Supplementary Information for

Quantitative assessment of chlorophyll types in cryo-EM maps of photosystem I adapted to far-red light

Christopher J. Gisriel[1,4,*], Hao-Li Huang[1,4], Krystle Reiss[1], David A. Flesher[2], Victor S. Batista[1], Donald A. Bryant[3], Gary W. Brudvig[1,2], and Jimin Wang[3,*]

[1]Department of Chemistry, Yale University, New Haven, CT 06520, USA.

[2]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA.

[3]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA.

[4]These authors contributed equally.

*To whom correspondence should be addressed: christopher.gisriel@yale.edu and jimin.wang@yale.edu

Supplementary Text

Supplementary Text 1. Validation of the cone scan method. To validate the cone-scan method, we calculated the energy profiles of the C8, C7, and C3 substituents as a function of orientation (torsion angle) in DFT energy-minimized Chl models (**Supplementary Fig. 6**) and compared them with the corresponding cone scans of the ESP from all modeled Chl sites in the two FRL-PSI structures (**Supplementary Fig. 7**).

Calculated energies of Chl substituents as a function of orientation. DFT energy-minimized structures of Chl *f* (**Supplementary Movie 1**) and Chl *a* (**Supplementary Movie 2**) were generated, and the calculated free energies associated with the C8, C7, C3, and C2 substituents at 72 different torsion angles were calculated (see **Supplementary Fig. 6** and **Methods**). Energies associated with the C8 ethyl moiety and C7 methyl moiety are similar for Chl *f* and Chl *a*. For the position of the ethyl moiety at C8, the energy is minimal at 95° or 265°, where the ethyl group is oriented perpendicular to the tetrapyrrole ring to minimize steric interactions. For the methyl group at C7, three small peaks are observed for preferential orientations of the hydrogen atoms, with small energy differences. The energy profile of the C3 vinyl substituent of Chl *a* is more complex compared to the other three substituents (**Supplementary Fig. 6**). The sp^2 hybridization of C3 and avoidance of steric clashing cause energy minima at 30° and 145° when the vinyl group is oriented toward one face of the ring and energy minima at 215° and 330° when it is oriented toward the other face. When the C3 formyl of Chl *a* is directed closer to C2 (145° or 220°), the energy of the methyl at C2 is slightly increased. While this is a trivial point when considering expectations of substituent orientations in cryo-EM maps containing Chl *a*, it suggests that in Chl *f*, the C3 vinyl and C2 formyl substituents may strongly influence one another. Our calculations agree with previous energy calculations that simulated the Chl *a* Q_y spectrum as a function of the orientation of substituents [1].

Indeed, an orientation dependence is observed in the energy scans for the C2 and C3 substituents of Chl *f*. The C2 position of Chl *f* favors orientations planar to the tetrapyrrole ring. When the C3 vinyl substituent is oriented at 140° or 210°, away from C2 (see the top row of **Fig. 1** for orientation), energy minima for C2 are at 0° and 180°. When the C3 vinyl substituent is oriented at 35° or 325°, toward C2, the C2 energy increases at both minima, and the 0° minimum is shifted to ~15° to avoid steric clashing. The C3 vinyl substituent energy scan of Chl *f* is similar to that of Chl *a* when the C2 formyl substituent is oriented at 0°, away from C3. However, the C3 energy increases overall when the C2 formyl is oriented at 180°, toward C3.

The DFT-calculated free energy analysis provides insight into intra-molecular interaction energies of Chl substituents but it neglects inter-molecular interactions with the protein environment that are likely to impact substituent orientation. For example, it has been pointed out that formyl substituents on Chl *b* in peripheral antenna complexes almost always accept H-bond donors from their environment [2], as is also the case for Chl *f* [3]; however, the energy scans allow for reasonable approximations for Chl substituent orientations. In general, these DFT energy scans suggest four main points that may be expected in the experimental cryo-EM maps: (1) the C8 ethyl moieties are likely to be observed perpendicular to the tetrapyrrole ring; (2) the C3 vinyl moieties are likely to exhibit a wider variation in their orientations compared to C8 ethyl; (3) the C2 formyl substituent of Chl *f* is likely to be found nearly in plane with the tetrapyrrole ring; and (4) the orientations of the C2 and C3 Chl *f* moieties strongly influence one another. Future studies

may be enhanced by including a variety of possible contributions from the protein environment that may contribute to its ESP, especially for use in generating Chl restraints for use in cryo-EM maps.

Cone scans of C8, C7, and C3 substituents, and comparison with DFT energy profiles. Cone scans of the C8, C7, and C3 substituents were generated from both structures (**Supplementary Fig. 7)** with scan parameters corresponding to their identity as found in **Supplementary Table 1** and binned by resolution. We note that the local resolutions cover a broader range in the *H. hongdechloris* data than the *F. thermalis* data, therefore we split the data sets into three and two local resolution bins, respectively.

The C8 ethyl and C3 vinyl cone scans exhibit average amplitudes greater than that of the C7 methyl cone scans, suggesting that ESP contribution from the additional carbon atom of ethyl or vinyl can be differentiated from a methyl moiety at all resolutions examined here. The average of the standard deviation for a group of scans at higher resolution is consistently greater than that at lower resolution (**Supplementary Fig. 7**). Despite the larger average standard deviation in higher resolution bins, the peak signals are more pronounced than at lower resolution; therefore, it is clear that higher resolutions better resolve substituent identities and orientations as expected.

At Chl sites where the local resolution is better than ~2.6 Å, which are only present in the *H. hongdechloris* data, the peak distributions appear inversely correlated with the DFT free energy profiles (**Supplementary Fig. 7**), suggesting that substituent orientations predicted to exhibit low energies by our DFT calculations (**Supplementary Fig. 6**) are consistent with their true orientations found in the sample. This also implies that ~2.6-Å resolution is a reasonable cutoff for being able to determine the orientation of an ethyl or vinyl moiety. At resolutions better than ~2.6 Å, the C8 ethyl substituent exhibits two easily discernable peak distributions out of plane with the tetrapyrrole ring and the C3 vinyl cone scans exhibit a broader peak distribution relative to the C8 ethyl cone scans, both of which were predicted by our DFT energy calculations. There is a tendency for the C3 vinyl to be oriented toward the C2 position which is probably because the majority of Chls in the structure are Chl *a* where there is more open space when the vinyl is oriented toward C2 rather than when it is oriented toward C5. At resolutions lower than ~2.6 Å, the standard deviation is relatively large, rendering the peak distributions relatively unreliable, and thus orientation challenging to determined.

Because the cone scan results resolve the additional ESP from an ethyl or formyl moiety versus a methyl moiety, it suggests that formyl moieties at C2 may also be differentiated from methyl substituents, possibly allowing for direct detection of Chl *f*-binding sites; however, the ESP from a partial negative charge on the formyl oxygen is theoretically less than that for a neutral carbon atom on an ethyl or formyl moiety or for a neutral O atom, which would increase the resolution requirement to differentiate the substituents as described previously [3] and as is apparent in our DFT-derived ESP calculations (**Fig. 1d**).

Supplementary Text 2. Sites that exhibit C2 peaks but are unlikely Chl *f* candidates. In the *H. hongdechloris* C2 cone scans, scans for sites A8, A17, B9, and A10 exhibit peaks above the methyl distribution but are unlikely Chl *f* sites. In their C2 cone scans, all four sites exhibit a broad peak centered at improbable angles for a formyl substituent. In all these cases, the C2 position is very near the tetrapyrrole ring of another Chl site (**Supplementary Fig. 11**). These false-positive

examples suggest that evidence for Chl *f* from cone scan results must also be consistent with a reasonable chemical environment.

