A new type of disordered CP12 protein in the marine diatom *Thalassiosira pseudonana*

Hui Shao
CNRS: Centre National de la Recherche Scientifique
Wenmin Huang
CAS
Luisana Avilan
University of Porthsmouth
Veronique Receveur-Brechot
CNRS
Carine Puppo
CNRS
Rémy Puppo
cnrs
Régine Lebrun
CNRS
Brigitte Gontero (bmeunier@imm.cnrs.fr)
CNRS https://orcid.org/0000-0003-1731-712X
Helene Launay
CNRS: Centre National de la Recherche Scientifique

Research

Keywords: coiled coil, diatom, intrinsically disordered protein IDP, Nuclear Magnetic Resonance, photosynthesis, Small angle X-ray Scattering

Posted Date: December 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-135099/v1

Version of Record: A version of this preprint was published at Cell Communication and Signaling on March 24th, 2021. See the published version at https://doi.org/10.1186/s12964-021-00718-x.

Abstract

Background

CP12 is a small chloroplast protein that is widespread in various photosynthetic organisms and is often involved in the redox metabolic on/off switch of the Calvin Benson Bassham (CBB) cycle. The gene encoding this protein is conserved in many diatoms, but the protein has been overlooked in these organisms, despite their ecological predominance and their complex and still enigmatic evolutionary background.

Methods

A combination of biochemical, bioinformatics and biophysical methods including electrospray ionization-mass spectrometry, circular dichroism, nuclear magnetic resonance and small X ray scattering spectroscopy, was used to characterize a diatom CP12.

Results

Here, we demonstrate that CP12 is expressed in the marine diatom *Thalassiosira pseudonana* constitutively in dark-treated and in continuous light-treated cells as well as in all growth phases. This CP12 behaves abnormally under gel electrophoresis, is heat resistant and lacks a structural core, all features of intrinsically disorder family similarly to its homologues in other species. By contrast, unlike other known CP12 proteins that are monomers, this protein is a dimer as shown by native electrospray ionization-mass spectrometry and small angle X-ray scattering. In addition, small angle X-ray scattering showed that this CP12 is an elongated cylinder with kinks. Circular dichroism spectra indicated that CP12, though it has features of disordered proteins, has a high content of α-helices. Nuclear magnetic resonance spectroscopy showed that these helices are unstable and dynamic within a millisecond timescale. Together with *in silico* predictions, these results suggest that *T. pseudonana* CP12 has both coiled-coil and disordered regions.

Conclusions

These findings bring new insights into the large family of intrinsically disordered proteins increasing the diversity of known CP12 proteins. This raises questions about the role of this protein in addition to the well-established regulation of the CBB cycle.

Background

CP12 (chloroplast protein of 12 kDa) is a small nuclear encoded protein of about 80 amino acid residues, hat occurs in many photosynthetic organisms [2, 3]. In higher

plants, green and red algae, and cyanobacteria, it is associated with two enzymes, phosphoribulokinase (PRK) and glyceraldehyde 3 phosphate dehydrogenase (GAPDH) from the Calvin Benson Bassham cycle that is responsible for CO_2 assimilation [4]. This ternary complex has been well-studied and its structure has been recently solved using cryo-electron microscopy in the cyanobacterium *Synechococcus elongatus* [5] and using X-ray diffraction in the model higher plant *Arabidopsis thaliana* [6]. CP12 proteins from different organisms have some highly conserved regions such as a AWD_VEEL motif [2] and in most cases have a pair of cysteine residues at the C-terminus and/or a second pair at the N-terminus. Although cysteine residues are order-promoting amino-acids, and likely to structure the molecule, CP12 shares some physico-chemical properties with intrinsically disordered proteins (IDPs). In agreement with predictors of intrinsic disorder, CP12 from the green alga *Chlamydomonas reinhardtii* [7] and later from the angiosperm *A. thaliana* [8] were shown by circular dichroism (CD) and Nuclear Magnetic Resonance (NMR) to contain little regular secondary structure in solution [9–11].

These proteins have been extensively studied in these organisms and in *C. reinhardtii* it is a jack–of-all-trades that can not only bind to PRK and GAPDH under its oxidized state, thereby downregulating their activity upon complex formation in the dark but can also perform other functions[12]. For example, it can bind metal ions[13], and can act as a specific chaperone-like for GAPDH [14]. In the tropical legume, *Stylosanthes guianensis*, higher expression of CP12 increases growth, plant height and photosynthesis rate [15]. Conversely, in the tobacco *Nicotiana tabacum* and in the mouse-ear cress *A. thaliana*, antisense suppression of CP12 reduces the rate of photosynthesis and increases the expression of proteins related to oxidative stress [16, 17]. Finally, in *A. thaliana*, one CP12 isoform is mainly expressed in non-photosynthetic tissues (roots and floral tissues) [18]. All these features suggest that CP12 proteins have other functions beyond the dark downregulation of CBB enzymes.

Little is known about CP12 in the diatoms, an ecologically important group of microalgae. Diatoms have a more complex evolutionary background than the other organisms mentioned above in which CP12 has been studied, and the regulation of their CBB enzymes is not fully understood [19]. The complex PRK-GAPDH-CP12 does not seem to be present [20–23] though there are some studies indicating a possible presence of CP12 in these organisms [20, 24]. For example, a CP12-like protein from *Thalassiosira pseudonana* has been shown to be expressed under stress conditions such as low CO_2 [25] or low nutrients (nitrogen, phosphorus, silicon) [26]. The aim of this manuscript was therefore to characterize the structural properties of this protein from the marine diatom *T. pseudonana*, using both *in silico* and a range of biophysical experimental approaches.

Material And Methods

Diurnal expression of *T. pseudonana* CP12 *in vivo*

Cells from *T. pseudonana* (strain CCMP 1335 from https://ncma.bigelow.org/) were grown in F/2 + Si medium (http://www.ccap.ac.uk/) under continuous light (50 µmol.photon.m^{-2}.s^{-1}) in an incubator (… 19 °C and shaken at 100 rpm. Growth of *T. pseudonana* was

monitored using the absorbance at 680 nm. When the cells reached the exponential phase, half of the culture was put in the dark, and half left in the light. After 24 hours, cells were collected by centrifugation for 15 min at 3275 g, 19 °C with a Beckman Allegra X15R centrifuge, and resuspended in 15 mM tris(hydroxymethyl)aminomethane (Tris), 4 mM ethylenediaminetetraacetic acid (EDTA), pH 7.9 with protease inhibitors (Sigma) 0.5 µg.mL^{-1}. The cells were sonicated (Sonic Ruptor 250, one ice, 4 cycles, 1 min sonication and 1 min rest), then centrifuged at 16 000 g for 20 min at 4 °C and the supernatant collected. Protein amount was measured by the Bradford protein assay using bovine serum albumin as a standard (Bio-Rad). Proteins were loaded to 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) that was stained with Coomassie Blue or immediately transferred onto a 0.45 µm nitrocellulose membrane (Thermo Fisher Scientific). The antibodies raised against recombinant His-tagged CP12 in rabbits were produced by Eurogentec (https://www.eurogentec.com/en/custom-antibodies). The membrane was incubated first with α-CP12 antibody diluted 1: 10 000, then with goat anti-rabbit IgG horse radish peroxidase (HRP, Invitrogen) diluted 1: 10 000. Finally, the membrane was revealed with luminol-based substrate (Amersham Enhanced Chemiluminescence western blotting kit detection reagent) using ImageQuant LAS 4000 biomolecular imager (GE Healthcare).

Expression of CP12 during different phases of growth

A preculture of *T. pseudonana* cells in F/2 + Si medium, under continuous light (50 µmol.photon.m^{-2}.s^{-1}), was first grown at 19 °C, shaken at 100 rpm under high CO_2 concentration (20 000 ppm) to increase biomass. After five days, the pellet of a 40 mL aliquot of this pre-culture obtained by centrifugation at 3275 g for 15 min at 4 °C was washed and re-suspended in fresh F/2 + Si medium. This was performed three times. These cells were then inoculated into fresh F/2 + Si medium at an initial absorbance at 680 nm of 0.2 (pathway length 1 cm) and grown under air-concentration of CO_2 (400 ppm). Growth of *T. pseudonana* was monitored using the absorbance at 680 nm. Every day, a volume of the culture was collected that was normalized to obtain 30 µg of total protein according to the following equation: $Volume(mL) = \frac{0.908}{ODA_{680nm}}$. A pellet of cells was obtained after centrifugation for one minute at 3275 g and re-suspended into SDS-PAGE loading buffer containing 1 mM dithiothreïtol. Cell lysis and protein denaturation were performed at 95 °C for 10 min. During the exponential phase and the beginning of the stationary phase, the expression of CP12 was monitored using western-blot analysis.

In silico bioinformatic analysis

The subcellular localization of CP12 from five species of diatoms was predicted using a signal peptide predictor dedicated to diatoms, HECTAR (https://webtools.sb-roscoff.fr/) [27]. The signal peptides of CP12 from other organisms were predicted using ChloroP (http://www.cbs.dtu.dk/services/ChloroP/, [28]), SignalP (http://www.cbs.dtu.dk/services/SignalP/ [29]) and TargetP (http://www.cbs.dtu.dk/services/TargetP/ [30]). These enabled the N-terminus of the mature chloroplast CP12 proteins to be determined. Disordered regions were predicted using PONDR VL-XT (http://www.pondr.com/) [31], DisEMBL_Remark-465 (http://dis.embl.de/) [32], IUPred2A

(https://iupred2a.elte.hu/) [33] and s2D (http://www-mvsoftware.ch.cam.ac.uk/index.php/s2D) [34]. Coiled coil regions were predicted using Paircoil (http://cb.csail.mit.edu/cb/paircoil2/paircoil2-like.html), with a minimum search window of 21 residues [35]. Amino acid frequency in the algal and diatom CP12 was analyzed by composition profiler (http://www.cprofiler.org/cgi-bin/profiler.cgi) [36] against the Protein Data Bank (PDB) (https://www.rcsb.org/).

