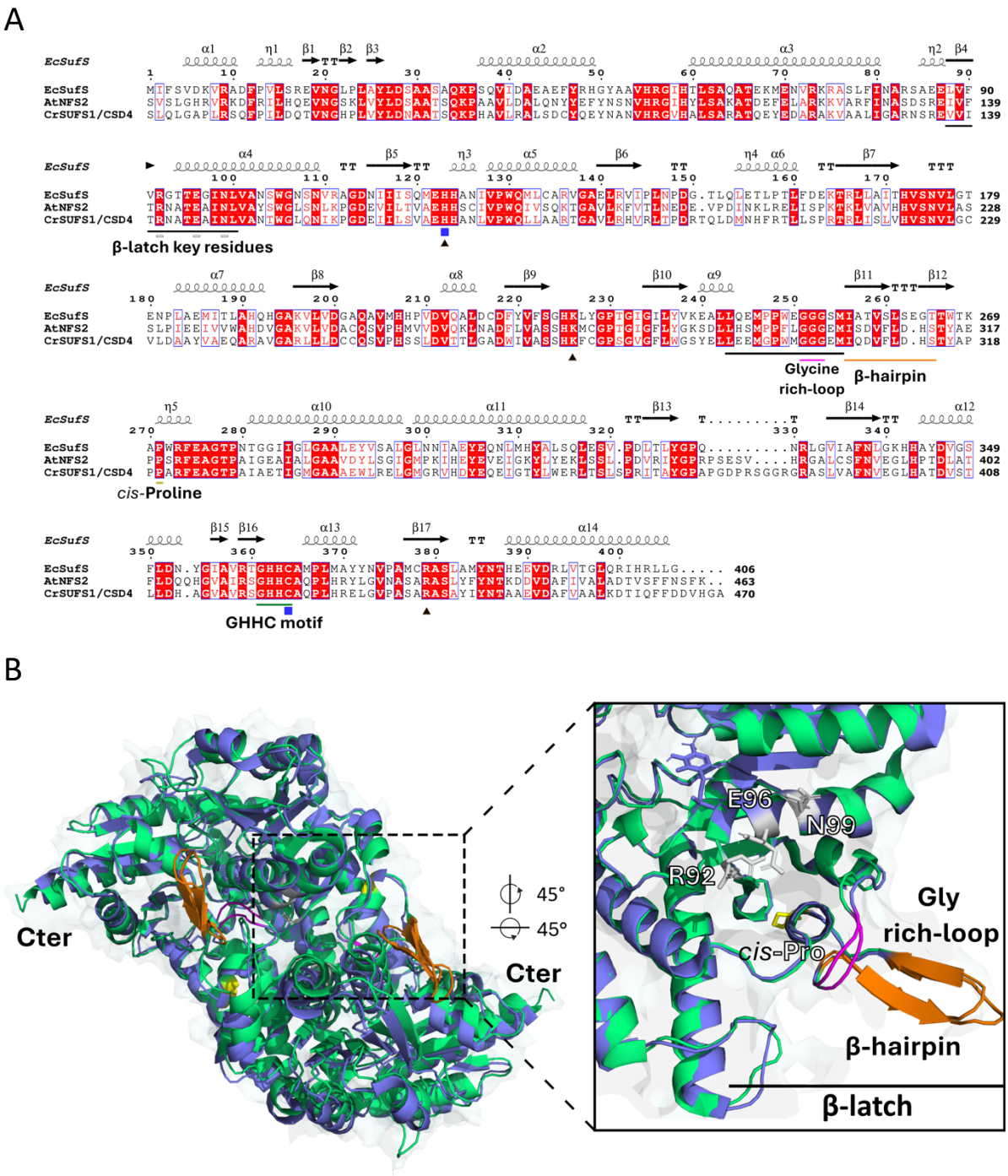
Supplementary material to the paper "Iron-sulfur cluster synthesis in chloroplasts by the SUF system: a mechanistic and structural perspective" by Antoine Kairis, Benjamin Das Neves and collaborators

Methods associated with sequence alignment, structural modelling and superimposition

Multiple sequence alignments were generated by Clustal Omega program from EMBL-EBI [1] and drawn with ESPript 3.0 [2]. Identical residues are represented on a red background. Residues written in red represent residues that are not strictly conserved among all sequences but are similar. AlphaFold2 [3] and AlphaFold Multimer [4] were run via the neurosnap web interface used to predict individual structures and complexes involving *A. thaliana* or *C. reinhardtii* SUF proteins, using 12 to 20 iterations. In the case of monomer, structure prediction was directly retrieved from AlphaFold2 database. For each predicted complex, five models were generated, and the best model was selected based on predicted Template Modelling score (pTM) and interface predicted Template Modelling score (ipTM) as following: 80*ipTM + 20*pTM to obtain a model confidence score described in Table S2. Only predicted complex models with an overall score > 70 were normally considered as relevant, unless otherwise discussed. Pairwise or multiple structural alignments for structural similarity search with proteins of SUF machinery from *E. coli* were performed with US-align for monomers and multimers [5]. We considered structural similarity significant, only if the pairs of structures show TM-score > 0.5. PyMOL v2.5.4 was used to visualize structures and to perform superimposition of multimer complex when US-align was unable to provide exploitable structures.

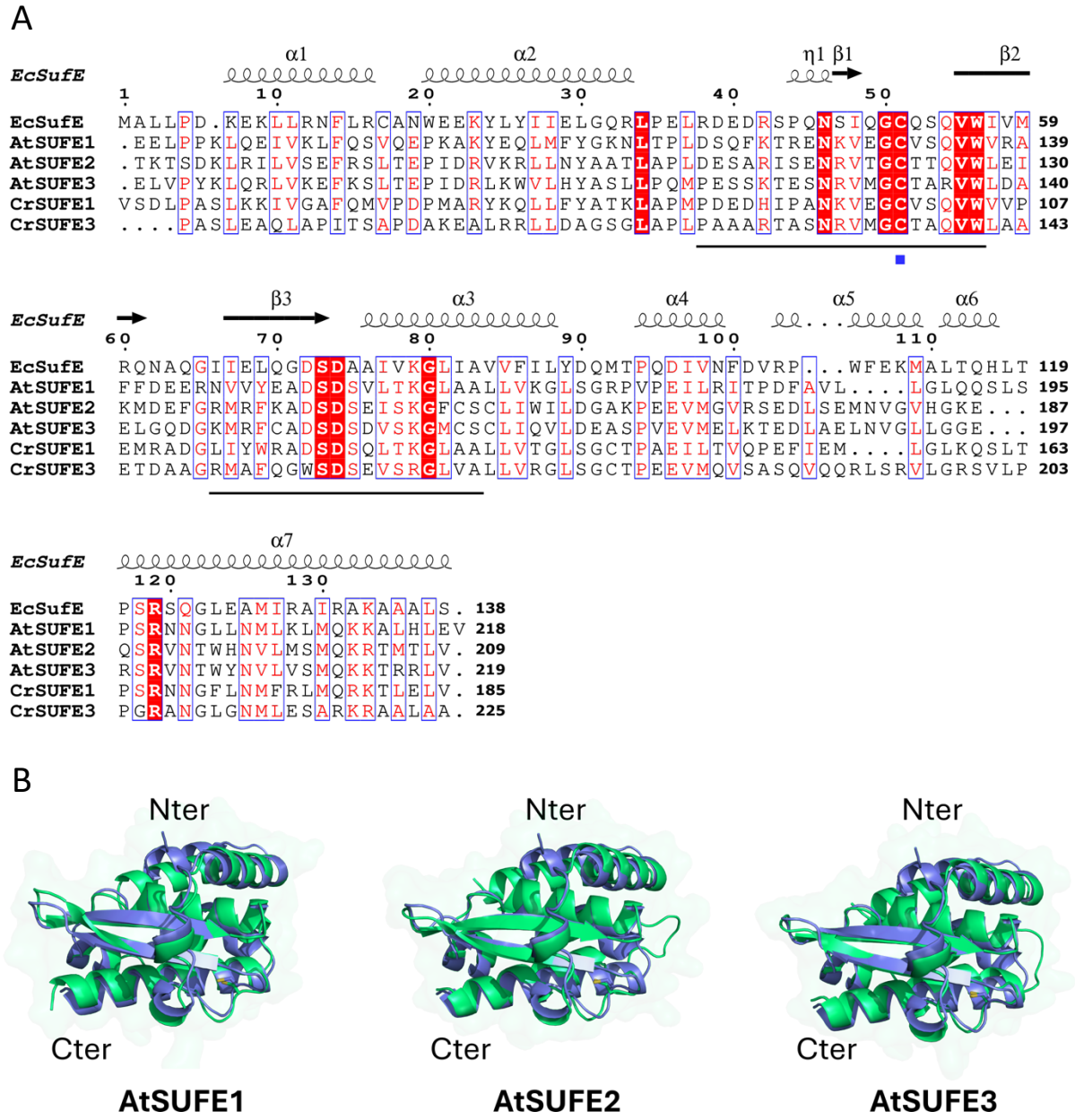


Supplementary figure 1. The structural features of class II cysteine desulfurases are present in NFS2 proteins.

