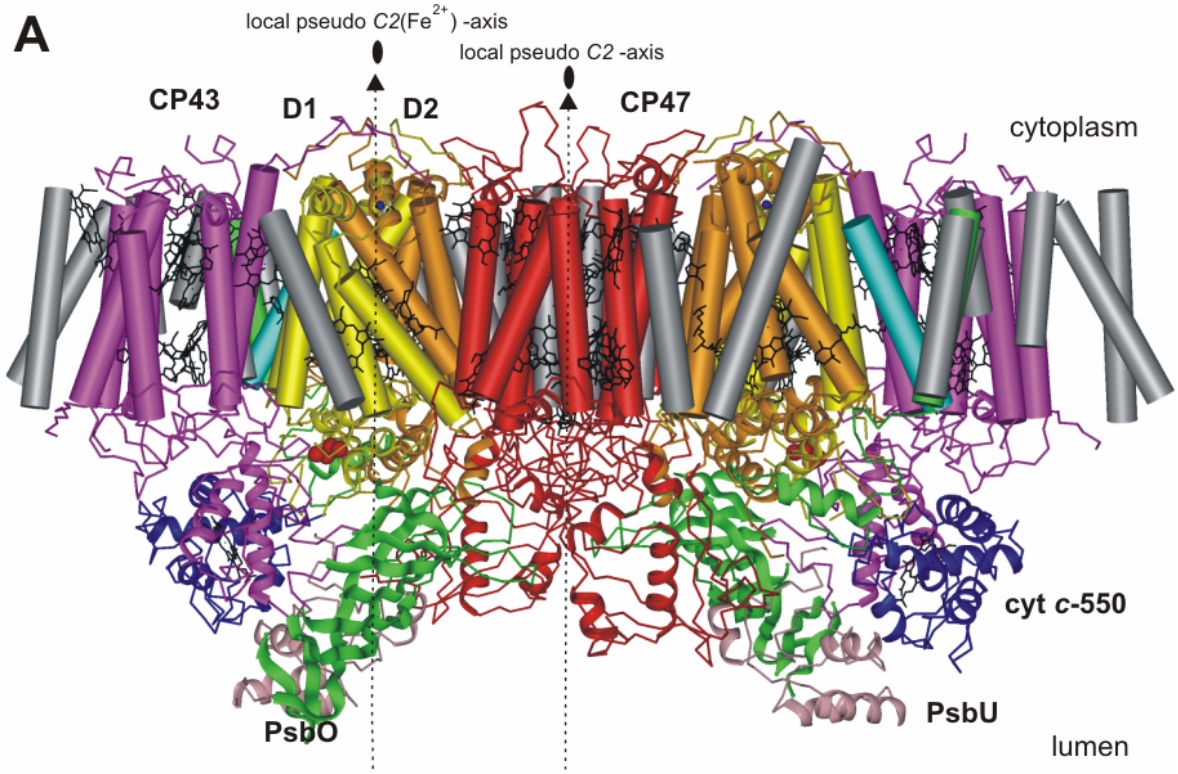
4 Membrane-intrinsic subunits of photosystem II

4.1 An overview - overall arrangement

In this chapter, the overall arrangement of the subunits as well as a detailed description of each assigned subunit will be illustrated.

Each monomer of the PSIIcc dimer with a molecular weight of 375 kDa is composed of 19 different subunits, 16 forming a field of 36 TMH per monomer. In the homodimer, a total of 44 000 atoms have been modelled that belong to protein and to 46 protein-embedded and identified cofactors. The 22 TMH of the central subunits D1 and D2 as well as the antenna subunits CP43 and CP47 are related by a local pseudo-*C2*(Fe^{2+}) symmetry (Fig. 4.1A and B). This axis is orientated parallel to the local-*C2* axis that relates the two monomers and passes through the non-haem iron (Fe^{2+}) on the cytoplasmic side of the membrane. The pseudo-*C2*(Fe^{2+}) axis relates D1 and D2 that are form the core of the complex (2 x 5 TMH) and the antenna subunits CP47 and CP43 (2 x 6 TMH). The low molecular weight subunits are located at the periphery and at the dimer interface. The three membrane-extrinsic and small membrane-intrinsic subunits break the local pseudo-*C2*(Fe^{2+})-fold symmetry of the monomer.



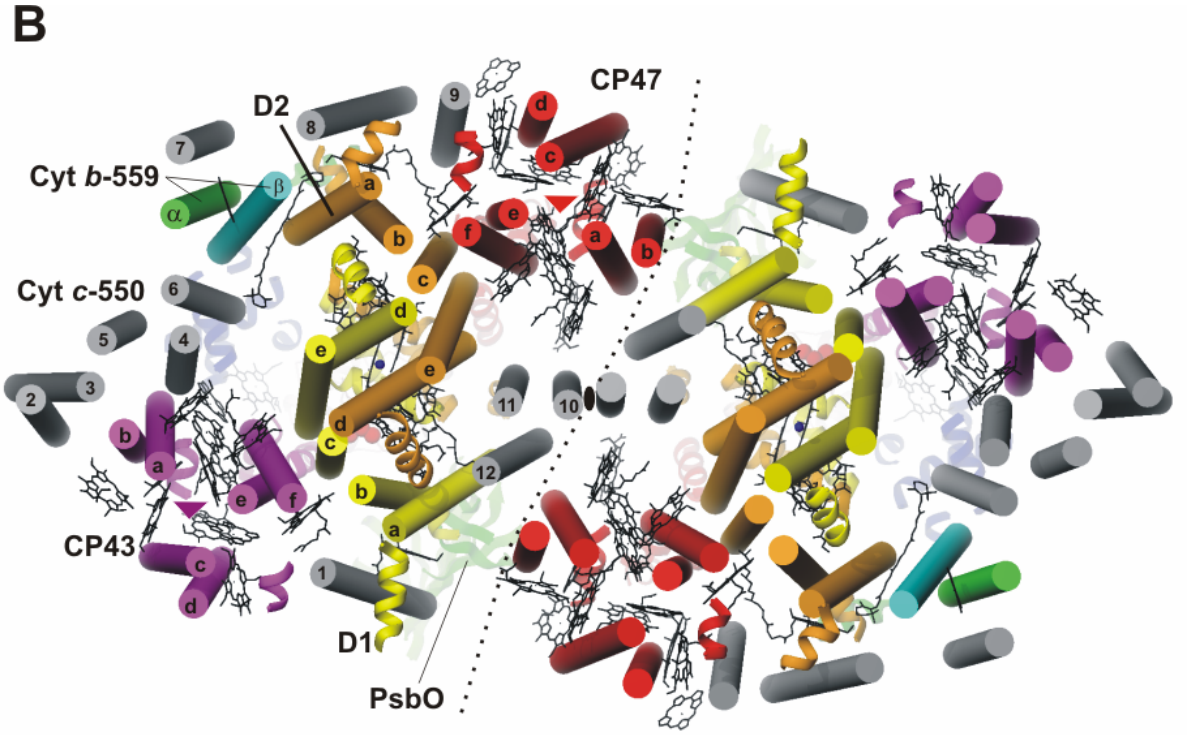


Fig. 4.1: Overall view of the structure of the dimeric PSIIcc. **(A)** View on the membrane plane. One monomer is related by the pseudo-*C2* axis to the other monomer. Assigned subunits in magenta (CP43), yellow (D1), orange (D2), red (CP47), green and cyan (cyt *b*-559, α- and β-chain), unassigned TMH (grey). Local pseudo-*C2*(Fe^{2+}) axis passes through the non-haem Fe^{2+} site (blue dot). Chl*a*, haem and Car are drawn as black wire models.. **(B)** Rotation of the monomer in Fig. 1A about the horizontal axis. View from the cytoplasmic side. Colouring code for the protein subunits as in Fig. 1A and membrane-extrinsic subunits in green (PsbO), pink (PsbU) and blue (cyt *c*-550) are on the bottom. Non-helical segments are not shown. The local pseudo-*C2* axis is indicated by a black ellipse. The dashed line indicated the monomer-monomer-interface of the dimer. Triangles (magenta and red) indicate positions of pseudo-threefold axes within CP47 and CP43

A total of 39 Chl*a* molecules were assigned per monomer. The core antenna subunits CP47 and CP43 harbour 16 Chl*a* in CP47 and 13 in CP43, arranged in two layers near the cytoplasmic and lumenal sides of the membrane. In addition we could locate a carotenoid close to D2 and cyt *b*-449. In cyt *b*-449 and cyt *c*-550 both haem porphyrins were positioned. Due to the insufficient resolution, Mn and Ca cations of the Mn-Ca-cluster could only tentatively be placed in the electron density.

4.2 Nomenclature

The nomenclature of the secondary structure elements of D1, D2, CP47 and CP43 is as follows. TMH are labelled with letters **a** to **e** according to their position in the amino acid

sequence of D1, D2 and in analogy for CP47 and CP43 with the letters **a** to **f**. Loops connecting α-helices are labelled with the letters of the flanking α-helices. Secondary structure elements lying between TMH are numbered (in parentheses) according to their position in the amino acid sequence. The first membrane-extrinsic α-helix located in the loop domain between TMH **a** and **b** would be labelled **ab(1)**. Secondary structure elements in the three membrane-extrinsic subunits are numbered according to their type from the N- to the C-terminus. Due to the fact that the 12 small subunits are not yet assigned and encode for only one TMH, they are numbered, starting from the first unassigned TMH close to D1 going along the periphery to the D2 side. The remaining unassigned TMH **10** to **12** are located at the monomer-monomer interface in the dimer.

Since the numbering of antenna Chl*a* in previous reports (Vasil'ev *et al.*, 2001; Zouni *et al.*, 2001) was not systematic, we introduced a new numbering within PSIIcc. It was introduced to allow fast assignment of Chl*a* to CP47 and to CP43. The new system takes into account that with higher resolution data, additional Chl*a* could be discovered and systematically integrated in the existing system. Numbering of Chl*a* in CP43 and CP47 reflects symmetry due to the pseudo-*C2*(Fe^{2+}) rotation axis, Chl*a*(N) in CP47 being related to Chl*a*(N + 20) in CP43.

Amino acid sequence alignments shown in this work were performed with the program CLUSTALW (Thompson *et al.*, 1994) applying BLOSUM62 similarity matrix and the standard penalty gaps of the program. All available amino acid sequences of cyanobacteria and higher plants were retrieved from GenBank database. To create the consensus sequence of subunits present in cyanobacteria sequences of the following species were used *Chlamydomonas reinhardtii*, *Synechococcus* sp. WH8102, *Synechocystis* sp. PCC6803, *Nostoc sp*. PCC7120 and *T. elongatus* BP-1, in various "algae": *Cyanophora paradoxa, Euglena gracilis, Chaetosphaeridium globosum, Cyanidioschyzon merolae, Cyanidium caldarium, Chlorella vulgaris, Porphyra purpurea, Guillardia theta, Nephroselmis olivacea, Odontella sinensis, Mesostigma viride* and from higher plants the sequences of coniferous plants: *Pinus thunbergii, Pinus koraiensis,* of angiosperms, monocotyledonous *Oryza sativa, Zea mays* and *Triticum aestivum* as well as dicotyledonous *Arabidopsis thaliana, Nicotiana tabacum, Spinacia oleracea, Lotus japonicus* and *Atropa belladonna* and of other plants like *Psilotum nudum, Adiantum capillus-veneris, Anthoceros formosae, Calycanthus fertilis, Odontella sinensis*. As the small subunits composition differs between cyanobacteria and higher plants, only sequences of cyanobacterial subunits were aligned. For clarity all

alignments are restricted to two cyanobacterial sequences (*T. elongatus* and *Synechocystis* sp. PCC6803) and two representatives of higher plants (*S. oleracea* and *A. thaliana*) and the consensus sequence. Other reasons for this choice of species are the large number of available mutation studies on *Synechocystis* sp. PCC6803 and on *S. oleracea* protein subunits in combination with spectroscopic studies providing plenty of information that is currently less profound than for *T. elongatus*.

The colour code used for all alignments is the following: hydrophobic amino acids (green), polar amino acids (orange), basic (blue), histidine (light blue), asparagine (dark blue), glutamine (grey), acidic (red), glycine (violet), cysteine (brown) and proline (yellow).

4.3 The reaction centre

The RC of PSII is formed by two subunits: PsbA (D1) and PsbD (D2). Their original naming D1/D2 derived from the two diffuse bands on SDS-polyacrylamide gels. The five TMH (**a** to **e**) of D1 and D2 are arranged in semicircles related by a local pseudo-*C2*(Fe^{2+}) axis normal to the membrane plane and passing through the non-haem Fe^{2+} (Fig. 4.1A and B), similar as in the well resolved (2.3 Å) PbRC (Deisenhofer *et al.*, 1995). Both subunits harbour all cofactors of the ETC and have structural homology to the PbRC subunits L and M. Based on the conservation of the amino acid sequence, corresponding functional regions were suggested for PSII (Michel and Deisenhofer, 1988). Furthermore D1 and D2 show similarity to the RC domain of PsaA and PsaB of PSI suggesting that gene duplications from a common ancestor gave rise to these subunits in PSII and PSI (Schubert *et al.*, 1998). Interestingly, the D1 and D2 proteins possess 32% identity (52% similarity), whereas the L and M subunits show 30% identity and 46% similarity. Besides the cofactors of the ETC, the PSII RC harbours two carotenoid molecules (Gounaris *et al.*, 1990; Bassi *et al.*, 1993), exclusively in *trans*-configuration (Fujiwara *et al.*, 1987). One function of the carotenoids is quenching of singlet oxygen and additionally they could be involved in secondary cyclic electron transfer *via* cyt *b*-559 (Schweitzer and Brudvig, 1997; Tracewell and Brudvig, 2003).

Both RC subunits are the best defined in the electron density. As both subunits are at the heart of PSII, they receive the greatest stabilisation, which is reflected in lower degree of freedom. Consequently, all side chains of amino acids belonging to these subunits could almost completely be modelled.

