

2 Methods

2.1 Cultivation, isolation and purification of PSIIcc

The biochemical work was performed at the Max-Volmer-Laboratorium at the Technische Universität in the group of Drs A. Zouni and K.-D. Irrgang. PSII was isolated from the thermophilic unicellular cyanobacterium *T. elongatus*. This species is strictly photoautotrophic, has an optimum growth temperature of 56 °C and inhabits hot springs. The cultivation was performed as described in (Schatz and Witt, 1984). Cells were grown in Castenholz Medium D (Castenholz, 1988) within seven days at constant pH of 7.8 achieved by addition of CO₂ and increasing illumination adapted to the optical density of the culture. Thylakoids were prepared from cells after lysozyme digestion, followed by cell disruption using a Yeda press. PSII complexes were solubilised using 0.6% n-dodecyl- β -D-maltoside (β -DM) and separated from membrane fragments by centrifugation. All further purification steps were conducted at 6°C in the dark or under dim green light.

Buffers used for protein chromatography were mixed from buffer A (40 mM MES-NaOH, pH 6.0, 20 mM CaCl₂, 0.02% (w/w) β -DM, 5% (w/w) glycerol) and buffer B (as buffer A including 50 mM MgSO₄). In the first purification step, the protein was diluted with the same volume of a mixture of 76% buffer A and 24% buffer B and loaded on an anion exchange column packed with Toyo Pearl DEAE 650S (TosoHas, Japan). After sample loading a long washing step at constant salt concentration was used (24% B, 5 CV) followed by a step to 50% B and a gradient from 50% to 80% B in 3 CV. Three main fractions were observed with increasing gradient: (1) phycobilisomes, (2) PSIIcc, contaminated with traces of PSI and ATP-synthetase, and (3) PSI.

Before the second purification step, the fractions containing PSIIcc were combined and diluted with buffer A to lower the salt concentration. As in the first purification step, Toyo Pearl DEAE 650S column material was used, packed in a column with different geometry. A washing step of 5 CV with 24% B was followed by a gradient from 24 to 80% B in 7 CV. In this purification step, peaks were separated into phycobilisomes and ATP-synthase, monomeric PSIIcc, dimeric PSIIcc and PSI. Fractions containing dimeric PSIIcc were pooled and concentrated. Immediately after concentration the dimeric PSIIcc solution was brought to a Chl_a concentration of 0.74 mM by dilution with buffer C (100 mM PIPES, pH 7.0, 5 mM

Methods

CaCl₂, 0.03% (w/w) β -DM). The protein solution was mixed with the same volume of buffer D (100 mM PIPES, pH 7.0, 5 mM CaCl₂ and 15% (w/w) polyethylen glycol (PEG) 2000). After incubation for 1 h, the resulting precipitate was centrifuged and the pellet was re-dissolved in buffer C to yield a concentration of 0.74 mM Chl_a. This step of precipitation was repeated twice with the same solution and decreasing final concentration of PEG-2000. In the third step micro crystals were obtained over night. These were re-dissolved in buffer C to yield a final Chl_a concentration of 4 mM (Kern *et al.*, 2004a).

2.2 Spectroscopic and biochemical characterisation of PSIIcc

The Chl_a and Car content of PSIIcc preparations were determined by UV-VIS spectroscopy after extraction with 80% (v/v) acetone. Oxygen evolution measurements were performed with a Clark type electrode. Artificial electron acceptors were added (2 mM ferricyanide and 0.2 mM p-benzoquinone or 2 mM 2,6-dichloro-p-benzoquinone). Lipid and β -DM content was investigated using 2D thin layer chromatography. The Mn content was determined by atom absorption spectroscopy.

The subunit composition was analyzed using SDS/Urea PAGE, Matrix assisted laser desorption ionisation time of flight mass spectrometry (MALDI-TOF MS) and N-terminal amino acid sequencing.

