

3 Structure determination

3.1 Spectroscopic and biochemical characterisation of PSIIcc

Each PSIIcc preparation was extensively characterised by spectroscopic and biochemical methods (Kern *et al.*, 2004a). Assuming the presence of 4 Mn per Mn-Ca-cluster and 36 Chl*a* per P680, the determined manganese content of 3.7 and 3.8 Mn/36 Chl*a* in monomeric and dimeric PSIIcc indicates that 93-95% of the centres contain a fully assembled Mn-Ca-cluster. Oxygen evolution measurements revealed that PSIIcc in solution as well as re-dissolved crystals are fully active in light-induced electron transfer, water splitting and oxygen release (Zouni *et al.*, 2000). From the UV-Vis spectra, the carotenoid content for dimeric PSIIcc was determined to be 9.1 ± 1 Car per 34 Chl*a*. Chromatography showed that about 2.3 plastoquinones are present per PSIIcc. The detergent shell is composed of about 240 β -DM and lipid molecules (from the thylakoid membrane) per dimeric PSIIcc.

The integrity of the whole complex was investigated by SDS-PAGE and MALDI-TOF MS (Kern *et al.*, 2004b). It was shown that the subunits PsbA, B, C, D, E, F, H, I, J, K, L, M, O, T, U, V, X, Y and PsbZ (*ycf9*) are present in monomeric, dimeric PSIIcc as well as in re-dissolved crystals (Kern *et al.*, 2004b). In addition, this study revealed the presence of possible posttranslational modifications. Although interpretation of MALDI-TOF spectra especially for some of the low molecular weight subunits was difficult and had to be supported in some cases by information gained by additional methods like N-terminal sequencing.

Table 3.1: Subunit composition their possible posttranslational modification of PSIIcc from *T. elongatus* as determined by MALDI-TOF MS (Kern *et al.*, 2004b). Calculated masses based on DNA-sequence (Nakamura *et al.*, 2002).

subunit	calculated masses [kDa]	experimental masses [kDa] ($M+H$) ⁺ \pm σ	deduced posttranslational modifications from MALDI-TOF MS
PsbA1	39737.5	38144 \pm 224	cleaved Met1
PsbA2	39738.8		
PsbA3	39757.6		
PsbB	56603.7	56485 \pm 108	
PsbC	51621.4	51704 \pm 87	

Structure determination

PsbD	39362.4	39290 ± 79	cleaved Met1
PsbE	9573.9	9446 ± 6	cleaved Met1
PsbF	5065.95	4938±4	cleaved Met1 and acetylation
PsbH	7355	7227±5	
PsbI	4406.3	4437±4	cleaved Met1
PsbJ	4105.9	4017±2	cleaved Met1 and acetylation
PsbK	5027.1	4103±4	N-terminal processed
PsbL	4298.1	4301±4	
PsbM	3980.7	4025±2	
PsbO	29607.6	26830±30	N-terminal processed
PsbT	3875.8	3906±4	
PsbU	15018	11649±8	N-terminal processed
PsbV	18028	15752±11	N-terminal processed and attached haem
PsbX	5233.4	4192±4	N-terminal processed
PsbY	4773.9	4617±4	cleaved Met1 and Gly2
PsbZ	6765.3	6798±5	

Concluding, it was shown that the protein material is highly homogeneous with similar subunit and pigment composition observed from different protein preparations, suggesting that the protein material is suitable for crystallisation experiments. The sensitive molecular apparatus of PSII_{cc} has not changed its composition during isolation, purification and crystallisation steps.

3.2 Crystallisation of PSII_{cc}

Crystallisation was performed as described in the chapter 2.4.1. Besides the crystallisation in the batch method, trials using the hanging drop method were set up to allow gentler harvesting of the crystals. Initial trials were not satisfactory as the crystals were inter-grown and with a size of around 30 to 50 µm too small for diffraction experiments. Therefore, the small crystals were used as seed for micro- and macro-seeding experiments.

Another attempt was the addition of small amphiphilic molecules to make the detergent micelle smaller (Michel, 1983). This approach was successful in the case of PbRC and light-harvesting complexes. A number of small amphiphilic molecules and the additive screen of Hampton Research were used. Crystals were growing under many different conditions by varying the pH, buffer system and using different PEG. X-ray diffraction experiments of these

crystals revealed that they had the same macroscopic properties, as crystals grown from the standard condition.

In an endeavour to grow crystals in a different space group, the precipitant was exchanged. Crystals were obtained under conditions using PEG-400 (15-18%), PEG-3350, PEG-4000 and PEG-6000. Again, initial crystals were checked for their unit cell parameters and space group. As all of them belonged to same orthorhombic space group ($P2_12_12_1$), no fine tuning of the crystallisation conditions was performed. The pH seems to have a minor influence on the crystallisation of PSIIcc, as crystals could be obtained under various buffer systems over a pH-range from 5.6 (citrate buffer) to 8.5 (Tris-HCl). All crystallisation experiments showed that PSIIcc seems to be sensitive to the final PEG concentration in terms of crystal size, whereas it is possible to obtain crystals with low and high molecular PEG as precipitant. Unfortunately, up to now no other precipitant was successfully employed to crystallise PSIIcc.

3.3 Macroscopic crystal properties

Crystals of dimeric PSIIcc from *T. elongatus* grow as dark green crystals in an elongated hexagonal shape. However, only few crystals had "perfect" shape and size. Nevertheless also smaller crystals and less perfectly grown crystal e.g. with a missing tip often showed similar diffraction limits as the perfect looking ones. Additionally, many of these crystals seemed to be rather bent, which could be due to their growth in quartz capillaries. These macroscopic properties have rather minor influence if at all on the diffraction quality.

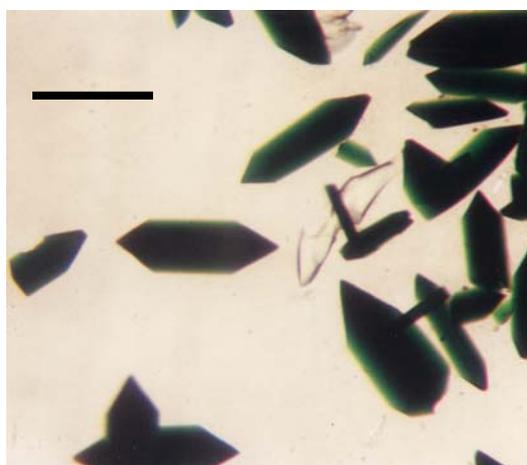


Fig. 3.1: Dark green crystals of dimeric PSIIcc. The length of the bar is equal to 500 μm .

Structure determination

If the crystals were observed under back light, the surface reflected light and on the surface intergrowth was visible. The dark green colour of the crystals alleviate the handling e.g. the mounting of the crystals, but on the other hand, crystal defects are difficult to detect. Even though some crystals had a maximum length of 1 mm, they showed higher mosaicity in the X-ray beam.

3.4 Cooling of PSIIcc crystals

Crystals were harvested in cryosolution A containing the same concentration as the mother liquor for crystallisation with an increased concentration of 10% PEG-2000. In four steps, the solution was exchanged by cryosolution B, containing additionally 25% glycerol. All intermediate steps were achieved by mixing cryosolution A and B in ratios 2:1, 1:1 and 1:2. The crystals were kept not longer than 5 min. in each solution apart from the final solution where the crystals were kept for 40 min.

A crystal was fished out of the cryosolution B with a cryo-loop and plunged into a tube filled with liquid propane. The tube was kept cooled in a bath of liquid N₂ and stored in a tank filled with liquid N₂ when propane solidified. For data collection, the tube was removed from liquid nitrogen, shortly warmed up to partially melt the propane. The "propane plug" was set up on the goniometer head and centred in the middle of the 100 K cold cyro-stream. Such procedure resulted in diffraction images characterised by high background scattering (diffuse scattering) and high mosaicity leading to problems during data reduction due to a high number of overlapping reflections (see chapter 3.5.7). An improved protocol was indispensable to achieve lower mosaic crystals in order to collect higher resolution data, with an appropriate number of separated Bragg reflections. A first attempt was to cool the crystals in liquid nitrogen at temperatures 100 K. The resulting diffraction frames showed the same mosaicity and strong background scattering as observed by freezing in liquid propane. Therefore a number of different cryoprotectants were screened that have been used successfully for other protein crystals. First choices were cryoprotectants with similar molecular weight and analogous functional groups that are able to diffuse into the crystal. The following chemicals were checked, their concentration (v/v) range is given in parentheses: glycerol (15-30%), ethylene glycol (15-30%), *iso*-propanol (20-30%), butane-2,3 diol (8-12%), PEG-400 (25-35%), PEG-400 (25-35%), MPD (20-30%) and Jeffamine[®] M-600 (5-20%). If characteristic ice rings at 3.9, 3.7 and 3.4 Å were visible, the concentration of the cryoprotectant agent was

increased. If no ice formation was visible, a crystal was transferred to the cryoprotectant solution and observed under a microscope by stepwise increase of the cryoprotectant. In case no macroscopic changes e.g. dissolution of the crystals or cracking were visible the crystal was frozen using the loop mounting technique and the X-ray diffraction was checked. The second class of molecules tested were high molecular weight components: maltose (20%), glucose (22%) and dried paraffin oil. All these latter components did not lead to better results. Consequently, the established protocol was maintained with the difference that the crystals were directly frozen in the N₂-stream at 100 K. If the crystals were transferred to the final cryoprotectant solution and stored at 4 °C, they were stable for at least 48 hours. This indicates that glycerol is the first choice in case of PSIIcc.

