Cryo-EM Structure of a Tetrameric Photosystem I from *Chroococcidiopsis* TS-821, a Thermophilic, Non-heterocyst-forming Cyanobacteria

Dmitry Semchonok

Groningen Biomolecular Sciences & Biotechnology Institute, University of Groningen

Jyotirmoy Mondal

University of Tennessee Knoxville

Connor Cooper

University of Tennessee Knoxville

Katrina Schlum

University of Tennessee Knoxville

Meng Li

University of Washington <https://orcid.org/0000-0001-9831-8575>

Muhamed Amin

5Department of Sciences, University College Groningen

Carlos Sanchez Sorzano

Biocomputing Unit, Centro Nacional de Biotecnología (CNB-CSIC) <https://orcid.org/0000-0002-9473-283X>

Emey Ramírez-Aportela

National Center for Biotechnology (CSIC)

Panagiotis Kastiris

Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg

Egbert Boekema

Biophysical Chemistry

Albert Guskov

University of Groningen <https://orcid.org/0000-0003-2340-2216>

Barry Bruce (✉ bbruce@utk.edu)

University of Tennessee at Knoxville <https://orcid.org/0000-0002-4045-9815>

Article

Keywords: Oxygenic Photosynthesis, Dimer-of-dimers Organization, Excitation Energy Pathway

Posted Date: October 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-81702/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Title

Cryo-EM Structure of a Tetrameric Photosystem I from *Chroococcidiopsis* TS-821, a Thermophilic, Non-heterocyst-forming Cyanobacteria.

Authors

Dmitry A. Semchonok^{1,#,+}, Jyotirmoy Mondal^{2,+}, Connor J. Cooper³, Katrina Schlum³, Meng Li^{2,4,†}, Muhamed Amin⁵, Carlos O. S. Sorzano^{6,7}, Erney Ramírez-Aportela⁶, Panagiotis Kastiris⁸, Egbert J. Boekema¹, Albert Guskov¹ and Barry D. Bruce^{2,3,4,9*}

Affiliations

¹Groningen Biomolecular Sciences & Biotechnology Institute, University of Groningen, Groningen, NLD.

²Biochemistry & Cellular and Molecular Biol. Dept., University of Tennessee, Knoxville, TN, USA.

³Program in Genome Science and Technology, University of Tennessee, Knoxville, TN, USA.

⁴Bredesen Center for Interdisciplinary Research & Education, University of Tennessee, Knoxville, TN, USA.

⁵Department of Sciences, University College Groningen, Groningen, NLD

⁶Biocomputing Unit, National Center for Biotechnology (CSIC), Darwin 3, Campus Univ. Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain.

⁷Univ. CEU San Pablo, Campus Urb. Montepríncipe, Boadilla del Monte, 28668 Madrid, Spain.

⁸Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg,
Halle/Saale, Germany.

⁹Microbiology Dept., University of Tennessee, Knoxville, TN, USA.

*Correspondence to: Barry D. Bruce (bbruce@utk.edu), Phone +1(865)974-4082

⁺Denotes equal contribution

[#]Current address: Institute of Biochemistry and Biotechnology, Martin Luther University Halle-
Wittenberg, Halle/Saale, Germany

[†]Current address: School of Oceanography, University of Washington, Seattle, WA, USA.

Abstract

Photosystem I (PSI) is one of two the photosystems involved in oxygenic photosynthesis. PSI of cyanobacteria exists in monomeric, trimeric, and tetrameric forms, which is in contrast to the strictly monomeric form of PSI in plants and algae. The tetrameric organization raises questions about its structural, physiological, and evolutionary significance. Here we report the ~ 3.9 Å resolution cryo-EM structure of tetrameric PSI from the thermophilic, unicellular cyanobacterium *Chroococcidiopsis* sp. TS-821. The structure resolves all 44 subunits and 448 cofactor molecules. We conclude that the tetramer is arranged via two different interfaces resulting from a dimer-of-dimers organization. The localization of chlorophyll molecules permits an excitation energy pathway within and between adjacent monomers. Bioinformatics analysis reveals conserved regions in PsaL subunit that correlate with the oligomeric state. Tetrameric PSI may function as a key evolutionary step between the trimeric and monomeric forms of PSI organization in photosynthetic organisms.

Introduction

Oxygenic photosynthesis is a unique energy conversion process performed by plants, algae, and cyanobacteria (1, 2) where photons from the sunlight are converted into chemically fixed energy by synthesizing carbohydrates, generating oxygen as a side-product of water splitting. Oxygen production and carbon dioxide fixation into organic matter performed by photosynthetic organisms determines the composition of Earth's atmosphere and provides all life forms with essential food and fuel (3, 4). Oxygenic photosynthesis of cyanobacteria, algae, and plants is catalyzed by four defined membrane complexes: PSI, photosystem II (PSII), cytochrome b_6/f complex, and CF₁-ATPase, which are the major components of the electron transport chain (ETC). Both PSI and PSII

are large multi-subunit membrane-embedded pigment-protein complexes composed of a core complex – the reaction center, where electron transport is initiated, and of a peripheral antenna system, which is essential for light-harvesting and regulation of photosynthetic activity (5). The reaction center and peripheral antenna system work in concert to carry out the light conversion steps that ultimately lead to the production of ATP by ATP-synthase and reduction of NADP^+ to NADPH with the concomitant release of oxygen as a result of water oxidation.

The two photosystems function in series to couple the oxidation of water to enable the generation of proton motive force for ATP synthesis and reduction of NADP^+ . NADPH and ATP then support carbon fixation via the Benson-Bassham-Calvin cycle that supports all life on the planet Earth (6). PSII is capable of generating powerful oxidation states that drive oxygen evolution via the water-splitting complex (7, 8). PSI is the second photosystem in ETC. PSI includes the special pair, P700, and between 96 to 112 antennae chlorophyll (Chl) *a* molecules (9, 10) that function to increase the optical cross-section for excitation with a subsequent high-efficiency energy transfer to the special pair. It receives electrons from the b_6/f complex via cytochrome *c* or plastocyanin in oxygenic photosynthetic organisms (1). Upon photoexcitation, photooxidized PSI transfers an electron from primary electron donor P700^+ (a special pair of Chl *a/a'* molecules) via its internal electron acceptors (A_0 , A_1 , F_x , F_A , and F_B) to the terminal one – ferredoxin (11). The P700^+ is then re-reduced via electron donation from its primary donor (plastocyanin or cytochrome *c*), connecting PSI to PSII via the b_6/f complex. The structure of the cyanobacterial PSI complex has been known for nearly two decades (9) and only recently, the structures of PSI from plants and algae have been resolved (12, 13).

Electron micrographs of PSI from the cyanobacteria *Synechococcus* sp. provided the first evidence for trimeric PSI over three decades ago (14). Further studies focused on the diverse filamentous

and unicellular cyanobacteria, including the most primitive known cyanobacterium, *Gloeobacter violaceus* PCC 7421 (14-17). Eventually, the trimeric PSI structure was resolved at 2.5 Å by X-ray crystallography from the thermophilic cyanobacteria *Thermosynechococcus elongatus* BP-1 (*T.e.* BP-1) (PDB ID: 1JB0), confirming the conclusions about the trimeric structure of PSI (9). As a result of these early seminal reports, PSI was initially believed to assemble into stable trimeric structures in cyanobacteria, as opposed to the monomeric form observed in all plants and algae. Since then, further evidence for trimeric PSI has been observed in the 2.5 Å crystal structure of the mesophilic cyanobacteria *Synechocystis* sp. PCC 6803 (*Syn* PCC 6803) (PDB ID: 5OY0) (18) and recently in atomic force microscopy analysis of multiple ecotypes of *Prochlorococcus* (19), suggesting a prevalent occurrence of trimeric PSI in cyanobacteria.

In cyanobacteria, a PSI protomer comprises twelve different subunits, and in many species, the total mass of a trimer is ~1 MDa (9, 10, 20). However, this initial belief of PSI trimer being the sole oligomeric state in cyanobacteria has recently been challenged by observation of the tetrameric form of PSI in two cyanobacteria *Anabaena* sp. PCC 7120 (21, 22) and *Chroococcidiopsis* sp. TS-821 (TS-821) (23). However, the PSI tetramer was not considered as a major oligomeric state in cyanobacteria. The physiological and evolutionary significance of this tetrameric state has yet to be elucidated, nor the mechanism driving this oligomerization state and sustaining its stability is known. The recent cryo-EM structure of PSI tetramer from TS-821 revealed that tetrameric PSI is actually a dimer-of-dimers with two different interaction interfaces between monomers (24). This structure suggests that subtle changes in the placement of the central PsaL subunit yield changes in helical bundling that have been implicated to be critical in the formation of PSI trimers.

To date, no crystal structure of the tetrameric PSI is available. Although two cryo-EM structures of the tetrameric form of PSI has been reported in a heterocyst-forming cyanobacterium, *Anabaena* (21, 22), our lab has observed that a tetrameric PSI organization is very widespread being found in most of the Heterocyst-forming cyanobacteria and their Close Relatives (HCR) after investigating 61 different cyanobacteria (25). These organisms have been frequently proposed as the likely chloroplast progenitor (26). However, other reports suggest that maybe a nitrogen-fixing unicellular cyanobacteria similar to *Chroococcidiopsis* or other members of Chroococcales may be the plastid progenitor (27). The apparent occurrence of tetrameric PSI oligomers in all forms (mentioned above) of organisms suggests the primordial existence of tetrameric PSI in the earliest plastid ancestor. It is perhaps an intermediate in the evolution of monomeric forms of PSI in algae and plants.

This work advances in the understanding of the structural basis of the tetrameric form of PSI in non-heterocyst forming cyanobacteria by elucidation its structure in TS-821 by cryo-EM. Structure analysis visualizes the dimer-of-dimers formation, defines the correlation between the structural changes in PsaL subunit with the variations in oligomeric state, and describes the structural relationship between the novel tetrameric PSI organization with the known trimeric one. Our study allows not only a direct comparison of the TS-821 PSI tetramer with previous trimeric PSI crystal structures from *T. e.* BP-1 and *Syn* PCC 6803, but also enable the first comparison of tetrameric PSI structures within two different cyanobacteria. Finally, the bioinformatic analysis revealed multiple conserved regions of PsaL that may be critical to the formation of these different oligomeric PSI states.