2.3 Amino acid sequences of PSIIcc subunits

By phylogenetic analysis of 16S rRNA sequences it was found that the meso- and thermophilic strains of *Synechococcus* could belong to different genera, therefore the former ones were assigned to a new genus *Thermosynechococcus* (Katoh *et al.*, 2001). The entire genome of the thermophilic unicellular cyanobacterium, *T. elongatus* BP-1, was recently sequenced. The genome consists of a circular chromosome 2,593,857 bp long. 63% of the *T. elongatus* genes show significant sequence similarity to those of both *Synechocystis* sp. PCC 6803 and *Anabaena* sp. PCC 7120, while 22% of the genes were unique to this species, indicating a high degree of divergence of the gene information among cyanobacterial strains (Nakamura *et al.*, 2002). In the frame of this dissertation all protein sequences were taken from *T. elongatus*.

2.4 Crystallisation and cooling

2.4.1 Crystallisation of membrane proteins

In contrast to the high number of about 26,000 3D structures of soluble proteins, the available number of integral membrane proteins is about 90. This is remarkable, since more than 2/3 of all gene products are expected to be integral membrane proteins. Of special interest are membrane proteins that have photosynthetic or bioenergetic properties, forming pores as well as transporters and receptors. During the last 3 years a number of medium to high resolution crystal structures of membrane proteins were published, for example: PSI (Jordan *et al.*, 2001), cyt *b₆f* complex (Stroebel *et al.*, 2003; Zhang *et al.*, 2003), protein-conducting channel (van den Berg *et al.*, 2004), myobacterial outer-membrane channel (Faller *et al.*, 2004), ATPase (Iwata *et al.*, 2004) and aquaporin water channel (Sui *et al.*, 2001).

Membrane proteins are difficult to handle due to their amphipathic nature. They possess a hydrophobic belt, where they interact with the alkyl chains of membrane integral lipids. In addition their polar amino acids close to the membrane surfaces interact with the polar head groups of the lipids and the aqueous phase. Therefore detergents are used at concentrations slightly above their critical micelle concentration to solubilise membrane proteins from their native membrane. The choice of the detergent is the most important factor apart from common requirements for the crystallisation of soluble proteins. The detergent micelle can mediate contacts between neighbouring detergent micelles and/or protein molecules and therefore contribute to the stability of the crystal lattice. Many results show that even small chemical differences in the detergent can cause essential differences in the crystallisation behaviour of these detergent-protein complexes. A shorter alkyl chain might cause significant changes in the crystallisation behaviour. Another approach is the addition of small amphiphilic molecules such as heptane-1,2,3-triol to reduce the size of detergent micelle (Timmins *et al.*, 1991; Iwata *et al.*, 2004). Instead of a smaller detergent micelle, one can try to increase the solubility of the membrane protein by binding of a specific monoclonal antibody (Iwata *et al.*, 1995).

Another problem is the availability of sufficient amounts of pure and homogenous protein. High yields from over-expression of membrane proteins, sometimes combined with refolding from inclusion bodies and/or reconstitution of chromophores and cofactors can complicate the

Methods

picture. As a result, many crystals of membrane proteins have been grown from natural sources.

PSIIcc crystals are grown using the traditional batch method. The advantage over the common vapour diffusion protocols is the higher initial concentration of the precipitant necessary for nucleation to avoid the growth of many small crystals. Secondly, during equilibration, the vapour pressure of the solution rises as the protein crystallises (Rayment, 2002). This batch method yields in larger crystals compared with all other crystallisation methods and opens the route for successful data collection recorded from only one crystal or for neutron diffraction experiments.

Due to small differences in the PSIIcc preparations, initial crystallisation trials were performed at different PEG-2000 concentrations to find the optimum PEG-2000 concentration to obtain crystals with suitable size for X-ray diffraction experiments. The precipitant solution, containing 3-10% PEG-2000, 100 mM PIPES pH 7.0, 5 mM CaCl₂, 0.03% β -DM, was 1:1 pre-mixed in PCR-tube with PSIIcc solution with an adjusted protein concentration of 4 mM Chl*a* in buffer C. With the aid of 10 μ L-pipette the solution was sucked up in a quartz-capillary (diameter 1.5 mm; Hirschmann Laborgeräte, Germany). The capillary was set upside down until the solution reached the middle of the capillary. Subsequently, both sides were sealed with wax and the capillaries were stored in an incubator at 20°C in dark.

