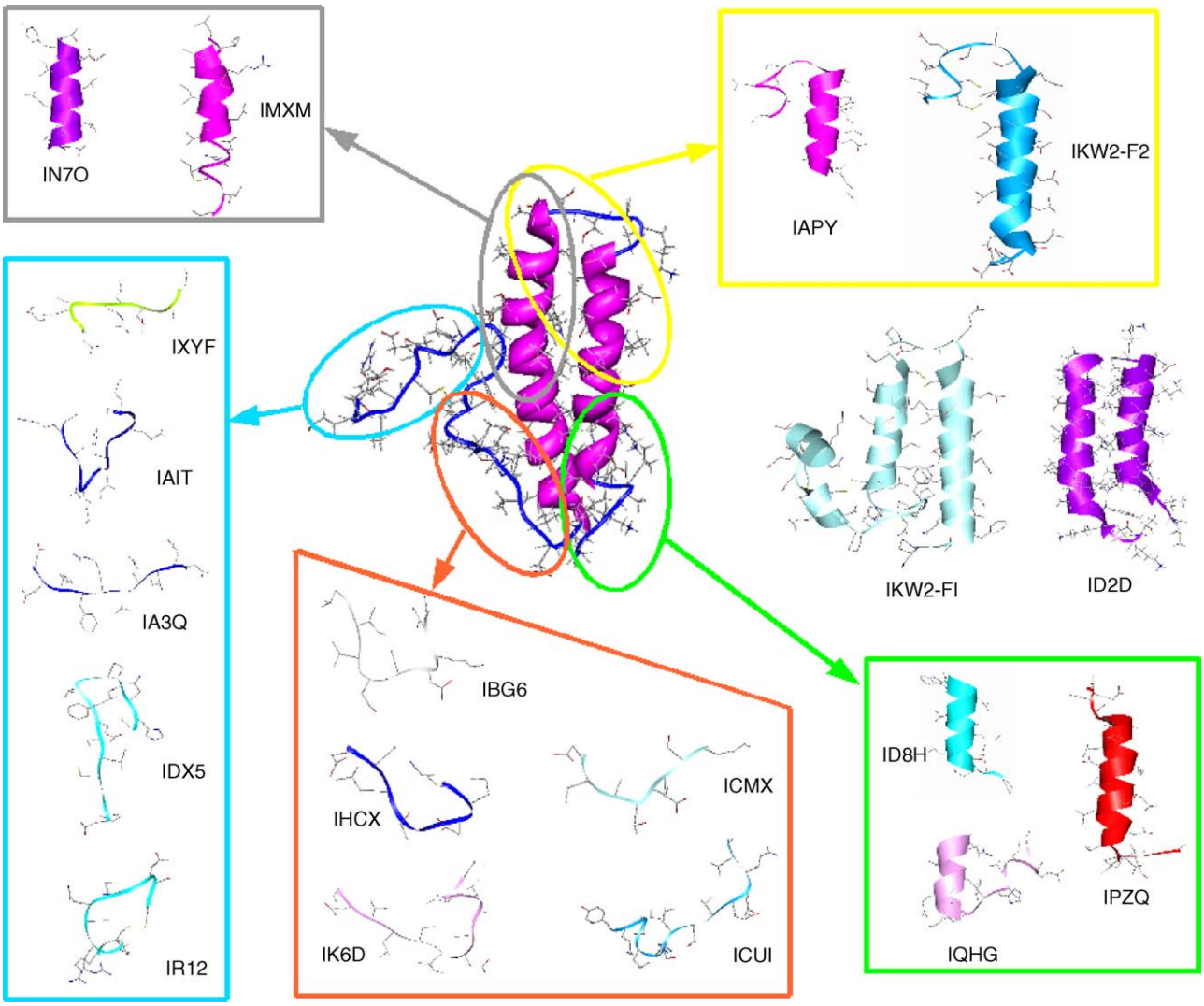


Fig. 10. Template fragments used for the construction of CP12 model as indicated in the alignment in Fig. 1(see text). These are represented in an exploded view starting from the centre of the figure where the backbone of the final CP12 model is illustrated. The identification codes are taken from the RCSB Protein Data Bank.

Besides, CP12 displays on its CP12*hp* face a large hydrophobic surface area and very localized surface patches with negative charges due to groups of carboxylate ions. Because such a surface tends to favour hydrophobic interactions, CP12 *pn* face may be involved either in the dimerization of CP12 through the pairing of two similar faces or in the binding with hydrophobic surfaces of target enzymes. On the other hand, the localized negative patches may also play a key role in binding positive entities namely metal ions. Experimental results show that CP12 binds divalent copper ions ($Cu^{2+}$) [29]. In particular, one of these patches that do not interact with any positive charges includes ASP-36, GLU-39 and GLU-40. We are currently investigating the role of these residues for the binding of $Cu^{2+}$ by site directed mutagenesis experiments.

