

# Feeling the Strain: Quantifying ligand deformation in photosynthesis

Chientzu Lin,<sup>†</sup> Yuval Mazor,<sup>‡</sup> and Mike Reppert<sup>\*,†</sup>

<sup>†</sup>*Department of Chemistry, Purdue University, West Lafayette, IN 47920, USA*

<sup>‡</sup>*School of Molecular Sciences, The biodesign institute, Arizona State University, Tempe, AZ 85281, USA*

E-mail: reppertm@purdue.edu

## Abstract

Structural distortion of protein-bound ligands can play a critical role in enzyme function by tuning the electronic and chemical properties of the ligand molecule. However, quantifying these effects is difficult due to the limited resolution of protein structures and the difficulty of generating accurate structural restraints for non-protein ligands. Here, we seek to quantify these effects through a statistical analysis of ligand distortion in Chlorophyll (Chl) proteins (CP), where ring deformation is thought to play a role in energy and electron transfer. To assess the accuracy of ring-deformation estimates from available structural data, we take advantage of the C<sub>2</sub> symmetry of Photosystem II (PSII), comparing ring-deformation estimates for equivalent sites both within and between 113 distinct X-ray and Cryogenic electron microscopy (CryoEM) PSII structures. Significantly, we find that several deformation modes exhibit considerable variability in predictions, even for equivalent monomers, down to 2 Å resolution, to an extent that probably prevents their utilization in optical calculations. We further find that refinement restraints play a critical role in determining deformation values to

resolution as low as 2 Å. However, for those modes that are well-resolved in the structural data, ring deformation in PSII is strongly conserved across all species tested, from cyanobacteria to algae. These results highlight both the opportunities and limitations inherent in the structure-based analysis of the bioenergetic and optical properties of CPs and other protein-ligand complexes.

## Introduction

Non-protein ligands play a central role in the function of many biological enzymes, catalyzing chemical and photophysical processes that cannot be achieved by the 20 natural amino acids directly. As a result, the characterization of protein-ligand interaction and binding is critical both for the mechanistic understanding of native protein function and for the design of new ligands and ligand-protein complexes in emerging fields such as computer-aided drug discovery,<sup>1-3</sup> protein engineering<sup>4,5</sup> and *de novo* enzyme design.<sup>6,7</sup>

While ligand-induced changes to protein structure are widely studied (often understood via the “induced fit” or “conformational selection” mechanisms<sup>3,8-11</sup>), the forces that drive ligand binding are equally capable of distorting the structure of the ligand itself. These distortions can play a mechanistic role in protein function, tuning the electronic and vibrational properties of the ligand to facilitate specific physical or chemical purposes.<sup>12-19</sup>

Unfortunately, quantifying the energetics of such distortion is complicated by the fact that (in contrast to the amino acid residues that comprise the protein scaffold), relatively few solved structures or validated molecular-mechanics forces fields exist for most ligands. Most structures solved in structural biology are not determined to atomic resolution. This implies that the resulting structural models, which are specified as atomic positions, rely heavily on prior chemical knowledge, implemented as model restraints in various refinement software.<sup>20-22</sup> For the twenty natural amino acids, a set of highly curated restraints were generated from atomic resolution structures.<sup>23-25</sup> For some ligands, reasonable chemical restraints may be available, but in many cases these restraints only recently became available.

In addition, many researchers generate their own restrains using the computational method available to them. Generating correct restrains for metal coordinating molecules (which are common enzyme cofactors) is difficult and constitutes its own, yet unresolved, challenge.<sup>26,27</sup> In addition, and in contrast to the twenty natural amino acids, high quality ligand validation in deposited structures is problematic and still under development.

A notable success in characterizing the functional significance of ligand deformation comes from the heme group found in oxygen-transport proteins such as hemoglobin and myoglobin and in redox-active proteins such as cytochromes. Building on a rich literature on porphyrin structure and spectroscopy,<sup>28</sup> the large number of available crystal structures for heme proteins, and their easily measured spectroscopic properties, Shelnutz and coworkers established a biologically conserved relationship between structural distortion of protein-bound heme groups and their electronic properties in various proteins.<sup>12,29,30</sup> More recently, heme deformation has been extensively characterized via a variety of spectroscopy techniques,<sup>13,31–37</sup> and the specific impact of heme deformation on functional metrics such as O<sub>2</sub> binding affinity<sup>38</sup> and electron transfer rates<sup>39</sup> have been investigated in detail. A variety of computational tools (including an online server<sup>40</sup>) has likewise been developed for automated heme deformation analysis.<sup>41–44</sup>

This success in characterizing heme distortion has prompted interest in the functional deformation of related ligands, particularly the Chlorophyll (Chl) and Bacteriochlorophyll (BChl) pigments that drive biological photosynthesis. Distortion of the (B)Chl macrocycle has long been suspected of playing a functional role in photosynthesis by tuning electron-transfer properties and “site energies” (local electronic transition frequencies) that, along with inter-pigment excitonic interactions, determine the optical absorption spectra of light-harvesting proteins.<sup>12,19,28,45–56</sup> For example, ultra-fast electron density differences around the special pair Bchl’s in the purple bacteria reaction center were observed. These were suggested to emanate from chlorophyll ring structural changes (among others) that contribute to the high efficiency of charge separation in this system.<sup>57</sup> However, structure based calculations

did not identify the expected changes in the redox potential of the Bchl’s in the reaction center.<sup>56</sup> These discrepancies may be attributed to the difficulty in capturing the observed map differences in structural models.

To quantify ring deformation in (B)Chl systems, most studies employ the normal-mode structural decomposition (NSD) approach developed by Shelnutz and coworkers for heme proteins.<sup>29,41</sup> Briefly, the NSD method projects the observed deformation of individual pigments from an experimental structure onto a basis of six harmonic normal modes: saddle (“sad”), ruffle (“ruf”), dome (“dom”), wave (“wav(x)” and “wav(y)” depending on the wave axis), and propeller (“pro”).<sup>41</sup> Relative to the “strain energy” often used to quantify ligand deformation,<sup>58,59</sup> the NSD approach is simpler and more physically transparent but less accurate in terms of computed energies. Whereas ligand deformation is often computed with high-level quantum-mechanical methods able to quantify anharmonic effects and ligand-specific chemistry, the NSD approach computes deformation energies for a variety of different porphyrin-like molecules (hemes, Chls, BChls, etc.) using a single set of harmonic normal modes and force constants computed for a standard (Cu(II) porphine) reference. The advantages of this approach are computational speed and, more importantly, the conceptual utility of having a single, transferable basis with which to quantify deformation in porphyrin-like systems. The recent observation that ligand strain energy trends are often quite well conserved across varying levels of theory<sup>59</sup> supports the idea that simple models like NSD analysis can offer useful physical insight, despite their simplicity.

A particularly detailed NSD study on the Photosystem II (PSII) complex has been reported by Saito et al., which examined both the nature and environmental origins of observed Chl ring deformations in the 1.9 Å crystal structure of Umena et al.<sup>60</sup> Although large deformations were observed for several PSII sites, QM/MM calculations indicated a relatively minor role for such distortions in determining the redox potential in the reaction center.<sup>50</sup> Zucchelli et al. adapted the NSD approach to predict shifts in both the HOMO → LUMO and (HOMO-1) → (LUMO+1) transition energies for Chl *a* and Chl *b* molecules.<sup>48,51</sup> Using

a simple model for electronic mixing between these two transitions, the NSD coefficients were then translated to predicted shifts in the  $Q_y$  transition frequencies (or site energies) of individual Chl molecules in LHCII. In terms of direct comparison against experimental data, however, the application of such NSD-based methods to predict spectroscopic or electronic parameters remains largely untested.

This lack of experimental validation is likely due in part to the scarcity of Chl-protein structures with sufficiently high resolution to make ring-deformation calculations feasible. Whereas an atomic displacement of a few tenths of an Angstrom will have little impact on a long-range electrostatic interaction, it can represent a significant perturbation on ring-deformation energies. Thus a rigorous assessment of the NSD-based methods (or any other approach focused on ring deformation) must simultaneously address *both* whether the specific deformation model is reliable *and* whether the available structural data is sufficiently precise.

The goal of this work is to identify what structural resolution is “good enough” that Chl ring deformation may be accurately assessed by NSD and related approaches. To accomplish this goal, we carried out an NSD-based survey of ring-deformation statistics across a library of 113 solved structures for a single, large CP: the Photosystem II (PSII) complex of oxygenic photosynthesis. The C2 symmetry of the complex (which allows for comparison of equivalent sites on both halves of a structure), the structural diversity of its Chl binding sites, and the large number of independent structures of PSII available (of varying resolution) offer a fertile testing ground for the reproducibility of ring-deformation estimates. These results for PSII are then benchmarked against an even larger ensemble of Chl- and Bacteriochlorophyll (BChl)-containing structures from the protein data bank (PDB) and the PDB-Redo server (which offers a refinement “redo” of X-ray data from the PDB),<sup>61,62</sup> exploring how the structural refinement protocol impacts Chl ring-deformation estimates.

At a basic level, our results provide a concrete assessment of the precision with which ring-deformation energetics can be characterized in Chl-binding proteins as a function of resolution. We anticipate that these observations will help to calibrate expectations of what

insights can and cannot be extracted from NSD analysis of a given Chl-protein structural model. More broadly, our analysis offers insight into the factors that determine the accuracy of ligand binding geometries, highlighting the importance of minimizing refinement bias by using structural models constructed using a unified refinement protocol, such as those publicly available from the PDB-redo database.<sup>61–63</sup>

## Approach

In this section, we provide a brief overview of our approach. Technical details are provided in the Methods section. Figure 1 provides a schematic overview of our workflow. First, a Basic Local Alignment Search Tool (BLAST) search was used to identify as many available PSII structures as possible. Second, home-built Python scripts were used to split each structure into monomeric PSII cores and to align each core against a single reference structure; this allowed us to automatically identify matching pigments across all structures, accounting for the fact that not all structures contain the same number of pigments. We then performed NSD calculations using the PigmentHunter app<sup>64</sup> to quantify out-of-plane deformation of each Chl and Pheo ligand in each structure. Finally, from this database on NSD coefficients, we analysed statistically the variability of NSD estimates both from structure to structure and from site to site within each structure.

## Selection of PSII Structures

To identify available PSII structures, we ran BLAST queries for the sequences of PSII chains A (D1 protein), B (CP47 protein), C (CP43 protein), and D (D2 protein) from 10 different organisms (see Table 1) against the pdbaa sequence database representing all entries in the RCSB Protein Databank. (See Methods for additional details.) Because the *Thermosynechococcus vestitus* and *Thermotichus vulcanus* sequences are nearly identical, only one search was performed for both species, based on the *T. vestitus* sequence. Note that the

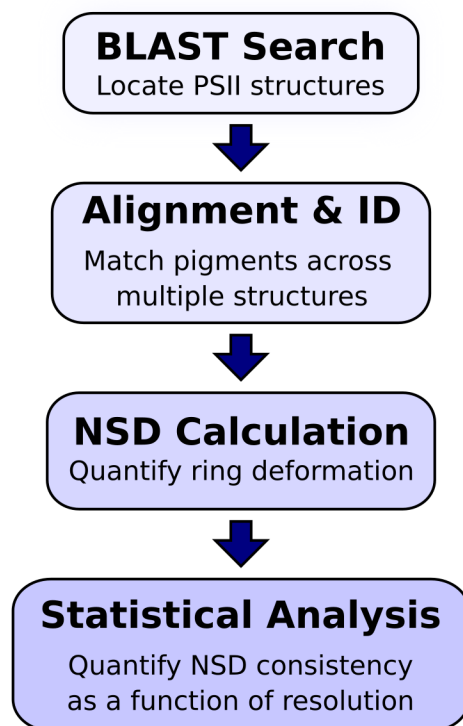


Figure 1: Workflow schematic for our NSD statistical analysis on PSII.

species names *Thermosynechococcus vestitus* and *Thermotichus vulcanus* have recently been updated, from *Thermosynechococcus elongatus* and *Thermosynechococcus vulcanus*, respectively, and many of the FASTA sequence files online still contain the old species identifiers.

For comparison against these “raw” PDB structures, we likewise downloaded the equivalent structure from the site [pdb-redo.eu](http://pdb-redo.eu), which hosts “re-done” versions of PDB models that have been rebuilt using a fixed refinement protocol.<sup>61–63</sup> Comparison of these “redo” structures against the native PDB entries offers a look at the impact of the refinement protocol on ring-deformation predictions. Note, however, that [pdb-redo](http://pdb-redo.eu) operates only on X-ray data; a similar server has recently been established for CryoEM data, but at the time of our analysis, the number of available chlorophyll-protein structures was not sufficient for robust statistical analysis.<sup>65</sup>

Since different PSII structures use different numbering conventions for the bound pigments, we performed a systematic alignment of each monomeric PSII structure against a

Table 1: Results from blastp search of the pdbaa sequence database against reference sequences for ten organisms. The column “# PDB entries” indicates the total number of PDB entries located that contained at least one copy of the D1/CP47/CP43/D2 core monomer. The column labeled “# monomers” indicates the total number of such core monomers from all PDB entries. (Most, but not all, PDB entries contain 2 monomers in a dimeric arrangement.)