2.4.2 Crystal mounting in capillaries

For neutron diffraction experiments, crystals were carefully transferred into quartz capillaries and the mother liquor was removed with paper tips to avoid slipping during data collection. On both sides the capillary was sealed with wax and mounted on goniometer head.

2.4.3 Cryoprotection and cooling

Data collection at cryogenic temperatures (100 K) has become routine in protein crystallography. It improves data quality as a result of reduced thermal motion (lower temperature factors), reduces radiation-induced damage of the protein in the crystal associated with loss of order and thus loss of intensity of high-order diffraction (dose/time-dependent as

well as resolution-dependent). A disadvantage of cryo-cooling is the possible disruption of the crystal lattice and in increased mosaicity of the crystals leading to blurred reflections. Other drawbacks are possible changes in unit cell constants (usually shrinking by several percent).

Since protein crystals have a solvent content between 40 and 60%, ice formation could destroy the protein crystal during freezing, as water transforms from its liquid to hexagonal ice phases near 273 K. By cooling rapidly below 150 K, water can be directly converted into amorphous glass phase. Warming up to 150-155 K, the glass phase changes to cubic ice and after further warming above 186 K leads to the formation of hexagonal ice (Kriminski *et al.*, 2002).

In practice, cooling rates available for large systems, like protein crystals, are too slow for glass transition of pure water that is a preferred state for cryo-crystallography, as it does not disrupt the integrity of the protein crystal. Crystalline ice formation will occur in most buffer systems even with the highest cooling rates and destroy the crystal lattice. Consequently cryoprotective additives (cryoprotectants) have to be used for cooling to a temperature range of 100-150 K. Further important parameters are: solvent content and solvent channel size of the crystal. The size of channels seems to influence the kinetics of solvent crystallisation. The larger the channels are, the quicker is the crystallisation process (Weik *et al.*, 2001).

Therefore the critical step in this approach is to prevent ice nucleation by formation of an amorphous glass upon cooling the crystal. The disordered water molecules (bulk solvent) and the water molecules on the protein surface (ordered solvent) have to be exchanged by a suitable cryoprotectant. Two groups of cryoprotectants could be roughly distinguished (Kriminski *et al.*, 2002). The first group is able to penetrate into the crystal and the second can not. Non-penetrating cryoprotectant like oils, sugars and high molecular weight PEG remove water covering the crystal surface and therefore suppress ice nucleation. Penetrating cryoprotectants like glycerol and other alcohols with short alkyl-chains as well as 2-methyl-2,4-pentandiol (MPD) and low molecular weight PEG prevent hexagonal ice formation within the crystal. There is no direct approach to find a cryoprotectant with satisfying quality.

Flash freezing protocols can be optimised by two different routes: (1) by increasing the cooling rates and/or (2) by decreasing the cooling rates required to achieve vitrification of internal and external solvent. The fastest cooling rates can be achieved by plunging the crystal

Methods

into liquid cryogen and by choosing conditions so that the liquid close to the crystal can not boil. Boiling can be avoided using cryogens having a huge difference between melting ($T = 86 \text{ K}$) and boiling temperature ($T = 231 \text{ K}$), as in the case of propane. Heat conduction through the crystals becomes more important for large crystals. For large crystals $>800 \mu\text{m}$ the use of liquid, fast cooling cryogen is in principle favourable (Kriminski *et al.*, 2002).

The lower helium temperature and heat transfer properties of helium provide a more efficient cooling regime. Although this does not directly affect the rate of radiation damage during data collection, it decreases the amount of cryoprotectant-agent required to obtain vitreous ice. As the temperature of a protein molecule is decreased below 293 K , the dynamic disorder decreases, until below around 150 K all that remains is the zero-point motion (theoretical limit at 0 K), the Debye-solid harmonics and the static disorder. Of these, only the Debye solid harmonics will decrease as the temperature is lowered further, and so the reduction of atomic B-factors, in going from 100 to 30 K is likely to be small.

