X-ray Crystallographic Study on the Mechanisms of Bacteriorhodopsin and the Sensory Rhodopsin/Transducer Complex

by **Rouslan Moukhametzianov**

M. Sc., Moscow Institute of Physics and Technology

for

Obtaining the PhD degree at

Mathematics and Natural Sciences Faculty of

Heinrich Heine University Düsseldorf

Supervisor:	Georg Büldt, Prof. Ph.D., Physics
Examining Board:	Jorg Labahn, Ph.D., Chemistry Georg Pretzler, Prof. Ph.D., Physics Dieter Willbold, Prof. Ph.D., Biology

February 2006

Contents

Abstract

Microbial rhodopsins belong to a family of seven-helical transmembrane retinal proteins which are found in Bacteria, Archaea and Eukaryota. They are considered to be the archetypes for ion transport and signal transduction, which use for these distinct functions a common structural design. The ion pumps Bacteriorhodopsin (BR) and Halorhodopsin (HR) operate as energy converters, whereas the photoreceptors Sensory rhodopsin I (SRI) and II (SRII) operate as light sensors providing the initial signal which via associated receptor-specific transducers (HtrI and HtrII) activate a two-component signalling cascade that moves the cell in response to light.

The archaeal rhodopsins are the best understood proteins among seven-helical receptors with respect to structural information from X-ray crystallography. High resolution structures of Sensory rhodopsin II from *Natronobacterium pharaonis* and Bacteriorhodopsin from *Halobacterium salinarum* as well as Halorhodopsin have already been obtained. The X-ray structure of the complex between *N.Pharaonis* SRII (pSRII) and HtrII at 1.94 Å resolution was reported.

In this thesis the structure of the ground state, the early K state and the signalling late M state of Sensory rhodopsin II and its cognate transducer as well as the structure of the ground state and the late M state of the Bacteriorhodopsin were investigated by means of X-ray crystallography. Crystals were grown in the lipidic cubic phase which provided data to a resolution of 1.9 Å for the ground state, 2.0 Å for K state and 2.2 Å for the late M state of the Sensory rhodopsin II/transducer complex and 1.35 Å for the ground state, 1.5 Å for the late M state of Bacteriorhodopsin. The occupancies of the intermediate states trapped in the crystals at cryo temperatures were estimated from the crystallographic analysis. The structure solution based on molecular replacement yielded atomic pictures of ground, K and M state of the pSRII/transducer complex and ground and M state of bacteriorhodopsin. The refinement scheme using simulated annealing for models consisting of two conformations (one accounting for the ground state and the other for the intermediate state) with corresponding occupancies were assessed and used for structure solution. Additionally it was investigated if experimental phases obtained by isomorphous replacement and anomalous scattering would help to determine the transducer structure including linker domain. Different models of the crystal were investigated to extract structural data for the complete fold of the transducer. Results provide insights in signal transfer from pSRII to the transducer in the membrane part of the protein complex and reveal the details of the evolvement of the proton translocation channel in BR.

The observed structural changes allow to propose a mechanism for the light induced activation of the complex: Upon light excitation retinal isomerization leads in K state to a rearrangement of a water cluster that partially disconnects two helices of pSRII. In the transition to late M the changes in the hydrogen bond network proceed further. The signalling state is established by tertiary structural changes induced by the new hydrogen bond pattern and the changed charge distribution. The two partially decoupled subdomains of the receptor show a relative displacement that is most significant between helices F and G which form the interface to TM2 of the transducer. The transducer responses to the receptor activation by a clockwise rotation of about 15° of helix TM2 and a displacement of this helix by 0.9 Å at the cytoplasmic surface.

The late M state structure of the wild type bacteriorhodopsin extends the knowledge of this important intermediate of the photocycle reported previously. The achieved resolution of 1.5 Å and the low twinning of the crystal enable a better definition of the model. Though this structure is very similar to the reported one it still provides new information particularly concerning the water molecule chain in the cytoplasmic part of the channel between Asp96 and Schiff base.

The revealed structures allow drawing the parallels between ion transport and sensory signalling by this family of proteins. Important structural differences and similarities related to the function of these proteins are underlined.

1 Introduction

Besides providing essential energy for life on earth, light provides signals and images that enable organisms to manage successfully in their environments. Light controls movement, growth, differentiation and development, circadian timing, and gene-expression responses in diverse organisms from primitive single cell microbes through higher plants and animals, including man. Organisms sense light intensity, color, direction, duration, and the pattern of light in the environment. All these functions are carried out by the complex organization of different molecules present in the cell. At the top of any cascade of reactions happening in response to light are photoreceptors.

Photoreceptors use the proteins which can absorb light and upon that undergo structural changes of their conformations which will be transmitted to the cell. These proteins utilize for their function different chromophors. Several groups of the proteins, selected in accordance with the chromophor they use, are known and well described. Among them is a family of the retinylidene proteins or rhodopsins (Spudich et al. 2000 [1]).

Over three hundred photochemically reactive proteins that use vitamin-A aldehyde (retinal) as their chromophor (rhodopsins) have been described in both prokaryotic and eukaryotic organisms. The common feature of these proteins is that they contain seven membrane-embedded α-helices that form an internal pocket in which the retinal is bound. Primary sequence alignment splits the rhodopsins into two clearly distinct families. One family (type I) consists of the archaeal-type rhodopsins first observed in the archaeon *Halobacterium salinarum*, a halophilic prokaryote, and now also found in eukaryotic microbes (Robb et al. 1995 [2]). Type I rhodopsins function as light-driven ion transporters (Bacteriorhodopsin and Halorhodopsin), phototaxis receptors (Sensory rhodopsin I and Sensory rhodopsin II), or have a yet undiscovered function (e.g. fungal rhodopsins). The type II rhodopsins family consists of the photosensitive receptor proteins in animal eyes, including human rod and cone visual pigments, receptor proteins in the pineal gland, hypothalamus and other tissues of the lower vertebrates. All type II rhodopsins so far reported belong to higher eukaryotes.

Rhodopsins are among few membrane proteins for which structural information is available from crystallography. Atomic resolution structures of Bacteriorhodopsin have resulted from electron and X-ray crystallography (Grigorieff et al. 1996, [3] Pebay-Peyroula et al. 1997 [4], Sass et al. 2000 [5], Luecke et al. 1998 [6]) The 1.55 Å structure reported (Luecke et al. 1999 [7]) is one of the most highly resolved of any membrane protein. An atomic resolution structure (1.8 Å) of Halorhodopsin has also been reported (Kolbe et al. 2000 [8]). Sensory rhodopsin I and II have been studied by electron crystallography (Kunyi et al. 2001 [9]). For Sensory rhodopsin II the structure was recently solved at 2.1 Å resolution (Pebay-Peyroula et al. 2001 [10]). Electron crystallography has also revealed the structures of the three rhodopsins of type II: bovine, frog and squid visual pigments (Shertler et al. 1995 [11], Davies et al. 1996 [12], Krebs et al 1998 [13]). Recently a 2.8 Å structure of the bovine rod rhodopsin was obtained from X-ray diffraction data (Palczewski et al. 2000 [14]).

The present work introduces new insights in the function and structural mechanism of Bacteriorhodopsin and Sensory rhodopsin II/Transducer complex. The two solved crystal structures of these proteins present a vast amount of the new information which reveals important details of functioning of molecular machines.

1.1 Chemotaxis and phototaxis in *H. salinarum*

Halobacterium salinarum cells swim back and forth by switching the rotational sense of their flagellar bundle, which is composed of 5-10 flagellar filaments (Alam et al. 1984 [15]). Without exogenous stimulus, spontaneous reversals of the swimming direction occur randomly on the time scale of tens of seconds. Stimulation with UV or blue light shortens the average duration of the current run by inducing the next switching event earlier, whereas an increase of orange light intensity prolongs the straight swimming period by transiently suppressing spontaneous motor switching.

If the cell moves in the direction of the positive gradient of the positive stimulus the two-component chain will send the signals to the flagellar motor to decrease the probability of the motor switch and the time during which the cell moves in this favorable direction is prolonged. The opposite situation happens when the cell moves in the direction of the negative gradient of the positive stimulus. The probability of the switch is increased and the direction of the motion will be switched to the other direction.

Excited cells adapt to the incident light intensity and resume their spontaneous behavior within the time period of one motor-switching event (Hildebrand et al. 1985 [16]). The interplay of excitation and adaptation accumulates the cells at those sites of their habitat that provide the best environmental conditions.

The stimulus implies also a gradients of the chemicals present in the cell environment. The cell respond to various chemicals, e.g. arginine, leucine, or dipeptides, as attractants and also to phenol, indole or benzoate as repellents. This is known as chemotaxis in comparison to the phototaxis which is the reaction to the light stimulus (Alam et al. 1989 [17]).

Both responses are mediated by the two-component signal transduction chain which uses phosphotransfer schemes involving two conserved components, a histidine protein kinase and a response regulator protein (Falke et al. 1997 [18]). The histidine protein kinase, which is regulated by environmental stimuli, autophosphorylates at a histidine residue, creating a high-energy phosphoryl group that is subsequently transferred to an aspartate residue in the response regulator protein. Phosphorylation induces a conformational change in the regulatory domain that results in activation of an associated domain that effects the response.

The basic model pathway consists of the sensory rhodopsin-transducer complex and the autophosphorylating histidine kinase CheA, which phosphorylates the response regulators CheY and CheB by catalytic transfer of the phosphoryl group, the adaptor protein CheW and protein CheR. It is shown schematically in Figure 1. CheY-P finally acts as the switching signal. Low concentration of CheY-P prolongs the swimming period between reversals. CheB functions as a methylesterase of methylated Asp or Glu residues that flank the signalling domain. An adapter molecule CheW is attached to the cytoplasmic domain of the transducer. CheR is the constitutively active methyltransferase, that re-methylates the carboxylates. Thereby, depending on the input of attractant and repellent impulses the methylation level is altered. This methylation/demethylation reactions are involved in the adaptation of the bacteria to constant stimuli. Upon methylation of the carboxylic residues of the transducer at methylation sites, the activity of histidine protein kinase CheA is increased. This in turn activates CheB due to the transfer of the phosphoryl group and therefore reduces the methylation of the transducer. When the activity of CheB is low because of a low activity of CheA, the constant level of the methyltransferase CheR will increase the level of the methylation and activate the histidine protein kinase CheA. Therefore the system always regulates the level of methylation to a certain level corresponding to the current stimulus. Furthermore the level of methylation is adjusted to bring the frequency of the flagellar motor switch to that present without the stimulus when the stimulus is constantly distributed in the space.

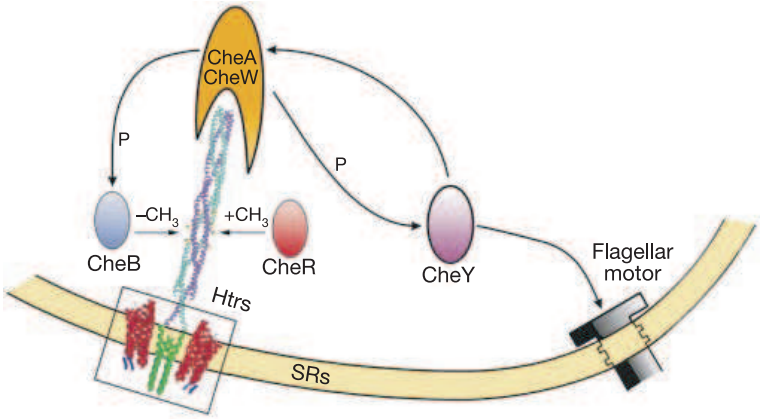

Figure 1: Two component signalling cascade. The activation of the transducer by its receptor leads to the conformational change that propogates to the coiled-coil cytoplasmic domain. The next step in signalling cascade involves CheA, CheW, CheY, CheB and CheR (from Gordeliy et al. [47])

The methylation of the transducer provides the bacteria with simple chemical or photo-memory, which is used to check whether the current direction of motion is favorable or not. The level of methylation is high when the concentration of the positive stimulus has been high in the recent past of the bacteria and low when the stimulus has been low. When the cell moves it compares the current environment conditions to those of the recent past which was recorded in the level of transducer methylation. In this way the correctness of the current direction is assessed.

It should be noted that because of the same components of the signal transduction chain in spite of the difference of the receptor and its cognate transducer there is united reaction for the photo- and chemo signals. The overall chemo- and phototactic signal integration occurs in *H.salinarum*.

1.2 Archaerhodopsins

The four archaeal rhodopsins and the three functions, proton transport, chloride transport, and phototaxis signaling, appear to account for retinal pigmentation and retinal-dependent functions in *H. salinarum*. Over 30 archaeal rhodopsins have been described and they all correspond in absorption spectra and functions to BR, HR, SRI, or SRII. The functions of proton pumping, chloride pumping, and sensory signaling are distinctly different; nevertheless the modifications of the same versatile phototransduction machinery are responsible for these diverse consequences of photon absorption (Oesterhelt et al. 1998 [19]).

When oxygen and respiratory substrates are plentiful, *H. salinarum* cells would be expected to avoid sunlight and potential photooxidative damage. To accomplish this, the cells synthesize the repellent receptor sensory rhodopsin SRII (also known as phoborhodopsin) as their only rhodopsin. SRII absorbs blue-green light in the energy peak of the solar spectrum at the Earth's surface. Hence, its wavelength sensitivity is tuned strategically to be maximally effective for seeking the dark.

A drop in oxygen concentration suppresses SRII production and induces the synthesis of BR and HR, enabling orange light absorbed by these pumps to be used as an energy source. Like the respiratory chain, BR pumps protons out of the cell, directly contributing to the inwardly

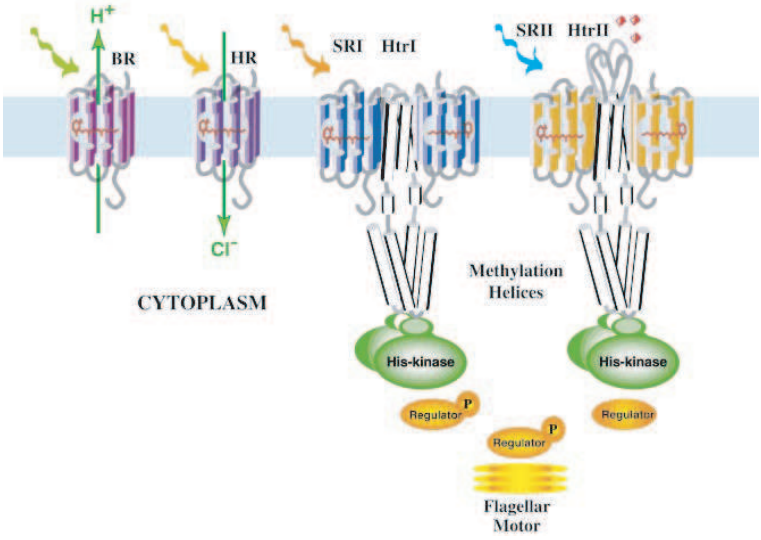

Figure 2: The four archaeal rhodopsins in *H.Salinarum*. The transport rhodopsins BR (a proton pump) and HR (a chloride pump) are shown in addition to the sensory rhodopsins SRI and SRII with components in their signal transduction chain (from Spudich et al. [1])

directed proton motive force needed for ATP synthesis, active transport, and motility. HR is an inwardly directed pump, transporting chloride into the cell. Like cation ejection, anion uptake hyperpolarizes the membrane positive outside. Therefore, the inward transport of chloride contributes to the membrane potential component of the proton motive force without loss of cytoplasmic protons. This transport helps to maintain pH homeostasis by avoiding cytoplasmic alkalization. Along with BR and HR, the production of SRI is induced. SRI mediates attractant responses to orange light, facilitating migration into illuminated regions where the ion pumps will be maximally activated. SRI is endowed with a second signaling activity to ensure it will not perilously guide the cells into higher energy light. Therefore, the signals from SRI attract the cells into a region containing orange light only if this region is relatively free of near-UV photons. When back in a rich aerobic environment, the *H. salinarum* cells switch off BR, HR, and SRI synthesis and switch on SRII production.

1.2.1 Bacteriorhodopsin (BR)

Bacteriorhodopsin (BR) was discovered in *H.salinarum* in 1966 (Stoeckenius et al. 1971 [20]). The presence of the retinal and its secondary structure which resembles that of the rhodopsin suggested its name. BR is a membrane protein of $26.7kDa$. The chromophor in BR is all-*trans* retinal that is bound via a Schiff base linkage to a Lys residue on helix G. The maximum of the light absorption by the retinal is at 569 nm in the visible range of spectra which accounts for the characteristic purple color of *H.salinarum*. BR is aggregated in the cytoplasm membrane in a two-dimensional lattice, the purple membrane. In this crystal structure the ratio of the protein molecules to the lipids is 1:10 with no other proteins present. (Racker et al. 1974 [21]). The protein in the purple membrane has an exceptional stability to the high temperatures (till $+55^0C$) and low ionic strengthes (it can function in the absence of salt in solution). BR is able to transport the protons through the membrane against a potential as high as $U = -280mV$.

The absorption of a photon by BR causes a cycle of structural changes in the protein. Dur-

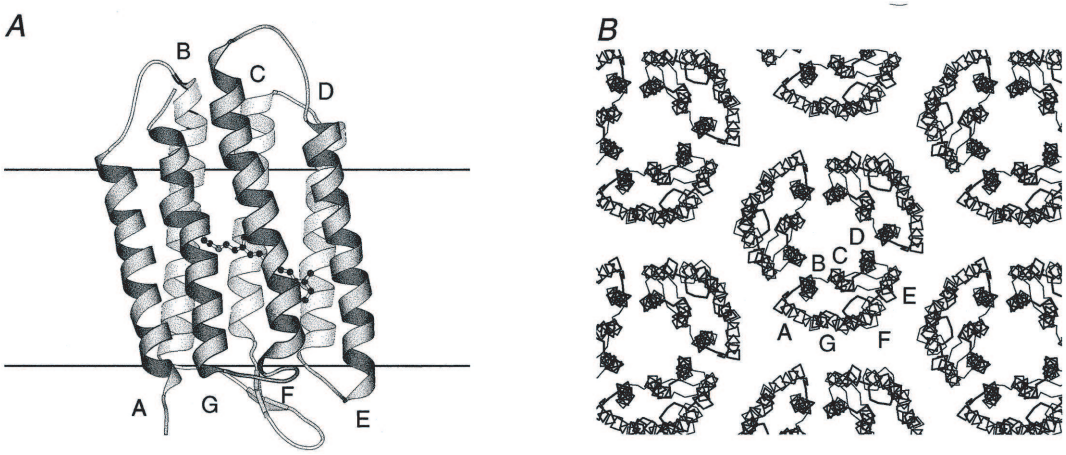

Figure 3: Bacteriorhodopsin and the purple membrane. (A) Side view of the BR monomer. (B) cytoplasmic view of the purple membrane. From Crebs et al. BBA 2000

ing its functioning the protein undergoes different states which can be distinguished by their absorption spectra in the visual range. The kinetic studies of the spectroscopic changes of the photocycle revealed the following intermediates of the photocycle: $BR_{570} - K_{590} - L_{550} - M_{410} - N_{520} - O_{640} - BR_{570}$ (the subscript indicates the wavelength of maximum light absorption of the respective intermediate).