Overexpression and purification of CP12 from *T. pseudonana*

Primers containing the *NdeI* and *BamHI* restriction sites were used to amplify and clone the CP12 gene in frame with the N-terminal histidine tag of the pET28a expression vector (Novagen) (forward primer 5' CATATGGCTGCCATTGAAGCTGCTCT 3' and reverse primer 5' GGATCCCTAACGGGAACCAAGGGCC 3'). This plasmid was used to transform *Escherichia coli* BL21(DE3) pLysS. Freshly transformed bacteria were grown in 2YT medium with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol at 37 °C until the absorbance at 600 nm reached 0.5 to 0.6 (1 cm pathlength). Cultures were cooled on ice for 30 min and then CP12 expression was induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Cells were cultured at 30 °C overnight in an incubator (Edmund Bühler GmbH, Fisher Bioblock Scientific), then centrifuged at 3275 g at 4 °C (Beckman Allegra X15R centrifuge). The pellets containing cells were re-suspended in 50 mM NaH_2PO_4/Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0 (Ni-NTA buffer). Cells were then broken by sonication (Sonic Ruptor 250, 1 min sonication and 1 min on ice, 4 cycles) and centrifuged at 27000 g for 20 min at 4 °C. The supernatant contained the recombinant histidine tagged CP12 (His-CP12) that was then purified by nickel ion affinity chromatography on Ni-NTA agarose column (Qiagen) (1.2 × 8 cm of resin). The column was equilibrated with Ni-NTA buffer. Contaminants were firstly washed out with 10 mM imidazole until the absorbance at 280 nm reached a minimum, then fractions were gradually eluted with an imidazole gradient (10 to 250 mM imidazole, 2 × 45 mL). Proteins elution was followed by absorbance at 280 nm. His-tagged CP12 was eluted with 150 mM imidazole, and dialyzed with 10 mM sodium phosphate buffer, pH 7.4 then stored at -20 °C. Size exclusion chromatography (SEC), electrospray ionization coupled to mass spectrometry (ESI-MS) and circular dichroism (CD) were performed on CP12 after His tag removal with thrombin (T4648, Sigma) (1 u for 100 µg of CP12) at room temperature for 18 h. The sample was concentrated using a 500 µL spin X-UF ultra centrifugal concentrator, Corning, 5 kDa cut-off. After His-tag removal, CP12 was stored at -20 °C.

Determination of redox state

The free thiol groups in CP12 were quantified using 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) (Sigma-Aldrich). 2 µM CP12 was mixed with 50 µM DTNB in 10 mM phosphate, 2 mM EDTA at pH 8 and the absorbance followed at 412 nm. A control without the protein was recorded in parallel. Number of free SH groups was calculated from N = $\Delta OD / \varepsilon CL$ (N: number of thiol; ε: DTNB extinction coefficient, 13600 $M^{-1}.cm^{-1}$; C: protein concentration in M; L: light pathlength in cm).

Native electrospray ionization-mass spectrometry (ESI-MS)

Prior to native ESI-MS analysis, CP12 without the histidine tag was dialyzed against 200 mM ammonium acetate buffer (pH 8.0) using 5 kDa cut-off concentrator columns (Spin-X, Corning). Experiments at 5 μM of CP12 were carried out on an electrospray Q-ToF mass spectrometer (Synapt G1 HDMS, Waters) using NanoLockSpray ionisation source with borosilicate emitter (NanoES spray capillaries, Thermo Scientific). Optimized instrument parameters were as follows: source pressure 5.3 mbar, source temperature 20 °C, capillary voltage 1.8 kV, sampling cone voltage 180 V, extractor cone voltage 4 V, trap collision energy 30 V and transfer collision energy 20 V. Mass spectrometer was calibrated in positive mode from 1000–5000 m/z with CsI (1 mg/mL) just prior acquisition.

Size exclusion chromatography

500 μL of CP12 after histidine tag removal, at 0.5 mM were loaded on a Hiload Superdex 200 (prep grade 26/60) equilibrated with 150 mM NaCl, 50 mM sodium phosphate buffer at pH 7.5. The column was calibrated with seven globular proteins of different molecular mass: thyroglobulin, ferritin, alcohol dehydrogenase, conalbumin, ovalbumin andcarbonic anhydrase. Throughout elution, the absorbance at 280 nm was monitored to determine the presence of proteins. The fractions were collected, concentrated (spin X-UF ultra centrifugal concentrator, Corning, 5 kDa cut-off) and stored at -20 °C.

Circular dichroism (CD)

10 μM CP12 without histidine tag in 50 mM sodium phosphate pH 6.5 was used to record CD spectra (Jasco 815CD spectrometer, 2 mm thick quartz cells) scanned from 190 to 260 nm at a speed of 10 nm/min (n = 3). The data were analyzed using Dichroweb standard algorithms (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) [37].

Nuclear magnetic resonance (NMR)

^{15}N labelled histidine tagged CP12 were produced using the enhanced M9 medium (protocol of the European Molecular Biology Laboratory, https://www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/n15/index.html) and purified as described above. The final sample was buffer exchanged in 50 mM sodium phosphate pH 6.5, 50 mM NaCl, 10% D_2O with traces of sodium trimethylsilylpropanesulfonate (DSS) and at a final protein concentration of 250 μM.

The data were recorded at 4 ºC and 15 ºC. Fast ^{1}H-^{15}N heteronuclear single quantum correlation (fHSQC [38]) spectra were recorded with a ^{1}H acquisition time of 243 ms, ^{15}N acquisition time of 42 ms and with 24 scans on a 600 MHz NMR spectrum equipped with a cryogenic probe (Bruker). Translational diffusion was measured using standard bi-polar stimulated echo experiment [39], with a diffusion delay (Δ) of 200 ms. Ten experiments were recorded in which the encoding and decoding pair of gradients are produced with squared 1.4 ms long gradients (δ) with strength (G) ranging from 2–98% of the maximum gradient strengths (5.1 $G.mm^{1}$). The data were processed using nmrPipe [40], plotted using Sparky[41]. ng Octave [42] from the integral of proton signals of the

methyl side chains of the protein from 1.2 to 0.7 ppm. The linear dependency of the logarithm of the integral as a function of the gradient strength was used to determine D as follows:

$ln\left(\frac{I_G}{I_0}\right) = D.\left[\Delta - \frac{\delta}{3}\right].\left(\delta.\mathrm{G}.\gamma_H\right)^2$, where I_G is the integral as a function of the gradient strength, I_0 is the integral in the absence of gradient, γ_H is the proton gyromagnetic ratio, and Δ, δ and G are defined above. The hydrodynamic radius associated with the diffusion coefficient was determined using the Stokes-Einstein equation:

$r_H = \frac{k_B.T}{6\pi.D.\eta_{(T)}}$ where k_B is the Boltzmann constant, T the temperature in Kelvin and $\eta_{(T)}$ the viscosity.

Small angle X ray scattering (SAXS)

SAXS experiments were performed on SWING beamline at the SOLEIL synchrotron using the online HPLC size exclusion chromatography facilities [43]. The sample-to-detector (CCD Aviex) distance was set at 2 m, leading to scattering vectors ($q = 4\pi/\lambda\sin\theta$, where 2θ is the scattering angle and λ the wavelength, equal to 1.033 Å) ranging from 0.01 to 0.46 Å. 50 µL of His-tagged CP12 (11 mg/mL) were injected into a pre-equilibrated size exclusion chromatography column (Agilent Bio-SEC-3 300 Å) upstream of the measurement capillary at a temperature of 15 °C. Frames of 990 ms with dead-time of 10 ms were recorded throughout the elution, with 100 frames recorded at the very first minutes of the elution to measure the buffer background. The protein concentration was monitored via the absorbance at 280 nm with an *in situ* spectrophotometer. The experiment was performed in 30 mM Tris, 50 mM NaCl, 2 mM EDTA, 1 mM Tris(2-carboxyethyl)phosphine (TCEP), pH 7.5.

Data reduction to absolute units and solvent subtraction were performed using FOXTROT, a dedicated in-house application. The frames recorded during the elution peak were carefully compared with each other, and data corresponding to identical scattering profiles and radius of gyration (Rg) 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 up to $q.Rg < 1.0$, and the distance distribution function P(r) was obtained via GNOM. The molecular mass of CP12 was determined using the forward scattering intensity I(0) of the frame corresponding to the top of the peak, at the highest protein concentration as described in [45]. *Ab initio* models of 3D envelopes corresponding to the scattering curve were constructed using DAMMIF (10 runs) and GASBOR (5 runs) with P1 and P2 symmetry [44].

3D-modelling

3-D modelling of the coiled coil domain has been performed using Swiss-model (https://swissmodel.expasy.org/ [46]) on a sequence including 5 residues before and 6 residues after the

predicted coiled coil sequence highlighted in bold (APIVD**SEYEAKVKSLSQMLTKTKAELDQVKALADDLKGVK**LASPSV), without any other input.

Results

Expression of a CP12 protein in *T. pseudonana*

Chen et al. [26] reported a hypothetical protein of uncharacterized function that was overexpressed when *T. pseudonana* cells were limited by N, P or Si (identification: XP_002286772). This protein is identical to our previously identified CP12 protein that is expressed under low CO_2 condition [25]. Here we showed that this CP12 protein was present in dark-treated and light-treated *T. pseudonana* cells (Fig. 1). Similarly, the expression level was stable during all phases of growth (Fig. 2).