2.5 Annealing

To improve the diffraction quality, flash cooled crystals can be warmed up and then cooled down again. Two different, principle protocols have been reported so far: in the first protocol flash frozen crystals are placed back for 3 min to the original cryoprotectant solution and then again flash-cooled. In the second, the cryo-stream is blocked until melting of the drop around the crystal becomes visible and then the crystal is flash-recooled (Harp *et al.*, 1999). By these protocols resolution could be improved and mosaicity diminished. The success of this method hardly depends on the size of the crystals and its solvent content. The higher the solvent content, the less likely annealing will be successful (Harp *et al.*, 1999). Annealing crystals reduces the lattice-spacing spread. The large stress applied during flash freezing can relax, resulting in local lattice reordering and domain growth. Probably it is the transformation of intra-crystalline water which is responsible for the well known effect of annealing.

2.6 Dehydration

Another post-crystallisation treatment is the dehydration of protein crystals to improve their diffraction quality. Crystal dehydration is applicable as reported by different groups applying altered methods. The most common protocols use dehydrating chemicals e.g. high molecular

PEG, dehydration by air or the influence of hygroscopic salts e.g. heavy atom solutions used for derivatisation (Heras *et al.*, 2003). The dehydration is associated with a different solvent content of the crystals, resulting in shorter cell constant. Recently, a device was developed to flash-freeze crystals under humidity control (Kiefersauer *et al.*, 2000). A systematic study showed that the dehydration is reversible and the crystal runs through different states with optima and minima and even lead to different space group (phase transition). Recently it was shown that dehydration could improve the resolution limit and diffraction quality of a chloride channel (Kuo *et al.*, 2003).

2.7 Data collection

When the crystals had been soaked according to the cryoprotection-protocol described in chapter 2.4.3, they were harvested with cryo-loops and latter were mounted on a goniometer head.

All X-ray reflection data were collected in the oscillation method, where the crystal is rotated in the X-ray beam so that all lattice points of the reciprocal lattice rotate through the Ewald sphere and give rise to reflections. This is done in small steps of 0.1° - 1.2° per diffraction image, that are chosen such that recorded reflexes are just separated and do not overlap. The size of the oscillation step strongly depends on the mosaicity spread, the resolution, the orientation of cell axis as well the divergence of the beam. A complete dataset consists of a consecutive series of these images covering a total rotation of 30° to 360° , depending on the symmetry (space group) of the crystal.

2.8 Data collection strategy

Besides the beamline optics, additional difficulties are connected with crystal properties. Crystals of large protein assemblies diffract weakly due to their large unit cell and produce closely spaced reflections (Ban *et al.*, 2000; Cramer *et al.*, 2001). Overlapping of these Bragg reflections present additional problems, and the larger reflection (spot) size, the more severe the overlap.

Large area detectors are crucial for efficient data collection from crystals with large unit cells to achieve optimal spot separation. If the crystal has one large unit cell dimension, spot overlap is reduced by mounting the crystal in this way that the long axis of the unit cell is

Methods

parallel to the spindle axis. If the longest axis does not coincide with the axis of highest symmetry, a larger wedge of data needs to be collected over a wider angular range to obtain a complete dataset. Crystals with large unit cells have huge resolution-dependent differences in diffracted intensity, often too large to be recorded with a single data collection pass when using CCD detectors having a slow dynamic range. Even with the use of the strongest X-ray source (ESRF and SLS synchrotrons) it might be necessary to use long exposure times to collect data at the highest possible resolution and consequently the amount of unique data from one crystal is limited due to radiation damage. This means that data from different crystals have to be collected and merged to obtain a complete dataset; this is advisable only after potential problems with non-isomorphism and crystal quality have been resolved.

The unit cell parameters of macromolecular crystals often show significant variation and this can cause problems when merging data from several crystals together. This is especially the case for large complexes located in biological membranes. Due to the hydrophobic character of such proteins the number of charged residues, which might be involved in crystal contacts, is limited. Furthermore, proper separation of high-resolution reflections often requires low mosaicity. Since the quality of protein crystals often depends on the preparation and purity of a protein, it means in practice that only a small fraction of a population of crystals may be suitable for data collection at high resolution.

