Title

Development of a TSR-based method for understanding structural relationships of cofactors and local environments in photosystem I.

Journal

BMC Bioinformatics, 26(1)

Authors

Luo, Lujun
Milon, Tarikul
Tandoh, Elijah
et al.

Publication Date

2025-01-14

DOI

10.1186/s12859-025-06038-y

Peer reviewed

BMC Bioinformatics

RESEARCH Open Access

Development of a TSR-based method for understanding structural relationships of cofactors and local environments in photosystem I

Lujun Luo[1], Tarikul I. Milon[1], Elijah K. Tandoh[1], Walter J. Galdamez[1], Andrei Y. Chistoserdov[2], Jianping Yu[3], Jan Kern[4], Yingchun Wang[5] and Wu Xu[1*]

[†]Lujun Luo and Tarikul I. Milon should be considered co-first authors.

*Correspondence: wxx6941@louisiana.edu

[1] Department of Chemistry, University of Louisiana at Lafayette, Lafayette, LA 70504, USA
[2] Department of Biology, University of Louisiana at Lafayette, Lafayette, LA 70504, USA
[3] Biosciences Center, National Renewable Energy Laboratory, Golden, CO 80401, USA
[4] Bioenergetics Department, MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[5] Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

Abstract

Background: All chemical forms of energy and oxygen on Earth are generated via photosynthesis where light energy is converted into redox energy by two photosystems (PS I and PS II). There is an increasing number of PS I 3D structures deposited in the Protein Data Bank (PDB). The Triangular Spatial Relationship (TSR)-based algorithm converts 3D structures into integers (TSR keys). A comprehensive study was conducted, by taking advantage of the PS I 3D structures and the TSR-based algorithm, to answer three questions: (i) Are electron cofactors including P700, A_{-1} and A_0, which are chemically identical chlorophylls, structurally different? (ii) There are two electron transfer chains (A and B branches) in PS I. Are the cofactors on both branches structurally different? (iii) Are the amino acids in cofactor binding sites structurally different from those not in cofactor binding sites?

Results: The key contributions and important findings include: (i) a novel TSR-based method for representing 3D structures of pigments as well as for quantifying pigment structures was developed; (ii) the results revealed that the redox cofactor, P700, are structurally conserved and different from other redox factors. Similar situations were also observed for both A_{-1} and A_0; (iii) the results demonstrated structural differences between A and B branches for the redox cofactors P700, A_{-1}, A_0 and A_1 as well as their cofactor binding sites; (iv) the tryptophan residues close to A_0 and A_1 are structurally conserved; (v) The TSR-based method outperforms the Root Mean Square Deviation (RMSD) and the Ultrafast Shape Recognition (USR) methods.

Conclusions: The structural analyses of redox cofactors and their binding sites provide a foundation for understanding the unique chemical and physical properties of each redox cofactor in PS I, which are essential for modulating the rate and direction of energy and electron transfers.

Keywords: TSR-based method, Photosystem I, Representation of cofactor 3D structures, Cofactor and protein interaction, Cofactor binding site and A and B branches

Introduction

Life on planet Earth is sustained largely by oxygenic photosynthesis. Oxygenic photosynthesis is a process in which higher plants, eukaryotic algae, and cyanobacteria convert CO_2 to chemical forms of energy, produce O_2 using sunlight and through that, they power the entire biological world [1, 2]. Virtually all oxygen in the atmosphere is thought to be generated through the photosynthetic process [3, 4]. This process can be divided into two chains of coordinated reactions: light reactions and dark reactions. In light reactions, sunlight is harnessed to synthesize ATP and NADPH from splitting H_2O into H^+, e^- and O_2. ATP and NADPH are utilized in the dark (i.e., light-independent) reactions to drive the synthesis of carbohydrates from CO_2. During light reactions, the four membrane-protein complexes, photosystem II (PS II), cytochrome b_6f, photosystem I (PS I) and ATP synthase, function in a coordinated way to initiate the photosynthetic process. PS I and PS II are involved in capturing sunlight and converting the absorbed energy into the energy of charge separation, i.e., act as natural "solar cells" that convert light into electrical current. The PS I complex of cyanobacteria contains twelve subunits (PsaA, B, C, D, E, F, I, J, K, L, M and X), chlorophylls (Chls) and carotenoid cofactors [5]. The electron transfer chain of PS I consists of six Chls, two phylloquinones and three [4Fe-4S] clusters. PsaA and PsaB are the core subunits that harbor the most antenna Chls, the primary electron donor P700 (a dimer of Chls), and a chain of electron acceptors A_{-1} (a Chl *a*), A_0 (a Chl *a*), A_1 (a phylloquinone) and F_X (a [4Fe-4S] cluster). The peripheral subunit PsaC binds the terminal electron acceptors F_A and F_B, two [4Fe-4S] clusters. Each individual electron transfer cofactor is labeled with a respective structural and spectroscopic name since they are located on both PsaA and PsaB sides of a pseudo-C2 axis of symmetry. There are two electron transfer chains starting from P700 ($P700_A$/$P700_B$), through the A branch (A_{-1A}, A_{0A} and A_{1A}) or the B branch (A_{-1B}, A_{0B} and A_{1B}) and converging at F_X. The antenna contains ~ 100 Chls [5].

Electron transfer is a fundamental process required for energy conversion in biological systems. Essential for electron transfer is the fine-tuning of the redox potentials of the electron acceptors and donors through interactions with the protein in which they are embedded [6] and the precise arrangement of cofactors with respect to each other. Therefore, it is critical to obtain a mechanistic understanding of interactions between cofactors, e.g., Chl and quinone, and between cofactors and their protein environments. The Triangular Spatial Relationship (TSR)-based method was developed for comparing molecular 3D structures [7] and probing drug and target interactions [8]. The input data for the TSR-based method are experimentally determined 3D structures from the Protein Data Bank (PDB) [9]. The first version of the TSR algorithm creates triangles with the C_α atoms of proteins as vertices. Triangles are constructed for every combination of three amino acids of a protein structure. A TSR key (an integer) is computed using geometric features such as length, angle, and vertex labels. Labels are determined by a rule-based assignment, which ensures consistent assignment of keys to identical TSRs across proteins, hence allowing a simpler but exact representation of protein structures [7]. Representation of 3D structures by TSR keys has its unique advantage of searching for similar substructures across structure datasets. In this study, we have developed a new version of the TSR-based method for understanding structural relationships of Chls

and quinones as well as structural relationships of Chl and quinone binding sites. The examples of electron cofactors (Chl and phylloquinone) used in this study are from PS I.

The crystal structure of PS I complex from the cyanobacterium *Thermosynechccocus elongatus* (thereafter *T. elongatus,* recently renamed to *T. vestitus*) was solved at 2.5 Å resolution [5]. This structure has been known for a long time and has had therefore a significant positive impact on functional studies of PS I. Plant and other cyanobacterial PS I structures were solved at 4.4 Å resolution [10], 3.4 Å resolution [11], 3.3 Å resolution [12], 2.8 Å resolution [13–15] and 2.6 Å resolution [16]. Over the last six years, the structural knowledge greatly increased with a large number of published structures (2018 [17, 18], 2019 [19–24], 2020 [25–32], 2021 [33–44], 2022 [45–52], 2023 [53–60], 2024 [61, 62]) from cyanobacteria and algae, some of them obtained under different light conditions and in different oligomeric states (monomer, trimer and tetramer forms). This wealth of information allows the architecture of pigments, cofactors and proteins to be accurately modeled at the atomic level. This study, by taking advantage of the available PS I 3D structures and the TSR-based algorithm, aims to answer three questions: (i) Cofactors of P700, A_{-1} and A_0 are Chl molecules. What are structural differences among P700, A_{-1} and A_0? (ii) What are structural differences between A-branch ($P700_A$, A_{-1A}, A_{0A} and A_{1A}) and B-branch ($P700_B$, A_{-1B}, A_{0B} and A_{1B}) cofactors and their corresponding binding sites? (iii) Are the amino acids in cofactor binding sites structurally different from the amino acids not in cofactor binding sites?

This study is organized into four sections. First, we discuss structural relationships of PsaA and PsaB polypeptides. Second, we report a method for representing 3D structures of Chl and phylloquinone and discuss the structural relationships of the pigments using such method. Third, we present the structural relationships of cofactor binding sites. Finally, we evaluate the TSR algorithm by comparing it with popular structural comparison methods. The main contribution to the method development includes a new representation of Chl and phylloquinone 3D structures. Key findings include correlations of cofactor structures or structures of cofactor binding sites with their functions. In summary, this work introduces a new computational method with advantages in understanding the structural foundation for determining the redox potentials of electron donors and acceptors. Through this extensive study of cofactor conformations and cofactor local protein environments, we have discovered unique substructures exclusively belonging to a certain type of cofactors or a specific binding site for a cofactor.

Experimental procedures

Key generation

Key generation method using C_α atoms, MaxDist and Theta was reported before [7]. Three vertices of triangle i are labeled as l_{i1}, l_{i2} and l_{i3} that are determined using a rule-based formula. MaxDist is defined as the distance of the longest edge of a triangle. Theta is defined as the angle that is $< 90°$ between the line from the midpoint of the edge of l_{i1} and l_{i2} to the opposite vertex l_{i3} and half of the l_{i1}—l_{i2} edge. The Python code for C_α key generation is available in the supplementary document.

Protein structural similarity and distance calculation

The Generalized Jaccard coefficient measure [63] was used for calculating pairwise similarity between any two protein structures in a dataset [7]. The distance matrix is derived from the similarity matrix [7]. Protein structure clustering is visualized based on Average Linkage Clustering [64]. The complexity of the multiple dimensional relations among 3D structures is reduced and represented by the Multidimensional Scaling (MDS) method [65]. Structural images were prepared using the Visual Molecular Dynamics (VMD) package [66].

Development of a new version of the TSR-based method for pigments

To quantify the structures of pigments including Chls and quinones, a new version of the TSR-based method has been developed where every possible triangle is constituted from all the atoms except hydrogen atoms in a pigment. The bin boundaries used for Theta were the same as those we reported for the TSR algorithm using C_α atoms [7]. Seventeen bins, about half the number of the MaxDist bins for the C_α TSR algorithm [8], with one angstrom as an interval were used for MaxDist. To generate TSR keys for pigments, information on PDB ID, chain and pigment name and ID is needed. Each cofactor of PS I complexes from different species was annotated by examining structures using VMD. Each type of atoms was assigned an integer. An atom filtering algorithm was developed to select specific atoms for TSR key generation.

Development of a TSR algorithm for quantifying structures of amino acids

The TSR concept was used to develop an algorithm for quantifying the structures of different amino acids and same amino acids at the different positions. All atoms except hydrogen atoms of every amino acid were used for TSR key generation. The bin boundaries used for Theta were the same as those we reported for the TSR algorithm using C_α atoms [7]. Fifty-eight bins with one angstrom as an interval were used for MaxDist. Normalized Jaccard coefficient measure is used for calculation of similarity between two amino acids.

Sequence alignment

The MUSCLE module of SnapGene was applied to conduct multiple sequence alignments. Phylogenetic studies of protein sequences were conducted using the MEGA software [67].

Dataset preparation

The datasets containing 3D structures of PS I complexes from plants, cyanobacteria and algae cultured under white or red light, normal or high light, and normal or high temperature conditions were prepared. All pigments and proteins in the datasets were selected from the PDB [9]. The PDB IDs, chains, pigment names and IDs can be found in Supplementary File 1.

Output files from key generation code and definition of different types of TSR keys

Two output files were generated from the key generation step for every molecule, either a protein, a cofactor or an amino acid. One output is named "key file" and data structure of a "key file" is an integer (TSR key) vector for representing a 3D structure of protein, cofactor or amino acid. The other output is referred as "triplet files" containing the details for three amino acids and their positions, MaxDist and Theta values and the key for each protein. If for cofactor or amino acids, a "triplet file" contains the details for three atoms, MaxDist and Theta values and the key. The keys using C_α atoms are called *CA* TSR keys. The keys for a pigment are called *Cofactor* TSR keys. The keys for amino acids are called *AA* TSR keys. For *CA, Cofactor and AA* TSR keys, they can be further divided into *distinct*, *total*, *distinct* and *total common*, and *distinct* and *total specific TSR keys* that were reported before [68] to reveal structural relationships. Calculations of every type of TSR key (*distinct*, *total*, *distinct common, total common*, *distinct specific* or *total specific*) is accomplished through integer search using "key files". A TSR key is an integer and only the integer is not biologically and chemically meaningful. If a key of interest is identified through key search, the key needs to be mapped into the triangle(s) with the details of three C_α atoms and MaxDist and Theta values for *CA* TSR or with the details of three atoms and MaxDist and Theta values for *Cofactor* or *AA* TSR through searching the key in the "triplet files".

Ultrafast shape recognition method

The Ultrafast Shape Recognition (USR) method was developed by Ballester's group [69]. In this method, the set of all atomic distances from four molecular locations are considered: the molecular centroid (ctd), the closest atom to ctd (cst), the farthest atom to ctd (fct), and the farthest atom to fct (ftf). Each set of four distances can be regarded as a distribution. The first three moments are used for each distribution. Therefore, USR encodes the shape of a molecule and creates pairwise similarity output using 12 descriptors. The pairwise similarity output is the input file for hierarchical cluster analyses. The Python codes to compare structures of proteins, cofactors and amino acids using the USR method are available in the supplementary document.

Root mean square deviation method

The Root Mean Square Deviation (RMSD) method calculates the minimum value of the root-mean-square distance between all possible one-to-one matchings between the atoms in the superimposed structures [70]. Pairwise structural differences for protein C_α atoms were calculated using the TM-align method [71]. The Python codes to compare structures of cofactors and amino acids using the RMSD method are available in the supplementary document. The pairwise distance output files from the TM-align method are the input files for hierarchical cluster analyses.

Statistical analyses

T-test was used to identify statistical differences between the different feature engineering methods' similarity values. A threshold of $p < 0.05$ was used to determine significance.

Results

3.1 The analysis of PsaA and PsaB structures using CA TSR keys has identified the specific substructures exclusively belonging to a certain organism or a particular cell-culture condition

The analysis of the entire PsaA and PsaB structures reveals high structural similarities of PsaA, PsaB and between PsaA and PsaB from different photosynthetic organisms

The hierarchical cluster analysis demonstrates that the TSR-based method can distinguish PsaA structures from PsaB structures and vice versa. One exception was observed where PsaA and PsaB from *Acaryochloris marina* form a separated small cluster besides a large PsaA cluster and a large PsaB cluster (Fig. 1a). As expected, both PsaA (an average of 80.3% among different structures) and PsaB (an average of 79.3% among different structures) have high structural similarities as well as those between

(See figure on next page.)

Fig. 1 Hierarchical cluster and MDS analyses of PsaA and PsaB from diverse organisms demonstrate the capacity of the TSR algorithm for distinguishing their 3D structural differences and the substructures of PsaA or PsaB exclusively belonging to either red or white light cultural conditions were identified. Panel **a**, the hierarchical cluster analysis shows 3D structural relationships of PsaA and PsaB. The numbers of PsaA and PsaB sequences used in the analysis are labeled; panel **b**, calculations of the overall structural similarity of PsaA and PsaB and structural similarities for individual PsaA and PsaB structures. The average similarity values, SDs, and 25/75 percentiles are indicated; panel **c**, the Venn diagram of the numbers of TSR keys exclusively belonging to the PsaA protein class or the PsaB protein class and commonly shared between the PsaA and PsaB protein classes; panel **d**, the same PsaA and PsaB structures, presented in panel (**a**), analyzed using the MDS method. The numbers of PsaA and PsaB structures and the PsaA and PsaB structures for the organism of *Acaryochloris marina* are labeled; panel **e**, the *specific* TSR keys identified for PsaA and PsaB of *H. hongdechloris* cultured under both red and white light culture conditions. Average values are labeled; panel **f**, the 3D substructures corresponding to the two *specific* TSR keys (9102616 and 9346188) that exclusively belong to PsaA of *H. hongdechloris* cultured under the red-light conditions. The triangle corresponding to the key 9102616 is constructed from the three C_α atoms from PsaA-H357, PsaA-N359 and PsaA-H363. The triangle corresponding to the key 9346188 is constructed from the three C_α atoms from PsaA-A358, PsaA-N359 and PsaA-H363; panel **g**, the 3D substructures corresponding to the three *specific* TSR keys (5038551, 6662534 and 9098508) that exclusively belong to PsaA of *H. hongdechloris* cultured under the red-light conditions. The triangle corresponding to the key 5038551 is constructed from the three C_α atoms from PsaA-C464, PsaA-H465 and PsaA-M478. The triangle corresponding to the key 6662534 is constructed from the three C_α atoms from PsaA-C464, PsaA-H465 and PsaA-T674. The triangle corresponding to the key 9098508 is constructed from the three C_α atoms from PsaA-C464, PsaA-H465 and PsaA-N466; panel **h**, the 3D substructures corresponding to the two *specific* TSR keys (3416607 and 8818371) which exclusively belong to PsaB of *H. hongdechloris* cultured under the red-light conditions. The triangle corresponding to the key 3416607 is constructed from the three C_α atoms from PsaB-H196, PsaB-H218 and PsaB-E220. The triangle corresponding to the key 8818371 is constructed from the three C_α atoms from PsaB-R216, PsaB-H218 and PsaB-P219; panel **i**, the 3D substructures corresponding to the two *specific* TSR keys (5103509 and 9183774) that exclusively belong to PsaB of *H. hongdechloris* cultured under the white light conditions. The triangle corresponding to the key 5103509 is constructed from the three C_α atoms from PsaB-K302, PsaB-M305 and PsaB-H308. The triangle corresponding to the key 9183774 is constructed from the three C_α atoms from PsaB-M305, PsaB-N306 and PsaB-H308; panel **j**, the 3D substructures corresponding to the two *specific* TSR keys (7071561 and 7194413) that exclusively belong to PsaA of *H. hongdechloris* cultured under the white-light conditions. The triangle corresponding to the key 7071561 is constructed from the three C_α atoms from PsaA-H253, PsaA-W255 and PsaA-L257. The triangle corresponding to the key 7194413 is constructed from the three C_α atoms from PsaA-H253, PsaA-W255 and PsaA-P259; For panels **f**–**j**, the IDs of the PDB and chlorophylls are labeled