Organism	# PDB entries	# monomers
<i>Thermosynechococcus vestitus</i> ( <i>Thermosynechococcus elongatus</i> )	58	112
<i>Thermostichus vulcanus</i> ( <i>Thermosynechococcus vulcanus</i> )	32	62
<i>Dunaliella salina</i>	5	14
<i>Chaetoceros gracilis</i>	5	10
<i>Synechocystis PCC 6803</i>	3	5
<i>Chlamydomonas reinhardtii</i>	3	6
<i>Pisum sativum</i>	3	6
<i>Arabidopsis thaliana</i>	2	4
<i>Cyanidium caldarium</i>	1	4
<i>Spinacia oleracea</i>	1	2
Total	113	225

single reference structure, the first PSII monomer (chains A-Z) present in the 1.9 Å 3WU2 structure.<sup>60</sup> (See Methods for detailed alignment protocol.) The choice of this reference is somewhat arbitrary, but we chose 3WU2 as a suitably high-resolution reference with a convenient numbering scheme for the bound Chl and pheophytin (Pheo) pigments. (For example, pigment ordering in 3WU2 is consistent with the proposed numbering scheme of Ref.<sup>66</sup>) The results of this procedure are illustrated graphically in Figure 2, where small colored spheres represent the Mg atoms for all Chl molecules in all PDB structures considered. Larger yellow spheres indicate additional pigments, not present in 3WU2, identified in a small number of structures. (The PDB identifiers of the corresponding structures are depicted in yellow.) These additional Chls come from PSII supercomplex structures, where they are bound between the core proteins and a peripheral antenna. Since they are present in only a few structures, we exclude them here from our statistical analysis. As can be seen in Figure 2, the positions of the remaining 35 Chl *a* molecules are strongly conserved across

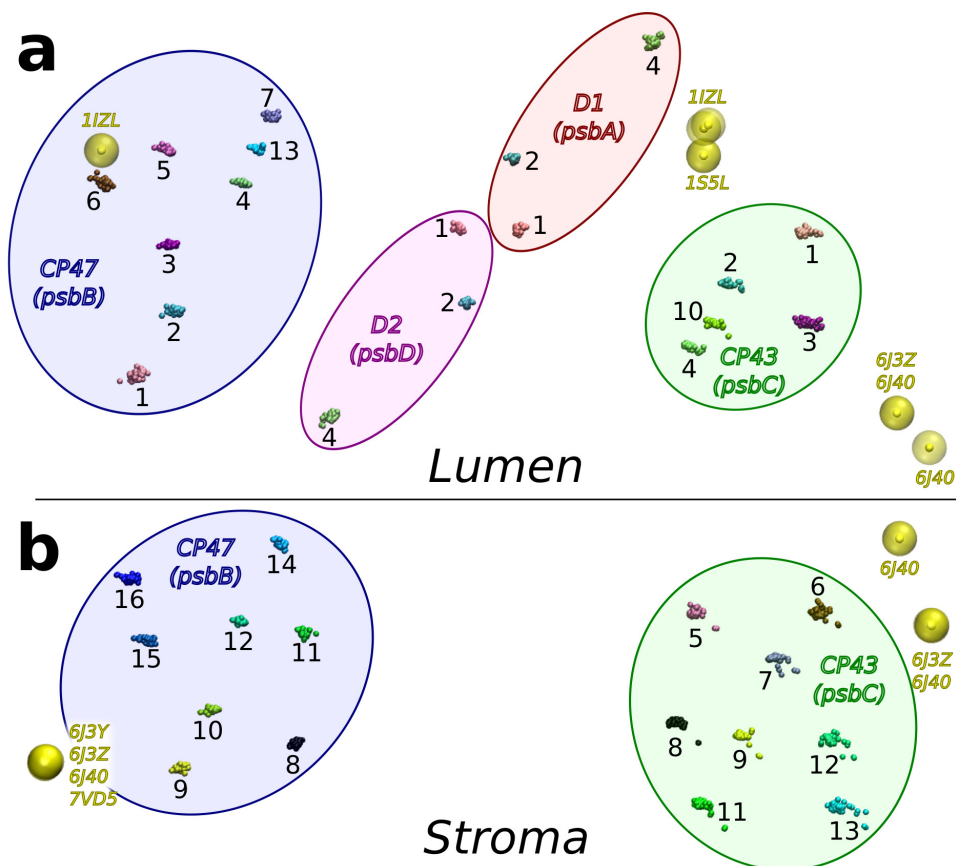


Figure 2: Alignment results for PSII structures from the PDB. For visualization purposes, the two pigment pools that are positioned toward the luminal and stromal sides of the membrane are separated into separate panels (A: Lumen; B: Stroma). Small colored spheres represent “core” Chl *a* molecules that are present in the 3WU2 reference structure. Larger yellow spheres indicate additional pigments located in PSII supercomplexes (PDB codes indicated in yellow). The colored ovals indicate the protein chain to which each pigment is assigned.

all structures, with Mg atom positions typically within 2 Å of the 3WU2 reference.

## NSD Calculation

To quantify ring deformation in bound Chl and Pheophytin (Pheo) molecules, we incorporated the normal coordinate structural decomposition (NSD) method developed by Shelnut and co-workers into the PigmentHunter application,<sup>64</sup> which automates the process of identifying and characterizing Chl pigments in protein structures. (See Methods for implementation details.) In the NSD approach, each pigment is first aligned against a fully planar

symmetry-adapted Copper(II) porphine reference structure. After alignment, normal mode deformation coefficients  $d_m$  are calculated by taking the scalar product between a given normal mode vector and the *difference* between the actual coordinates of the aligned pigment ring atoms and those of the Copper(II) porphine reference. Mathematically, for the  $k^{\text{th}}$  normal mode, we have

$$d_k = \hat{D}_k \cdot \vec{D}_{\text{obs}} \quad (1)$$

where  $\hat{D}_k$  is the normal mode displacement vector (with  $3N_{\text{atoms}}$  elements and unit norm) for normal mode  $k$ , and  $\hat{D}_{\text{obs}}$  is the vector of observed displacements between the aligned pigment and the Copper(II) porphine reference. Here and in what follows, we omit the superscript  $\Gamma$  used by Jentzen et al. to denote the symmetry group of each vibration, since we restrict our attention to the “minimal basis” consisting of the lowest-frequency out-of-plane deformation mode of each symmetry.<sup>41</sup> (The corresponding symmetry groups are noted in Table 2.) Higher frequency mode deformations are expected to be more difficult to quantify from structural data since they represent smaller displacements in real space.<sup>29,41</sup> For reference, Figure 3 illustrates the impact of deformation by  $d_k = 1 \text{ \AA}$  along each of these normal modes on the structure of an otherwise-planar porphyrin ring.

Table 2: Properties for the six deformation modes comprising the minimal basis for NSD analysis: index ( $k$ ), symmetry ( $\Gamma$ ), force constant ( $\kappa$ ), and frequency ( $\bar{\nu}$ ). Data is from Refs.<sup>41,48</sup>

$k$	Name	$\Gamma$	$\kappa \text{ (cm}^{-1}\text{\AA}^{-2}\text{)}$	$\bar{\nu} \text{ (cm}^{-1}\text{)}$
1	sad	$B_{2u}$	761	65
2	ruf	$B_{1u}$	1,379	88
3	dom	$A_{2u}$	3,461	135
4	wav(x)	$E_{g(x)}$	5,626	176
5	wav(y)	$E_{g(y)}$	5,626	176
6	pro	$A_{1u}$	19,929	335

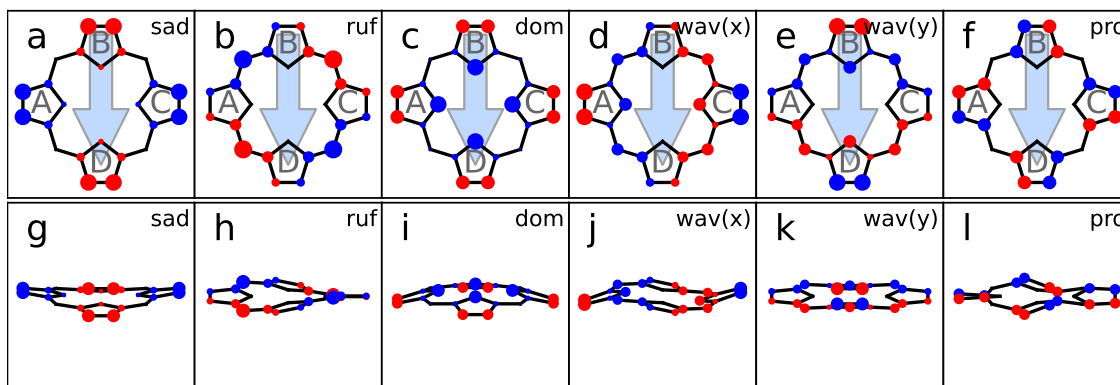


Figure 3: Geometries of a porphine ring displaced by  $1 \text{ \AA}$  along each NSD basis mode (see frame labels). Frames a - f are viewed from above the ring; red indicates positive displacement (toward the viewer), while blue indicates negative displacement (away from the viewer); the radius of each dot indicates the amplitude of the displacement of each atom. Frames g - l are viewed from the side, corresponding to an  $80^\circ$  rotation around the NA - NC axis relative to frames a - f. The blue block arrow indicates the approximate direction of the  $Q_y$  transition dipole moment.

## Methods

### BLAST Search

Searches were run using a local installation of the blastp tool, obtained from <https://blast.ncbi.nlm.nih.gov/Blast.cgi> against a copy of the pdbaa database last updated on August 9, 2023. The blastp output was formatted to print the percent identity of each sequence against the subject accession number ('sacc' flag), percentage of identical matches ('pident' flag), number of mismatches ('mismatch' flag), and all subject sequence IDs ('sallseqid' flag). The 'sallseqid' flag ensures that all matching PDB structures are reported, including distinct monomers within multimeric PSII structures. (Most PSII structures are dimeric.) This output was then post-processed using Python scripts to first identify PDB structures that contain at least 90% identity and fewer than 10 mismatches against each of the four chains corresponding to proteins D1, CP47, CP43, and D2 and then to download selected PDB structures and split them into individual PDB files for each PSII monomer. Only the chains corresponding to D1, CP47, CP43, and D2 were kept at this stage. The mismatch allowance

provides some redundancy, since, e.g., the *A. thaliana* and *S. Oleracea* sequence searches both include each other; only one copy of each structure was included in our database, however, and in subsequent analysis all structures were identified by the source organism specified in its own FASTA sequence.

## PSII Alignment

Structure alignment was complicated by the varying number (and order) of pigments present in each structure, particularly since even the chain assignments for pigments in the reaction center (RC) sometimes vary from structure to structure. To standardize the alignment process, we adopted the following three-step protocol:

1. Initial alignment against Pheo positions:

- Identify the two Pheo residues. Pheos were identified as any molecule that contains four atoms named ‘NA’, ‘NB’, ‘NC’, and ‘ND’ but does *not* include an atom named ‘MG’. In each structure queried, exactly two such residues were identified.
- Determine the identity of the two Pheos based on proximity to the D1 and D2 chains. This step is necessary because some PSII structures assign both Pheos to the same chain.
- Align the identified Pheo residues against the ring atoms of the corresponding residues in the reference structure.

2. First-round Chl *a* alignment: Although Pheo alignment brings the subject structure into approximate alignment with the reference, pigments far from the RC may remain poorly aligned. We thus repeated the alignment based on Chl Mg atom positions as follows:

- Match Chl molecules in the subject structure to corresponding Chls in the reference structure, based on the distance between their ‘MG’ atoms. Any subject

Chl that cannot be matched to within 4.0 Å of a reference pigment is ignored in the alignment.

- Align the structures based on the position of all matched Chl Mg atoms.
3. Final Chl *a* alignment: Step 2 is repeated on the new structures, with a tighter cutoff of 2.0 Å for matching Chl MG atom positions. This alignment helps to ensure that any pigments excluded from Step 2 due to the crudeness of the initial Pheo alignment are accounted for in the final alignment.

## Chl and BChl Ligand Search

For further statistical analysis, we searched the PDB for all structures containing either Chl or BChl ligands. Structures were identified by first searching the string “chlorophyll, bacteriochlorophyll” in the Advanced Search > By Chemical Attributes > By Chemical Name > “has exact phrase” feature at [www.rcsb.org](http://www.rcsb.org) and then retaining only those PDB structures for which a corresponding PDB-Redo structure could be located. In total, this search returned 261 pairs of PDB/Redo structural models.