Results

Initial characterization of the thylakoid membranes of TS-821 by blue native polyacrylamide gel electrophoresis (BN-PAGE) observed a larger PSI complex (**Fig. 1A**) (25). We have extended this BN-PAGE analysis using a range of 8 other common detergents, known for their ability to maintain membrane proteins in their native conformation (data not shown) where we observed the same tetrameric complex. This work confirms the presence of a tetrameric form of PSI in a non-heterocyst-forming cyanobacteria and argues against it being a detergent artifact. Additionally, using multiple imaging methods of TS-821 cells revealed its unicellular and non-heterocyst morphology (**Fig. 1B-E**). All of these images clearly show that these cells are not filamentous yet exist either as single cells or in some cases as 2, 4, and 8 cell aggregates. The shape of these cell aggregates suggests that it undergoes binary fission in multiple planes and can lead to highly symmetric octamers of cells that indicate possible coordination of cell divisions in multiple planes. The scanning electron microscopy (SEM) images (**Fig. 1B**) also revealed a thick fibrous material on the surface of the large aggregated cells that was not as abundant as on nonaggregate cells. This thick fibrous material is consistent with the appearance of polysaccharides under transmission electron microscopy (TEM) (**Fig. 1E**). TEM imaging showed a thin section of two cells revealing thick outer sheath composed of a fibrous outer cell wall layer or F-layer, as initially observed in the *Pleurocapsales* (28).

Observation of cells under bright field microscopy revealed that TS-821 cells exist primarily as one or two cells or in larger aggregates (**Fig. 1C**). Cells were also stained with DAPI, a DNA specific fluorescent dye, in order to observe the location of DNA within cells during division (**Fig. 1D**). The cell undergoing binary fission exhibits fluorescence throughout the entire interior portion of the cell, suggesting that DNA is distributed throughout the cell (**Fig. 1D**). However, distinct

regions of fluorescence were seen within adjacent cells dividing via multiple fissions. These distinct globular regions of DNA are most likely small daughter cells that are the result of multiple fissions. All of this imaging supports the original morphological classification of TS-821 as a member of the order Pleurocapsales (28).

To analyze the phylogeny of strain TS-821, a phylogenetic tree was generated as described in materials and methods. It is also important to note that genera representing all major orders within the phylum Cyanophyta are included within this phylogenetic analysis. **Fig. 1F** shows the resulting unrooted consensus tree with branches containing bootstrap values where values below 50% were excluded are shown. Bootstrap values are interpreted as the confidence level in the arrangement of the branch for which they are listed. Phylogenetic analysis of TS-821 revealed that TS-821 does not group close to the other *Chroococcidiopsis* *sp.*, but rather with other unclassified cyanobacteria with high confidence. However, it is distant from other members of the Pleurocapsales and the phylogenetic 16S tree indicates that TS-821 is much closer to heterocyst forming cyanobacteria such as *Nostoc*, *Anabaena*, and *Fischerella* as previously reported (23). However, the 16S sequence strongly suggests that it is distinct from the other more well-characterized members of the *Chroococcidiopsis* genus and may require reclassification in the future.

Cryo-EM and Model building

Although our earlier cryo-EM low-resolution structure suggested TS-821 tetramer was a dimer-of-dimers, this resolution prevented us from investigating the structural basis of this unique symmetry (24). To have a better insight into the organization of tetrameric PSI from TS-821 we conducted the single-particle cryogenic electron microscopy analysis (**Fig. S1A**). 2D particle classification of 325,648 particles yielded 16 different 2D classifications (**Fig. S1B**) which upon relion

3D classification/refinement yielded 63,131 particles (**Fig. S1C**). From this classification, we obtained ~3.9 Å resolution structure (**Fig. S1D-H**).

The local resolution of the final map varies from 3.5 Å to 5.5 Å (**Fig. S1E**) with high resolution within the transmembrane core of each monomer composed of multiple PsaA and PsaB helices, suggesting increased protein stability and less conformational flexibility. In addition, the interfacial subunits PsaL, PsaI and PsaM at one of the dimeric interfaces contain the best resolution distribution within the map, possibly reflecting higher stability.

To reconstruct a tetramer of PSI, the single protomer of cyanobacterial PSI (PDB ID: 1JB0) was manually placed in the cryo-EM map using Chimera (29). This rough placement was followed by the rigid-body refinement of each subunit in Phenix (30). For all subunits we mapped density for all known PSI subunits except for PsaX, which had a very fragmented and weak density. This could indicate that this subunit may have been partially lost during the sample preparation. Thus, PsaX was not included in further modeling. The sequence was manually adjusted during the modeling in Coot (31). The positions of Chl *a* molecules from 1JB0 structure were used as reference points to model Chl *a* in the tetramer. The map resolution was not sufficient to model the positions of carotenoids. The manually rebuilt model underwent several rounds of real-space refinement in Phenix (30), including NCS restraints, secondary structure restraints, and simulated annealing protocol. The positions of side chains were modelled according to the density, except for the ambiguous situations, where the most common rotamers were utilized.

Overall Structure and Placement of Subunits

As we have previously shown (24), the tetramer is organized as a dimer-of-dimers with two different interfaces: A-B and B-A' (**Fig. 2A, S3A-B**). Therefore, the structure is pseudotetrameric and it has C2 and not C4 symmetry. The obtained cryo-EM structure of the tetrameric PSI of TS-

821 resolves all 44 subunits: four each of PsaA, PsaB, PsaC, PsaD, PsaE, PsaF, PsaI, PsaJ, PsaK, PsaL, and PsaM (**Fig 2B-C**). The inner cavity is surrounded by four PsaL, two PsaM, and two PsaI subunits. The relative positions of the subunits are identical in the individual monomers. However, the tetramer has a dual symmetry where the positioning of the monomers A and B are identical to that of monomers A' and B' supporting the hypothesis that tetrameric PSI is a dimer-of-dimers. The PsaL subunits are closely associated with each other between monomers A-B and A'-B' (**Fig. 2B**). Interestingly, PsaL subunits in adjacent monomers B-B' and A-A' are not in contact with each other. PsaM and PsaI subunits of monomers A and A' are oriented towards the inside cavity of the tetramer. In monomers B and B' these two subunits are buried in the interface between monomer A-B and A'-B'.

Cofactor placements

Based on the densities and our molecular modelling, we are able to place all of the known light-harvesting Chls and electron transport components. There are three [4Fe-4S] clusters (F_A , F_B , and F_X) in each monomer. F_A is located in PsaA subunit and both F_X and F_B are located in the PsaC subunit (**Fig. 2D-E**). In the resolution of our model, we could not unambiguously identify any carotenoids or lipids and therefore not deposited in the PDB dataset. However, based on unresolved densities and comparative analysis of the carotenoids in the *T.e.* BP-1 PSI crystal structure (PDB ID- 1JB0), we have been able to putatively place both phylloquinones and carotenoids in our model. Each monomer has two phylloquinones and 16 β -carotene molecules that were manually placed (**Fig. 2D-E**). However, the identity of these carotenoids could not be matched to the prior chemically identified carotenoids (25). There are a total of 362 Chl *a* molecules in the tetramer, 92 in monomers A and A' and 91 in monomers B and B'. The special pair of Chl *a* is located at the center of the PsaA-PsaB interface, towards the luminal side (**Fig. 2D-E**).

Closely placed Chl *a* molecules with Mg²⁺-Mg²⁺ distance < 10 Å are found primarily within individual monomers but are not found in any of the four interfaces between monomers (**Fig. S2A**). A pair of Chl *a* molecules has one Mg²⁺-Mg²⁺ distance of less than 15 Å at both the A-B and A'B' interfaces (**Fig. S2B**). There are no Chl *a* molecules with Mg²⁺-Mg²⁺ distances < 15 Å across the B-A' and B'-A interfaces (**Fig. S2B**). There are many Chls within < 20 Å in all the interface (**Fig. S2C**). **Fig. S2D** shows all of these contacts superimposed. These closely placed Chl pairs are delineated by their monomer, chain ID and residue ID in **Fig. S2E**. Key Chl residues at the monomer interfaces are highlighted in **Fig. S2E** and shown in **Fig. S2F**. In monomers B and B' there are two Chl *a*'s oriented parallel to one another near the monomer interface (**Fig. S2F**).

Central cavity

The central cavity of the tetramer contains unresolved densities which may correspond to lipids, detergent molecules, or pigment/carotenoid molecules. A model cyanobacterial membrane containing 47% monogalactosyldiacylglycerol (MGDG), 23% digalactosyldiacylglycerol (DGDG), 21% sulfoquinovosyl diacylglycerol (SQDG), and 9% phosphatidylglycerol (PG) was generated with CHARMM-GUI. Tetrameric PSI was embedded in the membrane using VMD (**Fig. 3A-B**). CASTp was used to calculate the volume of the central cavity using PsaI, PsaL, and PsaM subunits. The central cavity is about 70 Å x 50 Å and has an approximate volume of 65,000 Å³ (**Fig. 3C-D**). Although we do not observe densities in this central cavity, the volume would suggest it could accommodate a bilayer with about 25-30 lipids per leaflet. Prior work has shown that TS-821 tetramer is enriched in novel carotenoids that are lost upon detergent destruction into dimers which may suggest that this central cavity contains carotenoids that maybe lost during isolation.

Differential Stability of Dimeric Interfaces

The two different dimeric interfaces, A-B and B-A', are shown in the luminal and stromal view of **Fig S3A-B** with the interfacial amino acid residues shown in space-filling. An enlarged view of these interfaces from the external membrane view is shown in **Fig S3B**. The residues contributing to the most stabilization of each interface is shown in spheres and also denoted in **Fig S3C**. To qualitatively assess the stability of these interfaces, we performed refinement of each interface with the HADDOCK, including parameterization of Chl *a* molecules. Results reveal a differential contribution of combined non-covalent forces governing each interface (**Fig. S3C-D**). Although the overall calculated energetics are similar in the HADDOCK score, electrostatic contributions are substantial in the A-B' interface as compared to the A-B interface. Thus, charge-charge interactions are of more considerable significance for A-B'. The decreased contribution of electrostatic energies is compensated by extensive van der Waals interactions in the A-B interface (**Fig. S3C**). The A-B interface contains more proximal Chl (11) than the B-A' interface (6). However, the B-A' interface includes more transmembrane domains (14) than the A-B interface (7). This compositional difference is reflected in the increased buried surface area of A-B (2,418 Å²) whereas the B-A' only has 1561 Å² (**Fig. S3D**). In addition, we compared interfaces with missing Chl *a* molecules to assess the impact of the Chl *a* molecules. Overall, Chl *a* molecules contribute to the solvation energies of both interfaces and the regulation of electrostatic energies. As expected, they also contribute to the larger surface area for the A-B and A'-B' interfaces, indicating that they are integral for their formation and stability.

Similar results were obtained by using Monte Carlo calculations to allow the sampling of amino acids' protonation states. To understand the stability of the interfaces between the different monomers, we extracted the amino acids that have their C α within 12 Å from the amino acids of corresponding monomers. Hence, Monte Carlo sampling is used to obtain Boltzmann distribution for

the protonation patterns. Then, we evaluated the electrostatic and van der Waals interactions between 78 and 119 amino acids for interface A-B' and A-B respectively based on Boltzmann occupancies. The sum of the Coulombic interactions between the amino acids in interface A-B' is -0.4 Kcal/mol, while the sum of vdw interactions is -14 Kcal. For interface A-B, the contribution from Coulombic and vdw interactions is 0.1 Kcal/mol and -16 Kcal/mol respectively. These results indicate that the structure of the dimer-of-dimers is mainly stabilized by the dispersion interactions (London forces).