### 3.4. CP12: a typical disordered protein linker

CP12 is thought to be an intrinsically disordered protein namely in the reduced state. However, and despite conserving its properties of flexibility, CP12 in the oxidized form is shown both experimentally and by prediction, to possess α-helices and is constrained by two disulphide bonds.

It is well recognized that functions associated with intrinsically disordered proteins involve, in many cases, protein–protein interactions [17]. One interesting feature of these kind of proteins is that, in absence of their targets, they behave as unstructured or molten-globules ensemble. However, binding can induce conformational changes whereby regular secondary structures (mainly α-helices) appear in these linkers; this can have significant functional consequences. One example is the kinase-inducible-transcriptional-activation domain (KID) of the cyclic-AMP-response-element-binding protein (CREB) which is known to be intrinsically disordered [18,19] but undergoes transition to a most ordered state to form a pair of orthogonal helices upon binding to its target domain in CREB-binding protein [20].

Interestingly and similarly to KID, the relative intrinsic disorder of some portions of CP12 can be predicted by disorder probability calculations, as can inherent helical propensity in two regions. By using DisEmbl [21], GlobPlot [22] or DISOPRED [23], sequence signatures that have high disorder probability could be identified in CP12 sequence (data available upon request) whereas JPRED calculations predict two long α-helices in this protein (vide supra).

As shown for our CP12 model, the interatomic contacts involving key conserved hydrophobic residues contribute to reducing their exposure to the solvent. But in such small polypeptide systems, these contacts are not sufficient to induce spontaneous folding into highly organized 3D structure. It can hence be deduced that CP12 is similar to those cases where compact but disordered molten-globule-like states can be formed and where local regions of the sequence show a propensity to adopt isolated and fluctuating elements of a secondary structure [24].

Thus, these characteristics support the idea that CP12 is a typical disordered protein linker that adopts locally regular secondary structures. Based on our results and on known previous examples [18–20,25–27], we can reliably hypothesize that the proposed model of CP12 would therefore represent the linker in its complexed state.

## 4. Conclusion

In the process of modelling CP12, we had to face two major concerns. In first instance, despite the absence of homologous structures from structural databases, we were able to identify useful templates that share low sequence similarity with CP12 which, when combined together, encompass the whole length of CP12. In second instance, we were able to derive an additional set of useful constraints from both predictive methods and experimental data thereby overcoming the fact that critical mass is lacking to form a hydrophobic core in small proteins and that the two existing disulphides are non-interlocked.

Interestingly, one of the models derived from comparative modelling in MODELLER and subsequently refined using energy minimization was validated and displayed several meaningful features: regular secondary structures, two faces with asymmetrical charge distribution, conserved residues engaged in non-bonded interactions. This model further brings into front new insights regarding the biology of the protein namely with respect to its function as protein linker. In fact, our work demonstrates the possible re-conciliation between the ambivalent propensities of CP12 to be disordered and to form regular structures. And, as quoted by Dyson and Wright [17], this further illustrates that ‘concepts in protein structure and function will now evolve to encompass a continuum from fully unstructured proteins that fold on binding to their targets through strings of protein domains that assemble on binding their target to relatively rigid proteins with mobile functional regions’.

Analysis of the model of CP12 proposed in this paper suggested further experimental investigations and simulations. These are mainly site-directed mutagenesis, further NMR experiments, and docking to its target enzyme GAPDH. We are currently investigating both experimentally and by simulations the formation of a complex between the enzyme and CP12. These results will be the subject of a forthcoming paper.

## Acknowledgements

R. Thangudu is currently supported by a PhD grant from the Conseil Régional de La Réunion. The authors are most thankful to Dr. R. Sowdhamini for fruitful discussions.

## Appendix A

The 3D structures of the 19 fragments that appear in the alignment used in the comparative modelling approach are presented in Fig. 10 (see also Fig. 1). The CP12 final geometry is crudely delimited to illustrate the relationship with the respective portions of the template structures.

## References


[1] E. Graciet, P. Gans, N. Wedel, S. Lebreton, J.-M. Camadro, B. Gontero, The small protein CP12: a protein linker for supramolecular complex assembly, Biochemistry 42 (2003) 8163–8170.

[2] N. Wedel, J. Soll, B.K. Paap, Cp12 provides a new mode of light regulation of Calvin Cycle activity in higher plants, Proc. Natl. Acad. Sci. USA 94 (1997) 10479–10484.

[3] E. Graciet, S. Lebreton, B. Gontero, Emergence of new regulatory mechanisms in the benson-calvin pathway via protein–protein interactions: a glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase complex, J. Exp. Bot. 55 (2004) 1245–1254.

[4] F. Mouche, B. Gontero, I. Callebaut, J.P. Mornon, N. Boisset, Striking conformational change suspected within the phosphoribulokinase dimer induced by interaction with gapdh, J. Biol. Chem. 277 (2002) 6743–6749.