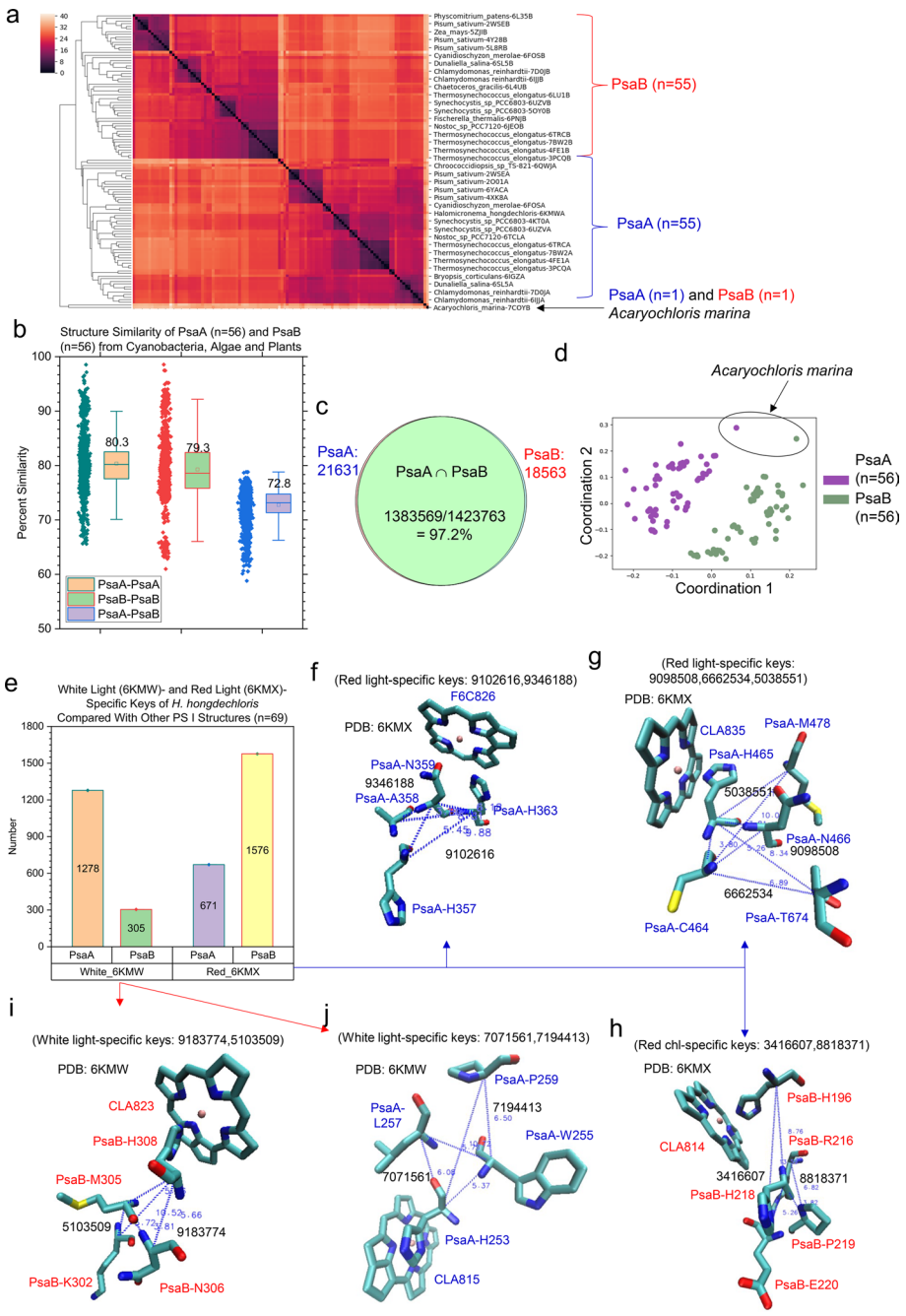

Fig. 1 (See legend on previous page.)

PsaA and PsaB (an average of 72.8%) (Fig. 1b). Such high structural similarities are supported by a high percentage of distinct ($78.1\% = 9.76\times10^5/1.25\times10^6$ for PsaA, $77.9\% = 9.66\times10^5/1.24\times10^6$ for PsaB, $72.3\% = 8.97\times10^5/1.24\times10^6$ for PsaA and PsaB,) and total ($98.5\% = 6.59\times10^7/6.69\times10^7$ for PsaA, $97.6\% = 6.43\times10^7/6.59\times10^7$ for PsaB, $97.6\% = 6.44\times10^7/6.60\times10^7$ for PsaA and PsaB) *common* keys (Supplementary Fig. 1). The Venn diagram provides additional evidence of a high structural similarity between PsaA and PsaB (Fig. 1c). The result from the MDS analysis of PsaA and PsaB structures

supports that from the hierarchical cluster analysis (Fig. 1d). Distinct and total *specific* keys were identified for PsaA and PsaB (Supplementary Fig. 2). Those keys represent unique substructures exclusively belonging to PsaA or PsaB.

The structural analyses identified substructures exclusively belonging to a certain oligomer form of PS I, a certain culture condition and a certain type of pigment containing organisms

In cyanobacteria, PS I exists as a trimer or monomer, and possibly a tetramer [72]. Depending on environmental conditions, trimer may be shifted to monomer and vice versa, suggesting that each form functions slightly differently, which may also translate into structural changes [73]. Specific substructures represented by *specific* TSR keys exclusively belonging to a trimer (PDB: 5OY0) [17] or a monomer (PDB: 6HQB) [73] (Supplementary Fig. 3a) were identified. One example of three trimer-*specific* keys is shown in Supplementary Fig. 3b. Those trimer-*specific* keys-associated triangles are close to three Chl molecules (CLA1218, CLA1219 and CLA1220) (Supplementary Fig. 3b) that could be a part of red pigments. Two examples of monomer-*specific* keys are shown in Supplementary Fig. 3c and 3d. One example containing two monomer-*specific* keys and their associated triangles are close to two Chl molecules (CLA1108 and CLA1109) (Supplementary Fig. 3c). The other monomer-specific triangles are not close to any Chl molecules (Supplementary Fig. 3d). As expected, the PsaA polypeptides from trimer and monomer have identical amino acid sequences (Supplementary Fig. 4). This is the case for PsaB as well (Supplementary Fig. 4). The *specific* keys identified for trimer or monomer demonstrate that the TSR keys can be used to quantify conformational changes induced by oligomerization or de-oligomerization. Because the conformational changes are close to Chls, it may explain absorption differences between trimers and monomers of PS I.

The data for cyanobacterium *Halomicronema hongdechloris* indicated that its Chl *f* functions to harvest the far-red light. This resulted in changes of the PS I gene expression favoring PsaA and PsaB for binding of Chl *f* [31]. The sequence alignment analysis of PsaA and PsaB from *Halomicronema hongdechloris* (6KMW: white light and 6KMX: far-red light) shows the difference in amino acid sequences under different light conditions (Supplementary Fig. 5). To understand the structural changes induced or partially induced by white or far-red light, we further investigated the structures of PsaA and PsaB from this cyanobacterium grown under these light conditions. The *specific* TSR keys exclusively for white light or far-red light conditions were identified for both PsaA and PsaB (Fig. 1e). The details of seven far-red-light-*specific* keys and four white-light-*specific* keys were analyzed. Two far-red-*specific* keys are close to a Chl *f* (F6C826) (Fig. 1f), three *specific* keys are close to a Chl *a* (CLA835) (Fig. 1g) and two *specific* keys are close to a Chl *a* (CLA814) (Fig. 1h). Two similar examples were identified for white-light-*specific* keys. One example shows two keys that are close to CLA823 (Fig. 1i) and another example shows that two different keys are close to CLA815 (Fig. 1j). Cyanobacterium *Acaryochloris marina* also has the ability in absorbing far-red light. The special pair in this cyanobacterium is a dimer of Chl *d* and its epimer Chl *d*' [34] rather than a dimer of Chl *a* and its epimer Chl *a*' found in other species. Also, the primary electron acceptor is pheophytin *a* [34] instead of Chl *a*. Like the situations for trimer vs. monomer and white light vs. far-red light, we were able to identify the substructures

exclusively belonging to only Chl *a*-containing, Chl *f*-containing and Chl *d*-containing organisms (Supplementary Fig. 6). The specific TSR keys discussed in this section reveal cofactor-specific protein environments that may contribute to absorption of a specific wavelength of light.

Development of the TSR algorithm for representing 3D structures of electron transfer cofactors and the hierarchical cluster analysis revealed that the clusters of the electron donor or acceptor generally match with their functions

Development of the TSR algorithm for representing 3D structures of electron transfer cofactors

There are ~ 100 Chl molecules and 2 phylloquinone molecules based on a high-resolution crystal structure of cyanobacterial PS I [74]. Six out of these Chl molecules function as either an electron donor or acceptor and they form the reaction center. Two out of 96 Chl molecules may function as the linkers to connect the reaction center with the rest of the antenna Chl molecules. These two Chl molecules are called connecting Chls (A_C). One is on PsaA side named A_{CA} and the other is on PsaB side named A_{CB} in this study. Similarly for the electron donors and acceptors, P700, A_{-1}, A_0 and A_1 on the PsaA side are named $P700_A$, A_{0A} and A_{1A} whereas they are named $P700_B$, A_{0B} and A_{1B} if they are on the PsaB side. A_{-1A} and A_{-1B} are the only electron transfer cofactors, in which PsaA binds A_{-1B} and PsaB binds A_{-1A}. $P700_A$, A_{-1A}, A_{0A} and A_{1A} are called A-branch electron transfer cofactors whereas $P700_B$, A_{-1B}, A_{0B} and A_{1B} are called B-branch electron transfer cofactors. To understand whether the electron donors, electron acceptors and

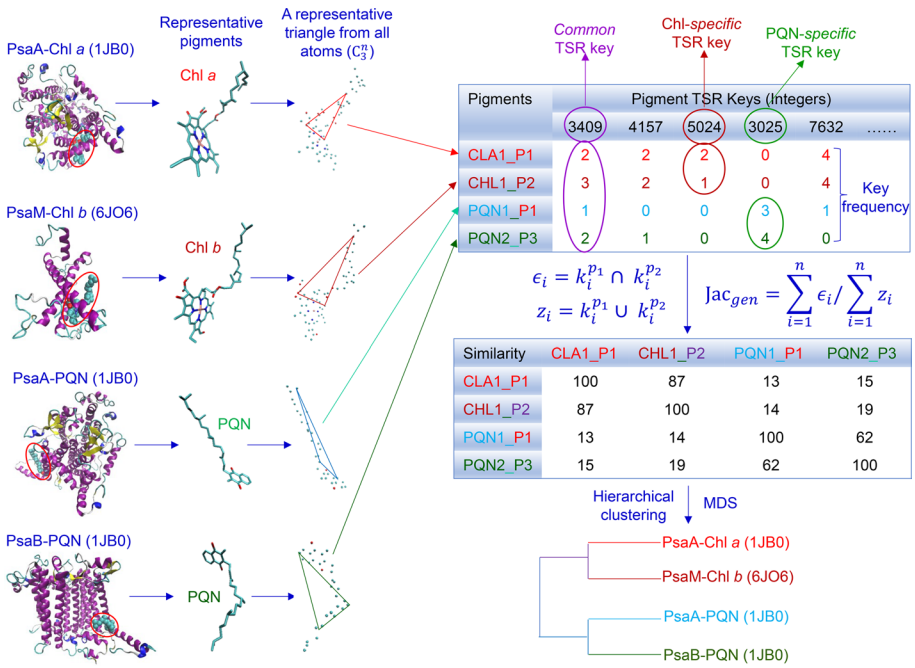

Fig. 2 The design of the TSR algorithm for representing 3D structures of chlorophyll and phylloquinone molecules. It illustrates the schema of how to decode chlorophyll and phylloquinone 3D structures to integers (TSR keys) and how to calculate pairwise structural similarities using the calculated TSR keys and the Generalized Jaccard coefficient approach. An example of *common* and *specific* keys is shown. A hypothetical hierarchical cluster result is also shown

connecting Chls have their specific structural characteristics, a novel method to represent their 3D structures was developed.

In this method, first, all atoms except hydrogen atoms of a pigment 3D structure are selected and all possible triangles constructed by the atoms are identified (Fig. 2). Second, three vertex labels are determined using the rule-based assignment. Third, TSR keys (integers) and key occurrence frequencies are calculated. Fourth, pairwise similarities between pigment 3D structures are calculated using the Generalized Jaccard similarity through computing identical and nonidentical keys, and their frequencies (Fig. 2). The 3D structures of pigments are represented by a vector of integers (*Cofactor* TSR keys). Such a representation for pigments is unique. The important objectives of this algorithm are to quantify structural similarities of pigments and provide insight into structural relationships through identifying *specific* and *common Cofactor* TSR keys (Fig. 2).

The clustering analysis has demonstrated that the electron donors and electron acceptors have unique structural characteristics

The selection of the structures of the electron donors and electron acceptors is based on the criteria: (i) whether it is a model structure of PS I; (ii) whether it comes from a model organism for photosynthesis research; (iii) whether PS I structures from a diversity of organisms are available and (iv) whether a high-resolution PS I structure is available. The crystal structure from *T. vestitus*, a type of thermophilic cyanobacteria, is the model PS I structure (PDB: 1JB0). *Chlamydomonas reinhardtii* (PDB: 6JO6), a eukaryotic green alga, and *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*) (PDB: 5OY0), a strain of unicellular and freshwater cyanobacteria, are the model organisms for photosynthesis research and numerous functional studies were conducted in these two organisms. PDB 6PNJ contains a PS I structure from filamentous true-branching cyanobacterium *Fischerella thermalis* and PDB 5ZJI contains a PS I structure from a plant (*Zea mays*). Therefore, five PDBs (1JB0, 5OY0, 6JO6, 6PNJ and 5ZJI) were selected for the structural study of cofactors.

