Photosystem I (PSI) is a large membrane protein complex vital for oxygenic photosynthesis, one of the most important biological processes on the planet. We present an “atomic” model of higher plant PSI, based on theoretical modeling using the recent 4.4 Å X-ray crystal structure of PSI from pea. Due to the lack of information on the amino acid sidechains in the X-ray structural model and the high cofactor content in this system, novel modeling techniques were developed. Our model reveals some important structural features of plant PSI that were not visible in the crystal structure, and sheds light on the evolutionary relationship between plant and cyanobacterial PSI.

Photosynthesis is one of the most important biological processes on Earth – one in which radiant energy from the sun is stored as chemical energy, which can be used as an energy source by all forms of life. (1) In both eukaryotes and prokaryotes, the initial steps of photosynthesis take place across the photosynthetic membrane. This membrane divides an internal space, known as the lumen, from the cytosol outside the membrane, known as the stroma. During photosynthesis, two large membrane protein complexes, photosystems I and II (PSI and PSII), harness the energy of incident photons and use it to drive a series of electron transfer reactions across the photosynthetic membrane that result in the establishment of a transmembrane electrochemical proton gradient that drives synthesis of ATP. PSII provides electrons for these redox reactions by splitting water into molecular oxygen and protons, while PSI provides the reducing power to reduce NADP⁺ to NADPH. The high-energy products ATP and NADPH are then used in the dark reactions of photosynthesis to fix CO₂ and synthesize the sugars.

Oxygenic photosynthesis is performed both by the chloroplasts of eukaryotic organisms such as plants and green algae, and by prokaryotes such as cyanobacteria. Approximately 1.5 billion years ago, organisms similar to cyanobacteria entered into an endosymbiotic relationship with anaerobic eukaryotic cells, becoming the ancestors of modern chloroplasts. Plant and cyanobacterial photosystems therefore share a common origin and perform very similar functions, but exist in different biological environments. (2) We present a model of plant PSI in which the functional differences between plant and cyanobacterial PSI can be correlated with structural features unique to each system. This can allow insight into the parallel evolution of these two highly optimized systems, as well as providing a guide for future experimental work that will further explore the structure and function of plant PSI.

Photosystem I catalyzes the light-driven electron transfer from the soluble electron carrier plastocyanin or cytochrome c₆, located at the luminal side of thylakoid membrane, to ferredoxin, which is located at the stromal side. In cyanobacteria, PSI exists as a trimer; each monomer is composed of 12
individual protein subunits, 96 chlorophyll α molecules, 22 carotenoids, three [4Fe4S] clusters, and two phylloquinones. (3) Plant PSI is a monomer containing fourteen subunits. Ten of these are very similar to corresponding subunits in the cyanobacterial structure: PsaA, PsaB, PsaC, PsaD, PsaE, PsaF, PsaI, PsaJ, PsaK, and PsaL. Plant PSI lacks two subunits that are unique to cyanobacteria (PsaX and PsaM), and contains four subunits that are absent in cyanobacteria (PsaG, PsaH, PsaN, PsaO).

Both plant and cyanobacterial PSI can bind external antenna systems when extra light-harvesting capacity is needed. The diversity of the external antenna systems is one of the most interesting differences between different photosynthetic organisms. The major external antenna system in cyanobacteria consists of large, membrane-extrinsic complexes known as the phycobilisomes. The phycobilisomes use bilin-based pigments to absorb light in the green spectral region, precisely where PSI’s antenna system does not absorb strongly (4).

Cyanobacteria subjected to low iron concentrations also synthesize an iron-stress-induced membrane protein called IsiA, which forms a symmetric 18-meric ring around the PSI trimer (4-6). The external antenna complex of PSI in plants is formed by the light-harvesting complex I and II proteins (LHCI and LHCII), which dock to the periphery of the monomeric PSI complex. The external antenna system in plants is asymmetric and heterogeneous, and is strongly modified under changing environmental conditions (e.g. light intensity, wavelength of light, oxygen supply, etc.).

Recently, a breakthrough in the understanding of plant photosynthesis has been achieved with the crystallization of a supercomplex of plant photosystem I from *pisum sativum* (garden pea) with its peripheral LHCI antenna. A medium-resolution structure of this PSI-LHCI supercomplex has been obtained by X-ray structure analysis at 4.4 Å resolution. (7). This was the first structure of a plant membrane protein to be determined by X-ray crystallography. The PSI-LHCI structure contains the subunits conserved between plants and cyanobacteria, along with the two of the non-conserved eukaryotic PS I subunits PsaG and PsaH, and four external light-harvesting complexes, tentatively assigned to Lhca1, Lhca2, Lhca3, and Lhca4.

One of the most remarkable features of this structure is the similarity of its conserved subunits to their counterparts in cyanobacterial PSI, whose structure was determined at 2.5 Å resolution (3,8). In most of the conserved subunits, the protein backbone conformation is virtually identical (see also Tab. 1). Because a 4.4 Å structure is able to reveal the location of the protein backbone but not the identity or conformation of the amino acid side chains, the plant PSI crystal structure answers many questions, but allows us to ask many more. If the backbone conformation is so similar, then what accounts for the functional differences between plant and cyanobacterial PSI? Where differences in the backbone conformation exist, what is their significance? What light can this structure shed on the results of biochemical studies of PSI function in plants and cyanobacteria?

To answer some of these questions and to provide detailed information for mutagenesis experiments and functional studies, we have derived an atomic-level computational model of plant photosystem I. We have used three different crystal structures in an effort to understand the structure of the plant PSI-LHCI complex at a more detailed level: the 2.5 Å structure of trimeric cyanobacterial PSI, the 4.4 Å structure of the plant PSI-LHCI supercomplex and the recent 2.72 Å structure of the LHCII complex (9). LHCII belongs to the same protein family as LHCI and therefore is helpful in the modeling of the LHCI proteins.

Materials and Methods

The computational model is based primarily on two structures: the 4.4 Å structural model from pea PSI-LHCI (7) and the 2.5 Å structure of cyanobacterial PSI (3). We have avoided “improving” upon the medium-resolution structural model – i.e. modifying the backbone conformation of
regions assigned in the X-ray structural model. This is also true for the chlorophyll molecules, where the position and macrocycle orientation from the medium resolution X-ray structure has been preserved in all cases.

Sequence data for the plant PSI subunits was obtained from the Swiss-Prot/TrEMBL sequence database (10). Homology models of each subunit were constructed individually, using a combination of the MOE software package (Chemical Computing Group, Inc.), the Swiss-Model online modeling server (11), and the Swiss-PDB viewer (12).

In addition to the structural information included in the plant X-ray structural model, the modeling of the LHCI proteins was based on the crystal structure of plant LHCII (9). The LHCI and LHCII proteins exhibit 25-35% sequence identity and share a similar fold, but form different oligomeric structures: LHCII forms trimers, while LHCI forms dimers.

In most cases, the sequence-structure alignment in our models of Lhca1 and Lhca4 follows the residue numbering in the plant X-ray structural model. This numbering was obtained by aligning LHCI and LHCII sequences from higher plants and placing the residues homologous to those coordinating chlorophyll in LHCII nearest the observed chlorophyll molecules. After adding amino acid sidechains into the structure, it became clear that, in a few cases, a shift of one turn in the helix (i.e. three amino acids) was needed to improve chlorophyll coordination geometry. In some cases (e.g. chlorophyll 1012 in Lhca4), the coordination geometry at one chlorophyll was compromised in order to improve coordination geometry for a neighboring chlorophyll.

Several loop regions were not visible in the plant PSI crystal structure. The MOE software package was used to obtain plausible starting conformations for these loops. A luminal portion of Lhca1 (Val 69-Leu 77) obtained its starting conformation from a weak, previously unassigned electron density. Lhca4 contains a rather long stromal loop that was not well-defined in the crystal structure and was minimized extensively to reach a suitable conformation, which is to be considered highly speculative (See Figure S1 in the Supplementary Information).

Some sidechain conformations in Lhca1 and Lhca4 were constrained by positioning them such that they could form hydrogen bonds with chlorophyll. Although the hydrogen-bond donors identified in the plant LHCII crystal structure were poorly conserved in the LHCI proteins, other residues were found which were well positioned to form hydrogen bonds and widely conserved among higher plants. (See Table S2 in the Supplementary Information) Not surprisingly, some of the hydrogen-bonding and ligation geometries are non-ideal. This could be a result of the fact that the stereochemical quality of the model is limited by the resolution of the template structure. It is also possible, of course, that these interactions would not be present in a higher-resolution structure.

The modeling of cofactors presented a unique challenge. Non-protein cofactors make up more than 30% of the total mass of plant PSI, yet traditional homology modeling approaches consider only polypeptides (13). Our approach to modeling cofactors is similar to one that has been described for buried water (14). An approach similar to ours has also been applied successfully in the modeling of protein and cofactors in the R. capsulatus reaction center (15).