## NSD Implementation

The NSD method was implemented as described in Ref.<sup>41</sup> with two minor adjustments, which were necessary to determine the sign of the displacement coefficients  $d_k$  for each mode. These adjustments have no impact on total deformation energies, the key quantity of interest in most NSD applications in the literature. First, in rotating the subject Chl into the plane of the reference porphyrin, we added a step to explicitly enforce that the normal vector of the subject Chl ring is aligned parallel and not *antiparallel* to the normal vector of the reference porphyrin. (The original Jentzen protocol guarantees that the two rings are coplanar, but does not distinguish between a “top” and “bottom” plane since, for symmetric porphyrins, the distinction is irrelevant.<sup>41</sup>) In our implementation, the subject normal vector

is always parallel to the vector  $\vec{r}_{\text{NB}} \times \vec{r}_{\text{NA}}$ , where  $\vec{r}_{\text{NA}}$  and  $\vec{r}_{\text{NB}}$  are the coordinates of the NA and NB nitrogen atoms after centering the Chl ring at the origin. (By this convention, the normal vector points out of the page toward the reader in frames a - f of Figure 3.) Second, we corrected what we believe to be a typographical error in Eq. B9 of the Supporting Information for Ref.:<sup>41</sup> we choose the final rotation angle  $\theta$  in the  $xy$  plane via the condition (in the notation of Ref.<sup>41</sup>)

$$\left(\vec{S}_{\text{ref}}^x \vec{S}_{\text{obs}}^y - \vec{S}_{\text{ref}}^y \vec{S}_{\text{obs}}^x\right) \sin \theta + \left(\vec{S}_{\text{ref}}^x \vec{S}_{\text{obs}}^x + \vec{S}_{\text{ref}}^y \vec{S}_{\text{obs}}^y\right) \cos \theta > 0. \quad (2)$$

The corresponding condition of Ref.<sup>41</sup> (Eq. B9) has the  $\cos \theta$  and  $\sin \theta$  switched; when we found this not to produce optimal alignment, we rederived Eq. 2 from Eqs. B7 and B8 of Ref.<sup>41</sup> We hasten to add that this change does not affect any literature results we are aware of since (as with the normal vector convention already noted) it changes only the sign on the deformation coefficients  $d_k$ ; deformation energies (which are quadratic in  $d_k$ ) are unchanged. To validate our algorithm, we compared NSD outputs from our PigmentHunter implementation to those of the Liptak server<sup>40,42</sup> for chain J of the 1RWT structure<sup>67</sup> of the light-harvesting complex II. Within output precision, the results match exactly up to changes in sign on the deformation coefficients  $d_k$ , presumably due to implementation differences such as those already noted.

## Site Energy Estimates

To assess the potential impact of ring deformation on pigment transition energies, we implemented a method developed by Zucchelli et al. to estimate site energy shifts from NSD coefficients. Briefly, the energy gaps  $\Delta E_{H \rightarrow L}$  and  $\Delta E_{H-1 \rightarrow L+1}$  for the HOMO  $\rightarrow$  LUMO and

HOMO-1  $\rightarrow$  LUMO+1 transitions for Chl pigments are calculated as

$$\Delta E_{H \rightarrow L} = (E_{e_g(x)}^0 - E_{a_{1u}}^0) + (\delta E_{e_g(x)} - \delta E_{a_{1u}}) \quad (3)$$

$$\Delta E_{H-1 \rightarrow L+1} = (E_{e_g(y)}^0 - E_{a_{2u}}^0) + (\delta E_{e_g(y)} - \delta E_{a_{2u}}) \quad (4)$$

where  $E_{\Gamma}^0$  is the undistorted (planar) reference energy of the electronic state with symmetry  $\Gamma$ , and  $\delta E_{\Gamma}$  is the distortion energy due to displacement of the macrocycle along normal modes in the  $\Gamma$  symmetry group. Since our calculations consider only the lowest-frequency mode of each symmetry, this is simply

$$\delta E_{\Gamma} = \kappa_k d_k \equiv \varepsilon_k \quad (5)$$

where  $\kappa_k$  is the force constant for displacement along the normal mode  $k$  which belongs to the symmetry group  $\Gamma$  (see Table 2). The quantity  $\varepsilon_k$  is the deformation energy for mode  $k$ , i.e., the change in the molecular potential energy due to distortion of the ring away from a full planar geometry. The  $Q_y$  transition frequency of each pigment is then calculated as

$$Q_y = \frac{1}{2} [\Delta E_{H \rightarrow L} + \Delta E_{H-1 \rightarrow L+1} - \sqrt{(\Delta E_{H-1 \rightarrow L+1} - \Delta E_{H \rightarrow L})^2 + 4C^2}] \quad (6)$$

Where  $C$  is a coupling coefficient that accounts for mixing between the HOMO  $\rightarrow$  LUMO and HOMO-1  $\rightarrow$  LUMO+1 transitions, and has been estimated by Zucchelli et al. as 6742  $\text{cm}^{-1}$  for Chl *a* and 6215  $\text{cm}^{-1}$  for Chl *b*.

## Monomer-to-Monomer Variation

To quantify variation across two or more monomers within a single PSII structural model, we use the metric

$$\sigma_{\text{mm}} = \sqrt{\frac{1}{N_{\text{sites}} - 1} \sum_{s=1}^{N_{\text{sites}}} \sigma_s^2}, \quad (7)$$

where

$$\sigma_s = \sqrt{\frac{1}{N_{\text{mon}} - 1} \sum_{m=1}^{N_{\text{mon}}} \left( \varepsilon_k^{(s,m)} - \bar{\varepsilon}_k^{(s)} \right)^2} \quad (8)$$

is the sample standard deviation in  $\varepsilon_k$  calculated for site  $s$  across all monomers  $m$  present in a given PDB entry. Here the superscript  $(s, m)$  indicates the deformation energy  $\varepsilon_k$  calculated for site  $s$  in monomer  $m$ , while  $\bar{\varepsilon}_k^{(s)}$  indicates the average deformation energy for site  $s$  across all monomers extracted from the same PDB entry. Expressed briefly,  $\sigma_{\text{mm}}$  is the sample standard deviation across all *sites* of the sample standard deviation in  $\varepsilon_k$  across each *monomer* within a given PDB structure. In the most common case that only two monomers are present,  $\sigma_{\text{mm}}$  reduces to simply the root-mean-square difference in deformation coefficients between the two monomers.

Note that our NSD calculations do *not* directly take into account the positional uncertainty associated with the finite resolution of each structure or variations in local resolution within a given structure. A single NSD coefficient is calculated for each site in each structure using the final reported coordinates of each atom; and the same global resolution reported in the PDB entry is assigned to all sites. Statistical metrics like  $\sigma_{\text{mm}}$  are then used to explore (as a function of reported resolution and refinement protocol) how reproducible are the ring deformation coefficients associated with each site in the PSII complex. In the future, it may be informative to examine to what extent these metrics are affected by variations in local resolution within each structure, although a preliminary check for such effects in our data

set did not reveal clear trends.

## Results and Discussion

### Ring Deformation in PSII structures from thermophilic Cyanobacteria

For a first look at ring-deformation in PSII, we focus on PDB structures derived from the two thermophilic cyanobacteria *T. vulcanus* and *T. vestitus*. Since the great majority of PSII structures derive from one of these two species, and since the sequences of their Chl-binding PSII subunits are nearly identical, their combined pool of PSII structures offers a useful test for structure-to-structure variability in ring-deformation assignments. In addition, since *T. vulcanus* PSII was also the focus of a previous NSD study by Saito et al., the deformation coefficients we analyze statistically here can be compared to the previously reported values, with the caveat that Saito et al. employ a modified NSD basis adapted for Mg-porphyrins.<sup>50</sup>

Figure 4 summarizes ring-deformation statistics for each pigment in PSII across a population of 134 PSII monomers extracted from 69 *T. vulcanus* and *T. vestitus* PDB entries with resolutions of 3 Å or better. The black dot for each site represents the average deformation coefficient  $d_k$ , while “whiskers” indicate plus and minus one standard deviation. The shaded gray area indicates the corresponding range (mean value plus or minus one standard deviation) in calculated ring-deformation energies  $\epsilon_k$  for each site. For comparison, the red dashed line indicates the thermal energy  $k_B T = 208 \text{ cm}^{-1}$  available at 300 K. Frame g shows the corresponding site energies calculated using the model of Zucchelli et al.<sup>48</sup>

This analysis reveals that ring-deformation along the two lowest-frequency modes – the “saddle” and “ruffle” deformations, with frequencies of 65 and 88  $\text{cm}^{-1}$ , respectively – are remarkably consistent across the entire structural ensemble; for both modes, the root-mean-square (RMS) average of the standard deviations across all sites is less than 0.12 Å, corresponding to less than 70  $\text{cm}^{-1}$  of uncertainty in deformation energy. For the higher-frequency

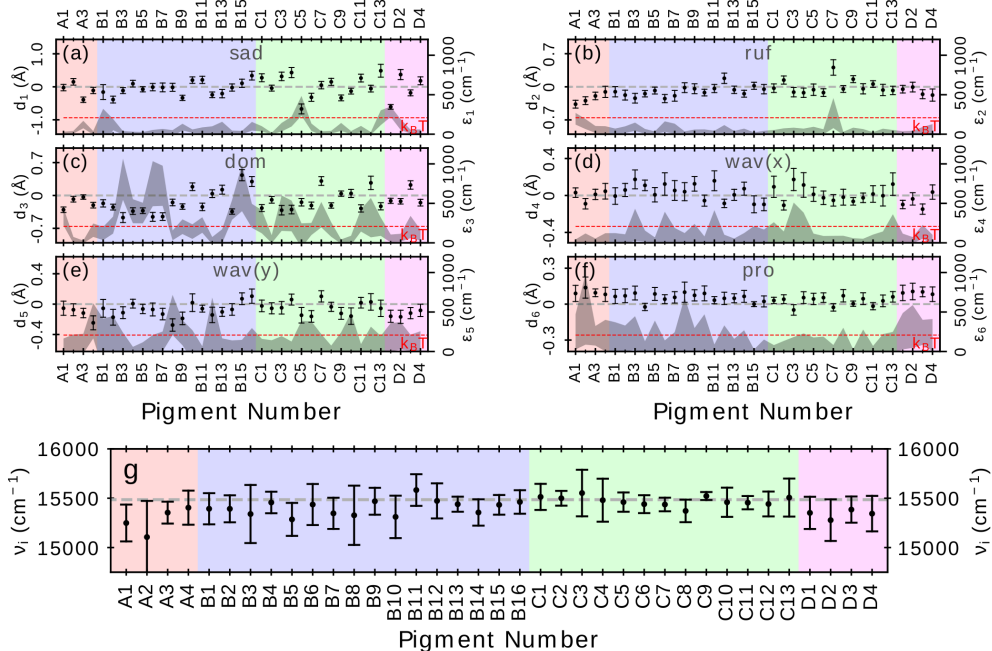


Figure 4: NSD coefficients (Frames a - f, left axis), deformation energies (Frames a - f, right axis), and estimated site energies (Frame g) for each pigment in PSII from thermophilic cyanobacterial structures with resolution better than 3.0 Å. Frames a - f: black dots represent the average value of each deformation coefficient  $d_k$  across all structures (scale on left axis); error bars indicate plus and minus one standard deviation. The grey dashed line indicates zero deformation. Markers along the x-axis indicate the protein chain (A = psbA, i.e., the D1 protein; B = psbB, i.e., CP47; C = psbC, i.e., CP43; D = psbD, i.e., the D2 protein) and pigment numbers from Figure 2. The colored backgrounds denote different protein chain as in Figure 2. Gray shaded areas indicate plus and minus one standard deviation in deformation energy  $\epsilon_k$  (right axis). The red dashed line indicates the thermal energy,  $k_B T = 208 \text{ cm}^{-1}$ , at 300 K. Frame g plots site energies calculated using the model of Zucchelli et al.<sup>48</sup> Black dots indicate average values; error bars indicate plus and minus one standard deviation.

modes, the absolute error in the coefficients  $d_k$  are similarly less than 0.1 Å. But the corresponding uncertainty in the deformation energy is much larger (on the order of 100 - 200  $\text{cm}^{-1}$ ) thanks to the quadratic scaling of  $\epsilon_k$  with normal mode displacement. This uncertainty translates into correspondingly large errors in the calculated site energies: the RMS average for the uncertainty at each site is 170  $\text{cm}^{-1}$ , compared with a standard deviation of only 93  $\text{cm}^{-1}$  between the average energies of each site. On average, the uncertainty at each site is thus larger than the site-energy variability *between* sites.

On the one hand, it is encouraging that in these results the absolute ring-deformation

errors appear to be substantially lower than the reported model resolution. Note here that the deformation coefficient  $d_k$  for a given mode represents an upper bound on the spatial displacement of any single atom, since the Cartesian displacements for each atom are obtained by multiplying  $d_k$  (with units of Å) against a normalized vector whose entries (representing the Cartesian displacements of each atom as the normal mode is displaced) each have absolute values less than one.

On the other hand, the large uncertainty in deformation energies for the higher-frequency modes bodes poorly for the accurate assessment of ring-distortion energetics from raw structural data. This observation is particularly troubling in that the total deformation energy for each site is a simple sum of the deformation energies along each mode; thus uncertainties in the energetics of each individual mode will be compounded in the total energetic uncertainty. This compounding of errors is even more severe when calculating site energies under the Zucchelli model, in which the high-uncertainty wav(x) and pro modes play a dominant role due to their matching of the HOMO and LUMO orbital symmetries (see Eq.(3) above).<sup>48</sup>

## Dependence on Resolution

This raises the obvious question: If 3 Å resolution is not “good enough” to accurately assess ring-deformation energetics, what resolution *would* be sufficient? In answering this question, care must be taken that variations in the experimental conditions under which different structures were obtained do not invalidate their comparison. For example, some structures in our ensemble are from low temperatures, while others were studied at room temperature; some PSII structures feature metal exchange sites, point mutations, or even subunit deletion, any of which could plausibly cause variations in the true ring-deformation values for each site. Other structures are part of an X-ray free electron laser (XFEL) series (see, e.g., Refs.<sup>68,69</sup>), where multiple structures are solved together as a function of the time-delay after some excitation pulse; including such series data could arguably either increase the heterogeneity of the overall ensemble (if the excitation pulse induces structural changes in the pigments<sup>57</sup>)

or decrease heterogeneity (since any input bias in the refinement procedure will be nearly identical for all structures.) On the other hand, excluding individual structural models from the data is somewhat subjective and rapidly decreases the quality of the statistical ensemble.