2.9 STRATEGY implemented in MOSFLM

Once the crystal symmetry and orientation has been determined with the program MOSFLM, it is possible to work out which rotation range is needed to collect an "optimum" dataset. The STRATEGY option allows a data collection strategy to be worked out semi automatically. It requires the parameters: the crystal symmetry (space group), orientation, distance, wavelength and detector type. The rotation range required for a complete dataset is determined from the crystal symmetry and orientation. The reflection list is compared to a list of all unique reflections for this space group and the completeness and multiplicity are calculated both as a function of rotation and resolution. If the total rotation angle (φ) in which data will be collected is specified and the number of discontinuous segments to be used, the program will determine start and end φ values for each segment that will give the highest possible completeness. If the crystal has to be translated during the data collection, it is possible to determine the best segments to collect data to maximum completeness. Furthermore it is

possible to compute the optimum oscillation angle - depending on the crystal orientation - in order to avoid overlapping reflections.

2.10 Data reduction with DENZO and XDISPLAYF

The first step in data reduction is to determine the precise crystal and detector orientation parameters for a single image in the collected series including the intrinsic crystal parameters (lattice type and unit cell dimensions). The crystal orientation parameters describe the orientation of the crystal lattice with respect to the coordinate system of the diffractometer. The detector and X-ray parameters include the wavelength, the crystal to detector distance, the precise coordinates of the direct beam, the detector missetting angles, and the internal scanner alignment parameters, such as the non-orthogonality of the scanner head motions or the scale of the y-pixels to the x-pixels of the scanned image. The use of a flexible, weighted, profile fitting algorithm for measuring the intensity of the spots provides highly accurate data processing. Crystal and detector parameters can be refined in any order and combination. Partially recorded reflections are used in the positional refinement of crystal orientations. The strategy in refinement is to start at low resolution, adjusting just a few parameters, and then to extend the refinement to a higher resolution, where more parameters can be fitted. The accurate prediction of spot positions is necessary to achieve a precise integration of Bragg reflections. The profile fitting increases the precision (decreases the statistical error) of the measurement, but it may introduce an error due to lack of accuracy of the predicted profiles. Therefore the profile of each reflection is averaged with profiles of neighbouring reflections, because they should have the similar profiles (Otwinowski and Minor, 1996).

XdisplayF allows to visualise the data in their original form as well as to observe the progress of data reduction. Displaying the raw data has the advantage to follow by eye the data reduction process and allows efficient recognition of what is correct or wrong. Finally, all of the images collected in the series using the parameters determined from indexing one frame, are processed. The user has full control over which parameters are refined in every cycle.

2.11 Data scaling with SCALEPACK

SCALEPACK accurately tracks and sums up partially recorded reflections over all frames of a dataset. The scaling and merging of different datasets as well as the global refinement of

Methods

crystal parameters (post-refinement) is performed by SCALEPACK. Unlike other procedures, in SCALEPACK the estimated error of the measurement is enlarged by a fraction of the expected, rather than the observed, intensity. This algorithm reduces the bias towards reflections with an integrated intensity below the average. Due to the correlation between crystal and detector parameters the values of the unit cell parameters refined from a single image may be quite imprecise. At the end of the data reduction process one would wish to get precise unit cell values. This is done in a procedure referred to as a global refinement or post-refinement. The implementation of this method in SCALEPACK allows for separate refinement of the orientation of each image, but with the same unit cell value for the whole dataset. In each batch of data different unit cell parameters may be poorly determined. Therefore a global refinement of e.g. crystal parameters is performed in the last steps. However, in a typical dataset there are enough orientations to determine precisely all unit cell lengths and angles. Global refinement is also more precise than the processing of a single image in the determination of crystal mosaicity and the orientation of each image (Otwinowski and Minor, 1996).