Another effort was the addition of 5 to 10% glycerol to the crystallisation trials to allow for better diffusion into the growing crystal. We obtained crystals which were then transferred to a cryosolution containing 20 to 30% glycerol. The diffraction quality was the same as obtained for the standard conditions.

Moreover, the temperature is an adjustable parameter at cryo-techniques. Inspired by the studies of Kiefersauer *et al.* (2000), showing that protein crystals undergo phase transitions under controlled humidity, PSIIcc crystals were cooled at different temperatures. To cool a crystal, the gas stream is temporarily shielded with a thin plastic card while the loop assembly is placed on the goniometer head (Hope, 1990). The deflector is then withdrawn rapidly, allowing the gas stream to flow over the crystal. This abrupt release of the cold stream maximises the cooling rate by avoiding passage of the crystal through the warm outer layers of the stream. As it was known that above 150-155 K the vitreous (glassy) ice changes to cubic ice, temperatures between 100 K and 150 K were applied. Surprisingly at 145-150 K an improved mosaicity was observed on the diffraction frames. Whereas at 100 K the mosaicity for a processed dataset of PSIIcc was about 0.8-1.0°, the mosaicity dropped down to about 0.4-0.6° at 145 K. This improvement drastically reduced the number of overlapping reflections and improved data statistics. Consequently all datasets were collected at 145 K except for some of the anomalous datasets, which were collected at 100 K. As the dynamic disorder of PSIIcc molecules in the crystal increases when raising the temperature to around 150 K, slightly higher temperature factors were accepted and they are not relevant at the resolution range of 3.2 Å. The advantages of lower mosaicity exceed the disadvantage of slightly higher B-factors and probable increased radiation damage.

3.5 Crystallographic studies

3.5.1 In-house data collection with an X-ray generator

Prior to freezing, the preparation of the crystals was done under green light, to avoid any exposure of the light-sensitive crystals to light. In-house test data collection was performed on a CuK α rotating anode X-ray generator (Enraf-Nonius, FR571) manufactured by Delft Instruments. Compared to the higher intensity and brilliance of synchrotron radiation, an exposure time of 30 min to 40 min. per frame was necessary to obtain an acceptable signal on the image-plate detector (MAR Research), about 30 to 40 times longer than at the synchrotron, e.g. at the ESRF, Grenoble.

3.5.2 Data collection at synchrotron sources

The advantage of synchrotron radiation is the high brightness of the beam in combination with very high photon flux on the sample. Additionally, the ability to focus the X-rays to very small spot sizes while retaining high brightness enables high spatial resolution. Tuneable beam lines offer the possibility to collect data at different wavelength.

Table 3.2: List of European synchrotron sources visited for X-ray diffraction experiments

source	beamline	insertion element	wavelength [Å]	detector
ESRF ¹	ID 14-1	undulator	0.9330	ADSC Q4R CCD
	ID 14-2	undulator	0.9330	ADSC Q4R CCD
	ID 14-3	undulator	0.9310	MAR165 CCD
	ID 29	undulator	0.8-2.4	ADSC Q4 CCD
SLS ²	PX-1	undulator	0.7-2.4	MAR-165 CCD
DESY ³	X13	undulator	0.8020	MAR-165 CCD
	BM6	undulator	0.7-2.4	MAR-165 CCD
BESSY ⁴	BL-1	undulator	0.7-2.0	MAR-165 CCD
	BL-2	undulator	0.7-2.0	MAR-345 IP

¹ European Synchrotron Radiation Facility, Grenoble, France

² Swiss Light Source, Villigen, Switzerland

³ Deutsches Elektronen Synchrotron, Hamburg, Germany

⁴ Berliner Elektronenspeicherring - Gesellschaft für Synchrotronstrahlung m.b.H., Berlin, Germany

Caused by the macroscopic properties (high mosaicity, long unit cell constants) of the PSII_{cc} crystals, data collection using synchrotron radiation was indispensable. In the frame of this dissertation four different synchrotron sources were used. A summary is given in Table 3.2.

3.5.3 Crystal selection

Since the resolution limit of PSII_{cc} crystals could not be determined in-house at rotating anode generators and the optimised freezing procedure (see chapter 3.4) did not permit transport of the crystals under liquid nitrogen, crystals had to be selected by screening all available crystals with a suitable size of at least 200-300 μm (Fig. 3.1) at the synchrotron beamline. Smaller crystals often have not enough diffracting power. Resolution was judged from two diffraction images taken 90° apart from each other to estimate the degree of anisotropy. Besides the diffraction limit, the mosaicity was estimated by the program MOSFLM. Based on these data the maximal possible resolution was guessed based on long term experience.

3.5.4 Crystallographic characterisation

PSII_{cc} crystals belong to the orthorhombic space group $P2_12_12_1$ with unit cell constants $a=127.5$, $b=224.6$ and $c=305.6$ Å with one PSII_{cc} dimer in the asymmetric unit. The unit cell parameters often showed non-isomorphism; in particular the c-axis diverged between different datasets. Since the PSII_{cc} samples were shown to be homogeneous (chapter 3.1) these changes in the unit cell parameters might reflect the loss of small membrane-intrinsic subunits as well as incomplete composition of the three membrane-extrinsic subunits during crystallisation. Another reason might be the sensitivity to PEG concentration resulting in different precipitant concentration for crystallisation. Furthermore the cooling speed of the crystals is hardly reproducible.

3.5.5 Anisotropic X-ray diffraction

PSII_{cc} crystals showed a strong anisotropy, which had a major influence on data collection. One could distinguish between two complementary orientations separated by 90° rotation about the spindle axis of the diffractometer. The "good" direction showing well defined

Structure determination

regular spots. The diffraction limit in the 011-plane was significantly higher (2.9 Å) accompanied with low mosaicity and good spot separation (Fig. 3.2). By contrast the "bad" direction showed higher mosaicity and a lower diffraction limit of 3.5 Å (Fig. 3.2). This can be clearly seen in Fig. 3.2A and B.