The symmetric structure of PSII offers a unique opportunity to circumvent this difficulty: instead of examining the variance in deformation energies at each site across the entire ensemble, we can instead examine the monomer-to-monomer variance *within* each PSII structure. To quantify such variability, we use a metric  $\sigma_{\text{mm}}$  (see Eq. (7) in Methods), which in the most common case that a PDB structure contains exactly two monomers is simply the root-mean-square deviation between deformation energies for equivalent sites across the two monomers. When more than two monomers are present,  $\sigma_{\text{mm}}$  generalizes to be a site-averaged sample standard deviation.

The advantage of this metric is that the calculated deviations  $\varepsilon_k^{(m,s)} - \bar{\varepsilon}_k^{(s)}$  are always taken between *equivalent sites within a single structure*, so that all sites are subject to the same experimental conditions and structural refinement methods. Since the two monomers of the PSII complex are both structurally and functionally equivalent (according to current knowledge),  $\sigma_{\text{mm}}$  provides a measure of the uncertainty in assigning ring-deformation energies across equivalent sites.

The results of this analysis are presented in frames a - f of Figure 5, where red dots represent PDB entries solved via X-ray diffraction and blue ‘×’ marks represent CryoEM structures. The data in Figure 5 extends over our entire PDB PSII ensemble (from all organisms), since the calculated  $\sigma_{\text{mm}}$  values are calculated independently for each structure and thus any true species-to-species structural variations should not affect the outcome.

Two observations bear immediate comment: first, that the monomer-to-monomer error  $\sigma_{\text{mm}}$  is near zero for almost all CryoEM structures and, second, that for most modes  $\sigma_{\text{mm}}$  actually *increases* systematically with improved resolution for PDB X-ray structures. The first result is something of a technical artifact: in a system with C2 symmetry, there is no *a priori* way to distinguish between individual monomers in raw CryoEM images; as

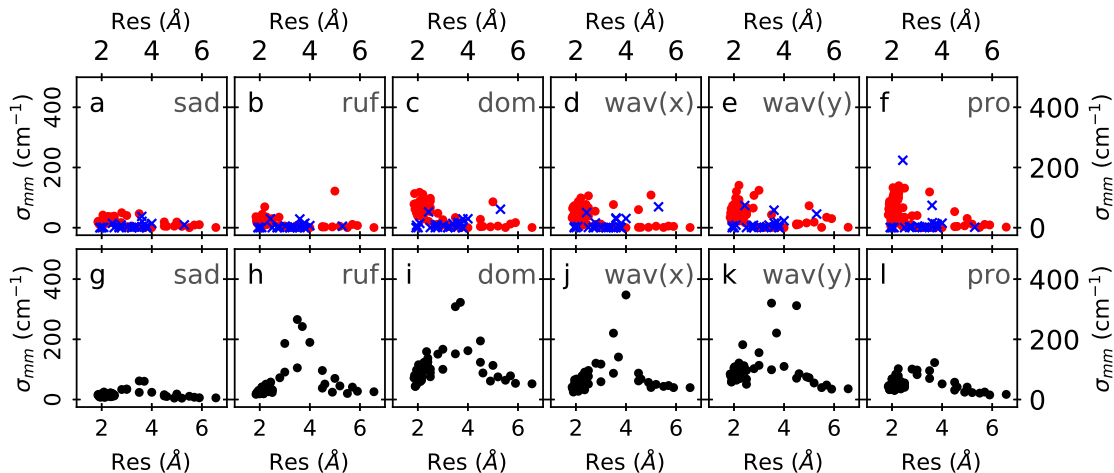


Figure 5: Cross-monomer deviation  $\sigma_{\text{mm}}$  for each deformation coefficient, tabulated for all available PSII structures that include at least two equivalent monomers. Frames a - f are for PDB structures; red indicates structures solved by X-ray, while blue indicates CryoEM structures. Frames g - l are for X-ray structures from the PDB-Redo server.

a result, CryoEM refinement protocols usually either directly enforce symmetry between equivalent monomers or “expand” the raw data set by including both the raw image and its 180° rotation. Because C2 symmetry is thus built into the refinement protocol itself,  $\sigma_{\text{mm}}$  is simply not a useful metric for examining prediction error in CryoEM structures. Note also that the resolution listed for CryoEM structures is the global resolution; the local resolution for individual sites may vary considerably.

The increasing  $\sigma_{\text{mm}}$  values in the X-ray structures is more informative; this trend may at first appear surprising, but it is not difficult to rationalize. For low-resolution X-ray structures, the raw data is not of sufficient quality to meaningfully constrain atomic positions, apart from showing the approximate location and orientation of the molecule. Indeed, sites near the periphery of the complex may not be fully occupied even for high-quality crystals, further limiting the information available to the experiment. Assigning the precise locations of each atom must thus rely heavily on predefined molecular topologies and energetic constraints that are introduced by the user and are the same for all sites. As a result, local molecular conformations in the two halves of the monomer are similar simply because both closely match the input parameters defined by the user. In a high-resolution structure, by

contrast, the assignment of atom positions relies much more heavily on the experimental data. Although this provides a more objective description of the actual atom locations, it also makes the structure much more sensitive to noise in the experimental data set.

## Dependence on the Refinement Procedure

Both of these observations call attention to the role played by the refinement protocol in making structural predictions. In order to examine this role more closely, we repeated the  $\sigma_{\text{mm}}$  analysis using structures pulled from the PDB-Redo repository ([www.pdb-redo.eu](http://www.pdb-redo.eu)), which hosts structural models rebuilt from the experimental diffraction data (taken from the PDB repository) using a single, systematic protocol.<sup>61-63</sup> Comparison between PDB and PDB-Redo structures offers a simple way to test the sensitivity of individual structural assignments to the input assumptions of a given refinement protocol.

Frames g - l of Figure 5 plot calculated  $\sigma_{\text{mm}}$  values for PDB-Redo structures. For both very high- and very low-resolution structures, the  $\sigma_{\text{mm}}$  values are generally similar for PDB-Redo and raw PDB structures, indicating that at both extremes, the performance of the various refinement protocols are similar. At intermediate resolution (3 to 5 Å), however, the PDB and PDB-Redo  $\sigma_{\text{mm}}$  values follow a qualitatively different trend: while the PDB  $\sigma_{\text{mm}}$  errors increase essentially monotonically with decreasing resolution, the PDB-Redo errors reach a peak near 4 Å and then decrease steadily as resolution improves. Although the PDB structures in our ensemble are refined under a variety of different protocols by various authors, this difference in the behavior of  $\sigma_{\text{mm}}$  would appear to indicate that the PDB-Redo protocol (which applies a similar protocol to all structures) either depends more heavily on the experimental data or imposes fewer symmetry constraints on the model at intermediate resolution. Although this produces larger monomer-to-monomer error at intermediate resolution, it also yields much more systematic performance as the resolution improves, which may reflect a more objective relaxation of restraints with increasing data quality.

To better elucidate the role of input bias in structural refinement, we next examined

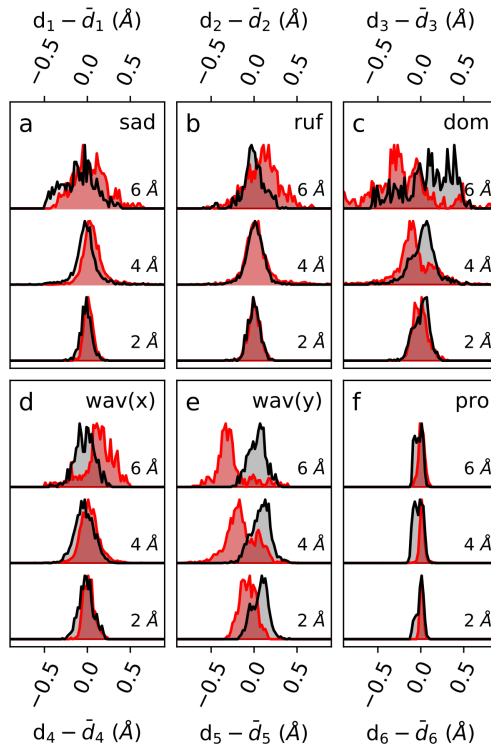


Figure 6: Error distributions across all thermophilic cyanobacterial PSII structures as a function of resolution. Gray curves are for structures from the PDB; red curves are for PDB-Redo structures. Each histogram is constructed by taking the average deformation coefficient  $\bar{d}_k^{(s)}$  across all structures with resolution better than 2.0 Å as the “true” deformation coefficient for each site  $s$  in the PSII structure and then tabulating deviations from this value for each site in each structural model.

the deviation of the normal mode deformation coefficients for every site in every structure from the ensemble average  $\bar{d}_k^{(s)}$  for that site across all structures with resolutions below 2 Å. The coefficient  $\bar{d}_k^{(s)}$  may be understood as the best available estimate of the true ring-deformation value for site  $s$ , so that the deviation  $d_k^{(s)} - \bar{d}_k^{(s)}$  indicates the deviation of any individual structure from that best estimate. Since we are here again comparing deformation values for individual sites across different structures, we restrict our attention to structures from the thermophilic cyanobacteria. We likewise consider only those structures for which a corresponding PDB-Redo entry exists, so as to allow for a direct comparison between PDB and PDB-Redo statistics. The best-estimate deformation  $\bar{d}_k^{(s)}$  is averaged across both PDB and PDB-Redo structures.

The results of this analysis are presented in Figure 6 in the form of histograms (across all structures and all sites) of the calculated deviations  $d_k^{(s)} - \bar{d}_k^{(s)}$  for each normal mode. The histograms are divided into 3 data sets, representing structures with resolutions in the ranges 0 to 2 Å, 2 to 4 Å, and 4 to 6 Å. Gray-shaded curves correspond to raw PDB entries, while red-shaded curves correspond to PDB-Redo structures. For the lowest-frequency mode (frame a), little “bias” in the refinement method is apparent, since all deformation-error histograms are essentially symmetric on either side of zero. For the remaining modes, however, the error histograms exhibit systematic shifts between the PDB and PDB-Redo predictions, which gradually disappear with improving resolution. This trend is most apparent in frame e, corresponding to the “wave(y)” mode, which corresponds to a wavelike deformation (see Figure 3) along the  $N_B$  to  $N_D$  molecular axis (the same direction as the  $Q_y$  electronic transition dipole moment). For this mode, PDB-Redo structures strongly favor distortions in the negative direction (relative to the best-estimate value  $\bar{d}_5$ ), while the raw PDB structures show no such systematic shift. The higher prevalence of systematic shifts in the PDB-Redo data set may be due to the fact that all PDB-Redo structures were constructed using a single protocol, whereas the corresponding PDB models were built by a variety of different authors using different protocols.

To further quantify the effects of such refinement bias, we next tabulated PDB vs. PDB-redo differences for equivalent pigments from an expanded data set that included all PDB models (for which corresponding PDB-Redo structures could be located) that included at least one chlorophyll or bacteriochlorophyll ligand. After matching corresponding pigments between the PDB and PDB-Redo structures using a Python script (since residue ordering is sometimes different between PDB and PDB-redo structures), we calculated the deformation difference

$$\Delta d_k = d_k^{(\text{PDB})} - d_k^{(\text{Redo})} \quad (9)$$

with  $d_k^{(\text{PDB})}$  the deformation coefficient calculated for a single pigment from the PDB repository and  $d_k^{(\text{Redo})}$  the deformation from the corresponding site of the PDB-Redo structure. While this expanded data set provides no reference for the true ring-deformation value for any given site, it does provide much better statistical accuracy for comparing PDB and PDB-Redo differences than does the more limited PSII structural ensemble analyzed so far.