The hierarchical cluster analysis shows that P700, A_{-1} and A_0 form their own clusters, indicating they have their individual structural characteristics. More structural diversity was observed for A_C. A_{CA} chlorophylls form their own cluster with an exception that one A_{CA} molecule is joined with the P700 cluster (Fig. 3a). A_{CB} structures also form their own cluster. However, the A_{CB} cluster is separated from the A_{CA} cluster (Fig. 3a). P700, A_{-1} and A_0 have similar structural similarities, but they have higher structural similarities than A_C (Fig. 3b). P700s have higher structural similarities among themselves than those when P700s were compared with A_{-1}, A_0 and A_C (Fig. 3b). It is true also for A_{-1}, A_0 and A_C (Fig. 3b). All these (Fig. 3b) support the clustering result obtained for P700, A_{-1} and A_0 (Fig. 3a). In addition, the analysis of structural similarity (Fig. 3b) demonstrates that A_C chlorophylls also have their structural characteristics. If we consider ten molecules of each type of cofactors as a group, P700, A_{-1}, A_0 and A_C groups share 78.2% of identical *Cofactor* TSR keys (Fig. 3c), suggesting that they have a high similarity percentage among four groups of cofactors as expected. If we consider individual cofactors, all forty P700, A_{-1}, A_0 and A_C pigments have 1350 distinct *common* keys (without considering key occurrence frequency) and 39,600 total *common* keys (with considering

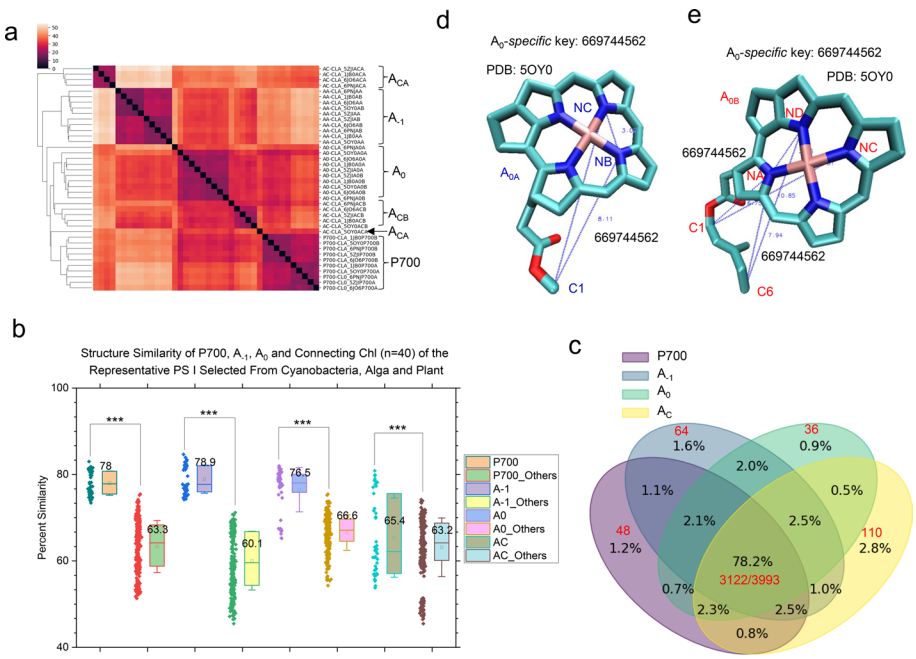

Fig. 3 Hierarchical cluster analysis of electron cofactors P700, A_{-1}, A_0 and connecting chlorophyll molecules demonstrating the capacity of the TSR algorithm, which can distinguish their 3D structural differences. Panel **a**, the hierarchical cluster analysis of 3D structural relationships between P700, A_{-1}, A_0 and connecting chlorophyll molecules. A_{CA} (A_{CB}) represents the connecting chlorophyll molecule on the PsaA (PsaB) side; panel **b**, the side-by-side structural comparisons between one type of cofactor pairs and between that type of cofactor and other types of cofactor pairs. The average similarity values, SDs, and 25/75 percentiles are indicated. *** means a *p* value is less than 0.001 using a *t*-test; panel **c**, the Venn diagram showing the numbers of TSR keys exclusively belonging to each type of cofactor class (the P700 cofactor class, A_{-1} cofactor class, A_0 cofactor class and the connecting chlorophyll (A_C) class), and the regions between any two, three and four cofactor classes; panels **d** and **e**, one A_0 *specific* TSR key (669744562). The triangle corresponding to this key is constructed from three atoms (C1, NB and NC) of A_{0A} and is shown in panel (**d**), and the triangle corresponding to the key 669,744,562 constructed from three atoms (C1, NA and NC or C6, NA and ND) of A_{0B} is shown in panel (**e**). The PDB ID is labeled

key occurrence frequency) on average (Supplementary Fig. 7). Those *common* keys can be found in each pigment of the 40 pigments. The 40 pigments have 2700 distinct and 43,400 total keys on average (Supplementary Fig. 7). P700, A_{-1}, A_0 and A_C share roughly 50% (1350/2700) to 91.2% (39,600/43,400) of common substructures. One *specific* key, 669,744,562, exclusively identified for A_0, is found in all A_0 molecules but not in P700, A_{-1} and A_C. This A_0-*specific* key is shown in Fig. 3d for A_{0A} with one occurrence frequency and in Fig. 3e for A_{0B} with two occurrence frequencies. To further examine the structural differences of the cofactors between the PsaA side and the PsaB side, we performed cluster analyses of each type of cofactors. The $P700_A$ cluster and the $P700_B$ cluster are distinct (Supplementary Fig. 8). In contrast, the A_{-1A} and A_{-1B} structures tend to cluster together (Supplementary Fig. 9). Two A_{-1} clusters were observed. However, each cluster is a mix of A_{-1A} and A_{-1B}, suggesting A_{-1} structural diversity among different species. The situations for A_{0A} and A_{0B} (Supplementary Fig. 10) as well as for A_{CA} and A_{CB} (Supplementary Fig. 11) are found to lie between P700 and A_{-1}.

We also performed the hierarchical cluster analysis of A_1 molecules. Most A_{1A} structures group together. It is also the case for most A_{1B} structures. However, two A_{1A} and six A_{1B} structures cluster together (Fig. 4a). The average structural similarities of A_{1A}

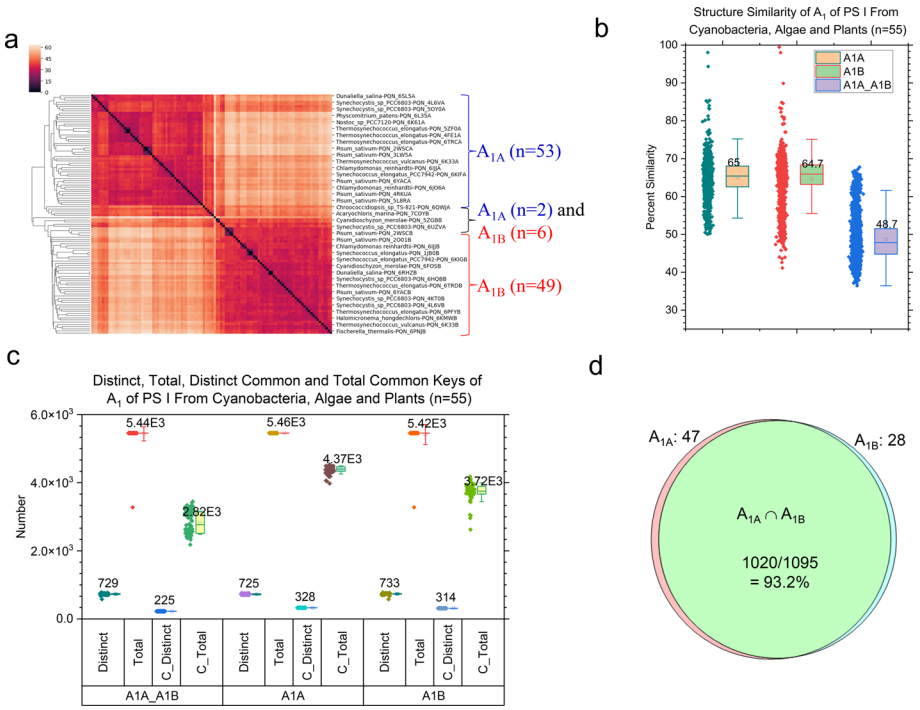

Fig. 4 Hierarchical cluster analysis of electron cofactors A_1 demonstrating the capacity of the TSR algorithm for distinguishing of their 3D structural differences. Panel **a**, the hierarchical cluster analysis of 3D structural relationships of A_{1A} and A_{1B}. The numbers of A_{1A} and A_{1B} chlorophylls used in the analysis are labeled; panel **b**, calculated overall structural similarities of A_{1A} and A_{1B} and structural similarities for individual A_{1A} and A_{1B} structures. The average similarity values, SDs, and 25/75 percentiles are indicated; panel **c**, the combined distinct, total, distinct common and total common TSR keys for A_{1A}, A_{1B} and A_{1A} and A_{1B}. The average numbers, SDs, and 25/75 percentiles are indicated; panel **d**, the Venn diagram showing the numbers of TSR keys exclusively belonging to the A_{1A} cofactor class or the A_{1B} cofactor class and commonly shared between the A_{1A} and A_{1B} cofactor classes

and A_{1B} are nearly the same. The average structural similarity between A_{1A} and A_{1B} are lower than that of A_{1A} as well as that of A_{1B} as expected (Fig. 4b). Taken together, the results suggest that A_{1A} and A_{1B} have their structural characteristics with their structural diversities. A_{1A} and A_{1B} share 30.9% (225/729) of distinct *common* keys and 51.8% (2820/5440) of total *common* keys (Fig. 4c). It indicates that A_{1A} and A_{1B} have 30.9% to 51.9% similar substructures. If we consider the keys from all A_{1A} molecules as a group and the keys from all A_{1B} molecules as another group, both groups share a high percentage of the same keys (Fig. 4d). Only very small portions of the keys were found exclusively belonging to either group (Fig. 4d).

A role of local environments for understanding the mechanisms underlying cofactor–protein interactions

Distance calculations reveal the difference in the arrangement of overall Chl molecules between cyanobacterial and eukaryotic photosynthetic organisms and specific arrangement of electron transfer cofactors

Both the chromophore—chromophore interaction strength and the chromophore—environment interaction coupling are important for modulating energy and electron

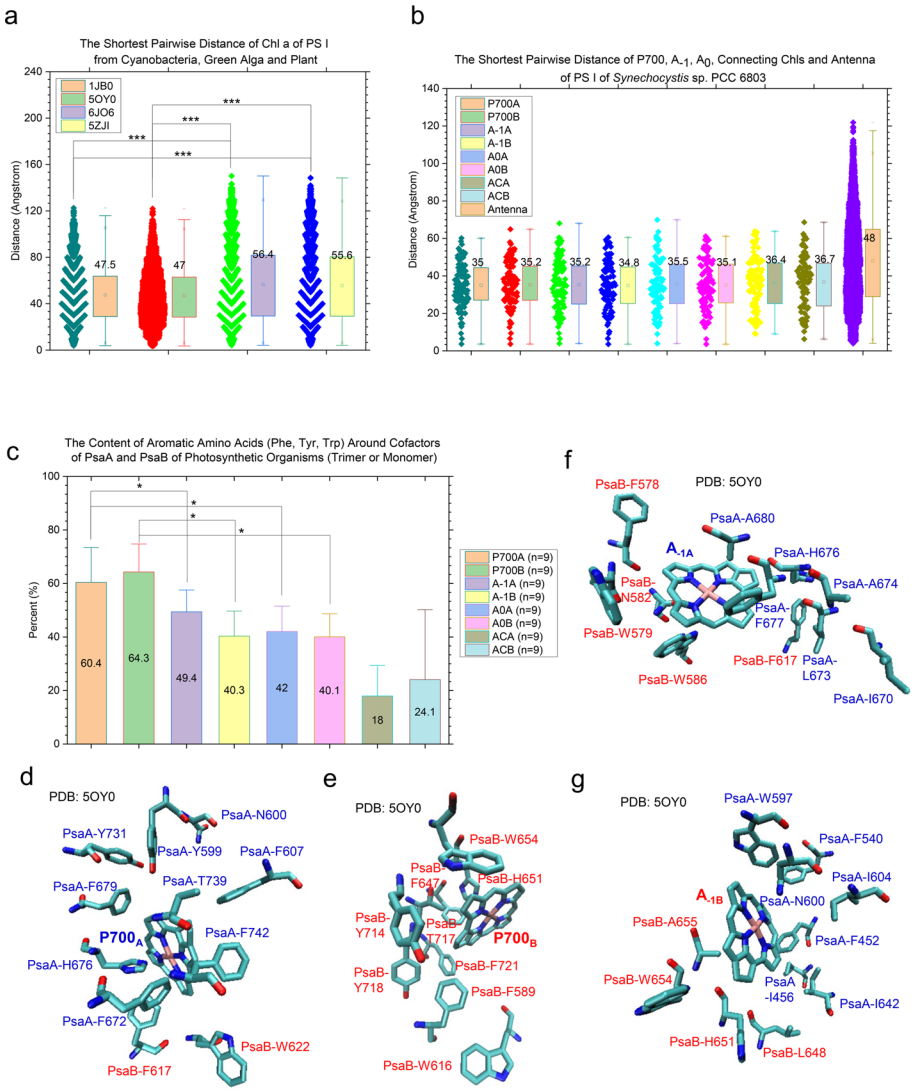

Fig. 5 The difference in global arrangement of chlorophyll molecules between representative cyanobacteria and alga/plant, structural characteristic of the reaction centers and more aromatic residues closely interacting with P700. Panel **a**, he shortest distances between chlorophyll—chlorophyll pairs were calculated and are present. The PDB IDs are labeled. *** means a *p* value is less than 0.001 in a *t*-test; panel (b), the shortest distances between $P700_A$—other chlorophyll pairs, $P700_B$—other chlorophyll pairs, A_{-1A}—other chlorophyll pairs, A_{-1B}—other chlorophyll pairs, A_{0A}—other chlorophyll pairs, A_{0B}—other chlorophyll pairs, A_{CA}—other chlorophyll pairs, A_{CB}—other chlorophyll pairs and chlorophyll pairs (including antenna). In panels **a** and **b**, the average values, SDs, and 25/75 percentiles are indicated; panel **c**, the numbers of aromatic residues (phenylalanine, tyrosine and tryptophan) that have close interactions with $P700_A$, $P700_B$, A_{-1A}, A_{-1B}, A_{0A}, A_{0B}, A_{CA} and A_{CB}. The percentages of those aromatic residues are present. The cutoff value for the close interactions is 3.5 Å. The average values and SDs are indicated. * means a *p* value is less than 0.05 using a *t*-test; Panels **d** through **g**, the representative surrounding amino acids of $P700_A$ **d**, $P700_B$ (**e**), A_{-1A} (**f**) and A_{-1B} (**g**) are illustrated. The PDB is 5OY0. The numbers of aromatic residues that have closely interactions with P700 is larger than those that closely interact with A_{-1} or A_0

transfers in PS I. Therefore, the shortest distances between each Chl pair were calculated and described herein. The description of the chromophore environment is discussed in the next section. These calculations show that eukaryotic photosynthetic organisms (green alga and plant) have larger pairwise distances between Chl pairs

than those of cyanobacteria (Fig. 5a), suggesting differences in the antenna arrangement between cyanobacterial and higher plant systems which agree with the data described earlier in the literature [75]. The reaction centers including P700, A_{-1} and A_0 of oxygenic organisms exhibit a common general architecture and share the same basic functional principles. Two connecting Chl molecules (A_{CA} and A_{CB}) are special as they structurally and perhaps functionally connect A_{-1} and A_0 of the electron transfer chains to the antenna. Therefore, we include the connecting Chl molecules in the analysis. The result reveals that P700, A_{-1}, A_0 and A_C have similar pairwise distances with the antenna Chls (Fig. 5b), suggesting unique positions of the redox factors in the cofactor—protein complex. It also suggests the special locations of two A_C molecules for connecting antenna Chls to the reaction center. In summary, the global arrangement of Chls is different between prokaryotic PS I and eukaryotic PS I and the locations of two A_C molecules are special that may indicate their specific functions.

Calculations of number and type of amino acids surrounding the cofactors reveal differences in local environments among different cofactors and between the A branch and the B branch

PS I are characterized by optimized structures where the protein scaffold acts on the energy and electron transfer cofactors, finely tuning their surroundings and modulating their properties and functionalities [76]. Numerous studies have addressed the contributions of individual amino acids to modulating the spectroscopic properties of bound redox cofactors. The subtle structural differences in cofactor binding sites have not been reported. First, the number of surrounding residues for the cofactors was investigated. Diverse cyanobacteria have similar average numbers of surrounding amino acids for Chl molecules while eukaryotic organisms have slightly lower numbers of surrounding amino acids (Supplementary Fig. 12). Significant numbers of Chl *a* and Chl *d* molecules were identified. Chl *d* molecules have slightly more surrounding amino acids than Chl *a* molecules (Supplementary Fig. 13). Second, the differences in cofactor binding sites between the A branch and the B branch were examined. The result shows that A_{-1A} has more surrounding amino acids than A_{-1B} (Supplementary Fig. 14). It is also true for the A_{CA} and A_{CB} binding sites (Supplementary Fig. 14). The redox cofactors (P700, A_{-1} and A_0) have more amino acids than A_C and the rest of Chls (Supplementary Fig. 14). For the P700 and A_0 binding sites, there is no difference between two branches (Supplementary Fig. 14). We found that the P700 binding sites have more aromatic residues than those of A_{-1} and A_0 (Fig. 5c). In contrast, A_{CA} and A_{CB} have less numbers of aromatic residues than the cofactors in the reaction center (Fig. 5c). Figure 5d, e, f and g, illustrate the representative examples of the binding sites of $P700_A$, $P700_B$, A_{-1A} and A_{-1B}, respectively. For the A_1 binding sites, the B branch has more amino acids than the A branch (Supplementary Fig. 15). As the references, we calculated the amino acid compositions including aromatic residues for PsaA and PsaB (Supplementary Figs. 16 and 17). The top three most abundant amino acids for PsaA (Supplementary Fig. 16) and for PsaB (Supplementary Fig. 17) are Leu, Gly and Ala. Interestingly, we observed more surrounding aromatic residues for P700 than A_{-1} and A_0 and a difference in the number of surrounding amino acids between the A branch and the B branch for A_{-1}, A_C and A_1.

The structural analysis using CA TSR keys demonstrates the differences of the redox cofactors between the A branch and the B branch

To study the binding sites of the redox cofactors, we included four more PS I structures for increasing diversity of cell-culture conditions, different oligomeric forms and the reaction centers. Two structures, 6KMW (white light) and 6KMX (far-red light) from *Halomicronema hongdechloris*, a cyanobacterium that produces Chl *f*; one monomeric PS I structure, 6HQB from the model photosynthetic organism *Synechocystis* and one structure, 7COY from far-red light utilizing PS I of *Acaryochloris marina* where Chl *d* and pheophytin are in the reaction center, are included in the study. The amino acids and their positions in PsaA and PsaB were labeled for different species. Because the amino acid positions for closely contacting the cofactors could be different for different species, for labeling the amino acid positions, the multiple sequence alignment analysis was performed (Fig. 6). The amino acids and their positions for $P700_A$ and $P700_B$ are summarized in Table 1 and those for A_{-1}, A_0 and A_1 are listed in Supplementary Tables 1, 2 and 3, respectively. The pigment numbers for $P700_A$, $P700_B$, A_{-1A}, A_{-1B}, A_{0A}, A_{0B}, A_{CA} and A_{CB} from different species are shown in Table 2. For the rest of the sections in this study, the nomenclature of *Synechocystis* is used.