Differences in structure with trimeric PSI

To elucidate the differences in Chl placements in the monomeric interfaces of the tetrameric PSI from the earlier elucidated trimeric structures, we have directly compared TS-821 PSI to *T.e.* BP-1 (PDB ID- 1JB0) and *Syn* PCC 6803 PSI (PDB ID- 5OY0) crystal structures (**Fig. 4 and S4**). These Chls may function in the energy transfer between adjacent PSI monomers. Interestingly, only the A-B and A'-B' interfaces are conserved in contacts and packing to the three identical interfaces found in the trimeric PSI (**Fig. 4A**). The enlarged view in **Fig. 4B** reveals the three parallel Chl *a* molecules observed at these interfaces of tetrameric PSI from TS-821 that are also observed in the trimeric PSI complexes. However, the two other dimer interfaces in the tetramer (A-B' and A'-B) form a distinct interface that is not observed in the interfaces of either trimeric PSI. These Chl *a* molecules are located much closer together at the A-B and A'-B' interfaces in the tetramer as compared to the A-B' and A'-B. The larger surface area of the A-B and A'-B' interfaces along with the conservation of the Chl *a* arrangement with trimeric structures suggests that these two interfaces likely form the dimers and the A-B and A'-B' interfaces are responsible for joining the two dimers together.

Role of PsaL in interface stability

Early work has shown that PsaL is key to the assembly and stability of the trimeric structure of PSI (32). Since the A-B interface in the tetrameric PSI resembles the trimeric PSI interface at the core, we investigated the interaction interface of PsaLs in PSIs of both *T.e.* BP-1 and TS-821 and other tetramers (PDB IDs 6JEO and 6K61) (**Fig. 5**). The three PsaL subunits form a central helical bundle in *T. e.* BP-1 PSI trimer (**Fig. 5A**). This central core contains a PsaL subunit from each PSI monomer and has virtually no cavity. In the trimer, the enlarged view highlights the residues that stabilize this PsaL bundle. These residues span the entire transmembrane region with mostly non-polar interactions. However, the dimeric PsaL bundle in all three tetramer superimposed (PDB IDs- 6QWJ, 6JEO and 6K61) shown in **Fig. 5B** reveals interactions only at the luminal and stromal regions. In the PsaL interface (A-B and A'-B') of the tetramer, only the luminal face has a few hydrophobic residues present (**Fig. 5B**). Interestingly, there is a large number of polar sidechains in the stromal and luminal faces in the PsaL interface of the tetramer.

Evolutionary differences in PsaL

Our prior work has highlighted that all or nearly all of the HCR contain a tetrameric form of PSI (21, 22, 25). Since previous work has shown that PsaL is driving the trimerization (32), it is not clear what determines the oligomeric state of the PSI tetrameric complex. However, prior work has identified changes in two regions of the PsaL primary sequence that seem to correlate with trimeric and tetrameric forms (25). To investigate that in greater details, we compared 113 PsaL orthologs from a broad range of cyanobacteria including those shown to be trimeric, far-red light-inducible Chl *f*-containing, marine (*Prochlorococcus/Synechococcus*) and putative tetrameric cyanobacteria. This phylogenetic approach is both broader and based on a different evolutionary trait from either the initial tree based on only 16S rRNA (**Fig. 1A**) or the previous tree utilizing 29

universally conserved genes (25). Moreover, this phylogenetic tree is focused on changes associated with the PsaL (**Fig. 6A**).

Our analysis of a non-redundant set of PsaL sequences yielded four putative monophyletic clusters (**Fig. 6A**). Each cluster does contain multiple species whose PSI oligomeric state/type is experimentally known (trimeric, tetrameric, or far-red), giving us further confidence in this phylogenetic approach to identify PSI oligomerization states. In some cases, organisms contained multiple PsaL proteins that fit into two or three groups. Classification of each group was based on existing structural data from representative members: using *Leptolyngbya* sp. strain JSC-1 (33) as an anchor for the far-red cluster, *T. e.* BP-1 as trimer cluster (9), *Chroococcidiopsis* sp. TS-821 as tetramer anchor (23, 34) and the Pro/Syn group does not yet have a high-resolution structure however a recent AFM study clearly indicated a trimeric form of PSI (19). Upon this agnostic phylogenetic classification, we identified some sequence/structure variation(s) found within the PsaL gene when each PSI cluster identified in **Fig. 6A** was analyzed by a LogoPlot of the region between the transmembrane domain (TMD) #2 and #3 (**Fig. 6B**) as well as the C terminus region (**Fig. 6C**). Both of these regions were previously identified as variable regions of PsaL in cyanobacteria (23).

The sequence LogoPlots of the trimer and far-red sequences are most alike based on conserved regions (CR) across CR-I and CR-II in the linker region (**Fig. 6B**). Similarly, the far-red and trimers show signs of common ancestry based on the c-terminal regions, specially CR-III and CR-IV. Interestingly, the tetramers were most similar to the *Prochlorococcus/Synechococcus* group based on their linker region with conserved TV/T/APNPP motif found in CR-V region (**Fig. 6C**). Unlike the trimers and far-red groups, there were no conserved regions between tetramers and the *Prochlorococcus/Synechococcus* group across the C-terminal regions. Three *Prochlorococcus* genomes of SS120 (low-light), MIT9313 (low-light), and MED4 (high-light) clustered with the

Prochlorococcus/Synechococcus group even though they are known to form trimeric PSI structures (19).

The LogoPlot identified conserved regions (CR I-V) were spatially placed in the known structures of the tetrameric TS-821 PSI and the trimeric PSI structures from *T. e.* BP-1 and *Syn* PCC 6803 (**Fig. 6D**). It is clear that these conserved regions form different associations based on the oligomeric state. Inspection of these models reveals apparent differences in where these CRs reside in the different PSI oligomers. For example, in tetramer, CR-II and CR-V, are located in the stromal face of both A-B and A'-B' interfaces. In trimeric PSI, CR-I and CR-II are located in the stromal face while CR-III and CR-IV are located in the luminal face. CR-II is found in both tetrameric (in the dimeric interfaces A-B and A'-B') as well as the trimeric PSI (central core), which suggests a crucial role of this region in oligomerization of PSI dimer in case of tetrameric PSI and PSI trimers. On the other hand, CR-III and CR-IV are only found in trimeric and far-red PSIs, which suggests a sole role in trimeric PSI. Interestingly, CR-III and CR-IV are not observed in the marine cyanobacteria (*Prochlorococcus/Synechococcus*) raising questions as to the oligomeric state of PSI in these organisms. This is supported by the confirmation and placement of CR-II and CR-V, both the tetramers and the marine PsalLs have two copies of CR-V.

Discussion

In this study, the structure of tetrameric PSI from TS-821, a non-heterocyst forming cyanobacteria, was solved by cryo-EM with a resolution of 3.9 Å resolution and has an organization of a dimer-of-dimers. The tetramer exhibits two types of interfaces- A-B (between monomer A and B) and B-A' (between monomer B and A'). The A-B interface of the tetramer resembles closely to that of the trimeric PSI of other cyanobacteria (*T.e.* BP-1 and *Syn* PCC 6803) whereas the BA' interface is novel and resembles the two reports of a tetrameric PSI in *Anabaena sp.* PCC 7120 (formerly

Nostoc), a heterocyst forming cyanobacteria (21, 22). This suggests that the tetrameric form of PSI is widespread, occurring in both heterocyst and non-heterocyst forming cyanobacteria. The Chls in the both tetramers show a distinct arrangements in the two interfaces as compared to the trimeric PSI. The Chl placements in the A-B interface are more similar to that observed in the trimeric interfaces but are clearly different the Chl positions in the B-A' interface. The three parallel Chls common to the A-B interface and trimer suggests that this dimeric interface most resembles the trimeric interface. Our extended phylogenetic analysis of PsaL confirms our prior observation that changes in the loop between TMD #2 and #3 and the C-terminus, may have an evolutionary and structural role in how subunit PsaL alters the oligomeric state of PSI. Structural analysis suggests that the PsaL core in trimeric PSI as compared to that of tetramer have multiple bulky residues, mostly aromatic residues all along the transmembrane domain. In tetramer, only the luminal face of PsaL in the AB interface has multiple bulky (mostly non-polar and aromatic) residues while the stromal face has polar residues. This is similar to that of the tetrameric PSI of heterocyst-forming cyanobacteria, where it was suggested that specific amino acids with large side chains may prevent formation of the trimer due to the steric hindrance provided by these bulky groups (22).

Our work adds a new tetrameric PSI cryo-EM structure for a cyanobacteria outside of the two reports in *Anabaena sp.* PCC 7120 (21, 22). This work coupled with our prior BN-PAGE and bioinformatic characterization (23, 24) suggest that most if not all members of the HCR group of cyanobacteria have a tetrameric form of PSI. However, the evolutionary role of this change in PSI structure is still elusive. Previous work has shown that in three different cyanobacteria, exposure to high light can induce the formation of tetrameric PSI and was also shown to induce the accumulation of more novel carotenoids in the thylakoid membranes (25). This might suggest that one role of the tetramer is to allow accumulation of photoprotective carotenoids with PSI when

exposed to high light environments. How these carotenoids are associated with PSI is unknown but their release upon dissociation into two dimers suggests that one possibility is an association within the central $\sim 65,000 \text{ \AA}^3$ central cavity.

The subunit PsaL was found to be important for formation of the trimeric form of PSI (32). Our bioinformatic work has identified small conserved regions in PsaL that correlate with this tetrameric symmetry by promoting PsaL dimerization versus a trimerization in the PSI trimers. According to the study by Netzer-El *et al.*, the C-terminus region of PsaL creates a short helix that associates with the PsaL subunits of the neighboring PSI monomers in the trimer, assisting in trimer stabilization, however addition of a terminal histidine disrupts this association resulting in largely monomeric form of PSI along with dislocation and structural differences in subunits PsaM and PsaI which are located in the trimerization region (35). In our study, we highlight the placement of PsaL in the formation of two dimeric interfaces and how the protein: protein interactions in these interfaces are quite different from those observed that stabilize the trimeric form. Phylogenetic analysis along with motif analysis has revealed several loosely conserved region (CRI-V) within the PsaL subunit that contributes to oligomerization of PSI tetramers. Our analysis suggests that cyanobacteria having strictly trimeric PSI along with far-red type PSI trimers have common ancestry with specific conserved regions. On the other hand, the PsaL from marine cyanobacteria (*Prochlorococcus/Synechococcus*) and those who have tetrameric PSI lack CR-III & IV but have a new conserved domain CR-V. The presence of semi-conserved CR-II in all four groups suggest a role in all PSI structures yet this region is the least conserved overall. These general observations may be complicated since some organisms are known to encode multiple copies of the *psaL* genes yet in most cases each PsaL is placed phylogenetically within one of the four groups.