Initial coordinates for most cofactors in the core of the complex were obtained from the cyanobacterial PSI structure (PDB accession code 1JB0). In general, chlorophyll positions assigned in the plant X-ray structural model were remarkably similar to those in the cyanobacterial structure, and it is reasonable to assume that, when position is conserved, orientation will be conserved also. A more exact fit to the data provided in the plant X-ray structural model was achieved by positioning each chlorophyll molecule such that the four nitrogen atoms in its macrocyclic head group aligned perfectly with the positions assigned for nitrogen atoms in the plant X-ray structural model. The positions of these nitrogen atoms were held fixed during minimization, forcing the rest of the molecule to fall into line.
Because the protein environment of cofactors differs between cyanobacteria and plants, extensive minimization was needed to optimize the cofactor conformations relative to their environment, especially in the flexible phytol tail.

Due to the limited resolution of the plant X-ray structural model, no beta-carotene positions were assigned. However, we found that most of the beta-carotenes in cyanobacteria were located in pockets of hydrophobic residues that are highly-conserved between cyanobacteria and plants. Other cofactors were also included in the plant PSI model. Three phospholipids and one galactolipid were included in the core complex, along with one phospholipid in each of the light-harvesting complexes. Two phylloquinones and three iron-sulfur clusters were placed at the same positions in which they were assigned in the plant X-ray structural model, as part of the electron transport chain.

Once an initial structure for the model was constructed and minimized we subjected it to a molecular dynamics simulation, in order to further optimize the structure and explicitly account for the water/lipid environment of PSI. The model was placed in a solvated bilayer of digalactosyldiacylglyceride (DGDG), which had been equilibrated at constant pressure in a molecular dynamics simulation. DGDG was chosen because it is a major lipid component of the thylakoid membranes in which PSI resides (16). The final system contained the modeled proteins and cofactors, 529 membrane lipid molecules, about 86,000 water molecules, and a sodium ion, needed for electrical neutrality. In order to maintain the secondary structure present in the starting structure, the protein alpha-carbon and chlorophyll nitrogen atoms which were assigned in the plant PSI X-ray structural model were bound to their starting positions by a harmonic potential \( k = 100 \text{ kcal mol}^{-1} \text{ Å}^{-2} \). All other atoms remained free to move, including the alpha-carbons of those regions unassigned in the X-ray model. In order to assess the progress of the MD simulation, the positions of these “flexible” alpha-carbon atoms were compared to their positions in the starting structure, and a global RMSD was calculated.

Once the system had been assembled, it was simulated at constant pressure at 300 K, with the protein and associated cofactors held fixed, so that the lipid and solvent could equilibrate. Once the volume remained stable, the atoms of the PSI model were released and could move freely, except for the restraints described above. Once a new stable volume was reached, solvent molecules more than 15 Å away from the protein were held fixed, in order to decrease the simulation time. The simulation was allowed to continue until the global RMSD of the loop regions had stopped increasing, indicating that the system had become conformationally stable. At this point, the system was repeatedly minimized (at \( T = 0 \)) and gradually warmed back up to 300 K, in order to find the best possible global minimum. The final simulation step was a simulated annealing process, in which the temperature of the system was gradually lowered from 300 to 0 K.

One consequence of this simulation in explicit solvent was that water molecules from the bulk solvent migrated into the pigment-protein complex in order to ligate chlorophyll molecules for which a ligand could not previously be found.

All minimization and equilibration steps were performed using NAMD 2.5 (17). Because the force field used (CHARMM27) lacks parameters for many of the non-protein cofactors present in plant PSI, parameter files needed to be developed, based on a combination of \textit{ab initio} calculations (Vaswani and Damjanović, personal communication, (18)) and estimates based on known parameters for chemically similar compounds (19).

A more detailed description of the modeling process can be found in the Supplementary Information accompanying this article.

Results and Discussion

The modeling of plant photosystem I presented a unique challenge for several reasons. The most obvious is the presence of
the numerous non-protein cofactors in the supercomplex. These are not addressed by traditional homology modeling techniques (13), and their arrangement is constrained by – and provides constraints on – their protein milieu. Correct assignment of cofactor location and conformation was of great importance to the correctness of the model.

Another challenge came from the nature of the template structure used in modeling. In the production of an “atomic” model based on medium-resolution experimental data, a balance must be found between stereochemical quality and fidelity to the experimental data. In several places, faithfulness to the medium-resolution crystal structure led to non-ideal backbone geometries, especially in the transmembrane helices of PsaG and the light-harvesting complexes (see Materials and Methods).

A further challenge for the modeling of the structure was the size of the PS I-LHCI supercomplex and its location in the membrane, where not only the aqueous environment but also the lipid bilayer has to be included in the modeling studies, which increases the size of the model system to 407,280 atoms. To limit the computation time we have used a two-step approach: the structure was modeled in vacuo in the first step, and the membrane/aqueous environment was included in the second modeling step.

The model of plant PSI and its comparison with the structure of cyanobacterial PSI is summarized in Table 1. The model allows, for the first time, an atomic-level examination of the evolution of photosystem I from two major kingdoms of life whose evolutionary branches have separated for more than one billion years. Table 1 provides an overview of the major functional and structural similarities and differences between plant and cyanobacterial PS I.

The most striking result of the comparison is that the core structure of photosystem I was conserved over more than one billion years of evolution. Not only the backbone conformation, but also the sidechains of the core complex are highly similar in the plant and cyanobacterial structures. The subunit C, which carries the terminal FeS clusters $F_A$ and $F_B$, shows the highest degree of homology, followed by the major subunits A and B. The homology is not restricted to the protein, but extends to the coordination site for the carriers of the ET chain and the antenna pigments. It is even more exciting that even the carotenoids and the lipids, that have been identified in the high resolution X-ray structure of cyanobacterial PS I, but have not been assigned in the medium resolution plant PS I structure may be conserved between plants and cyanobacteria.

As the core of PS I is well conserved, plant PS I has optimized the docking sites for the terminal electron acceptor ferredoxin and the plastocyanine/cytochrome $c_6$ docking site, and has evolved a different oligomeric structure. In contrast to cyanobacteria, where it is essentially unknown where the phycobilisomes dock to PSI, the plant structural model allows investigation of the functional interaction of PSI with its external LHC complexes at a previously impossible level of detail (20).

We will now discuss the major features of the plant structure and its implications for the function of plant PSI in more detail.

**Plastocyanin binding site** – The most important difference between plant and cyanobacterial PSI is that plant PSI has evolved a tighter docking site for plastocyanin (PC). Whereas cyanobacteria use cytochrome $c_6$, which is regarded to be the most ancient electron donor for photosynthetic systems (21) or plastocyanin, which contains Cu, plant PS I is can only use plastocyanin as an electron donor. Electron transfer between plant PSI and plastocyanin shows second-order kinetics and is 2-3 orders of magnitude faster than in cyanobacteria, indicating that a more stable PSI-PC complex is formed in plants. This kinetic improvement is associated with the N-terminal region of PsaF (22). Site-directed mutagenesis studies have identified several lysine residues in this N-terminal region (23) as well as an acidic patch on plastocyanin (24,25) as vital for the formation of the PSI-PC complex. The crystal structure of plant
PsaF shows that the N-terminal region of PsaF forms a helix-turn-helix loop on the lumenal side of PSI, but could not identify the sidechains. This loop is much longer in plants and green algae than in cyanobacteria.

Our model shows that, on this lumenal loop, eight positively-charged amino acid residues point directly into the lumen – only two of which are conserved in cyanobacteria. (Figs. 2a,3c,3d) We have thereby identified the amino acids that may be responsible for the electrostatic interaction with plastocyanin in plants.

Whereas the direct involvement of PsaF in the active docking of PC is well known, the lumenal loop of PsaB also plays an important role in the stabilization of the PC docking site. Site-directed mutagenesis studies (26) have shown that Glu 611 on PsaB is also crucial for the proper interactions between plastocyanin and PSI in C. reinhardtii. This finding can be explained by our model as it shows that this residue forms a strong salt bridge with Arg 17 of PsaF, suggesting that it is responsible for holding the lumenal loop of PsaF in the proper orientation for plastocyanin binding.

Ferredoxin binding site – The structural model can also provide us with valuable insight about the optimization of the interactions of PSI with its soluble electron acceptor, ferredoxin (Fd). In photosystem I from cyanobacteria and plants, the small subunits PsaC, PsaD, and PsaE form a stromal ridge on top of PSI, which interacts with ferredoxin. After reduction by PSI, ferredoxin is oxidized by ferredoxin:NADP⁺ oxidoreductase (FNR), which has been shown to interact with PsaE in plants (27), but not in cyanobacteria. For both plant and cyanobacterial PSI, flash absorption spectroscopy has revealed three different kinetic phases in the reduction of soluble ferredoxin by PSI, with half times of ca. 500 ns, 20μs, and 100μs (28). The ratio of the three phases shows significant differences between plant and cyanobacterial PSI, indicating that the docking site may have been further optimized during last 1 billion years of independent evolution. The two faster kinetic components do not depend on the concentrations of ferredoxin and PSI, consistent with an electron transfer process occurring within a PSI-ferredoxin complex that was formed before flash excitation. The slow component depends linearly on ferredoxin concentration, suggesting that this kinetic phase reflects the diffusion of ferredoxin to PSI after the flash excitation, followed by electron transfer. (29).