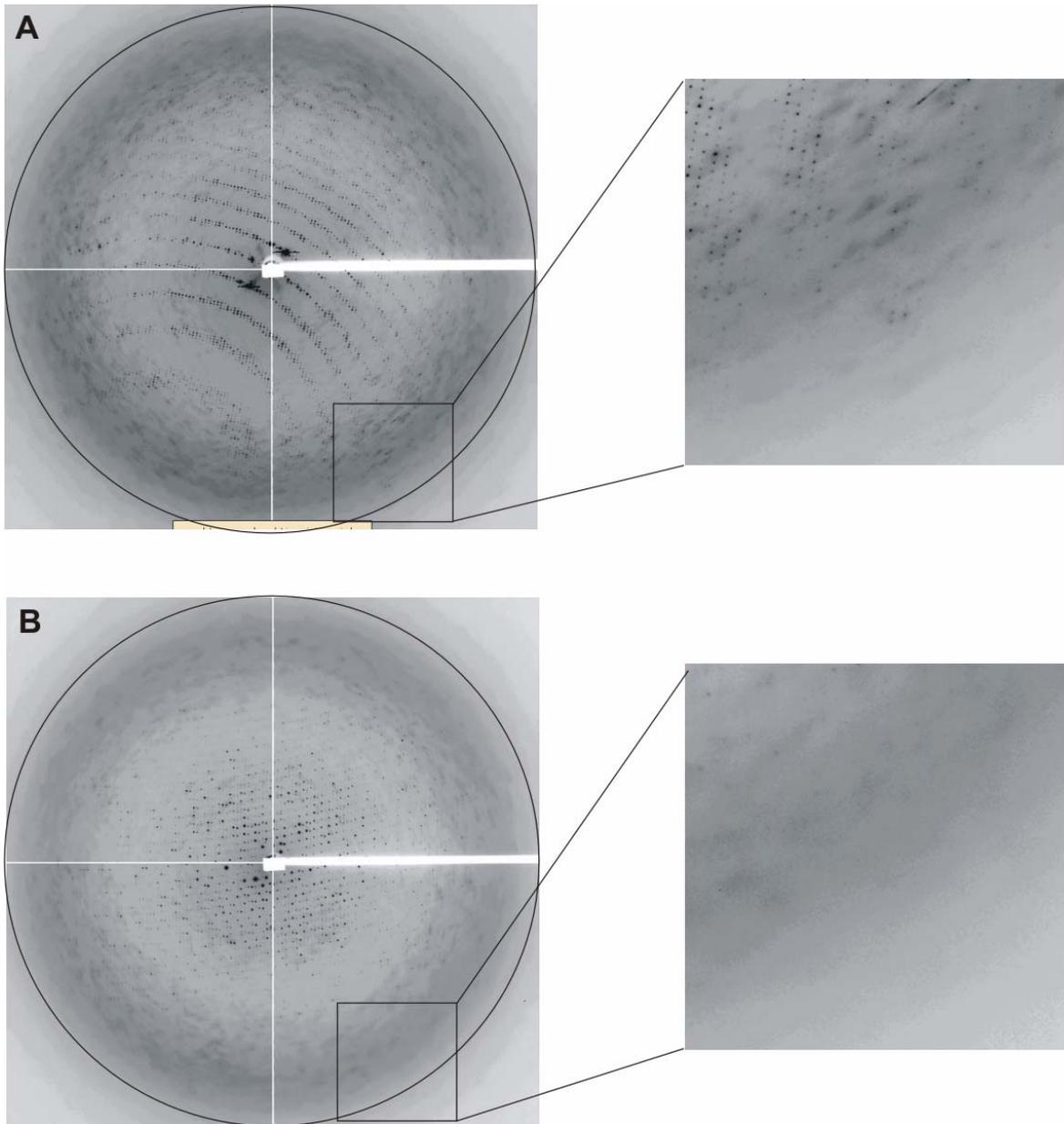


Fig. 3.2: Diffraction image of native PSIIcc in two orientations 90° apart from each other measured at the beamline ID14-2 at the ESRF with a wavelength of 0.933 Å. The dark circle corresponds to a resolution of 2.9 Å. On the right hand side, the lower right corner of the diffraction image A and B is enlarged. **(A)** Measured with $\Delta\phi = 1.0^\circ$. **(B)** Measured with $\Delta\phi = 0.3^\circ$.

For a typical data collection the oscillation ranges were set to 1.0° in the "good" direction, where as in the "bad" direction the oscillation range were set to $0.3-0.5^\circ$ to reduce the number of overlapping reflections. To achieve a complete dataset, it was not sufficient to collect data only around the "good" orientation; and parts of the "bad" orientation had to be included. During data collection the spatial resolution of closely spaced Bragg reflections and a high percentage of partially recorded reflections caused problems during data integration and scaling.

3.5.6 Radiation damage

In dependence of intensity and brilliance of the synchrotron source, radiation damage was observed not only after final data collection and processing, but also directly after the first exposure of a crystal to X-rays. The rectangular shape of the X-ray beam became visible on the surface of the crystals, changing the colour from dark green to dark-purple/black. This observation is in line with crystallographic studies on PSI (Dr. N. Krauss personal communication). The crystals of PSI changed their dark green colour to black, but no purple tone got visible. The higher content of Chl*a* gives the PSI crystals a much deeper dark green colour compared to PSIIcc crystals; hence colour-changes upon exposure to X-ray radiation are more difficult to follow in PSI.

The primary effects of X-ray irradiation of a protein crystal are the ejection of electrons from protein and solvent atoms through the photoelectric effect and through Compton scattering. The photons either interact with the surrounding water molecules and/or the protein and lead to radical formation. The released electrons remain mobile until they reach an electrophilic centre.

The most likely radical reactions are the opening of disulfide bridges, the decarboxylation of carboxylic acids (for a summary of possible radical reactions within protein crystals see (Burmeister, 2000)). Probably the decline of the diffraction power is the most obvious observation which becomes evident, if the diffraction frames during data collection are inspected by eye. In the case of proteins with embedded redox-active cofactors an additional problem may arise due to reduction of metal-centres by electrons generated by X-ray radiation. Along these lines, it was demonstrated that the copper centre in azurin II was reduced by X-rays and reduction became visible as the crystals lose their light-blue colour

Structure determination

(Debenham *et al.*, 1996). Other experiments showed that even at cryogenic temperatures, electron transfer reactions and proton-coupled electron transfer processes occur in protein crystals. These reactions are associated with small-scale movements of ligands and side-chains (Berglund *et al.*, 2002) and references therein). Thus in proteins where different redox-states are observed, the experiments have to be designed such the native state is still present at the end of data collection.

Radiation damage does not only affect solvent exposed residues, but also residues buried in the active sites of proteins as glutamine and histidine residues in the catalytic triad of acetylcholinesterase showed severe increase in B-factors (Weik *et al.*, 2000). Burmeister (2000) observed that movements in proximity of groups with a strong loss of electron density affected the active site of myrosinase.

All these problems have to be considered in PSII_{cc}, because it is involved in charge separation and a series of different electron transfer reactions occurring during water oxidation. It could well be that X-ray generated radicals lead to faster radiation damage in PSII_{cc} crystals and especially in the surrounding of the Mn-Ca-cluster, where oxidation states from Mn(II) to Mn(VI) have been postulated.

3.5.7 X-ray data collection

X-ray data were collected from crystals, whose size allowed at least two translations of the crystal parallel to the spindle axis. As the typical beam size at the ESRF is 100 x 100 μm^2 , suitable crystals for a complete dataset had to have 300 μm in length. Prior to data collection, a single diffraction image was indexed with MOSFLM and the mosaicity calculated. The knowledge of the spindle axis rotation angle φ of the "good" orientation, helped to find the best start and end degrees for the φ (Fig. 3.2). Typically 70° around the "good" orientation was collected in three segments (Fig. 3.3).

The maximum rotation angle (depending on the crystal orientation) was computed in order to reduce the number of overlapping reflections. Fig. 3.3 shows that at the start of data collection from the good orientation, $I/\sigma(I)$ is initially high and then drops down. This is repeated after each translation of the crystal. The $I/\sigma(I)$ quotient in the highest resolution shell decreases during data collection on each crystal segment. In the second segment the I/σ quotient drops

first down, passes through a minimum and increases again. This minimum indicates the strong anisotropy with medium diffraction power in combination with radiation damage.