In silico analysis of CP12 sequences

The sequence of this CP12 protein was aligned with other representative CP12 sequences from the angiosperms *A. thaliana*, *Pisum sativum* and *Spinacia oleracea*, the green alga *C. reinhardtii*, two red algae *Cyanidioschyzon merolae and Galdieria sulphuraria*, and the cyanobacterium *S. elongatus* (PCC 7942) (Fig. 3). Beside the highly conserved AWD_VEEL motif, one pair of cysteine residues at the C-terminus, another hallmark for CP12s, was also present. In contrast, the highly conserved proline positioned at the center of the 8-residues linker between these two cysteine residues is absent in CP12 from *T. pseudonana*. In the nuclear genome of other diatoms, *Phaeodactylum tricornutum*, *Thalassiosira oceanica*, *Fragilariopsis cylindrus* and *Pseudo-nitzschia multistriata*, a gene encoding a protein with high similarity with this CP12 protein was also found (Fig. 3). Furthermore, all these newly identified CP12 proteins from diatoms were predicted to be addressed to the chloroplast, in good agreement with the known localization of CP12 in other well-studied organisms.

In silico characterization of the propensity of disorder

Since CP12 proteins from other organisms are IDPs [9, 11, 47], we checked if this was the case for CP12 from *T. pseudonana*. The amino acid composition of this CP12 and that from *C. reinhardtii*, a well-known IDP, were compared to globular proteins. In the sequences of CP12 from *C. reinhardtii* and *T. pseudonana*, alanine and charged residues that promote disorder such as aspartate, glutamate and lysine residues (D, E and K), are more abundant than in globular structured proteins (Fig. 4). In contrast, order-promoting residues such as tryptophan, phenylalanine and tyrosine residues (W, F and Y) are less abundant than in structured proteins. In CP12 from *T. pseudonana*, cysteine residues are less abundant (two cysteine residues) than in the green algal CP12 (four cysteine residues). Titration with the Ellman's reagent, DTNB, of two thiol groups in our recombinant CP12 from *T. pseudonan*a showed that its two cysteine residues were not involved in a disulfide bridge, even under the atmospheric aerobic conditions. It has an isoelectric point of 4.5 like many proteins, and interestingly, a very high negative net charge − 14.6 strengthening the hypothesis that this protein may be an IDP.

The disorder propensity of CP12 from *T. pseudonana* was predicted using several algorithms. PONDR predicted a long disordered segment between residues 81–146, whereas the other predictors were more stringent (Fig. 5), in particular IUPred2A which predicts only a small number of disordered regions. Taking all predictions together, four regions of disorder were consensually predicted which encompass residues 30–42, 77–93, 108–117 and 129–139 (shaded residues in Fig. 5). Apart from the disordered regions, the algorithm s2D predicts the other regions to have a high propensity to form helices.

Experimental characterization of the propensity of disorder

Abnormal migration is often observed in SDS-PAGE for IDPs, with an apparent molecular mass higher than the theoretical one. The theoretical molecular mass of a monomer of CP12 from *T. pseudonana* is 17.6 kDa but it migrated as a protein of about 25 kDa in SDS-PAGE (Fig. 6). This abnormal migration resembles that of CP12 from *C. reinhardtii* where the molecular mass of the monomer with a histidine tag is ~ 11 kDa but migrates at about 20 kDa. Another key characteristic of IDPs is their thermal resistance; CP12 from *T. pseudonana*, like that from *C. reinhardtii*, was found in the supernatant after heat treatment at 95 °C. This indicates that even at high temperature this protein does not precipitate nor aggregate (Fig. 6).

Oligomeric state of CP12

Oligomeric state of CP12 was investigated using native ESI-MS. Monomer and dimer oligomeric states were observed (Fig. 7A). Experimental deconvoluted values of 17 642.1 (± 0.5) Da and 35 285.3 (± 0.7) Da fit well with theoretical values for the monomer and dimer species, 17 641.8 Da and 35 283.6 Da respectively. In parallel, size exclusion chromatography showed a single elution peak (Fig. 7B). The homogenous peak observed under size exclusion chromatography corresponds to a unique oligomeric state in solution (Fig. 7B). It is likely that during the ionization of the molecular complex, the dimer dissociates partially, indicating that it is not a covalent dimer. The discrepancy between the molecular mass determined by ESI-MS for the dimeric state and the size exclusion chromatography elution volume of CP12 (corresponding to an apparent molecular mass of 93 (± 4) kDa) is consistent with CP12 being an extended dimer.

CP12 is a dynamic extended dimer

The hydrodynamic size of CP12 in solution was confirmed by measuring the translational diffusion coefficient using DOSY-NMR (Fig. 8A). The translational diffusion coefficient was 3.8 (± 0.1) 10^{-11} $m^2.s^{-1}$ at 4 ºC, which corresponds to a hydrodynamic radius (Rh) of 3.4 (± 1) nm. For a folded dimer of 35.3 kDa, the theoretical Rh is 2.75 nm, while for an unfolded dimer it is 5.94 nm [48]. Therefore, the experimental Rh of 3.4 nm represents an intermediate between that expected for a fully disordered and a fully ordered dimer and is also compatible with an extended dimer.

^{1}H-^{15}N fast-HSQC spectrum of CP12 presents some characteristic features of that of a disordered protein with the [illegible] of ^{1}H chemical shift dispersion (Fig. 8B). 34 resonances present narrow linewidths

compatible with a purely disordered region. However, in contrast to true random-coil protein such as CP12 from *C. reinhardtii* (Fig. 8C, [9]), the linewidths of the observed resonance varied from narrow (< 20 Hz) to broad (> 60 Hz), indicating the presence of an intermediate chemical exchange on the NMR timescale. This result showed that reduced CP12 is highly dynamic on a range of timescale from ps to ms and does not possess a stable secondary structure.

We then performed SAXS experiment on CP12 to assess the overall structure and oligomeric state of disordered or partly disordered proteins (Fig. 9) [45]. The Guinier plot of the average spectrum was linear with a slight increase at very small q ($q < 0.14$ Å^{-1}), indicating the presence of very small traces of aggregates that were not detected by the absorbance at 280 nm (Fig. 9A). The molecular mass inferred from the forward scattering intensity is 41.5 (± 3.0) kDa, confirming that CP12 is dimeric, as the theoretical molecular mass of dimeric histidine-tagged CP12 is 39 kDa. The inferred radius of gyration of CP12 was 38.2 (± 0.4) Å. The maximum dimension of the protein was 135 (± 5) Å, indicating that the protein is elongated. The normalized Kratky plot exhibited a large peak at (q.Rg) = 3.4 and $(q.Rg)^2.I(q)/I(0) = 1.7$, followed by a decrease and then an increase of the curve, and is typical of a protein containing both globular folded domain(s) and disordered region(s) (Fig. 9A) [45]. Determination of the envelope of CP12 gave similar results using DAMMIF or GASBOR, applying a P1 or P2 symmetry, with excellent fits to the data ($\chi^2 = 1.61–1.66$ with DAMMIF, and 1.39–2.39 with GASBOR, Fig. 9B). The envelope is very elongated and forms a cylinder of ~ 20–25 Å diameter, with several kinks which may reflect the limits between different domains (Fig. 9C). This is also consistent with the presence of several disordered regions as previously mentioned (Fig. 5).

Secondary structure of CP12

Circular dichroism (CD) was used to determine the secondary structure of CP12. The CD spectrum had two minima at 208 and 222 nm (Fig. 10A), implying a high content of α helices (50%), plus 10% of β-sheets and 14% of random coil. This result is consistent with the prediction of both disordered and helical regions using the s2D predictor. To investigate further the conformation of CP12, the solvent 2,2,2-trifluoroethanol (TFE), known to stabilize helical structures, was added. The CD spectra with increasing TFE concentrations showed an increase of the content of α helices up to 79%, as revealed by lower values of ellipticity at 208 and 222 nm. We observed an isosbestic point at 201–202 nm, indicating the existence of two conformers in equilibrium (Fig. 10A). The presence of unstable helical structures in a dimeric and elongated protein is reminiscent of a coiled coil propensity, we thus used a predictor for coiled coil regions and found that the region encompassing residues 46–82 (Fig. 10B) has a very high propensity for coiled coil arrangement. Furthermore, 3-D modelling of this domain using Swiss-Model yielded a dimeric coiled coil structure that perfectly fits inside the SAXS envelope (Fig. 9C). The propensity to form coiled coil structures might be a characteristic shared with other CP12 proteins, and at least CP12 from *C. reinhardtii* has a weak propensity to form coiled coil in the region encompassing the AWD_VEEL motif (Fig. 10C). These *in silico* predictions are consistent with our experimental data; in particular the SAXS results are consistent with a putative central coiled coil domain flanked by other elongated domains likely short structural elements.

Discussion

CP12 proteins from different organisms are highly diverse at the protein sequence level [2] but have some characteristic features such as the AWD_VEEL motif and often a pair of cysteine residues, at the C terminus, separated by eight residues encompassing a proline residue. The CP12 proteins from *C. reinhardtii*, *A. thaliana* and *S. elongatus* have been extensively studied and shown to belong to the IDP family. IDPs, or in other words ductile, dancing, malleable or flexible proteins [49], are common in various proteomes [50–55] and occupy a unique structural and functional niche in which function is directly linked to structural disorder [47, 56–58]. The expression of CP12 from the diatom *T. pseudonana* is constitutively expressed under dark and light conditions, as well as during the growth, unlike CP12 from *A. thaliana* that is co-expressed with PRK and GAPDH in the light [59]. This indicates that CP12 may have other functions in diatoms compared to viridiplantae. Indeed, other studies have shown that it is up-regulated under stress-related conditions in *T. pseudonana* (nutrient deprivation [26], low CO_2 [25]).

