Summary of integrative structure determination of Chloroplastic protein CP12 from the diatom Thalassiosira pseudonana (tp-CP12, homodimer) (PDB ID: 9A9R | pdb_00009a9r)

1. Model Composition	
1.1. Entry composition	CP12 domain-containing protein: chain(s) A, B (163 residues)
1.2. Datasets used for modeling	- EPR data, Zenodo: 10.5281/zenodo.15117282 - SAS data, Zenodo: 10.5281/zenodo.15117282 - De Novo model, Zenodo: 10.5281/zenodo.15117282
2. Representation	
2.1. Number of representations	1
2.2. Scale	Atomic
2.3. Number of rigid and flexible segments	0, 2
3. Restraints	
3.1. Physical principles	Information about physical principles was not provided
3.2. Experimental data	- 5 unique DerivedDistanceRestraint: restraint type HarmonicDistanceRestraintEPR data - 1 unique SASRestraint: Assembly name: Chloroplastic protein CP12 Fitting method: None Multi-state: False
4. Validation	
4.2. Number of ensembles	0
4.3. Number of models in ensembles	Not applicable
4.4. Number of deposited models	1
4.5. Model precision	Not available
4.6. Data quality	Data quality has not been assessed
4.7. Model quality: assessment of atomic segments	- Clashscore: 1.45 - Ramachandran outliers: 6 - Sidechain outliers: 10
4.8. Fit to data used for modeling	Fit of model to information used to compute it has not been determined
4.9. Fit to data used for validation	Fit of model to information not used to compute it has not been determined
5. Methodology and Software	
1. 5.1. Method name	ab initio modeling

5.3. Method description	AlphaFold modelling: Structure prediction for the CP12 homodimer was performed using AlphaFold v2.1.1-Multimer (AF2) and the default databases The sequence of mature CP12 was used in the AF2 input file for multiple sequence alignments (MSAs) lookup and structural template matching. The predicted local distance difference test score (pLDDT,) was used to estimate the per-residue confidence of the models. The percentages of the secondary structure elements of the models were assessed using YASARA View22. Modification of the selected AF2 model was performed using PyMOL by a rotation around N-Ca bond of residue I134 and by respecting the Ramachandran plot. Structure prediction of T. pseudonana PRK was performed using AlphaFold3 webserver (Abramson, J et al. Accurate structure prediction of biomolecular interactions with AlphaFold3. Nature (2024) based on the sequence provided by UNIPROT (entry B8BZ40) devoid of the first 15 residues, which are predicted with a score of 0.96 to form the chloroplast transit peptide by HECTAR v1.3 (https://webtools.sb-roscoff.fr/root?tool_id=abims_hectar HECTAR: A Method to Predict Subcellular Targeting in Heterokonts", Gschloessl et al., BMC Bioinformatics, 2008 (PubMed ID 18811941)) specifically trained for heterokonts.
2. 5.1. Method name	sampling
5.2. Method type	harmonic restrained MD simulations
5.3. Method description	Harmonic restrained MD simulations were run with the Amber 20 package based on the modified AF2 model. The Amber ff19SB force field was used for the protein, which was immersed into a cubic periodic box filled with water solvent molecules modeled by the OPC model with the LEaP program of AmberTools20. All charged residues were considered in their standard protonation state at pH = 7.0. We added 28 sodium ions to neutralize the simulation box and the total number of atoms became 178352. Nonbonded particle-particle interactions were considered with a distance cutoff of 9 and the long-range electrostatic interactions were treated with the particle mesh Ewald (PME) method. After minimization, heating, and equilibration of the simulation system, several rounds of MD simulations were conducted imposing harmonic restraints on the CA-CA distances across the dimer for respective spin-labeled residues S39, S46, S56, S83, and C150 with a spring force of 30 kcal/(mol2). Each round ran for 100ps with time step of 2 fs at T = 300 K and P = 1 atm. The trajectories and snapshots were saved every 1 ps. The initial/target distances for each CA atom pair were 73 /68 for S39, 57 /43 for S46, 26 / 32 for S56, 51 /64 for S83, 12 /25 for C150, respectively.
5.4. Number of computed models	1
3. 5.1. Method name	sampling
5.2. Method type	restrained-ensemble MD simulation

5.3. Method description	The following reMD simulations were run by using a modified version of NAMD 2 with all-atom CHARMM36m protein force field. 25 copies of all-atom MTSL (CYR1 in CHARMM-GUI) spin labels were attached to residues S39, S46, S56, S83 and C150, respectively, with reMD Prepper in vacuum for saving the computational resources. The force field of all-atom CYR1 spin label is provided by CHARMM-GUI. Spatial overlap among the 25 copies of CYR1 spin labels was allowed by neglecting the interactions among them. Five independent all-atom reMD simulations were conducted at 303.15 K with different random number seeds using Langevin dynamics with a damping coefficient of 5 ps-1. Before each reMD production run, we performed minimization and equilibration, during which only sidechain atoms were relaxed, keeping the backbone atom positions with harmonic restraints of 2 kcal/(mol2) imposed on them. For these structural refinement, we switched off the harmonic restraints imposed on the backbone atoms, and conducted production runs of reMD simulations for 2 ns with a time step of 0.5 fs. The long-range electrostatic interactions were treated with the particle mesh Ewald (PME) method and the nonbonded interactions were truncated at a 10 distance cutoff. The distance distributions of each spin label pair were restrained with a force constant of 100 kcal/(mol2) toward the experimental distance distribution histograms with a bin width of 0.025 nm. The atomic coordinates of each reMD production run were saved every 1ps. Since each spin label has 25 copies, a total number of 625 distances for each spin label pair were obtained from a single snapshot of trajectories and a total of 1, 250, 000 data points were yielded for every spin label pair from one single reMD production run.
5.4. Number of computed models	1435
4. *5.1. Method name*	Not available
5.3. Method description	The saved atomic coordinates of reMD productions were then removed of CYR1 spin-labels, and then used as an ensemble of structures to simulate the SAXS data.
5. *5.1. Method name*	Not available
5.3. Method description	DEER distance distributions were extracted from raw experimental data using DeerAnalysis2019 software
5.5. Software	- AlphaFold Multimer (version 2.1.1) - Amber (version 20) - modified version of NAMD (version 2) - CRYSOL (version 3.1) - DeerAnalysis (version 2019)