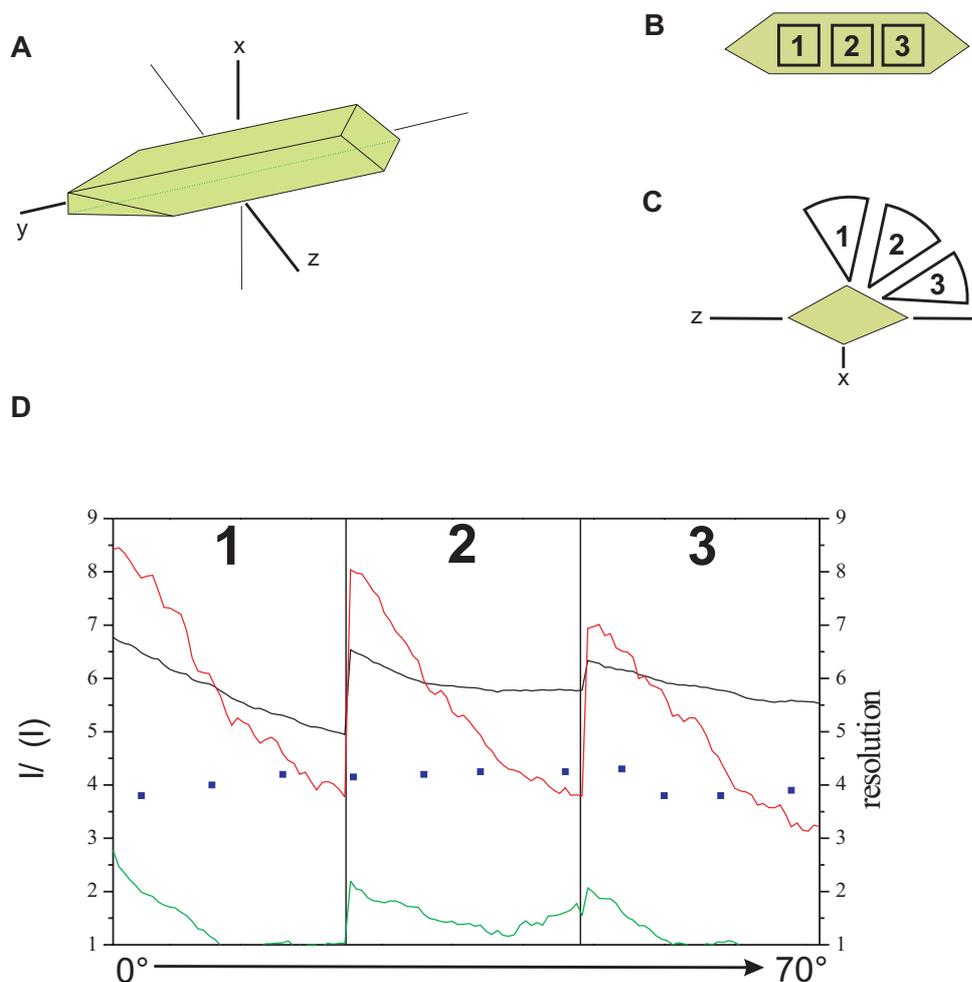


Fig. 3.3: Schematic drawing of a PSIIcc crystal (green). (A) The three axes (x, y, z) are indicated by black lines. (B) Numbering indicates the three beam positions for data collections. (C) View along the y-axis. The three segments of data collection are indicated corresponding to the position given in (B). (D) Scheme of dataset of 70 ° collected starting from the "good" orientation: I/σ for all shells in black, I/σ for the medium resolution shell (4.48-4.54 Å) in green and I/σ for the highest resolution shell (3.70-3.76 Å) in red. Blue dots indicate resolution limit.

Besides the anisotropy and radiation damage the large unit cell of PSIIcc caused problems during data collection. As a large unit cell corresponds to a small reciprocal cell, a huge number of reflections are lying on the Ewald sphere leading to a high number of closely spaced and sometimes overlapping reflections. One advantage of synchrotron radiation is better collimation of the X-ray beam, resulting in better separation of reciprocal-lattice points. Another approach to reduce overlapping reflections caused by the long cell axis, is mounting the longest cell axis parallel to the spindle axis (Fig. 3.3). In the case of PSIIcc such

Structure determination

experimental setup was not possible due to crystal shape, the b-axis being parallel to the long direction of the crystal. Therefore the second longest axis (y) was mounted parallel to the spindle.

A typical exposure time was 60 s per 1° at beamline ID14-2 (ESRF, Grenoble). This rather long exposure time made data collection time consuming. Due to the long exposure time, it could be not avoided to monitor over-saturated peaks on the detector. Such overloads had to be rejected during data processing, since inaccurate low-resolution intensities could have serious effects on the resulting electron density but are essential for correct electron density modification that demands completeness of data in the lower resolution shells. If there were too many overloads, an additional low resolution dataset was collected with reduced exposure time to avoid over-saturation of the detector and therefore in increase in the completeness. For data collection, the total radiation dose of one complete dataset was equally split on the three segments (Fig. 3.3).

The data collection on one single crystal (instead of several crystals) reduced the problems with completeness of data and non-isomorphism which sometimes change randomly between different crystals. In some cases, however, it was not possible to merge data collected at different positions of one single crystal. Especially the cone ends of the crystals showed significantly higher mosaicity compared to central part of the crystals.

Table 3.3: Summary of dataset for structure determination at 3.2 Å resolution. Values in parentheses apply to highest resolution shell. For beamline characteristics see Table 3.2.

Dataset	native
X-ray source	ESRF (Grenoble), ID14-2
Wavelength (Å)	0.933
Resolution ^a (Å)	50 - 3.2 (3.26 - 3.20)
Unique reflections	133,566
R_{sym} ^a	0.106 (0.681)
Completeness ^a (%)	91.3 (56.1)
$I/\text{sig}(I)$ ^a	12.8 (2.0)
Redundancy	4.7

$$^a R_{\text{sym}} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$$

3.6 Diffuse scattering

Crystals of PSII_{cc} show strong diffuse scattering, caused by X-rays being scattered into angles that do not satisfy the Bragg condition. Diffuse scattering can be caused by disorder in the crystal. The diffuse "solvent" ring is evident and noticeably anisotropic. Strong diffuse streaks can be seen which are approximately 3 mm long and symmetry related (Fig. 3.4A and B). Similar streaking has also been observed in images of X-ray diffraction from crystals of bovine brain calmodulin (Wall *et al.*, 1997; Telfer, 2002) and hen egg-white lysozyme (Doucet and Benoit, 1987). Beside the symmetry of the Bragg reflections there is clearly symmetry in the diffuse scattering visible (Fig. 3.4A). In Fig. 3.4B a part of the diffraction pattern is enlarged to illustrate the diffuse background scattering, having a head like shape. There is not only positional symmetry but the diffuse scattering is symmetry-related equally strong.

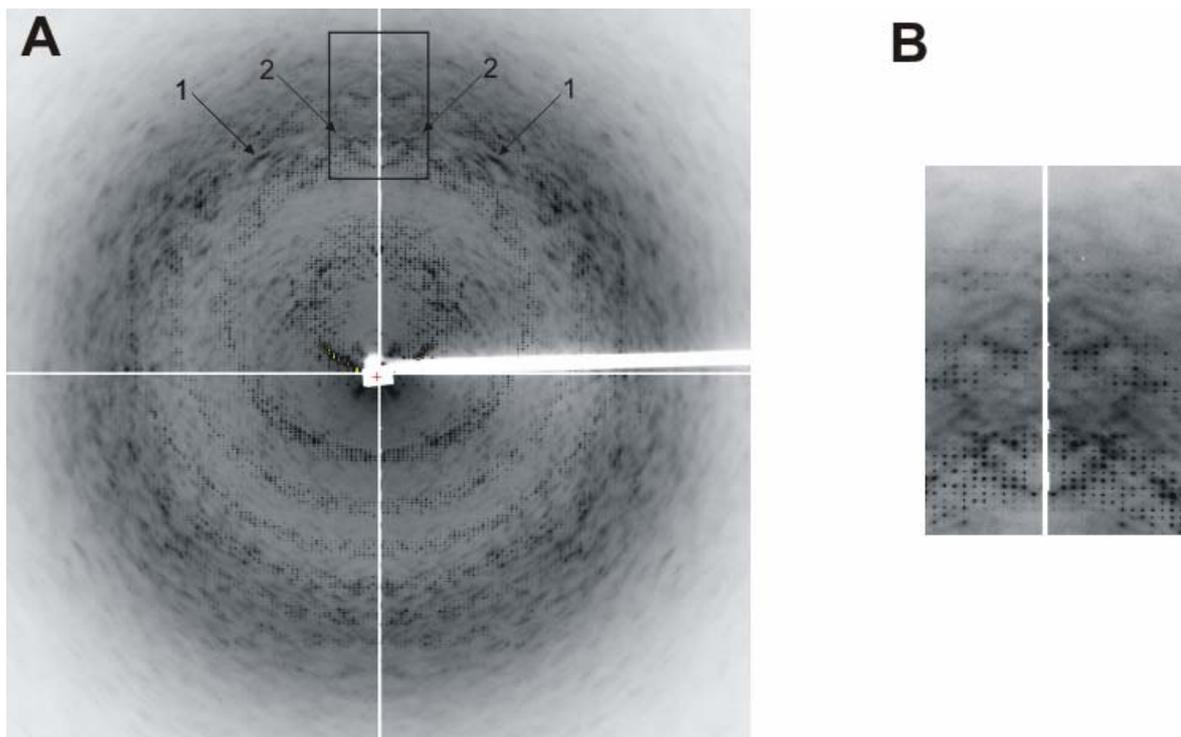


Fig. 3.4: (A) Diffraction image of native PSII_{cc} collected at the beamline ID14-2 at the ESRF with a wavelength of 0.933 Å and $\Delta\phi = 0.3^\circ$. (1) and (2) indicate symmetry related (B) Rectangle in (A) enlarged.

The diffuse background on an X-ray diffraction pattern may arise from several sources including thermal diffuse scattering, static disorder, solvent disorder, Compton scattering, fluorescence, scattering from mounting in tubes and air scattering (Glover *et al.*, 1991).

Structure determination

In case of PSII_{cc}, it might be that some subunits show positional and orientational disorder, leading to breakdown in the crystal translational symmetry and a reduction in Bragg intensities at high resolution. Some features of the diffuse scattering at low resolution could be explained by displacements of both monomers in the asymmetric unit as a whole rigid body (personal communication Dr. G. Bourenkov). Another possible source of the diffuse scattering might be fluctuations in the detergent shell around the protein. Sometimes also the incorrect choice of cryoprotectant leads to strong background scattering. Therefore all trials to optimise the cooling protocol (see chapter 3.4) were done under this aspect. A further possible reason of diffuse scattering might be a loose packing in PSII_{cc} crystals (see chapter 3.9). A possible candidate for a rather flexible part of the structure is PsbO. This subunit is very ill defined in the electron density maps, which might be due to a higher flexibility of this subunit or partial degradation. Strong background scattering leads to difficulties in correct integration of Bragg reflections.