These states are spectroscopic states, which reflect the changes in the immediate vicinity of the retinal. Structural changes far away from the retinal pocket do not produce changes in the absorption spectra. This can be observed in the photocycle for the M state that actually presents two protein states, M_1 and M_2. Structural changes occur in the tertiary structure between M_1 and M_2 states which corresponds to a silent transition in the absorption spectrum, that means the absorption spectra of M_1 and M_2 states do not show a traceable difference.

The starting point of the photocycle is the isomerization of the retinal. In the darkness the retinal of BR exists in two conformations. During several minutes of darkness the equilibrium between the all-*trans*, 15-*anti* and 13-*cis*, 15-*syn* establishes in the ratio 1:1. Conformations are different at the $C_{13} = C_{14}$ and $C_{15} = N$ bonds, see Figure 4A (Stoeckenius et al. 1980 [22]). After intensive illumination of the dark adapted state for several seconds, BR molecules are light adapted with an all-*trans* retinal.

During all the states of the photocycle from K to O the retinal is in the 13-*cis* conformation, and during the O to G state transition the relaxation to the all-*trans* state happens. The photocycle of the Bacteriorhodopsin together with the time constants is shown in Figure 4B.

In 1975 Henderson et al. [23] published the first three dimensional low resolution model of BR, which was obtained from electron microscopy of purple membranes. This model revealed the secondary structure of BR - 7 α-helices spanning the membrane. The development of a new method for crystallization of membrane proteins in lipidic cubic phase (Nollert et al. 2001 [24]) resulted in the high resolution structure of BR obtained by X-ray crystallography.

Three-dimensional structures of the BR at increasing resolutions reveal the positions of a number of charged residues and water molecules. They form the proton translocation channel that extends across the membrane through the center of the heptahelical bundle ([4], [6], [7]). Light induced structural changes cause these residues together with water molecules to pass the proton from one group to another in a ordered sequence of events, which have been characterized through detailed spectroscopic and mutational studies. These structural rearrangements of the

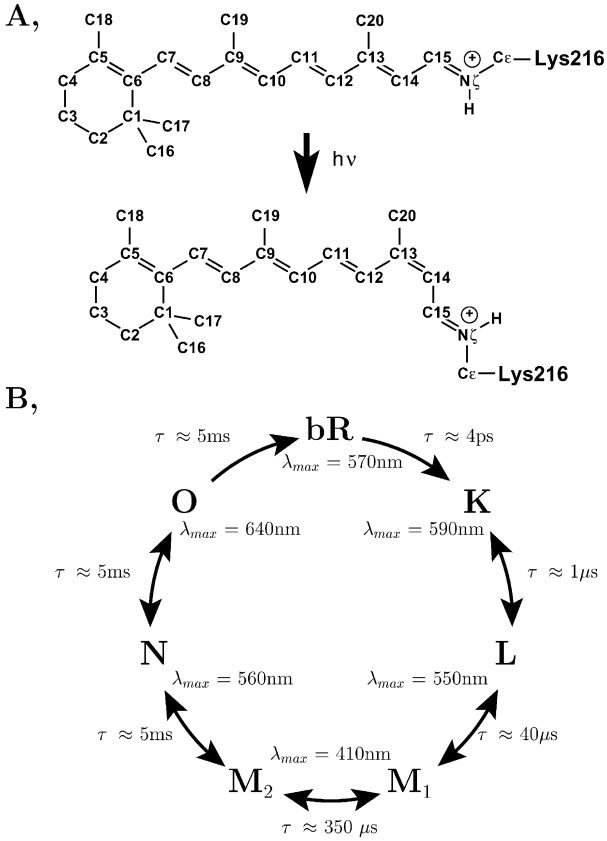

Figure 4: Schematic representation of light-induced retinal isomerization and the BR photocycle. (A) Following absorption of the photon the all-*trans* retinal is isomerized to 13-*cis* retinal configuration. (B) The spectral intermediates, their absorption maxima and their lifetimes at room temperature are shown. (From Neutze et al. BBA 2002)

protein serve both to define a pathway for proton exchange between charged groups for each step of this process and to manipulate the pK_a's of the key residues, so that the correct proton donor/acceptor relationship is achieved at the required point of the photocycle (Edman et al. 1999 [25], Royant et al. 2000 [26], Sass et al. 2000 [5], Schobert et al. 2002 [27], Schobert et al. 2003 [28]).

The reaction sequence with some details of this molecular machine is summarized in Figure 6. The red-shifted K-intermediate builds up within a few picoseconds at room temperature, followed by formation of the blue-shifted L-intermediate on a time scale of microseconds. The primary proton transfer event from the Schiff base to Asp85 defines the L-to-M spectral transition, which induces a large blue shift of the absorption band. Under physiological conditions a proton is released to the extracellular medium on a similar time scale. The proton release groups are believed to be formed by Glu194, Glu204 and a cluster of water molecules. The further spectrally silent transition from an early M (M_1) to a late M (M_2) is associated with large structural changes. These changes ensure the switch in accessibility of the retinal, so that the Schiff base can subsequently be reprotonated from Asp96 on the cytoplasmic side, corresponding spectrally to the M_2 to N transition. Asp96, in turn, is reprotonated from the cytoplasmic medium, and the absorption peak is further red-shifted as the retinal thermally reisomerizes to recover the all-*trans* configuration, associated with the O to G state transition. Finally the ground state is recovered

when a proton is transferred from Asp85 to the putative release group on the extracellular side via Arg82.

The detailed mechanism of proton pumping is not yet understood completely though there is a detailed picture of the structural changes which occur in the protein during the photocycle steps (Schobert et al. 2003 [28], Rouhani et al. 2001 [29]). X-ray crystallography has revealed plenty of the information concerning the mechanism. 20 models of Bacteriorhodopsin are presented in the Protein Data Bank, 11 of them are structures of the different intermediates of BR. Together with FTIR experiments data (Heberle et al. 2000, [30]) this allowed to proceed significantly in the understanding of the proton release from the Schiff base to the extracellular part. Current work is concerned with the step which happens after the proton release; the structure of the M state of BR is described in Chapter 5.

1.2.2 Ground state structure of Bacteriorhodopsin

Several important features have emerged from the X-ray structures of the ground state of BR at better than 2 Å resolution. Most strikingly, the detailed atomic nature of the Schiff base's complex counter ion, which stabilizes the high pK_a of 13.5 of the Schiff base and the low pK_a of 2.2 of the Asp85 was revealed (Luecke et al. 1999 [7]). This counter ion is stabilized by a network of water-mediated hydrogen bonds which extends from the Schiff base, the primary proton donor, to the extracellular surface (Figure 5). A key water molecule, Wat402, forms H-bonds to the Schiff base nitrogen and the two negatively charged aspartates (Asp85, the primary proton acceptor, and Asp212). Within the framework of this complex counter ion, this water molecule stabilizes the unusually high pK_a of the Schiff base. Two further water molecules (Wat400 and Wat401) combine to form a pentagon, and link the two negatively charged aspartates to the positively charged guanidinium group of Arg82, which is located further away towards the extracellular surface, but in the ground state structure is orientated towards the active site. The presence of three H-bonds to Asp85 ($O_{\delta 2}$ receives a H-bond from Thr89 and $O_{\delta 1}$ receives H-bonds from Wat401 and Wat402) as well as the position of the positive charge of Arg82 stabilize the negative charge on the carboxylate group to the extent that the pK_a of Asp85 is 2.2. Arg82 in turn is connected through a series of H-bonds to the putative proton release groups, formed by Glu194, Glu204 and water molecules.

Whereas several well-defined water molecules appear along the proton translocation channel on the extracellular side of the protein, on the cytoplasmic side there are relatively few. One conserved water (Wat501) appears between Trp182 and Ala215, and another crystallographic water (Wat502) was identified near the carbonyl oxygen of Lys216. An H-bond from Ala215 to Wat501 creates a local distortion of the H-bond pattern of helix G in this region, and this was recently assigned as a π-bulge, providing a more detailed description of the bend in helix G. With the exception of Asp96 and Thr46, which share a H-bond, the cytoplasmic half of the proton translocation channel consists predominantly of hydrophobic residues. These pack tightly against each other and create a hydrophobic barrier preventing the leakage of protons back across the membrane, which would otherwise dissipate the energy stored within the proton-motive potential. Nevertheless, it is necessary that a proton enters the proton translocation channel from the cytoplasm during the latter half of the photocycle. To assist this, the four aspartates at the cytoplasmic surface of BR (Asp36, Asp38, Asp102 and Asp104) serve to attract protons from the cytoplasm.

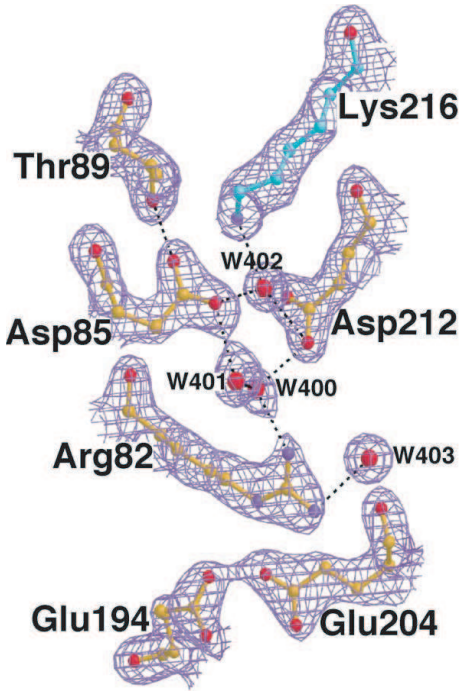

Figure 5: Refined structure and electron density of the extracellular half of BR in the ground state. $2F_{obs} - F_{calc}$ refined electron density map contoured at 1.2 σ shows the structure of the Schiff base counter ion formed by Asp85, Asp212, Wat400, Wat401, Wat402, Arg82. The proton release group is thought to be formed by Glu194, Glu204 and water molecules. (From Neutze et al. BBA 2002)

1.2.3 Proton pumping by Bacteriorhodopsin

Prior to the availability of high-resolution structures of BR in its ground and photocycle intermediate states, which provide exquisite details of its structural evolution, the mechanism of vectorial proton transport in BR was sketched in general terms. Frequently cited models include the isomerization/switch/transfer (IST) model and the local access model (Haupts et al. 1997 [31]). With the appearance of the intermediate state structures this picture was extended while new important conclusions where drawn from the available molecular structure (Edman et al. 1999 [25], Sass et al. 2000 [5], Schobert et al. 2003 [28], Edman et al. 2004 [32]). The outcome of the picture is shown in Figure 6. Structural details in the vicinity of the retinal must first define a pathway for proton transfer from the Schiff base to Asp85 on the extracellular side of the protein. Further structural rearrangements then break the accessibility of the Schiff base to Asp85 and create a pathway for proton transfer from Asp96 on the cytoplasmic side of the protein back to the Schiff base.

Light-induced retinal isomerization about the $C_{13} = C_{14}$ bond reverses the orientation of the N-H dipole of the Schiff base, which in the ground state is H-bonded to Wat402. It also causes a steric clash between the C_ϵ of Lys216 (the residue to which the retinal is bound) and Wat402. This strategic water molecule, which plays a key role in stabilizing the resting state, is consequently displaced, triggering a cascade of structural rearrangements on the extracellular side of BR. The network of water molecules becomes extensively disrupted and the guanidinium group of Arg82 reorients towards the extracellular medium. These changes dramatically perturb the pK_a values of key groups, and also facilitate a local flex of helix C which exaggerates an

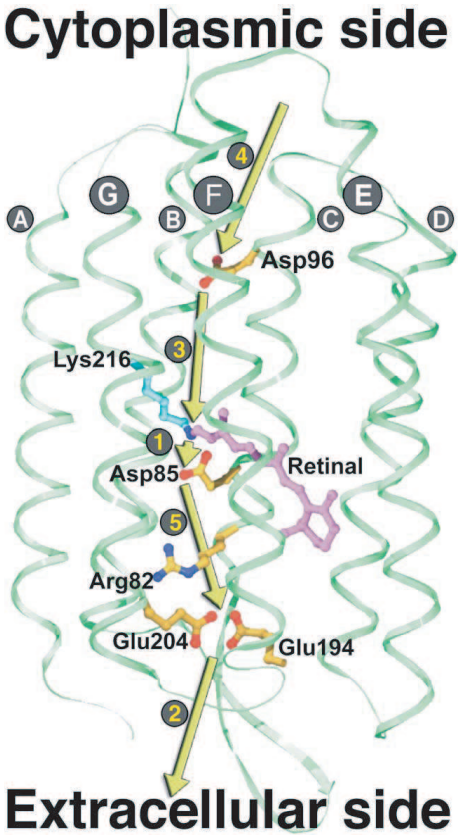

Figure 6: Schematic representation of the proton transfer steps in BR pumping cycle, overlaid on the ground state model. The primary proton transfer (1) is from the Schiff base to Asp85. A proton is released to the extracellular medium (2) by a proton release group, thought to be formed by Glu194 and Glu204 and/or water molecules. The Schiff base is subsequently reprotonated from Asp96 (3) which is then reprotonated from the cytoplasmic medium (4). The final proton transfer step (5) from Asp85 to the proton release group (via Arg82) restores the ground state. (From Neutze et al. BBA 2002.)

early movement of Asp85 towards the protonated Schiff base. This movement is driven in part by the mutual electrostatic attraction of the proton donor and acceptor. The movement of Asp85 towards the Schiff base creates a transient pathway for direct (or indirect) proton transfer. Proton transfer then cancels this electrostatic attraction. The release of strain on both the retinal and helix C enables the proton donor and acceptor to be drawn apart, thereby hindering the reversal of the proton transfer reaction. Proton release to the extracellular medium is assisted by the reorientation of the side chain of Arg82, as well as other movements which expose the proton release group to the extracellular surface. A number of events on the cytoplasmic side of the protein follow (see below), and the final proton transfer from Asp85 to the release group is delayed until the end of the photocycle, presumably occurring as the original network of water-mediated H-bonds on the extracellular side is restored.

The ground state structure of BR does not display an extended network of water molecules on the cytoplasmic side. The proton transport channel is characterized by an extended hydrophobic region which prevents the leakage of the stored proton-motive potential. Photoisomerization of the retinal results in a physical pulling on the side chain of Lys216 which induces, early on, a

small movement of its main chain. This displacement increases as the photocycle evolves, and a local unwinding of helix G in this region provides sites for water molecules to order transiently and thereby define a pathway for reprotonation from Asp96 almost as far as the Schiff base nitrogen. The final link facilitating Schiff base reprotonation from Asp96 is not well resolved and consequently not well understood in spite of several reported structures of the Bacteriorhodopsin M intermediate. Especially important M_2 intermediate should be considered to provide the information about the details of the Schiff base reprotonation pathway. Though the structure of this intermediate was already reported it could not provide the clear view of the reprotonation pathway because of the 2.25 Å resolution. (Sass et al. 2000 [5]).

A steric clash of the retinal's C_{13} methyl group with Trp182 following retinal isomerization provides the driving force for an outwards tilt of the cytoplasmic half of helix F which evolves later in the photocycle. This is assisted by proton transfer since the electrostatic attraction of the Schiff base towards Asp85 and Asp212 is released. A large outward movement of helix F opens the hydrophobic cavity, enabling Asp96 to be reprotonated from the cytoplasm. Finally, once the retinal thermally reisomerizes back to the all-trans configuration, the steric clash of the retinal's C_{13} methyl group with Trp182 is removed and this allows the structure to relax back to the ground state conformation.

1.2.4 Sensory rhodopsin II (pSRII)

In halophilic archaebacteria, sensory rhodopsins (SRI and SRII) are photoreceptors that relay light signals to their cognate transducer proteins (HtrI and HtrII). The small amount of SRII in the plasma membrane and its instability in detergents and low ionic strength solutions have precluded comprehensive studies of structure and function . On the other hand, the functional homologue of SRII from *Natronobacterium pharaonis* (pSRII) exhibits higher stability: it is stable in detergent solutions, and at low ionic strength (down to 50 mM) and high temperatures (range of 0-62 ^oC) (Tomioka et al. 1995, [33]). Most importantly, pSRII can be easily overexpressed in *Escherichia Coli* (Shimono et al. 1997 [34]).

Like other rhodopsins, pSRII displays seven transmembrane helices. Alignment of the amino acid sequence of these photoreceptors revealed a high degree of sequence homology in the transmembrane domains. In addition their sequences are highly homologous to the sequence of BR, especially for the residues which are in the pocket of the retinal. The homology of these residues is around 80% (Figure 7). So the retinal pocket is very similar to that of the BR.

The absorption maximum of pSRII from *Natronobacterium pharaonis* is at about 500 nm. This differs significantly from the absorption for the other three archaerhodopsins. The reason for this reduced opsin shift is not yet clear and should be elucidated on the basis of the structural information about the intermediates. The chromophor in pSRII is all-*trans* retinal and it is also bound via s Schiff base linkage to a Lys residue on helix G. pSRII does not exhibit light-dark adaptation, as only the all-*trans* retinal is extracted from pSRII in the ground state, which is consistent with the lack of binding of 13-*cis* retinal to the apoprotein opsin.

After light excitation pSRII thermally relax back to the original state through several intermediates. Generally this photocycle is quite similar to that of the BR, consisting of the intermediates K, L, M, N and O (Chizhov et al. 1998 [35]). The scheme of the photocycle is depicted in Figure 8. The K intermediate of pSRII is formed within 5 ps. The attributed structural changes are expected to be larger than those in the BR. The M intermediate exhibits a blue-shifted spectrum λ=390 nm, has a 13-cis retinal configuration and a deprotonated Schiff base. In pSRII Asp75 becomes protonated during the formation of the M-intermediate (this residue is the analogue of Asp85 in BR). The turnover rate of 1.2 s is much slower than that of BR probably because in the pSRII the Asp96 is replaced by an aromatic residue, thus interfering

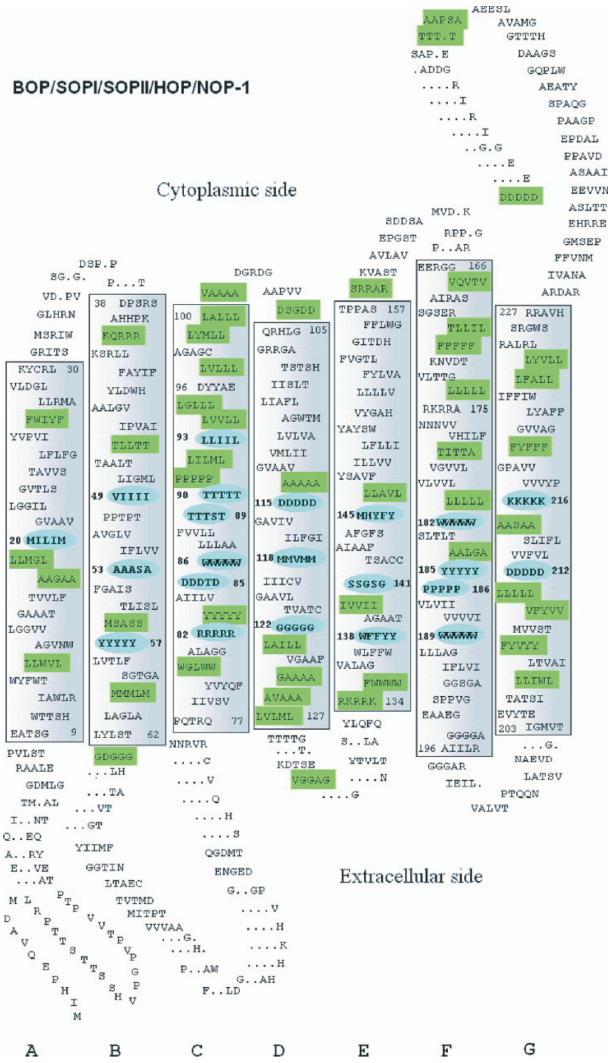

Figure 7: Sequence alignment of rhodopsins. Sequences for BR, SRI, SRII, HR from H.Salinarum and NOP-1 from *Neurospora crassa* are shown. Highly conserved residues are highlighted in green and less conserved residues are highlighted in blue. Spudich et al. [1]

with an optimal proton transfer from the cytoplasm to the Schiff base.