(A) Amino acid sequence alignment of NFS2 from *A. thaliana* (At1g08490) and *C. reinhardtii* (Cre12.g525650), devoid of their targeting sequences, with *E. coli* SufS (P77444). The secondary structures of EcSufS are depicted on top. Identical residues are represented on a red background. The key elements of EcSufS β-latch are represented under the alignment as well as the catalytic residues indicated by a blue square, the PLP-binding or substrate-binding residues indicated by a black triangle and the residues important for the dimer interface underlined in black. Noteworthy, the three conserved residues being part of the β-latch and described as part of the α6 helix in the main text are here represented at the level of the α4 helix because ESPript differentiates α-helices from 3_{10} helices (η).

(B) Structure superimposition of AtNFS2 (in limegreen, PDB entry 4Q75) and EcSufS E250A variant (in slate, PDB entry 6MRI) in a persulfidated state. The TM-score of 0.98 highlights the very high

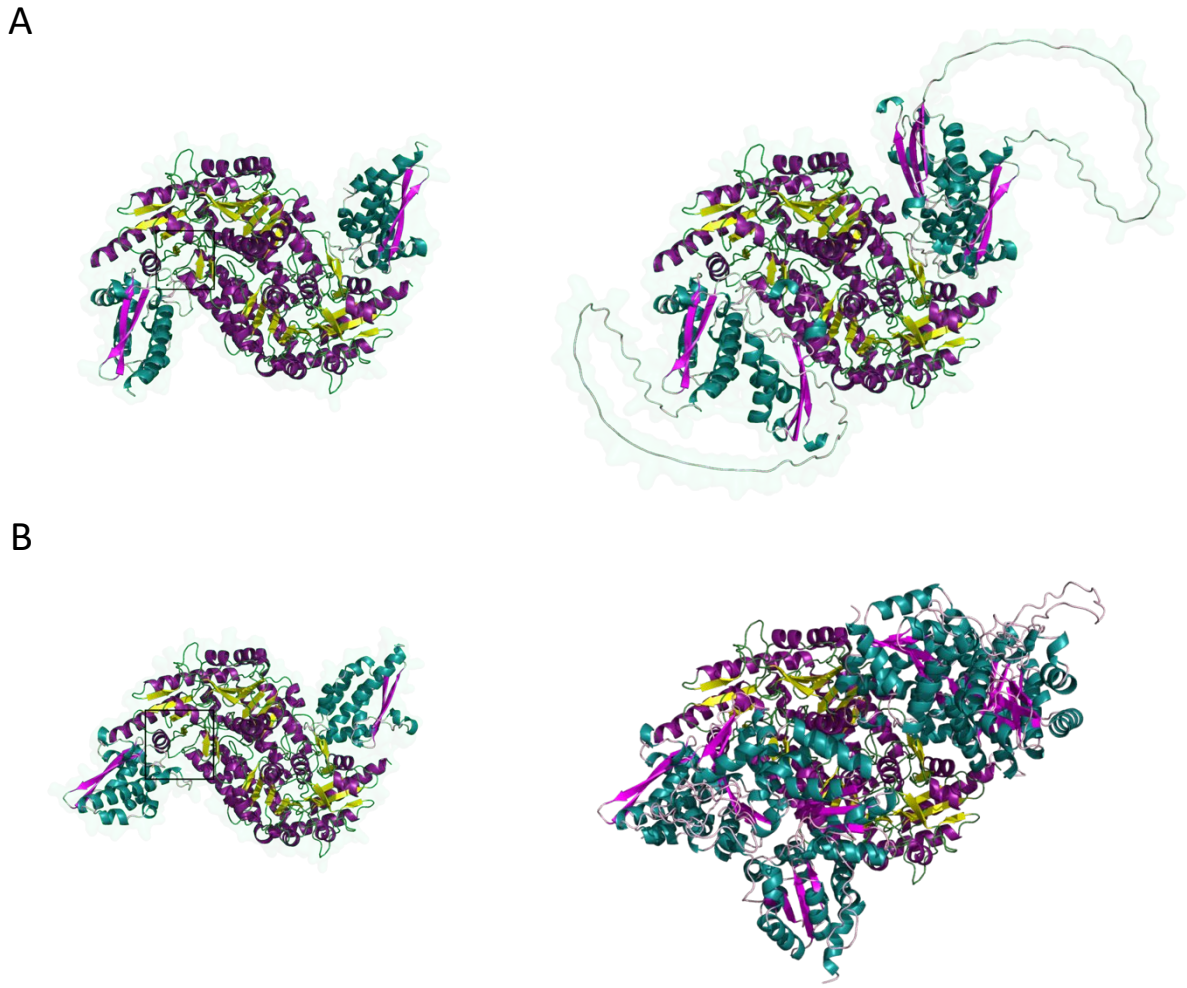
structural similarity between both proteins. A zoom on the β-latch structure of superimposed AtNFS2 and EcSufS is also shown.



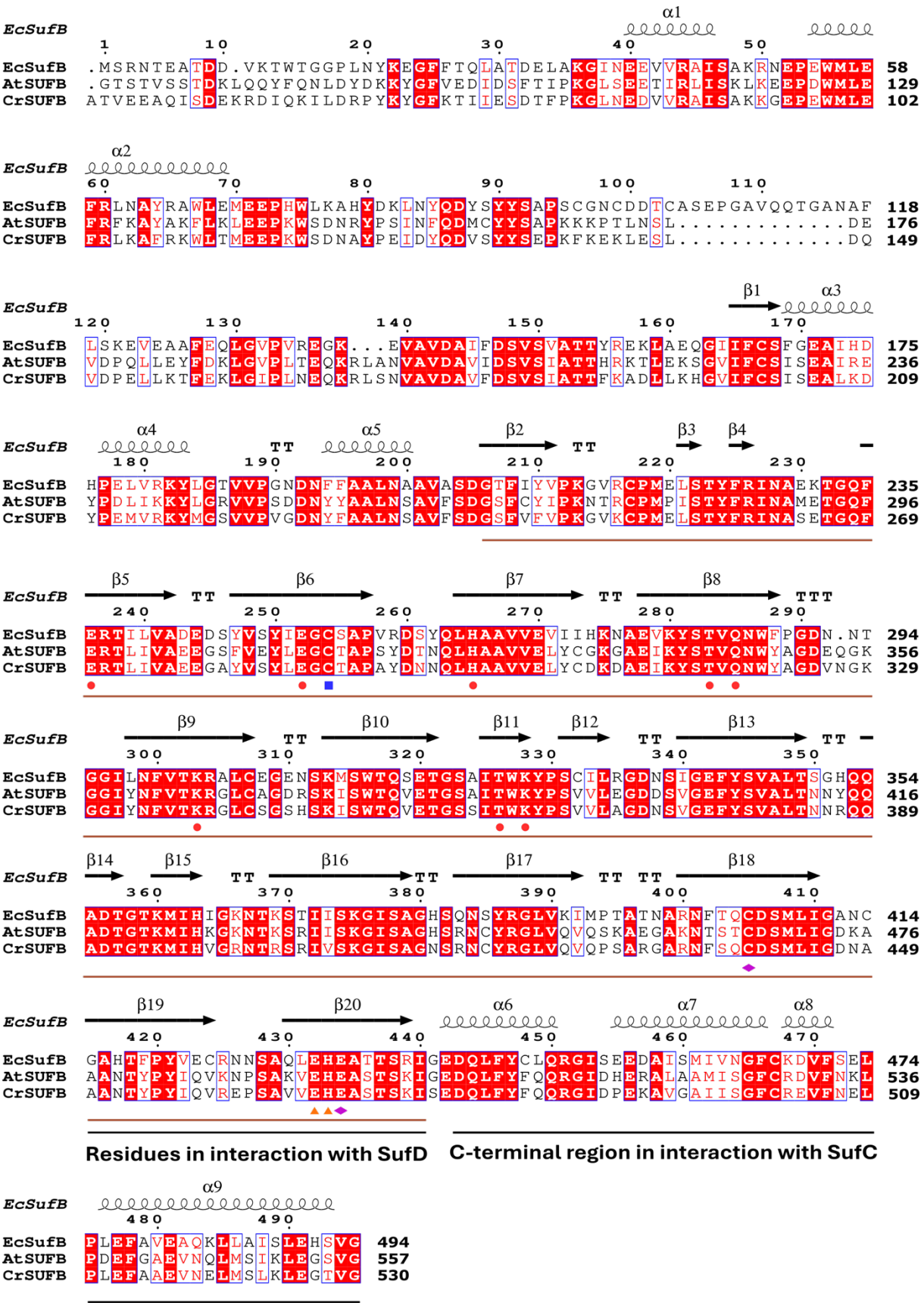
Supplementary figure 2. SufE proteins from photosynthetic organisms display an overall low sequence identity with *E. coli* SufE but a high structural similarity.