3.7 Single-wavelength anomalous scattering

The SAS measurements were performed at BW6 beamline of the Max-Planck Society at DESY or at the ID29 beamline at ESRF (Table 3.2). Prior to data collection, an EXAFS (fluorescence scan) was performed. The beamline setup allows to continuously changing the energy (wavelength), to detect the maxima in the fluorescence spectra of the crystal. From Fig. 3.5 three prominent maxima at 4.0381, 6.5390 and 7.1120 keV corresponding to Ca, Mn and Fe could be identified (Table 3.4).

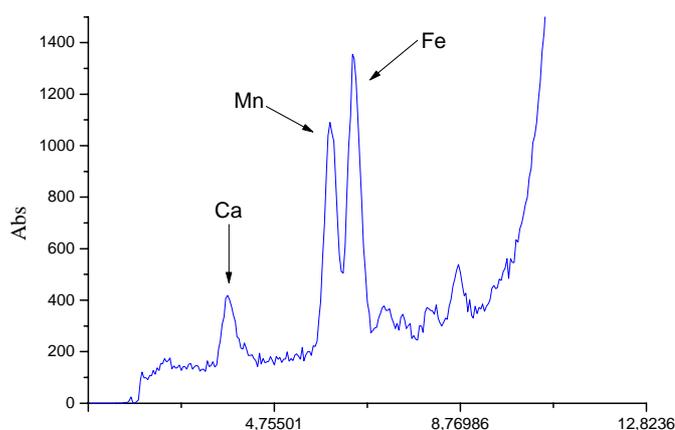


Fig. 3.5: Fluorescence scan recorded prior to experiments at beamline BW6 at the DESY.

Whereas the Ca and Mn are part of the Mn-Ca-complex, Fe^{2+} are bound to the haem groups of cyt *b*-449 and cyt *c*-550 and there is a non-haem Fe^{2+} . Therefore two experiments were set up. One dataset was collect at the Fe-edge (at 1.733 Å) and the other dataset was collect beyond the Mn-edge (at 1.910 Å). For dataset statistics see Table 3.5. Since the intensity of the beam at this long wavelength is reduced due to absorption effects, slightly longer exposure times than for normal data collection had do be used. Special care was taken to avoid fast crystal decay and radiation damage.

Table 3.4: Overview of elements expected to be bound to PSIIcc. The energy and the corresponding wavelength are given.

K-edge	keV	Å
Mg	1.3050	9.5007
S	2.4720	5.0155
Cl	2.8224	4.3929
Ca	4.0381	3.0704
Mn	6.5390	1.8961
Fe	7.1120	1.7433

Anomalous difference electron density maps of the dataset collected at the Fe-edge clearly showed four separated peaks. Two of them could be attributed to the haem group of cyt *c*-449 and cyt *b*-550. One single peak originated from the non-haem Fe^{2+} . The fourth peak was bulky and represented the Mn-Ca-cluster, probably composed of 4 Mn-cations and 1 Ca^{2+} .

Table 3.5: Summary of dataset for structure determination at 3.2 Å. Values in parentheses apply to highest resolution shell. For beamline characteristics see Table 3.2.

Dataset	Fe-edge	Mn-edge
X-ray source	DESY (Hamburg), BW6	DESY (Hamburg), BW6
Wavelength (Å)	1.733	1.910
Resolution ^a (Å)	40 - 4.0 (4.07 - 4.00)	40 - 4.0 (4.07 - 4.00)
Unique reflections	76,072	77,931
R_{sym}^a	0.060 (0.398)	0.064 (0.419)
Completeness ^a (%)	96.0 (91.9)	98.1 (90.2)
$I/\text{sig}(I)^a$	16.0 (2.7)	17.4 (3.5)
Redundancy	3.3	4.6

$$^a R_{\text{sym}} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$$

Structure determination

The anomalous electron density maps of the dataset beyond the Mn-edge showed only one single strong peak in the same location as the much larger peak originating from the Mn-Ca-cluster. In agreement with the EXAFS scan, it was attributed to Ca^{2+} .

3.8 Electron density maps – first orientation

Besides the computed crystallographic quality criteria (R-factor, figure of merit), a critical subjective analysis of the calculated electron density was essential. To get a first orientation and to distinguish between the cytoplasmic and luminal side, the high peaks for the Mn-ions in the electron density were exploited. Supplementary information to discriminate between the cytoplasmic and luminal side is the bulky electron density for the three extrinsic subunits on the luminal side attributed to PsbO, PsbU and PsbV.

3.9 Protein packing

The PSII_{cc} dimer consists of the two monomers labelled [A] and [B] in Fig. 3.6. The two monomers are related by a non-crystallographic C₂-axis, which is not parallel to any crystallographic symmetry axis. Each monomer [A] forms only contacts with monomer [B] from the neighbouring dimers and accordingly no [B]-[B] or [A]-[A] contacts occur between neighbouring dimers.

Altogether we can distinguish four different contacts shown in Fig. 3.6: (1) is between the cyt *c*-550 and the small luminal 12 kDa protein, (2) is between unassigned TMH 2 on the luminal side and the 33 kDa protein. Contact (3) is between the 12 kDa protein and unassigned TMH 7 on the cytoplasmic side of another PSII_{cc} molecule. Contact (4) is formed by the 33 kDa protein and the cytoplasmic side of CP43 of the other molecule.

The nature of these interactions is not obvious at the present resolution as for several subunits clear amino acid assignment is still elusive. Only relative small parts of the protein surface participate in crystal contacts. This seems to be a common characteristic in crystals formed by membrane proteins and might be one of the reasons why it is often difficult to obtain X-ray diffraction data at higher resolution.

This crystal packing is not in agreement with the packing model proposed by Kuhl *et al.* (2000).

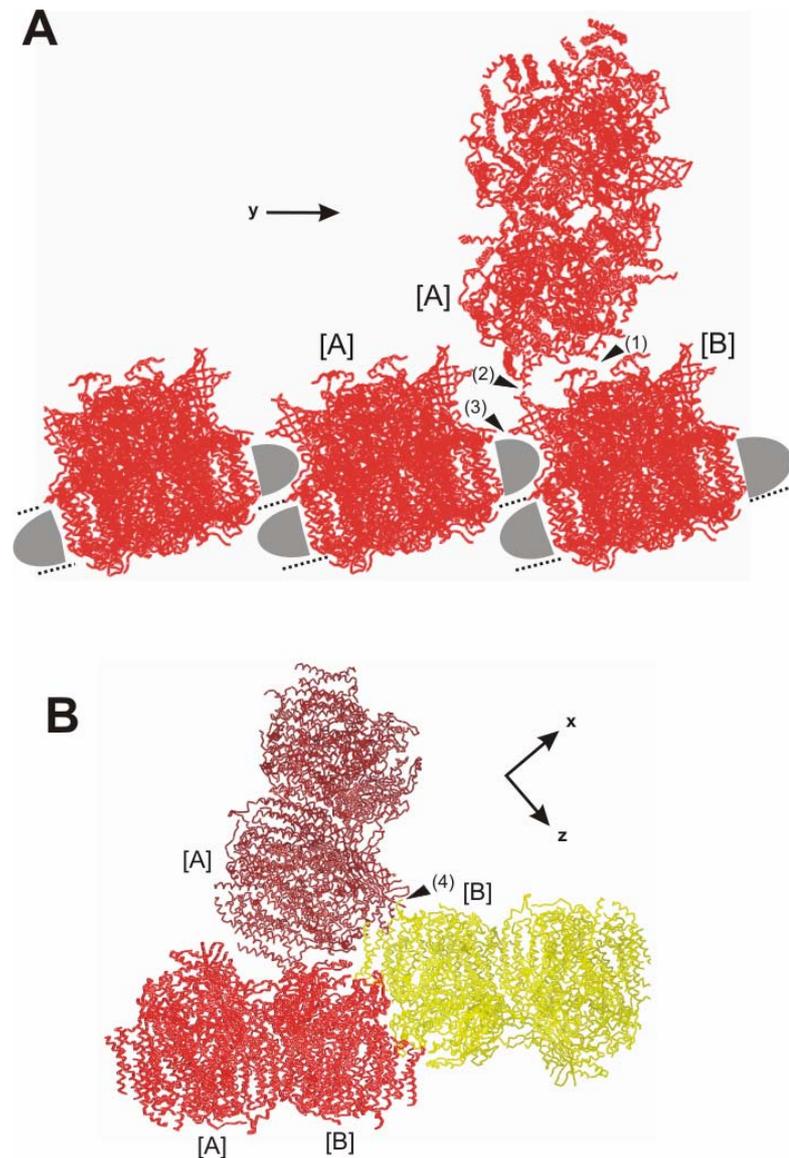


Fig. 3.6: Crystal packing of PSIIcc. The two monomers of the PSIIcc homodimer are labelled [A] and [B] (red and yellow). The intermolecular contacts are marked and numbered. (A) The dotted lines indicate the pseudolayer formed by the membrane neighbouring PSIIcc. The detergent shell is indicated in grey. (B) View onto the x,z-plane. Contacts (1) to (3) are not indicated for clarity.