4.3.1 PsbA - D1 protein

The *psbA* gene product has a number of trivial names in the literature: According to the binding of plastoquinone (Q_B) it is named Q_B-protein or due to its diffuse band on SDS-PAGE, D1 protein. Interestingly the genome of *T. elongatus* contains three different copies of the *psbA* gene (named *psbA1* to *psbA3*). For a more detailed discussion of the functional consequences of these three copies, see chapter 8.1.

The D1 subunit shows 33% sequence similarity to the L subunit of PbRC. It was shown that D1 correspond functionally to the L subunit by labelling D1 with azidoatrazine that binds to the same site as Q_B (Pfister *et al.*, 1981). Comparing the sequences of the D1 protein from *T. elongatus* and spinach, 85% identity and 94% similarity are found. This high degree of conservation indicates once more the importance of these proteins.

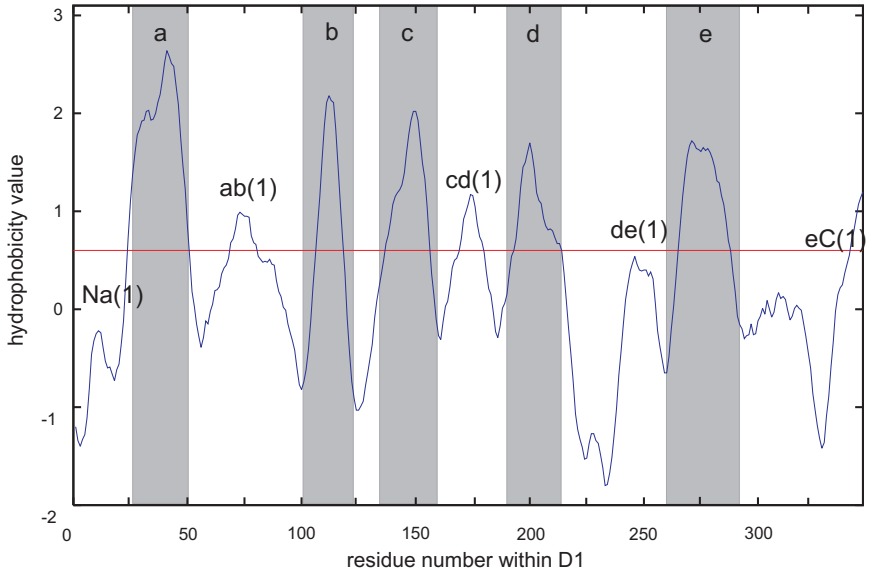


Fig. 4.2: Hydropathy plot of the D1 protein of *T. elongatus* according to (Kyte and Doolittle, 1982) with a window of 10 amino acids. The red line indicates the average hydropathy for all amino acid residues. Hydrophobic patches are marked are above this line. They coincide well with TMH-**a** to -**e** alphabetically ordered according to their occurrence in the sequence.

It is not surprising that high conservation is observed within the TMH as a number of coordinating residues are located within these regions. On the lumenal side of D1 significant differences are observed between the D1 and L protein. The D1 sequence shows an insertion in the amino acid sequence within the α-helices **cd** and **eC** (Fig. 4.3)

The hydropathy plot reveals clearly the presence of 5 TMH (Fig. 4.2 and Fig. 4.3) in analogy to the L subunit of the PbRC. The overall arrangement of TMH of D1 resembles the core of L subunit of PbRC. The outermost TMH-**a**, -**b** and -**d** are nearly straight, whereas TMH-**e** is smoothly curved (Fig. 4.6). The most distorted is TMH-**c**, with a kink of about 30° closer to C-terminal end of this TMH (Fig. 4.6). The kink is probably caused by D1-Pro162 which is conserved and found at the same position in PbRC.

The presence of five solvent exposed α-helices could be also derived from the hydropathy plot. For α-helices **ab**, **cd** and **de** a significantly increased Hydropathy value is observed compared to loop regions (Fig. 4.2), and α-helices at N- and C-termini (**Na** and **eC**) are also indicated, albeit with weaker hydrophobicity values than the other three α-helices. A more detailed look at these α-helices reveals that one side is mainly hydrophobic and interacts with the surface of D1 membrane, whereas the other side is more hydrophilic and solvent exposed.
Cyanobacterial D1 protein is synthesised in a precursor form containing a C-terminal extension of 16 amino acids (Nixon *et al.*, 1992). The D1 precursor protein is inserted into the thylakoid membrane, associates with other PSII components and is cleaved after Ala344 (Fig. 4.4) by carboxyl-terminal protease (CtpA; EC 3.4.21.102), a C-terminus processing protease, to yield the mature D1 protein (Anbudurai *et al.*, 1994). In a recent study using a Δ*ctpA* strain of *Synechocystis* sp. PCC6803, it was not possible to detect any manganese atoms associated with these PSIIcc complexes, indicating that D1 processing is required for stable interaction with manganese. The membrane-extrinsic polypeptides, PsbO, PsbU and PsbV, were also absent, suggesting that the C-terminal extension of non-processed D1 precludes their binding (Roose and Pakrasi, 2004). The mature D1 protein is predicted to have a molecular mass of 39.7 kDa (344 amino acids).

Characteristic for D1 is a high turnover rate (Aro *et al.*, 1993), which is possibly related to the fact that PSII is susceptible to photo-damage. This might be an argument for gene fission theory suggesting that the PsaA/B proteins of PSI were precursors for D1 and D2 of PSII. The PsaA/B proteins each contain a 6 TMH-long antenna domain and a 5 TMH-long RC-domain (Fig. 1.3), harbouring nearly all of the 128 cofactors and consequently replacement of only one RC proteins much more efficient than that of protein with molecular mass of ~ 81 kD. The D1 protein is suggested to be first degraded at the loop **d-e** close to the Q_B-binding site (de las Rivas *et al.*, 1992). It is not yet understood how the degraded D1 protein is removed from the PSII complex and how the freshly synthesised D1 protein is incorporated into PSII

complex. It must be a rather intriguing process as the PSII complex has to be partially disassembled, the freshly synthesised D1 protein has to be inserted into the complex and all cofactors have to be incorporated. The high turnover of the D1 protein is in contrast to D2. The heterodimeric organisation of the RC in D1 and D2 subunits is more economic than a homodimeric would be, because only one protein is degraded and has to be replaced. Consequently, nature limited protein damage to only one protein of the heterodimer.

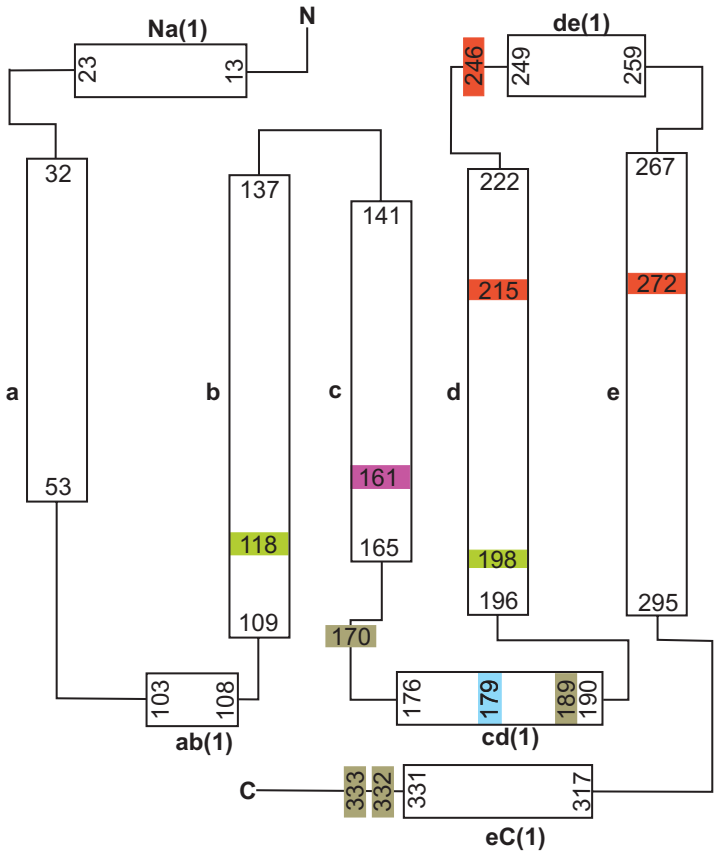


Fig. 4.3: Topography of D1: α-helices are indicated by rectangles, TMH are denoted a to e. The N-terminus (N) is on the cytoplasmic (top) side and C-terminus (C) on the lumenal side (bottom). Chl*a* coordinating histidine residues are marked with green background, other indirectly coordinating residues in D1 (Thr179) are indicated by light-blue background. His215, His272 and Tyr246 coordinating the non-haem iron are marked with red background. Residues coordinating the Mn-Ca-cluster are indicated by grey background and Tyr_Z (Tyr161) is marked with violet background.

Functionally, D1 provides the protein scaffold for the majority of cofactors involved in electron transport (Fig. 4.3). The topography in Fig. 4.3 illustrates the protein folding and location of cofactors involved in cofactor coordination. From the lumenal side to the cytoplasmic side the cofactors are: Mn-Ca-cluster, Tyr_Z, P_{D1}, Chl_{D1}, $Pheo_{D1}$, non-haem iron

and Q_B. This enumeration of cofactors reflects the relevance of D1, as all redox-active cofactors from the acceptor to the donor side are coordinated by its amino acid residues of this subunit. In turn, this explains the need for a high turnover rate of D1, as highly reactive radicals are produced in side reactions of the photosynthetic cycle which are able to destroy the protein matrix.

The N-terminus of D1 is located on the cytoplasmic side and folded to some extent into amphiphatic α-helix **Na** that solvent is exposed with its hydrophilic side. The quality of the electron density did not allow tracing the first seven amino acids of D1, probably due to a higher degree of flexibility of the N-terminus. The main secondary structure characteristics of D1 are five TMH-**a** to -**e** arranged in two semicircles interlocked in a handshake motif and related to the five TMH of D2 by the pseudo-C*2*(Fe^{2+}) axis.

A more detailed description of protein surrounding of cofactors within D1 will be presented and discussed in chapter 1 and the following chapters. TMH-**a** does not contain any coordinating residue, whereas all other TMH harbour at least one cofactor coordinating amino acid. TMH-**a** is located on the periphery of the field of TMH formed by D1 and D2 (Fig. 4.1B). It is close to unassigned TMH **1** and **12** as well as to TMH-**b**. Between TMH-**a** and TMH-**b** the longest lumenal loop region **ab** is found. It is partially folded into a short α-helix **ab(1)**, oriented perpendicular to the membrane plane. D1-His118 located on TMH-**b** coordinates the peripheral $Chlz_{D1}$. The cytoplasmic side of TMH-**b** is close to TMH-**f** of CP43 and stabilised by hydrophobic interaction.

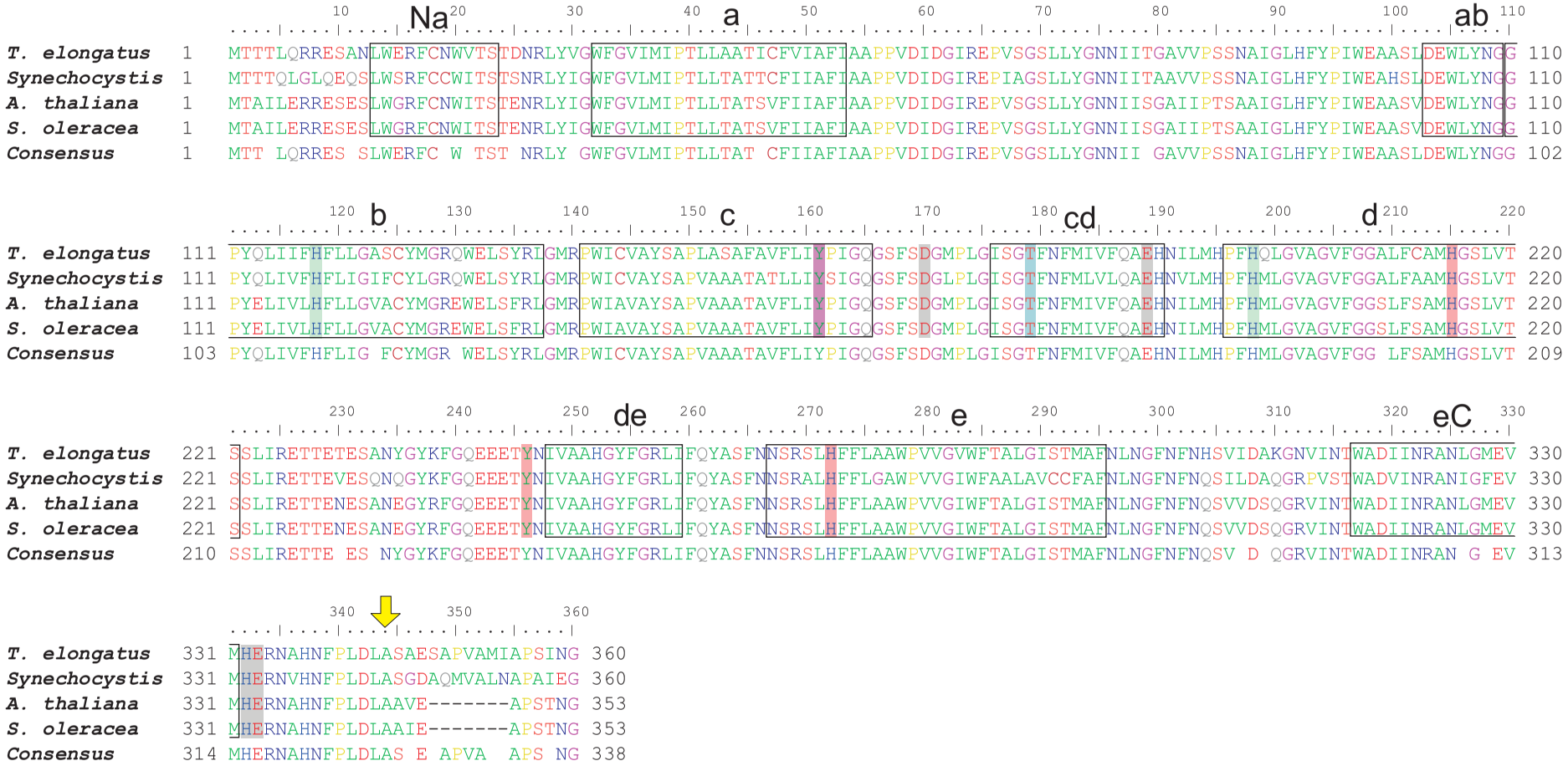


Fig. 4.4: Sequence alignment of D1 proteins. Rectangles define the position of α-helices in *T. elongatus*. The yellow arrow at position 344 indicates the cutting site of the carboxyl-terminal protease. Chl*a* coordinating histidine residues are marked with green background, other indirectly coordinating residues in D1 (Thr169) are indicated by light-blue background. His215, His272 and Tyr246 coordinating the non-haem iron are marked with red background. Residues coordinating the Mn-Ca-cluster are indicated by grey background and Tyr_Z (Tyr161) is marked with violet background. Colouring scheme for amino acids according to chapter 4.2.