The presence of two fast kinetic components in the electron transfer from PSI to ferredoxin suggests the existence of two distinct ferredoxin-binding conformations. The fastest component would correspond to a tightly bound ferredoxin, whose 2Fe-2S cluster is close to PSI’s distal 4Fe-4S cluster F₈. The intermediate kinetic component is thought to arise from a conformation in which ferredoxin is more loosely bound to the top of the stromal ridge and must settle into the lower conformation before efficient electron transfer can take place.

It has been suggested that plants and cyanobacteria might bind ferredoxin in completely different locations (30), but our structural model does not support this conclusion. Instead, we suggest that in both systems a distal and a proximal docking site may exist with binding constants that differ in plants and cyanobacteria. First, the 4.4 Å structure of plant PSI shows that the stromal subunits are, in their general structural outline, virtually identical to the stromal subunits of cyanobacterial PSI. This high degree of structural similarity suggests a similar binding site for ferredoxin. It is worth noting, however, that a higher-resolution structure may reveal some minor differences between the backbone conformation of the two proteins. Secondly, ferredoxin reduction in plants and cyanobacteria exhibits principally similar kinetic behavior, suggesting a common mechanism. Furthermore, plant ferredoxin is able to act as an electron acceptor for cyanobacterial PSI and vice versa. (29) If the biochemical data on PSI-ferredoxin interactions are taken into account (31), what does our structural model say about the ferredoxin binding site of plant PSI?

Site-directed mutagenesis studies have shown that Thr 15, Gln 16 (32), Lys 35 (33),
and Lys 37 (31) of PsaC, and Arg 68 (34) of PsaE (corresponding to Arg 39 in S. elongatus) are all located within the ferredoxin binding domain with ferredoxin at the position closest to F\(_B\). Glu 121 (35) and Lys 122 (36) of PsaD sit on top of the stromal ridge and have been shown to affect the affinity of ferredoxin for PSI, so they probably participate in the looser, more distal binding site. All of these residues are well-conserved between cyanobacteria and plants.

One region in which the plant and cyanobacterial ferredoxin binding sites differ is the loop that lies between Gly 47 and Gly 54 of PsaE. This loop sits below the lower binding site of ferredoxin; it would be reasonable to suspect that it also participates in ferredoxin binding. This loop region is unique to S. elongatus and may be related to the thermostability of the PSI-Fd interaction. Sequence alignments reveal that it is absent not only in plants and green algae, but also in the cyanobacteria Anabaena sp. (37) and Gloeobacter violaceus (38).

As most of the amino acids that are essential to the proximal docking site are well conserved, differences are found on the proposed distal binding site of ferredoxin at the top of the stromal hump. Differences in this site may be linked to differences in ferredoxin behavior between plants and cyanobacteria: in plants, Fd moves directly to an FNR bound to PsaE, while in cyanobacteria it diffuses away from PSI. The distal binding site may have a much more specific function in plants than in cyanobacteria. Detailed modeling studies of the interaction of ferredoxin with plant PSI to confirm this suggestion are in progress.

Structure and Function of the Unique Plant Subunits – The plant PSI crystal structure contains two of the four subunits that are unique to plants: PsaG and PsaH. We will now discuss their structure and possible function based on the “atomic” model.

Structure and Function of PsaH – PsaH is unique to plants and two functions have been suggested for this subunit: interaction with LHCII and the prevention of trimer formation. Because PsaH is unique to plants, the cyanobacterial crystal structure could not provide us with any clues about proper sequence-structure alignment; for details on the strategy used to find an alignment, see the Supplementary Information. The transmembrane helix prediction server TMHMM (39) predicted a transmembrane helix between residues Leu 51 and Ser 72. When the predicted transmembrane region is aligned with the transmembrane helix assigned in the plant X-ray structural model, chlorophyll 1801 is coordinated by Gln 35, which is conserved among all higher plant species. The plant-specific chlorophyll 1801 might be involved in the excitation energy transfer from LHCII to the core of PSI. It is coordinated in the anti-conformation, and allows an ester group on the porphyrin ring to form a hydrogen bond with Asn 32, which is also very well conserved.

If plant photosystem I contains subunits that are not present in cyanobacterial PSI, then the conserved subunits should exhibit structural and functional differences from their cyanobacterial counterparts. In most cases, these differences are not visible in the 4.4 Å structure, and become clear only when a more detailed structural model is compared with the high-resolution cyanobacterial structure.

The plant model reveals that a subtle modification of Psal allows the binding of the non-conserved subunit PsaH. (Fig. 4b) In cyanobacteria, Psal serves an important role in trimer formation (40), while in plants it interacts strongly with PsaH. Position 12 of cyanobacterial Psal is at the luminal end of its transmembrane helix, where it faces either the lipid bilayer (in PSI monomers) or the PSI-PSI interface (in trimers). It is poorly conserved among cyanobacteria, and is occupied by the bulky side chain tryptophan in S. elongatus. In plant PSI, the corresponding residue, which faces PsaH rather than the lipid bilayer, is a very well conserved serine (Ser 6). Unlike tryptophan, serine is small enough not to interfere with the binding between Psal and PsaH. The remaining contacts between the two subunits are formed by hydrophobic
residues that are conserved between cyanobacteria and plants. PsaH may also interact with the membrane-extrinsic subunits of the PSI core. Plant PsaD has an elongated N-terminus relative to cyanobacterial PsaD, and portions of this domain were assigned in the plant X-ray structural model. Because the electron density was not continuous with the remainder of PsaD, however, the protein backbone is not continuous and we were unable to assign a sequence to this portion of PsaD. Consequently, it was modeled as poly-alanine. It is known that this domain is responsible for the fact that plant stromal subunits are more strongly bound to the reaction center and more resistant to urea treatment than their cyanobacterial counterparts (41). This stabilizing effect is probably a result of the interactions between the N-terminal domain of PsaD and PsaH.

From the 4.4 Å structure of plant PSI, it is clear that the presence of PsaH precludes any trimerization of plant PSI analogous to that seen in cyanobacteria. Structural differences in PsaL may also play a role in the differing oligomerization behavior of the two complexes. Plant PsaL contains a large stromal loop that is much shorter in cyanobacteria, but that was not assigned in the plant X-ray structural model. In assigning a structure for this loop, we chose a conformation that was consistent with the packing of the PSI crystals (i.e. it would not clash with PsaF of the neighboring monomer), and that could interact with PsaH and potentially with the LHCII trimer during state transitions.

A more subtle difference between the two systems is in the residues of PsaL that form the trimerization domain in cyanobacteria. The cyanobacterial crystal structure reveals the hydrophobicity of this region – helix g is lined with non-polar residues that face into the trimerization domain, and the long C-terminus forms extensive hydrophobic interactions with neighboring monomers. In plants, several of these hydrophobic residues are replaced by polar ones, which would fail to promote these hydrophobic interactions. Cross-linking and immunoblotting studies suggest that PsaO interacts with PsaH and PsaL (42); it is possible that some of the non-conserved polar residues on PsaL interact with PsaO or LHCII.

Structure and Function of PsaG – In plants, PsaG and PsaK exhibit considerable sequence similarity and an identical fold, consisting of two transmembrane helices connected by a loop. It has been proposed that PsaG arose via gene duplication of PsaK and subsequent mutation, which allowed it to occupy a different – but symmetry-related – location in the complex (43-45). It has been suggested that the loop region of PsaG might be located in the lumen, unlike that of PsaK, which is located in the stroma (46). The plant crystal structure, however, shows the loop region of PsaG to be stromally-oriented, and our model retains this orientation. PsaG coordinates a single chlorophyll 1701, via Gln 27, which is highly conserved among plant species.

Modeling the LHCI proteins — The modeling of the LHC proteins was even more challenging than the modeling of the core subunits of PSI, because LHCI is less homologous to LHCII than the plant PSI core subunits are to their cyanobacterial counterparts. Furthermore, many structural features – such as the solvent-exposed loops – were missing from the X-ray structural model. Another complication was that the assignment of the three LHC's to Lhca1, Lhca2 and Lhca3 and Lhca4 was only tentative. The modeling strongly supports the assignment of Lhca1 and Lhca4 in the X-ray structure, the results on the assignment of Lhca2 and Lhca3 are less conclusive.

The interactions between Lhca1 and Lhca4 are mainly hydrophobic and confirm the importance of Trp residues for the interaction between Lhca's, which has been suggested by mutagenesis studies (47). Trp 4 of Lhca1 forms π-stacking interactions with Trp 106 of Lhca4, and Trp 185 of Lhca1 interacts with a hydrophobic pocket on the luminal side of Lhca4. Alignment between Lhca2 and Lhca1 reveals that the position containing Trp 185 in Lhca1 is occupied by a glycine in Lhca2. This might partially explain
the apparent weaker interaction between Lhca2 and Lhca3 compared to the Lhca1-Lhca4 heterodimer.