3.10 Model building

The structure models were built as described in the following paragraph. At the present resolution of 3.2 Å and with relatively low quality of experimental phases (acceptable to 4.2 Å resolution), the electron density shows only limited extent of details. All secondary structure elements, however, are well defined at the level of main chain conformation. Based on the model at 3.8 Å (Zouni *et al.*, 2001) the Ca-trace was stepwise replaced by a polyalanine model, starting from the regions with the best defined secondary structure, the

Structure determination

TMH. In the 3.8 Å model a field of 36 TMH was identified. Some of the TMH showed significant bending up to 30°. The TMH have a maximum length of 35 Å, spanning once through the thylakoid membrane and jutt out on both sides of the membrane. As found for other membrane proteins, the primary sequence shows hydrophobic residues within the membrane and on the membrane surface charged amino acids are located that interact with the polar head groups of the lipids. In initial steps idealised α -helices were fitted into the electron density of TMH. Torsion angle were adjusted according to the bending of the α -helices. The geometry of modelled helices was surveyed with a self-written program (HELIX-FIND, provided by Dr. J. Biesiadka). The peripheral TMH at the longest distances from the pseudo- $C2(Fe^{2+})$ axis of PSIIcc were less well defined in the electron density than those in the centre. One reason might be a partial loss of some low molecular weight subunits during solubilisation of the membrane intrinsic complex, reducing their occupancy. Another possible explanation could be a small rotation and dislocations of individual small subunits of the PSIIcc-dimer in the crystals.

Besides the membrane spanning TMH, small α -helices were observed nearly parallel to the membrane plane. In PbRC and PSI similar occurrence of small amphiphilic α -helices was described, having one hydrophilic side solvent exposed and the other hydrophobic side interacting with hydrophobic residues buried in the membrane. Several short α -helices could be identified, that were well defined in electron density. Their numerous occurrences seem to be characteristic for membrane proteins that are mainly α -helical. Connections between single TMH were subsequently traced, with difficulties due to breaks and branches in the electron density.

The amino acid assignment of some subunits was facilitated due to cofactor binding. PSIIcc contains two cytochromes, cyt *b*-449 and cyt *c*-550. Each contains one haem molecule with a central Fe^{2+} . The RC subunits D1 and D2 coordinate a Fe^{2+} positioned on the local pseudo- $C2(Fe^{2+})$ axis. Moreover, PSIIcc contains the catalytically active Mn-Ca-cluster on the luminal side built up of 4 Mn and 1 Ca cations. All metal cations could be located by contouring the electron density at high sigma levels or in electron density maps calculated with anomalous data on and beyond the Mn-edge.

3.10.1 Chlorophyll *a* and pheophytin *a*

Disk-shaped like electron density were assigned to Chl*a* or Pheo*a* molecules as they are the only known tetrapyrrol derivatives in the ETC and the antenna proteins of *T. elongatus*.

The Chl*a* is a derivative of phorbins, but in contrast to the fourfold rotation-symmetric phorbinsystem, significantly asymmetric. The asymmetry is introduced by a reduction of a double bond in ring IV and an additional fifth cyclopentanone-ring. The conjugated aromatic π -system of these chlorin rings results in a planar system. Due to the low resolution, they were modelled as porphyrins, in the early steps of refinement, because neither side chains nor the phytol chains could be unambiguously identified. In later steps of refinement, for most of the Chl*a* the orientation could be determined as positions of the side chains became clear (see chapter 7.2 and Table 10.4).

3.10.2 Carotenoid

The structure at 3.2 Å of PSIIcc revealed for the first time a β -carotene. Initially the central part of the isoprenoid tail without the head groups could be modelled into the electron density. Subsequent refinement cycles allowed extension of the isoprenoid tail and finally modelling of the two jonon rings. These observations are in line with modelling experience at higher resolution (Dr N. Krauss, personal communications and (Roszak *et al.*, 2004)) that for atoms in the central part of carotenes the B-factors are lower than for atoms of the head groups. This observation can be explained with a higher flexibility of the jonon ring compared to a more rigid structure of the conjugated polyene-chain. At the current resolution, no conclusion can be drawn about the structure of the head group. The jonon ring was modelled with two methyl groups as substituents, but an interpretation as a zeaxanthin molecule can not be excluded. As the pigment analysis revealed that PSIIcc contains mainly β -carotene (Kern *et al.*, 2004a), the carotenoid was modelled as β -carotene.

3.10.3 Unexplained electron density

In addition, there are several unconnected shorter, elongated patches of electron density in other places in the RC and antenna proteins that could represent β -carotenes, Chl*a* phytol chains, lipids or detergent molecules. At the current resolution we are not able to assign these

Structure determination

patches, as they show no branching points or regions with more bulky electron density. For refinement purposes, aliphatic tails were placed in these elongated patches.

3.10.4 Sequence assignment

The anomalous signal aided in the chain tracing by marking sulphur of cysteine and methionine residues. In fact, instead of looking for electron-density features that match large aromatic side chains as an aid in initially locating the primary sequence into the map, we recommend that one first looks for sulphur peaks found in an anomalous difference Fourier map (see chapter 3.7). This is analogous to finding the methionine positions in a MAD-phased map. Sulphur peaks also provide an important and easy information whether cysteine and methionine residues are correctly placed (Yang and Pflugrath, 2001). One should take into account however, the sensibility of sulphur-containing-residues to radiation damage. Disulfide bridges could be opened or even completely destroyed by the release of CH_3SH (Burmeister, 2000).

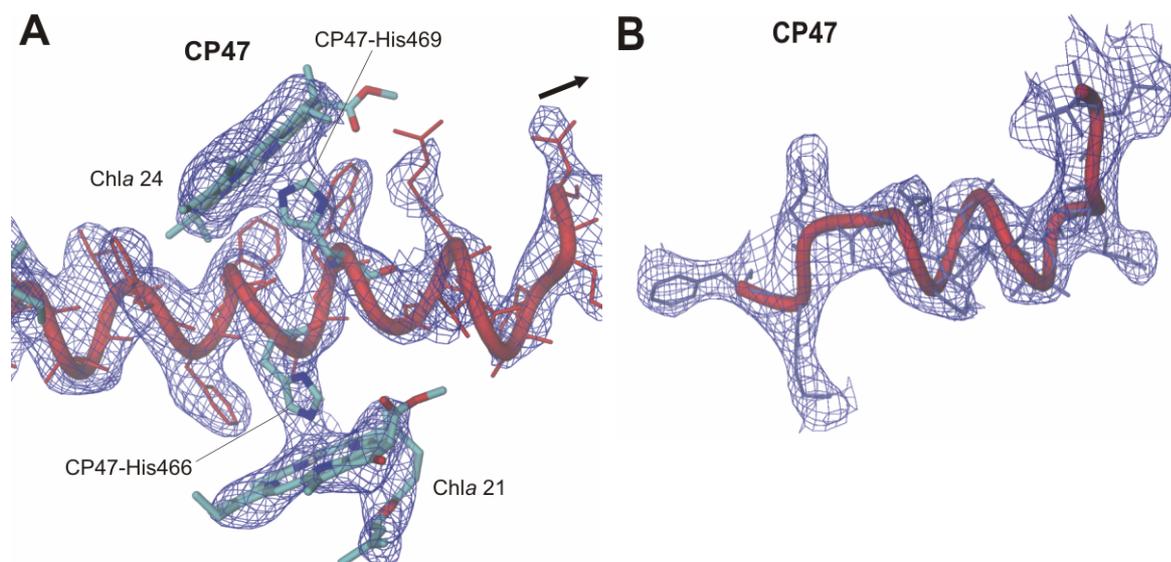


Fig. 3.7: Electron density (blue) is contoured at 1.0σ level. (A) TMH-f of CP47 (red) Two coordinating His residues (CP47-His466 and CP47-His469), the Chla molecules (Chla21 and Chla24) and the assigned amino acid sequence are shown in the electron density. The amino acid sequence is assigned (red side chains). (B) Luminal loop region of CP47: The amino acid sequence is tentatively assigned (dark side chains).