When not bound to HtrII, pSRII exhibits light-induced proton pumping function as BR, although at very low rates. This proton pumping activity is inhibited when pSRII is bound to the transducer HtrII; however, proton uptake and release still take place in the extracellular surface of pSRII (Sudo et al. 2001 [36]). Proton transport in pSRII has been analyzed in detail by Schmies et al 2001 [37]. Due to their slow photocycle turnover ($t \sim 1$ s) pSRII is a less effective pump than BR ($t \sim 10$ ms). The efficiency of the light-induced photocurrent is about 100 times weaker than that of the BR. The differences to BR happen in the second half of the photocycle after the rise of the M-state. Proton release happens like in the BR case but proton uptake is considerably slower because as sequence alignment shows the proton donor Asp96 in BR is replaced by aromatic residue in all SRs. The efficiency of the proton pumping can be enhanced by replacing Phe86 by an Asp residue or by the addition of the external proton donor

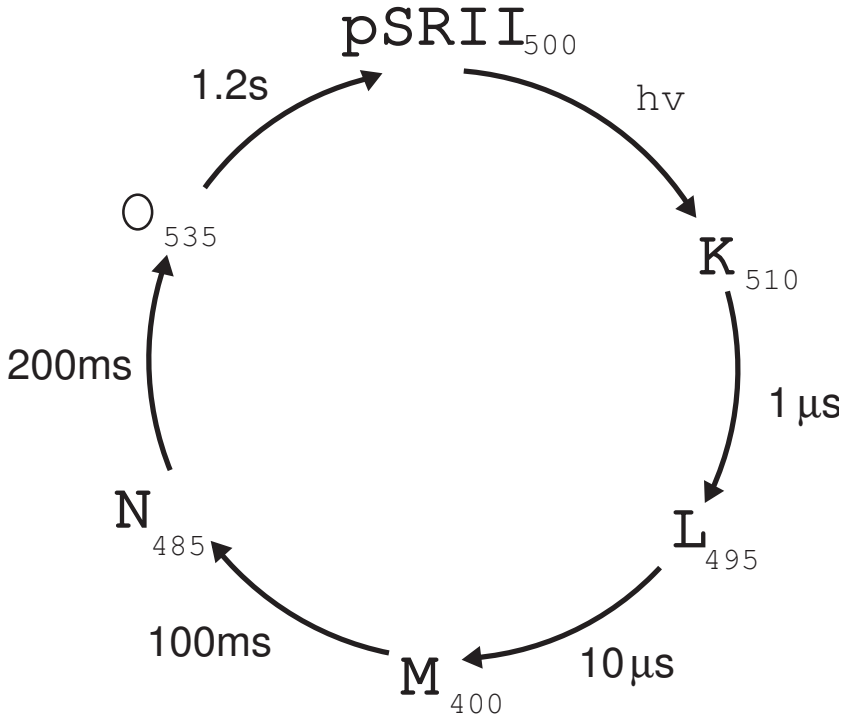

Figure 8: Sensory rhodopsin pSRII photocycle. The spectral intermediates, their absorption maxima and their lifetimes at room temperature are shown

like azide (Schmies et al. 2000 [38]).

1.2.5 Ground state structure of Sensory rhodopsin II

The structure of pSRII was solved at 2.1 Å resolution (Figure 9A) (Royant et al. 2001 [10]). This structure highlights similarities and differences in the retinal-binding pocket and along the putative proton pathway among archaerhodopsins of known structures. In difference to BR, the retinal chromophor in pSRII is largely unbent, which could contribute to the blue-shifted spectrum of pSRII. The hydroxyl group of Thr204 alters the polarity of the retinal-binding pocket, contributing to the blue-shifted spectrum of pSRII (Kandori et al. 2001 [39]). Unlike its orientation in BR, the guanidinium group of Arg72 in pSRII points toward the extracellular side of the membrane, resulting in a larger distance to Asp75 ([10]). This larger distance may account for the elevated pK_a of 5.6 for Asp75, as compared with a pK_a of 2.2 of the corresponding Asp85 in BR. The protonated Schiff base in pSRII is also H-bonded to a water molecule (W402), which in turn is H-bonded to two Asp residues, Asp75 and Asp201.

The surface of pSRII reveals a unique patch of charged and polar residues at the cytoplasmic ends of helices F (Lys157, Ser158, Arg162, Arg164, and Asn165) and G (Asp214) (Sudo et al. 2004 [40]). This patch is not present in BR or HR. It is, therefore, probable that these positively charged residues interact electrostatically with the negatively charged cytoplasmic domain Gly-Asp-Gly-Asp-Leu-Asp of transmembrane helix TM2 of HtrII, which is conserved among various species. The extraordinary stability of the pSRII-HtrII complex may be explained by this electrostatic interaction, because it is further enhanced by the low-dielectric environment of the lipid bilayer.

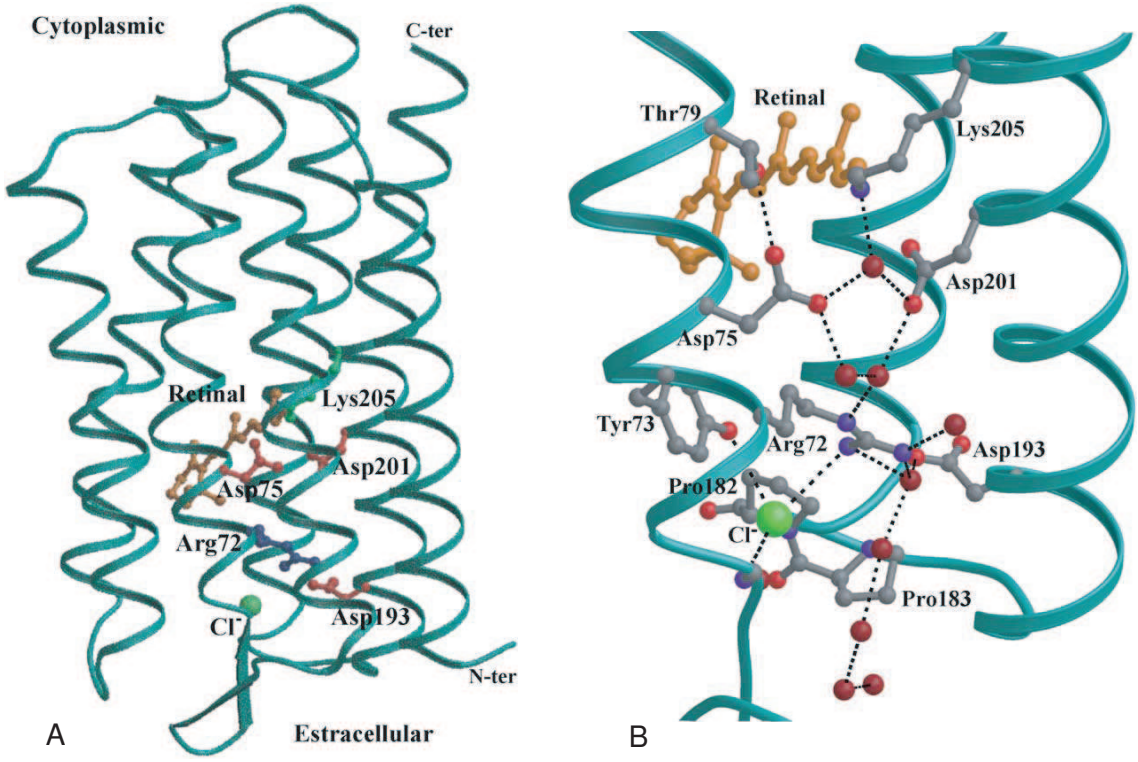

Figure 9: Sensory rhodopsin pSRII structure overview. (A) C_{alpha} trace representation. Key residues and the retinal are highlighted. (B) Hydrogen bonding network between the retinal binding site and extracellular surface. The retinal, water molecules and key residues are highlighted. From Pebay-Peyroula et al. BBA 2002.

Whereas the structures of the proton pathway in pSRII and BR are nearly identical from the Schiff base to W403, significant differences can be identified in the extracellular and cytoplasmic regions (Figure 9B). In contrast to Arg82 in BR's ground-state structure, the side chain of the corresponding Arg72 in pSRII is oriented towards the extracellular side. In addition, the proton release dyad in BR (Glu204 and Glu194) is replaced by a single negatively charged residue (Asp193) in pSRII. On the cytoplasmic side, pSRII is markedly more hydrophobic than BR, with Phe86 and Leu40 replacing Asp96 and Thr46 in BR, respectively.

1.3 Transducer molecule

Unfortunately the structure of the complete halobacterial transducer was not yet obtained. But the sequence homology to a family of receptors of the bacterial chemotaxis relates this protein to well evaluated structures (Le Moual et al. 1996 [41]). The aspartate receptor of *E.Coli* is an example of the protein which is as halobacterial transducers designed for the transmission of the stimulus to the histidine kinase at the cytoplasmic site. So far it is the best characterized protein from the chemotaxis receptors (Stoddard et al. 1992 [42]). It is approximately 60 kDa and exists as a stable homodimer. The receptor possesses the periplasmic ligand binding domain, a transmembrane region, and a cytoplasmic domain that possesses adaptation sites and a kinase docking surface (Figure 10).

The cytoplasmic region is highly conserved among all members of the chemoreceptor family

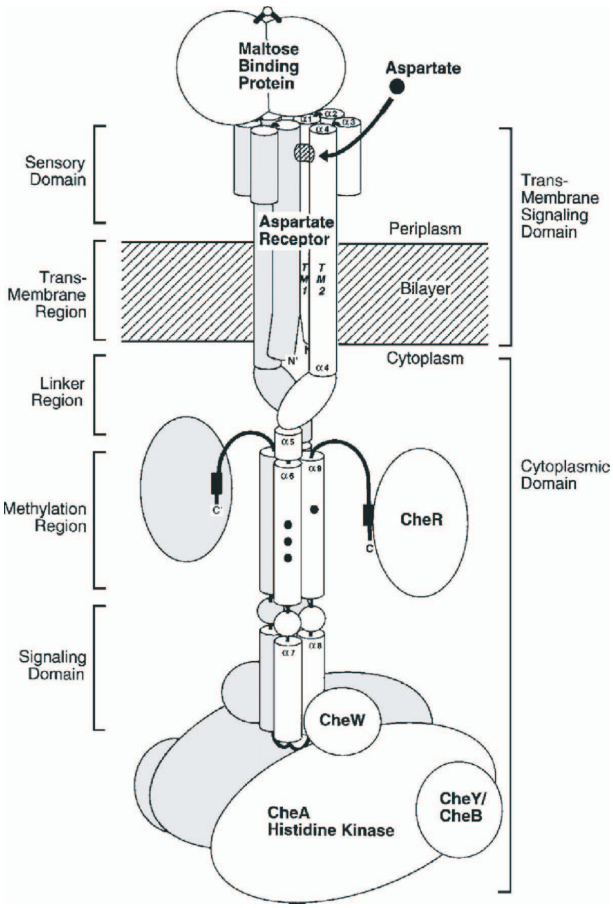

Figure 10: A typical receptor-kinase signalling complex illustrated by the aspartate receptor. Most of the chemosensory pathway components are associated with the supermolecular signaling complex. From Falke et al. [18]

reflecting a common function in CheA kinase regulation. This includes the transducer proteins HtrI and HtrII from Archaea. The amino acid identity is especially high for a distinct central region between sequences that contain the sites of methylated helices. Structure prediction algorithms indicate that the methylated helices are coiled coils and it can be hypothesized that almost the entire cytoplasmic region may participate in an extended coiled-coil structure ([41]). Amino acid sequence analysis of the cytoplasmic region reveals a pattern of amino acid characteristic of coiled-coil structures. A seven residue repeat $(a-b-c-d-e-f-g)$ with hydrophobic residues in positions a and d indicates this arrangement. The crystal structure of the cytoplasmic domain of a serine chemotaxis receptor of *E.Coli* was reported by Kim et al. 1999 [43]. It has indeed revealed a 200 Å long coiled-coil of two antiparallel helices connected by a 'U-turn'.

The experiments of Jung and Spudich, 2001 [44] explicitly show that the transducer molecules HtrI and HtrII share the same structure and as a consequence are expected to show the same signalling mechanism. Chimera protein was made by replacing the cytoplasmic portion of the transducer HtrII (from *Natronobacterium pharaonis*) with that of the Tar aspartic receptor from *E.Coli*. These chimera participate in a retinal-dependent phototaxis response in *E.Coli*.

1.4 Receptor/Transducer complex

The interaction of archaeal transducers with their cognate sensory receptors has been analyzed, and it could be proven that specificity is determined by their transmembrane helices. Spudich and co-workers prepared transducer chimeras between HtrI and HtrII in which different transducer domains were combined. After expression of the chimeric signalling complexes in *H.salinarum* the authors analyzed the phototaxis of these cells and concluded that the receptors interact specifically with their cognate transducers. Whereas the cytoplasmic domains can be exchanged, two transmembrane helices TM1 and TM2 are needed to mediate a correct physiological response (Spudich et al. 1999 [45]).

The binding of the cognate transducer to the receptor influences the proton pumping function of the receptor. The pSRII-F86D mutant, which displays an enhanced photocurrent, has shown no proton pumping activity upon the transducer binding in voltage-clump experiments made by Sudo et al 2001 [36]. The binding of the transducer to its receptor only affects the photostationary but not the transient photocurrent. Therefore, fast proton reactions are not inhibited and neutralization of the Schiff base-counter-ion pair can still occur. The specific interaction located in the cytoplasmic part of the membrane is considered as an interaction between the receptor and its transducer, which is responsible for the inhibitory effect on proton pumping. HtrII could close the cytoplasmic channel of pSRII which reduces the accessibility of the protons from the cytosol and thereby inhibits the proton pumping. It was shown (Engelhard et al. 2003 [46]) that the shortened transducer HtrII(1-114) is tightly bound to the pSRII ($K_D = 200\ nM$), thereby still being able to block the proton pump.

The structure of the complex of pSRII with this shortened transducer was recently obtained (Gordeliy et al. 2002 [47]). In the crystals the complex of pSRII and HtrII forms a homodimer. A dimer structure of HtrII from *N.pharaonis* in the native environment has been deduced already from electron paramagnetic resonance (Wegener et al. [53]). The expected dimer of the complex is formed by a crystallographic two-fold rotation axis, which is located in the middle of four transmembrane helices: TM1, TM2, TM1', TM2' (where a prime indicates the right-hand complex; Figure 11A). The transmembrane helices F and G of the receptor are in contact with the helices of the transducer. The interface between receptor and transducer is formed mainly by van der Waals (vdW) contacts and only a few hydrogen bonds. Whereas the straight TM2 is oriented parallel to helix G of the receptor, TM1 is kinked at Gly37 and bends away from the receptor.

Although only van der Waals contacts are observed between the four transducer helices themselves, defined cross-connections are observed between receptor and transducer. The F-G loop region affixes the transducer by several contacts as well as by three hydrogen bonds between Thr189 (pSRII), Glu43 (TM1) and Ser62 (TM2) (Figure 12). A second anchor point is observed in the middle of the membrane where, as mentioned above, the phenolic hydroxyl group of Tyr199 (helix G) bridges to Asn74 (TM2). A view from the cytoplasm down the binding domain (Figure 11A) reveals that closer contacts are between helix G and TM2, fixating these two transmembrane helices to one another. There are twice as many van der Waals contacts between helices G and TM2 than between F and TM2. The closer packing between G and TM2 can be quantified by an average van der Waals distance of 4.06 Å in comparison to a value of 4.22 Å between F and TM2 (Figure 11B). The four helices of the transducer in the dimer are packed against each other, thereby intercalating their bulky hydrophobic side chains.