A short three amino acids long loop connects TMH-**b** and **-c**. The TMH-**c** contains on the lumenal side the redox-active Tyr_Z (D1-Tyr161). The lumenal loop between TMH-**c** and **-d** harbours a number of functionally important amino acid residues. It encloses the 15 amino acids long α-helix **cd**, also found in PbRC that is parallel to the membrane plane and attached to the membrane surface. The amino acid sequence in this region of D1 differ significantly mainly due to insertions compared to the sequences in PbRC. On the **cd**-helix two important residues are located: D1-Thr179 might be involved in indirect (water mediated) coordination of Chl_{D1} and D1-Glu189 is involved in the coordination of the Mn-Ca-cluster. Another Mn coordinating residue (D1-Glu170) is located between TMH-**c** and **cd**. P_{D1} is coordinated by D1-His198 located on the lumenal side of TMH-**d**, whereas on the cytoplasmic side D1-His215 is found coordinating the non-haem Fe^{2+}.

The D1/D2 interface is formed by a four helix bundle (Fig. 4.6). The overall arrangement of TMH of D1 and D2 resembles the core of PbRC. TMH-**d** and **-e** of D1 as well as TMH-**d** and **-e** of D2 are arranged and constitute a rigid scaffold that incorporates the cofactors of the ETC. There are more interactions on the cytoplasmic side between TMH-**d** of D1 and TMH-**e** of D2 and *vice versa*, whereas on the lumenal side each TMH makes contacts with the nearest neighbour, because they are tilted such toward the membrane plane that they are closer to each other than on the cytoplasmic side. Towards the cytoplasmic side TMH-**d** of D1 and D2 approach each other, in vicinity of the non-haem Fe^{2+} (Fig. 4.5).

Towards the lumenal side, we identified D1-Cys212 and D2-Cys211 being involved in inter-subunit interaction. We identified linking electron density between D2-Cys211SG and the backbone oxygen of D1-Gly208O, however the distance of 3.8 Å indicates only a weak H-bond (Fig. 4.5). The side chain of D2-Met271 located on TMH-**e** points into the binding pocket towards the non-haem Fe^{2+}. D1-Cys212SG might interact with D2-Gly207 located on TMH-**d** (3.8 Å). D1-Gly208O located on TMH-**d** of D1 is interacting with D2-Cys211 located on TMH-**d** of D2 (3.7 Å). D2-His268ε is engaged in the coordination of the non-haem Fe^{2+} and D2-His268δ is stabilised by an H-bond to the backbone oxygen of D2-Lys264 (3.0 Å) and D1-His272 by an H-bond to backbone oxygen of D1-Ser268 (3.3 Å). Additionally, TMH-**d** of D1 and D2 are stabilised by van der Waals interactions especially around the non-haem Fe^{2+} site. Intersubunit H-bonds are between TMH-**d** of D1 and TMH-**c** of D2: D1-Ser217OG to the backbone oxygen of D2-Tyr141 (3.2 Å) and to D2-Asn142 (2.9 Å). The backbone oxygen of the non-haem Fe^{2+} coordinating residue D2-His268 is H-

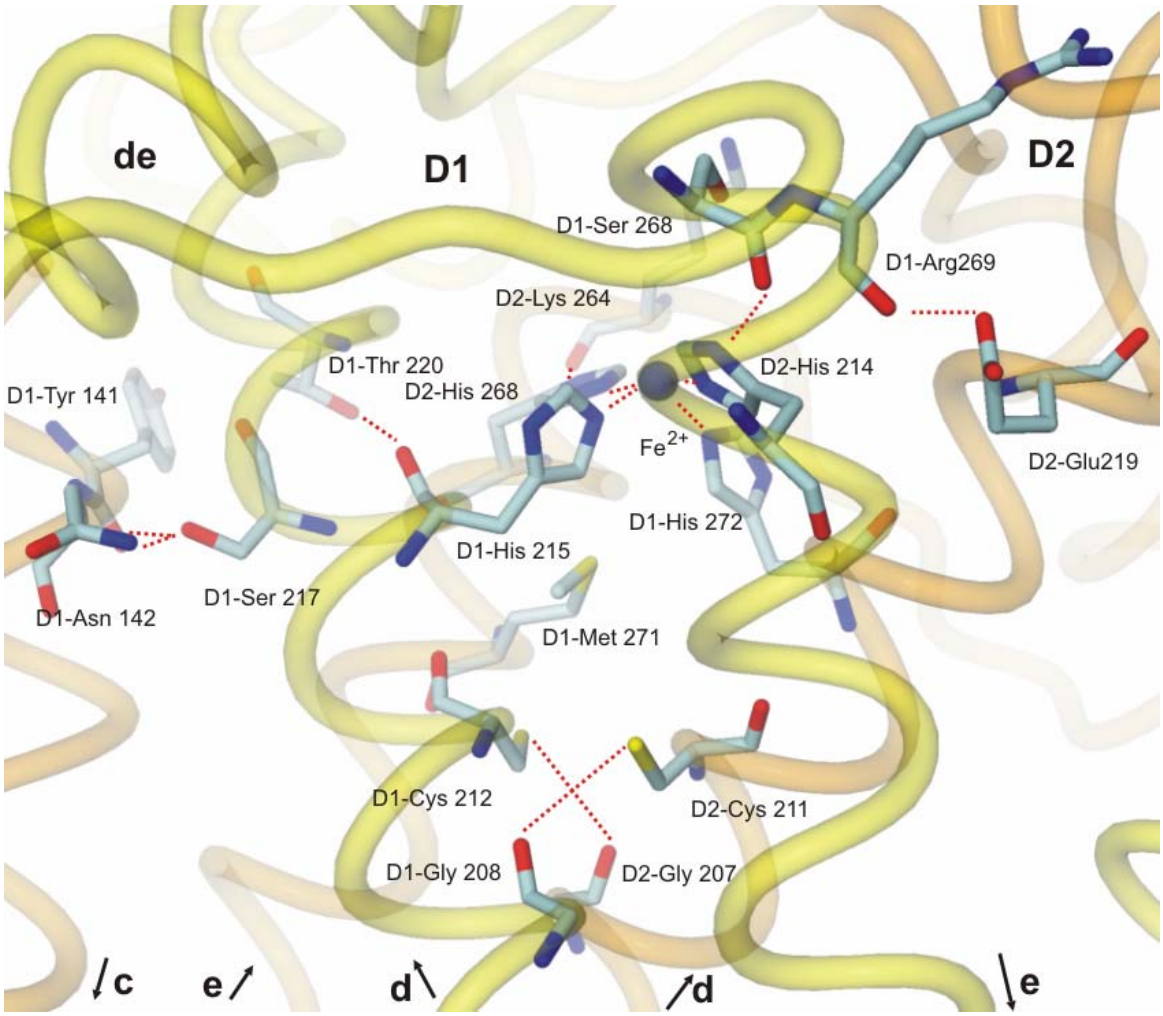


Fig. 4.5: Side view of the D1/D2-interface around the non-haem Fe^{2+} site (blue sphere): D1 is drawn in yellow and D2 in orange. H-bonding network and Fe^{2+} coordination are shown by red dotted lines. Distances of potential H-bonds are given in the text.

bonded to D1-Thr220 (3.3 Å), whereas D2-His214Nε is H-bonded to the backbone oxygen of D1-Ser268 (Fig. 4.5). For functional implications see 6.2.5. Furthermore there is an H-bond D2-Glu219Oε and D1-Arg264O (Fig. 4.5). H-bonding is completed by van der Vaals interactions. Interestingly, all these inter-subunit interactions are close to the non-haem Fe^{2+} site, fixing the geometry of TMH-**d** and -**e** of D1 and D2, respectively. This dense network of H-bonds within TMH is not observed in PbRC (Deisenhofer *et al.*, 1984), where only one intersubunit H-bond was found.

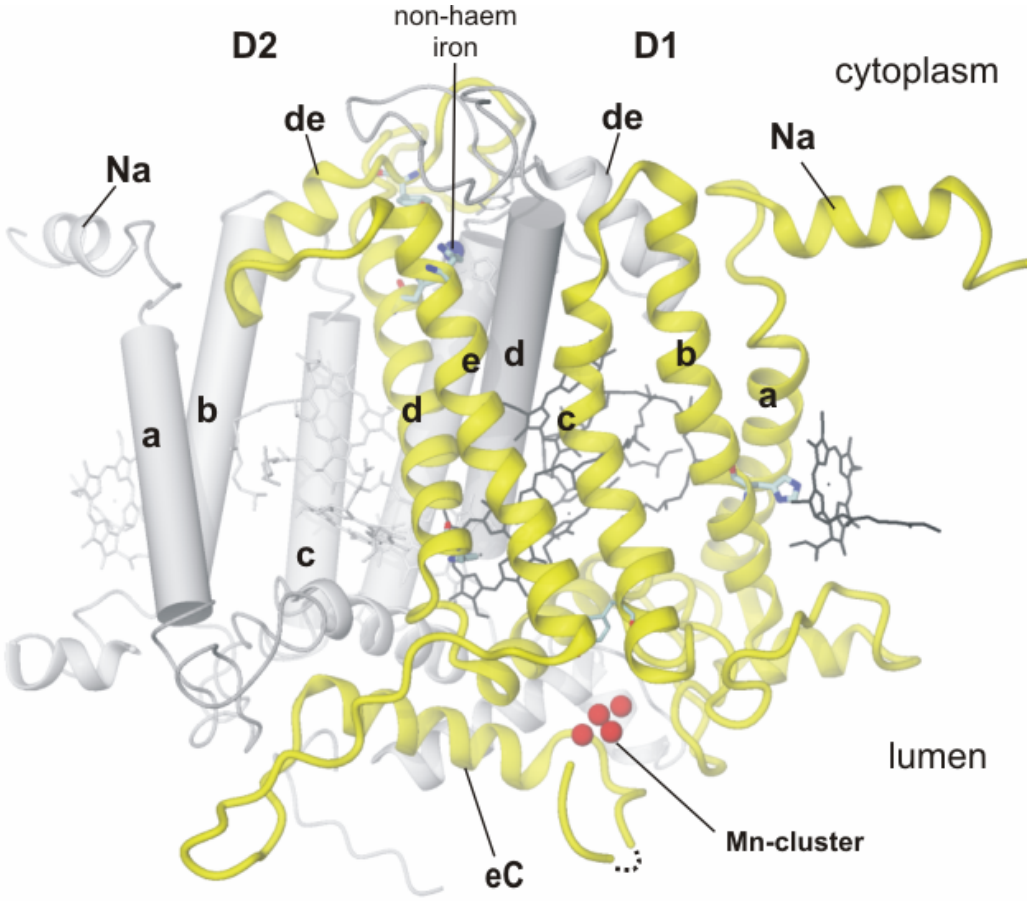


Fig. 4.6: Side view of the core subunits D1 in yellow and D2 in grey. TMH are denoted **a** to **e** The Chl*a* and Pheo*a* of D1 are drawn in black wire models and other coordinating residues as ball and sticks, whereas all cofactor embedded in D2 are shown in grey. The Mn-Ca-cluster (red) and the non-haem Fe^{2+} (blue) are represented by spheres.

TMH-**d** and -**e** are connected by a loop domain partially folded in α-helix **de** similar as in PbRC, but the segment connecting TMH-**d** and **de(1)** of D1 is longer than in PbRC, reflected by a low primary sequence similarity between L and D1. The Q_B site is formed exclusively by D1, by the C-terminal part of TMH-**d** (D1-Gly207 to D1-Val219), the cytoplasmic surface helix **de** and the loop connecting them as well as the N-terminal part of TMH-**e** (D1-Arg269 to D1-Phe274) (Fig. 6.4). In this region, the electron density is weak and consequently a number of residues are modelled as polyalanine. This observation is probably reflecting a higher degree of flexibility, caused by the absence of Q_B that was partially lost during purification and is not fixed in PSII as it has to leave the binding site after reduction to Q_BH_2. A further possibility is a probable degradation in this region.

In contrast to PbRC the C-terminus of D1 is around 100 residues longer. This has functional relevance, as D1-Glu332 and D1-His333 downstream of α-helix **eC** ligate the Mn-Ca-cluster (Fig. 4.3 and Fig. 4.6). The C-terminus is traced with side chains till residue D1-Arg334 followed by a gap and the last residues are traced as Cα.