**Interactions of the core proteins of PSI with the LHCI proteins** – Just as the conserved subunits in plant PSI showed structural adaptations to their new role in binding PsaH, the conserved subunits that interact with the LHCI proteins differ from the cyanobacterial proteins in ways that highlight their new roles. PsaK is a particularly dramatic example. With only 31% of its residues conserved between plant and cyanobacterial systems, it is somewhat generous to refer to PsaK as a conserved subunit. While the stromal loop in cyanobacteria is relatively short, the loop region in plant PSI is much longer. Although no location for this loop was assigned in the original plant X-ray structural model, we have assigned it a conformation that passes through a region of weak electron density above Lhca3 which was not assigned in the published crystal structure (48). It is probable that this loop is at least partially responsible for the well-documented interaction between PsaK and Lhca3 (45, 49).

The presence of Lhca3 points to another structural difference between plant and cyanobacterial PSI – the loop region extending from Ser 261 to Thr 267 in plant PsaA is not well-conserved between plants and cyanobacteria, and was not completely resolved in either structure. The conformation we suggest is partially supported by the electron density map from the plant PSI crystal structure, and is the most probable conformation. The loop’s functional importance probably lies in its close proximity to the C-terminus of Lhca3.

Another site for interactions between the PSI core and the light-harvesting complexes is the Lhca4-PsaF interface. PsaF has a fold that is unique among transmembrane proteins – in addition to its single transmembrane helix, there is a V-shaped structure formed by a helix that begins at the stromal side, extends halfway through the membrane, then forms a kink and emerges again on the stromal side. Given its location on the edge of the complex, this V-shaped domain is probably involved in the interactions between PSI and its external light-harvesting complexes (48, 50). It is very remarkable, that this very uncommon structural feature is conserved between plant and cyanobacterial PSI, despite the differences in their external light-harvesting complexes. It has been suggested that PsaF is involved in the docking of phycobilisomes in cyanobacteria (50). PsaF and PsaJ may therefore provide an “entrance gate” for excitation energy in both plants and cyanobacteria. Are there more subtle structural differences between plant and cyanobacterial PsaF that arise from the different functional role played by PsaF in these two complexes?

The V-shaped domain of PsaF may interact with the long stromal loop between residues Lys 122 and Leu 147 in Lhca4, which was mentioned above. Because the conformation of the stromal Lhca4 loop in our model is speculative, details of the interaction are difficult to elucidate. The N-terminal helix of Lhca4 (helix B, following the nomenclature of (51)) also comes close to the V-shaped domain of PsaF, and is visible in the crystal structure, allowing us to reach more definite conclusions about its interactions with PsaF.

This region contains a salt bridge between Arg 142 of PsaF and Glu 35 of Lhca4. These residues are well-conserved among higher plant species, but Arg 142 is replaced by a reasonably well-conserved lysine in cyanobacterial PsaF. In our model, a hydrogen bond is also visible between Asn 146 of PsaF and Glu 35 of Lhca4; these residues are conserved in many, but not all plant species. A more detailed understanding of the interactions between PsaF and Lhca4 awaits a more definitive conformation for the long stromal loop of Lhca4.

At the other end of the LHCI belt, the plant X-ray structural model deviates from the cyanobacterial structure between residues His 308 and Gly 318 of PsaB. The alteration of this loop may assist the binding of Lhca1, e.g. through the formation of a hydrogen bond between Arg 314, which forms part of this loop, and a stromal region on Lhca1.
Non-conserved chlorophylls – The unique functional characteristics of plant PSI arise not only from its protein structure, but also from the precise arrangement of the organic cofactors in its light-harvesting apparatus. The key players in coupling of core antenna system of PS I with the LHC I proteins are the “gap” chlorophylls. Ten chlorophylls were assigned as gap chlorophylls; they are observed in neither the cyanobacterial crystal structure nor the LHCII trimer. These molecules sit in the gap between PSI and the LHCI proteins, where they serve as a functional link by mediating energy transfer between the light-harvesting complexes and the core (48). Modeling of the excitation energy transfer based on our structural model reveals that there are 4 major “highways” for the excitation energy transfer that connect each of the LHCI proteins with the antenna system of the core. This tight coupling mediated by the gap chlorophylls leads to a preferential excitation energy transfer from the LHC chlorophylls to the core chlorophylls that is significantly faster than the excitation energy transfer from between the four LHCI complexes (20). If these chlorophylls are present in plants but not in cyanobacteria, what structural features of plant PSI enable it to bind them?

The non-conserved stromal loop of PsaB was mentioned earlier, in the context of its interactions with Lhca1. This loop has another function – the gap chlorophyll 4001 is coordinated by the carbonyl oxygen of Gly 312. Another change in PsaB with consequences for pigment binding is found at the loop between Ala 491 and Asn 497 of PSI in S. elongatus, which is replaced by a much shorter loop in plant PSI. This is associated with the re-positioning of chlorophyll 1233 in the plant structure, so that it is no longer parallel with chlorophylls 1231 and 1232. This pigment trimer is very strongly coupled in S. elongatus, and these chlorophylls are therefore thought to be among the most red-shifted in cyanobacterial PSI. A change in the orientation of chlorophyll 1233 would disrupt the excitonic coupling of this trimer, and decreases its red shift. Rather than being a systematic difference between plant and cyanobacterial PSI, the presence of this trimer may be a peculiarity of S. elongatus; substantially fewer red-shifted chlorophyll pigments have been spectroscopically observed in Synechocystis (52,53), and it is probable that this trimer is not present in all cyanobacterial species. Sequence comparisons, however, show that this loop seems to be systematically shorter in higher plants than in cyanobacteria.

Another gap chlorophyll, number 4003, is situated close to the lumenal side of PsaF, and is probably coordinated by the peptide carbonyl oxygen of Ser 74 in PsaF. The conformation assigned to this region in the plant X-ray structural model deviates slightly from its cyanobacterial counterpart, and sequence alignments reveal that this region is among those that are less well conserved between plant and cyanobacterial PsaF.

Another category of non-conserved chlorophylls in the plant PSI-LHCI supercomplex is the LHCI-linker chlorophylls. These are located at the interface between neighboring LHCI monomers, and have no counterparts in LHCII. In one case, we were able to identify a protein ligand for an LHCI-linker chlorophyll: chlorophyll 1032 of Lhca4 is clearly coordinated by His 99. An orientation was chosen that causes chlorophyll 1032 to be anti-coordinated, but no conclusive hydrogen-bonding partners seemed available. Coordination of non-conserved chlorophylls was not a criterion in the initial sequence-structure alignment; the proximity of His 99 to chlorophyll 1032 provides additional evidence that this alignment is correct.

The coordination situation is less clear with the other linker chlorophylls. Although widely conserved, His 99 of Lhca4 is replaced by a well-conserved glycine in Lhca3, so chlorophyll 1032 in Lhca3 must be coordinated by a different side chain, a backbone atom, or water.

Despite the varied and often incomplete nature of the information used in the building of this model, a substantial amount of functional information can be gleaned from it; this paper can only provide an overview of some of the most salient features.
Conclusions

The model presented here shows for the first time details of the unique structural and functional features of plant PSI. It provides specific details about protein-protein and protein-pigment interactions, estimated transition dipole moment orientations for antenna system chlorophylls, and some insights into pigment-pigment interactions. It will form the basis for mutagenesis studies that will further explore the interactions between the LHC and the PSI core. Work on the docking of PSI’s soluble partners, plastocyanin and ferredoxin is in progress. The estimated orientations of the chlorophyll transition dipole moments have been used in the theoretical modeling of energy transfer in the plant PSI-LHCI supercomplex (Sener et al., submitted).

At the same time, such a model should be used with caution. Because it is based on a 4.4 Å X-ray structural model, positions of atoms are far from precise, and geometries often non-ideal, so an X-ray structure of the complex with higher resolution may still reveal new, unexpected features of PSI. The computational model presented in this work may provide information on inter-subunit interactions or potential crystal contacts that could prove valuable in obtaining an improved crystal structure of plant PSI and be useful in phase determination of X-ray diffraction data.

Acknowledgements – The authors would like to thank Harsha Vaswani and Ana Damjanovic for help with the simulation of photosynthetic pigments, and Melih Sener for comments and suggestions on chlorophyll placement. CJ was supported by an NSF IGERT fellowship; this work was also supported by NSF grant MCB-0417142 and USDA grant 2003-35318-13573 to PF. NN is supported by a grant from the Israel Science Foundation.