[5] J.A. Cuff, M.E. Clamp, A.S. Siddiqui, M. Finlay, G.J. Barton, Jpred: a consensus secondary structure prediction server, Bioinformatics 14 (1998) 892–893.

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

[7] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, Procheck—a program to check the stereochemical quality of protein structures, J. Appl. Cryst. 26 (1993) 47–60.

[8] M. Wiederstein, P. Lackner, F. Kienberger, M.J. Sippl, Directed in silico mutagenesis, in: S. Brakmann, A. Schwienhorst (Eds.), Evolutionary Methods in Biotechnology, Wiley-VCH, 2004.

[9] G. Vriend, What if: a molecular modelling and drug design program, J. Mol. Graph. 8 (1990) 52–56.

[10] R.R. Thangudu, A. Vinayagam, G. Pugalenthi, A. Manonmani, B. Offmann, R. Sowdhamini, Native and modelled disulphide bonds in proteins: knowledge-based approaches towards structure prediction of disulphide-rich polypeptides, Protein Str. Func. Bioinf. 58 (2005) 866–879.

[11] L.A. Kelley, R.M. MacCallum, M.J.E. Sternberg, Enhanced genome annotation using structural profiles in the program 3D-pssm, J. Mol. Biol. 299 (2000) 499–520.

[12] Accelrys, formerly Molecular Simulation Inc., 9685 Scranton Road, San Diego, CA, 2001.

[13] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes, J. Comput. Phys. 23 (1977) 327.

[14] J.D. Thompson, D.G. Higgins, T.J. Gibson, Clustalw: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucl. Acid Res. 22 (1994) 4673–4680.

[15] F. Glaser, T. Pupko, I. Paz, R.E. Bell, D. Bechor-Shental, E. Martz, N. Ben-Tal, Consurf: identification of functional regions in proteins by surface-mapping of phylogenetic information, Bioinformatics 19 (2003) 163–164.

[16] L. Avilan, S. Lebreton, B. Gontero, Thioredoxin activation of phosphoribulokinase in a bi-enzyme complex from *Chlamydomonas reinhardtii* chloroplasts, J. Biol. Chem. 275 (2000) 9447–9451.

[17] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, Nat. Rev. 6 (2005) 197–208.

[18] I. Radhakrishna, G.C. Pérez-Alvarado, H.J. Dyson, P.E. Wright, Conformational preferences in the ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of creb, FEBS Lett. 430 (1998) 317–322.

[19] J.P. Richards, H.P. Bachinger, R.H. Goodman, R.G. Brennan, Analysis of the structural properties of creb and phosphorylated creb, J. Biol. Chem. 271 (1996) 13716–13723.

[20] I. Radhakrishnan, G.C. Pérez-Alvarado, D. Parker, H.J. Dyson, M.R. Montminy, P.E. Wright, Solution structure of the kix binding domain of cbp bound to the transactivation domain of creb: a model for activator:-coactivator interactions, Cell 91 (1997) 741–752.

[21] R. Linding, L.J. Jensen, F. Diella, P. Bork, T.J. Gibson, R.B. Russell, Protein disorder prediction. implications for structural proteomics, Structure (Camb) 11 (2003) 1453–1459.

[22] R. Linding, R.B. Russel, V. Neduva, T.J. Gibson, Globplot: exploring protein sequences for globularity and disorder, Nucl. Acid Res. 31 (2003) 3701–3708.

[23] J.J. Ward, J.S. Sodhi, L.J. McGuffin, B.F. Buxton, D.T. Jones, Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, J. Mol. Biol. 337 (2004) 635–645.

[24] V.N. Uversky, Natively unfolded proteins: a point where biology waits for physics. review, Protein Sci. 11 (2002) 739–756.

[25] A.A. Russo, P.D. Jeffrey, A.K. Patten, J. Massagué, N.P. Pavletich, Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A–Cdk2 complex, Nature 382 (1996) 325–331.

[26] A.H. Huber, W. Weis, The structure of the β-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by β-catenin, Cell 105 (2001) 391–402.

[27] M.K. Sorenson, S.S. Ray, S.A. Darst, Crystal structure of the flagellar sigma/anti-sigma complex sigma(28)/flgm reveals an intact sigma factor in an inactive conformation, Mol. Cell 14 (2004) 127–138.

[28] W. Kabsch, C. Sander, Definition of secondary structure of proteins given a set of 3D coordinates, Biopolymers 22 (1983) 2577–2637.

[29] A. Delobel, E. Graciet, S. Andreescu, B. Gontero, F. Halgand, O. Laprévote, Mass spectrometric analysis of the interactions between a chloroplast protein, CP12 and metal ions. A possible regulatory role within a PRK/GAPDH/CP12 complex, Rapid Commun. Mass Spectrom. 19 (2005) 3379–3388.