The results of this analysis are presented in Figure 7 in the form of  $\Delta d_k$  histograms, subdivided by the resolution of the corresponding structure (labeled on the left-hand side), from 1 Å to 7Å. (No models were located with sub-Angstrom resolution.) The sample standard deviation  $\sigma$  (a measure of the histogram width) is printed for reference alongside each curve. As may be expected by this point, all modes show considerable error for structures above 2.0 Å resolution, with the exception of the “propeller” distortion, which is presumably not sufficiently resolved (even below 2 Å resolution) for experimental structural models to make meaningful predictions. Interestingly, the “dome” and “wave” modes exhibit a pronounced bias in the 2 - 6 Å resolution range, with PDB-Redo structures favoring negative displacement along the dome and wave(y) deformations and positive displacements along the wave(x) deformation. This bias is particularly pronounced for the doming mode, even in the 2 - 3 Å range and corresponds to quite significant errors in calculated deformation

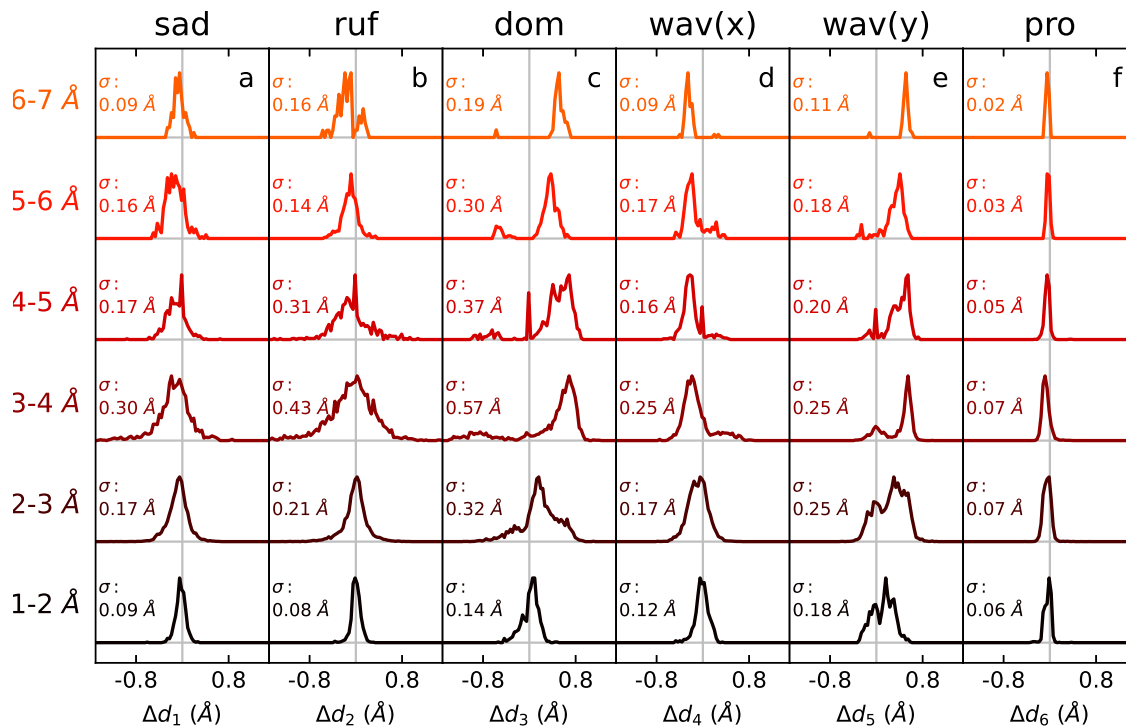


Figure 7: Distributions for the difference  $\Delta d_k$  between PDB and PDB-Redo predictions for all Chl and BChl molecules for which both PDB and PDB-Redo structures could be identified. Histograms are constructed by grouping structures by resolution in 1-Å intervals between 1 and 7 Å. Inset text indicates the sample standard deviation  $\sigma$  for  $\Delta d_k$  across all structures in each bin. Thin gray lines represent ordinate and abscissa axes for each histogram.

energy: the peak near 0.5 Å in the 2-3 Å doming histogram, for example, corresponds to a deformation energy of (see Table 2 and Eq. (5))

$$\epsilon_3 = \left(3,461\text{cm}^{-1}\text{Å}^{-2}\right) \times (0.5\text{Å})^2 = 825\text{ cm}^{-1} \quad (10)$$

roughly four times the thermal energy  $k_B T$  available at room temperature. Below 2.0 Å resolution, however, such systematic bias disappears almost completely: all distributions peak near 0 Å, albeit with much larger distribution widths for the dome, wave(x), and wave(y) modes (sample standard deviations: 0.14 Å, 0.12 Å, and 0.18 Å, respectively) compared to the apparently well-converged saddle and ruffle modes (0.09 Å and 0.08 Å standard deviations, respectively).

## Ring Deformation is Conserved in PSII

With these error estimates in place, we can begin to investigate the statistical significance of ring deformation in PSII. Specifically, for structures at 2 Å resolution (or better), we can consider ring deformation to be significant if it is larger than  $\sim 0.08$  Å in the saddle and ruffle modes, and greater than  $\sim 0.15$  Å for the dome and wave modes. As judged by the 2 Å-resolution histograms in Figure 7), any deformation values below these “cutoffs” are likely to be determined primarily by the structure-refinement process, rather than by the structural data itself. Propeller-mode distortions show no sign of convergence even below 2 Å resolution and thus should probably not be interpreted literally in existing structural models.

With these “guard rails” in mind, a look back at Figure 4 (from thermophilic cyanobacteria structures with 3 Å resolution or better) reveals statistically significant deformation in PSII for the saddle, ruffle, and dome deformations; deformations for the wave and propeller modes are generally smaller than their corresponding cutoffs. Of the three modes with significant deformation, only doming deformations correspond to energies larger than  $k_B T$  at

300 K: the saddle and ruffle modes are well-resolved, but are likely to be largely washed out by the thermal energy available at room temperature. The fact that doming is the only deformation (of the six considered here) that exhibits both statistically *and* thermally significant deformation in PSII suggests that doming in particular bears further investigation for potential functional significance in Chl proteins.

With respect to the prior PSII NSD study of Saito et al.,<sup>50</sup> our statistical analysis largely confirms their single-structure observations, at least for the saddle, ruffle, and doming modes. For example, the P<sub>D1</sub> pigment of the RC (site D1 in our numbering) exhibits a strongly conserved negative saddle deformation in our data, in agreement with the observations of Saito et al.<sup>50</sup> Note that there appears to be a sign change for the doming-mode between our study and Ref.,<sup>50</sup> which uses a different NSD basis: e.g., the mostly positive doming coefficients observed by Saito et al. are largely negative in our calculations, whereas the strong negative doming Ref.<sup>50</sup> reports for the last two sites in CP47 becomes strong positive doming in our analysis.

An obvious follow-up question is: To what extent are such deformations conserved across different photosynthetic organisms? To explore this question, Figure 8 compares ring-deformation coefficients for each PSII site across 8 different organisms: two thermophilic cyanobacteria (*T. vulcanus* and *T. vestitus*), the mesophilic cyanobacterium *S. PCC 6803*, the diatom *C. gracilis*, the red alga *C. caldarium*, two green algae (*C. reinhardtii* and *D. salina*), and three green plants (*S. oleracea*, *A. thaliana*, and *P. sativum*). The shaded gray area in each frame of Figure 8 represents the  $\mu \pm \sigma$  curve for thermophilic structures with resolutions better than 2 Å, with  $\mu$  and  $\sigma$  the sample mean and sample standard deviation, respectively. Colored lines depict mean-deformation estimates across *all* available PDB and PDB-Redo structures for each non-thermophilic organism; not enough structures were available for these 6 organisms (see Table 1) to construct meaningful error bars or to impose constraints on the experimental resolution. (Of the entire non-thermophilic data set, only the *Synechocystis* PCC 6803 structure 7N8O has better than 2 Å resolution.<sup>70</sup>) Given

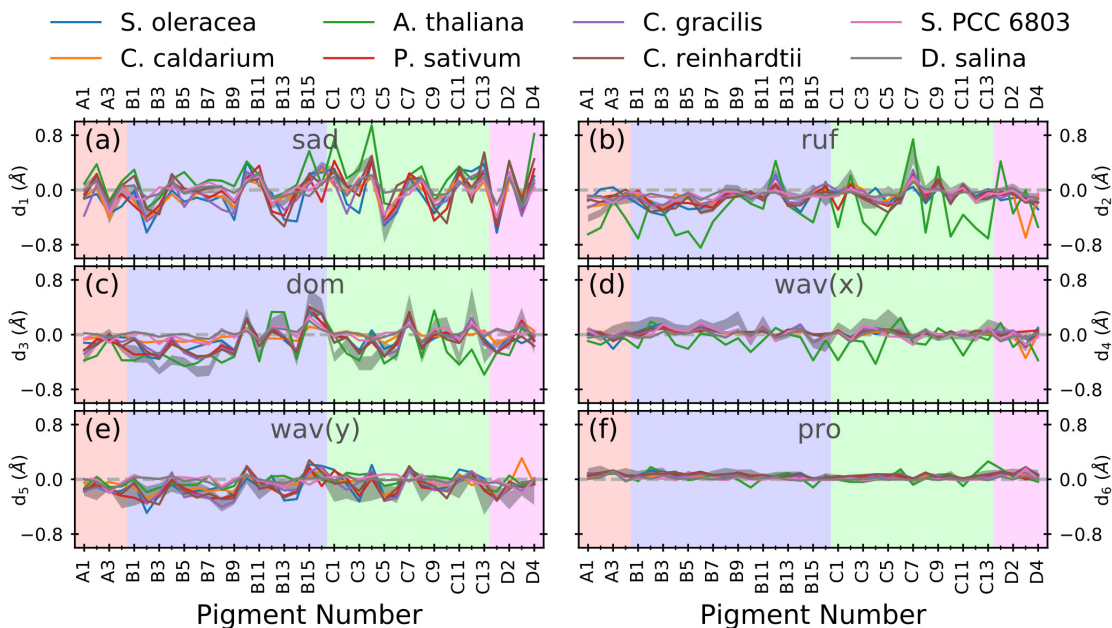


Figure 8: Comparison of deformation coefficients  $d_k$  across PSII across species. The gray shaded area in each plot indicates the  $\pm\sigma$  region for thermophilic structures better than 2.0 Å resolution. Colored lines indicate average values across all PSII structures found for that organism. Both PDB and PDB-Redo structures are included (with equal statistical weight) whenever available.

these limitations, considerable noise is to be expected in the non-thermophilic deformation estimates. Even with this caveat in mind, however, correlation amongst the deformation coefficients for the saddle, ruffle, and dome modes is apparent in the data. Indeed, for the saddle and ruffle modes, the thermophilic deformation data is almost completely obscured by curves from the other six species, which (with few exceptions) lie directly on top of it. The outliers from *A. thaliana* come from averaging over only two structures,<sup>71,72</sup> with resolutions of 2.79 Å and 5.3 Å in the intermediate range where ring-deformation is most sensitive to experimental noise (see Figures 5 - 7). Correlation in the doming mode is less perfect, but qualitative agreement is still apparent (i.e., dome-up sites in the thermophilic structures are typically domed up in all species, and dome-down sites are typically domed down in all species). Inter-species correlation for the wave modes is less clear, but this is to be expected from the fact that the site-to-site variation in these modes is of similar magnitude as the statistical error at each site individually.

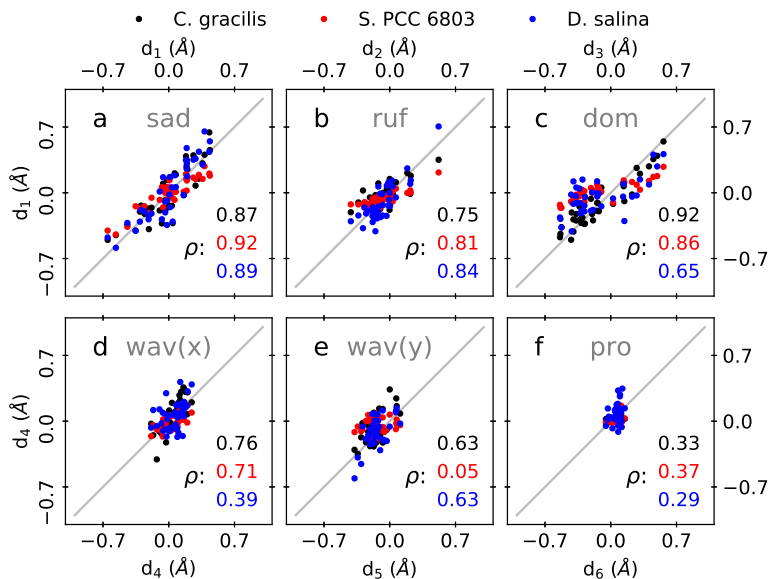


Figure 9: Scatter plot illustrating site-by-site correlation between average ring deformation coefficients from thermophilic cyanobacteria (horizontal axis; structures with 2.0 Å resolution or better) and structures from *C. gracilis* (black points), *D. Salina* (red points), and *S. PCC 6803* (blue points) with resolution better than 2.5 Å.

Figure 9 presents a more quantitative assessment of this correlation via scatter plots of thermophilic ring deformation against deformation in the three non-thermophilic species with PSII structures available at resolutions better than 2.5 Å. Colored text in each frame indicates the corresponding Pearson correlation coefficient, which ranges between -1 and +1;  $\rho = -1$  corresponds to perfect (linear) anticorrelation, while  $\rho = +1$  correspond to perfect correlation. Strong linear correlation ( $\rho > 0.87$ ) is observed for the saddle deformation; the quality of the correlation decreases for the higher frequency modes, but remains significant ( $\rho = 0.65$  to  $\rho = 0.92$ ) for the ruffle and dome modes. Note that the Pearson correlation coefficient indicates only correlation, not equality; thus while doming coefficients for *S. PCC 6803* are strongly correlated with the thermophilic values ( $\rho = 0.86$ ), the slope of the correlation plot is less than one: the *Synechocystis* structures systematically underestimate doming relative to the thermophilic reference. Based on our earlier discussion, this preference for planarity in the *Synechocystis* structures likely reflects the input assumptions of the refinement procedure. Taken together, these results indicate that ring deformation is indeed conserved

across photosynthetic organisms, at least for those modes that can currently be accurately resolved.