Our bioinformatic analysis of the PsaL protein of most *Prochlorococcus* strains (including MIT9313 and MED) did not cluster with trimeric group in the PsaL tree, yet was placed much closer to the tetrameric PSI-forming cyanobacteria. Surprisingly, a logo-plot of the loop region between TMD 2 & 3 of the marine cyanobacteria, identified CR-V that is enriched in Pro residues. There are actually two CR-V motifs that are also conserved with species known to have tetrameric form of PSI. Although, the structure of PSI from these marine cyanobacteria have not been studied by crystallography or cryo-EM, there was a recent AFM study two *Prochlorococcus* ecotype including a high-light (MED4) and low-light (MIT9313) ecotypes (19). Using AFM imaging of intact thylakoids, this group observed that non-trimeric forms of PSI (dimeric/monomeric) were greatly increased (~5x) relative to PSI trimers, when the cells were grown in high light conditions. This suggests that either the formation of trimers or the stability of existing PSI trimers is reduced upon high light exposure. Together this suggests that although the thylakoids of *Prochlorococcus*/*Synechococcus* group may contain PSI trimers, upon exposure to high light, there is a major shift towards dimeric/monomeric forms of PSI.

Our biochemical observation of tetrameric PSI in the HCR may reflect an increased stability of the tetramer, possibly due to the high local concentration of PSI and/or the lipid composition of the thylakoid region or sub-domain where PSI is localized. TS-821 is a thermophilic cyanobacteria that will have predominantly saturated fatty acids possibly due to limited fatty acid desaturase genes (FAD) as observed in *T. elongatus* (36). However the marine cyanobacteria (*Prochlorococcus*/*Synechococcus*) are mesophiles and have been shown to have highly unsaturated fatty acids due to their multiple FAD genes (36, 37) and possibly due to recently identified activity of a cyanophage-encoded lipid desaturases (38). Future work will be needed to see if the marine cyanobacteria contain a stable dimeric or tetrameric form of PSI when grown in high light.

Early analysis of 56 different cyanobacteria suggested that chloroplasts arose as a single monophyletic event from an organism that is most closely related to N₂-fixing unicellular cyanobacteria (Chroococcales) and possibly their sister group of closely-related heterocyst-forming cyanobacteria (Nostocales) ~2.1 BYA (27). It has been shown that members of the genus *Chroococcidiopsis* are the closest living relatives to the filamentous heterocyst-differentiating cyanobacteria (39). We and others have shown that the genus *Chroococcidiopsis* includes species with unique survival abilities under nitrogen-limiting conditions and can grow quite well in salt water (39). These authors speculated that *Chroococcidiopsis* was capable of surviving, following a sudden washout, into an increasingly saline environment, thereby providing a route for the evolution of open, ocean-dwelling cyanobacterial strains (40). Recently Sánchez-Baracaldo and colleagues proposed that closest relative of the chloroplast was an ancient freshwater cyanobacterium, *Gloeomargarita*. Their data suggests that the ancestor of the chloroplast and *Gloeomargarita* diverged about 2.1 billion years ago, which is relatively early in the evolutionary timeline of the cyanobacterial lineage (41, 42) and prior to when planktonic marine cyanobacteria diverged from freshwater ancestors (43). It will be interesting to see if *Gloeomargarita* has a tetrameric form of PSI, which may suggest that it could give rise to both marine cyanobacteria and chloroplasts. This early placement of a tetrameric PSI in cyanobacterial and possibly chloroplast evolution is supported by the observation that a tetrameric PSI was also found in *Cyanophora paradoxa* a member of the Archaeplastida and the most primitive group of photosynthetic eukaryotes known (44). Collectively, this work suggests that tetrameric PSI is a widely occurring form of PSI that may be an adaptation to high light conditions during cyanobacteria expansion and may be the key intermediate in the evolution of PSI structure in plants and algae.

Materials and Methods

Source of TS-821 cyanobacteria

Chroococcidiopsis TS-821 was originally isolated over 25 years ago from the Sankampaeng and Mac Fang hot springs near Chiang Mai in the northern part of Thailand (45, 46). All strains had vegetative cells surrounded by fibrous (F) layers. Early analysis by light and electron microscopy revealed that cell division occurred by binary fission, but neither motility nor mobile baeocytes were observed. Based on work by Woodbury and Stainier TS-821 (28) was assumed to belong to the genus *Chroococcidiopsis*. Further early work indicated that TS-821 was able to overgrow to dense cultures and was proposed to be a source for early biomass conversion (45, 47).

Growth of cyanobacteria

Chroococcidiopsis sp. TS-821 (TS-821) was cultured in a two-liter glass culture bottle with aeration at 45°C, with continuous white fluorescent light ~40 $\mu\text{mol}/\text{m}^2/\text{s}$ at bottle surface. Cells were harvested at the late log phase, and the wet cell mass harvested from a two-liter bottle culture is usually 3-4 g.

Cell lysis and isolation of cyanobacterial thylakoids

Thylakoid membrane isolation was done similar to earlier studies (48, 49). Homogenized suspensions of cells in ice-cold lysis buffer (50 mM MES-NaOH, pH 6.5, 5 mM CaCl_2 , 10 mM MgCl_2 , 0.5 M sorbitol) were ruptured by passing through the French press three times at 15,000 psi. After removing the intact cells by centrifugation 10,000 g for 5 min, the thylakoid membrane was pelleted after 30 min of 193,000 g (Type 50.2 Ti, Beckman) centrifugation. The pelleted membranes

were washed in wash buffer (50 mM MES-NaOH, pH 6.5, 5 mM CaCl₂, 10 mM MgCl₂) and pelleted again before final resuspension in storage buffer (50 mM MES-NaOH, pH 6.5, 5 mM CaCl₂, 10 mM MgCl₂, 12.5% v/v glycerol) and homogenized before storing at -20 °C or immediate solubilization.

Isolation of PSI tetramers

The TS-821 thylakoid membrane containing 1 mg/mL Chl *a* were solubilized in 1% β-DDM (Glycon Biochemicals, Luckenwalde, Germany). The solubilized membrane solution was loaded on a 10-30% sucrose gradient containing 0.01% β-DDM in the wash buffer. Two-step ultracentrifugation was used to purify PSI tetramer as described previously(48) except the following modifications. The first centrifugation was done using at 30,000 rpm (SW 32 Ti, Beckman) for 20 h. The concentrated and dialyzed PSI tetramer from the first spin was loaded on a 10-30% sucrose gradient again and centrifuged at 30,000 rpm for 24 h. PSI tetramers from the second gradient after spinning were dialyzed and concentrated for analysis.

DAPI imaging of TS-821 Cells

DAPI staining was used in order to observe DNA within intact cells. This staining allowed determination of the state of fission, binary or multiple; a particular cell was undergoing as well as the location of the dense nucleoid regions. Two microliters of DAPI stain was added to 1 mL of liquid cell culture, covered with aluminum foil, and incubated at room temperature with shaking for 1 h. The sample was observed using a Nikon Eclipse 80i Fluorescent Microscope under the DAPI filter.

Scanning electron microscopy

The samples were fixed in 3% glutaraldehyde buffered with 0.1 M sodium cacodylate. Following a 1 h incubation period, the samples were washed in cacodylate buffer 3 times, allowing 10 min per wash and then post-fixed in cacodylate buffered with 2% osmium tetroxide for 1 h. Samples were then washed three times with deionized water. During the final wash, small aliquots of the sample were allowed to settle onto a 2x3 mm silicon chip, which had been previously washed with poly-lysine. The sample was then dehydrated in a graded acetone series and critical point dried in CO₂ using a Ladd Research Critical Point Dryer. Dried samples were coated with gold with an SPI Sputter coater before examination in a Zeiss 1525 scanning electron microscope.

Transmission electron microscopy

The fixation protocol is the same as described above until the dehydration step. Samples were then washed in water 3X before dehydration in a graded ethanol series, and then finally dehydrated with propylene. Samples were then embedded in Spurr epoxy for 48 h prior to the final embedding and subsequent curing of the resin at 68°C for 24 h. For ultramicrotomy, samples were sectioned with a Reichert OMU3. Thin sections of approximately 70 – 90 nm were stained with methanolic uranyl acetate and lead citrate before examination in a Hitachi H800 transmission electron microscope operating at 75 KeV. Images were recorded on Kodak 4489 electron microscopic film.

DNA isolation and cloning of 16S RNA

DNA isolation was performed using a modification of the Saha *et al.* protocol (50). This method was modified to use mechanical cell lysis using a FastPrep®-24 Tissue and Cell Homogenizer and Orange Capped Lysis Matrix A Tubes, which are specific for DNA isolation. A large cell pellet

harvested from a dense liquid cell culture was resuspended in 1 mL 1x TE buffer (10 mM, Tris-HCl, pH 7.5, 1 mM EDTA) and homogenized at 4 m/s for 20 s. Lysate appeared blue due to the release of phycobilin proteins into solution. This tube was centrifuged for 1 min at 10,000 g, and the supernatant was immediately transferred to another 1.5 mL, microfuge tube. All additional steps were done, as previously described (50). Due to a significant amount of RNA contamination, 3 µL of RNase (10 mg/mL) was added to each tube and allowed to incubate at room temperature for 10 min. Amplification of the 16S rRNA was accomplished using these primers: forward (5'-AGAGTTTGATCCTGGCTCAG-3') and reverse (5'-AAGGAGGTGATCCARCCGCA-3'). The 50 µL reaction mixture consisted of 10 µL of 5X GoTaq Reaction Buffer, 1 µL dNTP mixture (10 mM each), 1 µL of each primer (10 pM/µL), 0.25 µL GoTaq Polymerase, 50 ng of template DNA and an appropriate amount of nuclease-free water to achieve 50 µL total reaction. The PCR conditions were as follows: 95° for 3 min, 30 cycles of 95°C for 1 min, 60°C for 1 min, 72°C for 1.5 min, and finally 72°C for 10 min subsequently followed by a 4°C hold. The resultant PCR product was cleaned up using QIAquick® PCR clean-up system and immediately ligated into the TOPO 2.1®(Invitrogen®) vector as directed and transformed into TOP10 chemically competent cells (Invitrogen®).

Phylogenetic analysis

Phylogenetic analysis was performed utilizing 16S rRNA sequences of similar length to the sequence obtained for TS-821. Sequences were chosen for comparison based on high sequence homology from the BLAST search results or as representatives of different genera within the phylum Cyanophyta. All sequences were obtained through GenBank. Once all taxa were chosen, the sequences were aligned via the ClustalX algorithm within MEGA 4.0 (51). The sequence alignment

was exported to Modeltest 3.7, which chose the most appropriate statistical model for the generation of a phylogenetic tree (52). Following the suggested model chosen by Modeltest 3.7, MEGA 4.0 generated a phylogenetic tree with the following parameters: Neighbor-Joining Method coupled with a Maximum Composite Likelihood of Nucleotides, complete deletions of all gaps and missing data within the sequence alignment, the heterogeneous pattern among lineages with a gamma parameter of 0.5087, and with 5000 Bootstrapping replicates (53). The resulting tree was condensed to a consensus tree containing bootstrap values located on the branches.