Most notably, the four helices of the transducer extend beyond the extracellular side by about three helical turns, comprising residues 44-59 (Figure 11B). This sequence is missing in HtrI from *Halobacterium salinarum*, as obtained from a sequence alignment with *N. pharaonis* HtrII, but is repeated at the N-terminal end of TM2 in *H. salinarum* HtrII (Zhang et al. 1996 [48]). Notably,

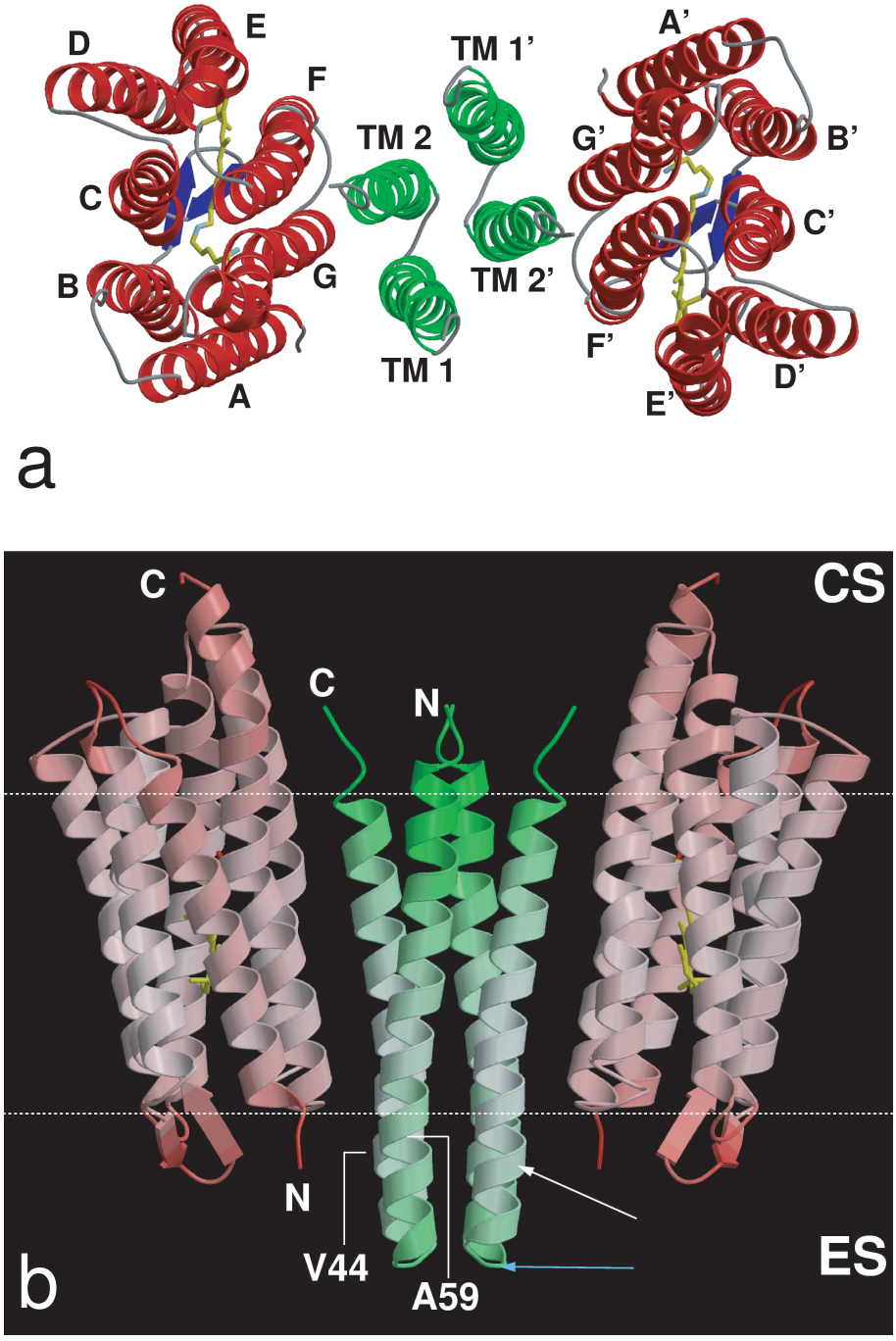

Figure 11: Fold of the receptor-transducer complex. (A)Ribbon diagram of the top view from the cytoplasmic side. (B) Side view of the complex. From Gordeliy et al. [47]

H. salinarum HtrII not only transmits the signal from the photoreceptor SRII but also operates as a chemoreceptor (Hou et al. 1998 [49]). This function is conferred by a serine-binding domain

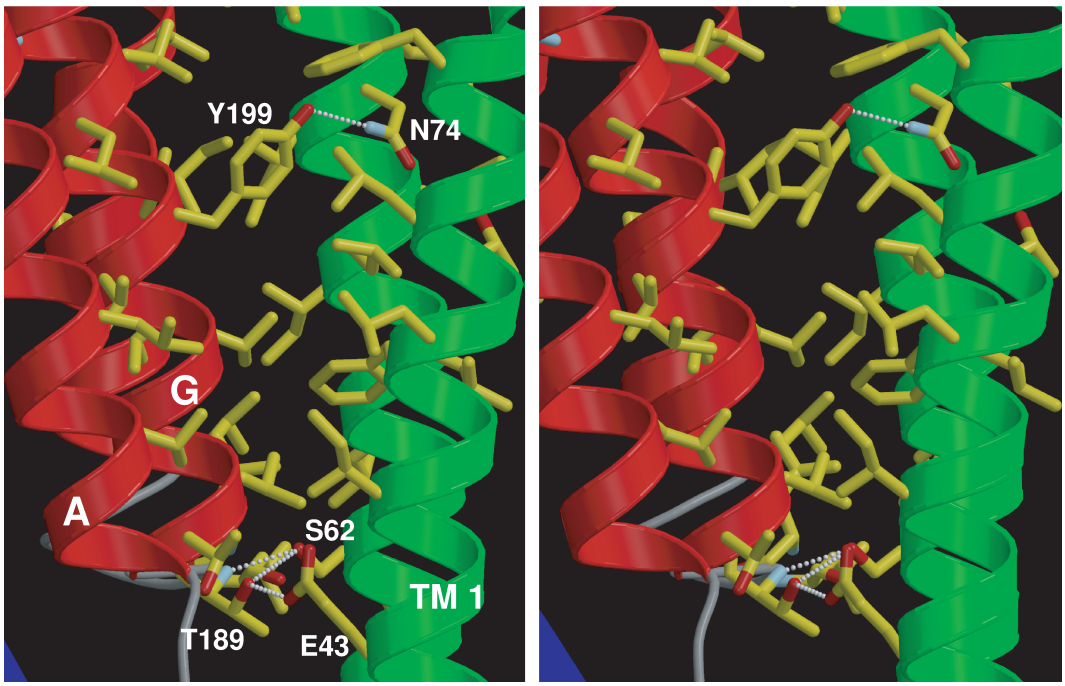

Figure 12: Stereo view of the hydrogen bonds and van der Waals contacts between receptor and transducer. The residues that are involved in hydrogen bonds are labelled. From Gordeliy et al. [47]

that is inserted in front of the sequence forming the stalk. Crystal structures of the ligand-binding domain of homologous eubacterial aspartate receptors display two extended helices in the dimer (4 and 4'), which could connect to a structural element like the stalk (Yeh et al. 1996 [50]). The observation of different degrees of periplasmic domain excision is in line with the evolution of the four archetypical halobacterial rhodopsins, which has been explained by two gene duplication events (Ihara et al. 1999 [51]). According to this hypothesis a proto-chemoreceptor gene has been acquired by the proto-sensor gene after the first duplication step.

1.5 Molecular mechanism of the signal transfer

The present structure and the observation that the flap-like movement of helix F induces a rotation of the cytoplasmic end of TM2 (Koch et al., Klare et al. 2004 [52]), gives rise for a probable mechanism of transmembrane signalling. Helix F contains a proline residue (Pro175; Fig. 4a) at an equivalent depth as Tyr199 (Figs 3 and 4a), which could function as a hinge for the light-activated movement of helix F (a similar role has been proposed for Pro185 in Bacteriorhodopsin). If the outward bending of helix F is in the same direction as observed for Bacteriorhodopsin, it should collide tangentially with TM2 (Figure 11 and 12), thereby inducing its rotary motion.

EPR measurements analyzed this reaction (Wegener et al. 2000 [53]). Specifically spin-labelled transducer and receptor Cys mutants positioned on helices F and G, as well as on TM1 and TM2 were prepared. Two observations were central to establish a model for receptor-transducer signal transfer (Figure 13). The mobility of the S158R1 (R1 denotes the spin-modified Cys side chain) and L159R1 is changed after light excitation of pSRII. S158R1, which faces TM2, experiences a transient immobilization the opposite is observed for L159R1, which is oriented

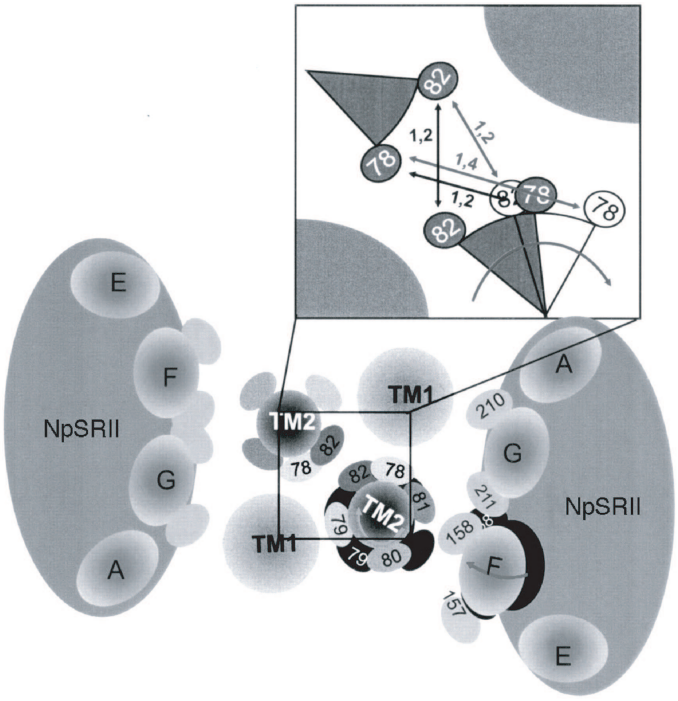

Figure 13: Schematic representation of the light induced conformational changes in 2:2 receptor-transducer complex. View from the cytoplasmic side. Arrows shows the conformational changes. Wegener et al. EMBO Journal 2001

towards the interior of pSRII. This implies the above mentioned motion of helix F. TM2,which was modified at positions 78 and 82, revealed relatively mobile unperturbed residues. The dipolar coupling allowed determination of distances between the two residues. On light excitation only the distance between V78 and V78' increases, but not that between L82 and L82'. This result is only compatible with the rotary motion of the TM2. However, it should be emphasized that a small piston-like movement of about 1 Å cannot be excluded which would - together with the rotation of TM2 - result in a screw-type motion.

The above mentioned putative mechanism of the signal transfer from receptor to transducer do not account for the signal transfer from the membrane to the CheA activation domain. This arises the question of how small conformational alterations such as rotation or piston stroke can be transmitted along large distances. In the case of transducer it would be almost 260 Å. The key role in understanding the mechanism of signal transfer will certainly be played by the linker domain which connects the transmembrane helices with the cytoplasmic four-helix bundle.

In connection with possible signal relay mechanism two more general models should be discussed. In the first model the small perturbation at the membrane domain, rotation or piston stroke, is transmitted to the signalling domain (Otteman et al. 1999 [54]). The rotation could invoke an unwinding of the coiled-coil domain. However these small alterations might be damped out by the helix dynamics during their way to the tip of the helix bundle. On the other hand, if the linker region can provide the amplification one could envision the transmission of the signal. A second possibility was proposed by Kim et al 2002 [55]. In their model the activation of the receptor domain decreases the dynamics of the cytoplasmic domain, thereby increasing phosphorylation of CheA. The adaptation process was also explained in the framework of this model as methylation of Glu residues would modify the dynamics of the transducer.

2 Materials and methods

2.1 Equipment

UV-VIS spectrophotometer	UV-2101PC, Shimadzu, Japan
Gel-electrophoresis	PowerPack P25, Biometra, Goettingen, Germany
Autoclave	Fedegari Autoclavis SPA, IntegraBioscience, Fernwald, USA
Centrifuges	Centrifuge 5417R, Eppendorf, Hamburg, Germany
	Megafuge 1.0R, rotor BS402/A, Heraeus, Osterode, USA
	Sorvall RC5C, rotors DuPont SS34, Wilmington, USA
	Beckman Optima L-70K, rotor Ti70, USA
French press	French Pressure Cell Press, SLM Aminco, Rochester, USA
	French Press Cell 40K, SLM Aminco, Rochester USA
Incubator	Multitron, Infos HAT, Einsbach, Germany
Sonicator	Transsonic digital, ELMA, Singen, Germany
pH-meter	MP220 Mettler Toledo, USA
Microscope	Olympus, Japan
Vacuum evaporator	Buchi Waterbath B480, Germany

2.2 Materials

Materials of columns	Ni-NTA agarose, Qiagen
for chromotography	DEAE-sephacel, Sigma, USA
Filters, concentrators	Centricon-50kDa, Centriplus-50kDa, Amicon, Beverly, USA
Chemicals	OG, n-octyl-D-glucopyranoside, Sigma, USA
	DDM, n-dodecyl-D-maltoside, Anatrace, USA
	MO, 1-monooleyol-rac-glycerol, Sigma, USA
	MV, monovaccenin, Nu-Chek Prep. Inc, USA
	MP, 1-monopalmitoyl-rac-glycerol, Sigma, USA
	NaH_2PO_4, Merck, USA
	KH_2PO_4, Merck, USA
	Na_2HPO_4, Merck, USA
	$MgCl_2$, $CaCl_2$, Merck, USA
	EDTA, Merck, USA
	Imidazole, Sigma, USA
	IPTG, Sigma, USA
	all-trans retinal, Sigma, USA
Ferments	protease inhibitor (Complete Coctail), Roche, Switzerland
	desoxiribonuclease I, Sigma, USA
Bio-beads	SM2 adsorbent, Bio-Rad Laboratories, USA
Bacteria growth media	dYT - 16g Bacto-tryptone, 10g Bacto-yeast extract, 5g NaCl per 1 liter
Additives to media	canamicin, Roche, Switzerland
	ampicillin, Roche, Switzerland
Buffer solutions	
buffer for cells	A - 150 mM $NaCl$, 25 mM $NaPi$, 5 mM $EDTA$, pH = 8.0
buffer for solubilization	B - 300 mM $NaCl$, 50 mM $NaPi$, 2% DDM, pH = 8.0
buffers for affinity	C - 300 mM $NaCl$, 50mM $NaPi$, 0.05 % DDM, pH=8.0
chromotography	D - 300 mM $NaCl$, 50 mM $NaPi$, 30 mM imidazol, 0.05 %

	DDM, pH = 8.0
	E - 300 mM $NaCl$, 50 mM $NaPi$, 200 mM imidazol, 0.05 % DDM, pH=8.0
buffers for DEAE chromotography	F - 10 mM Tris, 0.2 % DDM, pH = 8.0
	G - 30 mM $NaCl$, 10 mM Tris, 0.1% DDM, pH = 8.0
	H - 500 mM $NaCl$, 25 mM $NaPi$, 0.1 % DDM, pH = 8.0
reconstitution buffer	J - 1000 mM $NaCl$, 100 mM $NaPi$, 0.1 % DDM, pH = 8.0
resolubilization buffer	K - 150 mM $NaCl$, 25 mM $NaKPi$, 2% OG, pH = 8.0
crystallization buffer (for optimal crystal growth)	L - 150 mM $NaCl$, 25 mM $NaKPi$, pH=5.1, 0.8 % OG

2.3 Overexpression of pSRII in E.Coli

Selected cells (10 ml) was added in 800 ml of dYT medium containing antibiotic (50 μg/ml) and in 2 liter Erlenmeyer flasks placed in incubator (Multitron). Growth of bacteria was in the presence of oxygen at $+37^oC$ and extensive mixing (150 rpm). When OD_{578} reached 0.6, IPTG was added to the medium in concentration 1 mM, all-trans retinal was added to concentration 10 μm. After retinal addition, bacteria growth was performed further during 4 hours without light. Final culture was harvested from the solution by centrifugating at 6000 rpm (Beckman), pellet after centrifugation was collected and freezed at -80^oC.

2.4 Extraction and purification of pSRII protein

2.4.1 Cell lysis and solubilization of the membranes

Cells containing overexpressed protein were resuspended in buffer **A** (1 mg of cells to 3 ml of buffer), was cooled to $+4^oC$, in the solution was added 5-10 mg of complex protease inhibitor and 5-10 mg of deoxiribonuclease I (to 100 g of cells). After that lysis of the cell was performed in cooled to 0^oC French press cell (lysis procedure was repeated several times until the fraction of the destroyed cells was satisfying). Overexpressed protein is located in the outer membrane of *E.Coli*. Membrane containing fraction was selected by ultracentrifugation at 100000 g at $+4^oC$ (Beckmann, rotor Ti-70). After that membranes with the protein were resuspended in solubilization buffer **B** (25-30 ml). Solubilization was performed in the dark at $+4^oC$ during 8-12 hours with mild shaking of the solution.

2.4.2 Affinity chromatography

The protein purification is performed in one step, due to the presence of 7-histidine tag in the C-terminus of the protein (Bailon et al. 2000 [56]). Between this histidine residues and Ni^{2+} iones, immobilized at the ion-binding Ni-NTA agarose, specific binding happens. Upon addition of big amount of imidazole containing solution to the column with bound protein due to the binding concurrence between imidazole and histidines protein can be eluted.

The column containing Ni-NTA agarose at $+4^oC$ in the beginning equilibrated with 3 column volumes (volume of the material loaded in column = 25-30 ml) of buffer **C**. After that the protein in solubilization buffer **B** is loaded. 20-30 min protein is equilibrated in the column to bind to the column material. After that the content of the column is washed by 5 volumes of buffer **D** with low concentration (30 mM) of imidazole to remove the proteins which has bound to the column with their histidine residues with less efficiency as HisTag. Elution is performed with buffer **E** containing high concentration of imidazole (200 mM), eluted fractions are not selected.

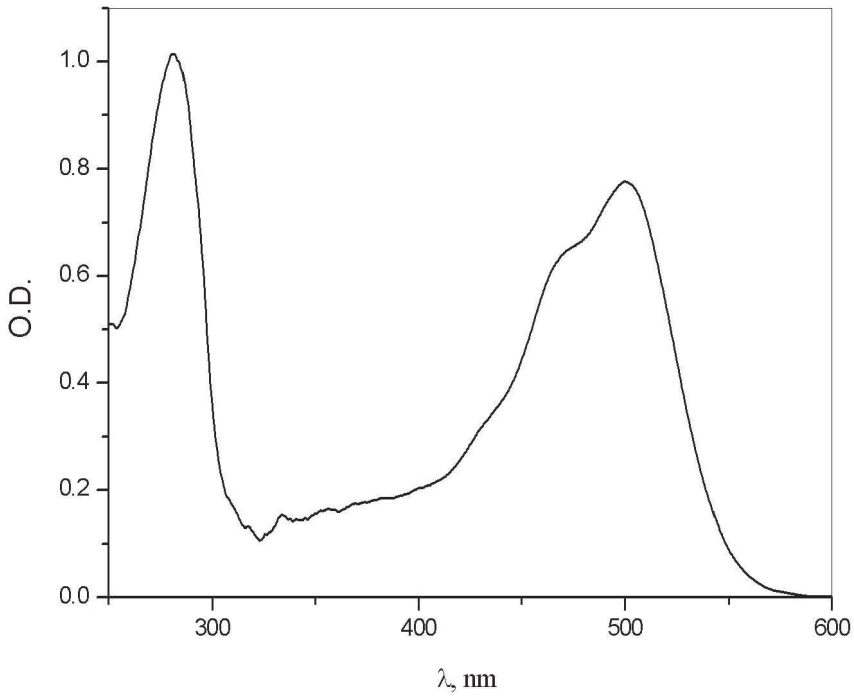

Figure 14: Absorption spectra of the pSRII in visual range. Peak at 500 nm corresponds to the absorption maximum of the retinal inside of the protein. Peak at 280 nm corresponds to the absorption maximum of aromatic aminoacids.

2.4.3 DEAE chromatography

To remove the excess amount of imidazole, buffer with eluate is supplied to the ion-exchange column, the concentration of the ions in buffer is reduced by dilution to 30 mM prior the application. Column material at room temperature equilibrated by 3-4 volumes of buffer **F**. After the protein is added to the column. The column with bound protein is washed with 5 volumes of buffer **G**. Elution is performed after loading the column with buffer **H** containing big concentration of ions (500 mM of $NaCl$).

2.5 Determination of the protein concentration

Concentration of the protein obtained was determined from the absorption spectrum in the visual range. Spectrum for pSRII is shown in Figure 14. Extinction value at the maximum absorption wavelength 500 nm is equal to 40000, respective concentration of the protein is calculated from

$$c = \frac{A_{500}\epsilon}{Md} \tag{1}$$

where A_{500} - absorption at 500 nm, ϵ - extinction coefficient, M - molar weight and d - width of the cuvette.

This spectra characterizes also the quality of the protein. The absorption value at 280 nm is proportional to the number of the aromatic groups in the protein. The absorption value at

500 nm is proportional to the number of functional proteins, because it is the absorption of the retinal in the correct protein environment. From the ratio of this two absorption values one can estimate the percentage of the physiologically active protein. The degraded protein will contribute only to peak at 280 nm. The ratio of 280nm/500nm is normally expected to be in the range 1.2-1.4 for pSRII.