## Breaking Symmetry in the PSII Monomer

The fact that ring deformation is conserved, however, does not necessarily imply functional significance. It is possible that ring deformation is conserved merely in order to preserve the structure of the complex as a whole (e.g., through maintaining pigment-protein hydrogen bonding patterns), rather than due to any particular significance of the deformation itself. While our structural analysis cannot alone demonstrate functional significance, it is nonetheless worth investigating whether any concrete correlation can be identified between specific ring deformation features in PSII and the spectroscopic (and by extension light-harvesting) properties of the complex. Since the relevance of ring-deformation for electron transfer in the RC has already been considered in some detail by Saito et al.,<sup>50</sup> we focus here on potential effects on site energies and energy transfer in the core antennas CP43 and CP47.

A simple place to look for such correlation is to again make use of symmetry. But rather than the complete C2 symmetry of the PSII core complex as a whole, we focus here on the pseudo-C2 symmetry that exists *within* each PSII monomer. Both the D1 and D2 proteins of the RC and the CP43 and CP47 core antennae form approximate C2 symmetry pairs; although the symmetry is not perfect (broken especially by the addition of 3 extra Chls in CP47 relative to CP43), the sequences within each pair are strongly homologous, and the pigment locations match closely upon 180° rotation around a pseudo-symmetry axis passing between the Chl D1/Chl A1 pair and orthogonal to the membrane plane.

This approximate symmetry is illustrated in Figure 10, which shows the locations of the Mg atoms of each Chl in PSII; the orientation and numbering of pigments are the same as in Figure 2, but the coloring is determined by the mean doming coefficient  $d_3$  calculated for each site from the thermophilic cyanobacteria structures with resolution better than 2.0 Å. Full red coloring indicates  $d_3 \approx -0.6$  Å, while full blue coloring indicates  $d_3 \approx +0.6$  Å;

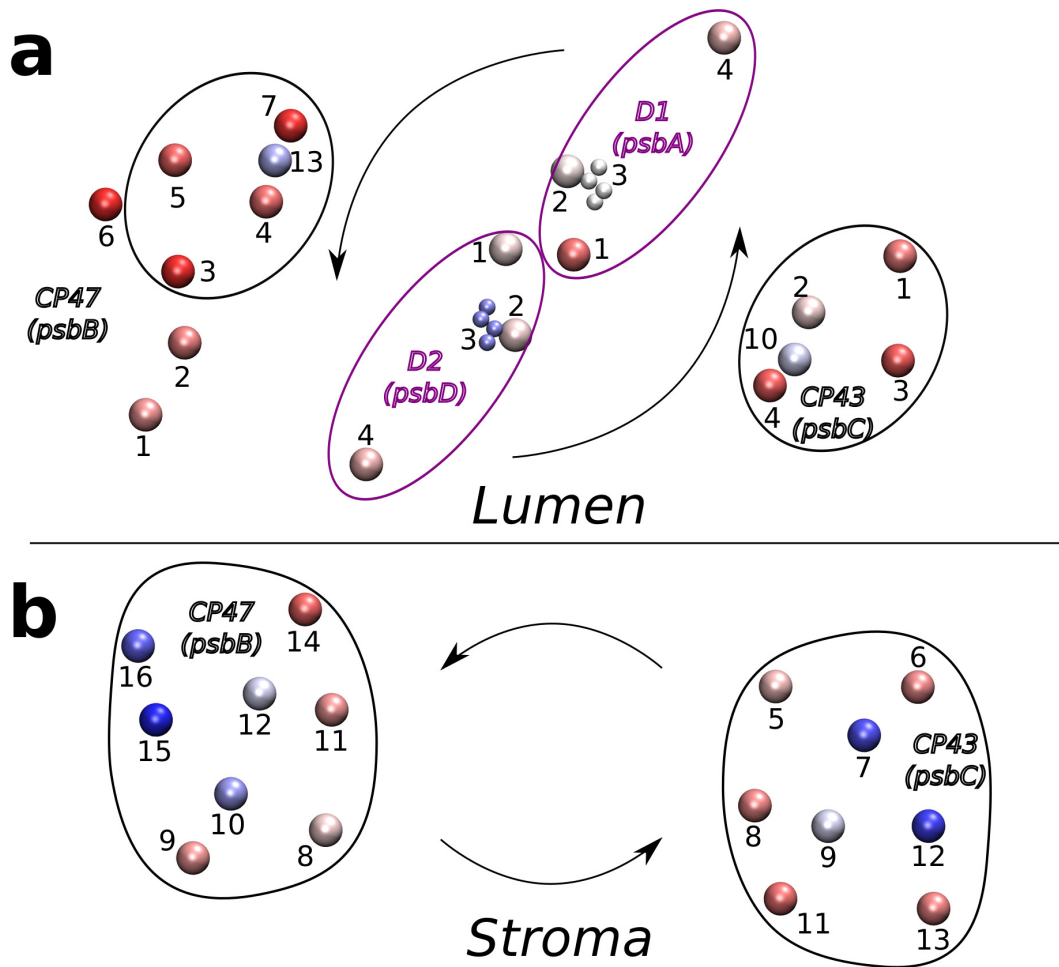


Figure 10: Arrangement of pigments in the lumenal (Frame A) and stromal (Frame B) layers of PSII from the 3WU2 structure.<sup>60</sup> Large spheres indicate Chl Mg atoms; smaller spheres indicate Pheo N atoms. Sphere color indicates doming deformation: red indicates negative doming ( $d_3 < 0$ ); blue indicates positive doming ( $d_3 > 0$ ).

white coloring indicates  $d_3 = 0$  Å. Each Pheo residue is indicated by four smaller spheres representing the four central N atoms. In the figure, pseudo-symmetric pigment clusters are enclosed by solid lines. Note that the approximate symmetry between CP43 and CP47 is broken by the addition of three “extra” Chls (numbers 1, 2, and 6) in CP47.

The broken structural symmetry of the PSII monomer is matched by functional asymmetry in the operation of both the RC and the core antennae. Within the RC, the D1 and D2 pigments are known to play distinct roles in the charge separation process, although specific assignments are somewhat controversial.<sup>66</sup> The asymmetry between CP43 and CP47 is simpler to characterize energetically, since CP47 is known to possess a low-energy electronic state (denoted “F695”) that fluoresces 695 nm light at low temperature; in contrast the lowest-energy states in CP43 emit near 685 nm, in approximate resonance with P680 state of the RC.<sup>66,73</sup> Significantly (and perhaps surprisingly), F695 is typically *not* believed to arise from the three “extra” pigments of CP47, but from a localized transition residing on Chl B16, the symmetry analog of Chl C13 of CP43.<sup>66,73–77</sup> This assignment is again the subject of considerable debate, however, due mostly to the difficulty of fitting available spectroscopic data under the assumption that site B16 is the lowest-energy pigment in the complex.<sup>66,73–81</sup> The functional significance of the F695 state is unknown, although photoprotection and far-red light absorption are both plausible assignments.

Two asymmetric features in the RC bear comment. First, Chl A1 of the D1 protein (Chl P<sub>D1</sub> in the notation of Ref.<sup>66</sup>) exhibits a negative doming of roughly  $-0.37$  Å, somewhat larger than its symmetry partner in the D2 protein (for which  $d_3 \approx -0.13$  Å). However, the asymmetry between these two coefficients is only slightly above the cutoff of  $\sim 0.15$  Å below which refinement model choices are likely to have a significant impact on observed structures. Moreover, the doming asymmetry within this pair has been pointed out before in conjunction with high-level quantum mechanics/molecular mechanics (QM/MM) analysis and was concluded to be unlikely to have a strong impact on RC function.<sup>50</sup> A somewhat stronger asymmetry is observed between the two pheophytins, with  $d_3 \approx 0.02$  for site A3

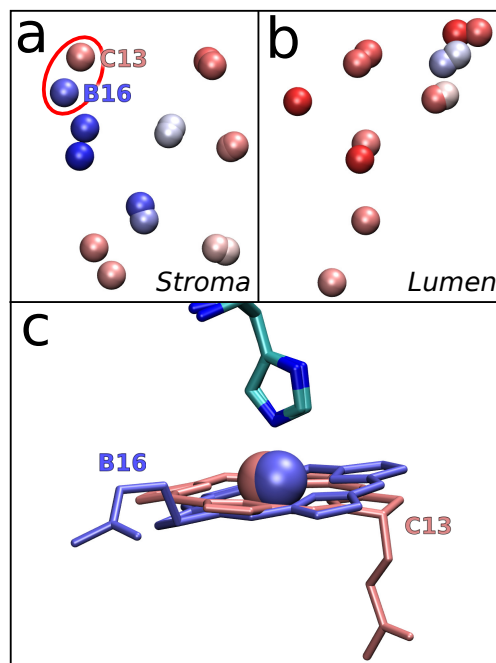


Figure 11: Ring deformation in CP43 compared with CP47 Chls. Frames a and b presents overlays of the stromal and lumenal CP43 and CP47 Chl pigments from Figure 10 after rotating the CP43 structure by 180° to bring it into alignment with CP47. Frame c compares the geometries of the B16 and C13 pigments after aligning their respective His ligands. A 180° flip of the Chl ring results in an inversion of the respective doming coordinates.

(Pheo D<sub>1</sub> in Ref.<sup>66</sup>) and  $d_3 \approx 0.31 \text{ \AA}$  for site D3 (Pheo D<sub>2</sub> in Ref.<sup>66</sup>). This difference is significantly larger than our estimated uncertainty of  $0.15 \text{ \AA}$  and may merit further investigation into potential functional significance.

However, the strongest asymmetry in PSII doming occurs outside the RC in the CP43 and CP47 core antennas. Figure 11 illustrates this feature by overlaying the CP43 and CP47 pigments after rotation by  $180^\circ$  around the central pseudo-C2 axis. Frame a shows the stromal pool of pigments, while Frame b presents the luminal layer. Note that the Mg atoms have not been aligned through any best-fit procedure here; only a  $180^\circ$  degree rotation is necessary to bring the Mg centers into close overlap so that corresponding pigment pairs may be identified directly by eye. (The three unpaired Mg atoms in the Luminal layer are the “extra” sites of CP47, 1, 2 and 6.)

As is visible in the figure, most CP43/CP47 pigment pairs exhibit nearly identical doming values; the sole exception is the B16/C13 pair, where strong positive doming of the CP47 site ( $d_3 \approx 0.41 \text{ \AA}$ ) contrasts strongly with negative doming in the corresponding CP43 site ( $d_3 \approx -0.32 \text{ \AA}$ ). This observation is particularly striking in light of the fact that it is the B16 site that has been implicated by mutagenesis experiments in 695 nm fluorescence, which is present in CP47 but absent from CP43.<sup>74,75,77</sup>

What causes this change in ring deformation? Comparison of the two binding sites reveals that, although the two pigments are ligated by homologous His residues, the two pigments are flipped inside the binding pocket, so that the phytyl tail appears on the opposite side relative to the His ligand. The significance of such “syn” and “anti” (or  $\alpha$  and  $\beta$ ) ligand binding geometries has previously been highlighted by Balaban et al.,<sup>82</sup> although no simple conclusions have yet been reached as to when such effects may (or may not) play a significant role in biological function. In the present context, it is clear that at least one consequence of the switch from “anti” ligation in CP43 to “syn” ligation in CP47 is the reversal of the doming geometry of the site. The fact that this change occurs specifically at the B16 site that has traditionally been assigned as the origin of the F695 state (and that doming is otherwise

largely conserved between the two homologs) likewise suggests that such geometric changes may indeed play an important role in determining the electronic properties of individual Chl pigments.

## Site Energy Modeling

To assess such claims quantitatively, it would be highly desirable to have a computational method capable of accurately “translating” calculated ring deformation values into predicted site energies. The model of Zucchelli et al. seeks to do exactly this, assigning energetic shifts to each Chl electronic state based on the deformation energy of the mode with matching symmetry.<sup>48</sup> Unfortunately, these symmetry considerations imply that site energies calculated under this model are particularly sensitive to deformations along the  $wav(x)$  and prop modes, since these match the respective symmetries of the LUMO and HOMO electronic states. Since the prop mode in particular shows large uncertainty even down to 2.0 Å resolution, our results suggest that currently available structures are not sufficiently resolved to accurately estimate site energies under this model.

To quantify this expectation, Figure 12 shows calculated site energies using the Zucchelli model for each site in PSII, obtained from thermophilic cyanobacterial structures (both PDB and PDB-Redo) with resolutions better than 2.0 Å. Although the results show some improvement over those shown in Figure 4 for 3.0 Å-resolution structures, the site-to-site variation ( $122\text{ cm}^{-1}$ ) is still smaller than the RMS uncertainty ( $144\text{ cm}^{-1}$ ) at each site individually. Note that because the Zucchelli model depends only on the total deformation energy  $\varepsilon_k$  of each mode, it is insensitive to changes in sign;<sup>48</sup> thus (as seen in Figure 12) the “flipping” of the  $d_3$  coefficient between sites B16 and C13 does not affect the calculated site energies in this model.