2.12 Single-wavelength anomalous scattering

When the energy of X-rays approaches the energy of an electronic transition from a bound atomic orbital, a resonance condition is established that amplifies the acceleration and perturbs the scattering. For standard diffraction experiments with energies in the range of 10000 eV equal to about $\lambda = 1 \text{ \AA}$, there are no electronic transitions for the elements building up naturally occurring amino acids.

Therefore only atoms with energies at longer wavelength show electronic transitions, e.g. bound metal ions in the case of proteins. The distinctive scattering from the anomalous centres can be distinguished by appropriate diffraction measurements and used to locate the positions of the anomalous scatters.

There are difficulties associated with the collection of diffraction data at longer wavelengths, mainly the increased absorption of X-rays at longer wavelengths, air scatter, the larger scattering angles resulting in a lower high resolution limit. In frame of this PhD thesis, single-wavelength anomalous scattering (SAS) was applied to locate the positions of the Mn and Ca cations of the catalytically active Mn-Ca-cluster.

2.13 Refinement and model building

2.13.1 Refinement

Refinement of the protein model was performed using the program package CNS (version 1.0; Brünger *et al.*, 1998)). The flow diagram in Fig. 2.1 illustrates the general refinement procedure applied to refine the PSIIcc structure.

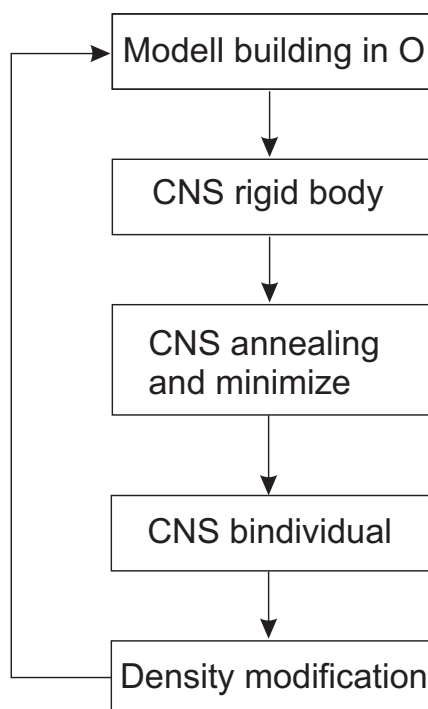


Fig. 2.1: Flow-diagram of consequent refinement cycles. Each refinement step is described in detail in the following paragraphs.

Rigid body refinement

The rigid-body method minimises the three rotational and three translational degrees of freedom for each specified group of atoms; e.g. the groups of atoms are treated as rigid bodies. The complete energy function is used for refinement. The rotational parameters are the three Eulerian angles for a rotation around the geometric centres of the rigid group. Rigid groups are e.g. single domains and/or subunits or secondary structure elements for instance a single TMH.

Methods

Simulated annealing

CNS (Brünger *et al.*, 1998) is dealing with the general problem of macromolecular crystallography that structure optimisation may suffer from multiple and false energy minima, arising from the high dimensionality of the parameter space. Simulated annealing - an optimisation protocol - is performed, when the phase information is rather inaccurate due to the resolution limit and phasing power e.g. of heavy atom derivatives. Unlike gradient descent methods, simulated annealing can overcome barriers between energy minima, and thus can explore a greater volume of the parameter space to find "deeper" and correct energy minima. The deepest energy minimum corresponds to the conformation of the molecule that best fits the diffraction data and maintains good covalent geometry and non-covalent interactions.

The efficiency of simulated annealing depends on the choice of the annealing schedule. Simulated annealing refinement is most useful when the initial model is relatively crude, with errors in backbone conformations or in side chain conformations.

Simulated annealing makes use of molecular dynamics, e.g. simulation of atom movements at certain temperatures. The standard schedule for molecular dynamics is the following: heating to 2500 K within 2-4 ps, slow ramp to 300 K (100 steps of 1 ps), constant temperature dynamics at 300 K (1000 steps of 1 ps) and finally positional refinement (Brünger *et al.*, 1990).