In the *F. thermalis* C2 cone scans, two Chl sites, A10 and A38, exhibit peaks above the methyl distribution but are unlikely Chl *f* candidates. Like the four false-positives from *H. hongdechloris*, the C2 position of A10 from *F. thermalis* is nearby Chl A18 (**Supplementary Fig. 12**) and its C2 cone scan peak is broad and centered ~105°, making it an unlikely Chl *f* candidate. The C2 cone scan for Chl A38 is even broader and centered ~85°. Unlike the rest of the Chl sites described in this section, the A38 C2 cone scan is not nearby the main ring of another tetrapyrrole. Instead, the additional ESP for the A38 C2 position appears to be merging with a nearby Val residue (**Supplementary Fig. 12**), suggesting that the C2 cone scan peak may arise from its close proximity to another nearby entity rather than arising from a formyl substituent.

Supplementary Text 3. Other previously proposed Chl *f* sites. Chl site A21 in both FRL-PSI structures is found in a π-stacked Chl dimer with Chl A20. Spectroscopic results suggest that FRL-PSI in various species likely have at least one Chl *f* dimer [4–6], and Chl *b* sites in LHC proteins are often found in a dimeric configuration [3]. To provide a possible Chl *f* dimer candidate, it was suggested that A20 was a Chl *f* in a dimeric arrangement with Chl A21 because the axial ligation of A21 was unclear in the 3.19-Å resolution structure of FRL-PSI from *F. thermalis*. Gisriel et al. proposed that if A20 is Chl *f*, its formyl substituent could be in a H-bonding network with a water that might provide axial ligation to A21 [7]. In the newly refined structure of FRL-PSI from *F. thermalis* and that from *H. hongdechloris*, it seems clear that the axial ligand of A21 is a water molecule on the opposite side of the tetrapyrrole ring from the previously suggested H-bonding network. Neither of the C2 cone scans for either structure exhibit peaks that suggest direct evidence of a formyl substituent, but this site is even lower resolution than A21, and thus does not exclude A21 from being occupied by Chl *f*, though there is presently no direct evidence to support it and the indirect evidence also seems weak.

In the FRL-PSI structure from *H. hongdechloris*, seven Chl *f* sites were assigned: A23, A24, A27, A29, B7, B22, and B37 [8]. Kato et al. made these assignments based on comparing experimental density differences compared to a structure of PSI from the same organism grown in white light in which all Chls are Chl *a*, and also by comparing the FRL-PSI ESP map to calculated ESP maps of Chl *a*. While this approach is logical, there were technical problems as discussed previously [3], and our quantitative cone scan analyses suggest that three of these sites, A23, B7 and B37, exhibit direct evidence for binding Chl *f* as discussed above. The other four Chls, A24, A27, A29, and B22, do not appear to have significantly different ESP signals compared to the methyl distribution, and thus are unlikely candidates for Chl *f* binding. This is consistent with them also having unlikely environments for a formyl substituent near their C2 position as noted previously [3]. Though we cannot rule out Chl *f* at low occupancy in these sites, there is no direct or indirect evidence for their assignment as binding Chl *f*.

Supplementary Figures

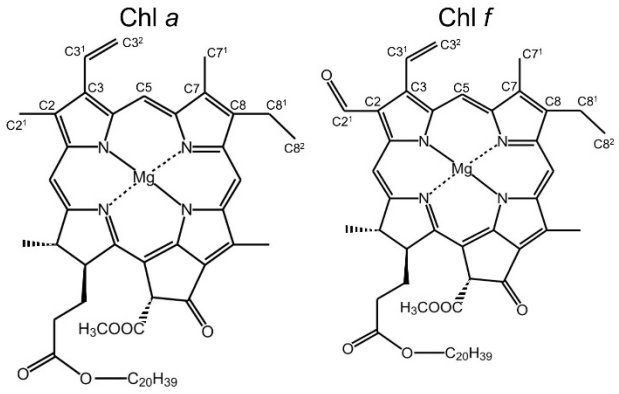

Fig. S1. Structures of Chl *a* and Chl *f*. Notable Chl positions discussed in the text are labelled on the Chl *a* model.

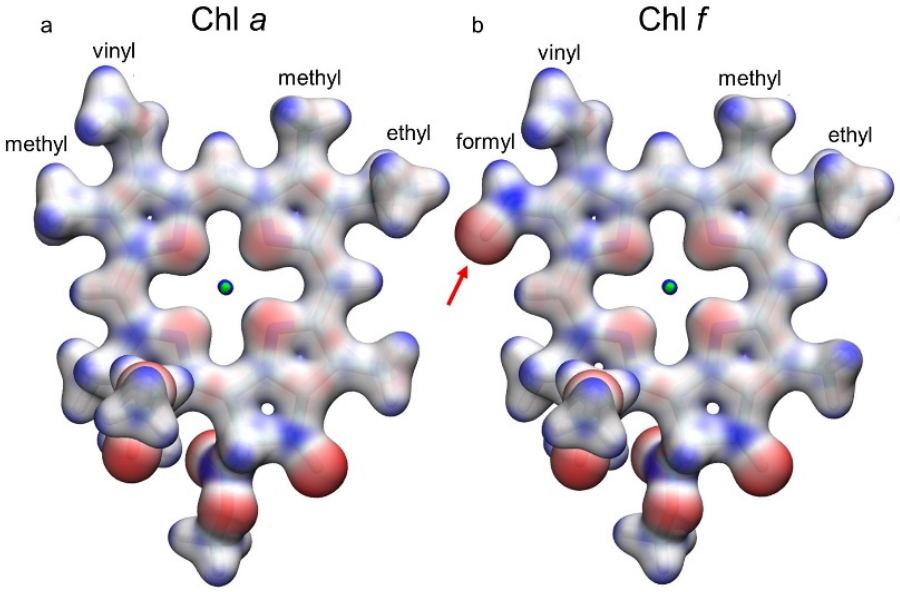

Fig. S2. Mapping DFT-derived ESP for Chl *a* (a) and Chl *f* (b) on van der Waals surfaces of neutral atoms of the molecules. ESP values with negative, close to zero, and positive values are in red, gray, and blue, respectively. The red arrow shows the O atom of the formyl group of Chl *f* where there are notable negative ESP values.