(**A**) Amino acid sequence alignment of the SUFE domain of all *A. thaliana* and *C. reinhardtii* SUFE proteins with *E. coli* SufE (P76194). Accession numbers are the following: AtSUFE1 (At4g26500), AtSUFE2 (At1g67810), AtSUFE3 (At5g50210), CrSUFE1 (Cre06.g309717), CrSUFE3 (Cre06.g251450). The secondary structures of EcSufE are depicted on top. Identical residues are represented on a red background. Cysteine involved in the sulfur moiety transfer is indicated by a blue square.

(**B**) Structure superimpositions of SUFE domains (in limegreen) from predicted structure of AtSUFE1 (AlphaFold DB entry AF-Q84W65-F1), AtSUFE2 (AlphaFold DB entry AF-Q9FXE3-F1) and AtSUFE3 (AlphaFold DB entry AF-Q9FGS4-F1) with one unit of EcSufE dimer (in slate, PDB entry 1MZG). The TM-scores of 0.85, 0.82 and 0.83 indicate the high structural similarity between the SufE domain of AtSUFE1/2/3 respectively and EcSufE.

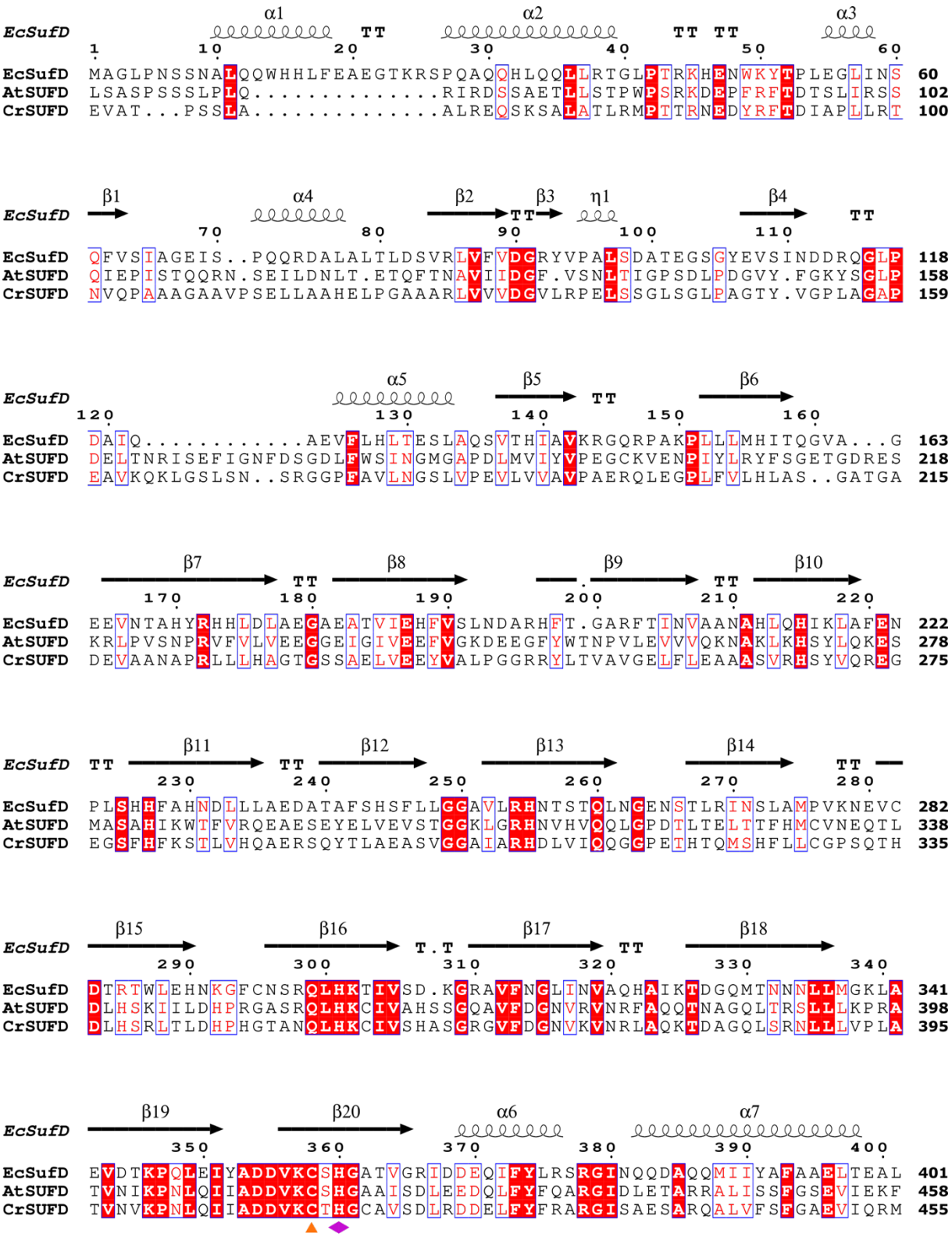


Supplementary figure 3. BolA domain of SUFE1 and NadA domain of SUFE3 are not compromising SufE domain orientation and interaction with NFS2. Best predicted models obtained using AlphaFold Multimer for the AtNFS2–AtSUFE1 complex (**A**) and AtNFS2-AtSUFE3 complex (**B**). On the left are represented models with BolA domain and NadA domain manually removed with PyMol and on the right are represented the full-length predicted complexes. AtNFS2-SUFE1/3 complexes are represented as ribbons. AtNFS2 is colored in dark purple, yellow and green, and AtSUFE1/3 are colored in cyan, purple and pink.

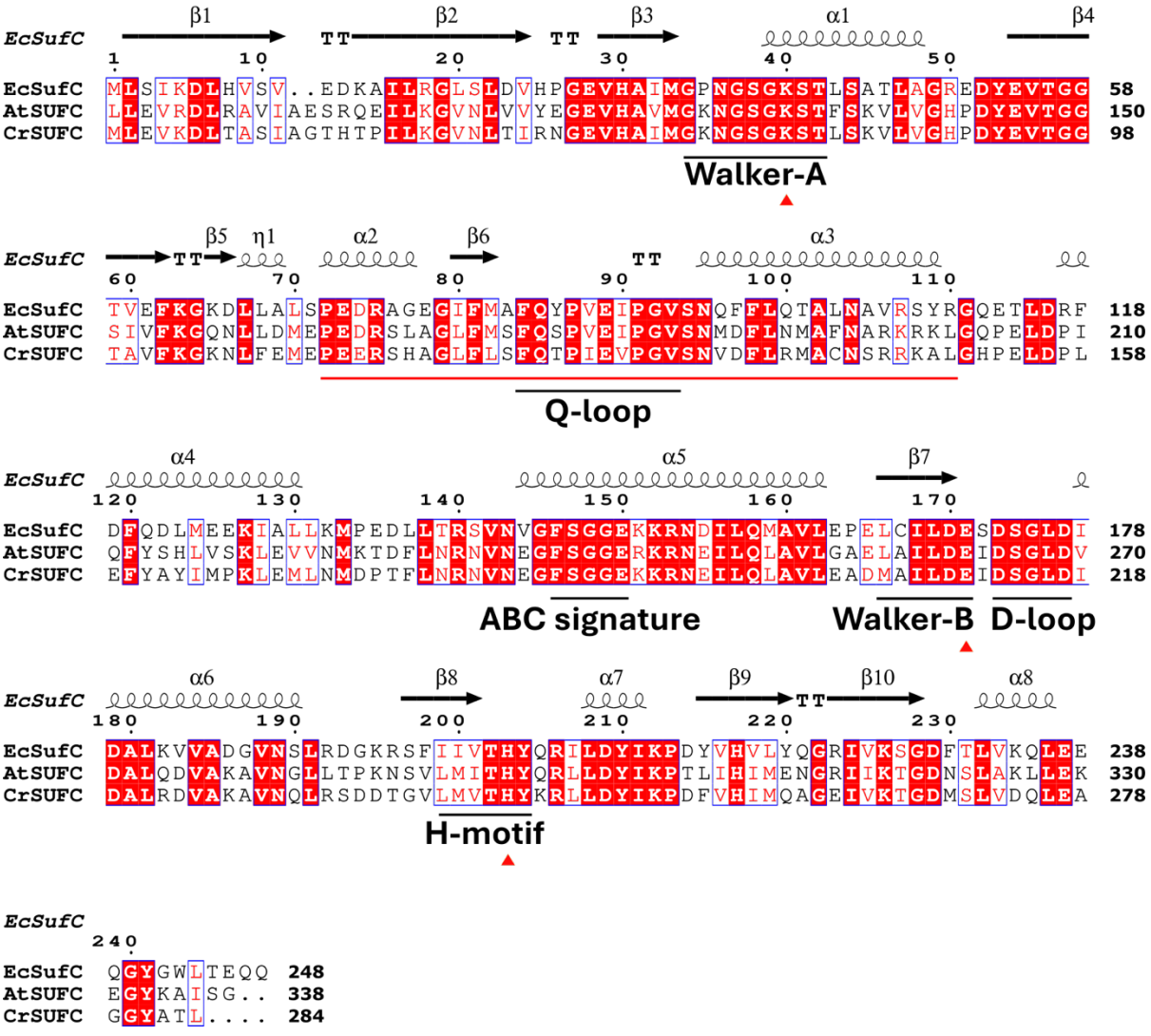


Supplementary figure 4. Sequence alignment of SufB proteins highlighting potential Fe-S cluster binding residues and residues in interaction within the SufC or SufD proteins. Amino acid sequence alignment of SUFB from *A. thaliana* (At4g04770) and *C. reinhardtii* (Cre15.g643600), devoid of their

targeting sequences, with *E. coli* SufB (P77522). The secondary structures of EcSufB are depicted on top. Identical residues are represented on a red background. The blue square indicates the primary cysteine acceptor of sulfur moiety. Purple diamonds indicate the two residues that are the main candidate ligands for Fe-S cluster binding whereas orange triangles highlight other potential ligands. Red dots mark residues with a putative function in the channeling of sulfur atoms from C254 to C405. The β-helix core domain is underlined in brown.