2.6 pSRII storage

Protein can be kept at $+4^oC$ during 2-3 month, after this time gradual degradation of the protein is observed. For longer terms protein can be frozen in a liquid nitrogen and kept at -80^oC. Protein is stable in range of pH from pH=4.0 to pH=10.0.

2.7 Extraction and purification of the truncated transducer

Expression, solubilization and purification of the truncated transducer was performed as for pSRII. Gene of the truncated transducer was cloned to pET27bmod vector and transducer was overexpressed in *E.Coli*. At the C-terminus overexpressed protein has the polyhistidine marker and is purified with affinity chromatography. Corresponding work on transducer was performed at collaborating Max-Plank Institute Dortmund.

2.8 pSRII-HtrII complex preparation

To make a complex of pSRII and truncated transducer both proteins were incorporated into lipid bilayers. It was shown also that solubilized proteins after incubation form a complex with stoichiometry 1:1, after the reconstitution to the lipid bilayer the stoichiometry of the complex is 2:2.

Different methods exist to incorporate the solubilized membrane protein into the lipid bilayer: reconstitution making use of organic solvents, mechanical approaches as ultrasonification, French press cell, removal of the detergent (Rigaud et al. 1995 [57]). Current approach consists in removal of the detergent. Choice of a certain method depends on the type of detergent. Detergents with high CCM (critical concentration of the micelle formation) can be easily removed by dialysis and gel filtration. Detergents with low CCM can be removed by absorption of the detergent on hydrophobic surfaces. Such hydrophobic surfaces are specially designed, for example SM2 Bio-Beads.

DDM detergent was used for solubilization of the protein. It is a mild detergent. It forms big micelles ($\approx$ 70 kDa), it has a low CCM = 0.1 - 0.3 mM and so Bio Beads are used for its removal (Figure 15A).

Mixture of the proteins in the ratio 1:1 and polar lipids (1M of pSRII to 35 M of polar lipids) transferred in buffer **J** - reconstitution buffer. Bio-Beads balls are added to the solution in the amount proportional to the detergent content (pSRII has concentration 3.5-4 mg/ml). Due to gradual removal of the detergent phase transition from micelles to bilayer happens (Figure 15B). During this phase transition protein incorporates to the bilayer and forms the complex.

Resconstitution is performed at $+4^oC$ in the nitrogen atmosphere with the mild shaking of the solution during 10-12 hours. After that Bio-Beads are removed and membranes containing reconstituted protein collected by centrifugation at 100000 g (Beckmann, rotor Ti-70) at $+4^oC$ during 1.5 hours.

Lipid membranes resolubilized further with OG (this is the detergent used in crystallization). Membranes are homogenized in the resolubilization buffer **K** by short sonification (10-15 seconds, Transsonic digital, ELMA). Solubilization is performed at $+4^oC$ in the dark with a mild shaking

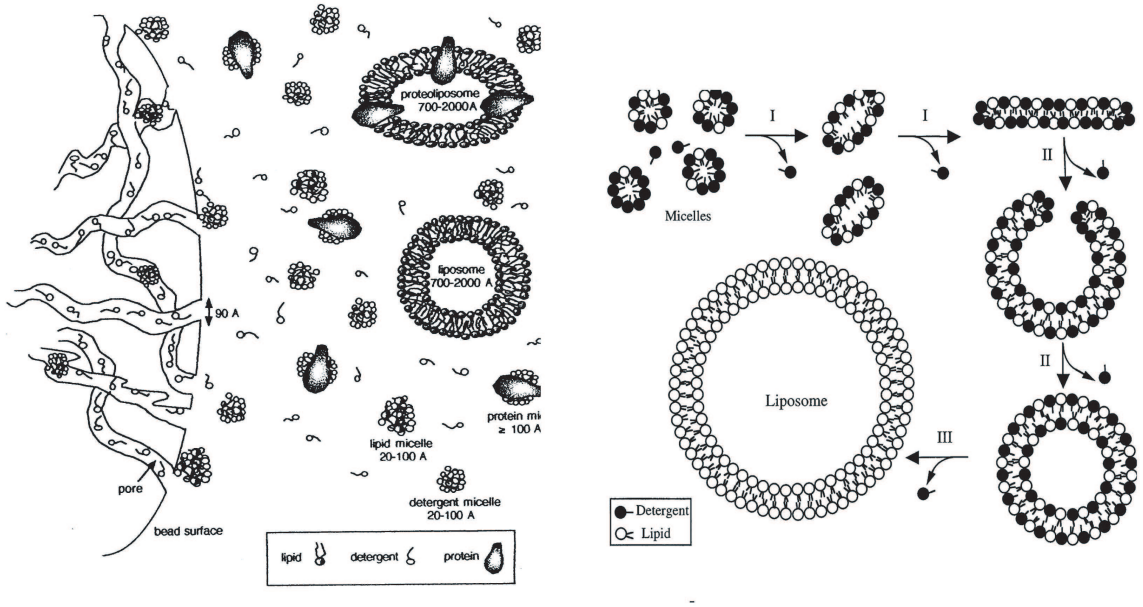

Figure 15: Reconstitution scheme of the membrane proteins by removal of the detergent. (A) Protein reconstitution into liposomes with the use of Bio-Beads. (B) Liposome formation upon the removal of the detergent. From Rigaud et al. [57]

during 8-10 hours. Obtained complex is crystallized or frozen in liquid nitrogen and stored at -80^oC.

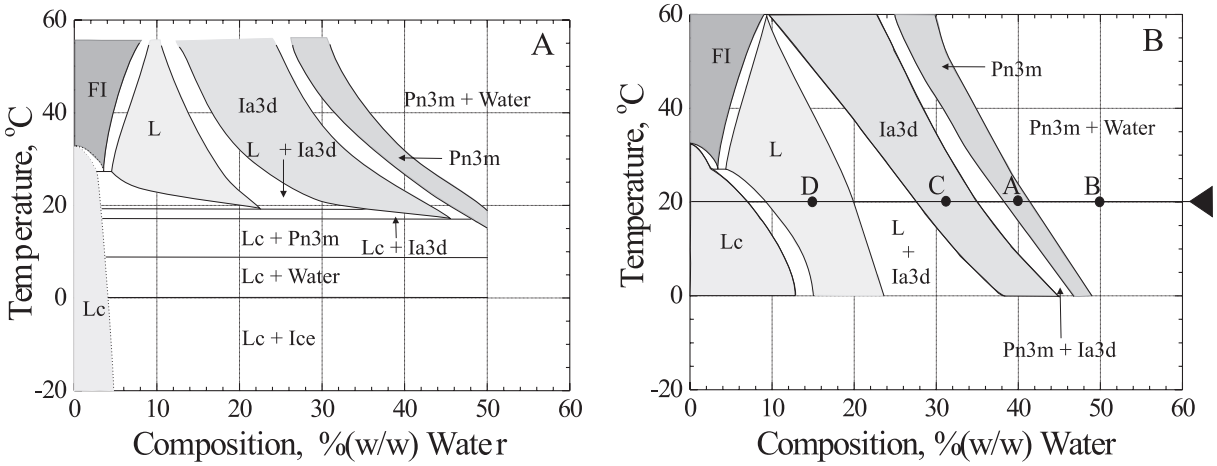

Figure 16: Monooleyl-water phase diagram. A. Equilibrium phase diagramm. B. Metastable phase diagramm. From Caffrey et al. Journal of Structural Biology. 2003

2.9 Crystallization in cubic lipidic phase

2.9.1 Cubic lipidic phase

To get the high resolution structure of the protein by X-ray diffraction method one should obtain well organized crystals of the protein. In the case of the membrane proteins the crystal production step is the main bottleneck. The number of unique membrane protein structures existing in Protein Data Bank ranges from 40 to 60. So membrane proteins represent a challenge in contrast to water-soluble proteins.

Diffraction-quality crystals are particularly difficult to prepare currently when a membrane source is used. The reason for this is a limited ability to manipulate proteins bearing hydrophobic/amphiphilic surfaces that are usually enveloped with membrane lipids. The removal of the membrane proteins in the case of the solubilization from the membrane environment into aqueous solution with the use of the detergent does the protein unstable; even very modest perturbations can bring to the loss of the native function of the protein or even to its denaturation.

This problems initiated the search for the conditions which let the protein crystal growth and have the similar to the original membrane environment properties. The main demands to such a system: (i) mechanical properties of the media (viscosity, elasticity) should be alike to the biological membranes properties; (ii) system should satisfy to both hydrophobic and hydrophilic surfaces of the protein; (iii) it should be able to incorporate sufficient big quantity of the protein and detergent without apparent change in its structure and without inducing the perturbations in the protein structure; (iv) system should have a necessary optical properties (transparency) for an easy detection of the process happening within the media; (v) it should let the formation of the crystals nuclei and their growth, e.g. should support the transport of the protein to the locations of the crystal nuclei.

According to this criteria the group of Landau and Rosenbusch, 1996 [58], came to the idea to use as such system one of the phases of the lipid/water system (Figure 16). Among the phase variety in this system there are two highly viscous cubic phases: micellar and bicontinuous. They correspond to stated above criteria. Both phases are macroscopically stable, sufficiently tough, transparent and capable to perform the transport of the protein molecules. Micellar cubic phase represents the organization of the micelles in the vertex of the cubic lattice surrounded by the

water. Bicontinuous cubic phase represents continuous non-polar and aqueous parts, which have the cubic symmetry; lipidic part divides the aqueous part to two channel nets which do not cross each other. This phase become indeed a suitable choice for growing the crystals of the membrane proteins.

2.9.2 Crystallization mechanism

The method of crystallization is sometimes referred to as *in cubo* as the success in the crystallization has happened in the bicontinuous cubic phase. This cubic phase has several possible symmetries: $Ia3d$ (body-centered) formed at low content of the water (less than 35%), $Pn3m$ (plane) formed at high water content, $Pm3n$, $Im3m$. Phase location with different symmetries in the phase diagram of the system, conditions for the transitions between the phases are strongly related to the geometry of the lipid forming the cubic phase.

Crystallization of the membrane proteins happens in $Pn3m$ cubic phase (at high water content). It is assumed that the protein to be crystallized is initially reconstituted into the lipid bilayer of the cubic phase. The cubic phase does not necessary remain stable throughout the crystal growing period and can transform with time to liquid phase depending on the precipitant used for crystallization.

It has been proposed that the initial extreme and local salt-induced dehydration provides the impetus and driving force for nucleation and subsequent crystal growth. It was shown in support of this that the lattice parameter of the cubic phase drops upon addition of precipitant salt to the crystallization mix. Less water should be encompassed by the cubic phase and the protein which presumably also loose the water tends to have new protein-protein contacts which leads to crystal growth.

The general mechanism of the crystallization in cubic phase can be suggested in a speculative way with some experimental evidence supplied from X-ray or neutron scattering (these studies were particularly made for Bacteriorhodopsin, Nollert et al. 2001 [24]) and is demonstrated in Figure 17. Some additional considerations should be stated. Because the number of MO and water molecules in this crystallization system far outnumber that of BR molecules, formation of a macroscopically undisturbed cubic phase is possible. On the microscopic level, there are, however, consequences due to the presence of one BR molecule in one out of five unit cells of the $Ia3d$ phase (unit cell dimension, a=145 Å), or in one out of 13 unit cells of the $Pn3m$ (a=104 Å) phase prior to crystallization. The molecular dimensions of BR and the matrix lipid MO (monooleyn) are markedly different, both in length (BR, ca. 33 Å; MO, 17.3 Å) and diameter (BR, 25 Å; MO, $\sim$ 4-5 Å) of the respective hydrophobic, roughly cylindrically shaped moieties. Therefore, the precise regular geometric pattern of the bicontinuous cubic phase cannot be maintained in unit cells harboring BR molecules. Protein may be considered to be in a taut state, presumably caused by a curvature-related hydrophobic mismatch at the lipid-protein boundary. This mismatch incurs the energy penalty. Such a penalty is presumably minimized by a local deviation of the membrane curvature from its regular pattern. The mismatch could become crucial with increasing in curvature, which is brought about by the addition of salt to the cubic phase matrix. It can be resolved by effecting a phase separation of the host lipidic cubic phase into a highly curved cubic phase lacking membrane defects, and a lamellar crystalline protein array. So crystal growth in the lipid-protein plane may be envisaged by lateral merging of protein and of small purple membrane-like units, guided by the matrix bilayer.

In the same time reconstitution into the bilayer of the lipidic cubic phase limits the mobility of the protein. Cubic phase-incorporated protein molecules are pre-oriented by the bilayer, such that transmembrane α-helices reside within the hydrophobic core of the membrane, and the loop regions are solubilized in the aqueous channels. The lateral diffusion pathways are defined by

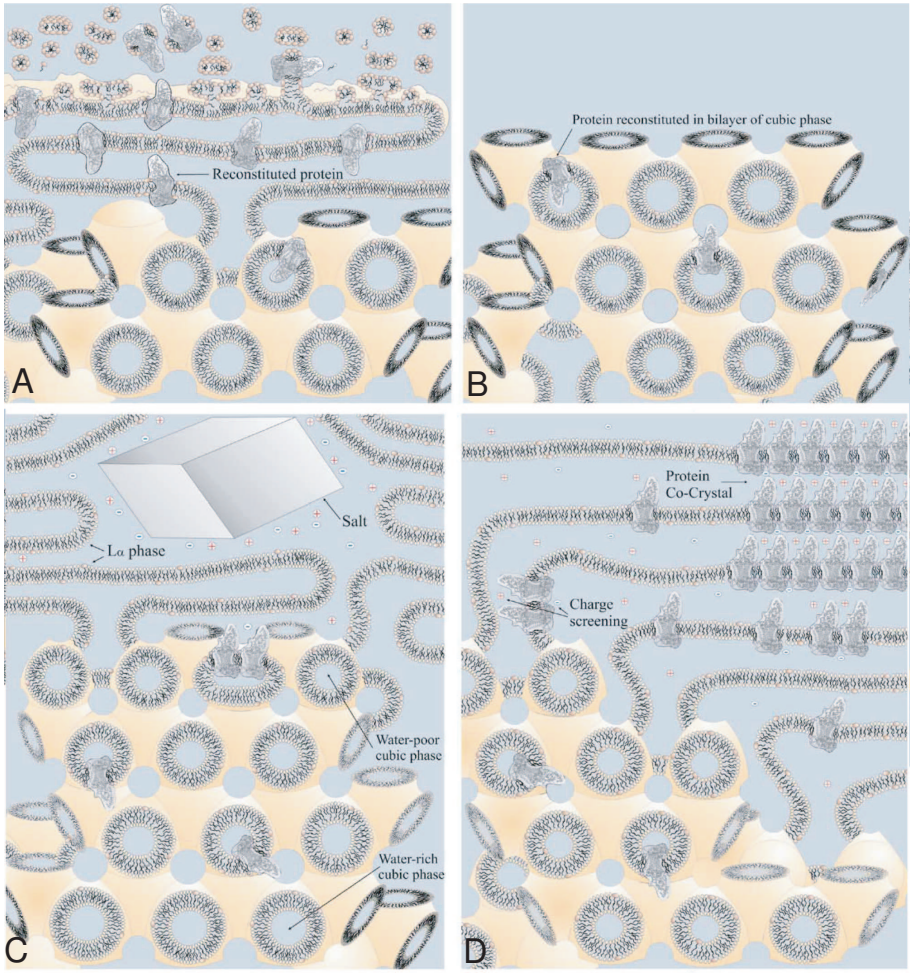

Figure 17: Crystallization mechanism in cubic lipidic phase. (A) Protein reconstitution and dispersion in mesophases. (B) Reconstituted protein in bilayer of cubic phase. (C) Addition of precipitant (salt) to initiate crystallization by water-withdrawing and charge screening effect. Bilayer curvature in cubic phase increases as water content drops. (D) Reversible crystallization of the protein from cubic phase through lamellar portal. From Caffrey et al. [59]

the structure of the curved lipid bilayer. This greatly increases the ratio of productive to non-productive protein encounters that are necessary for crystal nucleation and growth (Caffrey et al. 2004 [59]).

2.9.3 Crystallization of pSRII-HtrII complex and BR

Crystallization of the protein complex as well as BR was performed in cubic lipidic phase. Monooleyl (MO), monovaccenin (MV) and monopalmetoyl (MP) were used as lipids for the cubic phase formation. 4-6 mg lipids placed in 200 μl polypropylene tubes (Eppendorf) were melted during 30 minutes in thermostat (for MO - $+45^oC$, MV - $+55^oC$, MP - $+40^oC$). After that samples were cooled to room temperature and crystallization buffer containing the protein in the selected concentration were added to the lipids. Prepared samples were centrifuged for 1 hour at $+23^oC$ and 5000-10000 rpm in Centrifuge 5417R, Eppendorf. For more homogeneous cubic phase formation

the orientation of the probe in the centrifuge was varied during the centrifugation. Centrifugation was performed in following sequence: 5000 rpm (10 min) - 6000 rpm, $+90^o$ rotation of the sample (10 min) - 7000 rpm, $+90^o$ (1000 min) - ... - 10000 rpm (10 min). After centrifugation tubes were closed to avoid the evaporation of the water and placed in the thermostat at $+23^oC$ for 20-24 hours. During this time the lipidic cubic phase was formed and protein was reconstituted in it.

Crystallization was initiated by dry salt addition (0.4-2.0 M). Dry salt was curried through the sample from top to the bottom by centrifugation at varying rotation speed according to the scheme described above. After that the samples were placed in thermostat at $+23^oC$ where they were isolated also from mechanical perturbations and light. First crystals were observed in the cause of 1-3 weeks (depending on crystallization conditions).

2.9.4 Screening of the crystallization conditions

Screening of the multiple parameters of the system used for crystallization was performed to elucidate the conditions of crystal growth and later optimize the found conditions. Screening was performed on the following parameters:
monoglycerid type used for cubic phase formation (MO, MV, MP),
ratio between the monoglycerid and water amount (1:1, 1:1.2, 1:1.4; 1:1.6, 1:1.8; 1:2; 1:2.5),
concentration of the protein (from 12 to 30 mg/ml),
crystallization buffer content - concentration of NaCl,
concentration and content of phosphate salt (Na/Na Pi or Na/K Pi),
pH of the buffer (5.0-8.0),
precipitant salt content (NaCl, Na/Na Pi, Na/K Pi were used) - this defines final pH in the crystallization probe,
concentration of octylglucoside (OG) (0.8-1.2%).

2.9.5 Dissolution of the cubic phase

After the crystals were grown in the probe the cubic lipidic phase should be detached from the crystal. It is not possible to use the crystal with attached cubic phase for data collection because the cubic phase has its own diffraction which would be overlaid onto the diffraction pattern from crystal. Crystals are very fragile objects as far as they contain a big percentage of water including small molecules present in the probe, also the interactions between the proteins are not very strong. Fortunately it is possible to find conditions which do not influence the quality of the crystals and allow to dissolve the cubic lipidic phase. For cubic phase dissolution crystals were placed in the following buffer: 3 M Na/Na Pi, pH=5.1, 0.1% OG. Two days after crystals were subjected to this buffer, cubic lipidic phased was dissolved and crystal could be used for diffraction experiments.