Cryo-EM data acquisition

Aliquots of 3 μL of PSI samples were applied on glow-discharged holey carbon grids (Quantifoil R2/2, 400 mesh) coated with a continuous 2 nm carbon film. The grids were blotted and plunge-frozen using a Vitrobot Mark III (Thermo Fischer Scientific). Cryo-EM micrographs were recorded at liquid nitrogen temperature on a Titan/Krios transmission electron microscope (Thermo Fischer Scientific, USA) operating at 300kV at NECEN (Leiden, the Netherlands). Micrographs were recorded at a nominal magnification of 130,000x using a K2 Summit direct electron detector (Gatan, Inc.) using a pixel size of 1.108 $\text{\AA}$ /pixel, with a dose rate of ~ 4.3 electrons/ $\text{\AA}^2/\text{s}$ and defocus values ranging from -0.6 to -3 μm . The total exposure time was 12.0 s, and intermediate frames were recorded in 0.5 s intervals resulting in an accumulated dose of ~ 50 electrons per $\text{\AA}^2$ and a total of 24 frames per micrograph.

Cryo-EM image processing and 3D reconstructions

The Cryo-EM image processing was performed in SCIPION 2.0 (54) using the integrated protocols. 4845 raw movies were corrected for beam-induced motion using MotionCorr2 (55). A sum

of all frames, filtered according to exposure dose, in each image stack was used for further processing. CTF parameters for each micrograph were determined by CTFFIND4 (56) and XMIPP - CTF estimation (57, 58). Particle selection and 2D classifications were performed using xmipp3 - manual-picking / xmipp3 - auto-picking (57, 58), xmipp3 – cl2D (58, 59) relion - 2D classification protocols. 325,648 picked particles underwent the 2D classification. The side views particles were picked and processed separately from the others (59, 60). The initial C2 symmetry enforced model was calculated de novo based on the 2D class averages using xmipp3-ransac (58, 61). The 3D-classification was performed using relion - 3D classification (62) protocol. The good class from the 3D classification-relion contained 122,411 particles and was subjected to further refinement protocol using xmipp3 – highres (54, 58). The resulting cryo-EM map contained 66,130 particle projections that were sharpened using LocalDeblur (63), using as input the resolution map calculated with Monores (54, 64). This cryo-EM map was used for model building and then underwent another round of sharpening using sachselab – locscale (54, 65) protocol. Reported resolutions are based on the gold-standard Fourier shell correlation (FSC) using the 0.143 cutoff criterion (**Fig. S1**). The local resolution was determined using DeepRes (66) (**Supp. Fig. S1**).

Model building

The obtained cryo-EM map was used for manual model building in Coot using the single protomer of cyanobacterial PSI from Te-BP1 as a reference. The resolution of the map was sufficient to assign all protein subunits unambiguously and to model most of the Chl *a* molecules. Rounds of real space refinement were performed in Phenix and included simulated annealing protocol and NCS restraints. Coordinates were manually edited in Coot after each refinement cycle and subjected to further rounds of refinement. The final validation check was performed with Molprobit

and Phenix validation tools. Images were prepared with the open-source version of PyMol (<https://sourceforge.net/projects/pymol/>), and Chimera (<https://www.cgl.ucsf.edu/chimera/>).

HADDOCK analysis

The cryo-EM model was refined with HADDOCK (67), and calculations were performed as previously described (68). The OPLS force field (69) was used for topology and parameter file generation, and PRODRG (70) was used to parameterize the Chl *a* molecules. Energy calculations were performed for the two interfaces of the tetramer using the HADDOCK score along with its component energy terms. These terms include van der Waals (E_{vdW}) and Coulomb (E_{elec}) intermolecular energies represented as and non-bonded energies, an empirical desolvation term (E_{desolv}) (71), and buried surface area (BSA) upon complex formation in $\text{\AA}^2$.

Monte Carlo simulations

Sidechain rotamers and protonation states were generated using Multi-Conformer Continuum Electrostatics (MCCE) program (72). Then Monte Carlo sampling was used to generate Boltzmann distribution of the different conformations based on the electrostatic and van der Waals energies. The electrostatic interactions were calculated by solving the Poisson-Boltzmann equation using DELPHI software (73), and the vdw interactions were calculated based on Amber forcefield (74).

Central cavity analysis

The volume of the central cavity was calculated using the CASTp webserver (75) with a probe radius of 1.4 $\text{\AA}$. Only subunits that border the central cavity (PsaI, PsaL, and PsaM) were included due to file size limits on the webserver. CHARMM-GUI (76) was used to generate an MGDG,

DGDG, SQDG, and DPPG mixed lipid bilayer with a lipid composition of 47:23:21:9 based on averages from previously reported (77-80) cyanobacteria lipid compositions. The tetramer was then embedded in the membrane by removing lipids within 0.6 Å from the protein using VMD (81) to provide a rough estimate of how many lipids could fit in the central cavity.

Data availability

The data supporting the findings of this manuscript are available from the corresponding author upon reasonable request. A reporting summary for this article is available as a Supplementary Information file. Model coordinates and density maps are available in the Protein Data Bank (PDB ID 6QWJ) and the EM Data Bank (EMD 4659).

Bioinformatic Methods

a) Cyanobacterial genome selection. All 1,639 cyanobacteria genomes available as of January 25, 2019, were downloaded from NCBI based on keyword searches with excluding viruses and phages. Based on the whole genome sequences, a subset of these 1,639 genomes was generated by building a Mash tree (82) to reduce redundancy. We selected a cutoff value of 0.1 based on the first plateau in the plot (**Fig. S1**), which resulted in 295 clusters. In some of these clusters, we observe sets of highly related or even identical organisms (data not shown), so we selected a unique yet representative member from each of the 295 clusters, using a random number generation from the Python package NumPy (83). This allowed us to randomly choose a single genome from each cluster yet also significantly reduced redundancy from 1,639 to only 295 distinct non-redundant organisms.

b) *PsaL protein identification and selection*. The genes within these 295 genomes were then annotated using Prokka (84) with -Cdsrnaolap and default options. From this annotated set of 295 genomes, BLAST 2.7 (85) was used to identify PsaL orthologs using the PsaL sequence from *T. e. BP-1* (UniProt ID: Q8DGB4). Initially, over 1567 putative PsaL orthologs were identified. The list of putative orthologs was reduced by filtering based on sequence identity ($\geq 40\%$), sequence coverage ($\geq 35\%$), and e-value ($\leq 1e^{-10}$), resulting in 113 orthologs. These heuristic values were determined by plotting the percent identity range on the x-axis from 0 to 100 and the number of hits meeting the criteria on the y-axis for all PsaL vs. 295 genome sequences while keeping e-value $\leq 1e^{-10}$ and sequence length difference of ≥ 35 . These filters reduced the analysis to PsaL orthologs with a percent identity cutoff of $\sim 40\%$.

c) *Phylogenetics and motif analysis*. Using these 113 PsaL orthologs, the sequences were aligned using MUSCLE v3.8.31 (86) using default options. The genomes were further reduced to 83 representatives total by visually inspecting the alignments in Jalview (87), using the known anchors and selected genomes. The anchor genomes for the far-red group was the far-red JSC-1 genome with a crystal structure available (33). The second filter applied was all *psaL* sequences that were less than 180 amino acids long were included in analyses excluding the known *psaL* ortholog for JSC-1. The representative *psaL* orthologs were further reduced by the manual alignment of the sequences for each group. They removed outliers that had the length greater than the mean of the group analyzed. In addition, to these 83 randomly selected PsaL sequences, we also included 17 experimentally verified tetramer-forming PsaL sequences (25), 12 experimentally determined far-red forms of PSI (88), and 3 well-characterized *Prochlorococcus* strains- SS120, MIT9313, and MED4 (19). The trimeric group was selected based on manual alignment and visually inspecting

genomes that skewed alignment from known *psaL* sequence in the group from the trimer anchor genome *T. e.* BP-1 and *Syn* PCC 6803. The tetramer group was selected based on manual alignment and visually inspecting genomes that skewed alignment relative to the tetramer anchor genome's *psaL* sequence, *Chroococcidiopsis* sp. TS-821. The marine *Pro/Syn* group had three anchors, which included *Prochlorococcus* strain SS120, MIT9313, MED4 (19). With these selected 128 proteins, a maximum likelihood tree was built using FastTree Version 2.1.11 with default options. The four semi-monophyletic groups formed in the tree include members from four distinct groups: 18 trimeric PSI; 28 far-red PSI; 36 tetrameric PSI; and 26 marine PSI. For each group, a linker region and sequence logo plots for each group were generated by WebLogo (89).

References and Notes

1. P. Cardol, A. Krieger-Liszkay, From light capture to metabolic needs, oxygenic photosynthesis is an ever-expanding field of study in plants, algae and cyanobacteria. *Physiol Plant* **161**, 2-5 (2017).
2. K. Awai, H. Ohta, N. Sato, Oxygenic photosynthesis without galactolipids. *Proc Natl Acad Sci U S A* **111**, 13571-13575 (2014).
3. N. Nelson, A. Ben-Shem, The complex architecture of oxygenic photosynthesis. *Nat Rev Mol Cell Biol* **5**, 971-982 (2004).
4. N. Nelson, C. F. Yocum, Structure and function of photosystems I and II. *Annu Rev Plant Biol* **57**, 521-565 (2006).
5. S. Caffarri, T. Tibiletti, R. C. Jennings, S. Santabarbara, A Comparison Between Plant Photosystem I and Photosystem II Architecture and Functioning. *Current Protein & Peptide Science* **15**, 296-331 (2014).
6. K. Biel, I. Fomina, Benson-Bassham-Calvin cycle contribution to the organic life on our planet. *Photosynthetica* **53**, 161-167 (2015).
7. R. E. Blankenship, H. Hartman, The origin and evolution of oxygenic photosynthesis. *Trends Biochem Sci* **23**, 94-97 (1998).
8. Y. Umena, K. Kawakami, J. R. Shen, N. Kamiya, Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* **473**, 55-60 (2011).
9. P. Jordan *et al.*, Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411**, 909-917 (2001).