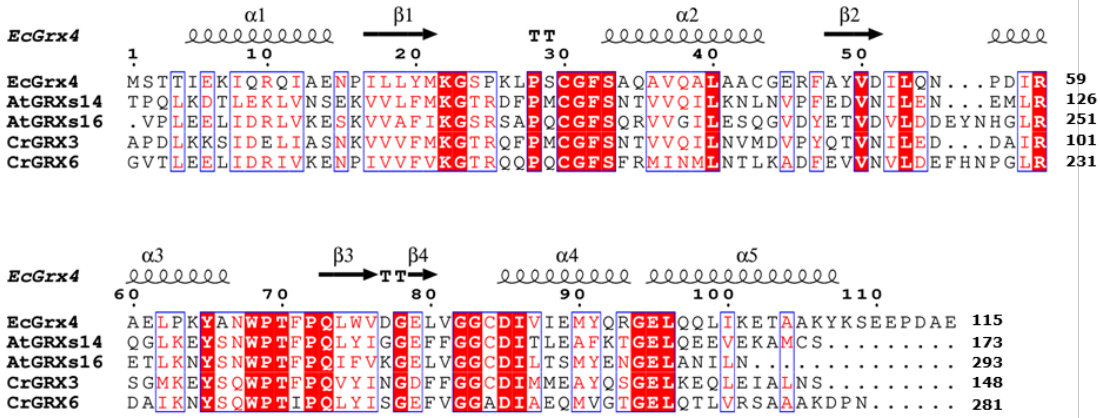

EcSufD
α1 α2 α3
TT TT TT
1 10 20 30 40 50 60
EcSufD MAGLPNSSNALQQWHHLFEAEGTKRSPQAQQHLQQLLRTGLPTRKHENWKYTPLEGLINS 60
AtSUFD LSASPSSSLPLQ.....RIRDSSAETLLSTPWPSRKDEPFRFTDTSLIRSS 102
CrSUFD EVAT...PSSLA.....ALREQSKSALATLRMPTTRNEDYRFTDIAPLLRT 100
β1 α4 β2 β3 η1 β4
TT TT TT
70 80 90 100 110
EcSufD QFVSIAGEIS..PQQRDALALTLDSVRLVFVDGRYVPALSDATEGSGYEVSINDDRQGLP 118
AtSUFD QIEPISTQQRN.SEILDNLT.ETQFTNAVIIDGF.VSNLTIGPSDLPDGVY.FGKYSGLP 158
CrSUFD NVQPAAAGAAVPSELLAAHELPGAAARLVVVDGVLRPELSSGLSGLPAGTY.VGPLAGAP 159
α5 β5 β6
TT
120 130 140 150 160
EcSufD DAIQ.....AEVFLHLTESLAQSVTHIAVKRGQRPAKPLLLMHITQGVA...G 163
AtSUFD DELTNRISEFIGNFDSGDLFWSINGMGAPDLMVIYVPEGCKVENPIYLRYFSGETGDRES 218
CrSUFD EAVKQKLGSLSN..SRGGPFAVLNGSLVPEVLVVAVPAERQLEGPLFVLHLAS..GATGA 215
β7 β8 β9 β10
TT TT TT
170 180 190 200 210 220
EcSufD EEVNTAHYRHHLDLAEGAEATVIEHFVSLNDARHFT.GARFTINVAANAHLQHIKLAFEN 222
AtSUFD KRLPVSNPRVFVLVEEGGEIGIVEEFVGKDEEGFYWTNPVLEVVVQKNAKLKHSYLQKES 278
CrSUFD DEVAANAPRLLLHAGTGSSAELVEEYVALPGGRRYLTVAVGELFLEAAASVRHSYVQREG 275
β11 β12 β13 β14
TT TT TT
230 240 250 260 270 280
EcSufD PLSHHFAHNDLLLAEDATAFSHSFLLGGAVLRHNTSTQLNGENSTLRINSLAMPVKNEVC 282
AtSUFD MASAHIKWTFVRQEAESEYELVEVSTGGKLGRHNVHVQQLGPDTLTELTTFHMCVNEQTL 338
CrSUFD EGSFHFKSTLVHQAERSQYTLAEASVGGAIARHDLVIQQGGPETHTQMSHFLLCGPSQTH 335
β15 β16 β17 β18
T.T TT
290 300 310 320 330 340
EcSufD DTRTWLEHNKGFCNSRQLHKTIVSD.KGRAVFNGLINVAQHAIKTDGQMTNNNLLMGKLA 341
AtSUFD DLHSKIILDHPRGASRQLHKCIVAHSSGQAVFDGNVRVNRFAQQTNAGQLTRSLLLKPRA 398
CrSUFD DLHSRLTLDHPHGTANQLHKCIVSHASGRGVFDGNVKVNRLAQKTDAGQLSRNLLLVPLA 395
β19 β20 α6 α7
350 360 370 380 390 400
EcSufD EVDTKPQLEIYADDVKCSHGATVGRIDDEQIFYLRSRGINQQDAQQMIIYAFAAELTEAL 401
AtSUFD TVNIKPNLQIIADDVKCSHGAAISDLEEDQLFYFQARGIDLETARRALISSFGSEVIEKF 458
CrSUFD TVNVKPNLQIIADDVKCTHGCAVSDLRDDELFYFRARGISAESARQALVFSFGAEVIQRM 455
Residues in interaction with SufB
C-terminal region in interaction with SufC
α8
410 420
EcSufD RDEGLKQQVLARIGQRLPGGAR..... 423
AtSUFD PNREIRDQARNHVKGLL..... 475
CrSUFD GHPVLQKRVQDDVSRTLLSAEPFAASA 482

Supplementary figure 5. Sequence alignment of SufD proteins highlighting potential Fe-S cluster binding residues and residues in interaction within the SufB or SufC proteins. Amino acid sequence alignment of SUFD from *A. thaliana* (At1g32500) and *C. reinhardtii* (Cre12.g513950), devoid of their targeting sequences, with *E. coli* SufD (P77689). The secondary structures of EcSufD are depicted on top. Identical residues are represented on a red background. The purple diamond indicates the main candidate ligand for Fe-S cluster binding whereas the orange triangle highlights another potential ligand.

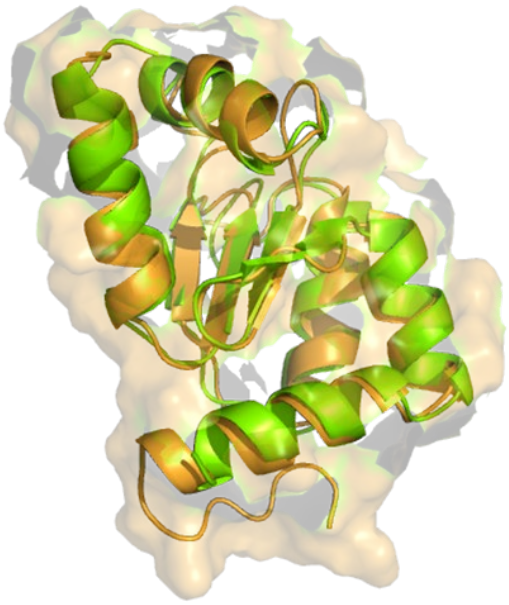


Supplementary figure 6. SUFC proteins from photosynthetic organisms display a high sequence identity with *E. coli* SufC.
Amino acid sequence alignment of SUFC proteins from *A. thaliana* (At3g10670) and *C. reinhardtii* (Cre07.g339700), devoid of their targeting sequences, with *E. coli* SufC (P77499). The secondary structures of EcSufE are depicted on top. Identical residues are represented on a red background. Typical motifs present in SufC are underlined, while strictly conserved residues required for ATPase activity are indicated by a red triangle.