The hierarchical cluster analysis clearly shows that the binding sites of each type of redox cofactors form their own clusters. The resulting four clusters: P700, A_{-1}, A_0 and A_1 can further be divided into two distinct subclusters: one for the binding sites in PsaA and the other for their corresponding sites in PsaB (Supplementary Fig. 18a). This result demonstrates the structural characteristics of the binding sites of each type of redox cofactors from either PsaA side or PsaB side and suggests a difference in redox potential of the cofactors between both sides. The pairwise structural

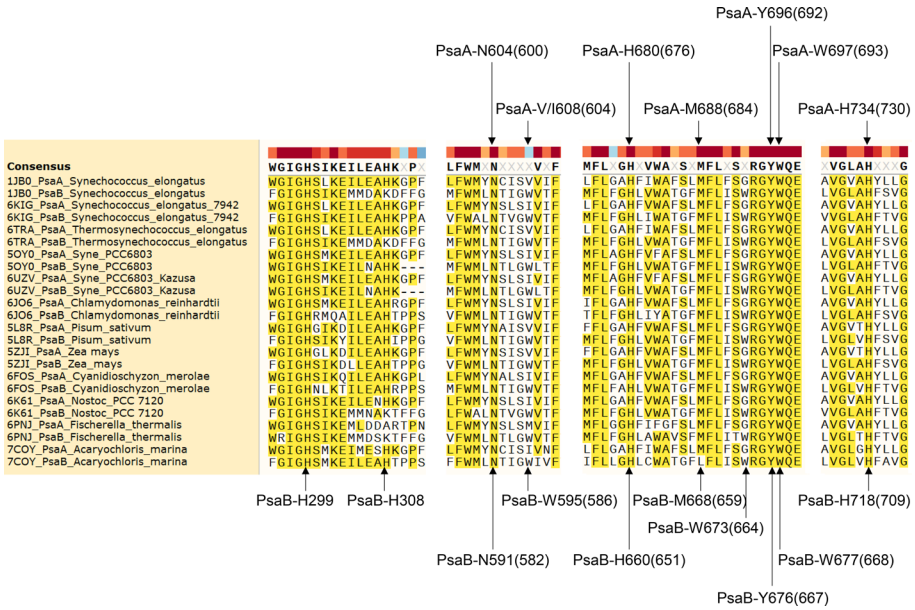

Fig. 6 The multiple sequence alignment of PsaA and PsaB from the representative organisms showing the conserved residues for closely interacting with P700, A_{-1}, A_0, A_C, A_1 and possibly red chlorophyll molecules. The critical residues for interacting P700, A_{-1}, A_0, A_C, A_1 and possibly red chlorophyll molecules are labeled. The nomenclatures for *Synechococcus elongatus* and *Synechocystis* sp. PCC 6803 (in the parenthesis) are used

Table 1 The Amino Acids that Interact with $P700_A$ and $P700_B$ and Their Positions

PDB	Organism and resolution	Cofactors	Axial ligand	Interacting amino acids		Comments
				PsaA	PsaB	
1JB0	*Synechococcus elongatus* (2.5 Å)	P700A	PsaA-H680	Y603, N604, F611, F676, W683, Y735, T743, F746	L626, W631	CyanobacteriumTrimer
		P700B	PsaB-H660		F598, W625, F656, W663, Y723, T726, Y727, F730	
5OY0	*Synechocystis sp. PCC 6803* (2.5 Å)	P700A	PsaA-H676	Y599, N600, F607, F672, F679, Y731, T739, F742	F617, W622	CyanobacteriumTrimer
		P700B	PsaB-H651		F589, W616, F647, W654, Y714, T717, Y718, F721	
6HQB	*Synechocystis sp. PCC 6803* (4.0 Å)	P700A P700B	PsaA-H676 PsaB-H651	Y599, N600, F607, F672, F679, Y731, T739, F742	F617, W622 F589, W616, F647, W654, Y714, T717, Y718, F721	CyanobacteriumMonomer
6JO6	*Chlamydomonas reinhardtii* (2.9 Å)	P700A P700B	PsaA-H676 PsaB-H655	Y600, N601, F608, F672, W679, Y731, T739, F742	L621, W626 F593, W620, F651, W658, Y718, T721, Y722, F725	Green Alga Trimer
6KMW	*Halomicronema hongdechloris C2206* (2.35 Å)	P700A P700B	PsaA-H689 PsaB-H662	Y612, N613, F620, F685, W692, F744, T752, F755	L628, W633 F600, W627, F658, W665, Y725, T728, Y729, F732	CyanobacteriumTrimer, Chl *f*, White light
6KMX	*Halomicronema hongdechloris C2206* (2.41 Å)	P700A A0B	PsaA-H709 PsaB-H664	Y632, N633, F640, F705, F712, Y764, T772, F775	F630, W635 F602, W629, F660, W667, Y726, T729, Y730, F733	CyanobacteriumTrimer, Chl *f*, Far-red light
6PNJ	*Fischerella thermalis PCC 7521* (3.19 Å)	P700A P700B	PsaA-H713 PsaB-H661	Y636, N637, F644, F709, F716, Y768, T776, F779	F627, W632 F599, W626, F657, W664, Y723, T726, Y727, F730	CyanobacteriumTrimer, Far-red light
7COY	*Acaryochloris marina MBIC11017* (2.5 Å)	P700A P700B	PsaA-H678 PsaB-H657	Y601, N602, F609, F674, W681, Y733, S741, F744	L623, W628 F595, W622, F653, W660, Y720, T723, Y724, F727	CyanobacteriumTrimer, Far-red light
5ZJI	*Zea mays* (3.3 Å)	P700A P700B	PsaA-H675 PsaB-H654	Y598, N599, F606, F671, W678, Y730, T738, F741	F620, W625 F592, W619, F650, W657, Y717, T720, Y721, F724	Plants, Trimer

similarities of the binding sites of all redox cofactors are shown in Supplementary Fig. 18b. The $P700_B$ binding sites are more conserved (Fig. 7a) and have a higher structural similarity (Fig. 7b) than the $P700_A$ binding sites. The $P700_A$ group and the $P700_B$ group share a small portion of the keys, suggesting a great difference between

Table 2 The Chl Chains and Numbers and the Amino Acids that Provide Axial Ligands to P700, A_{-1}, A_0 and A_C

PDB	Organism and resolution	P700		A-1		A0		AC	
		P700A	P700B	A-1A	A-1B	A0A	A0B	ACA	ACB
1JB0	*Synechococcus elongatus* (2.5 Å)	A, 1011, CLA, PsaA-H680	B, 1021, CLA, PsaB-H660	B, 1012, CLA, PsaB-N591	A, 1022, CLA, PsaA-N604	A, 1013, CLA, PsaA-M688	B, 1023, CLA, PsaB-M668	A, 1140, CLA, PsaA-H734	B, 1239, CLA, PsaB-H718
5OY0	*Synechocystis sp. PCC 6803* (2.5 Å)	A, 1011, CLA, PsaA-H676	B, 1021, CLA, PsaB-H651	A, 1012, CLA, PsaB-N582	B, 1022, CLA, PsaA-N600	A, 1013, CLA, PsaA-M684	B, 1023, CLA, PsaB-M659	A, 1140, CLA, PsaA-H730	B, 1239, CLA, PsaB-H709
6HQB	*Synechocystis sp. PCC 6803* (4.0 Å)	A, 1011, CLA, PsaA-H676	B, 1021, CLA, PsaB-H651	A, 1012, CLA, PsaB-N582	B, 1022, CLA, PsaA-N600	A, 1013, CLA, PsaA-M684	B, 1023, CLA, PsaB-M659	A, 1140, CLA, PsaA-H730	B, 1239, CLA, PsaB-H709
6JO6	*Chlamydomonas reinhardtii* (2.9 Å)	A, 801, CL0 PsaA-H676	B, 802, CLA, PsaB-H655	A, 803, CLA, PsaB-N586	A, 854, CLA, PsaA-N601	A, 802, CLA, PsaA-M684	B, 803, CLA, PsaB-M663	A, 842, CLA, PsaA-H730	B, 840, CLA, PsaB-H713
6KMW	*Halomicronema hongdechloris C2206* (2.35 Å)	A, 801, CL0, PsaA-H689	B, 803, CLA, PsaB-H662	B, 804, CLA, PsaB-N593	B, 801, CLA, PsaA-N613	B, 802, CLA, PsaA-M697	B, 805, CLA, PsaB-M670	A, 841, CLA, PsaA-H743	B, 841, CLA, PsaB-H720
6KMX	*Halomicronema hongdechloris C2206* (2.41 Å)	A, 801, CL0, PsaA-H709	B, 801, CLA, PsaB-H664	B, 802, CLA, PsaB-N595	A, 802, CLA, PsaA-N633	A, 803, CLA, PsaA-M717	B, 803, CLA, PsaB-M672	A, 843, CLA, PsaA-H763	B, 840, CLA, PsaB-H721
6PNJ	*Fischerella thermalis PCC 7521* (3.19 Å)	A, 1011, CL0, PsaA-H713	B, 1021, CLA, PsaB-H661	A, 1012, CLA, PsaB-N592	B, 1022, CLA, PsaA-N637	A, 1013, CLA, PsaA-M721	B, 1023, CLA, PsaB-M669	A, 1140, CLA, PsaA-H767	B, 1239, CLA, PsaB-H718
7COY	*Acaryochloris marina MBIC11017* (2.5 Å)	A, 3101, G9R, PsaA-H678	B, 3003, CL7, PsaB-H657	B, 3002, CL7, PsaB-N588	A, 3103, CL7, PsaA-N602	A, 3102, PHO, PsaA-M686	B, 3004, PHO, PsaB-L665	A, 3143, CL7, PsaA-H732	B, 3026, CL7, PsaB-H715
5ZJI	*Zea mays* (3.3 Å)	A, 801, CL0, PsaA-H675	B, 802, CLA, PsaB-H654	A, 803, CLA, PsaB-N585	A, 854, CLA, PsaA-N599	A, 802, CLA, PsaA-M683	B, 803, CLA, PsaB-M662	A, 842, CLA, PsaA-H729	B, 840, CLA, PsaB-H712

$P700_A$ and $P700_B$ environments (Fig. 7c). The *common* and *specific* keys for the binding sites of $P700_A$ and $P700_B$ were identified (Fig. 7d). Three $P700_A$-*specific* keys were shown in Fig. 7e (details in Supplementary 19a) whereas two $P700_B$-*specific* keys were illustrated in Fig. 7f (details in Supplementary 19b). The results obtained from the A_{-1A} and A_{-1B} binding sites (Fig. 8a-d) are similar to those from the $P700_A$ and $P700_B$ binding sites (Fig. 7a-d). Ten A_{-1A}-*specific* keys were identified and those ten triangles are from three residues (L673, H676 and F677) of PsaA and four residues (F578, W579, N582 and W586) of PsaB (Fig. 8e and Supplementary Fig. 20). No A_{-1B}-*specific* keys were found (Fig. 8d), suggesting more structural diversity for the binding sites of A_{-1A} than those of A_{-1B}. No *common* keys were identified for the binding sites of P700 (both $P700_A$ and $P700_B$) and for those of A_{-1} (both A_{-1A} and A_{-1B}). The results from the A_0 binding sites (Fig. 9a-d) are similar to the binding

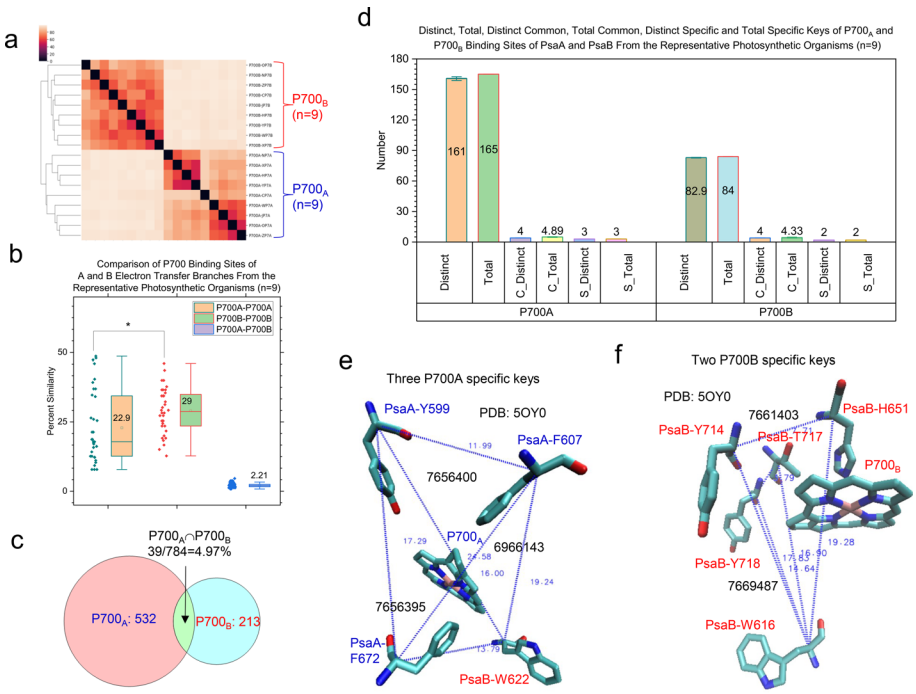

Fig. 7 Hierarchical cluster analysis of $P700_A$ and $P700_B$ molecules demonstrating the capacity of the TSR algorithm for distinguishing their 3D structural differences. Panel **a**, the hierarchical cluster analysis shows 3D structural relationships of $P700_A$ and $P700_B$. The numbers of $P700_A$ and $P700_B$ are labeled; panel **b**, structural similarities between $P700_A$—$P700_A$ pairs, $P700_B$—$P700_B$ pairs and $P700_A$—$P700_B$ pairs. The average similarity values, SDs, and 25/75 percentiles are indicated; panel **c**, the Venn diagram showing the numbers of TSR keys exclusively belonging to the $P700_A$ group, the $P700_B$ group and the intersection between the $P700_A$ group and the $P700_B$ group; panel **d**, the distinct, total, distinct *common* (C_Distinct), total *common* (C_Total), distinct *specific* (S_Distinct) and total *specific* (S_Total) TSR keys for $P700_A$ and $P700_B$. The average values and SDs are indicated; panel **e**, the three $P700_A$ *specific* TSR keys (6966143, 7656395, 7656400). The triangle corresponding to 6966143 is constructed from the three C_α atoms from PsaA-F607, PsaB-W622 and PsaA-F672. The triangle corresponding to the key 7656395 is constructed from the three C_α atoms from PsaA-Y599, PsaA-F607 and PsaB-W622. The triangle corresponding to the key 7656400 is constructed from the three C_α atoms from PsaA-Y599, PsaB-W622 and PsaA-F672; panel **f**, the two $P700_B$ *specific* TSR keys (7661403, 7669487) are shown. The triangle corresponding to the key 7661403 is constructed from the three C_α atoms from PsaB-W616, PsaB-H651 and PsaB-Y714. The triangle corresponding to the key 7669487 is constructed from the three C_α atoms from PsaB-W616, PsaB-T717 and PsaB-Y718. **e**–**f**, The PDB is 5OY0

sites of P700 and A_{-1}. One *common* key was identified for the binding sites of both A_{0A} (Fig. 9e) and A_{0B} (Fig. 9f). The A_{0A}-*specific* and A_{0B}-*specific* keys are shown in Supplementary Figs. 21 and 22, respectively. The binding sites of A_{1A} are more conserved (Fig. 10a) and have a higher structural similarity (Fig. 10b) than those of A_{1B}. The structural relationships of the binding site of the A_{1A} and A_{1B} groups combined as well as the individual A_{1A} and A_{1B} binding site groups are shown in Fig. 10c-d, respectively. One *specific* key (Fig. 10e) and three *specific* keys (Fig. 10f) (details in Supplementary Fig. 23) were identified for the binding sites of A_{1B} and A_{1A}, respectively. Therefore, the hierarchical clustering results demonstrate the difference of the cofactor binding sites between the A and B branches.