Minimize (restrained positioned refinement)

In an iterative process the co-ordinates were refined according to the protocol "*minimize*". A critical aspect of the refinement procedure is that the number of refined parameters (for each protein atom the position x, y, z) must not exceed the number of observations, that is the number of unique reflections collected from a crystal. The number of degree of freedom can be reduced indirectly by restraining the stereochemistry to values derived from small molecule crystallography (Engh and Huber, 1991). The outcome of such refinement is an atomic model of the protein with optimised stereochemistry and a minimised overall energy which best fits the collected X-ray data.

Bindividual (refinement of individual B-factors)

The protocol "bindividual" was used to refine the atomic B-factors followed by a second "minimize" procedure. For the first refinement steps it is possible to constrain B-factors in a certain range.

Refinement target

Refinement target was the maximum likelihood target (energy of bond lengths, angles, dihedral angles, improper line-to-plane angle, van der Waals contacts and Coulomb interactions), using structure amplitudes. The resolution range used for refinement was 12-3.2 Å. The low resolution limit was restricted as no bulk solvent correction was applied. The observed data cut off criteria (σ), applied to structure amplitudes was set to $\sigma = 2.0$.

Topology and parameter files (for the refinement of cofactors)

The parameter and topology libraries for the cofactors were prepared based on data of the HIC-Up server (Kleywegt and Jones, 1998). The topology and parameter files for each cofactor were manually manipulated.

2.13.2 R-factor

The progress of the refinement is judged by the crystallographic *R*-factor (R_{cryst}), which describes the deviation between the measured and calculated structure amplitudes. A reference factor, termed R_{free} , was introduced by Brünger, where a subset of the reflections (usually 5-10% of the experimental data) are excluded from refinement (test set, Brünger *et al.*, 1998). The R_{free} is then calculated as for R_{cryst} , but only against this test set.

2.13.3 Density modification

When an electron density map is not good enough to show the correct structure, then it could clearly show an incorrect structure. The fact that an experienced worker can distinguish an interpretable electron density map from an un-interpretable map suggests that it may be possible in some manner to quantify the quality of an electron density map. If we can modify the electron density map in such a manner to increase this quantity then it may be possible to

Methods

convert an un-interpretable electron density map into an interpretable one. Following methods are used to improve the electron density:

Solvent flattening

Macromolecular crystals contain a large part of disordered solvent whose contribution to low-resolution reflections is very important; an atomic macromolecular model without the contribution of the bulk solvent cannot correctly reproduce these diffraction data. The program DM (Cowtan, 1999; Cowtan, 2000) is an automated procedure to apply real space constraints based on known features of a protein electron density map in order to improve the approximate phasing experimentally obtained. Solvent flattening exploits the fact that the solvent region - in the case of membrane proteins the detergent is included - of an electron density is constant (flat) at medium resolution owing to the high thermal motion of atoms and disorder of the solvent in contrast to the ordered protein. The existence of a flat solvent region put strong constraints on the structure factor phases. The molecular envelope that separates protein from solvent is located by an automated convolution procedure first proposed by (Wang, 1985).

Histogram matching

Most proteins have fairly similar proportions of atomic types, distributed throughout the crystal unit cell according to known constraints of atomic spacing. As a result the distribution of protein density values in the protein region is fairly similar from protein to protein. This information is applied through the process of histogram matching (Zhang and Main, 1990). Histogram matching employs the predictable atomic composition of biological macromolecules to predict the histogram of density values in the protein region. The current electron density map can then be systematically modified to bring it into consistency with the predicted histogram. This technique is complementary to solvent flattening, since solvent flattening operates on the whole of the solvent region and histogram matching operates on the whole of the protein region. The known protein density histogram is a weaker constraint on the electron density than solvent flatness, but the volume involved is usually larger, and histogram matching has more power for phase extension than solvent flattening. The combination of solvent flattening and histogram matching usually converges in 10-20 cycles.

Non-crystallographic symmetry averaging

To increase the signal-to-noise ratio, crystallographically independent images of the same object can be averaged. In protein crystallography, often multiple copies of the protein are present in the asymmetric unit, related by non-crystallographic symmetry (NCS; Muirhead *et al.*, 1967). In contrast to crystallographic symmetry, NCS is local. For instance two monomers of a dimer could be related by a 2-fold rotation axis, to superimpose both monomers.