A



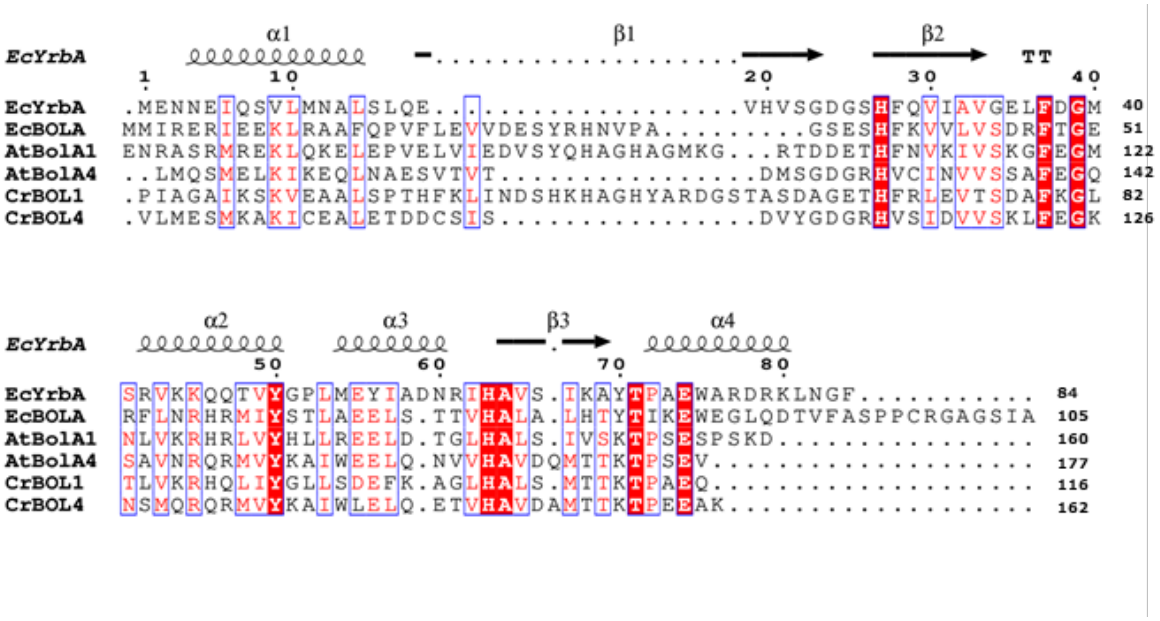
B



Supplementary figure 7. Glutaredoxins from photosynthetic organisms display a high sequence identity with *E. coli* Grx4.

(**A**) Amino acid sequence alignment of GRXS14 and S16 proteins from *A. thaliana* (At3g54900; At2g38270) and *C. reinhardtii* (Cre07.g325743; Cre01.g047800), devoid of their targeting sequences and of the N-terminal endonuclease domain in the case of GRXS16 and GRX6, with *E. coli* Grx4 (P0AC69). The secondary structures of EcGrx4 are depicted on top. Identical residues are represented on a red background.

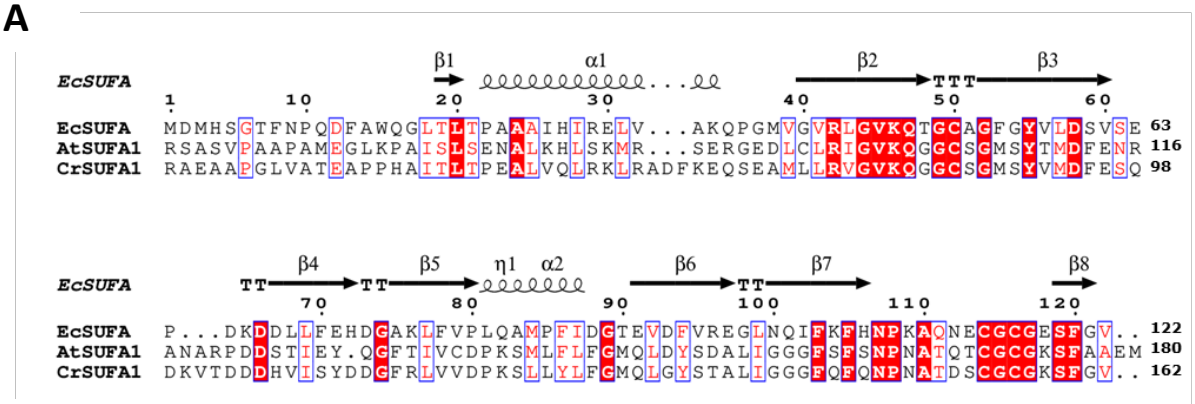
(**B**) EcGrx4 and AtGRXS14 structures superimpose almost perfectly. The experimental crystal structure of EcGRX4 (PDB: 1YKA) and NMR structure of AtGRXS14 (PDB: 2MMA) have been aligned with TM-align. A TM-score of 0.83 was obtained when normalized on AtGRXS14 length (without targeting sequence).



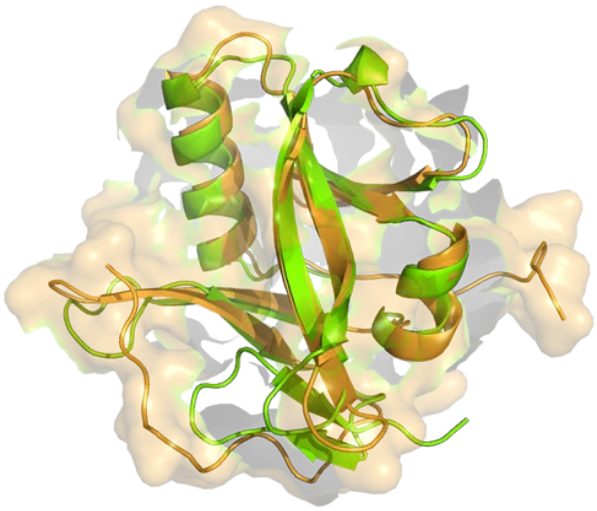
Supplementary figure 8. Only a few residues, including the Fe-S cluster histidine ligands, are conserved among BolA family members.

Amino acid sequence alignment of BOLA1 and BOLA4 proteins from *A. thaliana* (At1g55805; At5g17560) and *C. reinhardtii* (Cre03.g180700; Cre09.g394701), devoid of their targeting sequences, with *E. coli* BolA and YrbA (P0ABE2; P0A9W6). The secondary structures of EcYrbA are depicted on top. Identical residues are represented on a red background. The two histidine residues at position 27 and 63 (*E. coli* YrbA numbering), which serve for ligating the [2Fe-2S] cluster in GRX-BOLA heterodimers, are conserved among a few other residues.

A



B

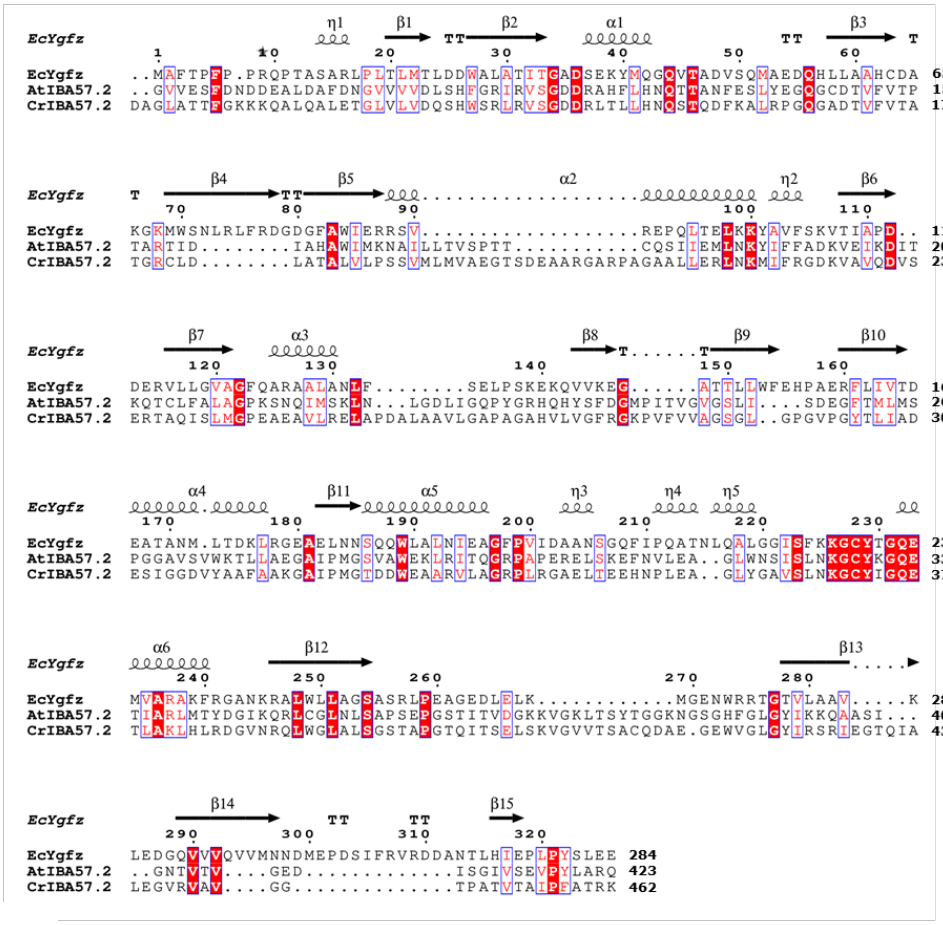


Supplementary figure 9. The regions encompassing the presumed Fe-S cluster cysteine ligands are conserved between SUFA proteins from photosynthetic organisms and *E. coli*.