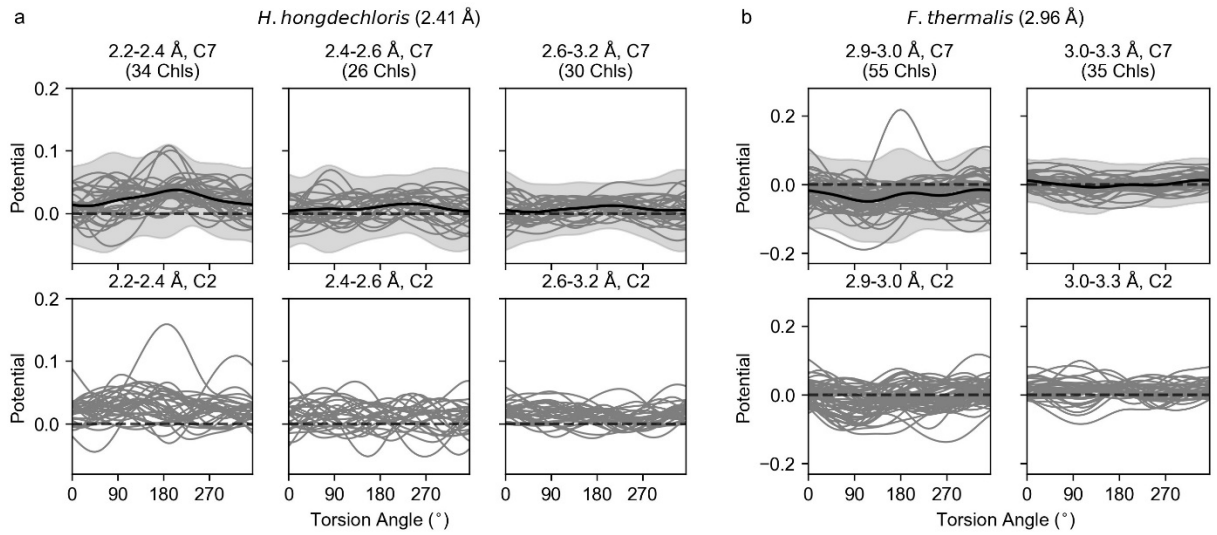

Fig. S3. C7 methyl cone scan distribution and C2 cone scans. Top rows show the methyl cone scan distribution derived from C7 traces. The gray lines are the C7 cone scans using formyl parameters, the black line is the mean of the scans, and the gray area covers the mean plus and minus three standard deviations ($\mu \pm 3\sigma$). Bottom rows show the C2 cone scans using formyl parameters. **a** shows the cone scans for Chl sites from the 2.41-Å global resolution *H. hongdechloris* structure and **b** shows the cone scans for Chl sites from the 2.96-Å global resolution *H. hongdechloris* structure *F. thermalis* on the right. Bin resolutions and the number of Chl sites per bin are labeled.

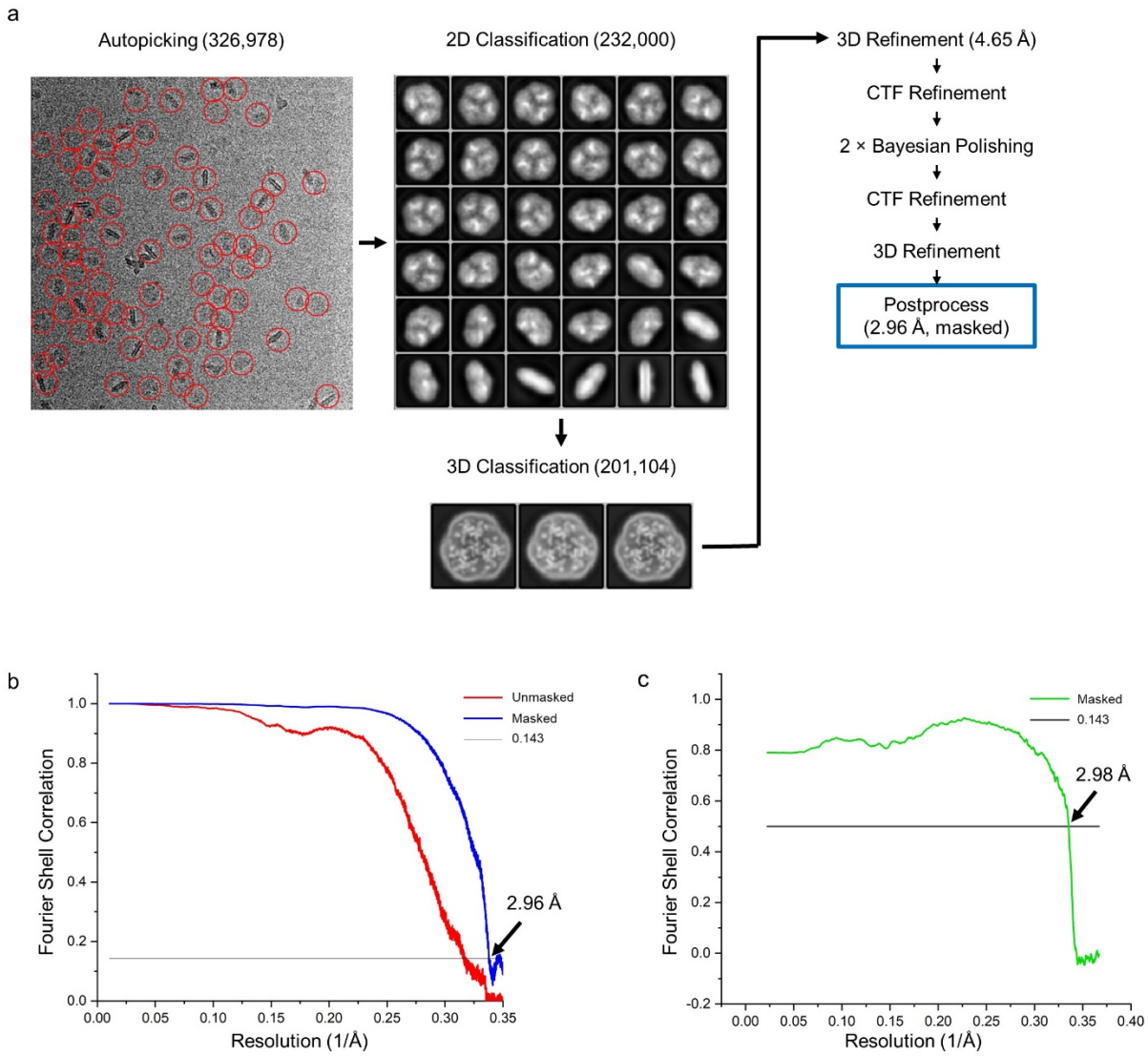

Fig. S4. Data processing for the re-processed *F. thermalis* FRL-PSI structure. **a** Data processing workflow. All data processing was performed in RELION 3.1 [9]. A subset of ~1,000 particles were picked manually and their corresponding 2D classes were used for automatic particle picking of the entire data set. Picking on all micrographs was inspected and edited manually resulting in 326,978 particle images. 36 of the 50 classes from the subsequent 2D classification were chosen resulting in a data set containing 232,000 particle images. Three of the four 3D classes from the subsequent 3D classification were chosen resulting in a data set containing 201,104 particle images. 3D refinement of this data set led to a map of 4.65-Å resolution. Iterations of CTF refinement and Bayesian polishing were performed to produce a map of 2.96-Å resolution when masked. **b** Fourier shell correlation of masked and unmasked half-maps. **c** Masked map-to-model Fourier shell correlation.

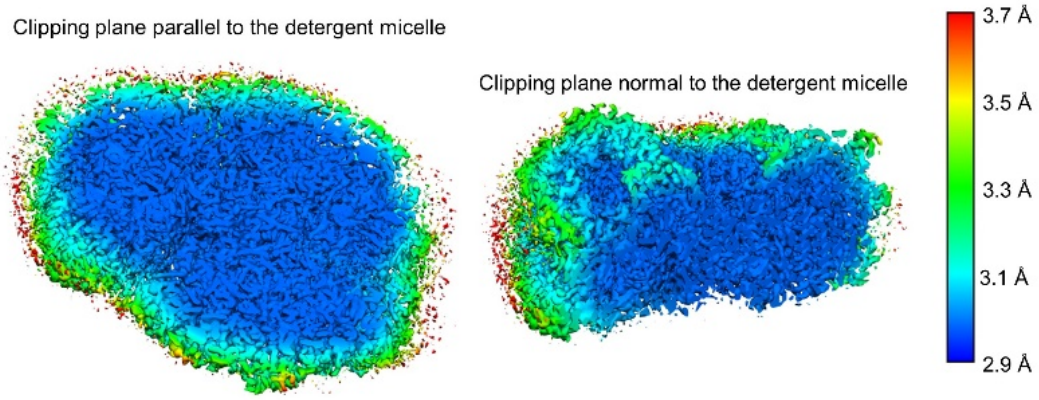

Fig. S5. Local resolution of the 2.96-Å resolution cryo-EM map of FRL-PSI from *F. thermalis*. The local resolution is shown by the heat map which spans from 2.9 to 3.7-Å resolution. The local resolution was calculated using the implementation of LocRes within RELION 3.1 [9].