10. E. El-Mohsnawy *et al.*, Structure and function of intact photosystem I monomers from the cyanobacterium *Thermosynechococcus elongatus*. *Biochemistry* **49**, 4740-4751 (2010).
11. J. Mondal, B. D. Bruce, Ferredoxin: the central hub connecting photosystem I to cellular metabolism. *Photosynthetica* **56**, 279-293 (2018).
12. Y. Mazor, A. Borovikova, I. Caspy, N. Nelson, Structure of the plant photosystem I supercomplex at 2.6 Å resolution. *Nat Plants* **3**, 17014 (2017).
13. A. Ben-Shem, F. Frolov, N. Nelson, Crystal structure of plant photosystem I. *Nature* **426**, 630-635 (2003).
14. E. J. Boekema *et al.*, Evidence for a Trimeric Organization of the Photosystem-I Complex from the Thermophilic Cyanobacterium *Synechococcus* Sp. *Febs Lett* **217**, 283-286 (1987).
15. E. J. Boekema *et al.*, A giant chlorophyll-protein complex induced by iron deficiency in cyanobacteria. *Nature* **412**, 745-748 (2001).
16. O. Almog, G. Shoham, D. Michaeli, R. Nechushtai, Monomeric and Trimeric Forms of Photosystem-I Reaction Center of *Mastigocladus-Laminosus* - Crystallization and Preliminary Characterization. *P Natl Acad Sci USA* **88**, 5312-5316 (1991).
17. D. Mangels *et al.*, Photosystem I from the unusual cyanobacterium *Gloeobacter violaceus*. *Photosynth Res* **72**, 307-319 (2002).
18. M. Antoshvili, I. Caspy, M. Hippler, N. Nelson, Structure and function of photosystem I in *Cyanidioschyzon merolae*. *Photosynth Res* **139**, 499-508 (2019).
19. C. MacGregor-Chatwin *et al.*, Membrane organization of photosystem I complexes in the most abundant phototroph on Earth. *Nat Plants* **5**, 879-889 (2019).

20. S. Y. Netzer-El, I. Caspy, N. Nelson, Crystal Structure of Photosystem I Monomer From Synechocystis PCC 6803. *Frontiers in Plant Science* **9**, (2019).
21. L. Zheng *et al.*, Structural and functional insights into the tetrameric photosystem I from heterocyst-forming cyanobacteria. *Nat Plants* **5**, 1087-1097 (2019).
22. K. Kato *et al.*, Structure of a cyanobacterial photosystem I tetramer revealed by cryo-electron microscopy. *Nat Commun* **10**, 4929 (2019).
23. M. Li, D. A. Semchonok, E. J. Boekema, B. D. Bruce, Characterization and evolution of tetrameric photosystem I from the thermophilic cyanobacterium Chroococcidiopsis sp TS-821. *Plant Cell* **26**, 1230-1245 (2014).
24. D. A. Semchonok, M. Li, B. D. Bruce, G. T. Oostergetel, E. J. Boekema, Cryo-EM structure of a tetrameric cyanobacterial photosystem I complex reveals novel subunit interactions. *Biochim Biophys Acta* **1857**, 1619-1626 (2016).
25. M. Li *et al.*, Physiological and evolutionary implications of tetrameric photosystem I in cyanobacteria. *Nat Plants* **5**, 1309-1319 (2019).
26. T. Dagan *et al.*, Genomes of Stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol Evol* **5**, 31-44 (2013).
27. L. I. Falcón, S. Magallón, A. Castillo, Erratum: Dating the cyanobacterial ancestor of the chloroplast. *The ISME Journal* **5**, 366-366 (2011).
28. J. B. W. a. R. Y. Stanier, Patterns of growth and development in pleurocapsalean cyanobacteria. *Microbiology Reviews* **42**, 2-44 (1978).
29. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).

30. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
31. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
32. V. P. Chitnis, P. R. Chitnis, PsaL subunit is required for the formation of photosystem I trimers in the cyanobacterium *Synechocystis* PCC 6803. *Febs Lett* **336**, 330-334 (1993).
33. F. Gan *et al.*, Extensive remodeling of a cyanobacterial photosynthetic apparatus in far-red light. *Science* **345**, 1312-1317 (2014).
34. M. Li *et al.*, Physiological and Evolutionary Implications of Tetrameric Photosystem I in Cyanobacteria. *bioRxiv*, 544353 (2019).
35. S. Y. Netzer-El, I. Caspy, N. Nelson, Crystal Structure of Photosystem I Monomer From *Synechocystis* PCC 6803. *Front Plant Sci* **9**, 1865 (2018).
36. X. Chi *et al.*, Comparative analysis of fatty acid desaturases in cyanobacterial genomes. *Comp Funct Genomics*, 284508 (2008).
37. S. Breton *et al.*, Unveiling membrane thermoregulation strategies in marine picocyanobacteria. *New Phytol* **225**, 2396-2410 (2020).
38. S. Roitman *et al.*, Cyanophage-encoded lipid desaturases: oceanic distribution, diversity and function. *ISME J* **12**, 343-355 (2018).
39. D. Fewer, T. Friedl, B. Budel, Chroococciopsis and heterocyst-differentiating cyanobacteria are each other's closest living relatives. *Mol Phylogenet Evol* **23**, 82-90 (2002).

40. A. J. Herrmann, M. M. Gehringer, An investigation into the effects of increasing salinity on photosynthesis in freshwater unicellular cyanobacteria during the late Archaean. *Geobiology* **17**, 343-359 (2019).
41. P. Sanchez-Baracaldo, J. A. Raven, D. Pisani, A. H. Knoll, Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc Natl Acad Sci U S A* **114**, E7737-E7745 (2017).
42. J. de Vries, J. M. Archibald, Endosymbiosis: Did Plastids Evolve from a Freshwater Cyanobacterium? *Curr Biol* **27**, R103-R105 (2017).
43. P. Sanchez-Baracaldo, Origin of marine planktonic cyanobacteria. *Sci Rep* **5**, 17418 (2015).
44. M. Watanabe, H. Kubota, H. Wada, R. Narikawa, M. Ikeuchi, Novel supercomplex organization of photosystem I in *Anabaena* and *Cyanophora paradoxa*. *Plant Cell Physiol* **52**, 162-168 (2011).
45. N. R. Hayashi, K. Terazono, N. Hasegawa, T. Kodama, Y. Igarashi, Identification and characterization of phycobiliprotein from a thermophilic cyanobacterium, *Chroococcidiopsis* sp. strain TS-821. *J Ferment Bioeng* **84**, 475-477 (1997).
46. N. R. Hayashi *et al.*, Isolation and cultivation of thermophilic cyanobacteria from hot springs of northern thailand. *J Ferment Bioeng* **78**, 179-181 (1994).
47. N. R. Hayashi, T. Ishida, Y. Peerapornpisal, Y. Igarashi, T. Kodama, Effect of carbon dioxide concentration on the growth and RubisCO activity of a thermophilic cyanobacterium, *Chroococcidiopsis* sp. strain TS-821. *J Ferment Bioeng* **80**, 507-509 (1995).

48. M. Li, D. A. Semchonok, E. J. Boekema, B. D. Bruce, Characterization and evolution of tetrameric photosystem I from the thermophilic cyanobacterium *Chroococcidiopsis* sp. TS-821. *Plant Cell* **26**, 1230-1245 (2014).
49. M. Watanabe, H. Kubota, H. Wada, R. Narikawa, M. Ikeuchi, Novel supercomplex organization of photosystem I in *Anabaena* and *Cyanophora paradoxa*. *Plant and Cell Physiology* **52**, 162-168 (2011).
50. S. K. Saha, L. Uma, G. Subramanian, An improved method for marine cyanobacterial DNA isolation. *World Journal of Microbiology & Biotechnology* **21**, 877-881 (2005).
51. K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596-1599 (2007).
52. D. Posada, K. A. Crandall, MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818 (1998).
53. P. I. Gordiichuk *et al.*, Solid-State Biophotovoltaic Cells Containing Photosystem I. *Advanced Materials* **26**, 4863-+ (2014).
54. J. M. de la Rosa-Trevin *et al.*, Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol* **195**, 93-99 (2016).
55. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).
56. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221 (2015).
57. C. O. Sorzano *et al.*, Semiautomatic, high-throughput, high-resolution protocol for three-dimensional reconstruction of single particles in electron microscopy. *Methods Mol Biol* **950**, 171-193 (2013).

58. C. O. S. Sorzano *et al.*, A new algorithm for high-resolution reconstruction of single particles by electron microscopy. *J Struct Biol* **204**, 329-337 (2018).
59. C. O. Sorzano *et al.*, A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J Struct Biol* **171**, 197-206 (2010).
60. S. Naithani, J. M. Hou, P. R. Chitnis, Targeted inactivation of the *psaK1*, *psaK2* and *psaM* genes encoding subunits of Photosystem I in the cyanobacterium *Synechocystis* sp PCC 6803. *Photosynthesis Research* **63**, 225-236 (2000).
61. J. Vargas, A. L. Alvarez-Cabrera, R. Marabini, J. M. Carazo, C. O. S. Sorzano, Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics* **30**, 2891-2898 (2014).
62. S. H. Scheres, Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579**, 125-157 (2016).
63. E. Ramirez-Aportela *et al.*, Automatic local resolution-based sharpening of cryo-EM maps. *Bioinformatics* **36**, 765-772 (2020).
64. J. L. Vilas *et al.*, MonoRes: Automatic and Accurate Estimation of Local Resolution for Electron Microscopy Maps. *Structure* **26**, 337-+ (2018).
65. A. J. Jakobi, M. Wilmanns, C. Sachse, Model-based local density sharpening of cryo-EM maps. *Elife* **6**, (2017).
66. E. Ramirez-Aportela, J. Mota, P. Conesa, J. M. Carazo, C. O. S. Sorzano, DeepRes: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. *IUCrJ* **6**, 1054-1063 (2019).

- 814 67. C. Dominguez, R. Boelens, A. M. Bonvin, HADDOCK: a protein-protein docking
815 approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731-
816 1737 (2003).
- 817 68. P. L. Kastritis, A. M. J. J. Bonvin, Are Scoring Functions in Protein-Protein Docking
818 Ready To Predict Interactomes? Clues from a Novel Binding Affinity Benchmark. *Journal*
819 *of Proteome Research* **9**, 2216-2225 (2010).
- 820 69. W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, Development and Testing of the OPLS
821 All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids.
822 *Journal of the American Chemical Society* **118**, 11225-11236 (1996).
- 823 70. D. M. van Aalten *et al.*, PRODRG, a program for generating molecular topologies and
824 unique molecular descriptors from coordinates of small molecules. *J Comput Aided Mol*
825 *Des* **10**, 255-262 (1996).
- 826 71. J. Fernandez-Recio, M. Totrov, R. Abagyan, Identification of protein-protein interaction
827 sites from docking energy landscapes. *J Mol Biol* **335**, 843-865 (2004).
- 828 72. Y. Song, J. Mao, M. R. Gunner, MCCE2: improving protein pKa calculations with
829 extensive side chain rotamer sampling. *J Comput Chem* **30**, 2231-2247 (2009).
- 830 73. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of
831 nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**,
832 10037-10041 (2001).
- 833 74. W. D. Cornell *et al.*, A Second Generation Force Field for the Simulation of Proteins,
834 Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **117**,
835 5179-5197 (1995).

836 75. W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, CASTp 3.0: computed atlas of surface
837 topography of proteins. *Nucleic Acids Res* **46**, W363-W367 (2018).

838 76. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface
839 for CHARMM. *J Comput Chem* **29**, 1859-1865 (2008).

840 77. H. W. Murata, *Membrane Lipids in Cyanobacteria*. In: Paul-André S., Norio M. (eds)
841 *Lipids in Photosynthesis: Structure, Function and Genetics. Advances in Photosynthesis*
842 *and Respiration*. (Springer, Dordrecht, 1998), vol. 6.