2.9.6 Selection of the crystals for the measurements at synchrotron

Preliminary tests of crystals were performed on the laboratory diffractometer with the rotating anode GX21, Fa. Elliot Marconi Avionics. Wavelength of the X-ray radiation was $\lambda = 1.542\mathring{A}$ (CuK_α). Working parameters of diffractometer - $U = 40kV$, $I = 55mA$, beam collimation $260\mu m \times 280\mu m$. For the registration of he diffraction Image-Plate Detector MAR2000 (MAR Research) was used.

The quality of the crystal diffraction was determined from one or two diffraction images, which were collected in oscillation mode. The oscillation angle was 1^o and time of acquisition - 1-1.5 hours. Crystals were first 'fished' out of the crystallization container by a cryoloop of a size

that is matched the crystal size (cryoloop should be somewhat bigger than the crystal to avoid the mechanical distortions of the crystal during the following freezing). Crystals were frozen in a nitrogen cryostream at 100 K (Oxford Cryostream). The buffer of the crystal do not contain any special cryoprotectors. The crystal freezing should be instantaneous ('flash' cooling). This is performed by placing crystal on the goniometer head with the cryostream being blocked and instantaneous opening of the cryostream. Images are collected with orientation of the crystal plate perpendicular to the beam and then 90^o apart.

A crystal which showed a good diffraction was frozen in liquid nitrogen and stored for synchrotron experiments.

3 X-ray diffraction method

3.1 X-ray diffraction by the crystals

X-ray diffraction method allows to obtain the structural information on the atomic resolution scale (Giacovazzo et al. 1992 [60]). The physical basis for that is the interaction of the X-ray beam with the periodic atomic structures.

The interaction of the X-ray with matter occurs by means of two processes:

- 1. photons of the incident beam are deflected without a loss of energy - scattered radiation, $\lambda = const$ or with a small loss of energy - Compton radiation $\lambda_{final} > \lambda_{initial}$
- 2. photons are absorbed by the atoms of the target

For the structure investigation by the means of X-ray diffraction method only the first process is important.

The scattering of the incident beam by the charged particle with electric charge e and mass m in the direction defined by the angle 2θ to the incident beam is given by Tompson formula

$$I_{eTh} = I_i \frac{e^4}{m^2 r^2 c^4} \frac{1 + cos^2 2\theta}{2} \tag{2}$$

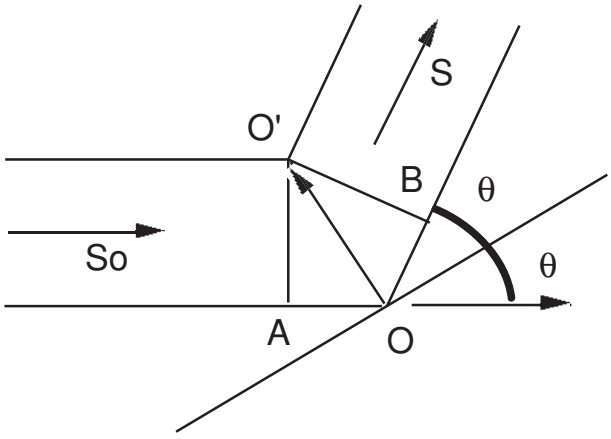

Figure 18: Diffraction of the scattered X-ray wave from two point object

In the case of the many scattering centers total scattering in a certain direction is given by the interference of the scattered waves. Two scattering centers at positions O and O' being excited by the plane wave become the sources of the secondary spherical waves (Figure 18). One can determine the phase difference of the waves scattered in the direction s. If s_o is the incident beam direction and r is a vector between O and O', then

$$\delta = \frac{2\pi}{\lambda}(\mathbf{s} - \mathbf{s}_o)\mathbf{r} = 2\pi\mathbf{qr} \tag{3}$$

where

$$\mathbf{q} = \frac{(\mathbf{s} - \mathbf{s}_o)}{\lambda} \tag{4}$$

The modulus of q equals to

$$q = \frac{2sin\theta}{\lambda} \tag{5}$$

For the N point scatterers along the path of the incident plane wave

$$F(\mathbf{q}) = \sum_{j=1}^{N} A_j exp(2\pi i \mathbf{q}\mathbf{r}_j) \tag{6}$$

where A_j is the amplitude of the wave scattered by the j-th scatterer. Or relating the scattering to the Thompson scattering with $f^2 = I/I_{eTh}$

$$F(\mathbf{q}) = \sum_{j=1}^{N} f_j exp(2\pi i \mathbf{q}\mathbf{r}_j) \tag{7}$$

This formula can be rewritten to the case of continuum scatterer. The wave scattered to the element $d^3\mathbf{r}$ containing $\rho(r)d^3r$ electrons is given by $\rho(r)d^3\mathbf{r}exp(2\pi i\mathbf{q}\mathbf{r})$. The total amplitude of the scattered wave will be

$$F(\mathbf{q}) = \int_V \rho(r) exp(2\pi i \mathbf{q}\mathbf{r}) d^3\mathbf{r} \tag{8}$$

This is a three dimensional Fourier transform of the electron density $\rho(r)$. This Fourier transform relates real space of the vector r and reciprocal space of the vector q. From the theory of the Fourier transforms

$$\rho = \int_{V*} F(\mathbf{q}) exp(-2\pi i \mathbf{q}\mathbf{r}) d^3\mathbf{r} \tag{9}$$

And this is how the knowledge of the structure factors of the scattered waves defines the electron density. Let us now consider the case of the molecules packed in the crystal. For the simplicity the one dimensional case. In this case the wave scattered from the first unit cell will be simply $F(\mathbf{q})$. The wave scattered from the N-th unit cell will be $F(\mathbf{q})exp(2\pi i(n-1)\mathbf{a}\mathbf{q})$. Hence the total wave scattered is

$$\sum_{n=1}^{T} F(\mathbf{q}) exp(2\pi i(n-1)\mathbf{a}\mathbf{q}) \tag{10}$$

The wave from each unit cell is out of phase with its neighbors by an amount of $2\pi\mathbf{a}\mathbf{q}$. When T becomes large the total wave becomes of the same order of magnitude as $F(\mathbf{q})$ except for the case when the phase difference scattered by the successive unit cells is equal to an integral multiple of 2π, i.e. $\mathbf{a}\mathbf{q} = h$, where h is integer. Under this circumstances a significantly scattered wave is formed proportional to $TF(\mathbf{q})$. Consequently the scattering will be observed only for a particular direction $\mathbf{q}$ which satisfies the diffraction conditions

$$\mathbf{a}\mathbf{q} = h \tag{11}$$

$$\mathbf{b}\mathbf{q} = k \tag{12}$$

$$\mathbf{c}\mathbf{q} = l \tag{13}$$

These equations are known as the Laue equations.

Equation (7) becomes now a definition of the coefficients in the Fourier series of the periodic function $\rho(\mathbf{r})$. So that

$$\rho(r) = \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{hkl} exp(-2\pi i(hx + ky + lz)) \tag{14}$$

where the sum is hold over all possible Miller indexes (h,k,l) combinations. Here $\mathbf{r}$ was substituted with the coordinates x, y, z and Laue equations were used.

The structure factor for a particular reflection F_{hkl} can be represented by its amplitude and phase. $\mathbf{F}_{hkl} = F_{hkl}exp(i\alpha_{hkl})$. In the recording of the diffraction patterns only the intensity of the scattered light rays are monitored $I_{hkl} = F_{hkl}F_{hkl}* = F_{hkl}^2$. And so only the amplitudes F_{hkl} are measured. All the information about the phases is lost.

One should know also the phases of the individual structure factors to define the $\rho(r)$ function . There are several ways to overcome the central problem of the crystallography - phase problem.

3.2 Solution of the phase problem

There are different methods by which the phase problem could be overcome (Taylor et al. 2003 [61]). This includes approaches which demand additional experiments or relay only to the observed I_{hkl} recorded for the investigated structure. This approaches include heavy atom isomorphous replacement in which a heavy atom is introduced into a light atom structure and is used as a marker atom to provide a phase information and anomalous scattering in which phase information is obtained from the information contained in the scattering by the atom whose natural absorption frequency is close to the wavelength of the incident radiation. Whereas the Patterson summation and direct methods relay on mathematical relationships between the reflections that can be used to provide the phase information. Furthermore the method of molecular replacement which uses the a priori information of a homologous structure can provide phase information. A short introduction to these methods used in the current work is given below.

3.2.1 Molecular replacement

The molecular replacement method is a technique which could supply the initial phases to the F_{hkl} reflections by positioning a homologous model of the investigated structure in the elementary unit cell of the crystal. This positioning takes into account all the copies of the molecule present in the unit cell of the crystal which are related to each other by noncrystallographic symmetry operators and also the crystallographic symmetry operators. The problem consists of finding all the operators which define the orientation of the subunits in the crystal. If the equivalent points of the two molecules are defined by vectors X_1 and X_2, then $X_2 = CX_1 + d$, where C is a rotation matrix and d is vector defining the translation. This operators of rotation and translation are searched in two stages. Lets consider the rotation first. To elucidate the operator of the rotation one should make use of the Patterson function which is defined as the convolution function of the electron density with itself. Patterson function can be calculated from the observed intensities by equation 15

$$P(uvw) = \frac{2}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{hkl}^2 cos2\pi(hu + hv + lw) \tag{15}$$

Given the discreet positions of the atoms the Patterson function will consists of the vectors which pairwise relate these positions. To define the orientation of the single subunit one should consider the set of vectors called the self-vector set. This set represents intramolecular vectors which all should lie within the largest intramolecular distance r from the origin of the Patterson function. The relative orientation of the molecules can be determined by comparing the self-Patterson functions which maybe considered to be within the sphere of radius r. As a criterion for the correspondence of the self-vector sets at different orientation following criteria could be chosen. It is based on the comparison of the values of the product function R where

$$R = \int_U P_2(X_2)P_1(X_1)dX_1 \tag{16}$$

for the Patterson P_1 and the rotated Patterson P_2 within the volume U. This will have a maximum value when the two self-vector sets are equivalently oriented. R is known as the rotation function (Navaza et al 2001 [62]). The rotation function could be expanded in the reciprocal space

The Patterson function of the first molecule is given by

$$P(X_1) = \frac{1}{V}\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 cos(2\pi \mathbf{h}\mathbf{X_1}) \tag{17}$$

and that of the second by

$$P(X_2) = \frac{1}{V}\sum_{\mathbf{p}} |F_{\mathbf{p}}|^2 cos(2\pi \mathbf{p}\mathbf{X_2}) \tag{18}$$

where $\mathbf{X_1}$ and $\mathbf{X_2}$ are the Patterson space vectors of peaks for molecules 1 and 2 such that $\mathbf{X_2} = C\mathbf{X_1}$ and the $\mathbf{h}$ and $\mathbf{p}$ are reciprocal space vectors so that $\mathbf{p} = C^T\mathbf{h'}$ where C^T is transpose of C and $\mathbf{h'}$ is a non-integral reciprocal lattice vector. Therefore we may rewrite

$$P(X_2) = \frac{1}{V}\sum_{\mathbf{p}} |F_{\mathbf{p}}|^2 cos(2\pi \mathbf{h'}\mathbf{X_1}) \tag{19}$$

Now the rotation function

$$R = \int P(\mathbf{X_2})P(\mathbf{X_1})dX_1 = \frac{1}{V^2}\sum_{\mathbf{p}}\sum_{\mathbf{h}} |F_{\mathbf{p}}|^2|F_{\mathbf{h}}|^2 \int_V cos(2\pi[\mathbf{h}+\mathbf{h'}]\mathbf{X_1})d\mathbf{X_1} \tag{20}$$

The integral is an interference function whose value maybe written as $U/V|G_{\mathbf{h},\mathbf{h'}}|$. If we assume that the protein molecule is approximately spherical then

$$|G_{\mathbf{h},\mathbf{h'}}| = \frac{3(sin(2\pi\mathbf{Hr}) - 2\pi\mathbf{Hr}cos(2\pi\mathbf{Hr}))}{(2\pi\mathbf{Hr})^3} \tag{21}$$

where $\mathbf{H} = \mathbf{h} + \mathbf{h'}$ Given in such form the rotation function still demands very time consuming computation. Some assumptions can be made to simplify the calculation. As the rotation function is approximately proportional to $|F|^4$ the smaller intensities will make little contribution to the sum and can be generally omitted. The interference function is significant only for $\mathbf{Hr}$ within the first zero of the function therefore only a cube in reciprocal space for which $|\mathbf{h}+\mathbf{h'}|$ is less than 1.5 should be considered. It is usually calculated by sampling the rotations by Eulerian angles $(\theta_1, \theta_2, \theta_3)$.

High resolution is required to define the rotation matrix precisely. The resolution which is required can be defined from the condition that rotation which moves the most reciprocal lattice point (h_{max}) through one reciprocal lattice translation is given approximately by $2sin^{-1}(1/2h)$.

Given the experimentally measured F_{hkl} values and the homologous model one could calculate the sum based on the amplitudes (20) for each plausible orientation of the model in order to find out the correct orientation in the crystal by examining the minimum of the sum.

With the knowledge of the molecular orientation determined from the self-Patterson, the translation vector $\mathbf{d}$ may be investigated by considering the cross-Patterson. The translation of the molecule in the unit cell is defined relative to the symmetry axis. The problem of finding the position of the molecule in a unit cell is to find its position perpendicular to any rotation axis. For each axis it is a two-dimensional problem. By the analogy with the rotation function the translation function (Tong et al. 2001 [63]) can be introduced by

$$T(\mathbf{t}) = \int P(\mathbf{u})P(\mathbf{u},\mathbf{t})d\mathbf{u} \tag{22}$$

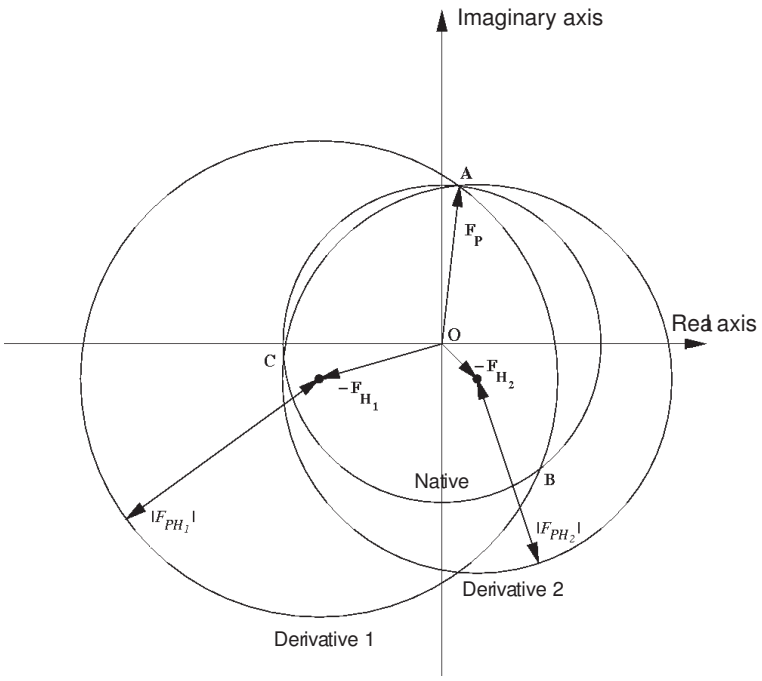

Figure 19: Harker construction for multiple isomorphous replacement

The maximum of this function finds the best fit of the computed and correctly oriented cross-Patterson $P(\mathbf{u}, \mathbf{t})$ with the Patterson $P(\mathbf{u})$. In the case of several copies of the subunits present in the unit cell one can position each of them step by step. For the putative structure correctly oriented and positioned in the unit cell the structure factors can be calculated. Such a model provides a set of phases which can be used in the conjunction with the observed structure amplitudes to obtain the electron density map.

3.2.2 Isomorphous replacement

The method of isomorphous replacement is the most widely used technique used for phase determination (Carvin et al. 2001 [64]). It uses the difference in scattering between two isomorphous atomic structures which differ in presence of some heavy atoms. In order to obtain phase information from isomorphous replacement it is necessary to locate the atomic position of the heavy atom scatterer.

Lets assume that the positions and occupancy of the site of the heavy atom binding have been determined as accurately as possible. If the structure factors of the protein crystal are $\mathbf{F}_P(\mathbf{h})$, of the isomorph are $\mathbf{F}_{PH}(\mathbf{h})$ and of the heavy atom $\mathbf{F}_H(\mathbf{h})$, then

$$\mathbf{F}_H(\mathbf{h}) = \mathbf{F}_{PH}(\mathbf{h}) - \mathbf{F}_H(\mathbf{h}) \tag{23}$$

In practise one can measure the structure amplitudes F_P and F_{PH}. From the heavy atom parameters, the corresponding structure factor $\mathbf{F}_H(\mathbf{h})$ can be calculated. Then phase ϕ of $\mathbf{F}_P(\mathbf{h})$ can be derived from the set of phase circles (Figure 19). From the origin O vector OA is drawn equal to $-\mathbf{F}_H$. Circles of radius F_P and F_{PH} are drawn about O and A respectively. The intersection of the circles at B and B' define two possible phase angles for F_P. This ambiguity may in principal be resolved by using a second heavy-atom isomorphous derivative or by utilizing

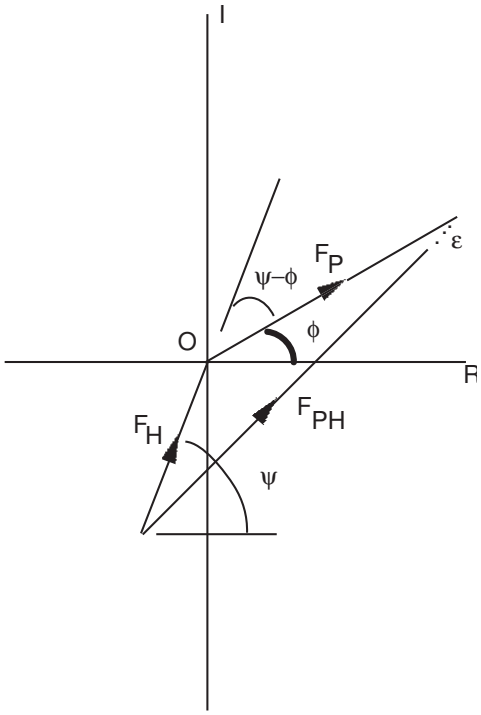

Figure 20: Introduction of the lack of closure of the triangle error to evaluate the phase probabilities

the anomalous scattering effect. The second isomorph will add to the picture third circle which will unambiguously define the correct phase.