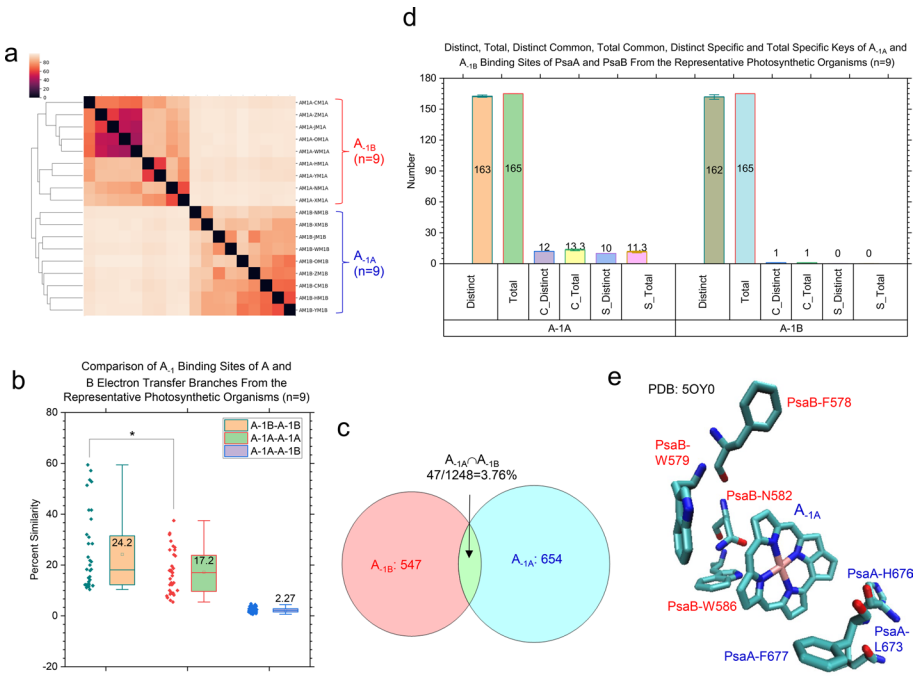

Fig. 8 Hierarchical cluster analysis of different types of A_{-1A} and A_{-1B} demonstrating the capacity of the TSR algorithm, which can distinguish their 3D structural differences. Panel **a**, the hierarchical cluster analysis of 3D structural relationships of A_{-1A} and A_{-1B}. The numbers of A_{-1A} and A_{-1B} are labeled; panel **b**, the structural similarity between A_{-1A}—A_{-1A} pairs, A_{-1B}—A_{-1B} pairs and A_{-1A}—A_{-1B} pairs. The average similarity values, SDs, and 25/75 percentiles are indicated; panel **c**, the Venn diagram showing the numbers of TSR keys exclusively belonging to the A_{-1A} group, the A_{-1B} group and the intersection between the A_{-1A} group and the A_{-1B} group; panel **d**, the distinct, total, distinct *common* (C_Distinct), total *common* (C_Total), distinct *specific* (S_Distinct) and total *specific* (S_Total) TSR keys for A_{-1A} and A_{-1B}. The average values and SDs are indicated; panel **e**, the ten A_{-1A} specific TSR keys. The amino acids associated with these ten keys are PsaA-L673, PsaA-H676, PsaA-F677, PsaB-F578, PsaB-W579, PsaB-N582 and PsaB-W586. The PDB is 5OY0

Evaluation of the TSR-based method for quantifying molecular 3D structures

A common approach to understand the functions of a protein is to compare it to other proteins [77]. The existing 3D structure comparison methods can be roughly divided into five categories [7]: sequence-, distance-, secondary structure-, geometry-, and network-based methods. The TSR algorithm is categorized as a geometry-based method. Therefore, we evaluate the TSR-based method in comparison with the RMSD [70] and USR [69] methods. RMSD is a popular measure of structural similarity between protein or drug 3D structures and involves alignment and optimal superposition between matched pairs of atoms [78]. It searches for the lowest RMSD result for both structures. Alignment or superposition is a complex problem because it is challenging to simultaneously optimize the number of equivalent residues and the global differences due to the fact that one may have to be optimized at the expense of the other [79]. An additional challenge can arise when two different global structures are similar in small local regions (e.g., Triad between chymotrypsin and subtilisin) that can be overlooked. USR is a shape similarity technique that is characterized as a non-superposition-based method [80, 81]. To provide a spectrum of comparisons, we decided to compare the methods for proteins at global and local structural levels as well as for electron redox cofactors.

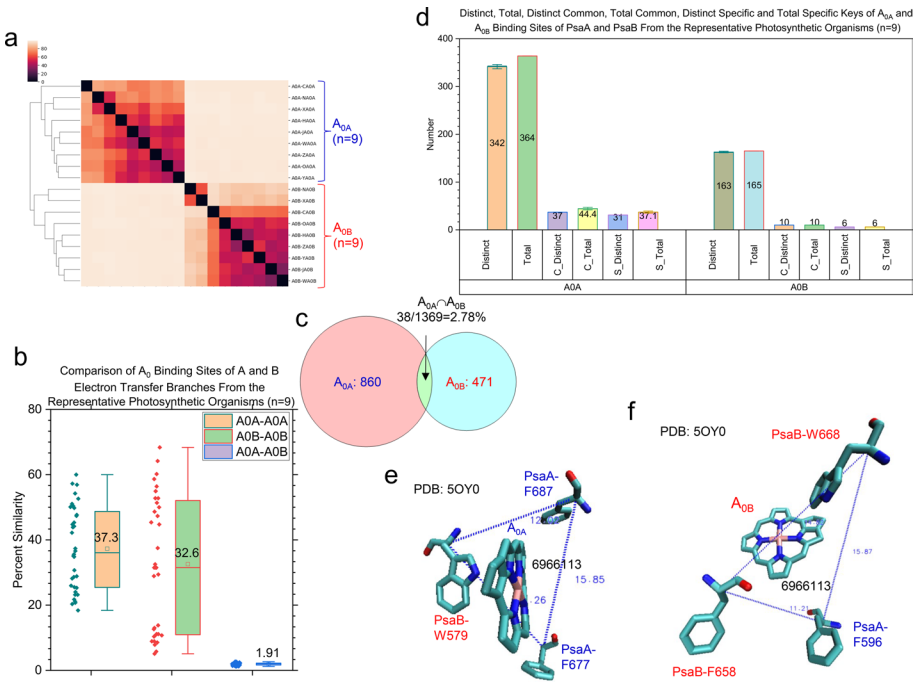

Fig. 9 Hierarchical cluster analysis of different types of A_{0A} and A_{0B} demonstrating the capacity of the TSR algorithm for distinguishing their 3D structural differences. Panel **a**, the hierarchical cluster analysis showing 3D structural relationships of A_{0A} and A_{0A}. The numbers of A_{0A} and A_{0B} are labeled; panel **b**, the structural similarity between A_{0A}—A_{0A} pairs, A_{0B}—A_{0B} pairs and A_{0A}—A_{0B} pairs. The average similarity values, SDs, and 25/75 percentiles are indicated; panel **c**, the Venn diagram showing the numbers of TSR keys exclusively belonging to the A_{0A} group, the A_{0B} group and the intersection between the A_{0A} group and the A_{0B} group; panel **d**, the distinct, total, distinct *common* (C_Distinct), total *common* (C_Total), distinct *specific* (S_Distinct) and total *specific* (S_Total) TSR keys for A_{0A} and A_{0B}. The average values and SDs are indicated; panels **e** and **f**, the A_0 *common* TSR key (6,966,113). The triangle corresponding to the key 6,966,113 for A_{0A} (**e**) is constructed from the three C_α atoms from PsaB-W579, PsaA-F677 and PsaA-F687. The triangle corresponding to the key 6,966,113 for A_{0B} **(f)** is constructed from the three C_α atoms from PsaA-F596, PsaB-F658 and PsaB-W668. The PDB is 5OY0

Comparison of the TSR-based method with the RMSD and the USR methods for global structures

PsaA and PsaB show a strong sequence homology [82] and have been suggested to evolve via gene duplication [83]. PsaA and PsaB are well preserved in the membrane integral parts while large differences between the two subunits are visible in the loop regions [84]. The TM-align software [71] was used to generate pairwise RMSD scores that were further used as an input for hierarchical clustering. The result clearly shows two clusters. One cluster contains nine PsaA structures and the other cluster contains nine PsaB structures (Supplementary Fig. 24). The result agrees with their functional classification as well as the protein sequence-based phylogenetic study (Supplementary Fig. 25). Adjusted Rand index (ARI) is frequently used in cluster validation, which measures agreement between two partitions: one partition is given by the clustering process and the other is defined by an external criterion. We used the functional classification as the external criterion in this study. The ARI values lie between 0 and 1 and should be interpreted as follows: ARI $\geq$ 0.90 excellent recovery; 0.80 $\leq$ ARI < 0.90 good recovery; 0.65 $\leq$ ARI < 0.80 moderate recovery; ARI < 0.65 poor recovery. As expected, the ARI for the clustering analysis of PsaA and PsaB using the RMSD method is 1.0. The clustering

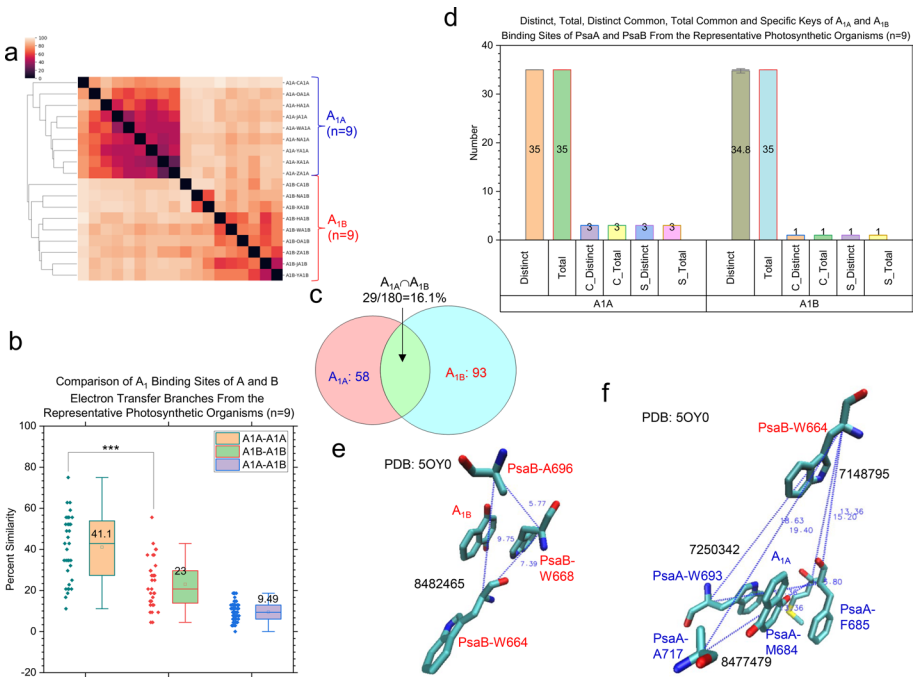

Fig. 10 Hierarchical cluster analysis of different types of A_{1A} and A_{1B} demonstrating the capacity of the TSR algorithm for distinguishing their 3D structural differences. Panel **a**, the hierarchical cluster analysis showing 3D structural relationships of A_{1A} and A_{1A}. The numbers of A_{1A} and A_{1B} are labeled; panel **b**, the structural similarity between A_{1A}—A_{1A} pairs, A_{1B}—A_{1B} pairs and A_{1A}—A_{1B} pairs. The average similarity values, SDs, and 25/75 percentiles are indicated; panel **c**, the Venn diagram showing the numbers of TSR keys exclusively belonging to the A_{1A} group, the A_{1B} group and the intersection between the A_{1A} group and the A_{1B} group; panel **d**, the distinct, total, distinct *common* (C_Distinct), total *common* (C_Total), distinct *specific* (S_Distinct) and total *specific* (S_Total) TSR keys for A_{1A} and A_{1B}. The average values and SDs are indicated; panel **e**, the A_{1B} *specific* TSR key (8482465) is shown. The triangle corresponding to the key 8482465 is constructed from the three C_α atoms from PsaB-W664, PsaB-W668 and PsaB-F696; panel **f**, the three A_{1A} *specific* TSR keys (7148795, 7250342, 8477479). The triangle corresponding to the key 7148795 is constructed from the three C_α atoms from PsaB-W664, PsaA-M684 and PsaA-F685. The triangle corresponding to the key 7250342 is constructed from the three C_α atoms from PsaB-W664, PsaA-F685 and PsaA-W693. The triangle corresponding to the key 8477479 is constructed from the three C_α atoms from PsaB-W664, PsaA-M684 and PsaA-A717; panels **e**–**f**, The PDB is 5OY0

analysis of the same structures using the USR method reveals that the PsaA structures cannot completely separated from the PsaB structures (Supplementary Fig. 26). The ARI value obtained from the USR method is 0.

The same PsaA and PsaB structures as those used in the RMSD and USR studies were used for the TSR-based analysis. The hierarchical clustering shows that eight PsaA structures are clustered together and eight PsaB structures are clustered together. However, the PsaA and PsaB structures from *Acaryochloris marina* are grouped into one cluster (Fig. 11a). The result does not perfectly match with their taxa classification (cyanobacterial PS I vs. green algal PS I vs. plant PS I) probably because the structures of PsaA (685 aa) and PsaB (658 aa) from *Acaryochloris marina* are smaller than the rest of PsaA (717–750 aa) and PsaB (727–740 aa) structures. It was reported that applying the amino acid-grouping algorithm improves the clustering result when two amino acids with similar structures and chemical properties are grouped together [85]. Applying the size-gap algorithm also improves clustering results when a small structure is compared with a

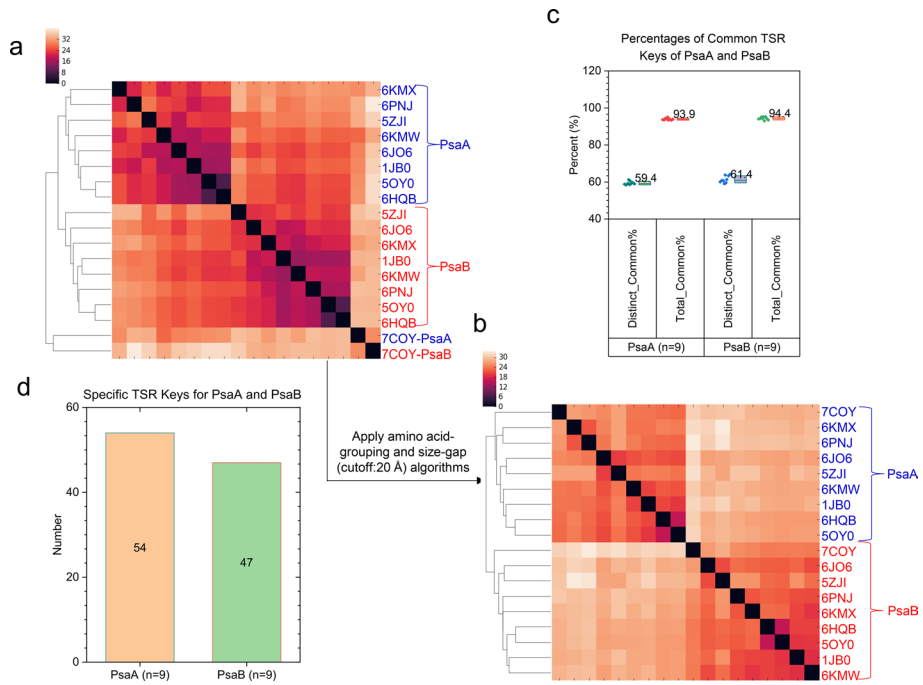

Fig. 11 Hierarchical cluster analysis of the representative PsaA and PsaB structures demonstrated the importance of applying the amino acid grouping algorithm and the size-gap algorithm in the TSR-based method. Panel **a**, the hierarchical clustering without applying amino acid-grouping algorithm and the size-gap algorithm; panel **b**, the hierarchical clustering with applying amino acid-grouping algorithm and the size-gap algorithm. The cutoff value for the size-gap algorithm is 20 Å; panels **a**–**b**, the PDB IDs, PsaA and PsaB are labeled. Blue represents PsaA structures and red represents PsaB structures; panel **c**, the distinct *common*, total *common*, distinct and total TSR keys for each structure of PsaA and PsaB were calculated and the percentages of distinct *common* and total *common* TSR keys are present. Percentage of distinct *common* TSR keys = No. of distinct *common* TSR keys/No. of distinct TSR keys * 100%. Percentage of total *common* TSR keys = No. of total *common* TSR keys/No. total TSR keys * 100%. The average values are labeled and the SDs are shown; panel **d**, the specific keys exclusively belonging to PsaA or PsaB were calculated and are shown; panels **c**–**d**, number of the structures for PsaA and PsaB are labeled

large structure [86]. To improve the clustering for the PsaA and PsaB structures, we have applied both the amino acid-grouping and the size-gap algorithms together, we observed an improvement of the cluster analysis (Fig. 11b) and the ARI value achieves 1.0. One of the uniqueness of the TSR algorithm lies in its ability to interpret clustering results using *common* and *specific* TSR keys and to offer valuable insights into the underlying hierarchical relationships of molecular structures within the dataset. It was reported that the common substructure motifs among different protein folds are of critical importance for biological function predictions [87]. Specific substructures exclusively belonging to a particular protein family can be considered as structural characteristics and could be structural foundation for drug development. *Common* (Fig. 11c) and *specific* (Fig. 11d) TSR keys were identified for PsaA and PsaB for a deeper understanding of their relationships.