843 78. I. Sakurai *et al.*, Lipids in oxygen-evolving photosystem II complexes of cyanobacteria and
844 higher plants. *J Biochem* **140**, 201-209 (2006).

845 79. H. S. Van Walraven, E. Koppelaar, H. J. Marvin, M. J. Hagendoorn, R. Kraayenhof, Lipid
846 specificity for the reconstitution of well-coupled ATPase proteoliposomes and a new
847 method for lipid isolation from photosynthetic membranes. *Eur J Biochem* **144**, 563-569
848 (1984).

849 80. L. L. Kiseleva, I. Horváth, Lás. Vigh, D. A. Los, Temperature-induced specific lipid
850 desaturation in the thermophilic cyanobacterium *Synechococcus vulcanus*. *FEMS*
851 *Microbiology Letters* **175**, 179-183 (1999).

852 81. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J Mol Graph* **14**,
853 33-38, 27-38 (1996).

854 82. B. D. Ondov *et al.*, Mash: fast genome and metagenome distance estimation using
855 MinHash. *Genome Biol* **17**, 132 (2016).

856 83. O. E. Jones E, Peterson P, *et al.*, SciPy: Open Source Scientific Tools for Python. (2001).

857 84. T. Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069
858 (2014).

- 859 85. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment
860 search tool. *Journal of Molecular Biology* **215**, 403-410 (1990).
- 861 86. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high
862 throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
- 863 87. M. Clamp, J. Cuff, S. M. Searle, G. J. Barton, The Jalview Java alignment editor.
864 *Bioinformatics* **20**, 426-427 (2004).
- 865 88. F. Gan, D. A. Bryant, Adaptive and acclimative responses of cyanobacteria to far-red light.
866 *Environ Microbiol* **17**, 3450-3465 (2015).
- 867 89. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: a sequence logo
868 generator. *Genome Res* **14**, 1188-1190 (2004).

869

Acknowledgments

General: We thank Jon Nguyen for culturing and harvesting TS-821 used in this work. We also thank Utrecht EM staff who assisted in sample preparation for TEM and imaging.

Funding: Support has been provided from the Gibson Family Foundation, the Bredesen Center for Interdisciplinary Research and Education, the Dr Donald L. Akers Faculty Enrichment Fellowship, and National Science Foundation (DGE-0801470 and EPS-1004083) support to B. D. B. In addition, J.M. and B. D. B. have been supported via a JDRD Award from UTK/ORNL Science Alliance to B. D. B., M. L. has been supported as a CIRE Fellow at University of Tennessee, Knoxville. C.J.C. was supported from an NSF GFRP Award (Grant No. 2017219379). K. S. has been supported by the Tickle College of Engineering, UTK.

Author Contribution: B.D.B and E. B. conceived the project, D.A.S, and B.D.B led the experimental design. M.L performed cell culturing and PSI isolations. D.A.S performed TEM analysis. D.A.S performed 2D particle analysis and 3D reconstruction. D.A.S, C.O.S.S., E. R.-A., and P.K. performed density mapping and refinement. A. G. performed model building and refinement. D.A.S. and P.K. performed HADDOCK calculations. M.A. performed Monte Carlo simulations. Structure analysis and comparison was guided by B.D.B and performed by J.M. and C.J.C. Genomic and bioinformatics analysis was performed by K.S. Figures were made by D.A.S, J.M., C.J.C., and B.D.B. The manuscript was written by D.A.S, J.M., C.J.C. and B.D.B. with input from all other authors.

Competing interests: The authors declare no competing interests.

893

894 **Data and materials availability:** If data are in an archive, include the accession number or a
895 placeholder for it. Also include any materials that must be obtained through an MTA.

896

898 **Fig. 1. *Chroococidiopsis* TS-821 cells and Isolation of PSI**

899 *Chroococidiopsis* TS-821 cells and Isolation of PSI. (A) BN-PAGE of the β -DDM solubilized
900 thylakoids of *T. elongatus* (left lanes) and TS-821 (right lanes) using increasing amounts of β -
901 DDM. The *T. elongatus* photosystems are identified on the left and the MW standards are shown
902 on the right. Asterisks (*) indicate the PSI-tetramer, PSI-dimer and PSI-monomer (top to bottom)
903 (B) Scanning electron micrographs of the TS-821 cells showing binary fission in multiple planes

yielding single, dimer/octamers, and multiple or aggregate cells. (C) Bright-field image of TS-821 cells showing carbohydrate sheath material. (D) DAPI staining of the TS-821 cells showing chromosomal DNA during cell division in multiple planes, also shows autofluorescence of extra-cellular sheath. (E) Transmission electron micrograph of a recently divided pair of cells showing the thylakoids, T; inner and outer membranes, M; and the F-layer, F. (F) Phylogenetic tree based on 16S rRNA analysis: the yellow box denotes different *Chroococcidiopsis* strains, the red box is TS-821 and the green box are other members of the order Pleurocapsales. The numerical value on the right is the percent identity with TS-821.

913 **Fig. 2. Tetrameric PSI structure of TS-821**

914

915 Tetrameric PSI structure of TS-821: (A) Surface view of PSI monomers (A, B, A' and B') that

916 make up the tetramer are shown in stromal (left), membrane (center) and luminal (right) views.

917 Each monomer is identical but colored differently for the purpose of visualization. (B) Stromal

918 (left), membrane (center) and lumenal (right) view of tetrameric PSI colored by chain. (C) Labelled
919 subunits of one of the monomers are shown in membrane (top) and stromal (bottom) view. (D)
920 Cofactors of one of the monomers are shown in both membrane (top) and stromal (bottom) views.
921 The three [4Fe-4S] clusters are depicted as spheres and colored by element. Individual subunits
922 are shown as surface. F_A is located in PsaA subunit while both F_X and F_B are located in PsaC
923 subunit. Phylloquinone molecules are shown in blue, beta-carotenoids in orange, and Chl *a* mole-
924 cules in green. Only pyrrole rings and Mg (gray) of Chl *a* molecules are show for simplicity. The
925 special pair of Chl *a* is shown in black. (E) Table showing all the prosthetic groups per subunit in
926 a PSI monomer.
927

928 **Fig. 3. Potential lipid capacity and central cavity size**

929

930 Potential lipid capacity and central cavity size: (A) Stromal and (B) luminal surface views of te-

931 trameric PSI embedded in a model cyanobacterial membrane generated with CHARMM-GUI and

932 embedded with VMD. The membrane contains 47% MGDG, 23% DGDG, 21% SQDG, and 9%

933 PG galactolipids. (C) Stromal and (D) luminal views of the central cavity (blue) calculated with

934 CASTp. Only chains that form the cavity are shown (PsaL on all four monomers, PsaI and PsaM

935 on monomers B and B').

936

Fig. 4. Interface Chl *a*'s between PSI monomers of TS-821 tetrameric PSI vs *T. e.* BP-1 trimeric PSI and *Syn.* PCC 6803 trimeric PSI

Interface Chl *a*'s between PSI monomers of TS-821 tetrameric PSI vs *T. e.* BP-1 trimeric PSI and *Syn* PCC 6803 trimeric PSI: (A) Tetrameric PSI of TS-821 (left), trimeric PSI of *T. e.* BP-1 (PDB-ID:1JB0) (center) and trimeric PSI of *Syn* PCC 6083 (PDB-ID: 5OY0) are shown in surface; only chain PsaA (blue) and PsaB (green) are shown as cartoon. In all three structures, the pyrrole rings of Chl *a* associated with PsaA at the interface of the PSI monomers are high-lighted as purple and

945 those associated with PsaB are high-lighted in orange. **(B)** A closer look of the interfaces: B'-A
946 and A-B: The black and red loops are part of PsaB with sequences- QPKFRPS and
947 MYRTNFGIGHS respectively. The distances are shown in yellow lines. The missing Chl *a* at the
948 B-A' interface is shown in purple dashed pyrrole ring at a relatively similar position as it is in the
949 mono-AB interface.
950

Fig. 5. Interaction interface of PsaL subunits in trimeric (*T. e.* BP1) and tetrameric (6JEO and 6K61)

954 Interaction interface of PsaL subunits in trimeric (*T. e.* BP1) and tetrameric (6JEO and 6K61).
955 **A)** highlights the polar and non-polar residues that participate in the interaction interface at the
956 interface of the three PsaLs of their respective monomers in *T. e.* BP-1. **B)** highlights the residues
957 participating in the interaction interface between the two PsaLs of the A-B or A'-B' interface,
958 RMSD- 0.708 (*Nostoc* and *Anabaena* aligned) and RMSD- 5.732 (TS-821 and *Nostoc* aligned).
959 **C)** Shows the PsaL from all three tetramers (TS821, *Nostoc* and *Anabaena*) A-B interface (stromal
960 and membrane view). It also shows the alignment of all the Chl *a* molecules in the three tetramers.
961

962 **Fig. 6. Phylogenetic Analysis of PsaL and motif-analysis**

963 **Phylogenetic Analysis of PsaL and motif-analysis.** **A)** This represents a maximum-likelihood tree
964 built using FastTree 2 on the multiple alignment of the 108 PsaL proteins using MUSCLE. The
965 color shaded regions delineate separate groups that are known to include different forms of PSI
966 including trimeric forms of PSI, far-red light forms of PSI, tetrameric forms of PSI, and a fourth
967 group of marine cyanobacteria including members of *Prochlorococcus* and *Synechococcus*. Some
968

969 species contain multiple copies of PsaL in their genome and these are denoted by colored circles
970 (by species) and numbers (by number of PsaL copies). **B)** Logo plot of the loop region between
971 the predicted TMD #2 and #3. The ending and beginning of the TMD are shown by colored boxes
972 below the Logo sequence. The bit score scale was set to 6 bits to allow the error bars to be visible.
973 The conserved motifs were shaded and named CR I-V. Within each CR the most conserved amino
974 acids were further indicated by an asterisk above the single letter code. **C)** Similar to panel B, we
975 display the Logo plots of the PsaL C-terminus of the four different groups also indicating the
976 conserved regions (CR III-IV). **D)** Comparison of the CRs in TS-821, *T. e.* BP-1 and *Syn* PCC
977 6803 PSIs.