It should be emphasized that these results do not invalidate the Zucchelli model: they simply indicate that current crystallographic data is not sufficiently accurate for it to be reliably applied to raw PSII structures. On the other hand, it is also possible that alternative

theoretical models could provide better performance, even at currently available resolutions, either through relaxation of the symmetry arguments on which the Zucchelli model relies (which are not strictly valid for Chl *a*) or through fortuitous correlations between low- and high-frequency deformations as has been observed for Hemes.<sup>32</sup>

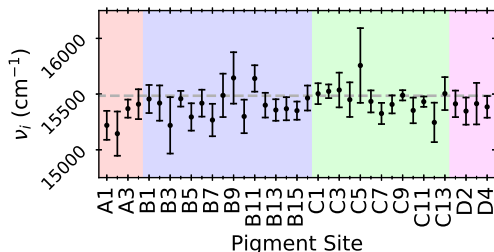


Figure 12: Pigment site energies in thermophilic cyanobacterial PSII estimated using the model of Zucchelli et al.<sup>48</sup> Black dots represent average values over all structures with resolution better than 2.0 Å; error bars indicate plus and minus one standard deviation.

## Conclusions

Perhaps the most significant point to emerge from this analysis is the importance of assessing input bias during the structure refinement process. The comparisons in Figures 5 to 7 between PDB and PDB-Redo structures highlight the fact that at moderate resolutions (2 - 3 Å), the geometries of individual ligands are strongly impacted by the refinement approach. In energetic terms, such discrepancies can easily represent hundreds or thousands of cm<sup>-1</sup> in predicted deformation energies. While no structure-refinement method can be entirely free of such biases, comparing the results of multiple refinement protocols – especially standardized protocols such as those offered by PDB-Redo – can at least serve to keep such input bias from going unnoticed and thus wrongly interpreted.

In terms of relevance to photosynthesis, by examining quantitatively the variability of ring deformation predictions across a large number of Chl protein structures, we have established approximate error bars for such predictions as a function of experimental resolution. Significantly, the relevant error bars are not the same for deformation along all ligand de-

degrees of freedom: higher-energy deformations tend to carry higher statistical error (in both real-space and energetic terms) than do lower-energy deformations, likely due to stronger sensitivity of the high-energy deformation coordinate values to user-defined input parameters during model refinement. Of the six modes examined here, only the three lowest-frequency modes (saddle, ruffle, and dome) carry error bars sufficiently low that robust conclusions can be drawn about their values in PSII and, even then, only for very high-resolution structures (better than 2 Å resolution).

Despite these limitations, several useful insights are obtained by comparing ring-deformation values across different species and between subunits within the PSII core. Specifically, ring deformation is found to be conserved across all species tested, at least for the three lowest-frequency modes; and the doming deformation is well conserved between the CP43 and CP47 homologs with the notable exception of the Chl B16/Chl C13 pair. Due to a “flipping” of the pigment plane, strong negative doming of Chl C13 in CP43 is converted into strong positive doming in CP47. Given the proposed (though not universally accepted) association of the B16 site of CP47 with 695 nm fluorescence, these observations support the notion that the Chl doming deformation may play a role in tuning Chl site energies. Unfortunately, given the sensitivity of existing site energy models (such as that of Zucchelli et al.<sup>48</sup>) to structural uncertainty, quantifying such effects must likely await either alternate theoretical and computational tools or even higher structural resolutions than are currently available.

## Acknowledgement

M.R. acknowledges support from Purdue University and from the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under award #DE-SC0022884. Y.M. acknowledges the support by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award #DE-SC0022956

## References

- (1) Sabe, V. T.; Ntombela, T.; Jhamba, L. A.; Maguire, G. E.; Govender, T.; Naicker, T.; Kruger, H. G. Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *European Journal of Medicinal Chemistry* **2021**, *224*, 113705.
- (2) Schneider, P.; Walters, W. P.; Plowright, A. T.; Sieroka, N.; Listgarten, J.; Goodnow, R. A.; Fisher, J.; Jansen, J. M.; Duca, J. S.; Rush, T. S.; et al. Rethinking drug design in the artificial intelligence era. *Nature Reviews Drug Discovery* **2020**, *19*, 353–364.
- (3) Wang, Z.; Sun, H.; Shen, C.; Hu, X.; Gao, J.; Li, D.; Cao, D.; Hou, T. Combined strategies in structure-based virtual screening. *Phys. Chem. Chem. Phys.* **2020**, *22*, 3149–3159.
- (4) Kazlauskas, R. J.; Bornscheuer, U. T. Finding better protein engineering strategies. *Nature Chemical Biology* **2009**, *5*, 526–529.
- (5) Chen, K.; Arnold, F. H. Engineering new catalytic activities in enzymes. *Nature Catalysis* **2020**, *3*, 203–213.
- (6) Baker, D. An exciting but challenging road ahead for computational enzyme design. *Protein Science* **2010**, *19*, 1817–1819.
- (7) Huang, P.-S.; Boyken, S. E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537*, 320–327.
- (8) Koshland Jr., D. E. The Key–Lock Theory and the Induced Fit Theory. *Angewandte Chemie International Edition in English* **1995**, *33*, 2375–2378.
- (9) Ma, B.; Kumar, S.; Tsai, C.-J.; Nussinov, R. Folding funnels and binding mechanisms. *Protein Engineering, Design and Selection* **1999**, *12*, 713–720.

- (10) Csermely, P.; Palotai, R.; Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* **2010**, *35*, 539–546.
- (11) Clore, G. M. Interplay between conformational selection and induced fit in multidomain protein–ligand binding probed by paramagnetic relaxation enhancement. *Biophysical Chemistry* **2014**, *186*, 3–12, Special issue : conformational selection.
- (12) A. Shelnutt, J.; Song, X.-Z.; Ma, J.-G.; Jia, S.-L.; Jentzen, W.; J. Medforth, C.; J. Medforth, C. Nonplanar porphyrins and their significance in proteins. *Chem. Soc. Rev.* **1998**, *27*, 31–42.
- (13) Li, D.; Stuehr, D. J.; Yeh, S.-R.; Rousseau, D. L. Heme Distortion Modulated by Ligand-Protein Interactions in Inducible Nitric-oxide Synthase. *Journal of Biological Chemistry* **2004**, *279*, 26489–26499.
- (14) Meijers, R.; Cedergren-Zeppezauer, E. A variety of electrostatic interactions and adducts can activate NAD(P) cofactors for hydride transfer. *Chemico-Biological Interactions* **2009**, *178*, 24–28, Enzymology and Molecular Biology of Carbonyl Metabolism.
- (15) Kubota, K.; Shingae, T.; Foster, N. D.; Kumauchi, M.; Hoff, W. D.; Unno, M. Active Site Structure of Photoactive Yellow Protein with a Locked Chromophore Analogue Revealed by Near-Infrared Raman Optical Activity. *The Journal of Physical Chemistry Letters* **2013**, *4*, 3031–3038.
- (16) Shingae, T.; Kubota, K.; Kumauchi, M.; Tokunaga, F.; Unno, M. Raman Optical Activity Probing Structural Deformations of the 4-Hydroxycinnamyl Chromophore in Photoactive Yellow Protein. *The Journal of Physical Chemistry Letters* **2013**, *4*, 1322–1327, PMID: 26282147.
- (17) MacGowan, S. A.; Senge, M. O. Computational Quantification of the Physicochemical

- Effects of Heme Distortion: Redox Control in the Reaction Center Cytochrome Subunit of *Blastochloris viridis*. *Inorganic Chemistry* **2013**, *52*, 1228–1237, PMID: 23320527.
- (18) Neumann, P.; Tittmann, K. Marvels of enzyme catalysis at true atomic resolution: distortions, bond elongations, hidden flips, protonation states and atom identities. *Current Opinion in Structural Biology* **2014**, *29*, 122–133, Catalysis and regulation / Multi-protein assemblies in signalling.
- (19) Senge, M. O.; MacGowan, S. A.; O'Brien, J. M. Conformational control of cofactors in nature – the influence of protein-induced macrocycle distortion on the biological function of tetrapyrroles. *Chem. Commun.* **2015**, *51*, 17031–17063.
- (20) Engh, R. A.; Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *urn:issn:0108-7673* **1991**, *47*, 392–400.
- (21) Vagin, A. A.; Steiner, R. A.; Lebedev, A. A.; Potterton, L.; McNicholas, S.; Long, F.; Murshudov, G. N. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta crystallographica. Section D, Biological crystallography* **2004**, *60*, 2184–2195.
- (22) Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta crystallographica. Section D, Biological crystallography* **2009**, *65*, 1074–1080.
- (23) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta crystallographica Section B, Structural science, crystal engineering and materials* **2016**, *72*, 171–179.
- (24) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhuntsev, A.; Zwart, P. H.; Adams, P. D. Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography* **2012**, *68*, 352–367.

- (25) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **2011**, *67*, 355–367.
- (26) Liebschner, D.; Moriarty, N. W.; Poon, B. K.; Adams, P. D. In situ ligand restraints from quantum-mechanical methods. *Acta crystallographica. Section D, Structural biology* **2023**, *79*, 100–110.
- (27) Long, F.; Nicholls, R. A.; Emsley, P.; Gražulis, S.; Merkys, A.; Vaitkus, A.; Murshudov, G. N. AceDRG: A stereochemical description generator for ligands. *Acta Crystallographica Section D: Structural Biology* **2017**, *73*, 112–122.
- (28) Renner, M. W.; Barkigia, K. M.; Zhang, Y.; Medforth, C. J.; Smith, K. M.; Fajer, J. Consequences of Oxidation in Nonplanar Porphyrins: Molecular Structure and Diamagnetism of the .pi. Cation Radical of Copper(II) Octaethyltetraphenylporphyrin. *Journal of the American Chemical Society* **1994**, *116*, 8582–8592.
- (29) Jentzen, W.; Ma, J.-G.; Shelnut, J. A. Conservation of the Conformation of the Porphyrin Macrocycle in Hemoproteins. *Biophysical Journal* **1998**, *74*, 753–763.
- (30) Ma, J.-G.; Zhang, J.; Franco, R.; Jia, S.-L.; Moura, I.; Moura, J. J. G.; Kroneck, P. M. H.; Shelnut, J. A. The Structural Origin of Nonplanar Heme Distortions in Tetraheme Ferricytochromes c3. *Biochemistry* **1998**, *37*, 12431–12442, PMID: 9730815.
- (31) Kiefl, C.; Sreerama, N.; Haddad, R.; Sun, L.; Jentzen, W.; Lu, Y.; Qiu, Y.; Shelnut, J. A.; Woody, R. W. Heme Distortions in Sperm-Whale Carbonmonoxy Myoglobin: Correlations between Rotational Strengths and Heme Distortions in MD-Generated Structures. *Journal of the American Chemical Society* **2002**, *124*, 3385–3394, PMID: 11916424.

- (32) Haddad, R. E.; Gazeau, S.; Pécaut, J.; Marchon, J.-C.; Medforth, C. J.; Shelnut, J. A. Origin of the Red Shifts in the Optical Absorption Bands of Nonplanar Tetraalkylporphyrins. *Journal of the American Chemical Society* **2003**, *125*, 1253–1268, PMID: 12553827.
- (33) Laberge, M.; Huang, Q.; Schweitzer-Stenner, R.; Fidy, J. The Endogenous Calcium Ions of Horseradish Peroxidase C Are Required to Maintain the Functional Nonplanarity of the Heme. *Biophysical Journal* **2003**, *84*, 2542–2552.
- (34) Huang, Q.; Medforth, C. J.; Schweitzer-Stenner, R. Nonplanar Heme Deformations and Excited State Displacements in Nickel Porphyrins Detected by Raman Spectroscopy at Soret Excitation. *The Journal of Physical Chemistry A* **2005**, *109*, 10493–10502, PMID: 16834304.
- (35) Schweitzer-Stenner, R.; Huang, Q.; Hagarman, A.; Laberge, M.; Wallace Static Normal Coordinate Deformations of the Heme Group in Mutants of Ferrocyanochrome c from *Saccharomyces cerevisiae* Probed by Resonance Raman Spectroscopy. *The Journal of Physical Chemistry B* **2007**, *111*, 6527–6533, PMID: 17508736.
- (36) Hagarman, A.; Wallace, C. J.; Laberge, M. M.; Schweitzer-Stenner, R. Out-of-plane deformations of the heme group in different ferrocyanochrome c proteins probed by resonance Raman spectroscopy. *Journal of Raman Spectroscopy* **2008**, *39*, 1848–1858.
- (37) Gruia, F.; Kubo, M.; Ye, X.; Ionascu, D.; Lu, C.; Poole, R. K.; Yeh, S.-R.; Champion, P. M. Coherence Spectroscopy Investigations of the Low-Frequency Vibrations of Heme: Effects of Protein-Specific Perturbations. *Journal of the American Chemical Society* **2008**, *130*, 5231–5244, PMID: 18355013.
- (38) Bikiel, D. E.; Forti, F.; Boechi, L.; Nardini, M.; Luque, F. J.; Martí, M. A.; Estrin, D. A. Role of Heme Distortion on Oxygen Affinity in Heme Proteins: The Protoglobin Case. *The Journal of Physical Chemistry B* **2010**, *114*, 8536–8543, PMID: 20524694.