FIGURE LEGENDS

Fig. 1: Plant PSI model. A: Stromal view, perpendicular to the membrane plane. B: View from within the membrane plane. Protein is color-coded by subunit; subunits are colored accordingly. Chlorophyll is shown in green, carotenoids are orange, lipids are pink, iron-sulfur clusters are shown as white and yellow spheres. Subunits conserved in cyanobacteria are shown in a cartoon representation with alpha-helices depicted as cylinders, those unique to plants are shown as thick ribbons. All figures were made using VMD (54) and rendered using Raster3D v2.6c (55).

Fig. 2: Interaction sites of PSI with its soluble electron carriers. A: Lumenal binding site of plastocyanin. The numerous positively-charged residues on the lumenal loop of PsaF (K9, K12, K19, K23, K30, K43, K50, and K58) may form electrostatic interactions with acidic patches on plastocyanin, while the salt bridge formed by R17, R16, and E611 (on PsaB) may help stabilize PsaF. B: Stromal docking site of ferredoxin. The residues depicted have been shown to be involved in one of the ferredoxin binding sites on PSI (see text for further details). The two terminal iron-sulfur clusters F_A and F_B are shown in purple; F_B is the final electron donor to ferredoxin.

Fig. 3: Charge distribution on the electron carrier docking sites of PSI. The surface of the ferredoxin and plastocyanin docking sites are shown, colored by amino acid charge – white for neutral, red for negative, and blue for positive charges. Amino acids whose charge is conserved between plant and cyanobacterial PSI are shown in muted colors; those with non-conserved charges are highlighted. A: Plant ferredoxin docking site. The proposed proximal and distal binding sites are shown by yellow ellipses. B: Cyanobacterial ferredoxin docking site. Note
that the charge distribution in the proximal (tight) binding site is largely conserved, and that most changes appear on top of the stromal hump, at the distal (loose) binding site. C: Plant plastocyanin docking site. The proposed docking site is shown by a yellow ellipse; note that nearly all of the non-conserved residues are part of the luminal loop of PsaF (see Figure 2A). The residues highlighted in yellow are PsaA:W658 and PsaB:W625, which are located immediately below P700 and interact directly with Pc (50). D: Cyanobacterial plastocyanin docking site. The highlighted residues are PsaA:W655 and PsaB:W631. Note the much smaller and decrease positive charge of the luminal loop of PsaF.

**Fig. 4:** A: Lhca1 and Lhca4 with their associated chlorophylls. Chlorophylls whose central ligand is identified in the model are shown in green, others are shown in yellow (see Table S2 in the Supplementary Information). Phytol tails have been omitted for clarity. Amino acids that coordinate chlorophyll have also been shown. B: Interactions between PsaH (green) and PsaI (white). Residues shown in yellow are from *S. elongatus*. Note the bulky non-conserved tryptophan at the luminal side of PsaI, which is replaced by a much smaller serine in plants.

**Table 1:** Summary of the major structural features of the plant PSI model. A more detailed description (especially of plastocyanin and ferredoxin docking) can be found in the text.

| Subunit | Sequence identity | Sequence homology | Conserved features | Non-conserved features |
|---------|-------------------|-------------------|-------------------|-----------------------|
| PsaA    | 78%               | 84%               | • 98% [98%]" of backbone positions conserved | • Lumenal region (S261-T267), which interacts with Lhca3 |
|         |                   |                   | • Positions of electron transfer cofactors (chlorophylls and phylloquinone) | • Lumenal region (G517-V521) |
|         |                   |                   | • Positions of antenna chlorophylls | |
|         |                   |                   | • C581 and C590, ligands to Fe-S cluster | |
|         |                   |                   | • E611, which forms a salt bridge with plastocyanin binding site on PsaF | |
| PsaB    | 76%               | 84%               | • 96% [98%]" of backbone positions conserved | • Stromal region (H308-G318), which interacts with Lhca1 and coordinates gap chl 4001 |
|         |                   |                   | • Positions of chlorophylls in electron transfer chain | • Luminal region (A488-N491), which interactions with Lhca1 |
|         |                   |                   | • Positions of most antenna chlorophylls | • Slight shift in position of phylloquinone in electron transfer chain |
|         |                   |                   | • C559 and C568, ligands to Fe-S cluster | • Inclusion of non-conserved chls 1240, 1241, and 1242 |
|         |                   |                   | • E611, which forms a salt bridge with plastocyanin binding site on PsaF | • Disruption of strongly-coupled chlorophyll trimer 1231-1232-1233, possibly caused by change in A488-N491 |
| PsaC    | 87%               | 94%               | • 100% [100%]" of backbone positions conserved | • K37, which interacts with ferredoxin (31) |
|         |                   |                   | • T15, Q16 and K35, which interact with ferredoxin (32,33) | |
|         |                   |                   | • C11, C14, C17, C21, C48, C51, C54, and C58, ligands to Fe-S clusters | |
| Protein | Cyanobacterial Conserved | Plant Conserved | Comments |
|---------|--------------------------|----------------|----------|
| PsaD    | 62% 75%                  | 99% [98%]     | Long N-terminal domain is absent in cyanobacteria, may interact with PsaH |
|         |                          | of backbone positions conserved | E121 and K122, which interact with ferredoxin (35,36) |
| PsaE    | 59% 73%                  | 87% [98%]     | Deletion of cyanobacterial loop region G47-G54, which interacts with the stromal side of PsaA. This loop is absent in other cyanobacteria and may be involved in ferredoxin docking. |
|         |                          | of backbone positions conserved | R68, which interacts with ferredoxin (34) |
| PsaF    | 47% 62%                  | 86% [77%]     | Lumenal loop region Q20-V46, which docks plastocyanin (22-25), is much longer than in cyanobacteria |
|         |                          | of backbone positions conserved | V-shaped domain P124-N146 |
| PsaI    | 50% 65%                  | 95% [93%]     | Cyanobacterial W12 replaced by Ser to allow binding of PsaH |
|         |                          | of backbone positions conserved | |
| PsaJ    | 54% 70%                  | 95% [93%]     | Removal of Chl 1303 |
|         |                          | of backbone positions conserved | Antenna chls 1301, 1302 |
| PsaK    | 31% 65%                  | 36% [28%]     | Location of N-terminal helix (T7-R23) shifted relative to cyanobacteria |
|         |                          | of backbone positions conserved | Stromal loop (F24-L57), which interacts with Lhca3 (45,48,49) |
|         |                          | Antenna chls 1401, 1402 | Addition of chlorophylls 1403, 1404 |
| PsaL    | 36% 58%                  | 83% [82%]     | Deletion of cyanobacterial C-terminus (L143-N154), which participates in trimerization |
|         |                          | of backbone positions conserved | Stromal loop (K109-L131), which may interact with PsaH and LHCII. |
|         |                          | Antenna chls 1501, 1502, 1503 | Addition of chlorophyll 1504 |

*a: The first number in each pair is the percentage of cyanobacterial residues whose position is identical in the plant PSI model; the second (in square brackets) is the percentage of plant residues whose position is identical in cyanobacteria. Differences in these numbers stem from insertions or deletions in the sequence, e.g. in PsaE or PsaF.
Figure 2:
Figure 4:
REFERENCES