4.3.2 PsbD - D2 protein

The genome of *T. elongatus* contains two identical copies of the *psbD* gene (*psbD1* and *psbD2*) encoding two proteins. The PsbD (D2) protein of *T. elongatus* shows 30% identity and 59% similarity to the D1 protein. The following cofactors are embedded in this subunit: Tyr_D, P_{D2}, Chl_{D2}, $Pheo_{D2}$, non-haem iron, Q_A and Car_{D2}.

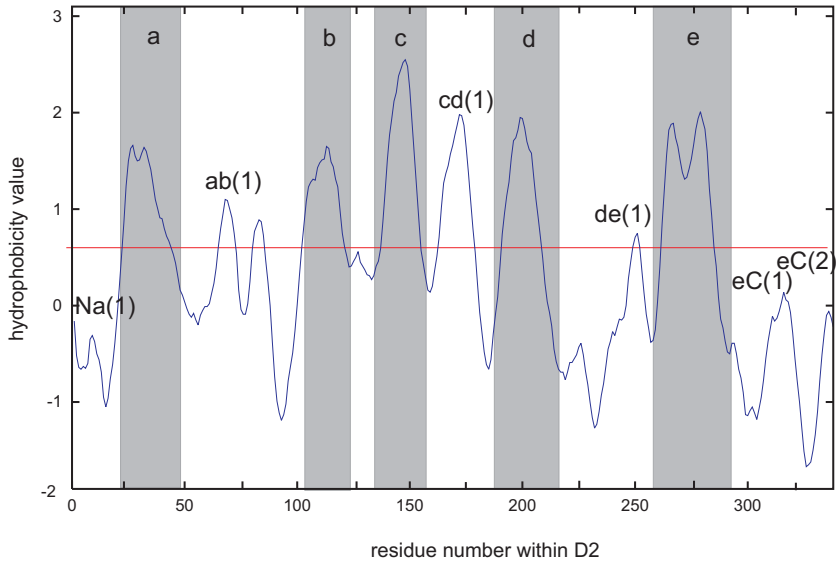


Fig. 4.7: Hydropathy plot of the D2 protein of *T. elongatus* according to (Kyte and Doolittle, 1982) with a window of 10 amino acids. The red line indicates the average hydropathy for all amino acid residues. Hydrophobic patches are marked are above this line. They coincide well with TMH-**a** to -**e** alphabetically ordered according to their occurrence in the sequence.

Like the D1 protein, D2 presents very high amino acid sequence identity (89%) and similarity (98%) to the spinach protein. It has a slightly lower molecular mass of about 39.4 kDa compared to 39.7 kDa for spinach. The identity and similarity within the transmembrane-spanning region is higher, whereas loop regions show insertions and deletions.

Hydropathy plots predict five TMH in D2 (Fig. 4.7 and Fig. 4.8), as in the M subunit of PbRC, and contains the same number of surface α-helices as predicted and shown for the D1 protein (see chapter 4.3.1).

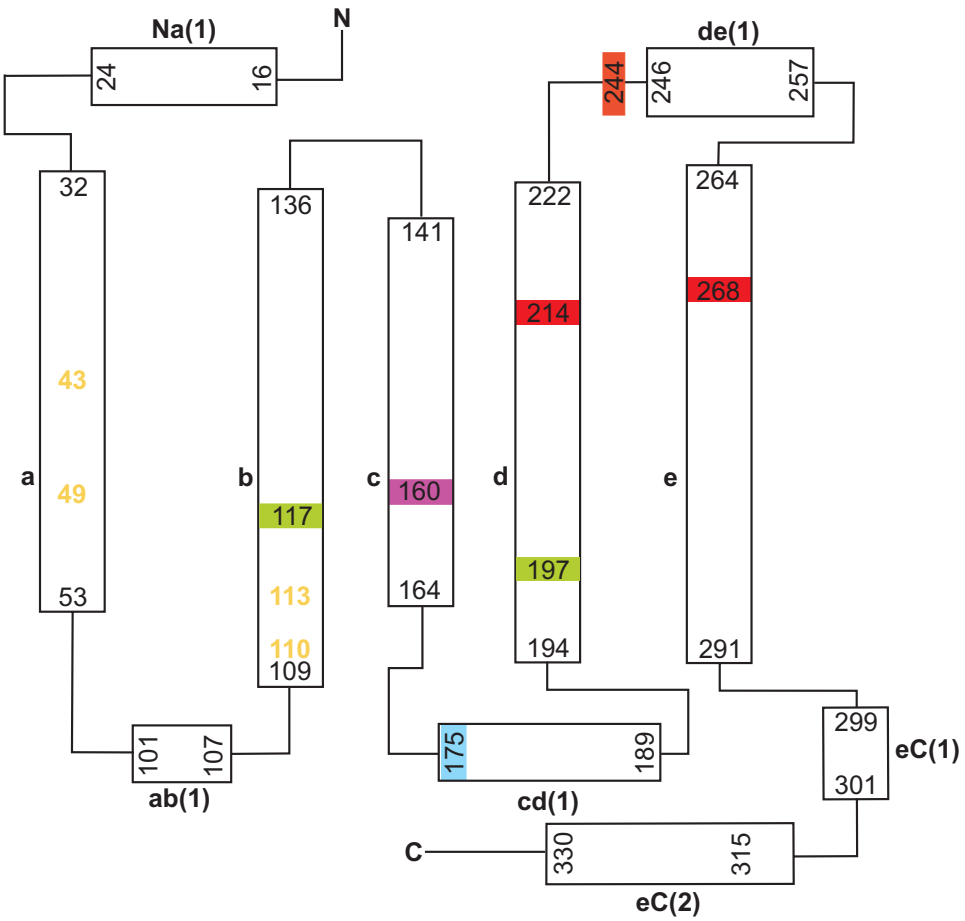


Fig. 4.8: Topography of D2: α-helices are indicated by rectangles and TMH denoted **a** to **e**. The N-terminus (N) is on the cytoplasmic (top) side and the C-terminus (C) on the lumenal side (bottom). Chl*a* coordinating histidine residues are marked with green background; other indirectly coordinating residues (Val175) in D1 are indicated by light-blue background. His214, His268 and Tyr244 coordinating the non-haem iron are marked with red background. Tyr_D (Tyr160) is marked with violet background. Hydrophobic residues in the binding pocket of Car_{D2} are given in bold orange numbers.

The overall fold of D1 and D2 is very similar with an rmsd of 1.5 Å for 295 pairs of Cα-atoms. Not only is the overall fold similar but also the location of cofactor coordinating amino acid residues on TMH, reflecting a possible common ancestor.

The backbone of D2 polypeptide is completely modelled from the N-terminus to C-terminus except for the C-terminal three residues. One carotenoid (Car_{D2}) of the whole PSIIcc was

located in D2 nestling in between TMH-**a** and **-b** and is bound by hydrophobic contacts. Functional implications are discussed in chapter 6.4. The primary plastoquinone (Q_A) in PSIIcc is inserted in a pocket flanked by the C-terminal part of TMH-**d,** N-terminal part of TMH-**e** and surface α-helices **de(1)** (Fig. 6.4). The arrangement of the Q_A binding pocket is similar to that of PbRC.

A notable difference between D1and D2 are two short α-helices in D2 (**eC(1)** and **eC(2)**) close to the C-terminus, whereas in D1 there was only one α-helix **(eC)** identified (Fig. 4.8 and Fig. 4.9).

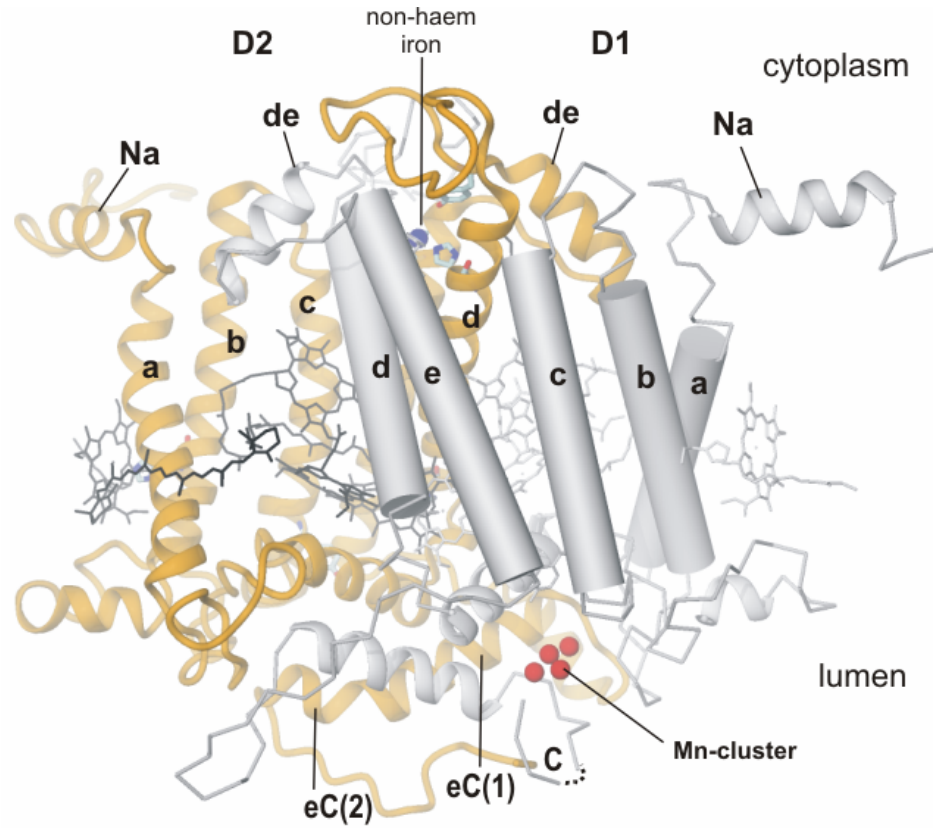


Fig. 4.9: View along the membrane plane. D2 subunit is shown in orange and D1 in grey. TMH are denoted a to e. Chl*a*, Pheo*a* and Car of D2 are drawn in black wire models and other coordinating residues as ball and sticks, whereas all cofactor embedded in D1 are shown in grey. The Mn-Ca-cluster (red) and the non-haem Fe^{2+} (blue) are represented by spheres.

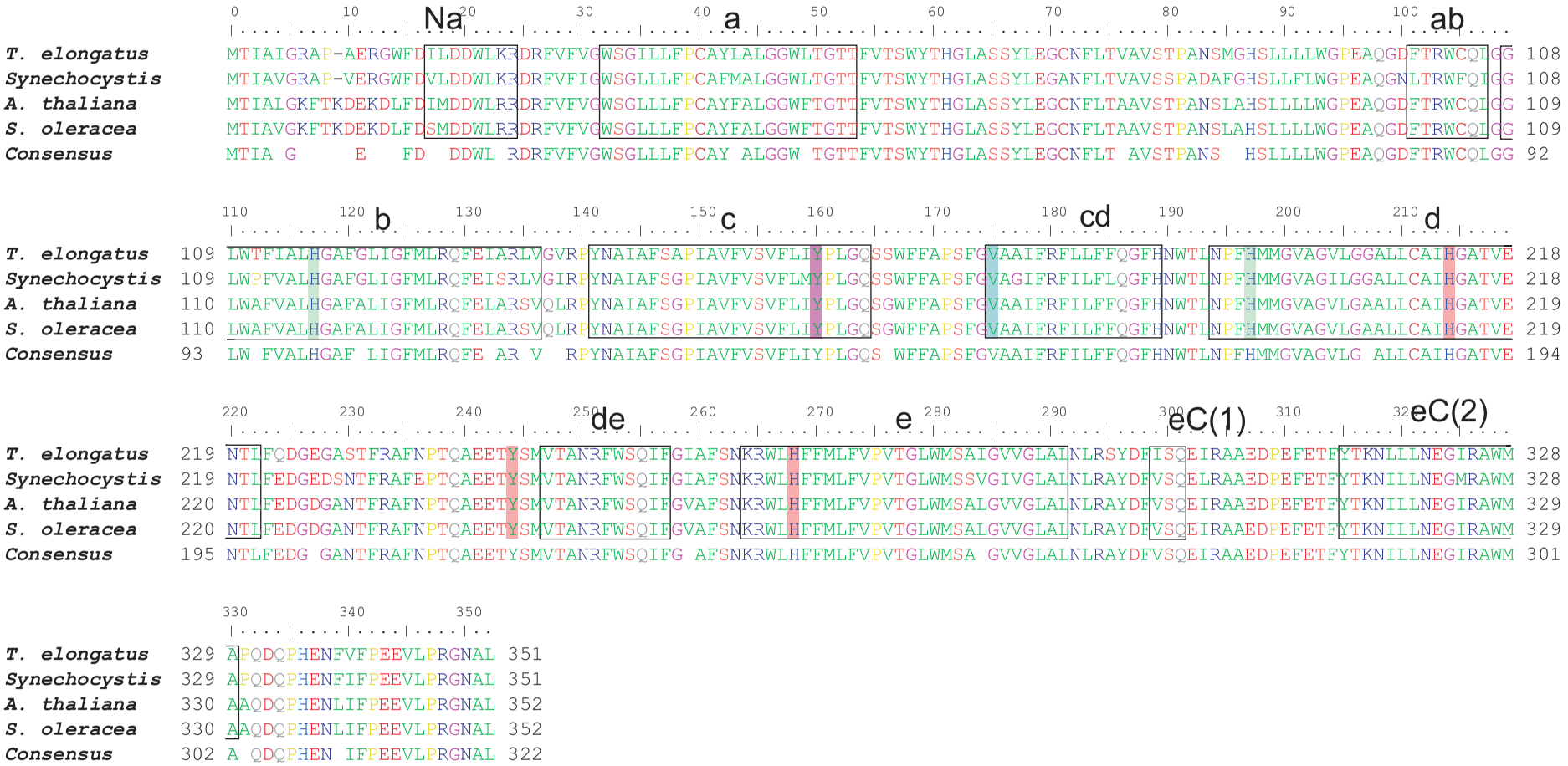


Fig. 4.10: Sequence alignment of D2 proteins. Rectangles define the position of α-helices. Chl*a* coordinating histidine residues are marked with green background, other indirectly coordinating residues in D1 (Val175) are indicated by light-blue background. His214, His268 and Tyr244 coordinating the non-haem iron are marked with red background. Tyr_D (Tyr160) is marked with violet background. Colouring scheme of amino acid according to chapter 4.2.