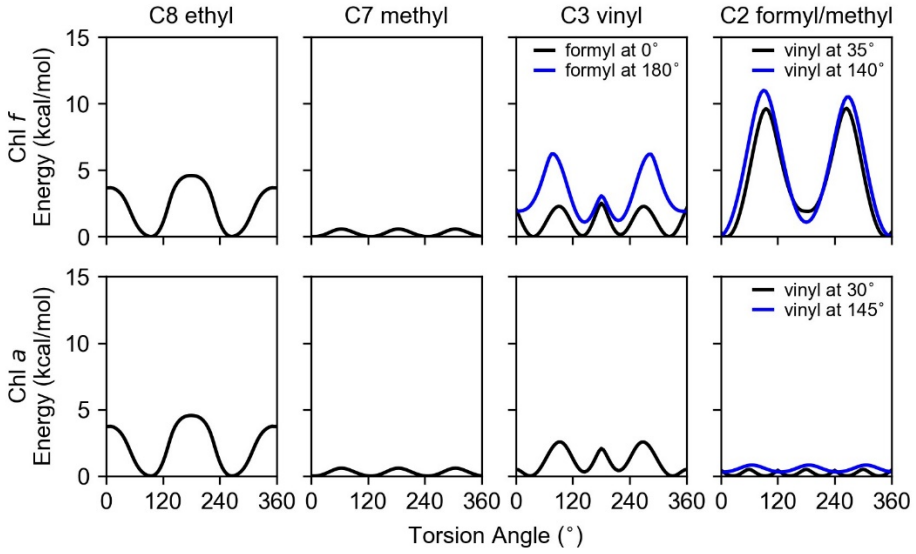

Fig. S6. Calculated free energies for Chl *f* and Chl *a* substituents. The free energies associated with specific substituent orientations are shown. For C2 methyl or formyl, the adjacent C3 vinyl substituent was held at each of two orientations corresponding to the C3 energy minima.

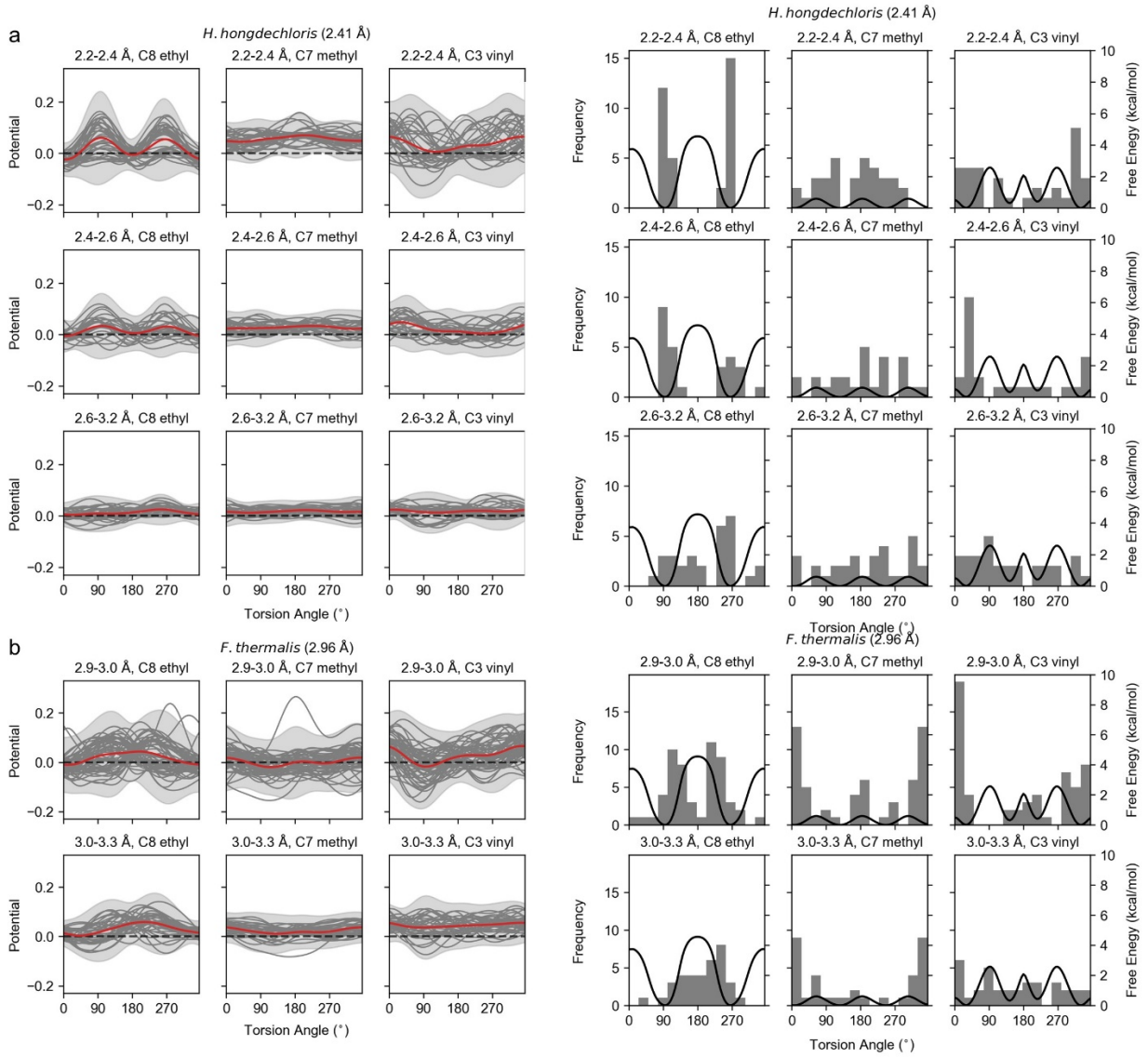

Fig. S7. Experimental cone scans of Chl substituents for C8 ethyl, C7 methyl, and C3 vinyl substituents divided into resolution bins. **a** and **b** show the cone scans (left) and peak distribution histograms of the cone scans (right) for *H. hongdechloris* and *F. thermalis*, respectively. Cone scan panels additionally include the average (red line). Cone scan panels also include a grey-shaded area that denotes mean plus and minus three standard deviations ($\mu \pm 3\sigma$) of the respective resolution bin.