The correct position in the amino acid sequence was subsequently based on bulky aromatic side chains, in particular on the localisation of Chla and on the assumption that they are axially coordinated predominantly by $\text{N}\delta$ or $\text{N}\epsilon$ of histidine. Chla coordination was used as

anchor points in polypeptide modelling and helped to assign the amino acid sequence for the reaction centre subunits D1 and D2 and the antenna domains CP47 and CP43 (Fig. 3.7).

At medium resolution it is only possible to discriminate between "short" and "long" side chains (Fig. 3.7). This enabled us only to assign the primary sequence in the transmembrane spanning regions. The further away from the C- or N-terminal ends of a TMH, the more difficult the correct assignment became. Therefore the amino acid sequence of short loops connecting two TMH could be assigned, whereas longer loops could only be modelled as polyalanine or as $C\alpha$ -trace.

Additional information to assign the sequence was the sequence homology of the TMH in the RC subunits D1 and D2 of PSII to the RC subunits L and M of PbRC (Fig. 3.8). Contrasting loop regions where the sequences of PSII and PbRC subunits differ due to deletions and insertions. For CP47 and CP43 the relation to the antennae domains of PsaA and PsaB of PSI helped to assign the sequence (Fig. 3.8). Beside this information retrieved from the primary sequence, the twofold pseudo-symmetry within the monomers of the PSII_{cc} dimer was applied to transfer parts of the model between the two monomers. By this procedure the expenditure to trace these main chains was reduced.

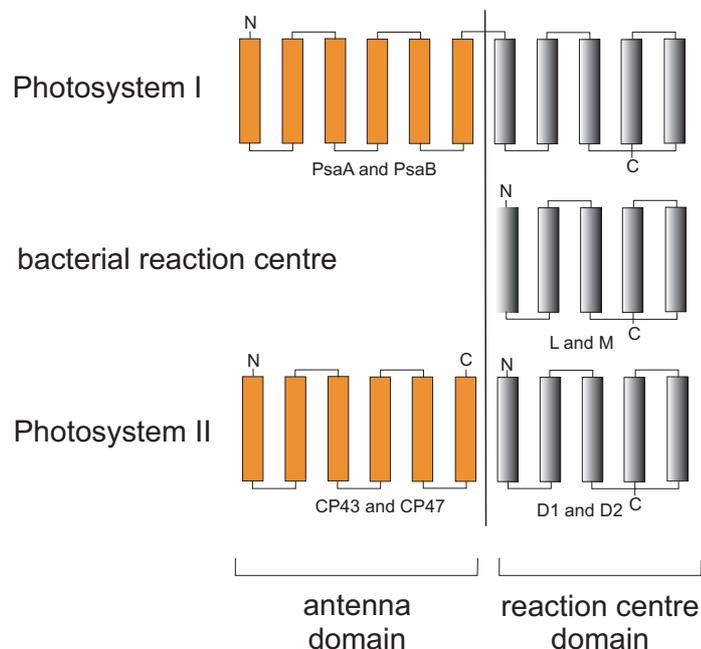


Fig. 3.8: Schematic representation of structurally related proteins/domains of PSI, bacterial RC and PSII. TMH are drawn as rectangles: antenna proteins/domains in orange and RC proteins/domains in grey. N- and C-termini are indicated.

3.11 Refinement

Crystallographic refinement was performed with the program CNS (Brünger *et al.*, 1998). All refinement steps were performed based on protocols of the program package. Prior to the geometric refinement a temperature-factor-correction of the X-ray data was done as well as a correction of amplitudes for diffraction of bulk solvent (bulk solvent correction). In the first step the model was refined according to the "rigid-body" protocol. The different subunits were defined as rigid groups and therefore refined as distinct rigid bodies. In the early stages of model building temperature factors (B-factors) were not refined, because this procedure destroyed the geometry of the model. In later steps of refinement, temperature factors were isotropically refined, according to the protocol "bindividual" in 30 cycles, but they were restricted to minimum and maximum values of 20 Å² and 90 Å², respectively. In the second step harsh mistakes in the model such as too short van der Waals contacts (side chain clashes of different subunits) were corrected by applying potential energy at constant temperature. In the following step a molecular dynamics refinement was performed according to the "simulated annealing" protocol based on the force field calculations of (Engh and Huber, 1991). The starting temperature was in a range of 1000-2000 K. The temperature was lowered in steps of 50 K. Subsequently, the refinement was finalised by 30 cycles at a constant temperature of 300 K.

The accuracy of the model was judged by the R_{free} -factor for which out of the 133,566 measured reflections, 1603 (1.2%) were excluded from refinement. Only 1.2% was used for the test set to avoid a low data to parameter ratio. This was especially important for the high resolution shells where the completeness was low (~56%). During the refinement the weighting term between the crystallographic and the geometric energy term was controlled by the R_{free} -factor and root mean square deviation (rmsd) of bond length and bond angles.

3.12 Architecture and characteristics of TMH

TMH exhibit a strong dipole moment due to strict orientation of their peptide bonds (Hol, 1985) and are consequently arranged to their approximately antiparallel within the membrane. The core of the PSII_{cc} RC is formed by TMH-**d** and -**e** of D1 and D2, respectively, that is packed as four-helix bundle and imparts electrostatic stabilisation. The distribution of positively charged amino acid residues on the TMH is characteristic and was

described as the "positive-inside-rule" (von Heijne, 1989). For example, arginine and lysine are in general much more abundant in the cytoplasmic domains than in periplasmic domains. The interior amino acids are almost exclusively non-polar. Polar and charged residues are located close to the membrane surface interacting with head groups of lipids and with aqueous solution.

Three-dimensional structures of biological macromolecules are mainly stabilised by H-bonding and van der Waals interactions. Among the "conventional" hydrogen bonds of the form X—H•••A (with X—H the donor and A the acceptor), those between peptide groups, N—H•••O=C, dominate as they define the secondary structure elements α -helix, β -pleated sheet and β -turn. These hydrogen bonds are complemented by hydrogen bonds with X = N, O, S and A = N, O that are formed between main-chain and side-chain atoms or between side-chain atoms. In the past two decades, hydrogen bonds with C—H as donor have been frequently observed and are now considered to be an important cohesive force in organic and biological structures (Desiraju and Steiner, 1999; Jiang and Lai, 2002). Systematic studies showed that C α —H•••O hydrogen bonds are frequently formed between closely packed α -helices (Senes *et al.*, 2001; Loll *et al.*, 2003). C α —H•••O hydrogen bonds can guide the specificity of hydrophobic attractions and thus increase the rigidity of the protein in lipophilic environment. Recently it was demonstrated that the occurrence of C α —H•••O seems to be of functional importance. They are not randomly distributed but many of them flank Chla coordinating amino acids in PSI, suggesting stabilisation of these structural segments (Loll *et al.*, 2003). If salt-bridges and/or hydrogen bonds are observed within TMH, they significantly contribute to stabilisation due to the low dielectric constant $\epsilon_p = 4$ for protein compared to $\epsilon_w = 80$ for water.

It has been generally observed that the motifs GxxxG (Senes *et al.*, 2000) and AxxxA (Kleiger *et al.*, 2002) are abundant in strong helix-helix C α —H•••O interactions, as small amino acid side chains facilitate close contacts between TMH. The role of these motifs is also supported by mutagenesis studies (Russ and Engelman, 2000). In the trimerisation domain PsaL of PSI, two TMH harbour these motifs resulting in tight packing of helices characterised by low B-factors (Loll *et al.*, 2003).

Structure determination

Helix dipole interactions in the TMH exhibit a substantial dipole moment due to the nearly parallel alignment of the dipole moments of the individual peptide bonds (Hol, 1985). In Fig. 3.9 a top view of PSI is drawn that illustrates the nearly perfect anti-parallel arrangement of TMH. This arrangement is not only restricted to single subunits but also observed to occur between subunits, and especially the two central subunits PsaA and PsaB fulfil these very well and the low molecular weight subunits on the periphery of PSI are mainly packed antiparallel.