Comparison of the TSR-based method with the RMSD and USR methods for amino acid structures

For the reaction center of PS I, the Mg^{2+} ion, a relatively hard acid, of each monomer of P700 is axially coordinated by a nitrogen atom of a histidine residue whereas A_{-1} and A_0

are coordinated to a water ligand and soft base sulfur ligand from a methionine residue, respectively [74]. In the X-ray crystal structure of PS I from *Synechocystis*, a water molecule serving as an axial ligand for A_{-1A} is bonded by two hydrogen bonds with PsaB-N582 and with PsaB-W586, which tightly arrange the A_{-1A} binding pocket. In contrast, the PsaB-W586 corresponding residue in PsaA, depending on the species, is Val or Ile and a water molecule that serves as an axial ligand for the A_{-1B} bonded only by one hydrogen bond with PsaA-N600 [17]. Upon the inspections of the structures of PS I [17], we found that PsaA-W597 is close to A_{0B} whereas PsaB-W579 is close to A_{0A}. To understand whether the amino acids that directly or indirectly participate in coordination bonds with the redox cofactors have their unique structural characteristics, we focused on His, Asn, Trp and Met. The structural similarities of those amino acids in PsaA and PsaB from different species are shown in Supplementary Figs. 27 (His), 28 (Asn), 29 (Trp) and 30 (Met). A trial analysis of one structure (PDB: 5OY0) indicates that the histidine residues (PsaA-H676 and PsaB-H651) coordinated to $P700_A$ or $P700_B$ are separated and are in different clusters (Supplementary Fig. 31). The asparagine residues (PsaB-N582 and PsaA-N600) hydrogen bonded to the water with the coordination to A_{-1A} or A_{-1B} are not next to each other (Supplementary Fig. 32). The methionine residues (PsaA-M684 and PsaB-M659) coordinated to A_{0A} or A_{0B} are next to each other and are in the same cluster (Supplementary Fig. 33), whereas tryptophan residues (PsaB-W579 and PsaA-W597) close to A_{0A} or A_{0B} are also next to each other and are in the same cluster (Supplementary Fig. 34). The trial analysis suggests that the amino acids that have close contact with the redox cofactors may (Met and Trp) or may not (His and Asn) have their unique structural characteristics. Therefore, histidine and asparagine residues were excluded from the follow-up analyses.

To verify the structural characteristics of Met and Trp, eight more PDB structures were included in the subsequent analyses. The hierarchical clustering result shows that fourteen out of seventeen methionine residues that coordinated with A_0 are grouped together. However, the rest three methionine residues are separated from the fourteen methionine residues. Therefore, we focus on the rest of the discussions only on tryptophan residues. Eighteen tryptophan residues are close to either A_{0A} or A_{0B}. Seventeen of all these 18 tryptophan residues are next to each other in the hierarchical cluster analysis (Fig. 12a). The examples of the close interactions between Trp and A_0 are illustrated in Fig. 12b (close interaction between PsaB-W579 and A_{0A}) and **12c** (close interaction between PsaA-W597 and A_{0B}). These 18 tryptophan residues have different MaxDist (Fig. 12d) and Theta (Fig. 12e) values compared with other tryptophan residues in PsaA and PsaB. Interestingly, we also found that the particular tryptophan residues from nine PDB structures group together. PsaB-Trp664 (PDB: 5OY0) separates two water clusters between A_{1A} and A_{1B} (Fig. 12f). These 9 tryptophan residues have their structural characteristics because they have unique MaxDist (Fig. 12d) and Theta (Fig. 12e) values. The Trp664 residues from different species have more *common* TSR keys than the tryptophan residues of 579 and 597 (Fig. 12g), suggesting that the Trp664 residues are structurally more conserved.

The corresponding residue of PsaB-Trp664 in PsaA is Gly689. PsaB-W664 has been suggested to play a role in the electron transfer acting as an electron acceptor between A_{1B} and F_X [88]. The functional studies demonstrated that (i) PS I with doubly protonated

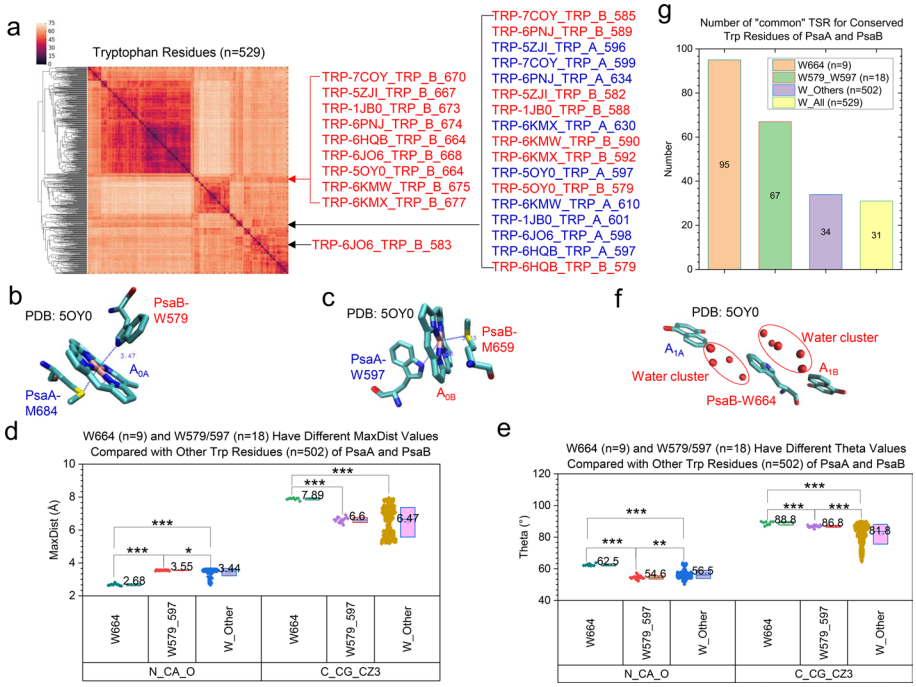

Fig. 12 The tryptophan residues closely interacting with A_0 and A_1 and their structural characteristics. Panel **a**, the hierarchical cluster analysis shows 3D structural relationships of the tryptophan residues of PsaA and PsaB from diverse organisms. The PDB IDs, the number of tryptophan and nine adjacent tryptophan residues as well as seventeen adjacent tryptophan residues are labeled. One tryptophan residue that is close to A_0 and separated from the remaining seventeen tryptophan is labeled too; panel **b**, the residues closely interacting with A_{0A}; panel **c**, the residues closely interacting with A_{0B}; panel **d**, the MaxDist values for the tryptophan residues that are close to A_0 and A_1 and for the rest of tryptophan residues were calculated; panel **e**, the Theta values for the tryptophan residues that are close to A_0 and A_1 and for the rest of tryptophan residues were calculated; panels **d**–**e**, * means a *p* value is less than 0.05 using a *t*-test, ** means a *p* value is less than 0.01 and *** means a *p* value is less than 0.001 using a *t*-test; panel **f**, the residues close to A_{1A}, A_{1B} and the water cluster; panels **b**, **c**, **f**, the PDB is 5OY0; panel **g**, numbers of common TSR keys for the tryptophan residues that are close to A_1, A_0, and the rest of tryptophan residues and all tryptophan residues were calculated. Average numbers are labeled

quinone in the A_1 binding site of the mutant with PsaB-W664F are not functional in electron transfer. However, the electron transfer functionality can be restored by incubating the light-treated mutant PS I sample in the presence of added phylloquinone [89]. (ii) PsaB-Trp664 is essential for the high-efficiency electron transfer between the phylloquinones and the iron-sulfur clusters [90]. Considering the structural analysis in this study and the published functional studies, we can link the structural characteristics of Try664 to their functions. As the result, we conclude that tryptophan residues, closely contacting with A_0, have their specific structural characteristics and the tryptophan residues, separating two water clusters between A_{1A} and A_{1B}, are structurally conserved.

To evaluate the performance of the TSR-based method, we used the RMSD and USR methods to study the same tryptophan residues that were analyzed by the TSR algorithm. The cluster analysis using the RMSD method shows that eight of 9 Trp664 are grouped together and thirteen of 18 Trp579—Trp597 are grouped next together (Supplementary Fig. 35). Based on the fact that all 9 Trp664 are grouped together and seventeen out of 18 Trp579—Trp597 are grouped when the TSR algorithm is used (Fig. 12a), we conclude that the TSR-base method is better for the cluster analysis of the tryptophan

residues than the RMSD method. The performance of the USR method on the tryptophan cluster analysis is worse (Supplementary Fig. 36) than the performances of the RMSD (Supplementary Fig. 35) and the TSR (Fig. 12a) algorithms.

Comparison of the TSR-based method with the RMSD and USR methods for structures of redox cofactors

Phytol tail of chlorophyll anchors the pigment to membranes of thylakoids and maintains the orientation of the pigment. The chlorin ring of redox cofactors, not the phytol tail, is directly involved in electron and energy transfer. The clustering result shows that the redox cofactors are grouped into the P700 cluster, the A_{-1} cluster and the A_0 cluster (Fig. 3a). In this study up to now, all the atoms of the cofactors have been used in generating the TSR keys. Chl *a*'s tail has 20 carbons. To further study the structural characteristics of the redox cofactors, we have developed a feature selection method in the TSR key generation step to filter out the carbons 6 to 20 of the phytol tails (Fig. 13a). The new hierarchical clustering result shows three clusters (ARI: 1.0): P700, A_{-1} and A_0 (Fig. 13b). The P700 cluster and A_0 cluster are joined into the P700—A_0 cluster that is then merged with the A_{-1} cluster (Fig. 13b). It reveals that P700 and A_0 from different species are structurally similar and A_{-1} are structurally different from either P700 or A_0.

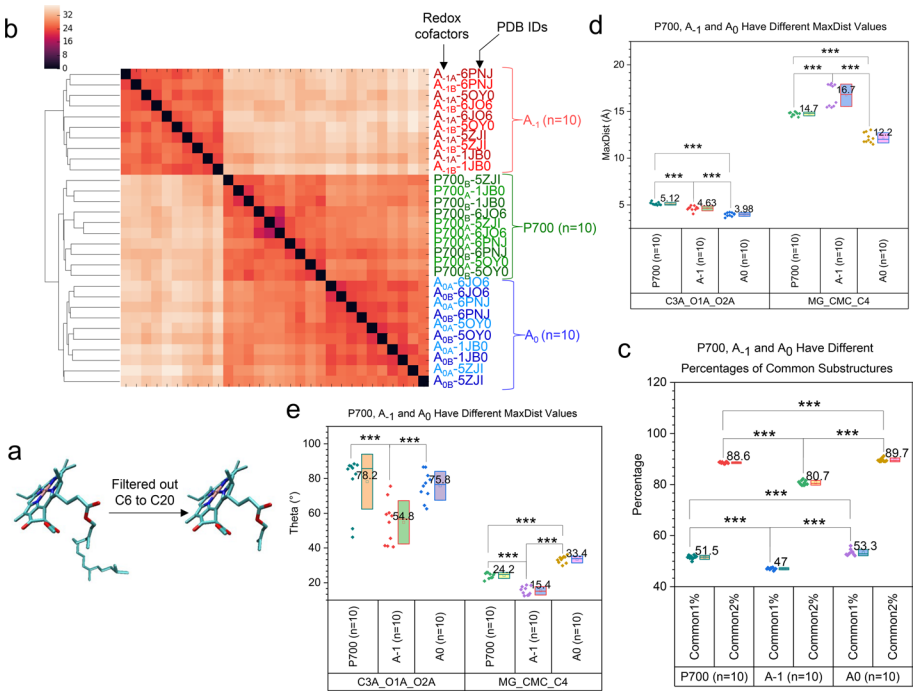

Fig. 13 Hierarchical cluster analysis of P700, A_{-1} and A_0 from different species. Panel **a**, the tails of chlorophyll molecules from carbon 6 to carbon 20 are not included in the study; panel **b**, the hierarchical cluster analysis of the redox cofactors: P700, A_{-1} and A_0. Redox cofactors, PDB IDs and numbers of structures are labeled; panel **c**, percentages of distinct *common* (Common1%) and total *common* (Common2%) TSR keys were calculated and are present. The definitions for percentage of distinct and total *common* TSR keys were defined in Fig. 13 panel **c**; panels **d**–**e**, the MaxDist (**d**) and Theta (**e**) values for two triangles (C3A-O1A-O2A and MG-CMC-C4) of P700, A_{-1} and A_0 were calculated and are present; panels **c**–**e**, *** means a *p* value is less than 0.001 using a *t*-test

It is unclear whether the structural similarity is related to their functions. $P700_A$ and $P700_B$ do not form their own clusters (Fig. 13b), suggesting $P700_A$ are not structurally different from $P700_B$ among different species. The same situation is observed for A_{-1A} and A_{-1B} as well as A_{0A} and A_{0B} (Fig. 13b). Interestingly, we have observed that A_{0A} and A_{0B} are grouped into a single cluster for each species (each PDB entry) (Fig. 13b). The hierarchical structural relationships of the redox cofactors can be described as (i) two A_0 clusters of thermophilic cyanobacterium *T. vestitus* and cyanobacterium *Synechocystis* are joined into a large cluster, and two A_0 clusters of filamentous true-branching cyanobacterium *Fischerella thermalis* and a single-cell green alga *Chlamydomonas reinhardtii* are joined into another large cluster; (ii) the two large clusters from (i) are merged into a large cluster; (iii) The larger cluster is merged with the A_0 cluster of the plant (Fig. 13b). The result indicates A_0 structures are species-specific. It is not the case for P700 and A_{-1} (Fig. 13b). To interpret the clustering results, we have calculated *common* and *specific* TSR keys for P700, A_{-1} and A_0. The percentages of *common* TSR keys are different for P700, A_{-1} and A_0 (Fig. 13c), suggesting different structural similarities for the redox cofactors. In addition, two triangles (C3A-O1A-O2A and MG-CMC-C4) have different MaxDist (Fig. 13d) and Theta (Fig. 13e) values for P700, A_{-1} and A_0, demonstrating the geometrical differences of different cofactors. Taken together, the *common* and *specific* TSR keys explain the clustering result (P700 cluster, A_{-1} cluster and A_0 cluster).

The cluster analysis of the redox cofactors using the RMSD method clearly shows three clusters: P700, A_{-1} and A_0 clusters (Supplementary Fig. 37) that agree with their functional classification (ARI: 1.00) and match with the three clusters from the TSR algorithm (ARI: 1.00). In contrast, the USR method can distinguish A_{-1} structures from those of P700 and A_0 (Supplementary Fig. 38). However, it cannot distinguish P700 structures from A_0 structures (Supplementary Fig. 38) (ARI: 0.544) (Supplementary Table 4). If we look closer at the subclusters from the RMSD method, there are two subclusters for the A_0 clusters: one is for A_{0A} (A branch) and the other is for A_{0B} (B branch) (Supplementary Fig. 37). The A_0 structures are branch-specific (A branch and B branch) for the RMSD method (Supplementary Fig. 37), not species-specific observed from the TSR algorithm (Fig. 13b). The P700 structures are also branch-specific for the RMSD method (Supplementary Fig. 37). However, it is not the case for the A_{-1} structures (Supplementary Fig. 37).

Discussions

Significant numbers of aromatic residues are found in the binding sites of P700, A_{-1}, A_0 and A_1. The exact functions of aromatic residues in the binding sites of the cofactors are unclear. We discovered that the tryptophan residues of PsaA and PsaB close to A_{0B} and A_{0A}, respectively, are structural conserved among different photosynthetic organisms. Trp has been reported to be involved in light-triggered electron transfer [91]. The studies highlight the generality of Trp-porphyrin electron transfer events in heme proteins [92]. The exact function of the tryptophan residues (PsaA-Trp597 and PsaB-Trp579) that are close to A_0 has not been studied by an experimental approach.

The early version of the TSR-based method has its limitation that can only quantify backbone structures of proteins. This new version allows studying structural complementarity between electron cofactors and their surrounding amino acids. This version of

the TSR algorithm was evaluated by comparing it with the RMSD and USR methods. For the quantitative comparisons of the global and local protein structures and the redox factor structures, the TSR-based and RMSD methods outperformed the USR method. For protein global structural and redox cofactor structural comparisons, the results of the hierarchical clustering using the TSR-based method or the RMSD method match with their functional classification. For tryptophan structural comparisons, the TSR-based method outperforms the RMSD method. In addition, the TSR-based method can interpret clustering results using *common* and *specific* TSR keys. In contrast, the RMSD and USR methods have their limitations in interpreting clustering results. Besides the advantages of interpreting results, the TSR-based method has two additional advantages. First, the RMSD method requires pre-alignment or pre-determination of equivalent residues for proteins or equivalent atoms for redox factors. Therefore, the RMSD method has its limitation in comparing two nonhomologous proteins (PsaA vs. PsaL) and two different types of redox cofactors (e.g., chlorophyll vs. phylloquinone). In contrast, the TSR-based method is an alignment-free algorithm. It can be used to quantify two completely different structures. Second, the unique representation of molecular 3D structures by TSR keys (integers) makes substructure search easy and effective. It would be useful if a computational method is able to search for functional substructures similar to catalytic sites, ligand binding sites and other interfacing residues [93]. Such an endeavor requires the availability of a method encoding molecular structures that are indicative of biological activity. Structural complementarity in molecular recognition events is an important indicator of a molecule's activity because favorable molecular interactions require such complementarity. The TSR algorithm has its uniqueness for quantifying structural complementarity (e.g., cofactor and cofactor binding sites).

Conclusions and future directions

Conclusions

A comprehensive study of PS I 3D structures brought the following main findings.

(i) A new version of the TSR-based method was developed to represent 3D structures of pigments and to quantify pigment structures.

(ii) The hierarchical clustering results using *Cofactor* TSR keys reveal that the redox cofactors, P700, A_{-1} and A_0 form their distinct clusters, suggesting their specific structural characteristics. For example, the two triangles (C3A-O1A-O2A and MG-CMC-C4) have different geometries for P700, A_{-1} and A_0.

(iii) The results using *Cofactor* and *CA* TSR keys demonstrate the structural differences of the redox cofactors, P700, A_{-1}, A_0 and A_1, as well as their binding sites between A branch and B branch.

(iv) Different types of TSR keys were used to show common substructures shared by different types of redox cofactors or their binding sites as well as unique substructures exclusively belonging to a certain type of cofactors or their binding sites.