1. Nelson, N., and Ben-Shem, A. (2004) *Nature Reviews Molecular Cell Biology* In press
2. Ben-Shem, A., Frolow, F., and Nelson, N. (2004) *FEBS Lett* 564, 274-280
3. Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W., and Krauss, N. (2001) *Nature* 411, 909-917
4. Bibby, T. S., Nield, J., and Barber, J. (2001) *J Biol Chem* 276, 43246-43252
5. Bibby, T. S., Nield, J., and Barber, J. (2001) *Nature* 412, 743-745
6. Boekema, E. J., Hifney, A., Yakushevska, A. E., Piotrowski, M., Keegstra, W., Berry, S., Michel, K. P., Pistorius, E. K., and Kruip, J. (2001) *Nature* 412, 745-748
7. Ben-Shem, A., Frolow, F., and Nelson, N. (2003) *Nature* 426, 630-635
8. Fromme, P., Jordan, P., and Krauss, N. (2001) *Biochim Biophys Acta* 1507, 5-31
9. Liu, Z., Yan, H., Wang, K., Kuang, T., Zhang, J., Gui, L., An, X., and Chang, W. (2004) *Nature* 428, 287-292
10. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O’Donovan, C., Phan, I., Pilbout, S., and Schneider, M. (2003) *Nucleic Acids Res* 31, 365-370
11. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) *Nucleic Acids Res* 31, 3381-3385
12. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* 18, 2714-2723
13. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) *Annu Rev Biophys Biomol Struct* 29, 291-325
14. Henriques, E. F., Ramos, M. J., and Reynolds, C. A. (1997) *Journal of Computer-Aided Molecular Design* 11, 547-556
15. Foloppe, N., Ferrand, M., Breton, J., and Smith, J. C. (1995) *Proteins* 22, 226-244
16. Pali, T., Garab, G., Horvath, L. I., and Kota, Z. (2003) *Cell Mol Life Sci* 60, 1591-1606
17. Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. (1999) *Journal of Computational Physics* 151, 283-312
18. Damjanovic, A., Kosztin, I., Kleinekathofer, U., and Schulten, K. (2002) *Physical Review E* 65, -
19. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wierkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) *Journal of Physical Chemistry B* 102, 3586-3616
20. Sener, M. K., Jolley, C., Ben-Shem, A., Fromme, P., Nelson, N., Croce, R., and Schulten, K. (2005) *Biophys J* Submitted
21. Hervas, M., Navarro, J. A., Diaz, A., Bottin, H., and De la Rosa, M. A. (1995) *Biochemistry* 34, 11321-11326
22. Hippler, M., Drepper, F., Rochaix, J. D., and Muhlenhoff, U. (1999) *J Biol Chem* 274, 4180-4188
23. Hippler, M., Drepper, F., Haehnel, W., and Rochaix, J. D. (1998) Proc Natl Acad Sci U S A 95, 7339-7344
24. Olesen, K., Ejdeback, M., Crnogorac, M. M., Kostic, N. M., and Hansson, O. (1999) Biochemistry 38, 16695-16705
25. Young, S., Sigfridsson, K., Olesen, K., and Hansson, O. (1997) Biochim Biophys Acta 1322, 106-114
26. Sommer, F., Drepper, F., and Hippler, M. (2002) J Biol Chem 277, 6573-6581
27. Andersen, B., Scheller, H. V., and Moller, B. L. (1992) Febs Letters 311, 169-173
28. Setif, P. Q., and Bottin, H. (1995) Biochemistry 34, 9059-9070
29. Setif, P. (2001) Biochim Biophys Acta 1507, 161-179
30. Ruffle, S. V., Mustafa, A. O., Kitmitto, A., Holzenburg, A., and Ford, R. C. (2000) J Biol Chem 275, 36250-36255
31. Setif, P., Fischer, N., Lagoutte, B., Bottin, H., and Rochaix, J. D. (2002) Biochim Biophys Acta 1555, 204-209
32. Fischer, N., Setif, P., and Rochaix, J. D. (1999) J Biol Chem 274, 23333-23340
33. Fischer, N., Hippler, M., Setif, P., Jacquot, J. P., and Rochaix, J. D. (1998) Embo J 17, 849-858
34. Barth, P., Guillouard, I., Setif, P., and Lagoutte, B. (2000) J Biol Chem 275, 7030-7036
35. Bottin, H., Hanley, J., and Lagoutte, B. (2001) Biochem Biophys Res Commun 287, 833-836
36. Hanley, J., Setif, P., Bottin, H., and Lagoutte, B. (1996) Biochemistry 35, 8563-8571
37. Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M., and Tabata, S. (2001) DNA Res 8, 205-213; 227-253
38. Nakamura, Y., Kaneko, T., Sato, S., Mimuro, M., Miyashita, H., Tsuchiya, T., Sasamoto, S., Watanabe, A., Kawashima, K., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N., Shimpo, S., Takeuchi, C., Yamada, M., and Tabata, S. (2003) DNA Res 10, 137-145
39. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) J Mol Biol 305, 567-580
40. Xu, Q., Hoppe, D., Chitnis, V. P., Odom, W. R., Guikema, J. A., and Chitnis, P. R. (1995) J Biol Chem 270, 16243-16250
41. Naver, H., Scott, M. P., Andersen, B., Moller, B. L., and Scheller, H. V. (1995) Physiologia Plantarum 95, 19-26
42. Jensen, P. E., Haldrup, A., Zhang, S., and Scheller, H. V. (2004) J Biol Chem 279, 24212-24217
43. Kjaerulff, S., Andersen, B., Nielsen, V. S., Moller, B. L., and Okkels, J. S. (1993) J Biol Chem 268, 18912-18916
44. Okkels, J. S., Nielsen, V. S., Scheller, H. V., and Moller, B. L. (1992) Plant Mol Biol 18, 989-994
45. Jansson, S., Andersen, B., and Scheller, H. V. (1996) Plant Physiology 112, 409-420
46. Jensen, P. E., Haldrup, A., Rosgaard, L., and Scheller, H. V. (2003) Physiologia Plantarum 119, 313-321
47. Schmid, V. H., Paulsen, H., and Rupprecht, J. (2002) Biochemistry 41, 9126-9131
48. Ben-Shem, A., Frolow, F., and Nelson, N. (2004) Photosynthesis Research 81, 239-250
49. Jensen, P. E., Gilpin, M., Knoetzel, J., and Scheller, H. V. (2000) Journal of Biological Chemistry 275, 24701-24708
50. Fromme, P., Melkozernov, A., Jordan, P., and Krauss, N. (2003) FEBS Lett 555, 40-44
51. Kuhlbrandt, W., Wang, D. N., and Fujiyoshi, Y. (1994) Nature 367, 614-621
52. Gobets, B., Vanamerongen, H., Monshouwer, R., Kruip, J., Rogner, M., Vangronnelle, R., and Dekker, J. P. (1994) Biochimica Et Biophysica Acta-Bioenergetics 1188, 75-85
53. Palsson, L. O., Flemming, C., Gobets, B., van Grondelle, R., Dekker, J. P., and Schlodder, E. (1998) Biophys J 74, 2611-2622
54. Humphrey, W., Dalke, A., and Schulten, K. (1996) J Mol Graph 14, 33-38, 27-38
55. Merritt, E. A., and Bacon, D. J. (1997) Macromolecular Crystallography, Pt B 277, 505-524
Supplementary information

Detailed Description of Materials and Methods:

The computational model was based on two structures: the 4.4 Å structural model from pea PSI-LHCI (1) and the 2.5 Å structure of cyanobacterial PSI (2). Sequence data for the plant PSI subunits was obtained from the Swiss-Prot/TrEMBL sequence database (3). For subunits that had been sequenced in pea, the available pea sequences were used for modeling. For some other subunits, only N-terminal sequence fragments have been published for pea. In these cases, an NCBI BLAST (4) search was performed, and the sequence chosen which was most similar to the pea sequence fragment. For PsaJ, PsaL, and Lhca1, published sequence data was insufficient to perform such a search and the sequences from Arabidopsis were used (see Table 2). As a very high degree of sequence homology exists between PSI subunit sequences in all higher plants, it is largely unimportant exactly which sequences are used.

For regions in which the backbone conformation assigned in the plant X-ray structural model matches that in the cyanobacterial structure, the cyanobacterial crystal structure was used as a template for homology modeling. The target sequences were aligned onto the templates using the default settings in the ClustalW program (5). Subunits PsaG and PsaH, as well as the light-harvesting complexes, are unique to plants and could not be aligned using this method. For these subunits, as well as regions in the conserved subunits whose backbone conformation in the plant structure (PDB accession code 1QZV) diverged from cyanobacteria (PDB accession code 1JB0) – such as the large luminal loop on PsaF and a stromal loop on PsaB – a poly-alanine chain based on the plant X-ray structural model served as a template. In Lhca1, the N- and C-termini from the LHCII crystal structure were added to this poly-alanine template. Because the 4.4 Å plant crystal structural model did not contain any sidechains, a structural template based solely on this structure could contain no sequence information upon which to base an alignment. Other criteria needed to be chosen.

In these cases, secondary structure prediction often proved to be a valuable first step. The TMHMM server (6) was used with default settings to predict the location of transmembrane helices. Although the predicted helices were approximately the same length as those shown in the crystal structure, secondary-structure prediction alone was not enough to provide an unequivocal sequence-structure alignment. Chlorophyll coordination criteria were very important in narrowing this rough alignment down to a more reliable one: only a nucleophilic atom with a free electron pair can coordinate the central magnesium atom in chlorophyll. If a particular chlorophyll is positioned to be coordinated by an amino acid sidechain, this provides substantial constraints on which sidechains are valid magnesium ligands. We further required that chlorophyll-coordinating residues be highly conserved between the PSI protein sequences for various higher plants. With these constraints in place, we were able to establish unique alignments for each subunit.

Homology models of each subunit were constructed individually, using a combination of the MOE software package (Chemical Computing Group, Inc.), the Swiss-Model online modeling server (7), and the Swiss-PDB viewer (8). The alignments used in modeling were adjusted manually, for the reasons described above. In several cases, especially in the light-harvesting complexes, large loop regions were not assigned in the plant X-ray structural model. The loop conformations suggested by the modeling programs usually required manual adjustment, because they clashed with adjacent subunits, cofactors, or the membrane plane. The presence of these elements provides constraints on possible loop conformations, making them somewhat less arbitrary.

The modeling of Lhca1 and Lhca4 was mentioned above; the biochemical evidence for the identification of the other two extrinsic light-harvesting complexes in the PSI-LHCI
supercomplex as Lhca2 and Lhca3 (9) was not conclusive enough to allow us to model them in detail at this stage. Instead, we have chosen to construct a poly-alanine model based on the alpha-carbon coordinates assigned in the plant X-ray structural model.

The modeling of cofactors presented another challenge. Non-protein cofactors make up more than 30% of the total mass of plant PSI, yet traditional homology modeling approaches consider only polypeptides (10). Our approach to modeling cofactors is similar to one that has been described for buried water (11), but to our knowledge this is the first application of this technique to organic cofactors.