First the correct NCS operator has to be found, based on information of multiple copies of the same molecule, subunit or domain. Because of the local character of NCS that applies only to a limited number of related objects, information about the shape of the objects whose electron density must be averaged is necessary. These are known as the envelope. Without this information, the averaging procedure could average density belonging to objects to which the operator does not apply. A wrongly assigned envelope leads to not well defined electron density of the periphery of the envelope and even parts of the protein could be truncated. A successful NCS averaging results in electron density maps of higher quality.

2.13.4 Electron density maps and model building

Several different procedures were applied to calculate electron density maps. In the early states of refinement, different kind of cross-validated F_0-F_C sigmaA-weighted maps were employed for model-building: F_0-F_C sigmaA-weighted maps calculated with phases obtained after refinement using the CNS-annealing protocol. In addition F_0-F_C sigmaA-weighted maps were calculated after density modification (see 2.13.3). In the later refinement process $2F_0-F_C$ sigmaA-weighted maps were calculated to verify the model-building procedure.

All steps of model building were performed with the program O (Jones *et al.*, 1991). The complete model was manually built, as at the resolution of 3.2 Å all algorithms for automated model building failed.

2.14 Analysis of the crystal structure and preparation of figures

The stereochemistry of the obtained structures of PSII_{cc} was checked with PROCHECK (Hooft *et al.*, 1996). Sequence alignments were performed using CLUSTALW (Thompson *et*

Methods

al., 1994) and displayed with SEAVIEW (Galtier *et al.*, 1996). Comparison of the structure to known structures was done via the DALI server (Holm and Sander, 1993). Superpositions of protein structures was performed using ALIGN (Cohen, 1997). Hydrogen-bonds (H-bond) were assigned with the program HBPLUS (McDonald and Thornton, 1994). Images of the PSIIcc were prepared with DINO (Visualizing Structural Biology (2002) <http://www.dino3d.org>) and rendered with POVray (<http://www.povray.org>).

2.15 Neutron scattering

Whereas single detergent molecules may be found in protein crystal structures, the bulk detergent can not be determined routinely using X-ray diffraction, owing on the fluidity of the detergent molecules and their low X-ray contrast. Knowledge of the detergent distribution provides insight how the detergent is involved in the crystallisation process and might be a useful approach to clarify how the detergent contributes to crystal packing by detergent-detergent and/or protein-detergent-interactions. The contradiction between the importance of detergent in the crystallisation of membrane proteins and the lack of information on the structure and distribution of the detergent phase in protein crystals may be addressed using contrast-matched neutron crystallography (Timmins *et al.*, 1994).

Neutron small angle scattering has been widely used to analyse biological structures. Hydrogen and deuterium have different scattering power as the neutron scattering length densities have opposite signs. Systematic molecular types have characteristic hydrogen densities, and these can be thought of as continuous over volume at low resolution ($> 16 \text{ \AA}$). The ratio of ($^2\text{H}_2\text{O}$)/($^2\text{H}_2\text{O} + ^1\text{H}_2\text{O}$) can be manipulated between 0% and 100%, which includes the hydrogen atoms of proteins, lipids and their head groups as well as the detergent head groups. In a neutron scattering experiment on a protein crystal, systematic volumes may therefore be matched out by an appropriate ($^2\text{H}_2\text{O}$)/($^2\text{H}_2\text{O} + ^1\text{H}_2\text{O}$) contrast (Timmins *et al.*, 1994). The contrast of bulk solvent region of a crystal is adjusted by soaking with a mother liquor analogue containing a defined fraction of ($^2\text{H}_2\text{O}$)/($^2\text{H}_2\text{O} + ^1\text{H}_2\text{O}$), termed contrast. The scattering length density varies with different contrasts for the detergent head groups, tails and protein volumes. In order to define and validate this relationship four or more dataset are generally used. Finally initial phases derived from X-ray experiments are used to estimate phases for each contrasted dataset.