Like CP12 from other organisms [7, 11, 50], CP12 from *T. pseudonana* is characterized by several features of IDPs, with a lower proportion of order-promoting residues and a higher proportion of charged residues than structured proteins. This high net charge and low overall hydrophobicity is concordant with its abnormal migration on SDS-PAGE, its heat-resistance capacity, the absence of rigid globular core as observed using NMR and with its high flexibility on the NMR-timescale. Another IDP characteristic is the ability to be involved in the formation of macromolecular complex, and this is the case for CP12 proteins from other organisms that bind to GAPDH and PRK [7, 8, 16, 60–62]. In contrast, the PRK-GAPDH-CP12 ternary complex has not been found in diatoms. This was attributed to the absence of cysteine residues on PRK in diatoms [22] but may also be the consequences of specific features of this CP12. Moreover, in a freshwater diatom, *Asterionella formosa*, GAPDH interacts with the ferredoxin-NADP reductase (FNR), from the primary phase of photosynthesis, and a small protein identified as a CP12 [63]. Diatom chloroplasts lack the oxidative pentose phosphate pathway, the main NADPH generating source in the dark. In the ternary complex GAPDH-CP12-FNR, GAPDH, a main NADPH consumer, is inhibited [20], thereby releasing the pressure on NADPH availability for other metabolic pathways.

While CP12 from *T. pseudonana* is predicted to contain at least a long disordered region, the proportion of α-helices in CP12 from *T. pseudonana* (50%) is much higher than in CP12 from *C. reinhardtii* [10] (30%) and in canonical IDPs. In addition, SAXS and native ESI-MS showed that CP12 from *T. pseudonana* is dimeric unlike CP12 from other species where it is monomeric in its free isolated state [9–11]. Dimeric oligomerization of very dynamic proteins is mediated by numerous but transient inter-molecular interactions, and this is a feature of peculiar IDPs that form coiled coil arrangement [64–66], but not exclusively [67, 68]. Indeed, CP12 from *T. pseudonana* was predicted to form α-helical coiled coils with a high probability, and this is consistent with the extended hydrodynamic radius experimentally observed by gel filtration and DOSY-NMR, and with the global shape derived from the SAXS data. Coiled coil is an omnipresent protein fold, accounting for about 3 to 10% of all protein-coding regions across all genomes [69–71]. Proteins having coiled coils can be regarded as a division of IDPs, contributing to an ever-

and coiled coils are one of the most ubiquitous protein-protein

interaction motifs [70]. Although CP12 from *C. reinhardtii* is not predicted to form coiled coil structures with the same high probability as *T. pseudonana* CP12, they both share a propensity to form unstable helices [10]. Such highly dynamic and adaptative biophysical properties is a common feature of moonlighting proteins [12].

Conclusion

The structural properties of CP12 from *T. pseudonana*, with putative dimeric coiled coil domain and disordered regions, suggest that it is not an alien as regard to other CP12s and might therefore have one-to-many functions. As the gene encoding this protein has also been found in other diatoms, it could be another facet of the enigmatic regulation of diatom metabolism [19, 23, 72]. These findings extend the context for dynamic and coiled coil proteins related to their functions in photosynthesis regulation and stress in diatoms.

Abbreviations

CBB: Calvin-Benson-Bassham

CD: Circular dichroism

CP12: Chloroplast Protein of 12 kDa

DOSY-NMR: Diffusion ordered spectroscopy - nuclear magnetic resonance

DTNB: 5,5'-dithiobis-(2-nitrobenzoic acid)

EDTA: ethylenediaminetetraacetic acid

ESI-MS: Electrospray ionization-mass spectrometry

FNR: Ferredoxin-NADP reductase

GAPDH: Glyceraldehyde phosphate dehydrogenase

HSQC: Heteronuclear single quantum coherence

IDP: Intrinsically disordered protein

NMR: Nuclear magnetic resonance

PRK: Phosphoribulokinase

SAXS: Small angle X-ray scattering

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC: Size exclusion chromatography

Tris: tris(hydroxymethyl)aminomethane

Declarations

Ethical approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availabitility of supporting data

Not applicable

Competing interests:

The authors declare that they have no competing interests.

Funding

This work was supported by the CNRS. BG is supported by the Agence Nationale de la Recherche grant ANR-15-CE05-0021-03. HL is supported by the A*MIDEX grant AAP-IM2B-NE-2020-02-Launay. HS scholarship was founded by the Chinese Scholarship Council (201704910795). WH was supported by a fellowship for a visiting scholar from the Chinese Academy of Sciences.

Authors' contributions

LA, VRB, BG, HL conceived and planned the experiments. HS, WH, VRB, CP, RP, RL, HL, BG conducted the experiment. VRB, RP, RL, HL, BG analysed the data. BG, HL wrote the manuscript. All authors provided critical feedback on research, analysis and manuscript writing. All authors read and approved the final manuscript.

Acknowledgements

We thank S.C. Maberly for useful discussions and manuscript corrections. We thank O. Bornet for access and maintenance of the IMM NMR platform We acknowledge SOLEIL for provision of synchrotron radiation facilities and we would like to thank the staff of the SWING beamline for assistance during the SAXS measurements.

References

1. Pohlmeyer K, Paap BK, Soll J, Wedel N. CP12: a small nuclear-encoded chloroplast protein provides novel insights into higher-plant GAPDH evolution. Plant Mol Biol. 1996;32:969–78.
2. Groben R, Kaloudas D, Raines CA, Offmann B, Maberly SC, Gontero B. Comparative sequence analysis of CP12, a small protein involved in the formation of a Calvin cycle complex in photosynthetic organisms. Photosynth Res. 2010;103:183–94.
3. Stanley DN, Raines CA, Kerfeld CA. Comparative analysis of 126 cyanobacterial genomes reveals evidence of functional diversity among homologs of the redox-regulated CP12 protein. Plant Physiol. 2013;161:824–35.
4. Gontero B, Avilan L, Lebreton S. Control of carbon fixation in chloroplasts. In: Control of Primary Metabolism of Plants. Editors: Plaxton WC, McManus MT. Oxford, UK: Blackwell Publishing Ltd; 2006. p. 187–218.
5. Mcfarlane C, Shah N, Kabasakal BV, Cotton CAR, Bubeck D, Murray JW. Structural basis of light-induced redox regulation in the Calvin cycle. Proc Natl Acad Sci. 2019;116:20984–90.
6. Yu A, Xie Y, Pan X, Zhang H, Cao P, Su X, et al. Photosynthetic phosphoribulokinase structures: enzymatic mechanisms and the redox regulation of the Calvin-Benson-Bassham cycle. Plant Cell. 2020;32:1556–73.
7. Graciet E, Gans P, Wedel N, Lebreton S, Camadro J-M, Gontero B. The small protein CP12: a protein linker for supramolecular complex assembly. Biochemistry. 2003;42:8163–70.
8. Marri L, Trost P, Trivelli X, Gonnelli L, Pupillo P, Sparla F. Spontaneous assembly of photosynthetic supramolecular complexes as mediated by the intrinsically unstructured protein CP12. J Biol Chem. 2008;283:1831–8.
9. Launay H, Barré P, Puppo C, Manneville S, Gontero B, Receveur-Bréchot V. Absence of residual structure in the intrinsically disordered regulatory protein CP12 in its reduced state. Biochem Biophys Res Commun. 2016;477:20–6.
10. Launay H, Barré P, Puppo C, Zhang Y, Maneville S, Gontero B, et al. Cryptic disorder out of disorder: encounter between conditionally disordered CP12 and glyceraldehyde-3-phosphate dehydrogenase. J Mol Biol. 2018;430:1218–34.
11. Fermani S, Trivelli X, Sparla F, Thumiger A, Calvaresi M, Marri L, et al. Conformational selection and folding-upon-binding of intrinsically disordered protein CP12 regulate photosynthetic enzymes assembly. J Biol Chem. 2012;287:21372–83.
12. Gontero B, Maberly SC. An intrinsically disordered protein, CP12: jack of all trades and master of the Calvin cycle. Biochem Soc Trans. 2012;40:995–9.
13. Erales J, Gontero B, Whitelegge J, Halgand F. 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. 2009;419:75–82, 4 p following 82.
14. Erales J, Lignon S, Gontero B. CP12 from *Chlamydomonas reinhardtii*, a permanent specific “chaperone-like” protein of glyceraldehyde-3-phosphate dehydrogenase. J Biol Chem.