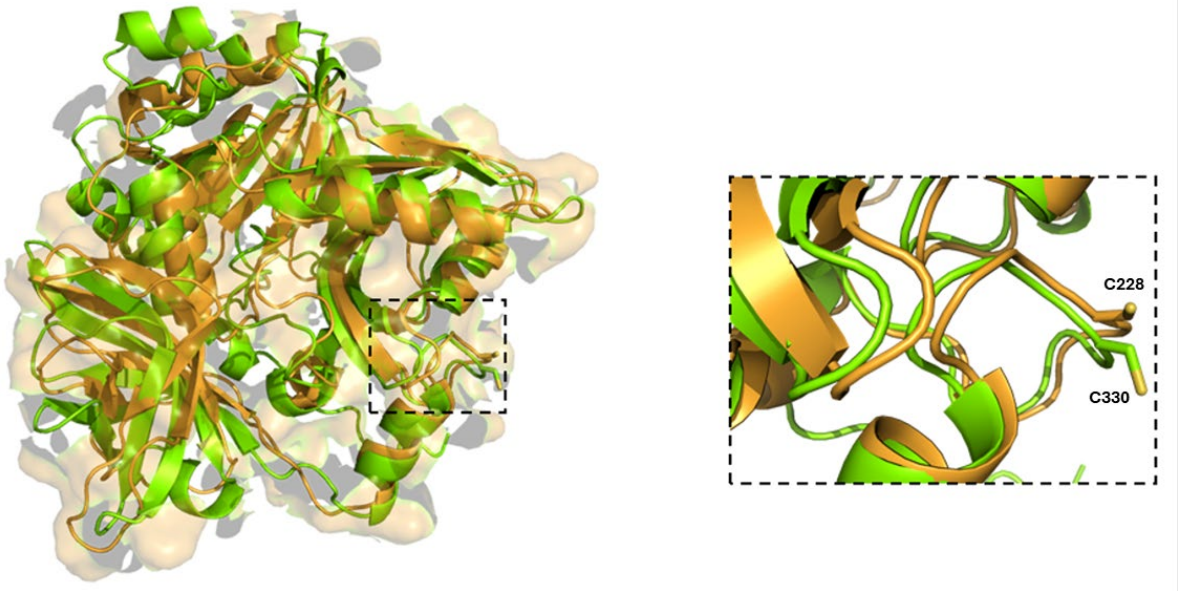
(**A**) The sequences of SUFA1 proteins from *A. thaliana* (At1g10500) and *C. reinhardtii* (Cre06.g800733), devoid of their targeting sequences, have been aligned with EcSufA (P77667). The secondary structures of EcSufA (PDB: 2D2A) are depicted on top. Identical residues are represented on a red background. The cysteine residues proposed to serve as ligands are conserved between the three sequences (C50, C114 and C116, *E. coli* numbering).

(**B**) Structure superimposition of a monomer of AtSUFA1 (in green) from the predicted structure of the AtSUFA1 dimer with one monomer of EcSufA dimer (in orange, PDB: 2D2A). The TM-score of 0.76 (normalized on EcSufA length) indicates the good structural similarity.