Initial coordinates for most cofactors in the core of the complex were obtained from the cyanobacterial PSI structure (PDB accession code 1JB0). In general, chlorophyll positions assigned in the plant X-ray structural model were remarkably similar to those in the cyanobacterial structure, and it is reasonable to assume that, when position is conserved, orientation will be conserved also. A more exact fit to the data provided in the plant X-ray structural model was achieved by positioning each chlorophyll molecule such that the four nitrogen atoms in its macrocyclic head group aligned perfectly with the positions assigned for nitrogen atoms in the plant X-ray structural model. These nitrogen atoms were then used to provide constraints during minimization, forcing the rest of the molecule to fall into line. Because the protein environment of cofactors differs between cyanobacteria and plants, extensive minimization was needed to optimize the cofactor conformations relative to their environment, especially in the flexible phytyl tail.

The cofactors in the light-harvesting complexes were similarly obtained from the crystal structure for LHCII, while the assignments of chlorophyll \( a \) and chlorophyll \( b \) follow the tentative assignments made in (12). Although LHCl and LHCII are closely related, their structures are not as similar as are the PSI core subunits in plants and cyanobacteria, given their 25-35% sequence identity. To assign orientations to the chlorophyll molecules in the LHCl proteins, the LHCl crystal structure was aligned with a monomer of LHCII, and the LHCl chlorophylls were given an orientation as close as possible to that of their counterpart in LHCII. The proper alignment was usually obvious, but in some cases the alignment chosen was less certain. The protein environment could not provide constraints on the phytyl tail conformations in Lhca2 and Lhca3, because these subunits were modeled only with poly-alanine; these tails were therefore omitted. In several cases, steric clashes and stereochemistry errors were corrected by manually moving atoms in VMD (13), followed by minimization.

Some chlorophyll molecules, in particular the gap chlorophylls 4001-4010, the linker chlorophyll 1031-1033 in the LHCl proteins, and the core chlorophylls 1240-1242, 1403-1404, 1504, 1701, and 1801 are unique to plant PSI and lack homologues in cyanobacterial PSI or plant LHCII. Although cofactors are not generally visible at 4.4 Å resolution, the macrocyclic head groups of the chlorophyll molecules were visible, due to a high electron density resulting from their conjugated pi-electron system. This tells us the chlorophyll molecule’s position and the plane in which this macrocycle lies, but gives us too little information to specify a unique orientation for the chlorophyll head group within this plane. The problem is analogous to that in the non-conserved protein subunits – the crystal structure supplies us with incomplete information, from which we must make inferences.

The first step in making these inferences was to examine the protein environment surrounding the position observed in the plant X-ray structural model for possible hydrogen-bond donors for the polar groups on the porphyrin ring, and hydrophobic pockets for the phytyl tail. Because these chlorophyll molecules are not conserved between cyanobacteria and plants, we suspected that they would interact with amino acid residues that are conserved among different plant species but not conserved between plants and cyanobacteria. We found that, in many cases, a pocket of non-conserved hydrophobic residues was located near these non-conserved chlorophylls and provided an ideal location for the chlorophyll phytyl tail.
Another clue in assigning orientations to chlorophyll molecules came from examining the preferred stereochemistry of chlorophyll-ligand complexes (14,15). It has been found, through analysis of crystal structures and ab initio modeling, that chlorophyll-ligand complexes are more stable when the chlorophyll ligand lies on the same side of the macrocyclic plane as the C13(2)-(R) methoxycarbonyl moiety. The nomenclature in this case is not yet well-established; we will follow what seems to be a sensible convention (16) in referring to this as the anti-conformation of the chlorophyll-ligand complex. This criterion, when considered together with the local protein environment of the chlorophyll molecule, allowed us to make reasonable assumptions about chlorophyll orientations.

Due to its limited resolution, no beta-carotene positions were assigned in the plant X-ray structural model. However, we found that most of the beta-carotenes in cyanobacteria were located in pockets of hydrophobic residues that are highly-conserved between cyanobacteria and plants. When the initial homology models of the plant proteins were constructed, the geometry of the binding pockets was largely retained. The carotenoids fit well, and minimization helped further improve their interactions with their protein environment. The double-bond orientations of the carotenoids were assumed to be conserved between cyanobacteria and plants. We strongly suspect that more carotenoids are present in plant PSI than we have included in our model, (especially in the interface between PSI and the LHCl proteins) but their position cannot be determined without a higher-resolution structure. The lutein molecules in the LHCl proteins were placed using similar means, with starting positions drawn from the LHClII structure.

Other cofactors were also included in the plant PSI model. Three phospholipids and one galactolipid were included in the core complex, along with one phospholipid each of the light-harvesting complexes. Two phylloquinones and three iron-sulfur clusters were placed at the same positions in which they were assigned in the plant X-ray structural model, as part of the electron transport chain.

All minimizations were performed using a steepest-descent algorithm in NAMD 2.5 (17). Because the force field used (CHARMM27) lacks parameters for many of the non-protein cofactors present in plant PSI, parameter files needed to be developed, based on a combination of ab initio calculations (Vaswani and Damjanović, personal communication, (18)) and estimates based on known parameters for chemically similar compounds (19).

Local, limited minimizations were conducted throughout the modeling process to eliminate bad contacts, move modeled loops into feasible conformations, and improve the stereochemical quality of the structure. During minimization steps, protein alpha-carbon and chlorophyll nitrogen atoms were constrained to the positions assigned in the plant X-ray structural model, in order to preserve secondary structure and chlorophyll orientation.

Once the model had been constructed in vacuum, we placed it in a solvated bilayer of digalactosyl-diacylglyceride (DGDG), which had been equilibrated at constant pressure in a molecular dynamics simulation. DGDG was chosen because it is a major lipid component of thylakoid membranes (20). The final system contained the modeled proteins and cofactors, 529 membrane lipid molecules, about 86,000 water molecules, and one sodium ion, needed for electrical neutrality. In order to maintain the secondary structure present in the starting structure, the protein alpha-carbon and chlorophyll nitrogen atoms were constrained to the positions assigned in the plant X-ray structural model, in order to preserve secondary structure and chlorophyll orientation.

Once the model had been constructed in vacuum, we placed it in a solvated bilayer of digalactosyl-diacylglyceride (DGDG), which had been equilibrated at constant pressure in a molecular dynamics simulation. DGDG was chosen because it is a major lipid component of thylakoid membranes (20). The final system contained the modeled proteins and cofactors, 529 membrane lipid molecules, about 86,000 water molecules, and one sodium ion, needed for electrical neutrality. In order to maintain the secondary structure present in the starting structure, the protein alpha-carbon and chlorophyll nitrogen atoms were constrained to the positions assigned in the plant PSI X-ray structural model were bound to their starting positions by a harmonic potential (k = 100 kcal mol⁻¹ Å⁻²), while all other atoms remained free to move, including the alpha-carbons of those regions unassigned in the X-ray model. In order to assess the progress of the MD simulation, the positions of these “flexible” alpha-carbon atoms were compared to their positions in the starting structure, and a global RMSD was calculated.

Once the system had been assembled, it was simulated at constant pressure at 300 K, with the protein and associated cofactors held fixed, so that the lipid and solvent could equilibrate. Once the volume remained stable, the atoms of the PSI model were released and could move freely, except for the restraints described above. Once a new stable volume was reached, those
residues more than 15 Å away from the protein were held fixed, in order to decrease the simulation time. The simulation was allowed to continue until the global RMSD of the loop regions reached a stable value, indicating that the system had become conformationally stable. At this point, the system was repeatedly minimized (at T = 0) and gradually warmed back up to 300K, in order to find the best possible global minimum. The final simulation step was a simulated annealing process, in which the temperature of the system was gradually lowered from 300 to 0 K.

The information used in modeling came from a variety of sources, and more information was available for some subunits than for others. Consequently, different portions of the model should be treated with varying degrees of skepticism. To provide a guide to the relative trustworthiness of different regions in the model, uncertainty scores – which we refer to as “B-factors” – were assigned to each atom in the structure. These should not be confused with the B-factors that are obtained during the refinement of high-resolution X-ray structures. This term is used merely to reflect potential uncertainty about the position of atoms, and to include this information in the PDB file. Note also that numbers only reflect uncertainty coming from the structure itself, not from the heterogeneous set of sequences used in modeling. For more details, see Table S1 and Figure S1.
Fig. S1: Stromal view of PSI model, with atoms colored by uncertainty score. Atoms whose location is reliable are colored in blue, those whose location is highly speculative are red, and intermediate cases are shown in white. Orientation of the molecule is the same as in Figure 1a.