15. Li K, Qiu H, Zhou M, Lin Y, Guo Z, Lu S. Chloroplast Protein 12 Expression Alters Growth and Chilling Tolerance in Tropical Forage *Stylosanthes guianensis* (Aublet) Sw. Front Plant Sci. 2018;9:1319.
16. Howard TP, Fryer MJ, Singh P, Metodiev M, Lytovchenko A, Obata T, et al. Antisense suppression of the small chloroplast protein CP12 in tobacco alters carbon partitioning and severely restricts growth. Plant Physiol. 2011;157:620–31.
17. Elena López-Calcagno P, Omar Abuzaid A, Lawson T, Anne Raines C. Arabidopsis CP12 mutants have reduced levels of phosphoribulokinase and impaired function of the Calvin–Benson cycle. J Exp Bot. 2017;68:2285–98.
18. Singh P, Kaloudas D, Raines CA. Expression analysis of the Arabidopsis CP12 gene family suggests novel roles for these proteins in roots and floral tissues. J Exp Bot. 2008;59:3975–85.
19. Launay H, Huang W, Maberly SC, Gontero B. Regulation of carbon metabolism by environmental conditions: a perspective from diatoms and other chromalveolates. Front Plant Sci. Frontiers; 2020;11:1033.
20. Mekhalfi M, Puppo C, Avilan L, Lebrun R, Mansuelle P, Maberly SC, et al. Glyceraldehyde-3-phosphate dehydrogenase is regulated by ferredoxin-NADP reductase in the diatom *Asterionella formosa*. New Phytol. 2014;203:414–23.
21. Michels AK, Wedel N, Kroth PG. Diatom plastids possess a phosphoribulokinase with an altered regulation and no oxidative pentose phosphate pathway. Plant Physiol. 2005;137:911–20.
22. Thieulin-Pardo G, Remy T, Lignon S, Lebrun R, Gontero B. Phosphoribulokinase from *Chlamydomonas reinhardtii*: a Benson-Calvin cycle enzyme enslaved to its cysteine residues. Mol Biosyst. 2015;11:1134–45.
23. Wilhelm C, Büchel C, Fisahn J, Goss R, Jakob T, LaRoche J, et al. The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae. Protist. 2006;157:91–124.
24. Erales J, Gontero B, Maberly SC. Specifity and function of glyceraldehyde-3-phosphate dehydrogenase in a freshwater diatom, *Asterionella formosa* (bacillariophyceae). J Phycol. 2008;44:1455–64.
25. Clement R, Lignon S, Mansuelle P, Jensen E, Pophillat M, Lebrun R, et al. Responses of the marine diatom *Thalassiosira pseudonana* to changes in CO_2 concentration: a proteomic approach. Sci Rep. 2017;7:42333.
26. Chen X-H, Li Y-Y, Zhang H, Liu J-L, Xie Z-X, Lin L, et al. Quantitative proteomics reveals common and specific responses of a marine diatom *Thalassiosira pseudonana* to different macronutrient deficiencies. Front Microbiol. 2018;9:2761.
27. Gschloessl B, Guermeur Y, Cock JM. HECTAR: A method to predict subcellular targeting in heterokonts. BMC Bioinformatics. 2008;9:393.
28. Emanuelsson O, Nielsen H, Heijne GV. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. 1999;8:978–84.

29. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37:420–3.
30. Armenteros JJA, Salvatore M, Emanuelsson O, Winther O, Heijne G von, Elofsson A, et al. Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance. Life Science Alliance; 2019;2:e201900429.
31. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta. 2010;1804:996–1010.
32. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. Struct Lond Engl 1993. 2003;11:1453–9.
33. Mészáros B, Erdos G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 2018;46:W329–37.
34. Sormanni P, Camilloni C, Fariselli P, Vendruscolo M. The s2D method: simultaneous sequence-based prediction of the statistical populations of ordered and disordered regions in proteins. J Mol Biol. 2015;427:982–96.
35. McDonnell AV, Jiang T, Keating AE, Berger B. Paircoil2: improved prediction of coiled coils from sequence. Bioinforma Oxf Engl. 2006;22:356–8.
36. Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. BMC Bioinformatics. 2007;8:211.
37. Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. Biopolymers. John Wiley & Sons, Ltd; 2008;89:392–400.
38. Mori S, Abeygunawardana C, Johnson MO, van Zijl PC. 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. 1995;108:94–8.
39. Stejskal EO, Tanner JE. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J Chem Phys. 1965;42:288.
40. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6:277–93.
41. Lee W, Tonelli M, Markley JL. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinforma Oxf Engl. 2015;31:1325–7.
42. Eaton J, Bateman D, Haubert S, Wehbring R. GNU Octave version 4.2.1 manual: a high-level interactive language for numerical computations. 2017. Available from: https://www.gnu.org/software/octave/doc/v4.2.1/
43. David G, Pérez J. Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. J Appl Crystallogr. 2009;42:892–900.
44. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, et al. New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr. 2012;45:342–

50.
45. Receveur-Brechot V, Durand D. How random are intrinsically disordered proteins? A small angle scattering perspective. Curr Protein Pept Sci. 2012;13:55–75.
46. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46:W296–303.
47. Launay H, Receveur-Bréchot V, Carrière F, Gontero B. Orchestration of algal metabolism by protein disorder. Arch Biochem Biophys. 2019;672:108070.
48. Nygaard M, Kragelund BB, Papaleo E, Lindorff-Larsen K. An efficient method for estimating the hydrodynamic radius of disordered protein conformations. Biophys J. 2017;113:550–7.
49. Dunker AK, Babu MM, Barbar E, Blackledge M, Bondos SE, Dosztányi Z, et al. What's in a name? Why these proteins are intrinsically disordered. Intrinsically Disord Proteins. 2013;1:e24157.
50. Zhang Y, Launay H, Schramm A, Lebrun R, Gontero B. Exploring intrinsically disordered proteins in *Chlamydomonas reinhardtii*. Sci Rep. 2018;8:6805.
51. Yruela I, Contreras-Moreira B. Protein disorder in plants: a view from the chloroplast. BMC Plant Biol. 2012;12:165.
52. Yruela I, Contreras-Moreira B, Dunker AK, Niklas KJ. Evolution of protein ductility in duplicated genes of plants. Front Plant Sci. 2018;9:1216.
53. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol. 2004;337:635–45.
54. Schad E, Tompa P, Hegyi H. The relationship between proteome size, structural disorder and organism complexity. Genome Biol. 2011;12:R120.
55. Kastano K, Erdős G, Mier P, Alanis-Lobato G, Promponas VJ, Dosztányi Z, et al. Evolutionary study of disorder in protein sequences. Biomolecules. 2020;10:1413.
56. Brodsky S, Jana T, Mittelman K, Chapal M, Kumar DK, Carmi M, et al. Intrinsically disordered regions direct transcription factor in vivo binding specificity. Mol Cell. 2020;79:459-471.e4.
57. Fuxreiter M, Tompa P. Fuzzy complexes: a more stochastic view of protein function. Adv Exp Med Biol. 2012;725:1–14.
58. Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, et al. Protein phase separation: a new phase in cell biology. Trends Cell Biol. 2018;28:420–35.
59. Marri L, Sparla F, Pupillo P, Trost P. Co-ordinated gene expression of photosynthetic glyceraldehyde-3-phosphate dehydrogenase, phosphoribulokinase, and CP12 in *Arabidopsis thaliana*. J Exp Bot. 2005;56:73–80.
60. Lebreton S, Graciet E, Gontero B. Modulation, via protein-protein interactions, of glyceraldehyde-3-phosphate dehydrogenase activity through redox phosphoribulokinase regulation. J Biol Chem. 2003;278:12078–84.
61. Oesterhelt C, Klocke S, Holtgrefe S, Linke V, Weber APM, Scheibe R. Redox regulation of chloroplast enzymes in *Galdieria sulphuraria* in view of eukaryotic evolution. Plant Cell Physiol. 2007;48:1359–

73.
62. Matsumura H, Kai A, Maeda T, Tamoi M, Satoh A, Tamura H, et al. Structure basis for the regulation of glyceraldehyde-3-phosphate dehydrogenase activity via the intrinsically disordered protein CP12. Struct. 2011;19:1846–54.
63. Boggetto N, Gontero B, Maberly SC. Regulation of phosphoribulokinase and glyceraldehyde 3-phosphate dehydrogenase in a freshwater diatom, *Asterionella formosa*. J Phycol. 2007;43:1227–35.
64. Lupas A. Coiled coils: new structures and new functions. Trends Biochem Sci. 1996;21:375–82.
65. Burkhard P, Stetefeld J, Strelkov SV. Coiled coils: a highly versatile protein folding motif. Trends Cell Biol. 2001;11:82–8.
66. Parry DAD. Fifty years of fibrous protein research: a personal retrospective. J Struct Biol. 2014;186:320–34.
67. Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB, et al. Extreme disorder in an ultrahigh-affinity protein complex. Nature. 2018;555:61–6.
68. Junglas B, Orru R, Axt A, Siebenaller C, Steinchen W, Heidrich J, et al. IM30 IDPs form a membrane-protective carpet upon super-complex disassembly. Commun Biol. 2020;3:595.
69. Rackham OJL, Madera M, Armstrong CT, Vincent TL, Woolfson DN, Gough J. The evolution and structure prediction of coiled coils across all genomes. J Mol Biol. 2010;403:480–93.
70. Newman JR, Wolf E, Kim PS. A computationally directed screen identifying interacting coiled coils from *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 2000;97:13203–8.
71. Liu J, Rost B. Comparing function and structure between entire proteomes. Protein Sci. 2001;10:1970–9.
72. Jensen E, Clément R, Maberly SC, Gontero B. Regulation of the Calvin-Benson-Bassham cycle in the enigmatic diatoms: biochemical and evolutionary variations on an original theme. Philos Trans R Soc. 2017;372.

Figures

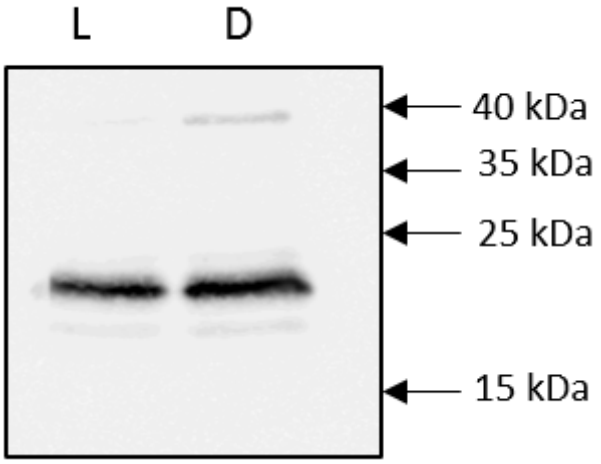

Figure 1

Western blot of CP12 expression under dark and light conditions. 20 µg crude extract from T. pseudonana cells grown under light (lane L) or dark (lane D). Molecular mass markers are shown on the right. After electrophoresis, the 12% SDS-gels are transferred onto nitrocellulose membranes and revealed with enhanced chemiluminescence with α-CP12 antibody 1:10 000.