A



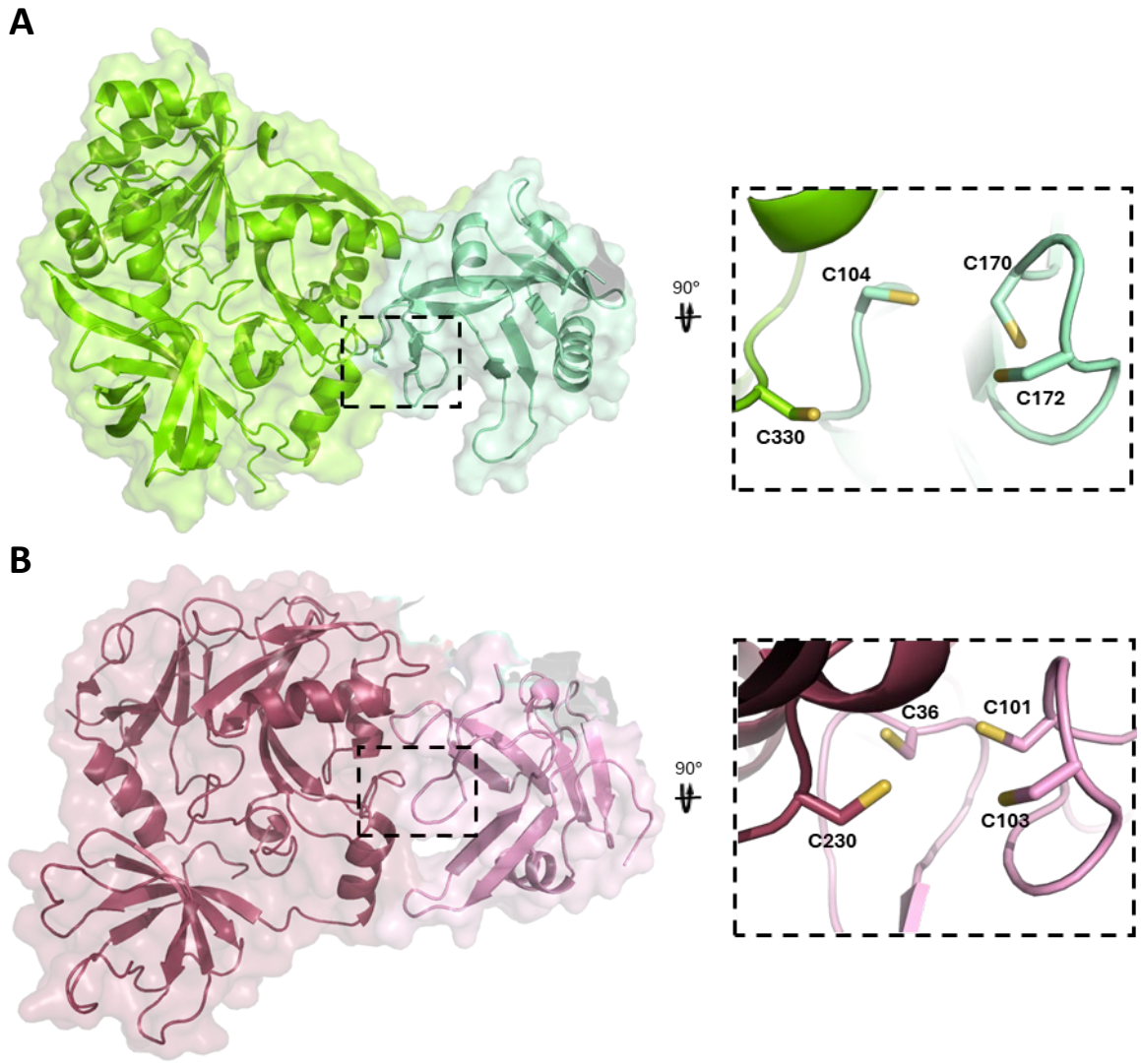
B



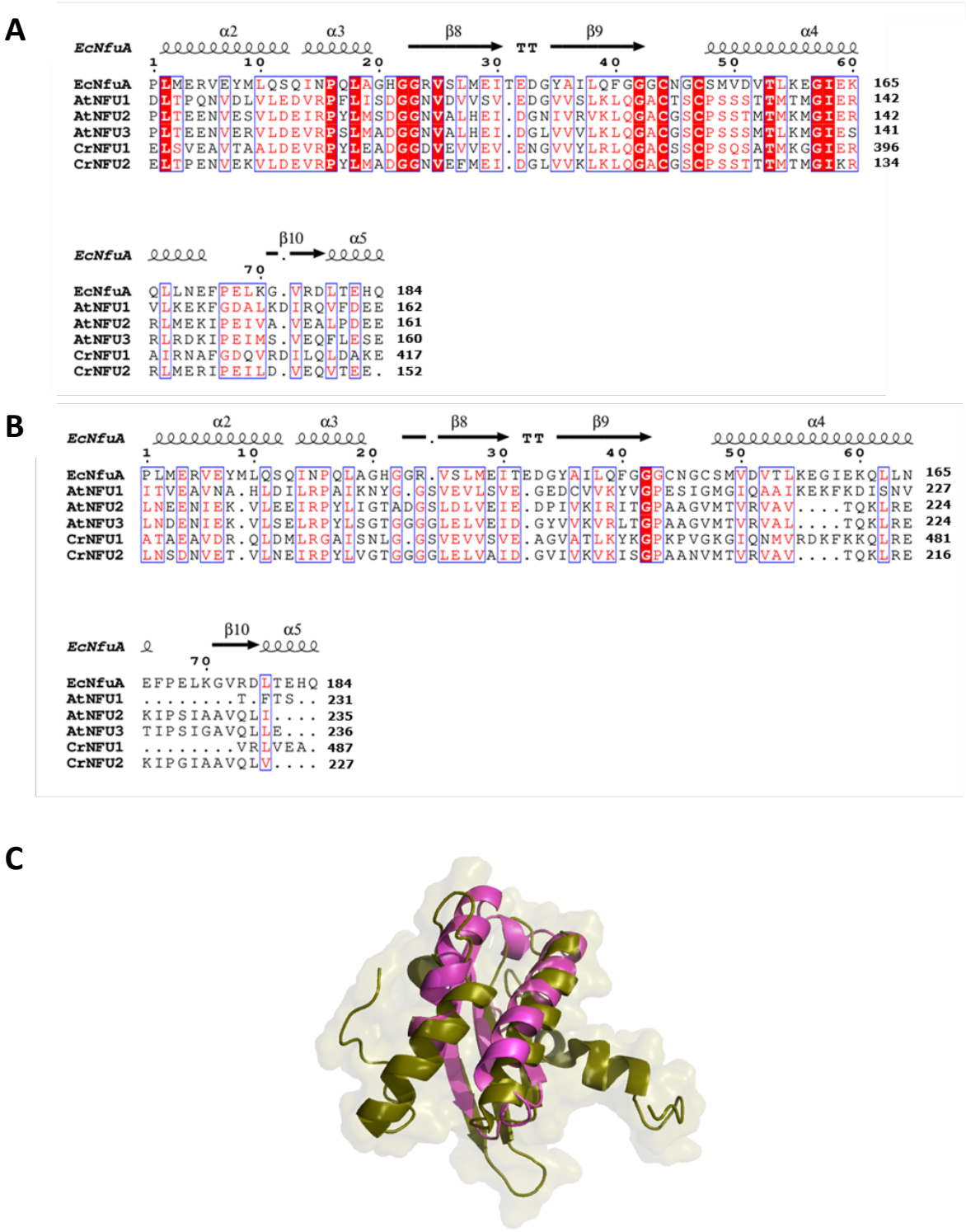
Supplementary figure 10. The overall structure of *Arabidopsis* IBA57.2 is similar to the one of *E. coli* YgfZ despite a low sequence identity.

(**A**) Amino acid sequence alignment of IBA57.2 from *A. thaliana* (At1g60990) and *C. reinhardtii* (Cre12.g552850), devoid of their targeting sequences, with *E. coli* YgfZ (P0ADE8). The secondary structures of EcYgfZ are depicted on top. Identical residues are represented on a red background. Apart

from the KGCYxGQE loop containing the invariant cysteine involved in Fe-S cluster ligation, the overall conservation is extremely low.
(**B**) Superimposition of the predicted structure of *Arabidopsis* IBA57.2 (predicted with AlphaFold2, in green) with the crystal structure of YgfZ from *E. coli* (PDB: 1NRK, in orange). The structures have been superimposed with TM-align and the TM-score is 0.84. The conserved cysteine that is essential for IBA57 function is found in the same region in AtIBA57.2 (C330) and in of YgfZ (C228).



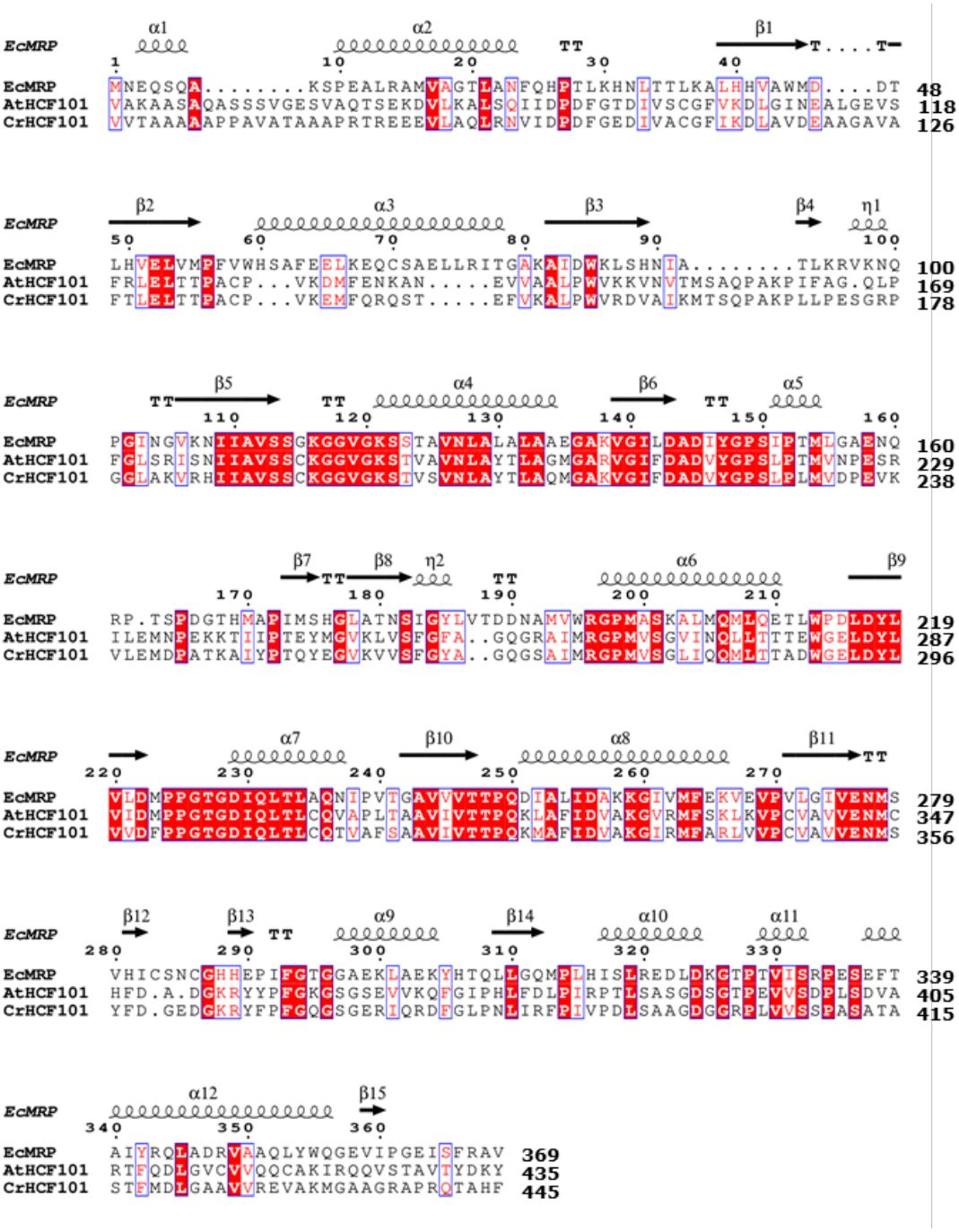
Supplementary figure 11. Comparison of AtIBA57.2/SUFA1 predicted structure and the human mitochondrial IBA57/ISCA heterodimer.
(**A**) Structural prediction of *Arabidopsis* SUFA1-IBA57.2 heterodimer with AlphaFold Multimer is the same as shown in Fig. 6C. IBA57.2 is in green and SUFA1 in turquoise.
(**B**) Structural model of the human IBA57/ISCA2 heterodimer determined by SAXS and docking experiments (SASDB code: SASDGF2). IBA57 is colored in fuchsia and ISCA in pink. Superimposition of the structures with TM-align gives a TM-score of 0.66 when normalized on the length of the human heterodimer. The cysteines of IBA57 and SUFA/ISCA proteins that allows the coordination of an Fe-S cluster by the human complex adopt globally the same position.



Supplementary figure 12. The C-terminal domain of NFU proteins from photosynthetic organisms is degenerated.

The sequences of *A. thaliana* NFU1 (At4g01940), NFU2 (At5g49940) and NFU3 (At4g25910) and of *C. reinhardtii* NFU1 (Cre17.g710800) and NFU2 (Cre18.g748447) devoid of their targeting sequences, have been aligned with EcNfuA (P63020). Identical residues are represented on a red background.
(**A**) Alignment of the N-terminal regular NFU domain with EcNfuA showing on average 28.7% identity.
(**B**) Alignment of the C-terminal degenerated NFU domain with EcNfuA showing on average less than 10% identity.

(**C**) Structural similarity between the regular NFU domain in *Arabidopsis* NFU1 (pink) and SufT (swamp green) from *Mycobacterium tuberculosis*. A TM-score of 0.6 (normalized on the length of the NFU domain, 76 aa) is obtained upon structure superimposition.
Numbers written at the end of each row correspond to the position of the last amino acid on said row as if in the native protein. The identity percentage was calculated as the ratio of conserved amino acids between each NFU domain and the sequence of EcNfuA then averaged.