(v) The hierarchical clustering results show that the tryptophan residues close to A_0 from different species were clustered together as well as the tryptophan residues splitting the water cluster near A_{1A} and A_{1B} binding sites were grouped together. The results demonstrate that the tryptophan residues close to A_0 are structurally conserved. The tryptophan residues splitting the water cluster are also structur-

ally conserved (e.g., N-CA-O triangle and C-CG-CZ3 triangle have different geometries between Trp664 and Trp579-Trp.). These structurally conserved residues imply their specific functional roles.

(vi) In term of hierarchical clustering results, the TSR-based method outperforms the RMSD and USR methods. In term of computational cost, the USR method runs faster than the RMSD and TSR methods for global protein structural comparisons.

In summary, this study of structural relationships of pigments and protein local environments provides new evidence for their unique chemical and physical properties of each redox cofactor that modulate the rate and direction of energy and electron transfer. This study builds a solid foundation for future functional studies of PS I complex using experimental approach as well as theoretical analyses, e.g., molecular dynamics simulations or QM/MM calculations. Understanding of the mechanisms underlying energy and electron transfer is essential for developing novel approaches for addressing two challenges being faced by the world: a need for energy sources, and a reduction of greenhouse gas emissions.

Future directions

The mechanism underlying the interactions between cofactors and protein environments is not fully understood. Thus, how to replicate the same mechanisms in artificial systems is still open to investigation [76]. > 50 PS I structures were included in this study. More PS I structures can be included in the future studies. The reaction centers of PS II have the arrangements similar to those of PS I. PS I and PS II structures can be studied together. In this study, we manually labeled the numbers of each redox factors and A_{CA} and A_{CB}, and numbers of the residues that coordinate with Mg^{2+} ions of the cofactors or the water molecules that have the coordination bond with the cofactors. PsaA and PsaB amino acid sequences and residue number assignment of each Chl molecules may not be the same across different species. Therefore, the manual labeling process for each cofactor and their corresponding residues is time-consuming. An algorithm needs to be developed for labeling each cofactor and their coordinating residues. We have developed a method for representing 3D structures of all twenty amino acids and quantifying their structures. Studies showed that His [94–99], Asn [100–103], Trp [88–90, 104, 105] and Met [106–113] play critical roles in modulating properties of the redox cofactors in PS I. Therefore, we discussed four amino acids (His, Asn, Trp and Met) with a focus on Trp in this study. As stated earlier, more structures and all other amino acids (Supplementary Fig. 39) can be included in the future studies.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-025-06038-y.

Additional file1.

Acknowledgements

This study is supported by NIH NIGMS (1R15GM144944-01) and the National Science Foundation 2313482. Most of this research was conducted with high-performance computational resources provided by the Louisiana Optical Network Infrastructure (http://www.loni.org). WX acknowledges Drs. Lyudmila V. Slipchenko and Sergei Savikhin for fruitful discussion on this project. Emily Sonnier, Kelli M, Frith and Davin J. Gardner helped to prepare the datasets. Here we want

to appreciate the LONI support team, especially Feng Chen, Jianxiong Li, and Oleg Starovoytov. The authors explicitly confirms that the work is original and has not been published elsewhere.

Author contributions
WX proposed and designed this study. LL, TIM, EKT, WJG and WX carried out the study, collected data and prepared the figures. TIM wrote the python code. AYC, JY and JK supervised students and contributed the discussion. All authors wrote the manuscript, read and revised the manuscript.

Funding
This study is supported by NIH NIGMS (1R15GM144944-01) and the National Science Foundation 2313482.

Availability of data and materials
These data were derived from the following resource available in the public domain: https://www.rcsb.org/. The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials. The source Python codes for this study are available for academic users on GitHub: https://github.com/WuXu26/PSI_TSR. Other TSR-related Python codes can be found from https://github.com/tarikulislammilon/TSR and https://github.com/WuXu26/Protein-3D-TSR.

Declarations

Ethics approval and consent to participate
This study does not involve human subjects and/or animals.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 2 September 2024 Accepted: 6 January 2025
Published online: 14 January 2025

References

1. De Marais DJ. Evolution. When did photosynthesis emerge on Earth? Science. 2000;289:1703–5.
2. Rye R, Holland HD. Paleosols and the evolution of atmospheric oxygen: a critical review. Am J Sci. 1998;298:621–72.
3. Allen JF. A redox switch hypothesis for the origin of two light reactions in photosynthesis. FEBS Lett. 2005;579:963–8.
4. Dismukes GC, Klimov VV, Baranov SV, Kozlov YN, DasGupta J, Tyryshkin A. The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis. Proc Natl Acad Sci USA. 2001;98:2170–5.
5. Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, Krauss N. Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution. Nature. 2001;411:909–17.
6. Olson TL, Williams JC, Allen JP. Influence of protein interactions on oxidation/reduction midpoint potentials of cofactors in natural and de novo metalloproteins. Biochim Biophys Acta (BBA) Bioenergetics. 2013;1827:914–22.
7. Kondra S, Sarkar T, Raghavan V, Xu W. Development of a TSR-based method for protein 3-D structural comparison with its applications to protein classification and motif discovery. Front Chem. 2021;8:602291.
8. Milon TI, Wang Y, Fontenot RL, Khajouie P, Villinger F, Raghavan V, Xu W. Development of a novel representation of drug 3D structures and enhancement of the TSR-based method for probing drug and target interactions. Comput Biol Chem. 2024;112:108117.
9. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucl Acids Res. 2000;28:235–42.
10. Ben-Shem A, Frolow F, Nelson N. Crystal structure of plant photosystem I. Nature. 2003;426:630–5.
11 Amunts A, Drory O, Nelson N. The structure of a plant photosystem I supercomplex at 3.4 A resolution. Nature. 2007;447:58–63.
12. Amunts A, Toporik H, Borovikova A, Nelson N. Structure determination and improved model of plant photosystem I*. J Biol Chem. 2010;285:3478–86.
13. Qin X, Suga M, Kuang T, Shen JR. Photosynthesis. Structural basis for energy transfer pathways in the plant PSI-LHCI supercomplex. Science. 2015;348:989–95.
14 Mazor Y, Borovikova A, Nelson N. The structure of plant photosystem I super-complex at 2.8 A resolution. Elife. 2015;4:e07433.
15 Mazor Y, Nataf D, Toporik H, Nelson N. Crystal structures of virus-like photosystem I complexes from the mesophilic cyanobacterium Synechocystis PCC 6803. Elife. 2014;3:e01496.
16 Mazor Y, Borovikova A, Caspy I, Nelson N. Structure of the plant photosystem I supercomplex at 2.6 Å resolution. Nat Plants. 2017;3:17014.
17 Malavath T, Caspy I, Netzer-El SY, Klaiman D, Nelson N. Structure and function of wild-type and subunit-depleted photosystem I in Synechocystis. Biochim Biophys Acta (BBA) Bioenergetics. 1859;2018:645–54.
18. Pan X, Ma J, Su X, Cao P, Chang W, Liu Z, Zhang X, Li M. Structure of the maize photosystem I supercomplex with light-harvesting complexes I and II. Science. 2018;360:1109–13.

19. Zheng L, Li Y, Li X, Zhong Q, Li N, Zhang K, Zhang Y, Chu H, Ma C, Li G, Zhao J, Gao N. Structural and functional insights into the tetrameric photosystem I from heterocyst-forming cyanobacteria. Nat Plants. 2019;5:1087–97.
20. Toporik H, Li J, Williams D, Chiu P-L, Mazor Y. The structure of the stress-induced photosystem I-IsiA antenna supercomplex. Nat Struct Mol Biol. 2019;26:443–9.
21. Gisriel C, Coe J, Letrun R, Yefanov OM, Luna-Chavez C, Stander NE, Lisova S, Mariani V, Kuhn M, Aplin S, Grant TD, Dörner K, Sato T, Echelmeier A, Villarreal JC, Hunter MS, Wiedorn MO, Knoska J, Mazalova V, Roy-Chowdhury S, Yang J-H, Jones A, Bean R, Bielecki J, Kim Y, Mills G, Weinhausen B, Meza JD, Al-Qudami N, Bajt S, Brehm G, Botha S, Boukhelef D, Brockhauser S, Bruce BD, Coleman MA, Danilevski C, Discianno E, Dobson Z, Fangohr H, Martin-Garcia JM, Gevorkov Y, Hauf S, Hosseinizadeh A, Januschek F, Ketawala GK, Kupitz C, Maia L, Manetti M, Messerschmidt M, Michelat T, Mondal J, Ourmazd A, Previtali G, Sarrou I, Schön S, Schwander P, Shelby ML, Silenzi A, Sztuk-Dambietz J, Szuba J, Turcato M, White TA, Wrona K, Xu C, Abdellatif MH, Zook JD, Spence JCH, Chapman HN, Barty A, Kirian RA, Frank M, Ros A, Schmidt M, Fromme R, Mancuso AP, Fromme P, Zatsepin NA. Membrane protein megahertz crystallography at the European XFEL. Nat Commun. 2019;10:5021.
22. Su X, Ma J, Pan X, Zhao X, Chang W, Liu Z, Zhang X, Li M. Antenna arrangement and energy transfer pathways of a green algal photosystem-I–LHCI supercomplex. Nat Plants. 2019;5:273–81.
23. Kato K, Nagao R, Jiang T-Y, Ueno Y, Yokono M, Chan SK, Watanabe M, Ikeuchi M, Shen J-R, Akimoto S, Miyazaki N, Akita F. Structure of a cyanobacterial photosystem I tetramer revealed by cryo-electron microscopy. Nat Commun. 2019;10:4929.
24. Suga M, Ozawa S-I, Yoshida-Motomura K, Akita F, Miyazaki N, Takahashi Y. Structure of the green algal photosystem I supercomplex with a decameric light-harvesting complex I. Nat Plants. 2019;5:626–36.
25. Kölsch A, Radon C, Golub M, Baumert A, Bürger J, Mielke T, Lisdat F, Feoktystov A, Pieper J, Zouni A, Wendler P. Current limits of structural biology: the transient interaction between cytochrome c6 and photosystem I. Curr Res Struct Biol. 2020;2:171–9.
26. Gisriel C, Shen G, Kurashov V, Ho M-Y, Zhang S, Williams D, Golbeck JH, Fromme P, Bryant DA. The structure of photosystem I acclimated to far-red light illuminates an ecologically important acclimation process in photosynthesis. Sci Adv. 2020;6:1–11.
27. Toporik H, Khmelnitskiy A, Dobson Z, Riddle R, Williams D, Lin S, Jankowiak R, Mazor Y. The structure of a red-shifted photosystem I reveals a red site in the core antenna. Nat Commun. 2020;11:5279.
28. Caspy I, Borovikova-Sheinker A, Klaiman D, Shkolnisky Y, Nelson N. The structure of a triple complex of plant photosystem I with ferredoxin and plastocyanin. Nat Plants. 2020;6:1300–5.
29. Akita F, Nagao R, Kato K, Nakajima Y, Yokono M, Ueno Y, Suzuki T, Dohmae N, Shen J-R, Akimoto S, Miyazaki N. Structure of a cyanobacterial photosystem I surrounded by octadecameric IsiA antenna proteins. Commun Biol. 2020;3:232.
30. Cao P, Cao D, Si L, Su X, Tian L, Chang W, Liu Z, Zhang X, Li M. Structural basis for energy and electron transfer of the photosystem I-IsiA–flavodoxin supercomplex. Nat Plants. 2020;6:167–76.
31. Kato K, Shinoda T, Nagao R, Akimoto S, Suzuki T, Dohmae N, Chen M, Allakhverdiev SI, Shen J-R, Akita F, Miyazaki N, Tomo T. Structural basis for the adaptation and function of chlorophyll f in photosystem I. Nat Commun. 2020;11:238.
32. Perez-Boerema A, Klaiman D, Caspy I, Netzer-El SY, Amunts A, Nelson N. Structure of a minimal photosystem I from the green alga *Dunaliella salina*. Nat Plants. 2020;6:321–7.
33. Çoruh O, Frank A, Tanaka H, Kawamoto A, El-Mohsnawy E, Kato T, Namba K, Gerle C, Nowaczyk MM, Kurisu G. Cryo-EM structure of a functional monomeric photosystem I from *Thermosynechococcus elongatus* reveals red chlorophyll cluster. Commun Biol. 2021;4:304.
34. Hamaguchi T, Kawakami K, Shinzawa-Itoh K, Inoue-Kashino N, Itoh S, Ifuku K, Yamashita E, Maeda K, Yonekura K, Kashino Y. Structure of the far-red light utilizing photosystem I of *Acaryochloris marina*. Nat Commun. 2021;12:2333.
35. Dobson Z, Ahad S, Vanlandingham J, Toporik H, Vaughn N, Vaughn M, Williams D, Reppert M, Fromme P, Mazor Y. The structure of photosystem I from a high-light-tolerant cyanobacteria. Elife. 2021;10:e67518.
36. Akhtar P, Caspy I, Nowakowski PJ, Malavath T, Nelson N, Tan H-S, Lambrev PH. Two-Dimensional electronic spectroscopy of a minimal photosystem I complex reveals the rate of primary charge separation. J Am Chem Soc. 2021;143:14601–12.
37. Keable SM, Kölsch A, Simon PS, Dasgupta M, Chatterjee R, Subramanian SK, Hussein R, Ibrahim M, Kim I-S, Bogacz I, Makita H, Pham CC, Fuller FD, Gul S, Paley D, Lassalle L, Sutherlin KD, Bhowmick A, Moriarty NW, Young ID, Blaschke JP, de Lichtenberg C, Chernev P, Cheah MH, Park S, Park G, Kim J, Lee SJ, Park J, Tono K, Owada S, Hunter MS, Batyuk A, Oggenfuss R, Sander M, Zerdane S, Ozerov D, Nass K, Lemke H, Mankowsky R, Brewster AS, Messinger J, Sauter NK, Yachandra VK, Yano J, Zouni A, Kern J. Room temperature XFEL crystallography reveals asymmetry in the vicinity of the two phylloquinones in photosystem I. Sci Rep. 2021;11:21787.
38. Caspy I, Schwartz T, Bayro-Kaiser V, Fadeeva M, Kessel A, Ben-Tal N, Nelson N. Dimeric and high-resolution structures of chlamydomonas photosystem I from a temperature-sensitive photosystem II mutant. Commun Biol. 2021;4:1380.
39. Caspy I, Neumann E, Fadeeva M, Liveanu V, Savitsky A, Frank A, Kalisman YL, Shkolnisky Y, Murik O, Treves H, Hartmann V, Nowaczyk MM, Schuhmann W, Rögner M, Willner I, Kaplan A, Schuster G, Nelson N, Lubitz W, Nechushtai R. Cryo-EM photosystem I structure reveals adaptation mechanisms to extreme high light in *Chlorella ohadii*. Nat Plants. 2021;7:1314–22.
40. Wang J, Yu LJ, Wang W, Yan Q, Kuang T, Qin X, Shen JR. Structure of plant photosystem I-light harvesting complex I supercomplex at 2.4 Å resolution. J Integr Plant Biol. 2021;63:1367–81.
41. Yan Q, Zhao L, Wang W, Pi X, Han G, Wang J, Cheng L, He Y-K, Kuang T, Qin X, Sui S-F, Shen J-R. Antenna arrangement and energy-transfer pathways of PSI–LHCI from the moss *Physcomitrella patens*. Cell Discov. 2021;7:10.
42. Huang Z, Shen L, Wang W, Mao Z, Yi X, Kuang T, Shen J-R, Zhang X, Han G. Structure of photosystem I-LHCI-LHCII from the green alga *Chlamydomonas reinhardtii* in State 2. Nat Commun. 2021;12:1100.