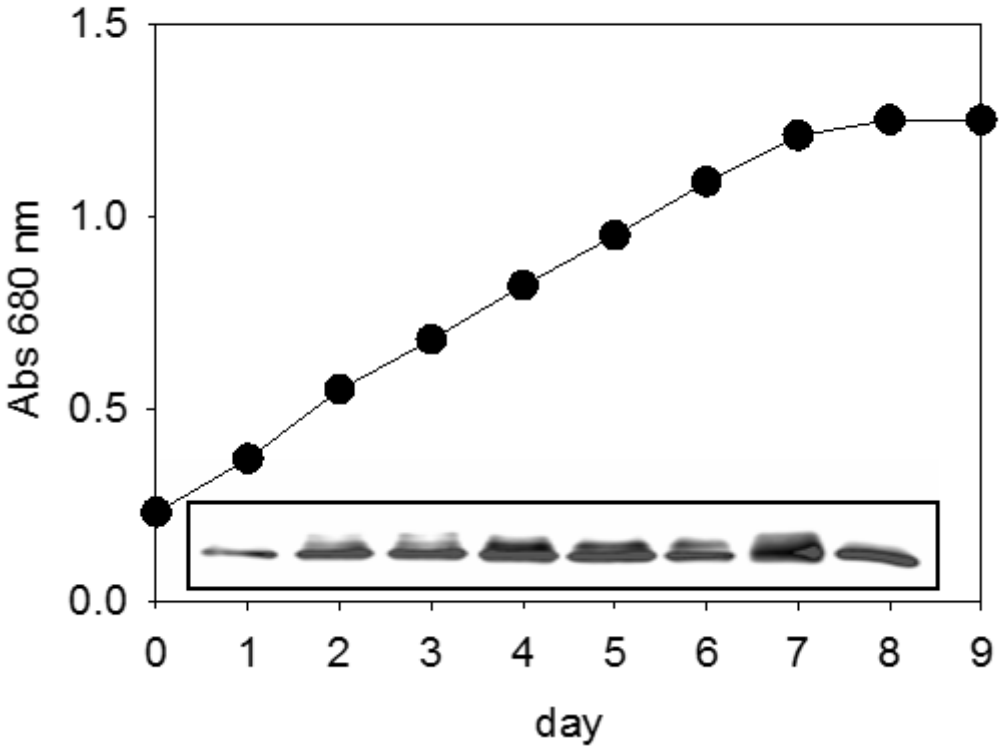

Figure 2

CP12 expression during the different growth phase. The growth of T. pseudonana in F/2+Si medium, under continuous light and with air-concentration of CO2 (400ppm) is followed by the absorbance at 680 nm. 30 µg of proteins from day 1 to 8 are loaded on SDS-PAGE gel. Proteins were transferred to a nitrocellulose membrane and revealed with antibodies raised against α-CP12 antibody.

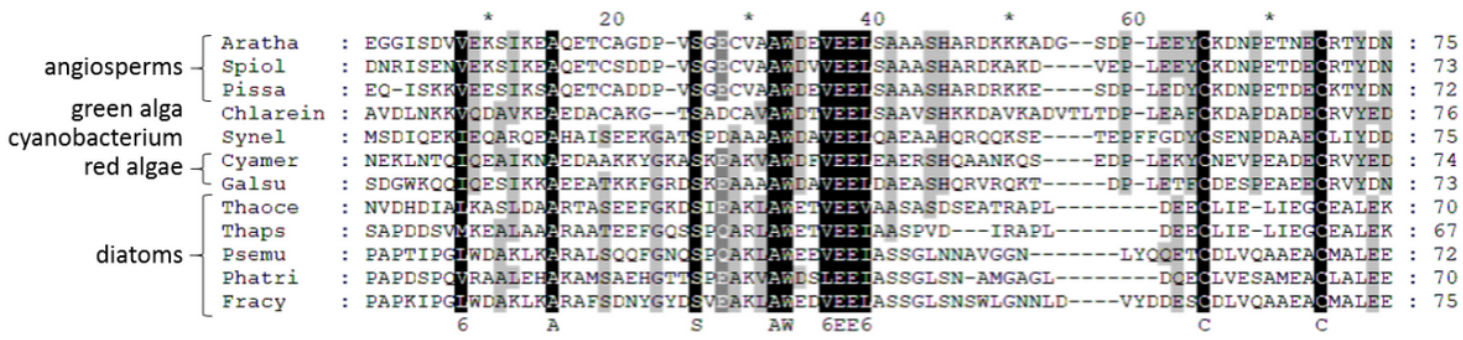

Figure 3

Alignment of CP12-like protein sequences from different species. Only partial sequences are shown. Species and protein IDs are the following: Aratha, A. thaliana (protein ID: NP_191800.1-NCBI, 131 aa); Spiol, Spinacia oleracea (protein ID: CAA96568.1-NCBI, 124 aa); Pissa, Pisum Sativum (protein ID: CAA96570.1-NCBI, 127 aa); Chlarein, Chlamydomonas reinhardtii (protein ID: XP_001694345.1-NCBI, 107 aa); Synel, Synechococcus elongatus (protein ID: Q6BBK3-UniProt, 75 aa); Cyamer, Cyanidioschyzon merolae (protein ID: BAM82525.1-NCBI, 122 aa); Galsu, Galdieria sulphuraria (protein ID: CAI34858.1-NCBI, 129 aa); Thaoce, Thalassiosira oceanica (protein ID: 95994-JGI, 165 aa); Thaps, T. pseudonana (protein ID: EED96413.1-NCBI, 197 aa); Psemu, Pseudo-nitzschia multistriata (protein ID: 292992-JGI, 256 aa); Phatri, Phaeodactylum tricornutum (protein ID: 44297-JGI, 329 aa); Fracy, Fragilariopsis cylindrus (protein ID: 244965-JGI, 234 aa). Alignments were performed with ClustalW, using MEGA7 software, and the figure was processed with GeneDoc (http://www.nrbsc.org/gfx/genedoc) with similar residues shaded using the conservation mode. The signal peptides are not included in the alignment.

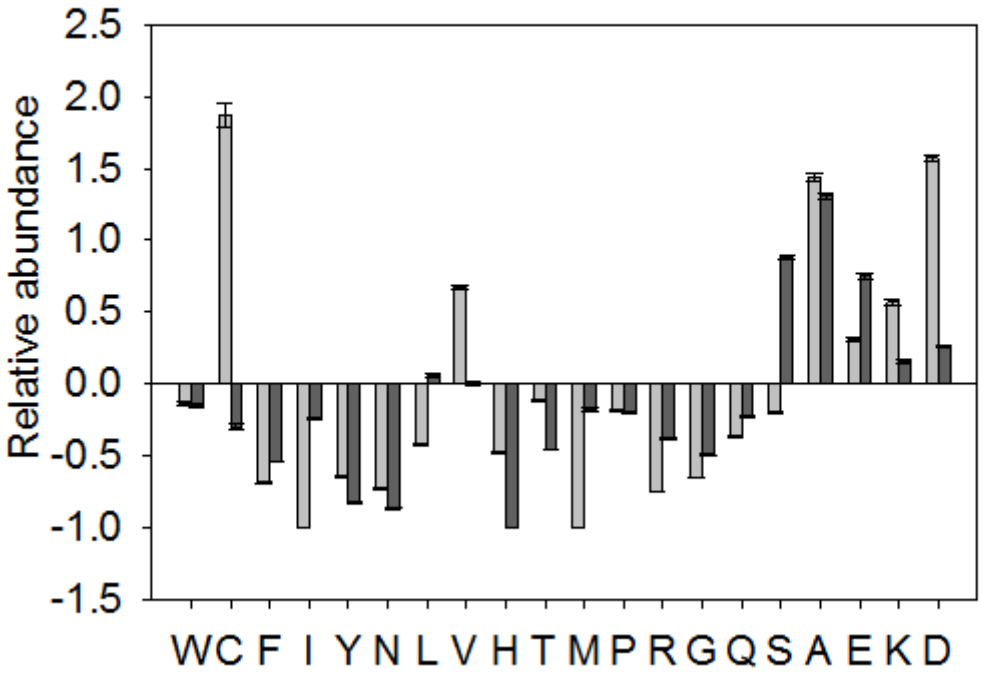

Figure 4

Amino acid composition of CP12s. The abundance of amino-acids of both T. pseudonana CP12 (black bars) and C. reinhardtii CP12 (grey bars) were compared against proteins from the PDB database using Composition Profiler (http://www.cprofiler.org/cgi-bin/profiler.cgi).