Supplementary figure 13. Only the P-loop ATPase domain of HCF101 proteins from photosynthetic organisms is conserved when compared to *E. coli* Mrp ortholog.
The sequences of HCF101 from *A. thaliana* (At3g24430) and *C. reinhardtii* (Cre01.g045902), devoid of their targeting sequences, have been aligned with EcMrp/ApbC (P0AF08). Identical residues are represented on a red background. Unlike the P-loop which exhibits an overall good conservation, the N-terminal DUF59 domain (first 100 amino acids) is poorly conserved. Noteworthy the C102 and C128 of HCF101 which are present in many SufT proteins are not conserved/aligned properly with EcMrp.

Protein names	AGI numbers	*E. coli* orthologs
NFS2/ CpNifS	At1g08490	SufS
SUFE1	At4g26500	SufE
SUFE2	At1g67810	-
SUFE3	At5g50210	-
SUFB/ NAP1	At4g04770	SufB
SUFC/ NAP7	At3g10670	SufC
SUFD/ NAP6	At1g32500	SufD
SUFA1/ CpIscA	At1g10500	SufA
GRXS14	At3g54900	Grx4
GRXS16	At2g38270	-
BOLA1	At3g54900	BolA
BOLA4	At2g38270	YrbA/IbaG
NFU1	At4g01940	NfuA
NFU2	At5g49940	-
NFU3	At4g25910	-
HCF101	At3g24430	Mrp
IBA57.2	At1g60990	YgfZ

Table S1. ***Arabidopsis thaliana*** **members of the plastidial SUF machinery and orthologs in** ***E. coli.***

Complex	Model Rank	Mean pLDDT	Max PAE	pTM	ipTM	Confidence score
AtSUFE1-AtNFS2-AtNFS2-AtSUFE1	1	82.54	31.69	0.78	0.71	72.4
	2	82.46	31.69	0.78	0.71	72.4
	3	82	31.7	0.78	0.71	72.4
	4	81.97	31.64	0.78	0.7	71.6
	5	81.62	31.7	0.78	0.69	70.8
AtSUFE2-AtNFS2-AtNFS2-AtSUFE2	1	84.46	31.66	0.86	0.82	82.8
	2	84.58	31.7	0.86	0.82	82.8
	3	84.01	31.7	0.86	0.82	82.8
	4	84.53	31.7	0.86	0.81	82
	5	84.02	31.7	0.86	0.81	82
AtSUFE3-AtNFS2-AtNFS2-AtSUFE3	1	78.62	31.72	0.55	0.45	47
	2	79.57	31.72	0.54	0.44	46
	3	76.73	31.7	0.53	0.42	44.2
	4	79.58	31.72	0.51	0.4	42.2
	5	78.37	31.72	0.5	0.39	41.2
AtSUFBCD	1	83.61	31.7	0.83	0.82	82.2
	2	83.03	31.67	0.79	0.77	77.4
	3	82.17	31.66	0.79	0.75	75.8
	4	83.08	31.69	0.77	0.74	74.6
	5	82.33	31.7	0.77	0.72	73
AtSUFE1BCD	1	79.94	31.7	0.79	0.78	78.2
	2	78.78	31.67	0.78	0.75	75.6
	3	79.59	31.66	0.77	0.74	74.6
	4	78.33	31.69	0.74	0.71	71.6
	5	79.36	31.69	0.74	0.71	71.6
AtSUFE2BCD	1	80.03	31.7	0.79	0.77	77.4
	2	79.35	31.69	0.77	0.73	73.8
	3	78.07	31.69	0.76	0.72	72.8
	4	79.35	31.69	0.75	0.71	71.8
	5	77.37	31.69	0.73	0.68	69
AtSUFE3BCD	1	77.57	31.72	0.63	0.57	58.2
	2	77.02	31.7	0.6	0.53	54.4
	3	75.16	31.72	0.6	0.52	53.6
	4	75.34	31.72	0.58	0.51	52.4
	5	77.75	31.72	0.58	0.5	51.6
AtSUFE1-AtNFS2 tetrameric AtSUFBCD	1	76.17	31.7	0.52	0.45	46.4
	2	74.23	31.72	0.51	0.43	44.6
	3	77.58	31.72	0.5	0.42	43.6
	4	74.21	31.7	0.49	0.42	43.4
	5	75.16	31.72	0.49	0.41	42.6
AtSUFE1-AtNFS2 dimeric AtSUFBCD	1	78.64	31.73	0.63	0.57	58.2
	2	79.42	31.7	0.6	0.53	54.4
	3	77.7	31.7	0.6	0.52	53.6
	4	76.62	31.7	0.59	0.51	52.6
	5	77.04	31.7	0.54	0.45	46.8

Complex	Model Rank	Mean pLDDT	Max PAE	pTM	ipTM	Confidence score
AtBolA4-AtGRXs14	1	82.67	31.7	0.77	0.87	85
	2	83.08	31.72	0.77	0.87	85
	3	81.75	31.72	0.76	0.86	84
	4	77.24	31.7	0.65	0.49	52.2
	5	76	31.7	0.61	0.4	44.2
AtNFU1-AtNFU1	1	92.14	29.02	0.81	0.81	81
	2	88.91	29.22	0.74	0.7	70.8
	3	86.04	29.08	0.59	0.53	54.2
	4	86.36	29.33	0.56	0.51	52
	5	83.6	30.34	0.47	0.37	39
AtNFU2-AtNFU2	1	95.57	19.66	0.85	0.84	84.2
	2	93.1	26.53	0.84	0.83	83.2
	3	92.68	25.78	0.83	0.82	82.2
	4	91.46	28.52	0.83	0.81	81.4
	5	92.17	28.73	0.81	0.79	79.4
AtNFU3-AtNFU3	1	92.33	30	0.81	0.8	80.2
	2	91.47	27.94	0.8	0.79	79.2
	3	92.3	28.92	0.8	0.79	79.2
	4	88.25	30.58	0.69	0.68	68.2
	5	83.35	30.33	0.53	0.5	50.6
CrNFU1-CrNFU1*	1	64.52	31.69	0.32	0.29	29.6
	2	62.74	31.69	0.31	0.27	27.8
	3	60.4	31.7	0.3	0.16	18.8
	4	61.11	31.63	0.28	0.14	16.8
	5	60.28	31.72	0.28	0.13	16
CrNFU2-CrNFU2	1	91.94	29.89	0.76	0.75	75.2
	2	91.46	27.67	0.76	0.75	75.2
	3	89.57	28.17	0.73	0.7	70.6
	4	88.64	27.84	0.67	0.65	65.4
	5	87.77	29.97	0.62	0.59	59.6
AtSUFA1-AtSUFA1	1	84.7	28.88	0.74	0.7	70.8
	2	80.1	30.63	0.7	0.66	66.8
	3	78.15	29.88	0.7	0.65	66
	4	78.53	28.83	0.67	0.56	58.2
	5	75.7	30.22	0.63	0.51	53.4
AtIBA57.2-AtSUFA1	1	93.05	30.98	0.91	0.81	83
	2	92.97	31.25	0.9	0.81	82.8
	3	90.8	31.02	0.88	0.72	75.2
	4	90.95	31.08	0.87	0.69	72.6
	5	89.12	31.16	0.81	0.43	50.6

Table S2. Structural predictions of different protein complexes, using AlphaFold2 [3] and AlphaFold Multimer [4]. The protein complexes tested are found in the first column. Each run

resulted in 5 models, ranked from 1 to 5. The mean predicted Local Distance Difference Test (pLDDT) is shown for each structure, a score above 70 indicates a high confidence in the local structure. The maximum Predicted Aligned Error (PAE), the expected error at residue X in Ångströms (Å) when aligned at residue Y, higher PAE scores correspond to the confidence of Alphafold concerning the relative placement of the domains. The Predicted Template Modelling (pTM) and Interface Predicted Template Modelling (ipTM) scores are metrics proper to the multimeric predictions: the first corresponds to a measure concerning the overall structure while ipTM assesses the accuracy of predicted relative positions of subunits in the protein-protein complex. The confidence score is calculated as [80*ipTM + 20*pTM]. *At* = *Arabidopsis thaliana*, *Cr* = *Chlamydomonas reinhardtii*.

References

[1] F. Madeira, M. Pearce, A.R.N. Tivey, P. Basutkar, J. Lee, O. Edbali, N. Madhusoodanan, A. Kolesnikov, R. Lopez, Search and sequence analysis tools services from EMBL-EBI in 2022, Nucleic Acids Research, 50 (2022) W276–W279.

[2] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Res, 42 (2014) W320–W324.

[3] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, et al., Highly accurate protein structure prediction with AlphaFold, Nature, 596 (2021) 583–589.

[4] R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim, O. Ronneberger, S. Bodenstein, M. Zielinski, A. Bridgland, A. Potapenko, A. Cowie, K. Tunyasuvunakool, R. Jain, E. Clancy, P. Kohli, et al., Protein complex prediction with AlphaFold-Multimer, bioRxiv, (2021) 2021.10.04.463034.

[5] C. Zhang, M. Shine, A.M. Pyle, Y. Zhang, US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes, Nat Methods, 19 (2022) 1109–1115.