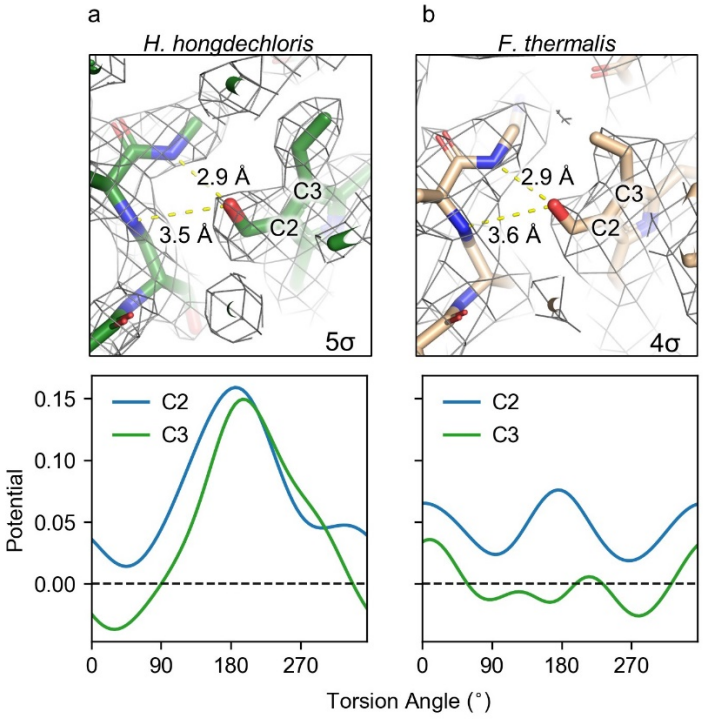

Fig. S8. Maps, models, and cone scans of Chl *f* in the B37 sites of the two FRL-PSI structures. **a** and **b** show the *H. hongdechloris* and *F. thermalis* data, respectively. Top panels show the model and map in the vicinity of the C2/C3 position. Bottom panels show the cone scans for the C2 (blue, prospective formyl) and C3 (green, vinyl) positions.

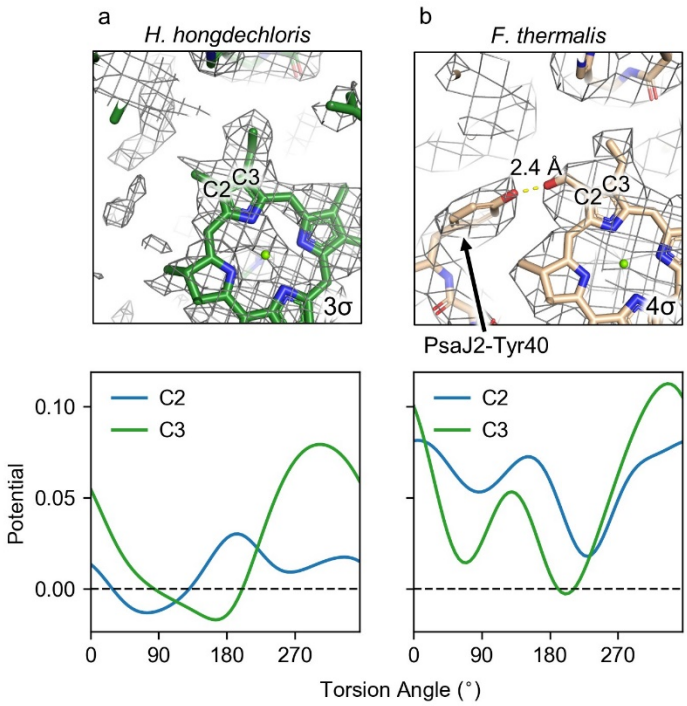

Fig. S9. Maps, models, and cone scans of Chl *f* in the B30 sites of the two FRL-PSI structures. **a** and **b** show the data for the *H. hongdechloris* and *F. thermalis* data, respectively. Top panels show the model and map in the vicinity of the C2/C3 position. Bottom panels show the cone scans for the C2 (blue, prospective formyl) and C3 (green, vinyl) positions. Note that PsbJ2 was presumed to be lost in sample preparation from the *H. hongdechloris* FRL-PSI structure.

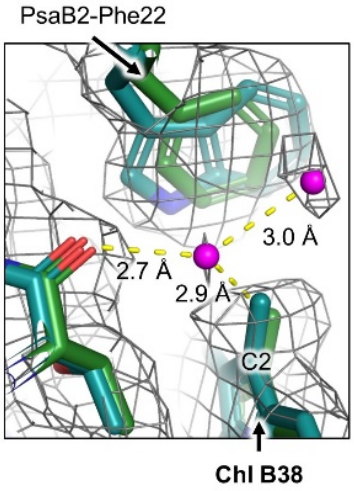

Fig. S10. Possible Chl *f* at site B37. The structure of FRL-PSI from *H. hongdechloris* is shown (green) within its corresponding ESP map at 2.0σ. The structure of PSI adapted to white light also from *H. hongdechloris* is shown (teal), superimposed. Prospective water molecules are shown in magenta.

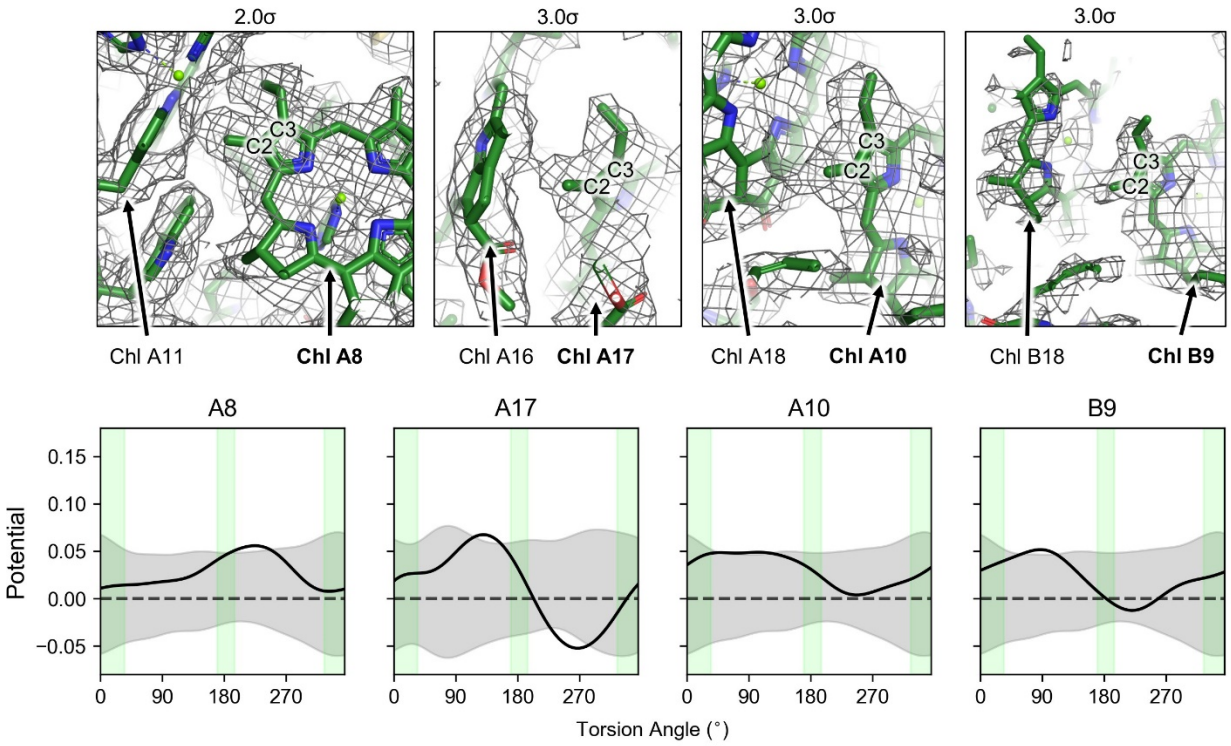

Fig. S11. Unlikely Chl *f* candidates in *H. hongdechloris* FRL-PSI whose C2 substituents exhibit signal above the background. The bold-labeled Chl sites are those for which the cone scans differ significantly from the methyl distribution.