In practice there will be errors in the observed amplitudes F_P and F_{PH} and in the heavy atom parameters. Also the isomorphism may be imperfect. As a result the intersection of the three phase circles may not coincide. The method of Blow & Crick was introduced to account for all this possible errors in the data. The phase angle ϕ can never be determined with complete certainty. Rather, there is a finite probability $P(\phi)$ that the arbitrary phase angle ϕ is the correct phase of $\mathbf{F}_P$. All errors are assumed to be associated with $|F_{PH}|_{obs}$. Owing to errors, the triangle formed by F_P, F_{PH} and F_H fails to close (Figure 20). The lack of closure error ϵ is a function of the calculated phase angle ϕ_P:

$$\epsilon(\phi_P) = |F_{PH}|_{obs} - |F_{PH}|_{calc} \tag{24}$$

Following Blow & Crick, if E is the r.m.s. error associated with the measurements., and the distribution of errors is assumed to be Gaussian, then the probability $P(\phi)$ of the phase ϕ being the true phase is

$$P(\phi) = Nexp\left(-\frac{\epsilon^2}{2E^2}\right) \tag{25}$$

where N is a normalizing factor such that the sum of all probabilities is unity. The two most probable phase angles ($\phi = \phi_1$ and $\phi = \phi_2$) are the alternative phases of F_P for which the phase triangle is closed. Individual probability distributions for the additional heavy-atom derivatives are derived in an analogous manner and may be multiplied together to give an overall probability distribution. The phases has the minimum error when the best phase ϕ_{best}, i.e. The centroid of

the phase distribution

$$\phi_{best} = \int \phi_P P(\phi_P) d\phi_P, \tag{26}$$

is used instead of the most probable phase. The quality of the phases is indicated by the figure of merit m, where

$$m = \frac{\int P(\phi_P) exp(i\phi_P) d\phi_P}{\int P(\phi_P) d\phi_P} \tag{27}$$

Normally phases are first calculated from the knowledge about the initial set of heavy-atom positions. To get the more accurate heavy-atom parameters it is necessary to refine these parameters. This can be achieved through the minimization of

$$\sum_{\mathbf{S}} \frac{\epsilon^2}{E} \tag{28}$$

where E is the estimated error $(\simeq)\langle(|F_{PH}|_{obs} - |F_{PH}|_{calc})\rangle)^2$. Phase refinement is generally monitored by three factors:

$$R_{Cullis} = \frac{\sum ||F_{PH} + F_P| - |F_H|_{calc}|}{\sum |F_{PH} - F_P|} \tag{29}$$

for noncentrosymmetric reflections only; acceptable values are between 0.4 and 0.6;

$$R_{Kraut} = \frac{\sum ||F_{PH}|_{obs} - |F_{PH}|_{calc}|}{\sum |F_{PH}|_{obs}} \tag{30}$$

which is useful for monitoring convergence; and the

$$phasing\, power = \frac{\sum |F_H|_{calc}}{\sum ||F_{PH}|_{obs} - |F_{PH}|_{calc}|} \tag{31}$$

which should ideally be greater than 1 (if less than 1 then phase triangle can not be closed via F_H) To locate the heavy-atom site one should make use of the difference Patterson function (Matthews et al. 2001 [65]). This is the Patterson function with amplitudes $(F_{PH} - F_P)^2$. It relies on the crude assumption that the desired scattering amplitude of the heavy atoms, F_H, can be approximated by

$$|F_H| \simeq |F_{PH} - F_P| \tag{32}$$

Given the well behaved isomorphous heavy-atom derivative, and accurately measured data, experience has shown that a map with coefficients $(F_{PH} - F_P)^2$ can give an excellent representation of the desired heavy-atom-heavy-atom vector peaks.

3.2.3 Anomalous scattering

The scattering of X-rays by an isolated atom is described by the atomic scattering factor, f^0, based on the assumption that the electrons in the atom oscillate as free electrons in response to X-ray stimulation (Blundell et al. 1976 [66]). In reality, electrons in an atom do not oscillate freely because they are bound in atomic orbital. Deviation from the free-electron model of atomic scattering is known as anomalous scattering. The total atomic scattering factor, f is a sum of 'normal' and 'anomalous ' components:

$$f = f^0 + f' + if'' \tag{33}$$

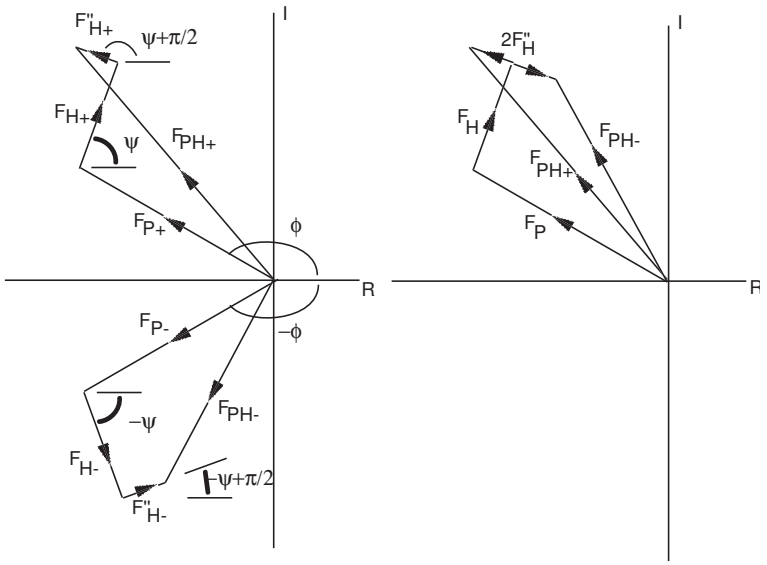

Figure 21: Anomalous difference within the Bijvoet pair

f' and f'' are expressed in electron units, as is f^0. The real component of anomalous scattering, f' is in phase with the normal scattering, f^0, whilst the imaginary component, f'', is out of phase by $\pi/2$. The imaginary component of anomalous scattering, f'', is proportional to the atomic absorption coefficient of the atom, μ_a at X-ray energy E:

$$f''(E) = \frac{mc}{4\pi e^2 \hbar} E \mu_a(E) \tag{34}$$

Anomalous scattering is present for all atomic types at all X-ray energies. However the magnitudes of f' and f'' are negligible at X-ray energies far removed from the resonant frequencies of the atom. This includes all light atoms (H,C,N,O) of biological macromolecules at all X-ray energies commonly used for crystallography. In contrast the heavy atoms often has the significant magnitudes of f' and f'' in the used range of energies. f'' can be obtained experimentally from the X-ray absorption spectra. And f' can be derived from the f'' by the Kramers-Kronig transform:

$$f'(E) = \frac{2}{\pi} P \int_0^\infty \frac{E' f''(E')}{E^2 - E'^2} dE' \tag{35}$$

where P represents the Cauchy principal value of the integral. The anomalous scattering of a heavy atom is always considerably less than the normal scattering.

Suppose that two isomorphous crystals are differentiated by N heavy atoms of position $\mathbf{r}_n$ and scattering factor $(\mathbf{f}'_n) + i\mathbf{f}''_n$. Then for the reflection hkl the calculated structure factor of the N atoms is

$$F_H(\mathbf{h}) + iF''_H(\mathbf{h}) = \sum_{n=1}^{N} f'_n(\mathbf{h}) exp(2\pi i \mathbf{h} \mathbf{r}_n) + i \sum_{n=1}^{N} f''_n(\mathbf{h}) exp(2\pi i \mathbf{h} \mathbf{r}_n) \tag{36}$$

If the heavy atoms are all of the same type, i.e. They all have the same ratio of f'/f'', then F_H and F''_H are orthogonal, and $F''_H = F_H/k$ The relation between the structure factors of the reflection hkl and its Friedel mate (reflection with inverted indexes) is illustrated in Figure 21. The Friedel's law breaks down. The differences

$$\Delta_{PH} = F_{PH+} - F_{PH-} \tag{37}$$

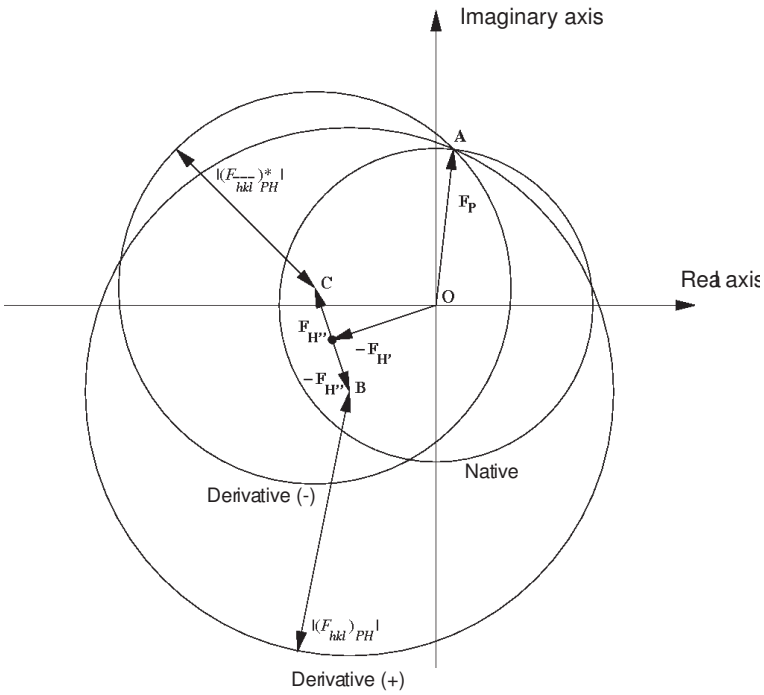

Figure 22: Harker construction for the case of anomalous scattering

can be referred as the Bijvoet differences, or simply the anomalous-scattering differences. The Harker phase circles are shown in Figure 22. In the absence of error the three phase circles will meet at a point, resolving the phase ambiguity and giving the unique solution for the phase of F_P. The anomalous scattering method complements the isomorphous replacement method.

On the average, the experimentally measured isomorphous replacement difference, $F_{PH} - F_P$, will be larger than the anomalous scattering difference, $(F_{PH+} - F_{PH-})$. But the former relies on measurements from a different crystals and is susceptible to the errors of the non-isomorphism. The latter can be obtained from the measurements on the same crystal and is not affected by non-isomorphism. As in the case of isomorphous replacement the lack of closer of the triangle can be considered and the probability of phase ϕ being the true phase of $\mathbf{F_P}$ can be estimated. The overall probability distribution obtained by combining the anomalous scattering data with isomorphous replacement data is given by

$$P(\phi) = NP_{iso}(\phi)P_{ano}(\phi) \tag{38}$$

To approximate the anomalous heavy atom scattering amplitude a relation can be used

$$|F''_H| \simeq \frac{1}{2}|F_{PH+} - F_{PH-}| \tag{39}$$

If all the heavy atoms are the same, $F_H = kF''_H$. Thus a Patterson function with coefficients $(F_{PH+} - F_{PH-})$ should also show the desired heavy-atom-heavy-atom vector peaks.

3.3 Synchrotron source radiation

The European Synchrotron Radiation Facility (ESRF) essentially comprises of three main components. The pre-injector is a 16 m long linear accelerator (linac) which produces electrons and accelerates them to an energy of 200 MeV. The booster synchrotron is a 10 Hz cycling synchrotron with a circumference of 300 m and increases the energy of the electrons to 6 GeV. In the main storage ring, which has a circumference of 844.4 m, bunches of electrons circulate at 6 GeV providing very intense X-rays at each passage inside the bending magnets or insertion devices (undulators or wigglers). The radiation is emitted in a continuous-wavelength spectrum around a characteristic wavelength λ_C (Å) which is inversely proportional to the square of the energy E (GeV) of the electron beam and to the field of the magnet B (T),

$$\lambda_C = \frac{18.65}{E^2 B} \tag{40}$$

For an ESRF bending magnet with $B = 0.8\ T$, then $\lambda_C = 0.65$ Å. The radiation is confined to a narrow cone with an opening angle of order $1/\gamma = mc^2/E$ (where m is the mass of an electron and c is the velocity of light). For a 6 GeV storage ring this angle is smaller than 20 arcsec ($\sim 80\ \mu$rad). The radiation is also polarized linearly in the orbit plane and elliptically outside the plane.

Insertion devices are comprised of an array of closely spaced magnets of opposite polarity which deflect the electrons by small amounts from a linear path, thus giving rise to synchrotron radiation at each position of deflection. There are several types of insertion devices. Multipole wigglers have somewhat smaller deflections and act as multiple sources for which the phase relationships are random, so that the total intensity is the sum of the intensities from the individual poles and the spectrum comprises a continuum of wavelengths. In an undulator, the deflection is kept within the natural angle of emission ($1/\gamma$), so that the radiation from the different poles is 'in-phase' for a particular energy (and therefore wavelength); the net intensity is then the sum of the squares of the individual amplitudes. Undulators therefore produce the X-ray sources with the highest brilliance.

Table 1

ID2	Monochromatic data collection, large unit cell	IP
ID9	Time-resolved Laue studues	CCD, IP
ID13	Micro-focus, small crystals	CCD, IP
ID14	Quadriga, principal MX station	
EH1	monochromatic	CCD
EH2	monochromatic	CCD
EH3	Large proteins and viruses, limited tunability	CCD, IP
EH4	Multiwavelength anomalous diffraction (MAD) and monochromatic data collection; very high brilliance	CCD
BM14	MAD	CCD, IP
ID29	MAD	CCD, IP

The source characteristics of the ESRF are ideally matched to a number of difficult problems in macromolecular crystallography (Lindley et al. 1999 [67]). The high brilliance and high degree of collimation allow intensity data to be collected from small weakly diffracting crystals, whilst the wavelength tunability readily enables the technique of multiple-wavelength anomalous dispersion

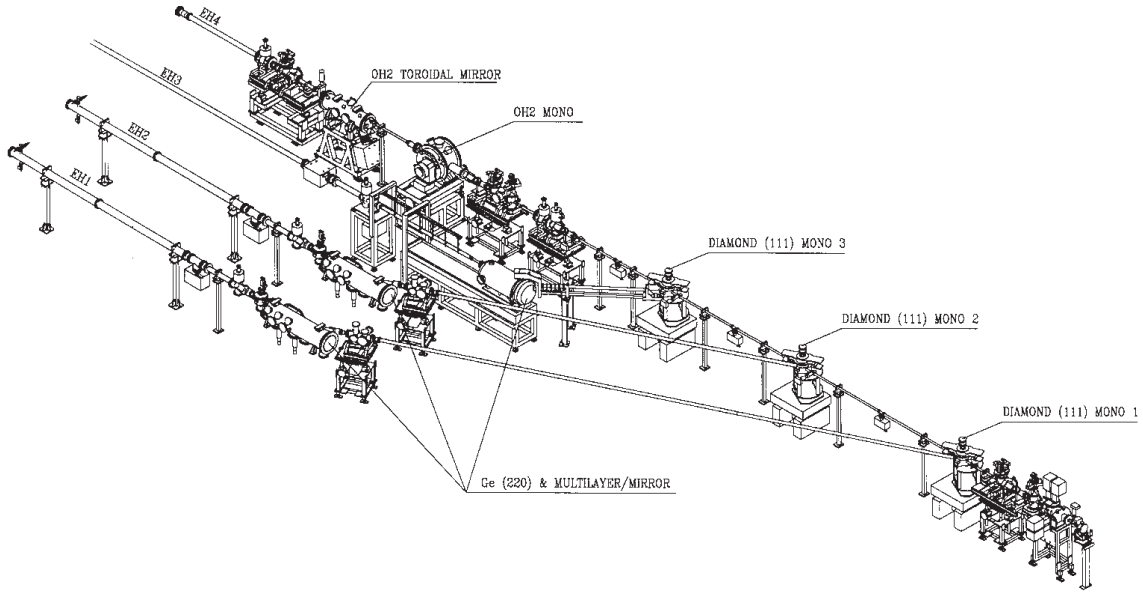

Figure 23: Schematic view of the Quadriga beamline ID14 at the ESRF. From Lindley [67]

to be used for phase determination over a wide range of absorption edges. Macromolecular crystallography at the ESRF is handled by a number of different beamlines, but the principal ones are shown in Table 1.

The layout of the typical beamline for the macromolecular crystallography is presented below. The Quadriga beamline ID14 (Wakatsuki et al., 1998) is designed to have four simultaneously operating stations: an end station EH4 equipped with a fast-scan double-crystal monochromator and three side stations EH1-3 deriving their radiation through the use of thin diamond crystals. Two undulators, each with a length of 1.65 m and minimum achievable gap of 16 mm, provide the high-brilliance X-ray source. The first undulator, with a periodicity of 42 mm, is tuneable over a wide wavelength range (7-35 keV), whereas the second, with a periodicity of 23 mm, is a single line undulator with very little tunability but optimized to give high brilliance around 13.5 keV (0.92 Å). The second undulator induces only a small heat load on the optical components and is operated in tandem with the first. Schematically the overall layout of the Quadriga beamline is shown in Figure 23.

Lets consider in details the equipment of one of these side stations - EH2. This station is conceived for standard experiments in macromolecular crystallography using a wavelength around 0.93 Å (below most interesting absorption edges). The flux is around 1.5×10^{12} photons/sec with a 200 μm beam. A thin (100 μm) synthetic diamond monochromates the white beam. Plane Ge(220) crystal as the second element diffracts the beam back almost parallel to the white beam and reduces harmonics. A 35 cm long Rh coated torroidal mirror focuses the beam and further removes higher harmonics.

The experimental hutch is shown in Figure 24. The X-ray beam exits the final collimation via a nose piece around 10mm from the sample position. The intensity monitor just before the nose piece checks the beam strength whilst the second, further upstream, allows the automatic realignment of the beam. The goniometer is the single axis ϕ spindle with motorized x, y and z axis. An ADSC Q4 CCD based detector is mounted on a rapid translation base. The detector has front active area of 188 mm square coupled to a CCD cheap with 2304 pixels of 81.6 μm

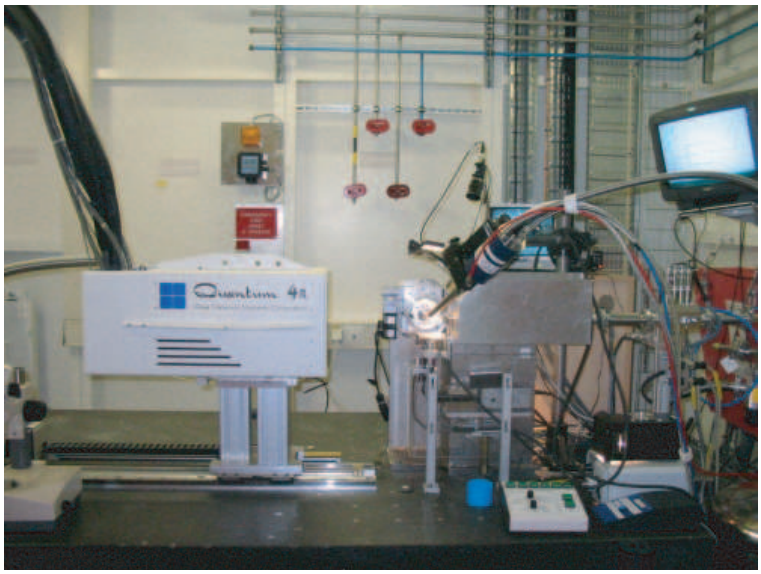

Figure 24: Beamline ID14-1 experimental hutch overview. From http://www.esrf.fr/UsersAndScience/Experiments/MX/ID14-1/

by an optical fibre taper. The beam size can be easily matched to the sample dimensions (in order to minimize the background or to take data from the several sections of the sample) using motorized slits. The beam is normally used at 100 μm square, but 50 μm to 200 μm are standard and beam down to 20 μm has been used.