- (39) Sun, Y.; Benabbas, A.; Zeng, W.; Kleingardner, J. G.; Bren, K. L.; Champion, P. M. Investigations of heme distortion, low-frequency vibrational excitations, and electron transfer in cytochrome c. *Proceedings of the National Academy of Sciences* **2014**, *111*, 6570–6575.
- (40) The Liptak Research Group. <https://mliptak.w3.uvm.edu/nsd.html>, Accessed: 2023-07-31.
- (41) Jentzen, W.; Song, X.-Z.; Shelnut, J. A. Structural Characterization of Synthetic and Protein-Bound Porphyrins in Terms of the Lowest-Frequency Normal Coordinates of the Macrocycle. *The Journal of Physical Chemistry B* **1997**, *101*, 1684–1699.
- (42) Graves, A. B.; Graves, M. T.; Liptak, M. D. Measurement of Heme Ruffling Changes in MhuD Using UV–vis Spectroscopy. *The Journal of Physical Chemistry B* **2016**, *120*, 3844–3853, PMID: 27035523.
- (43) Kondo, H. X.; Kanematsu, Y.; Masumoto, G.; Takano, Y. PyDISH: database and analysis tools for heme porphyrin distortion in heme proteins. *Database* **2020**, baaa066.
- (44) Kondo, H. X.; Fujii, M.; Tanioka, T.; Kanematsu, Y.; Yoshida, T.; Takano, Y. Global Analysis of Heme Proteins Elucidates the Correlation between Heme Distortion and the Heme-Binding Pocket. *Journal of Chemical Information and Modeling* **2022**, *62*, 775–784, PMID: 35157473.
- (45) Tronrud, D.; Schmid, M.; Matthews, B. Structure and X-ray amino acid sequence of a bacteriochlorophyll a protein from *Prosthecochloris aestuarii* refined at 1.9 Å resolution. *Journal of Molecular Biology* **1986**, *188*, 443–454.
- (46) Barkigia, K. M.; Chantranupong, L.; Smith, K. M.; Fajer, J. Structural and theoretical models of photosynthetic chromophores. Implications for redox, light-absorption properties and vectorial electron flow. *Journal of the American Chemical Society* **1988**, *110*, 7566–7567.

- (47) Hutter, M. C.; Hughes, J. M.; Reimers, J. R.; Hush, N. S. Modeling the Bacterial Photosynthetic Reaction Center. 2. A Combined Quantum Mechanical/Molecular Mechanical Study of the Structure of the Cofactors in the Reaction Centers of Purple Bacteria. *The Journal of Physical Chemistry B* **1999**, *103*, 4906–4915.
- (48) Zucchelli, G.; Brogioli, D.; Casazza, A. P.; Garlaschi, F. M.; Jennings, R. C. Chlorophyll Ring Deformation Modulates Qy Electronic Energy in Chlorophyll-Protein Complexes and Generates Spectral Forms. *Biophysical Journal* **2007**, *93*, 2240–2254.
- (49) MacGowan, S. A.; Senge, M. O. Conformational control of cofactors in nature—functional tetrapyrrole conformations in the photosynthetic reaction centers of purple bacteria. *Chem. Commun.* **2011**, *47*, 11621–11623.
- (50) Saito, K.; Umena, Y.; Kawakami, K.; Shen, J.-R.; Kamiya, N.; Ishikita, H. Deformation of Chlorin Rings in the Photosystem II Crystal Structure. *Biochemistry* **2012**, *51*, 4290–4299, PMID: 22568617.
- (51) Zucchelli, G.; Santabarbara, S.; Jennings, R. C. The Qy Absorption Spectrum of the Light-Harvesting Complex II As Determined by Structure-Based Analysis of Chlorophyll Macrocycle Deformations. *Biochemistry* **2012**, *51*, 2717–2736, PMID: 22417459.
- (52) Senge, M. O.; Ryan, A. A.; Letchford, K. A.; MacGowan, S. A.; Mielke, T. Chlorophylls, Symmetry, Chirality, and Photosynthesis. *Symmetry* **2014**, *6*, 781–843.
- (53) Bednarczyk, D.; Dym, O.; Prabakar, V.; Peleg, Y.; Pike, D. H.; Noy, D. Fine Tuning of Chlorophyll Spectra by Protein-Induced Ring Deformation. *Angewandte Chemie International Edition* **2016**, *55*, 6901–6905.
- (54) MacGowan, S. A.; Senge, M. O. Contribution of bacteriochlorophyll conformation to the distribution of site-energies in the FMO protein. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2016**, *1857*, 427–442.

- (55) Lahav, Y.; Noy, D.; Schapiro, I. Spectral tuning of chlorophylls in proteins – electrostatics vs. ring deformation. *Phys. Chem. Chem. Phys.* **2021**, *23*, 6544–6551.
- (56) Nishikawa, G.; Sugo, Y.; Saito, K.; Ishikita, H. Absence of electron-transfer-associated changes in the time-dependent X-ray free-electron laser structures of the photosynthetic reaction center. *eLife* **2023**, *12*, RP88955.
- (57) Dods, R.; Båth, P.; Morozov, D.; Gagnér, V. A.; Arnlund, D.; Luk, H. L.; Kübel, J.; Maj, M.; Vallejos, A.; Wickstrand, C.; et al. Ultrafast structural changes within a photosynthetic reaction centre. *Nature* **2021**, *589*, 310–314.
- (58) Fu, Z.; Li, X.; Merz Jr., K. M. Accurate assessment of the strain energy in a protein-bound drug using QM/MM X-ray refinement and converged quantum chemistry. *Journal of Computational Chemistry* **2011**, *32*, 2587–2597.
- (59) Vant, J. W.; Lahey, S.-L. J.; Jana, K.; Shekhar, M.; Sarkar, D.; Munk, B. H.; Kleinekathöfer, U.; Mittal, S.; Rowley, C.; Singharoy, A. Flexible Fitting of Small Molecules into Electron Microscopy Maps Using Molecular Dynamics Simulations with Neural Network Potentials. *Journal of Chemical Information and Modeling* **2020**, *60*, 2591–2604, PMID: 32207947.
- (60) Umena, Y.; Kawakami, K.; Shen, J.-R.; Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* **2011**, *473*, 55–60.
- (61) Joosten, R. P.; Long, F.; Murshudov, G. N.; Perrakis, A. The *PDB-REDO* server for macromolecular structure model optimization. *IUCrJ* **2014**, *1*, 213–220.
- (62) van Beusekom, B.; Touw, W. G.; Tatineni, M.; Somani, S.; Rajagopal, G.; Luo, J.; Gilliland, G. L.; Perrakis, A.; Joosten, R. P. Homology-based hydrogen bond information improves crystallographic structures in the PDB. *Protein Science* **2018**, *27*, 798–808.

- (63) Joosten, R. P.; Womack, T.; Vriend, G.; Bricogne, G. Re-refinement from deposited X-ray data can deliver improved models for most PDB entries. *Acta Crystallographica Section D* **2009**, *65*, 176–185.
- (64) Ahad, S.; Lin, C.; Reppert, M. E. Photosynthetic Protein Spectroscopy Lab. October 1, 2020; <https://nanohub.org/resources/pigmenthunter>.
- (65) Liebschner, D.; Afonine, P. V.; Moriarty, N. W.; Poon, B. K.; Chen, V. B.; Adams, P. D. CERES: a cryo-EM re-refinement system for continuous improvement of deposited models. *Acta crystallographica. Section D, Structural biology* **2021**, *77*, 48–61.
- (66) Müh, F.; Zouni, A. Structural basis of light-harvesting in the photosystem II core complex. *Protein Science* **2020**, *29*, 1090–1119.
- (67) Liu, Z.; Yan, H.; Wang, K.; Kuang, T.; Zhang, J.; Gui, L.; An, X.; Chang, W. Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature* **2004**, *428*, 287–292.
- (68) Kern, J.; Chatterjee, R.; Young, I. D.; Fuller, F. D.; Lassalle, L.; Ibrahim, M.; Gul, S.; Fransson, T.; Brewster, A. S.; Alonso-Mori, R.; et al. Structures of the intermediates of Kok’s photosynthetic water oxidation clock. *Nature* **2018**, *563*, 421–425.
- (69) Bhowmick, A.; Hussein, R.; Bogacz, I.; Simon, P. S.; Ibrahim, M.; Chatterjee, R.; Doyle, M. D.; Cheah, M. H.; Fransson, T.; Chernev, P.; et al. Structural evidence for intermediates during O<sub>2</sub> formation in photosystem II. *Nature* **2023**, *617*, 629–636.
- (70) Gisriel, C. J.; Wang, J.; Liu, J.; Flesher, D. A.; Reiss, K. M.; Huang, H.-L.; Yang, K. R.; Armstrong, W. H.; Gunner, M. R.; Batista, V. S.; Debus, R. J.; Brudvig, G. W. High-resolution cryo-electron microscopy structure of photosystem II from the mesophilic cyanobacterium, *Synechocystis* sp. PCC 6803. *Proceedings of the National Academy of Sciences* **2022**, *119*, e2116765118.

- (71) van Bezouwen, L. S.; Caffarri, S.; Kale, R. S.; Kouřil, R.; Thunnissen, A.-M. W. H.; Oostergetel, G. T.; Boekema, E. J. Subunit and chlorophyll organization of the plant photosystem II supercomplex. *Nature Plants* **2017**, *3*, 17080.
- (72) Graça, A. T.; Hall, M.; Persson, K.; Schröder, W. P. High-resolution model of Arabidopsis Photosystem II reveals the structural consequences of digitonin-extraction. *Scientific Reports* **2021**, *11*, 15534.
- (73) Kalaji, H. M.; Schansker, G.; Brestic, M.; Bussotti, F.; Calatayud, A.; Ferroni, L.; Goltsev, V.; Guidi, L.; Jajoo, A.; Li, P.; et al. Frequently asked questions about chlorophyll fluorescence, the sequel. *Photosynthesis Research* **2017**, *132*, 13–66.
- (74) Shen, G.; Vermaas, W. F. J. Mutation of Chlorophyll Ligands in the Chlorophyll-Binding CP47 Protein As Studied in a Synechocystis sp. PCC 6803 Photosystem I-less Background. *Biochem.* **1994**, *33*, 7379–7388.
- (75) de Weerd, F. L.; Palacios, M. A.; Andrizhiyevskaya, E. G.; Dekker, J. P.; van Gronnelle, R. Identifying the Lowest Electronic States of the Chlorophylls in the CP47 Core Antenna Protein of Photosystem II. *Biochem.* **2002**, *41*, 15224–15233, PMID: 12484760.
- (76) Raszewski, G.; Renger, T. Light Harvesting in Photosystem II Core Complexes Is Limited by the Transfer to the Trap: Can the Core Complex Turn into a Photoprotective Mode? *J. Amer. Chem. Soc.* **2008**, *130*, 4431–4446, PMID: 18327941.
- (77) D’Haene, S. E.; Sobotka, R.; Bučinská, L.; Dekker, J. P.; Komenda, J. Interaction of the PsbH subunit with a chlorophyll bound to histidine 114 of CP47 is responsible for the red 77K fluorescence of Photosystem II. *Biochim. Biophys. Acta, Bioenerg.* **2015**, *1847*, 1327 – 1334.
- (78) Reppert, M.; Acharya, K.; Neupane, B.; Jankowiak, R. Lowest Electronic States of the CP47 Antenna Protein Complex of Photosystem II: Simulation of Optical Spectra and Revised Structural Assignments. *J. Phys. Chem. B* **2010**, *114*, 11884–11898.

- (79) Hall, J.; Renger, T.; Müh, F.; Picorel, R.; Krausz, E. The lowest-energy chlorophyll of photosystem II is adjacent to the peripheral antenna: Emitting states of CP47 assigned via circularly polarized luminescence. *Biochim. Biophys. Acta (BBA) - Bioenergetics* **2016**, *1857*, 1580 – 1593.
- (80) Reinot, T.; Chen, J.; Kell, A.; Jassas, M.; Robben, K. C.; Zazubovich, V.; Jankowiak, R. On the Conflicting Estimations of Pigment Site Energies in Photosynthetic Complexes: A Case Study of the CP47 Complex. *Anal. Chem. Insights* **2016**, *11*, 35–48, aci-11-2016-035[PII].
- (81) Jassas, M.; Reinot, T.; Kell, A.; Jankowiak, R. Toward an Understanding of the Excitonic Structure of the CP47 Antenna Protein Complex of Photosystem II Revealed via Circularly Polarized Luminescence. *J. Phys. Chem. B* **2017**, *121*, 4364–4378, PMID: 28394609.
- (82) Balaban, T. S.; Fromme, P.; Holzwarth, A. R.; Krauß, N.; Prokhorenko, V. I. Relevance of the diastereotopic ligation of magnesium atoms of chlorophylls in Photosystem I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2002**, *1556*, 197–207.

# TOC Graphic