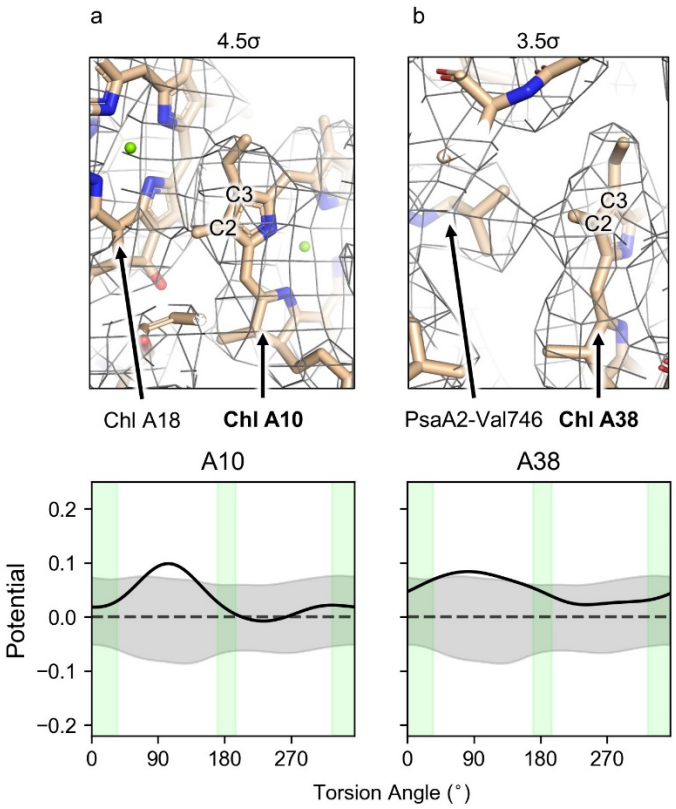

Fig. S12. Unlikely Chl *f* candidates in *F. thermalis* whose C2 substituents exhibit signal above the background. The bold-labeled Chl sites are those for which the cone scans differ significantly from the methyl distribution.

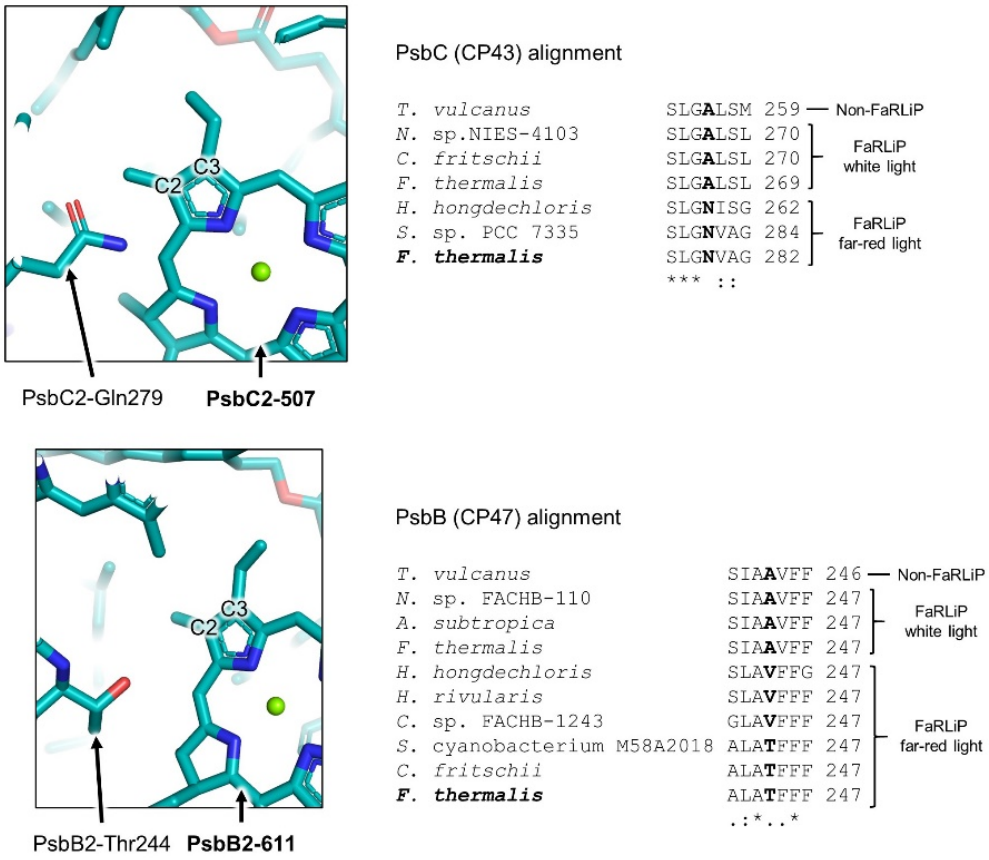

Fig. S13. Suggested Chl *f* sites in FRL-PSII. Left panels show the chemical environment of the C2 and C3 positions for the two antenna Chl sites described in the text. FRL-specific sidechains that are suggested to H-bond to the C2 formyl substituents of the Chls are labelled. Right panels show the corresponding sequence alignments. For the top panels describing CP43-507, the sequence alignment shows one non-FaRLiP sequence, three FaRLiP sequences where the isoform is expressed in white light, and three FaRLiP sequences where the isoform is expressed in far-red light. For the top panels describing CP47-611, the sequence alignment shows one non-FaRLiP sequence, three FaRLiP sequences where the isoform is expressed in white light, and six FaRLiP sequences where the isoform is expressed in far-red light. Of the six FaRLiP-exhibiting sequences where the isoform is expressed in far-red light, three are shown that exhibit the possible H-bonding Thr residue, and three instead exhibit a Val in that position, suggesting possible species-specific variation. For both sequence alignments, the FRL-PSII sequence pertaining to the homology model is in bold. Sequence alignments were performed using Clustal Omega [10]. Note that protein sidechains in the homology model are unlikely to be reliable.

Supplementary Tables

Table S1. Cone scan parameters for Chl substituents

Substituent scan parameters	Distance from CX^1 atom (Å)	Angle from $CX\text{-}CX^1$ bond (°)
formyl	1.23	120
vinyl	1.44	120
methyl	1.00	109.5
ethyl	1.57	109.5

X corresponds to substituent position (e.g. formyl substituent is at position C2 in the tetrapyrrole ring therefore X=2)

Table S2. Updated cryo-EM data collection, refinement, and validation statistics for FRL-PSI (PDB ID = 7PX0, EMDB ID = EMD-23563).

Data collection and processing	
Magnification	x47,600
Voltage (kV)	300
Electron exposure	57.3
Defocus range (μm)	-1 to -3
Pixel size (Å)	0.53
Symmetry imposed	C3
Initial particle images (no.)	326,978
Final particle images (no.)	201,104
Map resolution (Å)	2.96
FSC threshold	0.143
Map resolution range (Å)	2.9-3.7
Refinement	
Initial model used (PDB code)	6PNJ
Model resolution (Å)	2.98
FSC threshold	0.5
Map resolution range (Å)	2.9-3.7
Map-sharpening *B* factor ($Å^2$)	-36
Model composition	
Non-hydrogen atoms	72,462
Protein residues	6,699
Ligands	447
B factors ($Å^2$)	
Protein	32
Ligands	27
R.m.s. deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.900
Validation	
MolProbity	2.30
Clashscore	15.02
Rotamer outliers (%)	2.03
Ramachandran plot	
Favored (%)	94.25
Allowed (%)	5.63
Disallowed (%)	0.12

Table S3. Conversion of Chl site names from PDB file names.