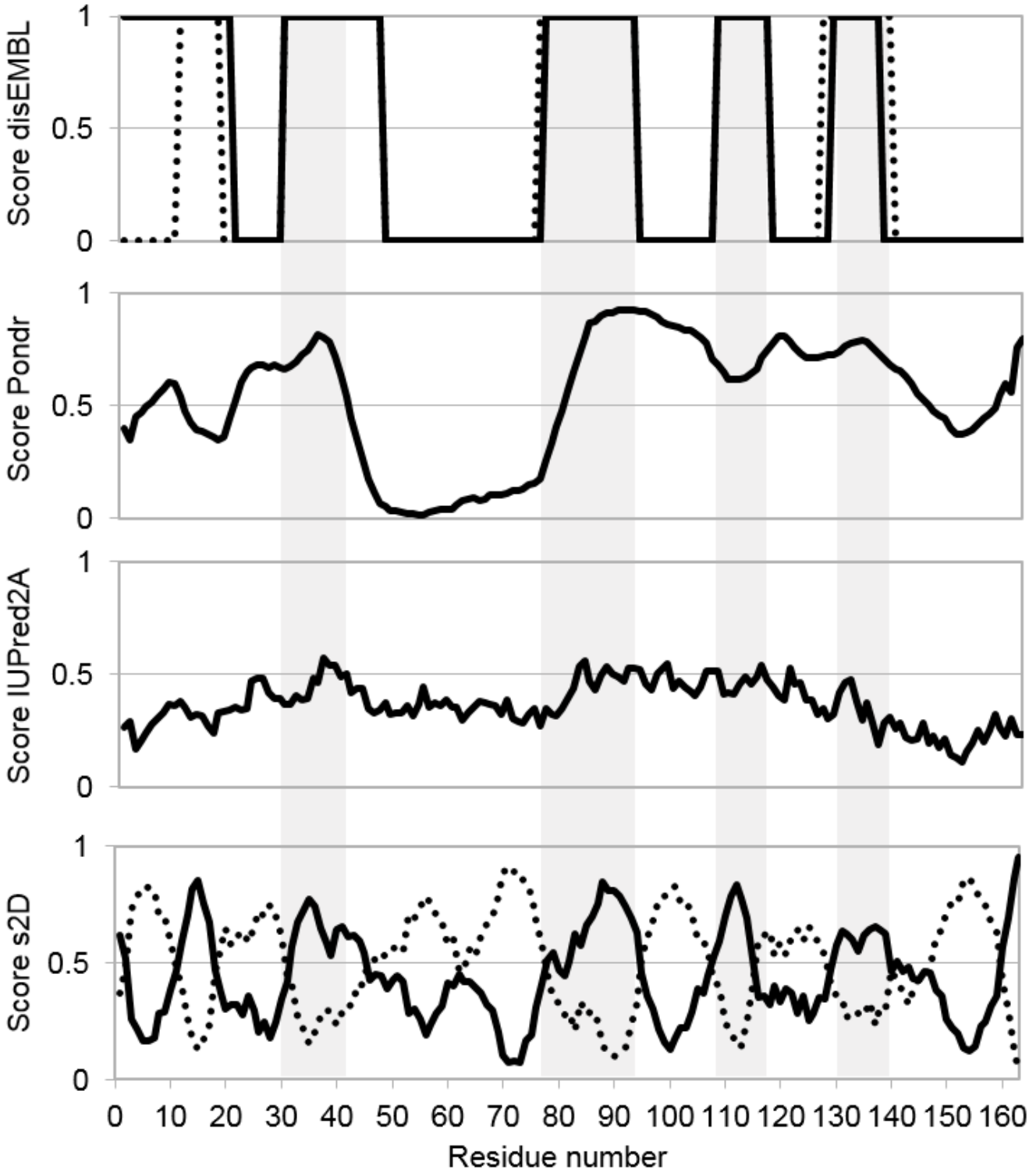

Figure 5

Prediction of disordered region. The sequence on the x-axis is numbered according to the protein without tag and starting at AAIEAA. The residues propensity to be in a disordered region is shown as a score varying from 0 to 1, and these are predicted using, from top to bottom: disEMBL (for hot loop in continuous line and for coil in dotted line), PONDR-FIT, IUPred2A and s2D (continuous line). For most

pensity to form α-helices shown with dotted line.

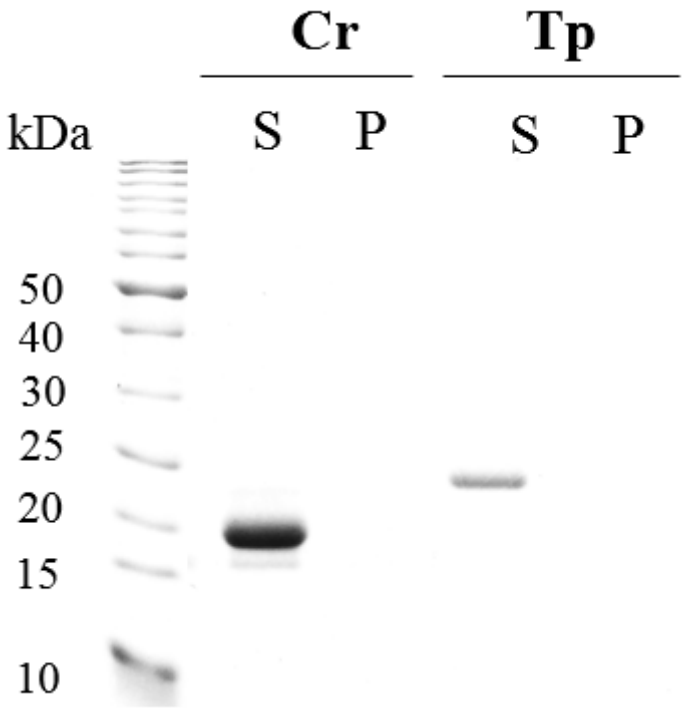

Figure 6

Heat resistance of CP12 from C. reinhardtii (Cr, left) and T. pseudonana (Tp, right). 1 µg of recombinant CP12 was heated at 95°C for 5 minutes, and centrifuged. The pellets (P) were resuspended with the same volume as supernatant (S), and loaded on 12% SDS-PAGE. Molecular weight markers are shown on the left.

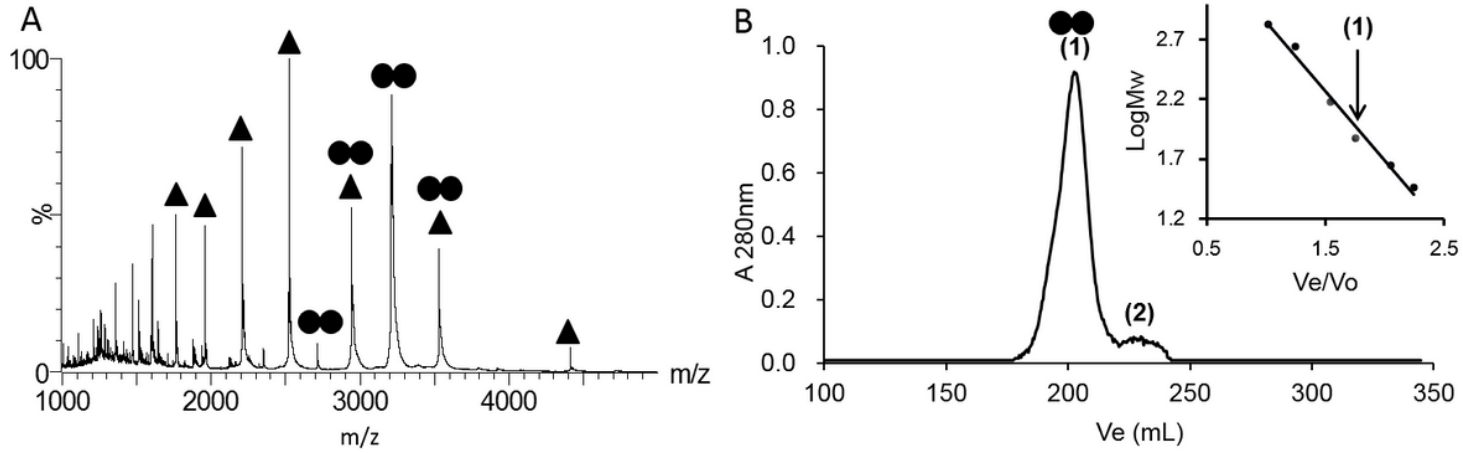

Figure 7

Oligomerization of CP12 revealed by native ESI-MS and size-exclusion chromatography. (A) ESI-MS spectrum of CP12. Signals labelled with a single triangle are assigned to monomeric CP12 (envelope of m/z 1765.2 to 4411.4 with charged states ranging from 4 to 10) , and signals labelled with two circles are assigned to dimeric CP12 (envelope of m/z 2715.1 to 3529.4 with charged states ranging from 10 to 13).

(B) Elution profile of CP12 on a S200 size-exclusion column. The single elution peak is assigned to dimeric CP12 (labelled with 1). A small proportion of a smaller particle is assigned with the labelled (2) and is ascribed to impurities. The column was calibrated with proteins of known relative molecular mass (insert): Thyroglobulin (670 kDa), Ferritin (440 kDa), Alcohol dehydrogenase (150 kDa), Conalbumin (75 kDa), Ovalbumin (44 kDa), carbonic anhydrase (30 kDa). The linear fit (R2 = 0.98) indicates a dependency of the MW as a function of the elution volume (Ve) as: MW= ⟦10⟧^(-1.16 Ve/Vo+4) with Vo being the elution volume of blue dextran (void volume of 110.3 mL). The elution volume of CP12 is indicated with an arrow.