43. Pan X, Tokutsu R, Li A, Takizawa K, Song C, Murata K, Yamasaki T, Liu Z, Minagawa J, Li M. Structural basis of LhcbM5-mediated state transitions in green algae. Nat Plants. 2021;7:1119–31.
44. Xu C, Zhu Q, Chen J-H, Shen L, Yi X, Huang Z, Wang W, Chen M, Kuang T, Shen J-R, Zhang X, Han G. A unique photosystem I reaction center from a chlorophyll d-containing cyanobacterium *Acaryochloris marina*. J Integr Plant Biol. 2021;63:1740–52.
45. Semchonok DA, Mondal J, Cooper CJ, Schlum K, Li M, Amin M, Sorzano COS, Ramírez-Aportela E, Kastritis PL, Boekema EJ, Guskov A, Bruce BD. Cryo-EM structure of a tetrameric photosystem I from chroococcidiopsis TS-821, a thermophilic, unicellular, non-heterocyst-forming cyanobacterium. Plant Commun. 2022;3:100248.
46. Naschberger A, Mosebach L, Tobiasson V, Kuhlgert S, Scholz M, Perez-Boerema A, Ho TTH, Vidal-Meireles A, Takahashi Y, Hippler M, Amunts A. Algal photosystem I dimer and high-resolution model of PSI-plastocyanin complex. Nat Plants. 2022;8:1191–201.
47. K. Kato, R. Nagao, Y. Ueno, M. Yokono, T. Suzuki, T.-Y. Jiang, N. Dohmae, F. Akita, S. Akimoto, N. Miyazaki, J.-R. Shen, Structural insights into an evolutionary turning-point of photosystem I from prokaryotes to eukaryotes, bioRxiv : the preprint server for biology (2022) 2022.2001.2003.474851.
48. Gorski C, Riddle R, Toporik H, Da Z, Dobson Z, Williams D, Mazor Y. The structure of the *Physcomitrium patens* photosystem I reveals a unique Lhca2 paralogue replacing Lhca4. Nat Plants. 2022;8:307–16.
49. Shen L, Tang K, Wang W, Wang C, Wu H, Mao Z, An S, Chang S, Kuang T, Shen J-R, Han G, Zhang X. Architecture of the chloroplast PSI–NDH supercomplex in Hordeum vulgare. Nature. 2022;601:649–54.
50. Su X, Cao D, Pan X, Shi L, Liu Z, Dall'Osto L, Bassi R, Zhang X, Li M. Supramolecular assembly of chloroplast NADH dehydrogenase-like complex with photosystem I from *Arabidopsis thaliana*. Mol Plant. 2022;15:454–67.
51. Li J, Hamaoka N, Makino F, Kawamoto A, Lin Y, Rögner M, Nowaczyk MM, Lee Y-H, Namba K, Gerle C, Kurisu G. Structure of cyanobacterial photosystem I complexed with ferredoxin at 1.97 Å resolution. Commun Biol. 2022;5:951.
52. Gisriel CJ, Flesher DA, Shen G, Wang J, Ho M-Y, Brudvig GW, Bryant DA. Structure of a photosystem I-ferredoxin complex from a marine cyanobacterium provides insights into far-red light photoacclimation. J Biol Chem. 2022;298:101408.
53. Nagao R, Kato K, Hamaguchi T, Ueno Y, Tsuboshita N, Shimizu S, Furutani M, Ehira S, Nakajima Y, Kawakami K, Suzuki T, Dohmae N, Akimoto S, Yonekura K, Shen J-R. Structure of a monomeric photosystem I core associated with iron-stress-induced-A proteins from Anabaena sp. PCC 7120. Nat Commun. 2023;14:920.
54. You X, Zhang X, Cheng J, Xiao Y, Ma J, Sun S, Zhang X, Wang H-W, Sui S-F. In situ structure of the red algal phycobilisome–PSII–PSI–LHC megacomplex. Nature. 2023;616:199–206.
55. Schwartz T, Fadeeva M, Klaiman D, Nelson N. Structure of photosystem I supercomplex isolated from a *Chlamydomonas reinhardtii* cytochrome b6f temperature-sensitive mutant. Biomolecules. 2023;13:537.
56. Gerle C, Misumi Y, Kawamoto A, Tanaka H, Kubota-Kawai H, Tokutsu R, Kim E, Chorev D, Abe K, Robinson CV, Mitsuoka K, Minagawa J, Kurisu G. Three structures of PSI-LHCI from *Chlamydomonas reinhardtii* suggest a resting state re-activated by ferredoxin. Biochim Biophys Acta (BBA) Bioenergetics. 2023;1864:148986.
57. Zhao L-S, Wang P, Li K, Zhang Q-B, He F-Y, Li C-Y, Su H-N, Chen X-L, Liu L-N, Zhang Y-Z. Structural basis and evolution of the photosystem I–light-harvesting supercomplex of cryptophyte algae. Plant Cell. 2023;35:2449–63.
58. Harris D, Toporik H, Schlau-Cohen GS, Mazor Y. Energetic robustness to large scale structural fluctuations in a photosynthetic supercomplex. Nat Commun. 2023;14:4650.
59. Zhang S, Tang K, Yan Q, Li X, Shen L, Wang W, He Y-K, Kuang T, Han G, Shen J-R, Zhang X. Structural insights into a unique PSI–LHCI–LHCII–Lhcb9 supercomplex from moss *Physcomitrium patens*. Nat Plants. 2023;9:832–46.
60. Zhao Z, Vercellino I, Knoppová J, Sobotka R, Murray JW, Nixon PJ, Sazanov LA, Komenda J. The Ycf48 accessory factor occupies the site of the oxygen-evolving manganese cluster during photosystem II biogenesis. Nat Commun. 2023;14:4681.
61. Li X, Li Z, Wang F, Zhao S, Xu C, Mao Z, Duan J, Feng Y, Yang Y, Shen L, Wang G, Yang Y, Yu LJ, Sang M, Han G, Wang X, Kuang T, Shen JR, Wang W. Structures and organizations of PSI-AcpPCI supercomplexes from red tidal and coral symbiotic photosynthetic dinoflagellates. Proc Natl Acad Sci U S A. 2024;121:e2315476121.
62. Kato K, Hamaguchi T, Kumazawa M, Nakajima Y, Ifuku K, Hirooka S, Hirose Y, Miyagishima SY, Suzuki T, Kawakami K, Dohmae N, Yonekura K, Shen JR, Nagao R. The structure of PSI-LHCI from Cyanidium caldarium provides evolutionary insights into conservation and diversity of red-lineage LHCs. Proc Natl Acad Sci U S A. 2024;121:e2319658121.
63. Jaccard P. Etude comparative de la distribution florale dans une portion des Alpes et des Jura. Bull Soc Vaudoise Sci Nat. 1901;37:547–79.
64. Ackerman M, Ben-David S. A characterization of linkage-based hierarchical clustering. J Mach Learn Res. 2016;17:8182–98.
65 Kruskal JB, Wish M. Multidimensional Scaling. California: Thousand Oaks; 1978.
66. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14:33–8.
67. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.
68. Sarkar T, Chen Y, Wang Y, Chen Y, Chen F, Reaux CR, Moore LE, Raghavan V, Xu W. Introducing mirror-image discrimination capability to the TSR-based method for capturing stereo geometry and understanding hierarchical structure relationships of protein receptor family. Comput Biol Chem. 2023;103:107824.
69. Ballester PJ, Richards WG. Ultrafast shape recognition to search compound databases for similar molecular shapes. J Comput Chem. 2007;28:1711–23.
70 Kufareva I, Abagyan R. Methods of protein structure comparison. Methods Mol Biol (Clifton, NJ). 2012;857:231–57.
71. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucl Acids Res. 2005;33:2302–9.
72 Semchonok DA, Li M, Bruce BD, Oostergetel GT, Boekema EJ. Cryo-EM structure of a tetrameric cyanobacterial photosystem I complex reveals novel subunit interactions. Biochim Biophys Acta (BBA) Bioenergetics. 1857;2016:1619–26.

73 Netzer-El SY, Caspy I, Nelson N. Crystal structure of photosystem I monomer from synechocystis PCC 6803. Front Plant Sci. 2019;9:1865.
74 Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, Krauß N. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. Nature. 2001;411:909–17.
75. Santabarbara S, Heathcote P, Evans MCW. Modelling of the electron transfer reactions in Photosystem I by electron tunnelling theory: The phylloquinones bound to the PsaA and the PsaB reaction centre subunits of PS I are almost isoenergetic to the iron–sulfur cluster FX. Biochim Biophys Acta (BBA) Bioenergetics. 2005;1708:283–310.
76. Fresch E, Meneghin E, Agostini A, Paulsen H, Carbonera D, Collini E. How the protein environment can tune the energy, the coupling, and the ultrafast dynamics of interacting chlorophylls: the example of the water-soluble chlorophyll protein. J Phys Chem Lett. 2020;11:1059–67.
77. Li SC. The difficulty of protein structure alignment under the RMSD. Algorithms Mol Biol. 2013;8:1.
78. Sapundzhi F, Popstoilov M, Lazarova M. Calculations for Comparing Protein Three-Dimensional Structures. In: Georgiev I, Datcheva M, Georgiev K, Nikolov G, editors. Numerical Methods and Applications. Cham: Springer Nature Switzerland; 2023. p. 279–88.
79. Zemla A. LGA: A method for finding 3D similarities in protein structures. Nucl Acids Res. 2003;31:3370–4.
80. Ballester PJ, Finn PW, Richards WG. Ultrafast shape recognition: evaluating a new ligand-based virtual screening technology. J Mol Graph Model. 2009;27:836–45.
81. Ballester PJ. Ultrafast shape recognition: method and applications. Future Med Chem. 2011;3:65–78.
82. Fish LE, Kück U, Bogorad L. Two partially homologous adjacent light-inducible maize chloroplast genes encoding polypeptides of the P700 chlorophyll a-protein complex of photosystem I. J Biol Chem. 1985;260:1413–21.
83 Fromme P, Schubert W-D, Krauß N. Structure of Photosystem I: Suggestions on the docking sites for plastocyanin, ferredoxin and the coordination of P700. Biochim Biophys Acta (BBA) Bioenergetics. 1994;1187:99–105.
84 Fromme P, Jordan P, Krauß N. Structure of photosystem I. Biochim Biophys Acta (BBA) Bioenergetics. 2001;1507:5–31.
85. Sarkar T, Raghavan VV, Chen F, Riley A, Zhou S, Xu W. Exploring the effectiveness of the TSR-based protein 3-D structural comparison method for protein clustering, and structural motif identification and discovery of protein kinases, hydrolase, and SARS-CoV-2's protein via the application of amino acid grouping. Comput Biol Chem. 2021;92:107479.
86. Kondra S, Chen F, Chen Y, Chen Y, Collette CJ, Xu W. A study of a hierarchical structure of proteins and ligand binding sites of receptors using the triangular spatial relationship-based structure comparison method and development of a size-filtering feature designed for comparing different sizes of protein structures. Proteins. 2022;90:239–57.
87. Wu S, Zhang Y. Recognizing protein substructure similarity using segmental threading. Structure. 2010;18:858–67.
88. Ali K, Santabarbara S, Heathcote P, Evans MCW, Purton S. Bidirectional electron transfer in photosystem I: replacement of the symmetry-breaking tryptophan close to the PsaB-bound phylloquinone (A1B) with a glycine residue alters the redox properties of A1B and blocks forward electron transfer at cryogenic temperatures. Biochim Biophys Acta (BBA) Bioenergetics. 2006;1757:1623–33.
89. Agarwala N, Makita H, Luo L, Xu W, Hastings G. Reversible inhibition and reactivation of electron transfer in photosystem I. Photosynth Res. 2020;145:97–109.
90. Kurashov V, Milanovsky G, Luo L, Martin A, Semenov AY, Savikhin S, Cherepanov DA, Golbeck JH, Xu W. Conserved residue PsaB-Trp673 is essential for high-efficiency electron transfer between the phylloquinones and the iron-sulfur clusters in photosystem I. Photosynth Res. 2021;148:161–80.
91. Liu B, Liu H, Zhong D, Lin C. Searching for a photocycle of the cryptochrome photoreceptors. Curr Opin Plant Biol. 2010;13:578–86.
92. Monni R, Al -Haddad A, van Mourik F, Auböck G, Chergui M. Tryptophan-to-heme electron transfer in ferrous myoglobins. Proc Natl Acad Sci. 2015;112:5602–6.
93 Ghani NSA, Emrizal R, Moffit SM, Hamdani HY, Ramlan EI, Firdaus-Raih M. GrAfSS: a webserver for substructure similarity searching and comparisons in the structures of proteins and RNA. Nucl Acids Res. 2022;50:W375–83.
94. W. Xu, Y. Wang, E. Taylor, A. Laujac, L. Gao, S. Savikhin, P.R. Chitnis, Mutational analysis of photosystem I of Synechocystis sp. PCC 6803: the role of four conserved aromatic residues in the j-helix of PsaB, PLoS One 6 (2011) e24625.
95. Krabben L, Schlodder E, Jordan R, Carbonera D, Giacometti G, Lee H, Webber AN, Lubitz W. Influence of the axial ligands on the spectral properties of P700 of photosystem I: a study of site-directed mutants. Biochemistry. 2000;39:13012–25.
96. Webber AN, Su H, Bingham SE, Käss H, Krabben L, Kuhn M, Jordan R, Schlodder E, Lubitz W. Site-directed mutations affecting the spectroscopic characteristics and midpoint potential of the primary donor in photosystem I. Biochemistry. 1996;35:12857–63.
97. Breton J, Xu W, Diner BA, Chitnis PR. The two histidine axial ligands of the primary electron donor chlorophylls (P700) in photosystem I are similarly perturbed upon P700+ formation. Biochemistry. 2002;41:11200–10.
98. Redding K, MacMillan F, Leibl W, Brettel K, Hanley J, Rutherford AW, Breton J, Rochaix JD. A systematic survey of conserved histidines in the core subunits of photosystem I by site-directed mutagenesis reveals the likely axial ligands of P_{700}. EMBO J. 1998;17:50–60.
99. Ramesh VM, Guergova-Kuras M, Joliot P, Webber AN. Electron transfer from plastocyanin to the photosystem I reaction center in mutants with increased potential of the primary donor in *Chlamydomonas reinhardtii*. Biochemistry. 2002;41:14652–8.
100. Cherepanov DA, Shelaev IV, Gostev FE, Nadtochenko VA, Xu W, Golbeck JH, Semenov AY. Symmetry breaking in photosystem I: ultrafast optical studies of variants near the accessory chlorophylls in the A- and B-branches of electron transfer cofactors. Photochem Photobiol Sci Off J Eur Photochem Assoc Eur Soc Photobiol. 2021;20:1209–27.
101. Cherepanov DA, Shelaev IV, Gostev FE, Petrova A, Aybush AV, Nadtochenko VA, Xu W, Golbeck JH, Semenov AY. Primary charge separation within the structurally symmetric tetrameric Chl2APAPBChl2B chlorophyll exciplex in photosystem I. J Photochem Photobiol, B. 2021;217:112154.

102. Kirpich JS, Luo L, Nelson MR, Agarwala N, Xu W, Hastings G. Is the A-1 pigment in photosystem I part of P700? A (P700+–P700) FTIR difference spectroscopy study of A-1 mutants. Int J Mol Sci. 2024;25:4839.
103. Badshah SL, Sun J, Mula S, Gorka M, Baker P, Luthra R, Lin S, van der Est A, Golbeck JH, Redding KE. Mutations in algal and cyanobacterial photosystem I that independently affect the yield of initial charge separation in the two electron transfer cofactor branches. Biochim Biophys Acta (BBA) Bioenergetics. 1859;2018:42–55.
104 Xu W, Chitnis PR, Valieva A, van der Est A, Brettel K, Guergova-Kuras M, Pushkar YN, Zech SG, Stehlik D, Shen G, Zybailov B, Golbeck JH. Electron transfer in cyanobacterial photosystem I: II. Determination of forward electron transfer rates of site-directed mutants in a putative electron transfer pathway from A0 through A1 to FX. J Biol Chem. 2003;278:27876–87.
105. Guergova-Kuras M, Boudreaux B, Joliot A, Joliot P, Redding K. Evidence for two active branches for electron transfer in photosystem I. Proc Natl Acad Sci. 2001;98:4437–42.
106. Dashdorj N, Xu W, Cohen RO, Golbeck JH, Savikhin S. Asymmetric electron transfer in cyanobacterial Photosystem I: charge separation and secondary electron transfer dynamics of mutations near the primary electron acceptor A0. Biophys J. 2005;88:1238–49.
107. Cohen RO, Shen G, Golbeck JH, Xu W, Chitnis PR, Valieva AI, van der Est A, Pushkar Y, Stehlik D. Evidence for asymmetric electron transfer in cyanobacterial photosystem I: analysis of a methionine-to-leucine mutation of the ligand to the primary electron acceptor A0. Biochemistry. 2004;43:4741–54.
108. Santabarbara S, Kuprov I, Poluektov O, Casal A, Russell CA, Purton S, Evans MCW. Directionality of electron-transfer reactions in photosystem I of prokaryotes: universality of the bidirectional electron-transfer model. J Phys Chem B. 2010;114:15158–71.
109. Sun J, Hao S, Radle M, Xu W, Shelaev I, Nadtochenko V, Shuvalov V, Semenov A, Gordon H, van der Est A, Golbeck JH. Evidence that histidine forms a coordination bond to the A0A and A0B chlorophylls and a second H-bond to the A1A and A1B phylloquinones in M688HPsaA and M668HPsaB variants of *Synechocystis* sp. PCC 6803. Biochim Biophys Acta (BBA) Bioenergetics. 1837;2014:1362–75.
110. van der Est A, Chirico S, Karyagina I, Cohen R, Shen G, Golbeck JH. Alteration of the axial met ligand to electron acceptor A0 in photosystem I: an investigation of electron transfer at different temperatures by multifrequency time-resolved and CW EPR. Appl Magn Reson. 2009;37:103.
111. Savitsky A, Gopta O, Mamedov M, Golbeck JH, Tikhonov A, Möbius K, Semenov A. Alteration of the axial met ligand to electron acceptor A0 in photosystem I: effect on the generation of P700·+A1·−radical pairs as studied by W-band transient EPR. Appl Magn Reson. 2010;37:85–102.
112. Gorka M, Gruszecki E, Charles P, Kalendra V, Lakshmi KV, Golbeck JH. Two-dimensional HYSCORE spectroscopy reveals a histidine imidazole as the axial ligand to Chl(3A) in the M688H(PsaA) genetic variant of Photosystem I. Biochim Biophys Acta. 2021;1862:148424.
113. McConnell MD, Sun J, Siavashi R, Webber A, Redding KE, Golbeck JH, van der Est A. Species-dependent alteration of electron transfer in the early stages of charge stabilization in Photosystem I. Biochim Biophys Acta. 1847;2015:429–40.

Publisher's Note