Site name used here	1JB0		6PNJ			6KMX		
	Chain Name	Residue Number	Chain Name	Residue Number	Local Resolution of Central Mg (Å)	Chain Name	Residue Number	Local Resolution of Central Mg (Å)
P_A	A	1011	A	1011	2.98	aA	801	2.234
A_{-1A}	B	1012	B	1022	2.98	aB	802	2.364
A_{0A}	A	1013	A	1013	2.98	aA	803	2.39
P_B	B	1021	B	1021	2.98	aB	801	2.318
A_{-1B}	B	1023	B	1023	2.98	aB	803	2.253
A_{0B}	A	1022	A	1012	2.98	aA	802	2.259
A1	A	1101	A	1101	2.999	aA	804	2.607
A2	A	1102	A	1102	2.99	aA	805	2.547
A3	A	1103	A	1103	2.98	aA	806	2.44
A4	A	1104	A	1104	2.98	aA	807	2.42
A5	A	1105	A	1105	3.008	aA	808	2.668
A6	A	1106	A	1106	2.984	aA	809	2.497
A7	A	1107	A	1107	3.013	aA	810	2.615
A8	A	1108	A	1108	3.06	aA	811	2.716
A9	A	1109	A	1109	3.04	aA	812	2.854
A10	A	1110	A	1110	3.055	aA	813	2.668
A11	A	1111	A	1111	2.997	aA	814	2.506
A12	A	1112	A	1112	3.028	aA	815	2.626
A13	A	1113	A	1113	3.108	aA	816	2.725
A14	A	1114	A	1114	3.147	aA	817	2.786
A15	A	1115	A	1115	3.055	aA	818	2.598
A16	A	1116	A	1116	3	aA	819	2.464
A17	A	1117	A	1117	2.989	aA	820	2.444
A18	A	1118	A	1118	3.033	aA	821	2.648
A19	A	1119	A	1119	2.992	aA	822	2.443
A20	A	1120	A	1120	3.025	aA	823	2.575
A21	A	1121	A	1121	3	aA	824	2.461
A22	A	1122	A	1122	2.99	aA	825	2.38
A23	A	1123	A	1123	2.985	aA	826	2.421
A24	A	1124	A	1124	2.98	aA	827	2.304
A25	A	1125	A	1125	2.982	aA	828	2.379
A26	A	1126	A	1126	2.98	aA	829	2.388
A27	A	1127	A	1127	2.98	aA	830	2.397
A28	A	1128	A	1128	2.98	aA	831	2.39
A29	A	1129	A	1129	2.98	aA	832	2.323
A30	A	1130	A	1130	2.98	aA	833	2.284
A31	A	1131	A	1131	2.98	aA	834	2.253

A32	A	1132	A	1132	2.98	aA	835	2.273
A33	A	1133	A	1133	2.99	aA	836	2.412
A34	A	1134	A	1134	3.018	aA	837	2.504
A35	A	1135	A	1135	2.98	aA	838	2.34
A36	A	1136	A	1136	2.98	aA	839	2.297
A37	A	1137	A	1137	2.98	aA	840	2.289
A38	A	1138	A	1138	3.009	aA	841	2.645
A39	A	1139	A	1139	3.002	aA	842	2.643
A40	A	1140	A	1140	2.98	aA	843	2.432
B1	B	1201	B	1201	2.98	aB	804	2.346
B2	B	1202	B	1202	2.984	aB	805	2.401
B3	B	1203	B	1203	2.98	aB	806	2.356
B4	B	1204	B	1204	2.982	aB	807	2.347
B5	B	1205	B	1205	2.98	aB	808	2.338
B6	B	1206	B	1206	2.98	aB	809	2.304
B7	B	1207	B	1207	2.98	aL	202	2.274
B8	B	1208	B	1208	3.003	aB	810	2.518
B9	B	1209	B	1209	3.051	aB	811	2.644
B10	B	1210	B	1210	2.995	aB	812	2.531
B11	B	1211	B	1211	3.006	aB	813	2.554
B12	B	1212	B	1212	3.071	aB	814	2.701
B13	B	1213	B	1213	3.061	aB	815	2.803
B14	B	1214	B	1214	3.002	aB	816	2.642
B15	B	1215	B	1215	2.983	aB	817	2.525
B16	B	1216	B	1216	3	aB	818	2.667
B17	B	1217	B	1217	3.069	aB	819	2.708
B18	B	1218	B	1218	3.116	aB	820	2.951
B19	B	1219	B	1219	3.091	aB	821	2.863
B20	B	1220	B	1220	3.031	aB	822	2.703
B21	B	1221	B	1221	2.993	aB	823	2.571
B22	B	1222	B	1222	2.986	aB	824	2.403
B23	B	1223	B	1223	3	aB	825	2.563
B24	B	1224	B	1224	2.98	aB	826	2.342
B25	B	1225	B	1225	2.98	aB	827	2.383
B26	B	1226	B	1226	2.982	aB	828	2.37
B27	B	1227	B	1227	3.048	aB	829	2.777
B28	B	1228	B	1228	3.064	aB	830	2.759
B29	B	1229	B	1229	3.024	aB	831	2.64
B30	B	1230	B	1230	3.02	aB	832	2.661
B31	B	1231	B	1231	3.032	aB	833	2.725
B32	B	1232	B	1232	3.075	aB	834	2.865
B33	B	1233	B	1233	3.216	aB	835	3.2
B34	B	1234	B	1234	2.993	aB	836	2.49

B35	B	1235	B	1235	3.025	aB	837	2.604
B36	B	1236	B	1236	3.007	aB	838	2.523
B37	A	1237	B	1237	2.98	aA	844	2.246
B38	B	1238	B	1238	2.98	aB	839	2.263
B39	B	1239	B	1239	2.98	aB	840	2.26
F1	F	1301	N/A	N/A	N/A	N/A	N/A	N/A
F2	J	1302	N/A	N/A	N/A	N/A	N/A	N/A
F3	J	1303	N/A	N/A	N/A	N/A	N/A	N/A
K1	K	1401	N/A	N/A	N/A	aK	103	2.595
K2	A	1402	K	1401	3.215	aK	101	2.803
L1	L	1501	L	1501	2.98	aL	204	2.247
L2	L	1502	L	1502	2.98	aL	205	2.24
L3	L	1503	L	1503	2.98	aL	206	2.267
M1	M	1601	N/A	N/A	N/A	N/A	N/A	N/A
X1	X	1701	X	1701	3.099	N/A	N/A	N/A
lip1	A	1801	N/A	N/A	N/A	N/A	N/A	N/A

Table S4. Axial ligands and H-bonding donors to Chl *b* and *f* molecules.

Species	Structure	Pigment identity	Pigment number	Axial ligand	H-bond donor to C3/C2
P. sativum	PSI-LHCI	Chl *b*	Lhca1-514	Lhca1-Glu146 (S)	N/A
P. sativum	PSI-LHCI	Chl *b*	Lhca1-512	Lhca1-Gln109 (S)	Lhca1-Glu138 (S)
P. sativum	PSI-LHCI	Chl *b*	Lhca1-521	Lhca1-Trp41 (B)	N/A
P. sativum	PSI-LHCI	Chl *b*	Lhca4-313	N/A	Lhca4-Glu145 (S)
P. sativum	PSI-LHCI	Chl *b*	Lhca4-314	Lhca4-W408	Lhca4-Arg155 (S)
P. sativum	PSI-LHCI	Chl *b*	Lhca4-316	N/A	Lhca4-CHL313 (P)
P. sativum	PSI-LHCI	Chl *b*	Lhca4-317	Lhca4-Asp169 (S)	Lhca4-Tyr151 (S)
P. sativum	PSI-LHCI	Chl *b*	Lhca2-512	N/A	Lhca2-CHL515 (P)
P. sativum	PSI-LHCI	Chl *b*	Lhca2-513	N/A	Lhca2-Arg166 (S)
P. sativum	PSI-LHCI	Chl *b*	Lhca2-515	Lhca2-W604	Lhca2-CHL513 (P)
P. sativum	PSI-LHCI	Chl *b*	Lhca2-516	Lhca2-Asp180 (S)	Lhca2-Trp162 (S)
P. sativum	PSI-LHCI	Chl *b*	Lhca2-526	Lhca2-Trp67 (S)	Lhca2-W602
P. sativum	PSI-LHCI	Chl *b*	Lhca3-314	Lhca3-Arg171	N/A
P. sativum	LHCII	Chl *b*	609	Tyr24 (B)	N/A
P. sativum	LHCII	Chl *b*	610	W2009	Gln131 (S)
P. sativum	LHCII	Chl *b*	611	W2008	Leu148 (B)
P. sativum	LHCII	Chl *b*	612	Glu139 (S)	Gln131 (S)
P. sativum	LHCII	Chl *b*	613	N/A	W2009
P. sativum	LHCII	Chl *b*	614	Val119 (B)	Ser123 (B)
H. hongdechloris	FRL-PSI	Chl *f* (direct)	B7	PsaB1-Gln95 (S)	PsaB1-Ala94 (B)
H. hongdechloris	FRL-PSI	Chl *f* (direct)	B37	PsaA1-W923	PsaB1-Gly700 (B)
H. hongdechloris	FRL-PSI	Chl *f* (direct)	A23	PsaA1-W953	PsaA1-Gln366 (S)
H. hongdechloris	FRL-PSI	Chl *f* (indirect)	B30	PsaB1-His446	N/A
H. hongdechloris	FRL-PSI	Chl *f* (indirect)	A21	PsaA1-W924	PsaA1-Leu334 (B)
H. hongdechloris	FRL-PSI	Chl *f* (indirect)	B38	PsaB1-W951	N/A
F. thermalis	FRL-PSI	Chl *f* (direct)	B7	PsaB2-Gln95 (S)	PsaB2-Ala94 (B)
F. thermalis	FRL-PSI	Chl *f* (indirect)	B37	N/A (assumed W)	PsaB2-Gly697 (B)
F. thermalis	FRL-PSI	Chl *f* (direct)	B30	PsaB2-His443 S or W	PsaJ2-Tyr40 (S)
F. thermalis	FRL-PSI	Chl *f* (indirect)	A21	N/A (assumed W)	PsaA2-Leu337 (B)
F. thermalis	FRL-PSI	Chl *f* (indirect)	B38	N/A (assumed W)	N/A

P. sativum PSI-LHCI complex: 2.6-Å X-ray crystal structure, 5L8R
P. sativum LHCII: 2.5-Å X-ray crystal structure, 5L8R
H. hongdechloris FRL-PSI complex: 2.41-Å cryo-EM structure, 6KMX
F. thermalis FRL-PSI complex: 2.96-Å cryo-EM structure, 6PNJ
S = sidechain
B = backbone
P = pigment
direct = direct evidence from C2 cone scans
indirect = high confidence Chl *f* sites based on indirect evidence

Table S5. Atoms comprising dihedral angles scanned in substituent energy analysis.

Chl type	Scanned Group	Atom 1	Atom 2	Atom 3	Atom 4
a	methyl (C2)	C1	C2	C21	H21
f	formyl (C2)	C1	C2	C21	O21
a and *f*	vinyl (C3)	C2	C3	C31	C32
a and *f*	methyl (C7)	C8	C7	C71	H71
a and *f*	ethyl (C8)	C7	C8	C81	C82

Supplementary Movies

Supplementary Movie 1. DFT simulation of Chl *f* rotating C2 formyl (external file).

Supplementary Movie 2. DFT simulation of Chl *a* rotating C3 vinyl (external file).

Supplementary Movie 3. DFT simulation of Chl *a* rotating C7 methyl (external file).

Supplementary Movie 4. DFT simulation of Chl *a* rotating C8 ethyl (external file).

Supplementary Codes

Supplementary Code 1. Source code for the cone scan method (external file).

Supplementary References

[1] K. Leiger, J.M. Linnanto, A. Freiberg, Establishment of the Qy absorption spectrum of chlorophyll a extending to near-infrared, Molecules. 25 (2020) 3796. https://doi.org/10.3390/molecules25173796.

[2] J.K. Hoober, L.L. Eggink, M. Chen, Chlorophylls, ligands and assembly of light-harvesting complexes in chloroplasts, Photosynth. Res. 94 (2007) 387–400. https://doi.org/10.1007/s11120-007-9181-1.

[3] C.J. Gisriel, J. Wang, G.W. Brudvig, D.A. Bryant, Opportunities and challenges for assigning cofactors in cryo-EM density maps of chlorophyll-containing proteins, Commun. Biol. 3 (2020) 408. https://doi.org/10.1038/s42003-020-01139-1.

[4] D.J. Nürnberg, J. Morton, S. Santabarbara, A. Telfer, P. Joliot, L.A. Antonaru, A. V. Ruban, T. Cardona, E. Krausz, A. Boussac, A. Fantuzzi, A.W. Rutherford, Photochemistry beyond the red limit in chlorophyll *f*–containing photosystems, Science 360 (2018) 1210–1213. https://doi.org/10.1126/science.aar8313.

[5] D.A. Cherepanov, I. V Shelaev, F.E. Gostev, A. V Aybush, M.D. Mamedov, G. Shen, V.A. Nadtochenko, D.A. Bryant, A.Y. Semenov, J.H. Golbeck, Evidence that chlorophyll *f* functions solely as an antenna pigment in far red-light Photosystem I from *Fischerella thermalis* PCC 7521, Biochim. Biophys. Acta - Bioenerg. 1861 (2020) 148184.

[6] M. Tros, V. Mascoli, G. Shen, M.-Y. Ho, L. Bersanini, C.J. Gisriel, D.A. Bryant, R. Croce, Breaking the red limit: Efficient trapping of long-wavelength excitations in chlorophyll-*f*-containing photosystem I, Chem. 7 (2020) 155–173. https://doi.org/https://doi.org/10.1016/j.chempr.2020.10.024.

[7] C. Gisriel, G. Shen, V. Kurashov, M.-Y. Ho, S. Zhang, D. Williams, J.H. Golbeck, P. Fromme, D.A. Bryant, The structure of Photosystem I acclimated to far-red light illuminates an ecologically important acclimation process in photosynthesis, Sci. Adv. 6 (2020) eaay6415. https://doi.org/10.1126/sciadv.aay6415.

[8] K. Kato, T. Shinoda, R. Nagao, S. Akimoto, T. Suzuki, N. Dohmae, M. Chen, S.I. Allakhverdiev, J.-R. Shen, F. Akita, N. Miyazaki, T. Tomo, Structural basis for the adaptation and function of chlorophyll *f* in photosystem I, Nat. Commun. 11 (2020) 238. https://doi.org/10.1038/s41467-019-13898-5.

[9] J. Zivanov, T. Nakane, B.O. Forsberg, D. Kimanius, W.J. Hagen, E. Lindahl, S.H. Scheres, New tools for automated high-resolution cryo-EM structure determination in RELION-3, Elife. 7 (2018) e42166. https://doi.org/10.7554/elife.42166.

[10] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 1–6. http://msb.embopress.org/content/7/1/539.abstract.