4.4 The core antenna proteins: CP47 and CP43

PSIIcc carries an integral antenna system composed of Chl*a* and carotenoids. In addition, this system is complemented by extrinsic membrane-attached phycobilisomes (Schirmer *et al.*, 1987; Boichenko *et al.*, 2001; Rakhimberdieva *et al.*, 2001; Zolla *et al.*, 2002). The purpose of both systems is the efficient absorption of visible light.

The inner antenna system is formed by the membrane-instrinsic subunits encoded by the genes *psbB* and *psbC*. For both subunits there are trivial names in the literature. Based on their ability to bind chlorophyll molecules they are named chlorophyll-binding proteins (CP) and the number indicates the molecular weight as estimated from SDS-polyacrylamide gels. In the following the names CP47 (PsbB) and CP43 (PsbC) will be used.

CP47 and CP43 show both high homology to the antennae domains of PsaA and PsaB of PSI (Schubert *et al.*, 1998). CP47 and CP43 are positioned on opposite sides of the heterodimeric PSIIcc-RC (D1/D2) (Fig. 4.1 B). CP43 is located at the periphery of the dimeric core adjacent to D1 whereas CP47 is at the interface region of two PSIIcc monomers and close to D2. It was shown that CP43 is easier to remove from PSIIcc than CP47 (Akabori *et al.*, 1988). Those biochemical data are supported by the structure, as a protein on the periphery should be easier to disassemble. Hydropathy analysis predicts 6 TMH (Bricker, 1990) for each, CP47 and CP43. The N- and C-termini of both proteins are located on the cytoplasmic side of the membrane and extended membrane-extrinsic, hydrophilic loop regions extending far into the lumen is characteristic for CP47 and CP43. In the current structure we could identify in total 26 Chl*a* molecules in CP47 and CP43. These system of antenna Chl*a* are spatially separated from the ETC, and transfer excitation energy to the RC (Deisenhofer *et al.*, 1985). For a more detailed description and discussion of the antenna system see chapter 7.

4.4.1 CP47 (PsbB)

The antenna subunit CP47 consists of 510 amino acids and has a calculated molecular mass of 56.6 kDa. A number of photoactive Chl*a* and *β*-Car are bound by CP47. Biochemical studies suggested that there are either 10-12 Chl*a* (Tang and Satoh, 1984; Barbato *et al.*, 1991) or 20-25 Chl*a* and 2–3 *β*-car (Hankamer *et al.*, 1997) associated with CP47. CP47 is localised at the interface of the two monomers. An important structural function of CP47 is the interaction of

the long membrane-extrinsic domain with the oxygen evolving protein (PsbO). The positive charges of Arg384 and Arg385 (Fig. 4.14) in CP47 are conserved among all PSII from different organisms and have been examined intensively. Replacement of these residues either individually or in tandem leads to a significant loss in oxygen-evolving capability (Putnam-Evans and Bricker, 1992; Putnam-Evans *et al.*, 1996; Qian *et al.*, 1997). Crosslinking studies (Bricker *et al.*, 1988; Enami *et al.*, 1992), antibody binding (Bricker and Frankel, 1987) and site-directed mutagenesis (Haag *et al.*, 1993; Gleiter *et al.*, 1994; Gleiter *et al.*, 1995) also showed that CP47 and PsbO are close together. For a detailed overview about biochemical studies on interaction of CP47 with PsbO see Bricker and Frankel (Bricker and Frankel, 2002).

The hydropathy plot in Fig. 4.11 shows that CP47 consists of 6 TMH. The two C-terminal TMH (TMH-**e** and TMH-**f**) are linked by an extended loop domain of about 180 amino acids (Fig. 4.13 and Fig. 4.18).

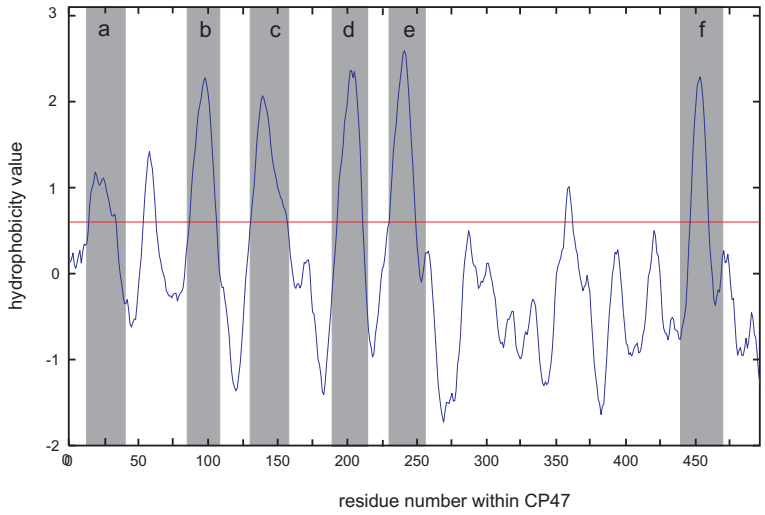


Fig. 4.11: Hydropathy plot of the CP47 protein of *T. elongatus* according to (Kyte and Doolittle, 1982) with a window of 10 amino acids. The red line indicates the average hydropathy for all amino acid residues. Hydrophobic patches are marked are above this line. They coincide well with TMH-**a** to -**f** alphabetically ordered according to their occurrence in the sequence.

Histidine residues were found as ligands of the central Mg^{2+} cation of Chl*a* but alternative amino acids, water or backbone carbonyls could also serve as ligands (Jordan *et al.*, 2001; Kühlbrandt *et al.*, 1994). A total of 12 histidine residues are located in the amino acid sequences of the TMH except for TMH-**e**. We could identify 15 Chl*a* of which 12 are

coordinated by histidyl ligands. All histidine residues have been targeted by mutagenesis. These showed that all 12 histidine residue found here indeed serve as ligands (Eaton-Rye and Vermaas, 1992; Shen *et al.*, 1993; Shen and Vermaas, 1994). So far, we could not identify any Chl*a* in loop regions. This could be due to the fact that the electron density in the loop regions is less defined than in the transmembrane-spanning regions, which makes the identification of Chl*a* more delicate.

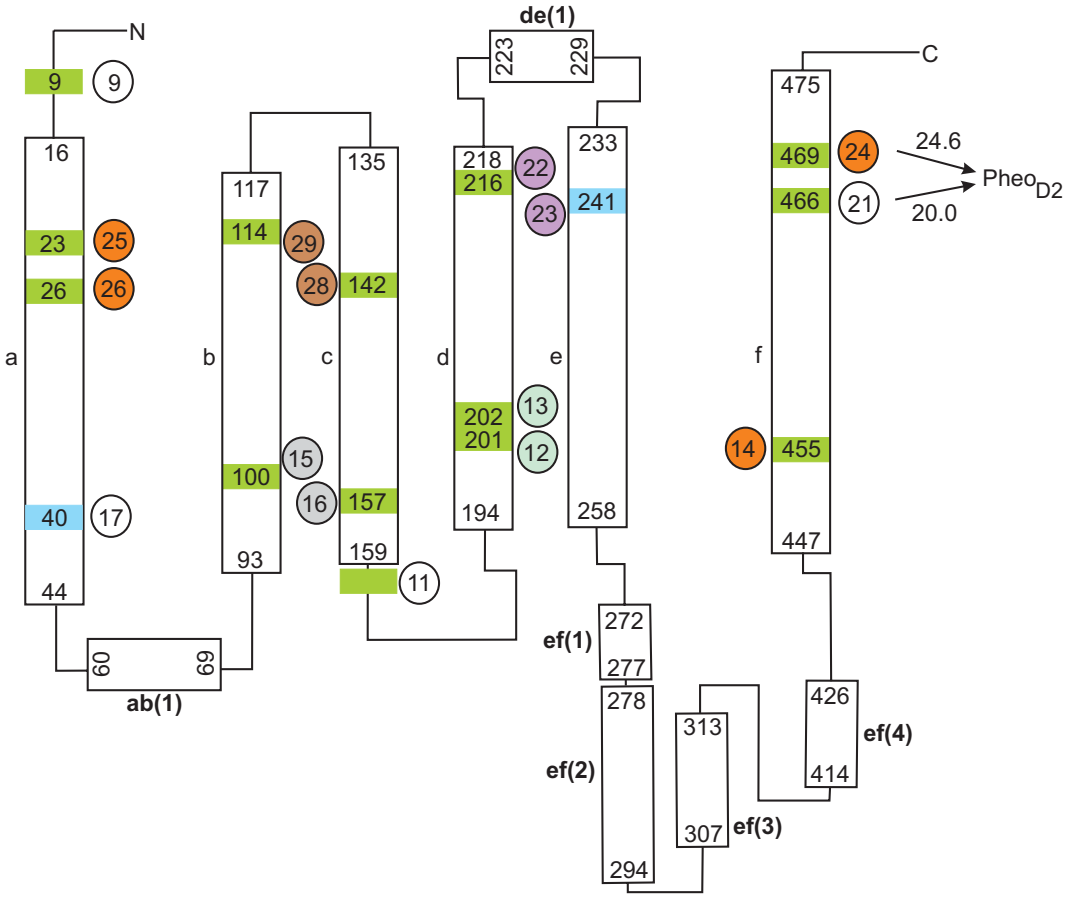


Fig. 4.12: Topography of CP47: TMH are indicated by rectangles and denoted a to f. α-helices are indicated by rectangles. The N-terminus (N) and C-terminus (C) are on the cytoplasmic (top) side. Chl*a* coordinating histidine residues are marked with green background, other indirectly coordinating residues in CP47 (Tyr40, Ser241) are indicated by light-blue background. Groups of Chl*a* in CP47 closer than 10.5 Å are encircled. Circles with the same background colour resemble Chl*a* belonging to one group. Chl*a* on white background do not belong too one group. The number in the centre of circles represents the numbering of Chl*a* according to chapter 4.2.

The field of six TMH-**a** to -**f** in CP47 are arranged as dimer of trimers. The dimers are formed by TMH-**a** and -**b**, TMH-**c** and -**d** as well as TMH-**e** and -**f**. All TMH exhibit an internal pseudo-threefold symmetry interrupted by the membrane-extrinsic domains. In addition CP47

contains two short α-helices (**ab(1)** and **de(1)**) oriented almost perpendicular to the membrane plane on the surface of the membrane as well as four α-helices (**ef(1)** to **ef(4)**) being part of the long membrane-extrinsic domain between TMH-**e** and TMH-**f** (Fig. 4.12 and Fig. 4.13). There are only few interhelical H-bonds, all of them found within the TMH dimers. Two H-bonds are observed between TMH-**a** and -**b**, one between TMH-**c** and -**d** and none between TMH-**e** and TMH-**f**. The few number of H-bonds between adjacent TMH agrees with the findings of that almost 80% of TMH pairs have only one or two interhelical H-bonds (Adamian and Liang, 2002). Interestingly the H-bonds are flanking Chl*a* co-ordinating amino acids and suggest stabilisation of these structural segments, in accordance with a study of H-bonds and Cα—H•••O hydrogen bonds in PSI (Loll *et al.*, 2003).

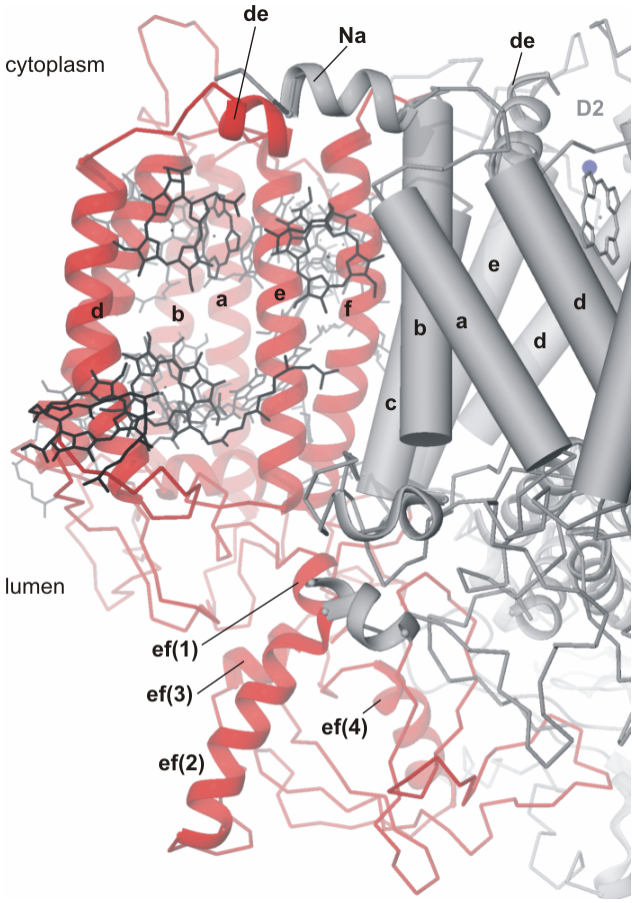


Fig. 4.13: View along the membrane plane. The six TMH of CP47, denoted **a** to **f**, are shown in red. Chl*a* bound to CP47 are drawn as black wire models. All other subunits and cofactors are drawn in grey.

CP47 is close to the local pseudo-*C2* of the PSIIcc dimer at the monomer-monomer interface and interacts with D2. The lumenal and cytoplasmic sides of TMH-**f** of CP47 are in van der Waals interaction with TMH-**e,** and TMH-**c** as well as TMH-**b** of D2.