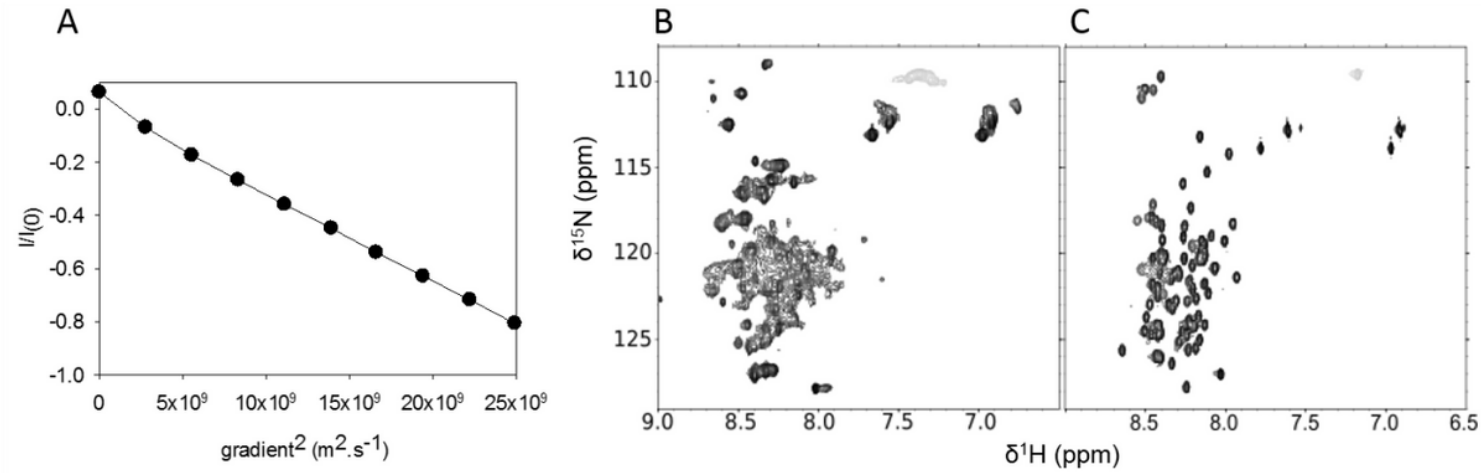

Figure 8

NMR spectrometric characterization of CP12: a dynamic dimer. (A) DOSY-NMR analysis of CP12, on the x-axis is the gradient strength and on the y axis is the normalized intensity on a ln scale (refer to Material and Methods section). A linear dependency indicates a homogenous oligomerization of the protein. The linear fit indicates a diffusion coefficient of 3.8 (± 0.1) 10-11 m2.s-1, which relates to a hydrodynamic radius (Rh) of 3.4 (± 1) nm. (B) and (C) show the 1H-15N fHSQC spectra of CP12 from T. pseudonana and C. reinhardtii [9] respectively.

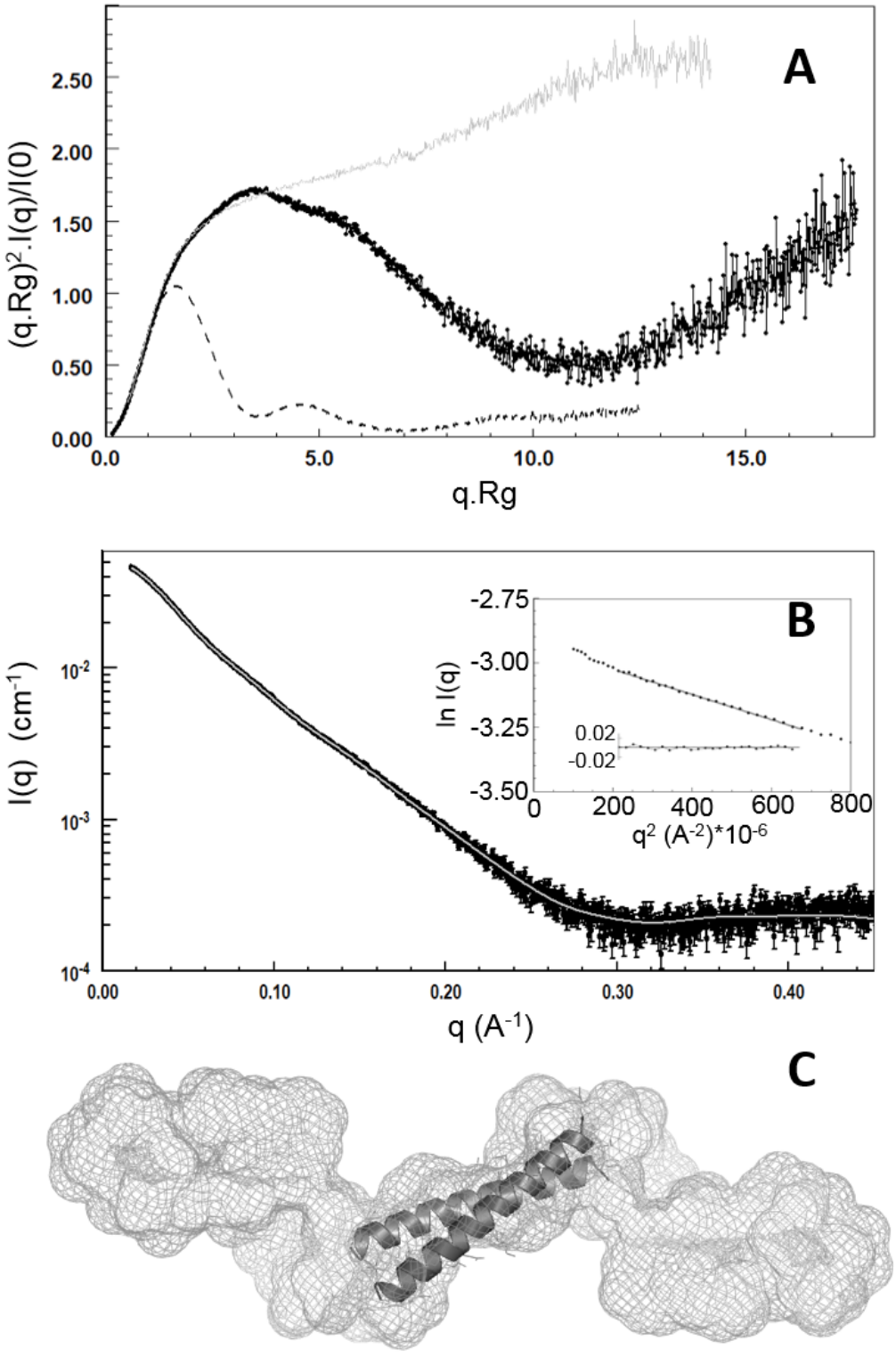

Figure 9

SAXS characterisation of CP12: an extended dimer. (A) Normalised and dimensionless Kratky plot. The data for CP12 is shown in black, and can be compared with that for C. reinhardtii CP12 [9] (grey), a representative of a random-coil polymer. The scattering data for C. reinhardtii GAPDH is shown with dotted line and is representative of what is expected for a globular protein. (B) Scattering intensity for erlaid with the DAMMIF fit (grey) related to the shape shown

in (C). The coiled coil structure calculated by Swiss-Model is inserted in the SAXS envelope (C). The insert in B shows the Guinier analysis together with the residuals.

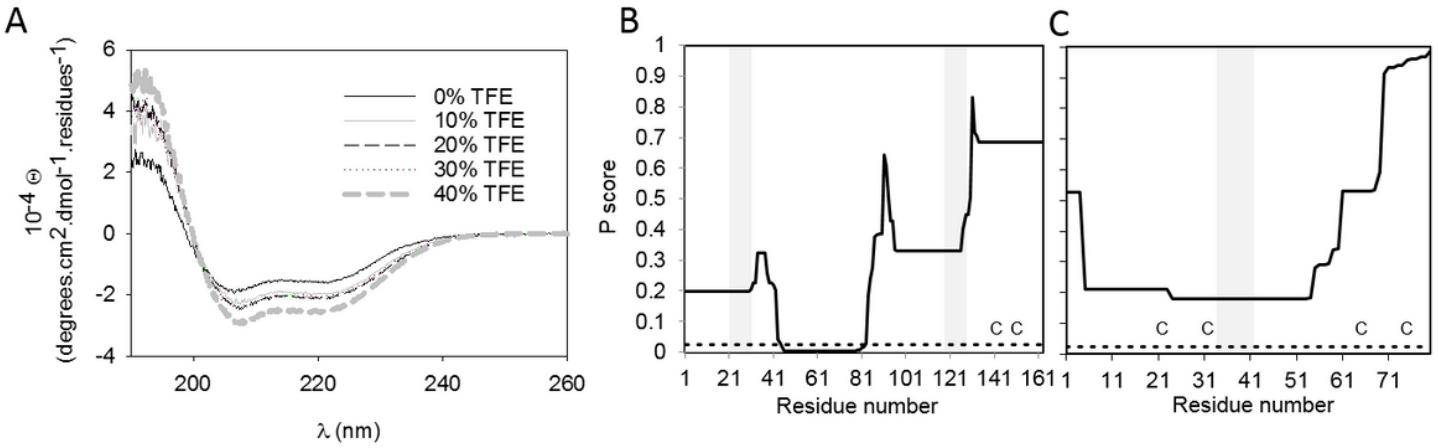

Figure 10

CD characterisation of CP12: high helical content and putative coiled coil arrangement. (A) CD spectra of 10 µM CP12 in 50 mM phosphate buffer pH 6.5 with increasing concentration of TFE. These spectra indicate a high proportion of helical structure, which increases further in the presence of TFE. (B and C) The Pscore of each T. pseudonana CP12 residues (B) and C. reinhardtii CP12 residues (C) to be in a coiled coil region is shown as a score varying from 0 to 1 (a score of 0 indicates a high propensity to be in a coiled coil region). The threshold for a region to have a high probability to form a coiled coil structure is indicated with a dotted line (Pscore < 0.025). The AWD_VEEL regions are highlighted in grey, and the position of the cysteine residues are indicated with a C.