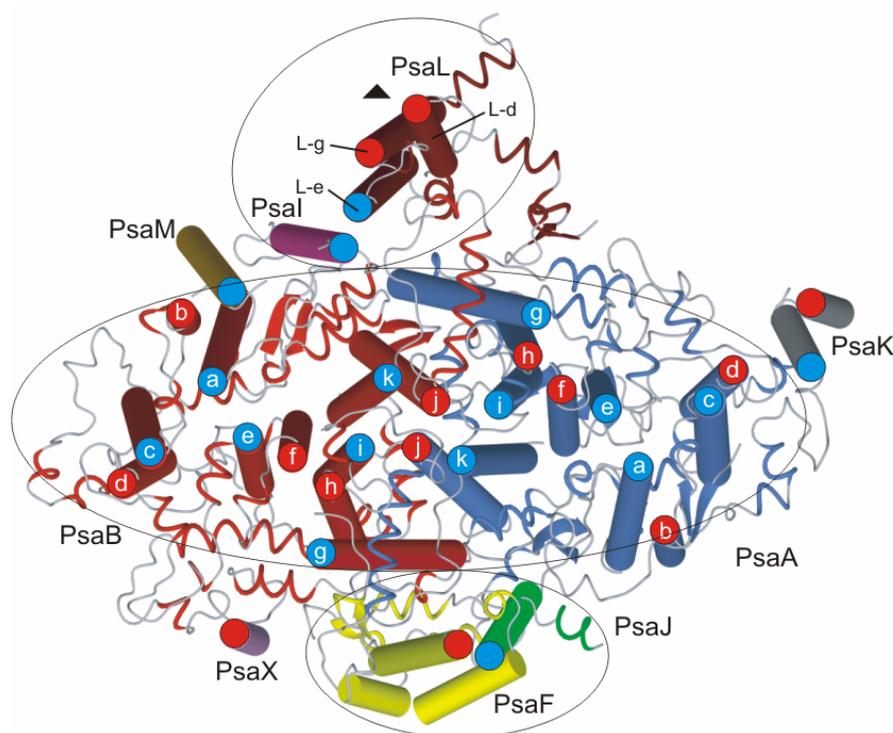


Fig. 3.9: Top view of PSI at 2.5 Å resolution (Jordan *et al.*, 2001) from the cytoplasmic side. The membrane-extrinsic subunits (PsbC, PsbD and PsbE) are omitted. TMH of PsaA a to k (blue), PsaB a to k (red), PsaF (yellow), PsaI (violet), PsaJ (green), PsaK (grey), PsaL (brown), PsaM (orange) and PsaX (pink) are drawn as cylinders. The helix-dipole is indicated on the top of each TMH: red circle (N-terminus) and blue circle (C-terminus). The grey ellipses indicate the gene clusters: PsaA and PsaB, PsaF and PsbJ as well as PsaI and PsaL.

3.13 Shape and charge distribution

The longest dimensions of PSc measure 200 x 100 Å² (Fig. 3.10). The membrane-integral part is 40 Å thick and extends above the cytoplasmic side of the membrane surface by no more than 20 Å. The extended loop domains of CP47, CP43 and the membrane-extrinsic subunits PsbO, PsbU and PsbV extend by around 78 Å into the lumen, the longest extension

being formed by PsbO. The membrane-intrinsic hydrophobic core is clearly separated from the solvent exposed regions that are dominated by hydrophilic and charged amino acids thereby increasing the solubility of PSIIcc.

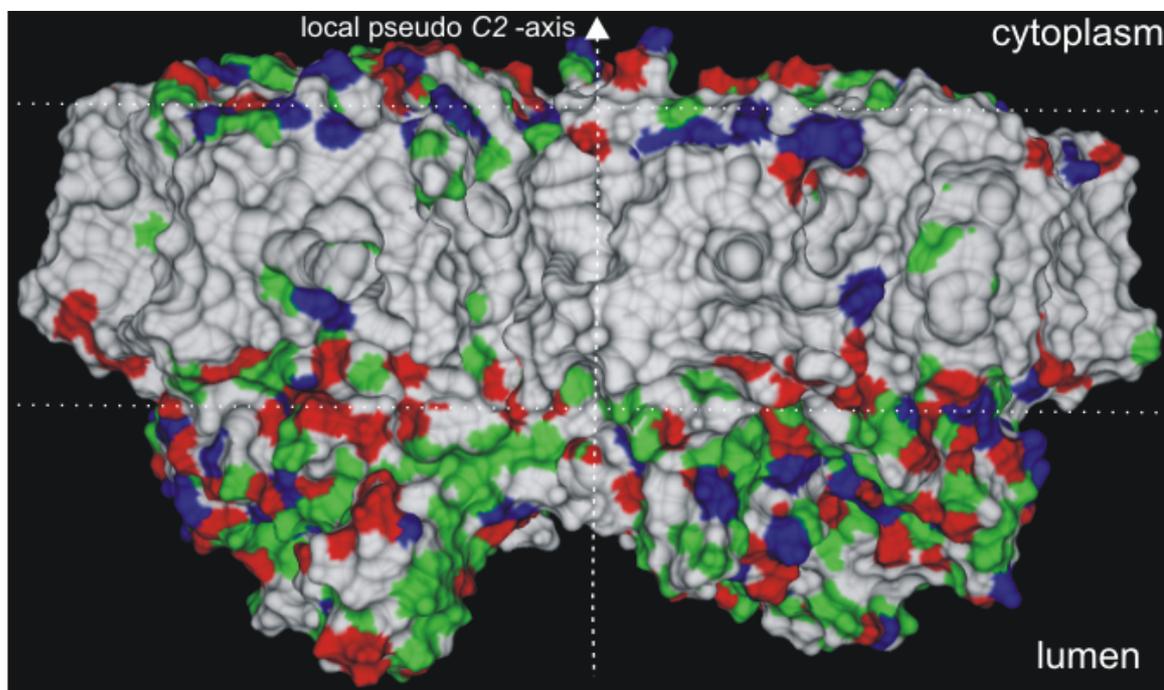


Fig. 3.10: Surface representation of the dimeric PSIIcc. The thylakoid membrane is indicated by dotted lines. The two monomers are related by a pseudo C2-axis (dashed line). The horizontal dotted line (white) indicates the membrane region Polar residues in green, positively charged residues in blue and negatively charged residues in red.

3.14 Neutron scattering

The largest PSIIcc crystals with dimensions (1000 x 300 x 100 μm) were soaked for at least one week in 100% ($^2\text{H}_2\text{O}$)/($^2\text{H}_2\text{O}$ + $^1\text{H}_2\text{O}$) and mounted in quartz capillaries. A small drop of mother liquor was left in the capillary to avoid slipping and of drying out of the crystal during data collection.

Neutron diffraction data for PSIIcc crystals were collected at room temperature on a 4-circle diffractometer (instrument DB21) at the Institute Laue-Langevin (ILL) in Grenoble (France). The sample was illuminated with collimated monochromatic neutrons supplied by the ILL nuclear reactor's cold source. The neutron wavelength was fixed at 7.53 \AA by reflection from a K-intercalated graphite monochromator. One test dataset was collected at 60 \AA resolution to

Structure determination

check whether the PSIIcc crystals have sufficient size. The sample to detector distance was 250 mm and data were collected in steps of 0.2° steps and an exposure time of 30 min per image. Reflections shown in Fig. 3.11 were monitored with a scintillation-detector of the Anger camera type. The $^2\text{H}_2\text{O}$ content of the mother liquor from which crystals were harvested was measured by neutron transmission. Unfortunately the volume of the crystals was not large enough to collect a dataset with about 30-40 Å resolution.

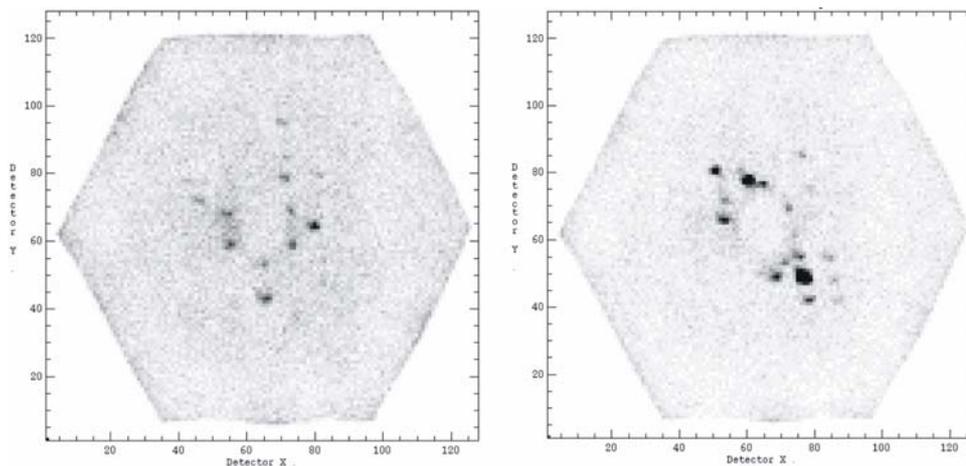


Fig. 3.11: Two different neutron diffraction images collect on beamline DB21 at the ILL in Grenoble. Experimental setup: $\lambda = 7.53 \text{ \AA}$, sample to detector distance 250 mm, $\Delta\phi = 0.2^\circ$ exposure time of 30 min.