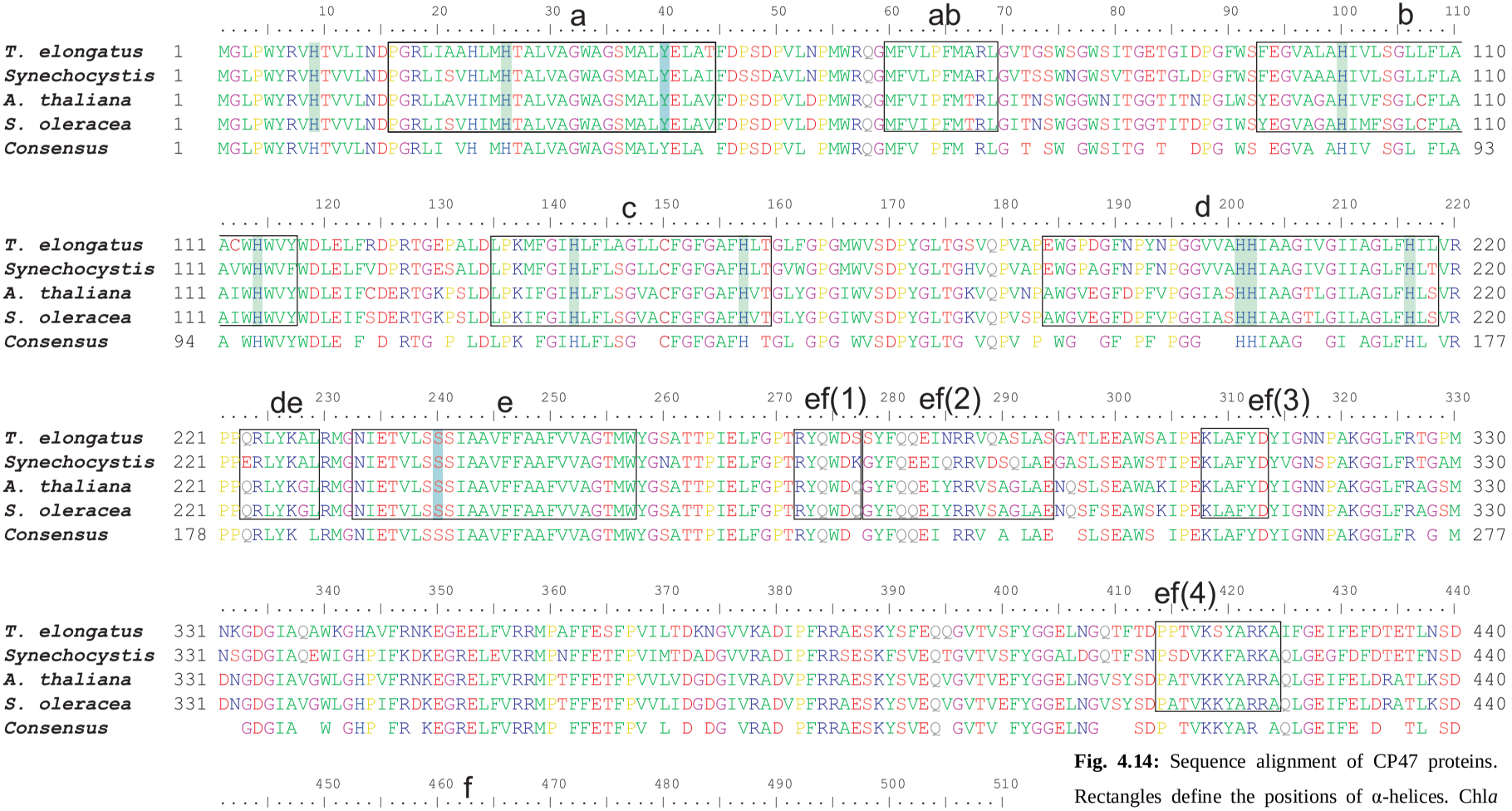


Fig. 4.14: Sequence alignment of CP47 proteins. Rectangles define the positions of α-helices. Chl*a* coordinating histidine residues are marked with green background; other indirectly coordinating residues in CP43 are indicated by light-blue background. Colouring scheme of amino acid according to chapter 4.2.

The extended domain between TMH-**e** and -**f** interacts with the loops of D2 and with PsbO. The connectivity within this domain is not yet clear as various loops of different subunits contact each other. Therefore in lumenal domains, the amino acid sequence is only partially assigned.

4.4.2 CP43 (PsbC)

The antenna subunit CP43 consists of 473 amino acid residues and has a molecular mass of 51.6 kDa. CP43 can be much easier disassembled from PSII than CP47. Additionally it was demonstrated that PSII can assemble in the absence of CP43 albeit with drastically reduced quantities and with restricted function (Vermaas *et al.*, 1988; Rögner *et al.*, 1991). The arrangement of CP43 at the periphery of the PSII dimer would explain this partial assembly of PSII. The well known high turnover rate of D1 (see chapter 4.3.1) connected with the disassembly and re-association of CP43 and repaired D1, would indeed favour a position of CP43 at the periphery of the whole complex (Fig. 4.1). Deletion of CP43 leads to the loss of photoautotrophic growth and oxygen-evolving activity (Rochaix *et al.*, 1989).

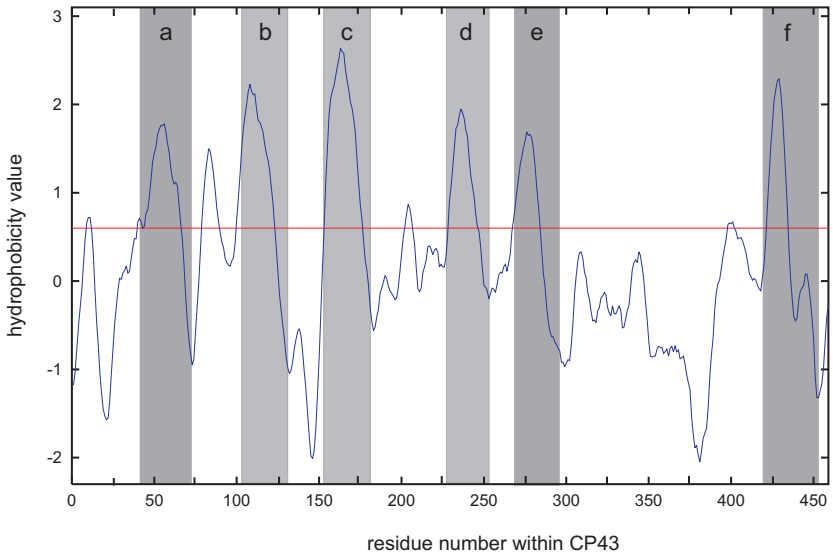


Fig. 4.15: Hydropathy plot of the CP43 protein of *T. elongatus* according to (Kyte and Doolittle, 1982) with a window of 10 amino acids. The red line indicates the average hydropathy for all amino acid residues. Hydrophobic patches are marked are above this line. They coincide well with TMH-**a** to -**f** alphabetically ordered according to their occurrence in the sequence.

Based on biochemical studies it was suggested that CP43 interacts with the oxygen evolving complex. Deletions of single amino acids in region from CP43-Gln298 to CP43-Ala386 led to completely loss of oxygen activity or electron transfer form Tyr_Z to Q_A (Kuhn and Vermaas, 1993). These studies make clear that this region is crucial for the correct assembly and function of PSII. For a detailed discussion of biochemical studies on CP43 see Brickel and Frankel (Bricker and Frankel, 2002). Furthermore, mutations in the extended lumenal domain leading to significant defects suggest that these regions might more directly interact with the Mn-Ca-cluster (Rosenberg *et al.*, 1999).

As for CP47, a large number of cofactors is predicted for this subunit: 10-12 Chl*a* or 20-25 Chl*a* (Tang and Satoh, 1984; Barbato *et al.*, 1991) and 2 – 3 *β*-car (Tang and Satoh, 1984; Barbato *et al.*, 1991).

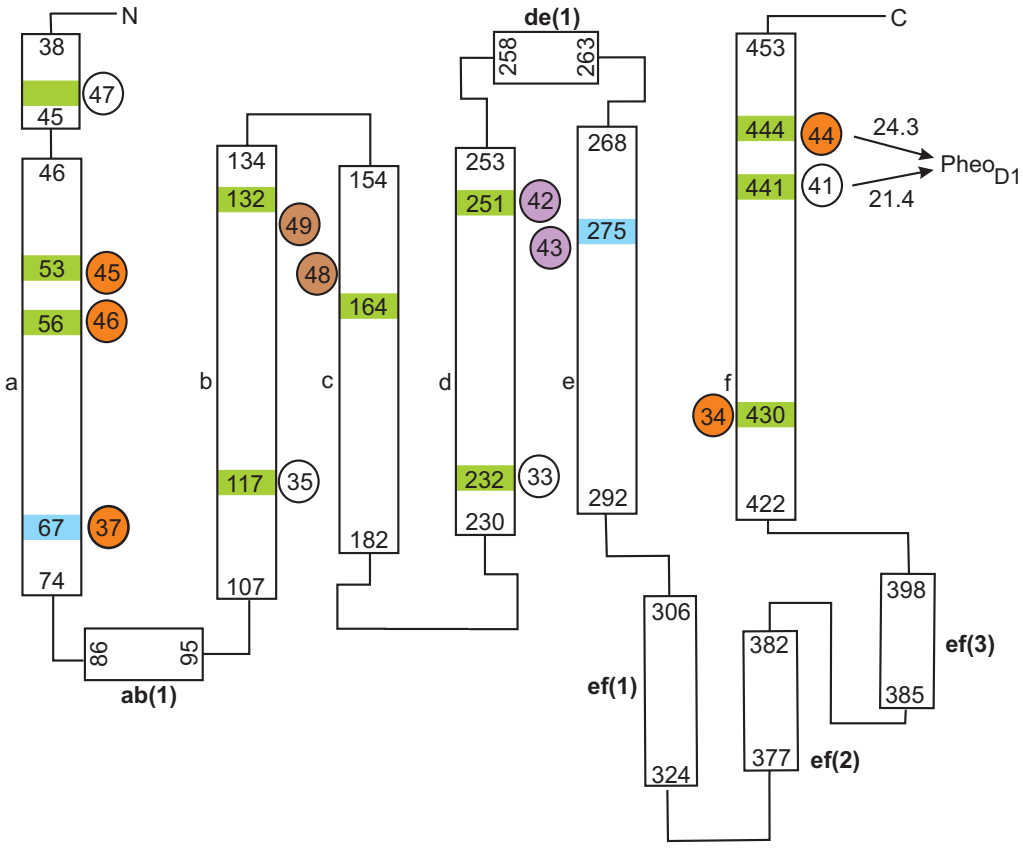


Fig. 4.16: Topography of CP43: α-helices are indicated by rectangles and TMH denoted a to f. The N-terminus (N) and C-terminus (C) are on the cytoplasmic (top) side. Chl*a* coordinating histidine residues are marked with green background, other indirectly coordinating residues in CP43 (Met67, Ser275) are indicated by light-blue background. Groups of Chl*a* in CP47 closer than 10.5 Å are encircled. Circles with the same background colour indicate Chl*a* belonging to one group. Chl*a* on white background do not belong to any group. The number in the centre of circles identifies Chl*a* according to chapter 4.2.

As predicted form hydropathy plots CP43 is composed of 6 TMH (Fig. 4.15). The overall arrangement of CP43 is very similar to CP47 (see chapter 4.4.1).

Besides the 6 TMH, CP43 contains a short N-terminal α-helix (**Na**), one α-helix **ab(1)** on the lumenal side and one α-helix (**de(1)**) on the cytoplasmic side. An extended lumenal domain connecting TMH-**e** and -**f** contains 3 short α-helices (**ef(1)**, **ef(2)** and **ef(3)**).

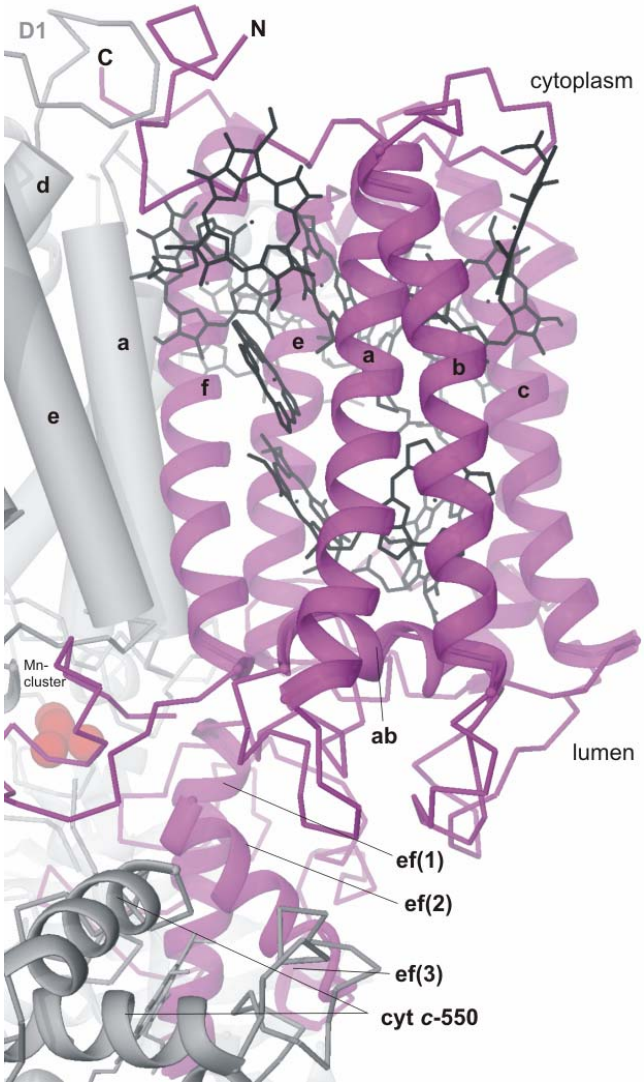


Fig. 4.17: View along the membrane plane. The six TMH of CP43, denoted **a** to **f**, are shown in magenta. Chl*a* bound to CP47 are drawn as black wire models. All other subunits and associated cofactors are drawn in grey.