Table 1: Uncertainty scores for different portions of the plant PSI model. Lower scores indicate regions of high confidence; atoms whose positions are highly speculative were assigned higher scores.

| Uncertainty score | Atoms assigned this score |
|-------------------|---------------------------|
| 0                 | Atoms whose positions are assigned in the plant X-ray structural model, i.e. protein alpha-carbons, iron-sulfur clusters, and the conjugated ring structures of chlorophyll and phylloquinone. |
| 1                 | Sidechains in regions conserved between plants and cyanobacteria |
|                   | Alpha-carbon atoms from weak electron densities |
|                   | Chlorophyll head groups in core |
| 2                 | Sidechains in PsaH |
|                   | Side chains in most experimentally-observed regions of light-harvesting complexes. |
|                   | Chlorophyll phytyl tails in core |
| 3                 | Sidechains in areas where alignment was not entirely certain (i.e. transmembrane helices of PsaG, helix C of Lhca4, C-terminal helix of PsaK) |
|                   | Short reconstructed loops (four residues or fewer) |
|                   | Head groups of chlorophylls conserved between LHCI and LHCII |
| 4                 | C- and N-terminal loops of Lhca1 |
|                   | Head groups of chlorophylls with guessed orientations |
|                   | Non-chlorophyll cofactors in the light-harvesting complexes |
|                   | Phytyl tails of chlorophylls conserved between LHCI and LHCII |
| 5                 | Long reconstructed loops (more then four residues) |
|                   | Phytyl tails of chlorophylls with guessed orientations |
|                   | Water molecules |

Table 2: Potential chlorophyll binding partners in Lhca1 and Lhca4. Chlorophyll-ligand interactions not seen in LHCII are shown in boldface. All magnesium ligands are well-conserved among plant species.

| Chlorophyll number | Ligand in Lhca1 | H-bonding partners in Lhca1 | Ligand in Lhca4 | H-bonding partners in Lhca4 |
|--------------------|------------------|-----------------------------|-----------------|-----------------------------|
| 1011               | E147             | R48                         | E153            | R49, S114                   |
| 1012               | N150             | K146                        | N156            | K152                        |
| 1013               | Q164             |                              | Q170            | H185                        |
| 1014               | E43              | Y39, K40                    | E44             | R142 (PsaF)                 |
| 1015               | H46              | Y39, S42                    | N47             | Y141 (PsaF)                 |
| 1016               | None             | T83, L85                    |                 |                             |
| 1017               | **LUT 5621**     | A71                         | **Water 5001**  |                             |
| 1021               | None             |                              |                 |                             |
| 1022               | LHG 5630         | N150                        | LHG 5630        | W50, R104, R105             |
| 1023               | H186             |                              | H185            | K152                        |
| 1025               | E109             | K122, Y133                  | E102            | W106                        |
| 1026               | **E101**         |                              | **E94**         | K74                         |
Table 3: Sequences used in the modeling of plant PSI. In the cases where the sequence from *Pisum sativum* (garden pea) were complete, these sequences were used. If the pea sequence was only fragmentary, the complete sequence was chosen which was most similar to the pea sequence, based on a BLAST search. Where no pea sequence data was available, sequences from *A. thaliana* were used.

| Subunit | Accession code in pea | Accession code used in model | Organism | References |
|---------|-----------------------|-----------------------------|----------|------------|
| PsaA    | P05310                | P05310                      | *Pisum sativum* | (21)       |
| PsaB    | P05311                | P05311                      | *Pisum sativum* | (21)       |
| PsaC    | P10793                | P10793                      | *Pisum sativum* | (22)       |
| PsaD    | P20117                | P29302                      | *Nicotiana sylvestris* | (23,24) |
| PsaE    | P20118                | P13194                      | *Hordeum vulgare* | (23,25) |
| PsaF    | P20119                | Q9XQB4                      | *Phaseolus aureus* | (23)       |
| PsaG    | P20120                | P12357                      | *Spinacia oleracia* | (23,26) |
| PsaH    | P20121                | O65101                      | *Zea mays* | (23)       |
| PsaI    | P17227                | P17227                      | *Pisum sativum* | (27,28) |
| PsaJ    | P17229                | P56769                      | *Arabidopsis thaliana* | (29,30) |
| PsaK    | P17226                | Q9ZT05                      | *Medicago sativa* | (29)       |
| PsaL    | -                     | Q9SU14                      | *Arabidopsis thaliana* | (31)       |
| Lhca1   | -                     | Q01667                      | *Arabidopsis thaliana* | (32)       |
| Lhca4   | Q9SQL2                | Q9SQL2                      | *Pisum sativum* | (33)       |
References:

1. Ben-Shem, A., Frolow, F., and Nelson, N. (2003) Nature 426, 630-635
2. Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W., and Krauss, N. (2001) Nature 411, 909-917
3. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O’Donovan, C., Phan, I., Piblout, S., and Schneider, M. (2003) Nucleic Acids Res 31, 365-370
4. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res 25, 3389-3402
5. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res 22, 4673-4680
6. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) J Mol Biol 305, 567-580
7. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) Nucleic Acids Res 31, 3381-3385
8. Guex, N., and Peitsch, M. C. (1997) Electrophoresis 18, 2714-2723
9. Jensen, P. E., Gilpin, M., Knoetzel, J., and Scheller, H. V. (2000) Journal of Biological Chemistry 275, 24701-24708
10. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) Annu Rev Biophys Biomol Struct 29, 291-325
11. Henriques, E. F., Ramos, M. J., and Reynolds, C. A. (1997) Journal of Computer-Aided Molecular Design 11, 547-556
12. Sener, M. K., Jolley, C., Ben-Shem, A., Fromme, P., Nelson, N., Croce, R., and Schulten, K. (2005) Biophys J Submitted
13. Humphrey, W., Dalke, A., and Schulten, K. (1996) J Mol Graph 14, 33-38, 27-38
14. Oba, T., and Tamiaki, H. (2002) Photosynthesis Research 74, 1-10
15. Balaban, T. S., Fromme, P., Holzwarth, A. R., Krauss, N., and Prokhorenko, V. I. (2002) Biochim Biophys Acta 1556, 197-207
16. Balaban, T. S., Holzwarth, A. R., and Schaffner, K. (1995) Journal of Molecular Structure 349, 183-186
17. Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. (1999) Journal of Computational Physics 151, 283-312
18. Damjanovic, A., Kosztin, I., Kleinekathofer, U., and Schulten, K. (2002) Physical Review E 65, -
19. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Workiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) Journal of Physical Chemistry B 102, 3586-3616
20. Pali, T., Garab, G., Horvath, L. I., and Kota, Z. (2003) Cell Mol Life Sci 60, 1591-1606
21. Lehmann, J., Stummann, B. M., and Henningsen, K. W. (1987) Nucleic Acids Res 15, 3630
22. Dunn, P. P. J., and Gray, J. C. (1988) Plant Molecular Biology 11, 311-319
23. Dunn, P. P., Packman, L. C., Pappin, D., and Gray, J. C. (1988) FEBS Lett 228, 157-161
24. Yamamoto, Y., Tsuji, H., Hayashida, N., Inoue, K., and Obokata, J. (1991) Plant Mol Biol 17, 1251-1254
25. Okkels, J. S., Jepsen, L. B., Honberg, L. S., Lehmbeck, J., Scheller, H. V., Brandt, P., Hoyer-Hansen, G., Stummann, B., Henningsen, K. W., and von Wettstein, D. (1988) FEBS Lett 237, 108-112
26. Steppuhn, J., Hermans, J., Nechushtai, R., Ljungberg, U., Thummler, F., Lottspeich, F., and Herrmann, R. G. (1988) FEBS Lett 237, 218-224
27. Smith, A. G., Wilson, R. M., Kaethner, T. M., Willey, D. L., and Gray, J. C. (1991) Curr Genet 19, 403-410
28. Nagano, Y., Matsuno, R., and Sasaki, Y. (1991) Curr Genet 20, 431-436
29. Ikeuchi, M., Hirano, A., Hiyama, T., and Inoue, Y. (1990) FEBS Lett 263, 274-278
30. Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999) DNA Res 6, 283-290
31. Mayer, K., Schuller, C., Wambutt, R., Murphy, G., Volckaert, G., Pohl, T., Dusterhoft, A., Stiekema, W., Entian, K. D., Terryn, N., Harris, B., Ansorge, W., Brandt, P., Grivell, L., Rieger, M., Weichselgartner, M., de Simone, V., Obermaier, B., Mache, R., Muller, M., Kreis, M., Delseny, M., Puigdomenech, P., Watson, M., and McCombie, W. R. (1999) Nature 402, 769-777
32. Jensen, P. E., Kristensen, M., Hoff, T., Lehmbeck, J., Stummann, B. M., and Henningsen, K. W. (1992) Physiologia Plantarum 84, 561-567
33. Brosche, M., Fant, C., Bergkvist, S. W., Strid, H., Svensk, A., Olsson, O., and Strid, A. (1999) Biochim Biophys Acta 1447, 185-198
Structure of plant photosystem I revealed by theoretical modeling
Craig C. Jolley, Adam Ben-Shem, Nathan Nelson and Petra Fromme

J. Biol. Chem. published online June 14, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M500937200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts