nature plants

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41477-020-0611-9

In the format provided by the authors and unedited.

Structure of a minimal photosystem I from the green alga *Dunaliella salina*

Annemarie Perez-Boerema[1,2,4], Daniel Klaiman[3,4], Ido Caspy[3,4], Sigal Y. Netzer-El[3], Alexey Amunts[1,2] ✉ and Nathan Nelson[3] ✉

[1]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Solna, Sweden. [2]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [3]Department of Biochemistry and Molecular Biology, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel. [4]These authors contributed equally: Annemarie Perez-Boerema, Daniel Klaiman, Ido Caspy. ✉
e-mail: amunts@scilifelab.se; nelson@post.tau.ac.il

Supplementary Table 1. X-ray data collection and refinement statistics.

	PDB 6QPH
Data collection	
Space group	P1 21 1
Cell dimensions	
a, *b*, *c* (Å)	158.881, 100.506, 191.04
α, β, γ (°)	90 91.7335 90
Resolution (Å)	49-3.4
R_{sym} or R_{merge}	1.435
I / σI	0.8
Completeness (%)	99.6
Redundancy	7.3
Refinement	
Resolution (Å)	49-3.4
No. reflections	688522(33502)
R_{work} / R_{free}	0.336/0.356
No. atoms	
Protein	21861
Ligand/ion	9009
Water	0
B-factors	
Protein	120.97
Ligand/ion	119.45
Water	0
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.876

*Values in parentheses are for highest-resolution shell.

The X-ray dataset was generated by merging two independent datasets with similar unit cell dimensions. Combination of the two datasets and an inherint anisotropicity caused the Rfree and clashscore values to be higher than expected, but still within a reasonable range for the reported resolution.

Supplementary Table 2. Cryo-EM data collection and model building.

	EMDB-4883 **PDB 6RHZ**
Data collection and processing	
Magnification	130000
Voltage (kV)	300
Electron exposure (e–/Å^2)	41.6
Defocus range (μm)	0.9-3.0
Pixel size (Å)	1.065
Symmetry imposed	No
Initial particle images (no.)	470,368
Final particle images (no.)	132,017
Map resolution (Å)	3.2
FSC threshold	0.143
Map resolution range (Å)	~2.7-4.0
Refinement	
Initial model used (PDB code)	5L8R
Model resolution (Å)	3.2
FSC threshold	0.143
Model resolution range (Å)	
Map sharpening *B* factor (Å^2)	-70.5
Model composition	
Non-hydrogen atoms	31,888
Protein residues	2767
Ligands	197
B factors (Å^2)	
Protein	15.52
Ligand	21.57
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.175
Validation	
MolProbity score	1.17 (99th Percentile)
Clashscore	1.51 (99th percentile)/5
Poor rotamers (%)	0
Ramachandran plot	
Favored (%)	96.2
Allowed (%)	3.73
Disallowed (%)	0.07

Supplementary Table 3. Peptides detected in Mass-spectroscopy analysis

Mass-spectroscopy analysis of various Dunaliella salina subunits. Underlined sequences indicate peptides detected in MS.

>chain ' 1'

MAHQYHWELKSFPSCACSNFAPGSEPKEYLNDLPGNFNFDPLELGKEKGTLQRYREAELIHCRWAMLGAAGCLAVEVLGLGNWYDAPLWAVTGDKPTWFGIEVPFDIATILGVEVVAMAVAEGLRNDNQDMEKRLYPGGAFDPLGFSKDPKSFEDKKLKELKNGRLAMVACLGFAGQHAATGKPILAALGDHLSSPFFNNFATNGVSVPGL

>chain ' 2'

MSALVNKSVSLQAVTKAPLAARSIAPRSVASRTRNVALRAGADRPLWSPGSEPPAWLDGSLAGDYGFDPLHLSEEPEMRKWMVQAELVHARWAMLGVAGILFTSIAAKTGAPFPDWYDAGKEAIKTSHPSLIFTELLLFGWVETKRLYDLRNPGSQGDGSFLGITDGLKGKENGYPGGMSKNEASYKEAKVKEIKNGRLAMLAFVGFIAQHHATHKSPIDNLLDHVADPFHVTFATNGVSIPHFTEF

>chain ' 3'

SKDFLYVGSDAAALKYLDGTLPGDYGFDPLGLLDPTVSNGQGAGGFVNPRWLQYSVIHARWAMLGAAGCIAPEILGKAGVIPAETAVDWFRTGVIPPAGVYKDFWADPFTLFFIEVVAIQFAELKRLQDYKNPGSQSRQYFLGLEGLFKGSDNPAYPGGPFFNFANFGKTEAEMKKLKLNEIKNGRLAMLAMFGYGAQAVITGDGPFDNLLAHLADPTGANLITNLGGKFGQ

>chain ' 4'

DREMWYPGATPPAHLDGSMLGDYGFDPLRLGTNPDRMKWFREAELTNGRWAMAAVVGILFTDVFTSIGLVGLPKWWEAGAQTYPIDNQTLAIIEFLLFGWVETKRLYDLRNPGSQGDGSFLGITDGLKGKENGYPGGMSKNEASYKEAKVKEIKNGRLAMLAFDGFIAQHHATHKSPIDNLLDHVADPFHVTFATNGVSIPHFTEF>

chain ' C'

MAHVVKIYDTCIGCTQCVRACPLDVLEMVPWDGCKAAQMASSPRTEDCVGCKRCETACPTDFLSVRVYLGNESTRSLGLAY

>chain ' D'

VRAEAAPPAAGAPPEPKAAGAPPAAPKKKAPPPPWKQPELDPDTPSPIFGGSTGGLLRKAQVEEFYVITWESPKEQIFEMPTGGAAIMRKGPNLLKFARKEQCMALTTQLRSKFRQTPCFYRVYADGKVQYLHPKDGVCPEKVNAGRVGVNQNMRSIGKNVDPIKVKFTGSEPFEI

>chain ' E'

MLAAKPQRTRLIVRAEGEAPPAAPKQAPAEAKAAPKGEAKAAPKKKEIGPKRGSLVKVLRPESYWYNQVGKVVSVDQSGIRYPVVVRFENQNYAGVSTNNYALDEIEVTDPPSK

>chain ' F'

PAQNLGAVACATALALTMGLTADVQPASADIAGLTPCSESKAYNKLERKELKVLDKRLKQYEPGSAPYLALQATKERTENRFKTYAKQGLLCGNDGLPHLISDPGLALRFNHAGEVFITFGFLYVAGYIGHVGRQYIILSKEDAKPTDKEIILDVPLALKLAFQGWAWPLASIQELRNGSLLEKDENITV

Supplementary Table 4. Model information.

Protein	Chain ID	Accession Number	Built crystal model	Built cryo-EM model	Remarks
Lhca1	1	ACN94453	36-220	32-228	In database annotated as Lhca2
Lhca2	2		62-254	64-269	Mass spec & GenBank: NSFN01000125.1
Lhca3	3	ACN94454	74-272	73-282	Mass spec & GenBank: NSFN01000125.1
Lhca4	4		123-328	123-331	Mass spec & GenBank: NSFN01000125.1 Gap
PsaA	A	YP_005089772	13-751	13-751	
PsaB	B	YP_005089813	2-735	6-735	
PsaC	C	YP_005089787	2-81	2-81	
PsaD	D		71-212	69-209	Mass spec & GenBank: NSFN01000125.1
PsaE	E		65-128	65-128	Mass spec & GenBank: NSFN01000125.1
PsaF	F		78-240	78-240	Mass spec & GenBank: NSFN01000125.1
PsaJ	J	YP_005089780	1-41	1-40	

Supplementary Table 5. Sequence alignment of Lhca proteins.
D. salina (Ds) Lhca1-4 were aligned against *C. reinhardtii* (Cr) Lhca1-8. The table summarizes the percentage of identity and similarity (in parenthesis). Ds-Lhca2 and Ds-Lhc4 appear to be similar to Cr-Lhca7.

Ds/Cr	Lhca1	Lhca2	Lhca3	Lhca4	Lhca5	Lhca6	Lhca7	Lhca8
Uniprot code	A8J249	A8IKC8	A8JF10	A8I000	Q75VY8	Q75VY6	A8ISG0	Q75VY7
Lhca1	65%(75%)	38%(49%)	36%(48%)	38%(53%)	39%(52%)	36%(48%)	39%(53%)	39%(50%)
Lhca2	41%(53%)	41%(51%)	41%(54%)	45%(59%)	46%(59%)	46%(53%)	63%(77%)	57%(69%)
Lhca3	36%(43%)	33%(45%)	71%(83%)	40%(55%)	38%(52%)	36%(47%)	42%(55%)	40%(51%)
Lhca4	38%(50%)	60%(67%)	38%(53%)	44%(59%)	39%(55%)	45%(57%)	53%(68%)	48%(64%)

Supplementary Table 6. LHCI chlorophylls nomenclature.

Chain	Ligand	*D. salina*	5ZJI	4XK8
	LUT	501	617	316
	XAT	502	618	317
	BCR	503	619	318
	CLA	601	610	310
	CLA	602	612	312
	CLA	603	613	313
	CLA	604	602	303
	CLA	605	603	304
1	CLA	606	604	305
	CLA	607	611	311
	CLA	608	614	314
	CHL	609	601	302
	CHL	610	607	307
	CLA	611	608	308
	CLA	612	609	309
	CLA	613	606	306
	CLA	615	616	315
	LHG	801	620	319

Chain	Ligand	*D. salina*	5ZJI	4XK8
2	LUT	501	619	615
	XAT	502	620	616
	BCR	503	621	617
	CLA	601	610	609
	CLA	602	612	611
	CLA	603	613	612
	CLA	604	602	602
	CLA	605	603	603
	CLA	606	604	604
	CLA	607	611	610
	CLA	608	614	613
	CHL	609	601	601
	CHL	610	607	606
	CHL	611	608	607
	CLA	612	609	608
	CHL	613	606	605
	CLA	616	1002(N)	
	LHG	801	622	618

Chain	Ligand	*D. salina*	5ZJI	4XK8
3	LUT	501	618	316
	XAT	502	619	317
	BCR	503	620	318
	BCR	504		
	BCR	506		
	CLA	601	610	309
	CLA	603	613	312
	CHL	604	602	302
	CLA	605	603	303
	CLA	606	604	304
	CLA	607		310
	CLA	608	614	313
	CLA	609	615	
	CLA	610	607	306
	CLA	611	608	307
	CLA	612	609	308
	CLA	613	606	305
	CLA	614	612	311
	CLA	615	617	314

Chain	Ligand	*D. salina*	5ZJI	4XK8
	LUT	501	619	616
	XAT	502	620	617
	BCR	503	621	618
	CLA	601	610	609
	CLA	602	612	611
	CLA	603	613	612
	CLA	604	602	602
	CLA	605	603	603
4	CLA	606	604	604
	CLA	607	611	610
	CLA	608	614	613
	CLA	609	601	601
	CHL	610	607	606
	CHL	611	608	607
	CLA	612	609	608
	CHL	613	606	605
	CLA	616		

Supplementary Table 7. cDNA primers designed according to the *C. reinhardtii* sequences.

Gene name	Primers	completion
lhca8	Fwd: GCATGGCTGGACGGCAGC Rev: AATCGGCCGTTCTTGATCTCCTTCAG	partial
lhca2	Fwd: GACGGCAGCTTGGCAGGC Rev: GGGGACTTGTGGGTGGCG	partial
lhcb1	Fwd: AAAGATGCATGGCTGGACGGCA Rev: GGCCGTTTTGATCTCCGCG	partial
lhcb2	Fwd: GAAAGATCCCTGAACCTGGGCGC Rev: GGGGTCAGCCAGGTGGTCC	partial

Lhca8 (partial)
GCATGGCTGGACGGCAGCAAGGTGCCTGCCCACCTGCAGAACGCCAAGCT
GCCAGGCAACTTTGGCTTTGACCCCCTGAACCTGGGCGCTGAGCCTGAGG
CCCTGCGCTGGTACCAGCAGGCTGAGCTGATCCACTCCAGGACTGCCATG
ATGGCCGTGGCAGGCATCATCATCCCAGGGGTGTTCATCAAGCTGGGCGC
CCTGAAGCTGCCCCAGTGGTACGAGGCTGGCAAGGTGTACATCAGCCAGC
CAGGCGCCATCCCCTTCGGCACCCTGCTGATGAGCATGTTCTTTGCCTACG
CCTTCGTGGAGGGCAAGAGGTGGCAGGACTTCCGCAAGCCTGGCAGCCAG
GCTGAGCCAGGCACCTTCTTCGGCCTGGAGAGCCAATTCAAGGGCACCGA
GGTGGGCTACCCCGGCGGCCTGTTCGACCCCCTGGGCTACTCCAAGACCA
GCCCCGAGAAGCTGGATGAGCTGAAGCTGAAGGAGATCAAGAACGGCCG
ATT

Lhca2 (partial)
GACGGCAGCTTGGCAGGCGATTATGGCTTTGACCCACTGCACCTGAGCGA
GGAGCCTGAGATGAGGAAGTGGATGGTGCAGGCCGAGCTGGTGCACTGC
CGCTGGGCCATGCTGGGCGTGGCAGGCATCCTGTTCACCTCTATTGGCGCC
AAGGCGGGCGGCAACTTCCCTGACTGGTACGACGCGGGCAAGGAGCTGC
AGAAGAACTCTGACATTCCCCTAGGCTCCCTGATCTTCACCGAGCTGCTGC
TGTTCGGATGGGTGGAGACCAAGCGCCTGTATGACCTGCGCAACCCCGGA
TCCCAGGGCGATGGCTCCTTCCTGGGCATCACTGACGGCCTGAAGGGCAA
GGAGAACGGCTACCCTGGTGGCCTGTTCGACCCCATGGGCATGAGCAAGA

ACGAGGCCAGCTTCAAGGAGGCCAAGCAGAAGGAGGTCAAGAACGGCCG
TCTGGCCATGCTAGCCTTCGTGGGCTTCATCGCGCAGCACCACGCCACCCA
CAAGTCCCCA

Lhcb1 (partial)
AAAGATGCATGGCTGGACGGCACCCTGCCTGGCGACCGTGGCTTCGATCC
TCTGGGCCTGGCCAAGCCCAAGGAGTTCGTGCTGATCGGCGTGGATGAGG
AGGACCAGAACAAGGCCAAGAACCAGAAGGGCAGCGTGGAGGCCACCTT
CGTGCCCGACTCCGACTCCATTGACACGGAGCAGAACAAGCTGGCGCCCT
ACTCCGAGGTCTTCGGCCTGGCGCGCTTCCGCGAGACCGAGGTCATCCAC
AGCCGATGGGCCATGCTGGCCGCGCTGGGCGTGTTCGTCAGCGAGGCCAG
CACCGGCGTGTCCTGGGTGGACGCTGGCAAGGTGGAGCTGGATGGCGCAA
GCTACCTGGGCTTCCCTCTGCCCTTCTCCATCACCCAGCTGCTGTGGATTG
AGGCCATCCTCATGGGTGGCGTGGAGATCCTGCGCAACAACGAGCTGGAC
CTAGAGAAGCGCATCTACCCTGGCGGTGCCTTCGACCCCCTGAACCTGGC
CTCTGAGGACGACCCCGAGCGCGCCTtCCGCCTGAAGACCGCGGAGATCA
AAACGGCC

Lhcb2 (partial)
GAAAGATCCCTGAACCTGGGCGCTCTGTCTGGCGAGCCCCCTTCCTACCTG
ACTGGCGAGTTCCCTGGTGACTACGGATGGGACTCCGCTGGCCTGTCTGC
GGACCCTAACACCTTCGCCCGCTACAGGACCATCGAGCTGATCCACGCTC
GCTGGGCTCTGCTGGGTGCTCTGGGCATCGTCACCCCTGAGCTGCTGGCCA
AGAACGGCGTGCCTTTCTCCGAGGACGCTGCAGTCTGGTTCAAGGCTGGC
TCTGAGATCTTCAAGGAGGGCGGCCTGAACTACCTGGGCAACGAGAACCT
GATCCACGCCCAGAACATTGTGGCCACCCTGGCTGTCCAGGTCATTGTGA
TGGGCGCAGCTGAGGGCTTCCGCGCCAACGGCGAGGCTCCAGGTGTGGAG
GGCGTTGAGGACCCCCTGTACCCTGGTGGCCCCTTCGACCCCCTGGGCCTG
GCTGACGACCCCGACACCTTCGCTGAGCTGAAGGTCAAGGAGATCAAGAA
CGGCCGCCTGGCTATGTTCTCATGCTTCGGCTTCTTCGTGCAGGCCATCGT
CACTGGCAAGGGCCCCATCCAGAACCTGCAGGACCACCTGGCTGACCCC