As observed in CP47, several interhelical H-bonds stabilise the arrangement of TMH. The structure shows one H-bond between TMH-**c** and -**d**, two H-bonds between TMH-**e** and -**f**. The interaction of CP43 with D1 of the RC is similar to the interactions of CP47 and D2.

TMH-**e** of CP43 interacts on the lumenal side with TMH-**e** and -**c** of D1. A weak H-bond might be possible between CP43-Trp291Nε1 and D1-Gly164O (3.3 Å). TMH-**f** interacts on the lumenal side with TMH-**e**, TMH-**c** and -**e** as well as on the cytoplasmic side with TMH-**c** and -**b** of D1.

The domain from CP43-354 to CP43-363 is part of the extended lumenal region and close to the Mn-Ca-cluster. As the amino acid sequence contains mainly residues with long side chains (Fig. 4.18) and the electron density is ill defined, an unambiguous assignment of the sequence was not possible. The importance of this region is supported by biochemical data. An amino acid exchange at position 352 from glutamate to glutamine severely affected PSII activity. This mutant failed to grow photoautotrophically and exhibited essentially no capacity for oxygen evolution. Measurements of total variable fluorescence yield indicated that this mutant did not assemble to functional PS II centres (Rosenberg *et al.*, 1999). Beside a possible direct interaction with the Mn-Ca-cluster, the lumenal domain interacts with loop regions of D1, PsbO and cyt *c*-550.

10 Chl*a* were located that are ligated by histidine residues in the transmembrane-spanning region. All 10 histidine residues within the membrane-spanning region have been mutated (Manna and Vermaas, 1997). However we can not rule out to identify more Chl*a* molecules during further refinement. It should be noted that none of the ligating Chl*a* are neither located on TMH-**e** of CP43 nor on TMH-**e** of CP47 (Fig. 4.18and Fig. 4.16). For a more detailed discussion of the arrangement of the antenna system see chapter 7.2.

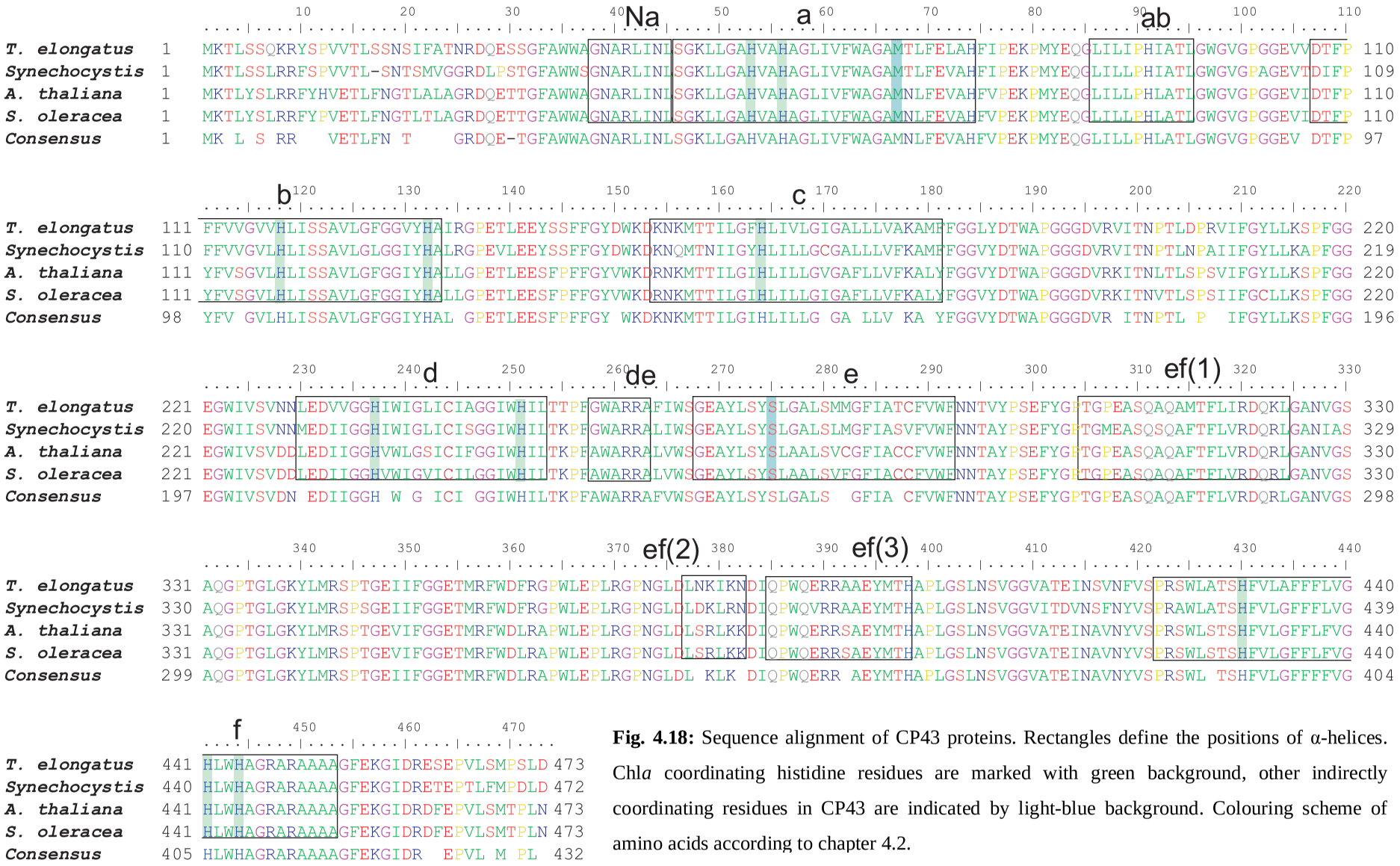


Fig. 4.18: Sequence alignment of CP43 proteins. Rectangles define the positions of α-helices. Chl*a* coordinating histidine residues are marked with green background, other indirectly coordinating residues in CP43 are indicated by light-blue background. Colouring scheme of amino acids according to chapter 4.2.

4.5 PsbE and PsbF - cytochrome b-559

The PsbE and PsbF proteins contain 84 and 45 amino acids (Fig. 4.19) of with a theoretical molecular weight of 9.6 kDa and 5.1 kDa, respectively. MALDI-TOF MS revealed a possible loss of Met1 for both subunits and a possible acetylation of PsbF (Table 3.1 and (Kern *et al.*, 2004b). In the literature the polypeptide chains PsbE and PsbF are referred to as α- and β-chain. Hydropathy plots (not shown) suggest that each subunit forms one single TMH. The dispute whether PSII monomer contains one or two cyt *b*-559 is clear for *T. elongatus* and *T. vulcanus* as biochemical data and both crystal structures located only one cyt *b*-559 (Zouni *et al.*, 2001; Kamiya and Shen, 2003). The *psbE* and *psbF* genes occur in all known genomes as one group (together with the genes for the small subunits *psbL* and *psbJ*) leading to the conclusion that during evolution the gene was probably duplicated. The primary sequences of PsbE and PsbF show an identity of 15% and similarity of 31% (Fig. 4.19), whereas the similarity to PsbL and PsbJ is low.

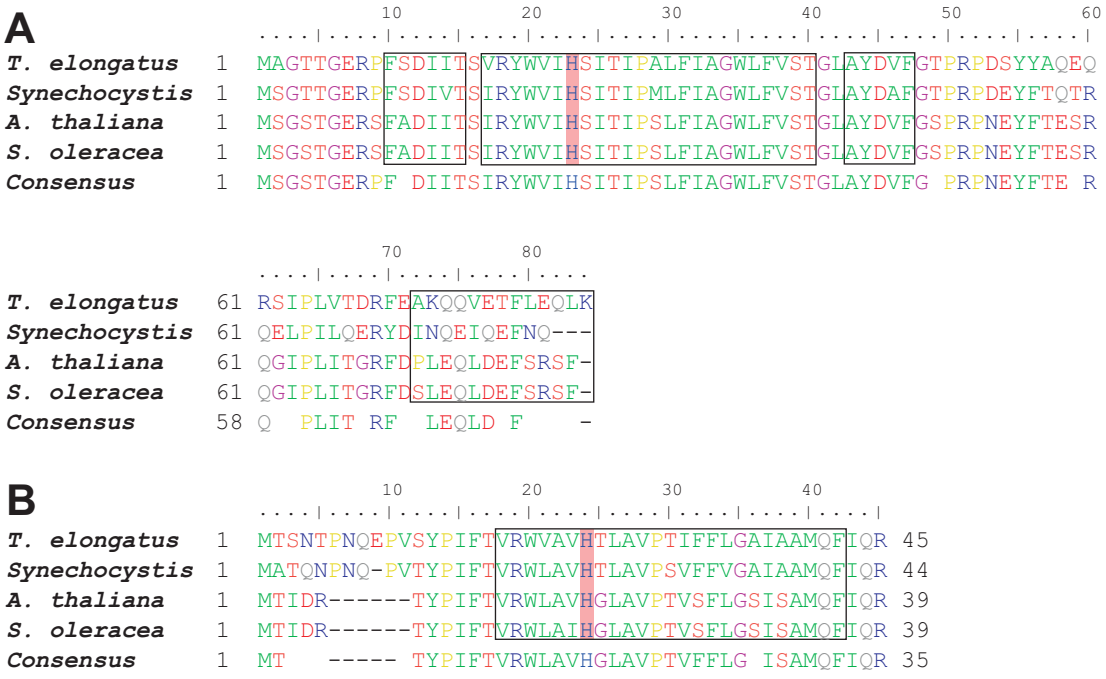


Fig. 4.19: Alignment of (**A**) PsbE and (**B**) PsbF. Rectangles define the positions of α-helices. Haem coordinating histidine residues are marked with red background (α-His23, β-His24). Colouring scheme of amino acids according to chapter 4.2.

A characteristic property of cyt *b*-559 is the occurrence of two forms with high and low midpoint potentials (+390 mV and +275 mV, respectively at pH 6.5). Compared with other *b*-

type cyt the high midpoint redox potential of cyt *b*-559 is unusual. This is associated with the unique coordination of the haem group in a heterodimeric, hydrophobic protein matrix and with H-bonds formed by both of the coordinating imidazoles to peptide oxygens (Fig. 4.21B). In the low potential form one of the H-bonds is possibly broken (Berthomieu *et al.*, 1992; Roncel *et al.*, 2003). The function of high potential cyt *b*-559 in PSII is still under debate. It was proposed to be involved in proton translocation and secondary electron transfer processes that may play a protective role under light-stress conditions and, in the absence of a functional Mn-Ca-cluster, it could transfer electrons to $P680^{+\bullet}$ (Stewart and Brudvig, 1998).

Cyt *b*-559 is located at the periphery of PSIIcc and close to TMH-**a** of D1 and unassigned low molecular weight subunit TMH **6** in a hydrophobic environment (Fig. 4.21A and Fig. 4.22). On the cytoplasmic side, the first seven N-terminal residues of PsbE are not defined in the electron density. In front of the TMH starting at PsbE-Val17, a short amphiphilic α-helix is located parallel to the membrane plane (Fig. 4.20) whose C-terminus is close to the N-terminus of unassigned TMH **6**. The Fe^{2+} coordinating α-His23 is not in the middle of the membrane depth, but closer to the cytoplasmic side. On the lumenal side of the α-chain, two short α-helices are found parallel to the membrane plane. Both short α-helices might help to anchor the TMH of the α-chain in the membrane. PsbE has a 44 residue long lumenal extension that is fully traced in our model (Fig. 4.20).

The C-terminus could be barely traced as α-helix and was modelled as polyalanine. The loop connecting the C-terminal α-helix interacts with the **a-ab** loop of D1 and is very weakly defined in the electron density and therefore modelled as Cα-trace. α-Gln58 in this loop is in H-bonding distance to PsbV-Glu2 and PsbV-Thr4.

Except for the first eight N-terminal residues, PsbF is completely traced. On the cytoplasmic side of the TMH, PsbF interacts with the unassigned TMH 7. β-His24 is the 6^{th} ligand of the central Fe^{2+} of the haem group, located in the same membrane-depth like its counterpart α-His23. The C-terminal part of the TMH of PsbF is in close to unassigned TMH **6**. A notable difference is the shorter C-terminal loop of PsbF in comparison to the long and structured C-terminus of PsbE.

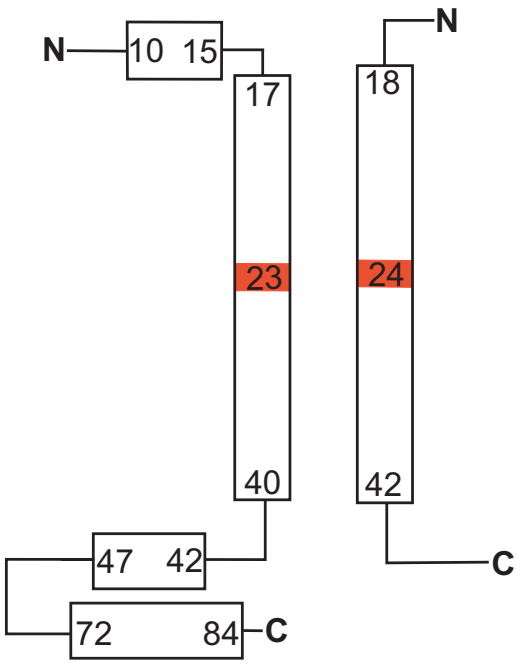


Fig. 4.20: Topography of cyt *b*-449 composed of the α-chain (PsbE) and the β-chain (PsbF): α-helices are indicated by rectangles and haem coordinating histidine residues are marked with red background. N-termini are on the cytoplasmic side and C-termini on the lumenal side.