978

Supplementary Materials

Fig. S1. Single particle 2D classification and 3D model reconstruction

(A-D) Flow chart of cryo-EM image processing of tetrameric photosystem I from TS-821. (A) Single particles picked manually and using xMipp3 – picking protocol – 325648 particles. (B) Representative good 2D class averages of PSI tetramer obtained by xMipp3 - cl2D and relion 2D classification protocols (scale bar = 25 nm). In total 240 000 particles. All of them were processed in relion 3D classification. The good 3D class after relion 3D classification contained 122411

988 particles. (C) Representative 2D projections of the refined 3D model D (**scale bar = 25 nm**). (D)
 989 Refined 3D model of PSI tetramer (**scale bar = 25 nm**) In total 66130 particles. (E - H) Local
 990 resolution and the FSC of the TS-821 tetrameric Photosystem I Cryo-EM map. (E) Resmap reso-
 991 lution color slice through the unsharpened Cryo-EM map. (F) DeepRes local resolution histogram
 992 of the unsharpened tetrameric Cryo-EM map. (G) Gold-standard FSC curves of the final 3D re-
 993 constructions after refinement with xmipp3 - highres of the tetramer validated at
 994 the <https://www.ebi.ac.uk/pdbe/emdb/validation/fsc/>, (H) Sharpened map of the TS-821 PSI te-
 995 tramer. The Cryo-EM map was sharpened using LocalDeblur. It has been colored based on
 996 DeepRes local resolution. In this case, DeepRes was applied on the sharpened map. From left to
 997 right respectively: stromal (n)-side; side view; luminal (p) – side. (**scale bar = 25 nm**).
 998

999 Fig. S2. Distances between Mg²⁺-Mg²⁺ of adjacent pyrrole rings of Chl *a*.

1000

1001 Distances between Mg^{2+} - Mg^{2+} of adjacent pyrrole rings of Chl *a*: Mg-Mg distances within (A) 10
1002 Å (blue), (B) 15 Å (orange), and (C) 20 Å (yellow) (D) Combined Mg-Mg distances (10, 15, and
1003 20 Å). (E) Table showing Mg-Mg distances of Chl *a* molecules at the interface between PSI mon-
1004 omers. (F) Mg-Mg distances of Chl *a* molecules at the interface between PSI monomers: Chl *a*
1005 pyrrole rings are colored based on the associated subunit - green for PsaA, blue for PsaB and
1006 yellow for PsaK. The surface depicts the respective monomers separated by dashed gray line and
1007 labelled.

1008 **Fig. S3. Structure and stability of the two dimer interface**

1009

1010 The stability of interfaces in tetrameric PSI. (A) Two distinct interfaces are observed in tetramer
1011 (labeled as A-B and A'-B); each individual PSI is colored differently. The residues that contribute
1012 significantly to the stabilization energy are shown as spheres. This contribution is dominated by
1013 the dispersion forces (B) The close up of the two interfaces, A-B and B-A'. (C) the dispersion
1014 energies contribution of individual amino acids into the stabilization of interfaces calculated using
1015 Amber forcefield. (D) The electrostatics and van der Waals energies and the accessible surface
1016 areas of both interfaces.
1017

Fig. S4. Comparison of Chl *a* arrangements of trimeric PSI (PDB-ID: 1JB0) from *T. e.* BP-1 and tetrameric PSI from TS-821

Comparison of Chl *a* arrangements of trimeric PSI (PDB-ID: 1JB0) from *T. e.* BP-1 and tetrameric PSI from TS-821: (A) The Chl *a* pyrrole rings are depicted in green and the distances (<20Å) between Mg-Mg are shown in yellow. Chl *a*'s at the interface between monomers are circled in black. (B) Distances between parallel Chl *a*'s at the interface of the monomers are shown and labelled.

1026 **Fig. S5. Alignment of tetrameric PSIs of TS-821(PDB ID- 6QWJ) and *Nostoc* (PDB ID-6JEO)**

1027

1028 Alignment of tetrameric PSIs of TS-821(PDB ID- 6QWJ) and *Nostoc* (PDB ID-6JEO). Comparing
1029 the tetramers of TS-821 and *Nostoc* by performing different alignment strategies- all four mono-
1030 mers; monomer A only; monomer A and B only; and PsaL. Pymol was used to generate these
1031 alignments.
1032

1033 **Table S1. Statistics of data collection, processing and refinement.**

PDB ID	6QWJ
EMDB ID	4659
<u>Data collection and Image processing</u>	
Microscope	FEI Titan Krios G2
Detector	K2 Summit direct electron detector (Gatan, Inc.)
Magnification	~130,000
Voltage (kV)	300
Defocus range (μm)	-0.6 to -3.0
Pixel size (Å/pix)	1.108
Total electron dose (e-/Å ²)	~50
Exposure time (s)	12
Number of frames per movie	24
Number of micrographs/movies	4,845
Initial particle images (no.)	325,648
Final particle images (no.)	66,130
Map resolution (Å)	3.9
Applied symmetry	C2
<u>Refinement</u>	
Initial Model used (PDB code)	1JBO
Gold-standard FSC threshold	0.143
Model resolution (Å)	3.9
No. of Protein subunits (total/tetramer)	44
No. of chlorophyll	364
No. of carotenoid	64
No. of phaeophytin	8
No. of Fe-S centers (4Fe-4S)	12
No. of Prosthetic groups/tetramer	448
No. of atoms (Protein)	69,176
No. of atoms (cofactors)	20,740
No. of atoms (total)	89,916
B-factor- protein (Å ²)	94.46
B-factor- Cofactors (Å ²)	114.27
Number of missing residues (from coding sequence)	209/2413
% total a.a. placement	8.66
r.m.s. deviations- Bond lengths (Å)	0.005
r.m.s. deviations- Bond angles (°)	0.845
<u>Structure Validation</u>	
MolProbity Score	3.06
ClashScore	14.89
Rotamer Correctness (%)	89.90
<u>Ramachandra Plot Analysis</u>	
Favored (%)	87.33
Allowed (%)	12.53
Disallowed (%)	0.14
<u>Model vs Data</u>	
CCmask	0.79
CCbox	0.84
CCpeaks	0.70
CCvolume	0.79
Mean CC for ligands	0.80

1034

Figures

Figure 1

Chroococcidiopsis TS-821 cells and Isolation of PSI. (A) BN-PAGE of the β -DDM solubilized thylakoids of *T. elongatus* (left lanes) and TS-821 (right lanes) using increasing amounts of β -DDM. The *T. elongatus* photosystems are identified on the left and the MW standards are shown on the right. Asterisks (*) indicate the PSI-tetramer, PSI-dimer and PSI-monomer (top to bottom) (B) Scanning electron micrographs of the TS-821 cells showing binary fission in multiple planes yielding single, dimer/octamers, and

multiple or aggregate cells. (C) Bright-field image of TS-821 cells showing carbohydrate sheath material. (D) DAPI staining of the TS-821 cells showing chromosomal DNA during cell division in multiple planes, also shows autofluorescence of extra-celular sheath. (E) Transmission electron micrograph of a recently divided pair of cells showing the thylakoids, T; inner and outer membranes, M; and the F-layer, F. (F) Phylogenetic tree based on 16S rRNA analysis: the yellow box denotes different Chroococcidiopsis strains, the red box is TS- 821 and the green box are other members of the order Pleurocapsales. The numerical value on the right is the percent identity with TS-821.

Figure 2

Tetrameric PSI structure of TS-821: (A) Surface view of PSI monomers (A, B, A' and B') that make up the tetramer are shown in stromal (left), membrane (center) and luminal (right) views. Each monomer is identical but colored differently for the purpose of visualization. (B) Stromal (left), membrane (center) and luminal (right) view of tetrameric PSI colored by chain. (C) Labelled subunits of one of the monomers are shown in membrane (top) and stromal (bottom) view. (D) Cofactors of one of the monomers are shown in both membrane (top) and stromal (bottom) views. The three [4Fe-4S] clusters are depicted as spheres and colored by element. Individual subunits are shown as surface. FA is located in PsaA subunit while both FX and FB are located in PsaC subunit. Phylloquinone molecules are shown in blue, beta-carotenoids in orange, and Chl a molecules in green. Only pyrrole rings and Mg (gray) of Chl a molecules are shown for simplicity. The special pair of Chl a is shown in black. (E) Table showing all the prosthetic groups per subunit in a PSI monomer.

Figure 3

Potential lipid capacity and central cavity size: (A) Stromal and (B) luminal surface views of tetrameric PSI embedded in a model cyanobacterial membrane generated with CHARMM-GUI and embedded with VMD. The membrane contains 47% MGDG, 23% DGDG, 21% SQDG, and 9% PG galactolipids. (C) Stromal and (D) luminal views of the central cavity (blue) calculated with CASTp. Only chains that form the cavity are shown (PsaL on all four monomers, PsaI and PsaM on monomers B and B').

Figure 4

Interface Chl a's between PSI monomers of TS-821 tetrameric PSI vs *T. e.* BP-1 trimeric PSI and *Syn* PCC 6803 trimeric PSI: (A) Tetrameric PSI of TS-821 (left), trimeric PSI of *T. e.* BP-1 (PDB ID:1JB0) (center) and

trimeric PSI of Syn PCC 6083 (PDB-ID: 5OY0) are shown in surface; only chain PsaA (blue) and PsaB (green) are shown as cartoon. In all three structures, the pyrrole rings of Chl a associated with PsaA at the interface of the PSI monomers are high-lighted as purple and those associated with PsaB are high-lighted in orange. (B) A closer look of the interfaces: B'-A and A-B: The black and red loops are part of PsaB with sequences- QPKFRPS and MYRTNFGIGHS respectively. The distances are shown in yellow lines. The missing Chl a at the B-A' interface is shown in purple dashed pyrrole ring at a relatively similar position as it is in the mono-AB interface.

Figure 5

Interaction interface of PsaL subunits in trimeric (T. e. BP1) and tetrameric (6JE0 and 6K61). A) highlights the polar and non-polar residues that participate in the interaction interface at the interface of the three PsaLs of their respective monomers in T. e. BP-1. B) highlights the residues participating in the interaction interface between the two PsaLs of the A-B or A'-B' interface, RMSD- 0.708 (Nostoc and Anabaena aligned) and RMSD- 5.732 (TS-821 and Nostoc aligned). C) Shows the PsaL from all three tetramers (TS821, Nostoc and Anabaena) A-B interface (stromal and membrane view). It also shows the alignment of all the Chl a molecules in the three tetramers.

Figure 6

Phylogenetic Analysis of PsaL and motif-analysis. A) This represents a maximum-likelihood tree built using FastTree 2 on the multiple alignment of the 108 PsaL proteins using MUSCLE. The color shaded regions delineate separate groups that are known to include different forms of PSI including trimeric forms of PSI, far-red light forms of PSI, tetrameric forms of PSI, and a fourth group of marine cyanobacteria including members of *Prochlorococcus* and *Synechococcus*. Some species contain multiple copies of PsaL in their genome and these are denoted by colored circles (by species) and numbers (by number of PsaL copies). B) Logo plot of the loop region between the predicted TMD #2 and #3. The ending and beginning of the TMD are shown by colored boxes below the Logo sequence. The bit score scale was set to 6 bits to allow the error bars to be visible. The conserved motifs were shaded and named CR I-V. Within each CR the most conserved amino acids were further indicated by an asterisk above the single letter code. C) Similar to panel B, we display the Logo plots of the PsaL C-terminus of the four different groups also indicating the conserved regions (CR III-IV). D) Comparison of the CRs in TS-821, T. e. BP-1 and Syn PCC 6803 PSIs.