The haem is located near the cytoplasmic side of PSIIcc tilted by 86° to the membrane plane (Fig. 4.21). α-His23 and β-His24 coordinate the central Fe^{2+} with Nε (Table 10.1), and their NδH groups form H-bonds to peptide oxygen α-Tyr19O and β-Trp20O at 3.0 and 2.8 Å, respectively, in agreement with FT-IR spectroscopic studies (Berthomieu *et al.*, 1992). Since the electron density is bulky for the coordinating histidine side chains, we can not discriminate their χ2 angles. H-bonding of both histidines to the TMH backbone should be energetically favourable. Most probable, we have a mixture of the high and low potential form in our crystals. Additionally, α-His23 imidazol π-stacks with the aromatic ring of α-Tyr19 and β-His24 imidazole π-stacks with the indol group of β-Trp20. The propionate groups of the haem are not defined in the electron density (Fig. 4.21). There are van der Waals interactions between TMH-**a** of D1 (Leu49, Thr50, Thr53) and the TMH of the *β*-chain of cyt *b*-559 (Phe33, Ile37). Car_{D2} is nestled between these two TMH and van der Waals interactions are due to D2Gly46, D2-Leu49, D2-Tyr50, D2-Phe113, β-Phe33, β-Pro29, β-Thr30, β-Leu34 and β-Ile37. The lumenal end of TMH-**a** of D2 is in loose contact with the C-terminal loop region of the α-chain of cyt *b*-559.

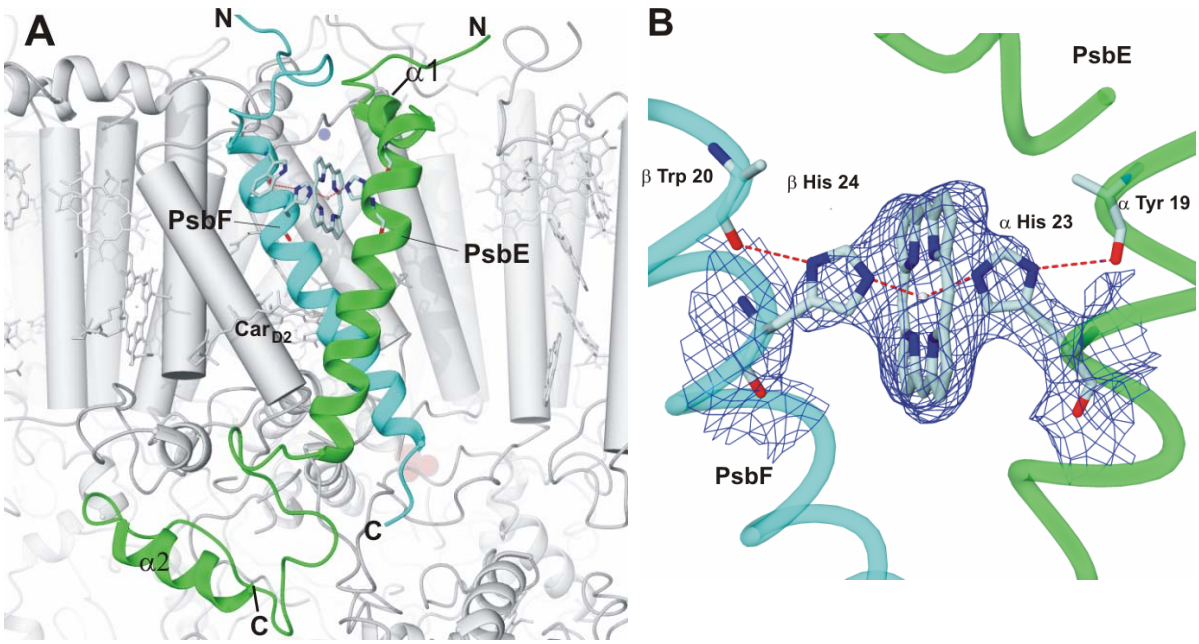


Fig. 4.21: View along the membrane plane; cyt *b*-559 α-chain in green and β-chain in cyan. The haem and coordinating histidines (α-His23 and β-His24) are shown. (**A**) All other subunits and cofactors are shown in grey. (**B**) Electron density (blue) around the haem is contoured at 1.0 σ level, H-bonds formed between imidazoles and peptide oxygens as red dotted lines.

Since the exact role of cyt *b*-559 is still not clear, several functions based on its redox and acid-base properties were proposed: involvement in proton translocation processes and a cyclic-electron transfer that plays a protective role (Stewart and Brudvig, 1998) under light-stress conditions (Ortega *et al.*, 1999). The high potential form of cyt *b*-559 is suggested to protect against donor-side induced photoinhibition by electron transfer to $P680^+$ in the absence of a functional Mn-Ca-cluster (Thompson and Brudvig, 1988), while the low potential form might prevent acceptor-side induced photoinhibition by oxidation of $Pheo^-$ (Nedbal *et al.*, 1992).

A model was suggested where cyt *b*-559 serves as a molecular switch between different redox forms to regulate donor and acceptor side photoinhibition (Whitmarsh and Pakrasi, 1996). The exact molecular mechanism involved in the conversion from high-potential to the low-potential forms is still under debate.

4.6 Structural summary of low molecular weight subunits

The function of the small subunits is not yet elucidated. None of them binds any cofactor, contrasting PSI, where the small subunits are involved in cofactor binding or are in contact with cofactors and their functions are well understood (Jordan, 2001; Jordan *et al.*, 2001). Nevertheless some of the small cofactors that are located in the dimerisation domain of the PSIIcc monomers could be crucial for the correct assembly of the PSIIcc dimer. The open question remains, why PSII needs so many small subunits? Some of them may contribute to the stability of the whole complex whereas some could support the correct folding and/or the incorporation of cofactors and it can not be excluded, that some might play a role in energy transfer. In all current models (Zouni *et al.*, 2001; Kamiya and Shen, 2003; Ferreira *et al.*, 2004) of PSIIcc complexes, there is only one copy of each subunit (Fig. 4.22).

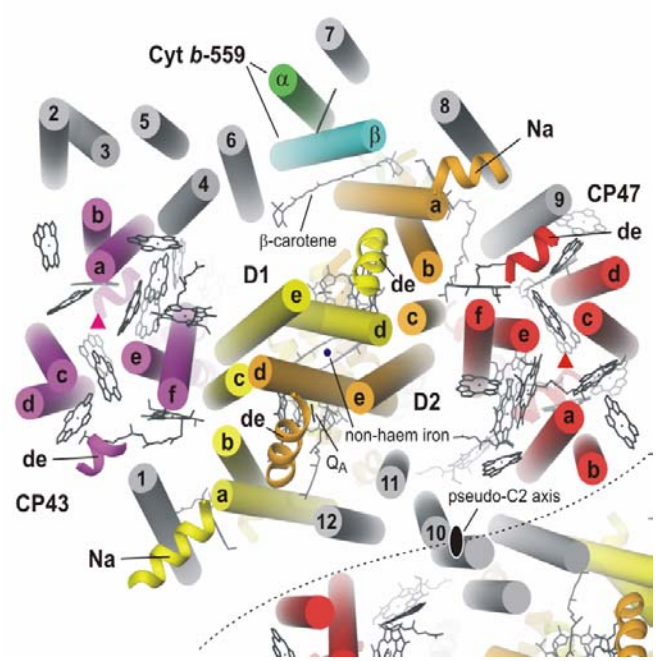


Fig. 4.22: View on the membrane plane from the cytoplasm, membrane-extrinsic subunits and non-helical segments are not shown. One monomer drawn completely, the other monomer in the dimer (related by the pseudo-*C2* axis, black ellipse) only partially. Assigned subunits in magenta (CP43), yellow (D1), orange (D2), red (CP47), green and cyan (cyt *b*-559, α and β). Unassigned TMH (grey) labelled 1 to 12. Local pseudo-*C2*(Fe^{2+}) axis passes through the non-haem Fe^{2+} site (blue dot). Triangles indicate positions of pseudo-threefold axes within CP47 and CP43.

Many investigations have been undertaken to assign the small subunits, based on chemical cross-linking studies, immunological analyses and some "guesswork". A more detailed look on the solvent exposed loop regions of PSIIcc reveals that crosslinks within almost every

subunit are possible, as the loops of various subunits interact with each other (Fig. 5.1B). A successful crosslink depends mainly on the length of the crosslinking reagent. Therefore all crosslinking studies have to be treated with caution. Some of them are even contradictory. This might be the reason for the different assignment of the small subunits in the different models (Zouni *et al.*, 2001; Kamiya and Shen, 2003; Ferreira *et al.*, 2004). These uncertainties are not surprising as all structures are determined to low or to medium resolution. MALDI-TOF MS on our protein samples and re-dissolved crystals showed that we have at least 19 subunits in our structure (Table 3.1), 10 of them belonging to the low molecular weight subunits (Kern *et al.*, 2004b). We could not detect PsbH in our samples, which is in contrast to the proposed assignment of PsbH in the 3.5 Å structure (Ferreira *et al.*, 2004).

PSII contains several small subunits in the range of 10 to 3.2 kDa. A common structural feature of all small subunits is the occurrence of one TMH, with merely short extensions on the cytoplasmic or lumenal side of the thylakoid membrane (Fig. 4.23). The prediction of transmembrane-spanning regions was performed with the program PHDhtm implemented in the program package "predict protein" (Rost *et al.*, 1996).

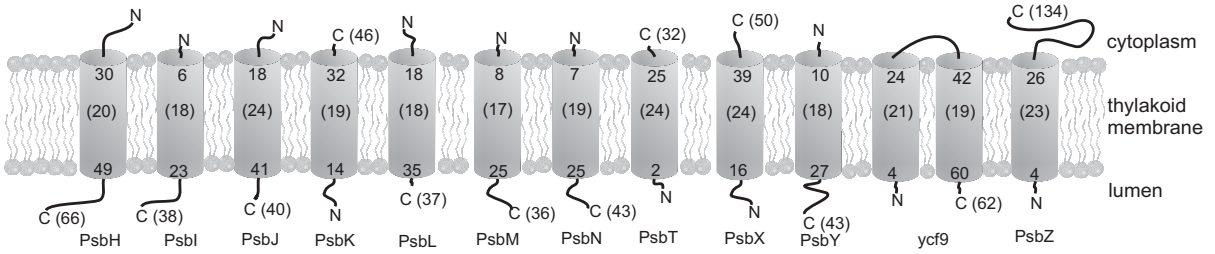


Fig. 4.23: Schematic drawing of 11 possible low molecular weight subunits embedded in the thylakoid membrane of *T. elongatus*. Grey cylinders represent TMH. The start and end of TMH is given on both side of the cylinder. The numbers in parentheses correspond to the number of amino acids in the membrane spanning region. At the C-termini, the total length of the polypeptide chains is given in parentheses.

For the ycf9 polypeptide two TMH were predicted by all algorithms. ycf9 is present in our protein preparation (Kern *et al.*, 2004b), whereas it is not clear for the PsbZ-like protein. There is some confusion in the literature concerning the designation of PsbZ. The *ycf9* gene is conserved from cyanobacteria to higher plants. However there exists a *psbZ*-like gene in cyanobacteria which reveals some sequence similarities. It might be that PsbZ consists of one TMH and one shorter α-helix perpendicular to it, lying on the surface of the membrane. An increased value of hydropathy was observed for membrane attached α-helices of D1 and D2,

hence higher values in the case of PsbZ could indicate such kind of α-helix. This hypothesis is supported by careful inspection of the primary sequence of PsbZ, as typical amino acid residues of a TMH are absent. Every fourth to fifth residue is polar, corresponding to one helix turn, which would be solvent exposed.

Fig. 4.23 shows that all the low molecular weight subunits except for ycf9 contain all one TMH, have similar length and do not contain any characteristic structural feature. The cytoplasmic and lumenal extensions of the small subunits lack any secondary structure element and consequently they are more disordered which makes the low molecular weight subunits even more similar. Discrimination and assignment of the low molecular weight subunits based on the primary structure is difficult and even impossible with the electron density maps at 3.2 Å resolution, as the transmembrane-spanning regions are mainly composed of small hydrophobic residues and typical landmarks are rare. For typical features of TMH architecture and characteristics see chapter 3.12.

Table 4.1: Summary of the unassigned TMH. The numbering of TMH refers to Fig. 4.22. The total length, the length as well as the orientation, from the N- to the C-terminus, of each TMH are given.

TMH	**total number of residues**	**length of TMH**	**orientation (N → C) terminus**
1	35	24	L[a] → C[a]
2	30	30	C → L
3	28	28	L → C
4	38	26	L → C
5	20	20	L → C
6	27	27	C → L
7	31	22	C → L
8	41	35	L → C
9	25	22	L → C
10	28	27	C → L
11	23	24	C → L
12	27	22	L → C

[a] L correspond to lumenal and C to cytoplasmic side.

Due to all mentioned difficulties related to low molecular weight subunits, we did not assign any of them in our current structure. Most probably unassigned TMH **2** and **3** are formed by the ycf9 protein. All other TMH are merely modelled as polyalanine and therefore not more

than the direction of the helix dipole is indicated. The unassigned TMH can be divided in three groups (Fig. 4.22): TMH **1** and **2** stabilise the peripheral $Chlz_{D1}$ and $Chlz_{D2}$ and are approximately symmetrically related. Flanked by CP43 and cyt *b*-559, TMH **2** to **6** are located at the periphery of PSIIcc. The remaining TMH **10** to **12** are involved in dimer formation. The closest contacts of the dimer-dimer interface are between TMH **10** of both monomers. Table 4.1 summarises the total number of amino acid residues and the length of each single TMH.