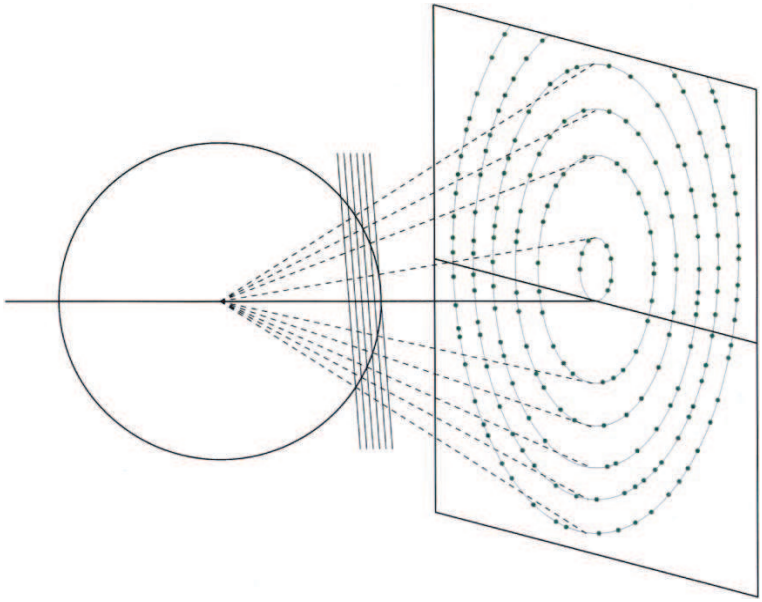

Figure 25: A still exposure with a stationary crystal contains only a small number of reflections arranged in a set of narrow ellipses. From Dauter Acta Crystallographica D. 1999

3.4 Data collection strategy

3.4.1 Rotation method

Lets consider the Ewald sphere which illustrates the Bragg's law of the diffraction in three dimensions (Blundell et al. 1976 [66]). The radiation of the wavelength λ is represented by a sphere with radius $1/\lambda$ centered on the X-ray beam. The crystal is represented by the reciprocal lattice, with its origin at the point of the Ewald sphere were the direct beam leaves it. If the reciprocal-lattice point lies on the surface of Ewald sphere the following condition equivalent to the Bragg's law is fulfilled:

$$\frac{1}{2d} = \frac{1}{\lambda} sin\theta \tag{41}$$

Therefore when a reciprocal-lattice point with the indexes (hkl) lies on the surface of the Ewald sphere, the interference condition to that particular reflection is fulfilled and it gives rise to a diffracting beam directed along the line joining the sphere center with the reciprocal lattice point on the surface. For any particular orientation of the crystal only few reflections can be in the diffracting position. The number depends on the density of the reciprocal lattice and hence on the unit cell dimensions of the crystal. For macromolecule crystals the unit cell are much larger than the wavelength of the radiation used and several reciprocal-lattice points will lie on the surface of the Ewald sphere. To observe the diffraction the reciprocal-lattice points must be moved to the surface of the Ewald sphere. For the selected wavelength (monochromatic radiation) this requires the crystal to be rotated to bring the successive reflections into diffraction. In the rotation method the crystal is rotated only about the single axis. This is the most common procedure used for recording of the diffraction data in the macromolecular crystallography.

The reciprocal lattice consists of points arranged in planes. This planes are densely populated in relation to the size of the Ewald sphere. The plane intersects with the sphere giving the small circle which projects onto the detector as an ellipse since all rays diffracting from the same plane

form a cone. When the crystal is not rotated during the X-ray exposure or rotates only very little the diffraction pattern will consist of spots arranged in a set of concentric ellipses originating from one family of the parallel planes in the reciprocal lattice. When the crystal is rotated the start and end orientations of the plane form two intersecting ellipses with the reflections recorded between them in the form of lunes. All reflections within the same lune originate from the same reciprocal lattice layer. Because reflections are arranged in families of the parallel planes there will be a family of the concentric lunes on the detector (Figure 25).

3.4.2 Crystal mosaicity and beam divergence

In practice, the incident radiation is not directed precisely along one line. The X-ray beam can be focused and collimated to be parallel within a small angle about 0.2^0 on the rotating anode source and somewhat smaller on synchrotron beamlines. Crystals also are composed of small mosaic blocks slightly misoriented with respect to one another which adds some divergence to the total rocking curve, that is to the amount of the rotation during which individual reflection diffracts. The Ewald sphere has two limiting orientations which results in a defined active width and the reciprocal lattice points can be represented by disks extended angularly, mosaicity does not extend them radially since the diffraction angle θ remains constant. The final value of the rocking curve has consequences for the diffraction pattern. In a rotation method images are exposed in a continuous series of narrow crystal rotations defined by the oscillation angle ϕ. Some reflections come into diffracting position during the one exposure and finish during the next. Consequently part of their intensity will be recorded on one image and another part on the next. If the rotation range per image is small compared with the rocking curve, individual reflections can be spread over several images. Such reflections are termed partially as opposed to fully recorded, the latter has all their intensity present on the single image. If the mosaicity increases the lunes become wider because there are more partial reflections. When the mosaicity reaches the value of the rotation range, there are almost no fully recorded reflections at all. Usually the diffraction images are collected with so called wide slicing when the rotation range is wider than the rocking curve. This saves the time for data collection. As a drawback wide slicing do not allow a good estimation of the three dimensional profile of the reflection which can be done when the finite slicing is used.

For wide slicing a few factors should be taken into account for selection of the rotation range per single exposure. The rotation range should be small enough to avoid the overlap of neighboring lunes. A simple formula can be derived and used to estimate the maximum permitted rotation range:

$$\Delta\phi = 180d/\pi a - \eta \tag{42}$$

where the factor $180/\pi$ converts radians to degrees, η is the angular width of the reflection, d is high resolution limit and a is the length of the primitive unit cell dimension along the direction of the X-ray beam (Figure 26). This is not a very strict requirement.

3.4.3 Crystal-to-detector distance

The larger the crystal-to-detector distance the better the signal-to-noise ratio in the recorded diffraction pattern, since the background area decreases with the square of the distance. The distance therefore should be adjusted to match the maximum resolution limit of the diffraction. To define this resolution limits usually two images 90^0 apart should be inspected as some crystals display anisotropy and diffract further in one direction that in the other.

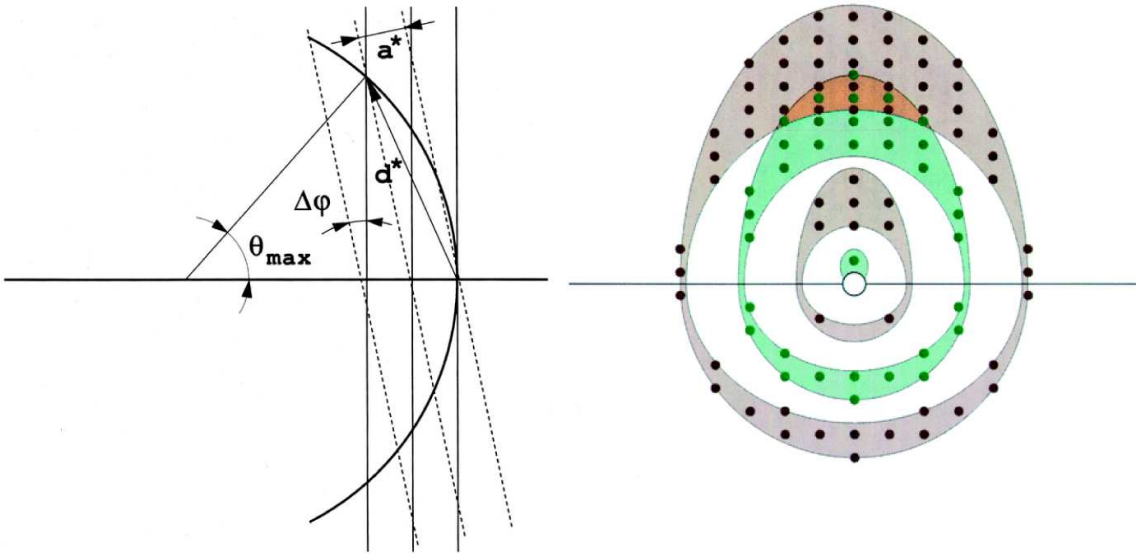

Figure 26: Origin of the reflections zones overlap and calculation of the limiting rotation range. From Dauter Acta Crystallographica D. 1999

3.4.4 Wavelength

At the most synchrotrons wavelength below 1 Å are used as this minimizes absorption of the radiation by the crystal and it's mother liquor and the air scatter. Short wavelength are advantageous for collecting very high resolution data since it decreases the maximum recordable 2θ angle and minimizes the blind region. The blind region exists since not all the reflections can be brought in diffracting positions. Reflections lying close to the rotation axis will never cross the Ewald sphere even after 360^0 rotation. The advantage of longer wavelength is a stronger interaction with crystals, leading to enhanced intensity of diffracted rays.

3.4.5 Total rotation range

Selection of the total rotation range appropriate for the crystal symmetry is the most important factor influencing the completeness of the data. Available beam time can be limited, especially at a synchrotron, and minimization of the time of the experiment is taken into account in the praxis of data collection. Analysis of the symmetry of the crystal in the relation to the geometry of the rotation method allows to specify conditions leading to the minimal complete dataset. In the cases when the maximum of the resolution should be obtained even frozen crystals can not survive the exceedingly intense radiation and to reach high completeness as soon as possible is beneficial. For example when the orthorhombic crystal is rotated around any of its two fold axes the required rotation range is 90^0. When the orthorhombic crystal is oriented around the arbitrary axis which is not in the symmetry plane, more than 90^0 of rotation is required. Collection of 2 images separated by 90^0 allows to determine the orientation of the crystal and then to decide which minimal range of the rotation can give the complete dataset.

3.4.6 Exposure time

Exposure time is a factor which most strongly influences the reflection intensities. In principal, the higher the intensities are the higher is the signal-to-noise ratio and therefore the data quality is

better. This is a simple consequence of counting statistics. Doubling the intensity enhancing the signal-to-noise ratio by $2^{1/2}$. But there are also restrictions for the exposure time. The dynamic range of the detector should be taken into account. Each detector has a certain saturation level. Pictures which have more counts per pixel are overloaded and can not be used for accurate evaluation of that reflection intensity. The contrast between the intensities of the strongest and weakest reflections is very large. It is therefore inevitable that if the exposure is adjusted to adequately measure the intensities of the weak reflections, some of the strongest one will be overloaded. They should be measured in a separate rotation path with shorter exposures adjusted to adequately cover the strong reflections below the overload limit. If the intensity is very high and the speed of the rotation is limited then instead of the decreasing time of the exposure one could attenuate the incident radiation beam intensity. The second pass does not need to extent to the same resolution limit, it is sufficient to cover the overloads zone. The distance can be increased as well as the rotation range per image.

3.5 Integration of the diffraction data

Data integration is the process of obtaining estimates of the diffracted intensities from the raw images recorded by the X-ray detector (Leslie et al. 2001 [68]). To integrate the data a number of parameters should be determined. The most important are unit cell parameters which should be determined to an accuracy of the tenth of the percent. Post-refinement procedures which make use of the estimated ϕ (crystal rotation angle) centroids provides more accurate estimates than methods based on the spot positions. Data from widely separated regions of the reciprocal space should be included in order to determine all unit cell parameters accurately. The crystal orientation also should be known to the accuracy which corresponds to a few percent of the reflection width. It is important to allow for movement of the crystal by continuously updating the crystal orientation during the integration. Non-orthogonality of the incident beam and the rotation axis or an off-centered crystal will also give rise to apparent changes in crystal orientation with spindle rotation. The crystal-to-detector distance, the detector orientation and the direct beam position must be refined and continuously updated during the integration using observed spot positions. The crystal-to-detector distance can vary during the data collection if the crystal is not exactly centered on the rotation axis, and the direct beam position on the detector can move after a beam refill at synchrotron. With appropriate care, it is normally possible to predict reflection positions on the detector to an accuracy of 20-30 μm or a fraction of the pixel size.

There are two quite distinct procedures available for the determining the integrated intensities: summation integration and profile fitting. Summation integration involves simply adding the pixel values for all pixels lying within the area of a spot and then subtracting the estimated background contribution of the same pixels. Profile fitting assumes that the actual spot shape or profile is known and the intensity is derived by finding a scale factor which, when applied to a known profile, gives the best fit to the observed spot profile. In practice profile fitting requires to separate steps: determination of the standard profiles and the evaluation of the profile fitting intensities.

X-ray scattering from air, the sample holder and the specimen itself give rise to a general background in the images which have to be subtracted in order to obtain the Bragg intensities. Background is determined using pixels immediately adjacent to the spot. The pixels to be used for the determination of the background (background pixels) and those to be used for evaluating the intensities (peak pixels) are defined using a 'measurement box' (Figure 27). This is the rectangular box of pixels which is centered on the predicted spot position. Each pixel within a box is classified as being a background or a peak pixel. It is generally assumed that the background can be adequately modelled as a plane, and the plane constants are determined

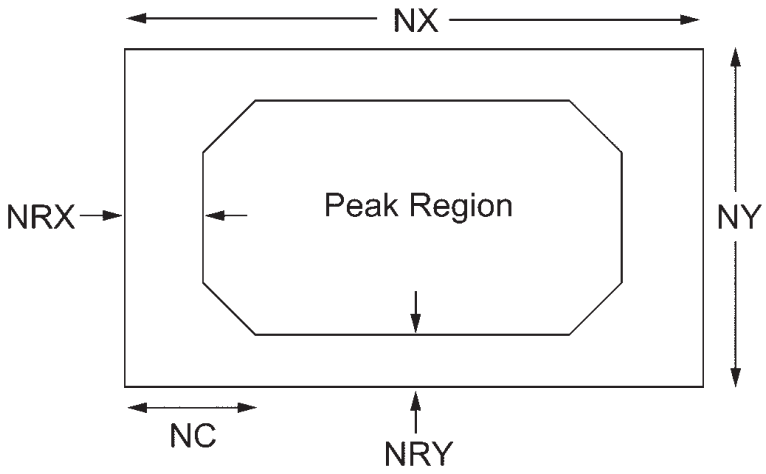

Figure 27: Definition of the peak and background regions for summation integration and profile fitting. Listed are the parameters to define the profile. From Leslie [68]

using the background pixels. This allows a background to be estimated for the peak pixels so that the background-corrected intensity can be calculated.

3.5.1 Integration by simple summation

The background plane constants are determined by minimizing

$$R_1 = \sum_{i=1}^{n} w_i(\rho_i - ap_i - bq_i - c)^2 \tag{43}$$

where ρ_i is the total number of counts at the pixel with coordinated (p_i, q_i) with respect to the center of the measurement box and the summation is over the n background pixel, w_i is a weight which ideally should be the inverse of the variance of ρ_i. The summation-integration intensity I_s is given by

$$I_s = \sum_{i=1}^{m}(\rho_i - ap_i - bq_i - c) \tag{44}$$

where the summation is over the m pixels in the peak region of the measurement box. If the peak region has mm symmetry this simplifies to

$$I_s = \sum_{i=1}^{m}(\rho_i - c) \tag{45}$$

To evaluate the standard error

$$I_s = \sum_{i=1}^{m} \rho_i - (m/n) \sum_{j=1}^{n} \rho_j \tag{46}$$

where the second summation is over the n background pixels.

The variance in I_s is

$$\sigma^2_{I_s} = \sum_{i=1}^{m} \sigma_i^2 + (m/n)^2 \sum_{j=1}^{n} \sigma_j^2 \tag{47}$$

From Poisson statistic, this becomes

$$\sigma^2_{I_s} = \sum_{i=1}^{m} G\rho_i + (m/n)^2 \sum_{j=1}^{n} G\rho_j = G[I_s + I_{bg} + (m/n)(m/n) \sum_{j=1}^{n} \rho_j] \tag{48}$$

where I_{bg} is the background summed over all peak pixels. We can also write

$$I_{bg} \simeq (m/n) \sum_{j=1}^{n} \rho_j \tag{49}$$

Then

$$\sigma^2_{I_s} = G[I_s + I_{bg} + (m/n) I_{bg}] \tag{50}$$

This expression shows the importance of the background (I_{bg}) in determining the standard error in the intensity.

3.5.2 Integration by profile fitting

In order to apply the profile-fitting method, the first requirement is to derive a 'standard ' profile which accurately represents the true reflection profile. Empirical profiles are defined by summing many different spots. The optimum profile is that which provides the best fit to all the contributing reflections, i.e. that which minimizes

$$R_2 = \sum_h w_j(h)[K_h P_j - \rho_j(h)_{corr}]^2, \tag{51}$$

where P_j is profile value for the j th pixel, $rho_j(h)_{corr}$ is the observed background corrected counts at that pixel for reflection h, K_h is a scale factor and $w_j(h)$ is a weight for the j th pixel of reflection h. The summation extends over all reflections contributing to the profile. The weight $w_j(h)$ is given by

$$w_j(h) = 1/\sigma^2_{hj} \tag{52}$$

and, from Poisson statistics, the expectation value of the counts at pixel j is given by

$$\sigma^2_{hj} = K_h P_j + (a_h p_j + b_h q_j + c_h) \tag{53}$$

Given an appropriate standard profile, the reflection intensity for fully recorded reflections is evaluated by determining the scale factor K and background-plane constants a, b and c which minimize

$$R_3 = \sum w_i (K P_i + a p_i + b q_i + c - \rho_i)^2 \tag{54}$$

where the summation is over all valid pixels in the measurement box. A before the expectation value of the counts at pixel i is given by

$$\sigma^2_i = a p_i + b q_i + c + J P_i \tag{55}$$

The summation-integration intensity is used to evaluate the scale factor J using

$$I_s = J \sum_i P_i \tag{56}$$

Minimizing R_3 allows to determine a, b, c and K. Profile fitted intensity I_P is given by

$$I_p = K \sum_i P_i \tag{57}$$

The standard error in the profile-fitted intensity is given by

$$\sigma_{I_p}^2 = \sigma_K^2 (\sum_i P_i)^2 \tag{58}$$

Minimizing R_3 with respect to the scale factor K then gives

$$I_p = K \sum P_i = \left(\sum w_i P_i \rho_i - a \sum w_i P_i p_i - b \sum w_i P_i q_i - c \sum w_i P_i \right) \times \\ \times \left(\sum P_i / \sum w_i P_i^2 \right) \tag{59}$$

where all summations are over the peak region only. For very strong reflections, the background level is very small, and reduces to

$$I_p \simeq \sum w_i P_i \rho_i (\sum P_i / \sum w_i P_i^2) \tag{60}$$

the weights are given by

$$w_i \simeq 1/JP_i \tag{61}$$

Substituting for w_i gives

$$I_p \simeq \sum \rho_i \tag{62}$$

This shows that for correctly weighted profile fitting, the profile-fitted intensity reduces to the summation-integration intensity for very strong intensities.

For very weak reflections, all pixels will have very similar counts and, therefore, all the weights will be the same. Equation 59 reduces to

$$I_p = \sum p_i (\rho_i - a p_i - b q_i - c)(\sum P_i / \sum P_i^2) \tag{63}$$

The last term in this equation depends only on the shape of the standard profile. This shows that the intensity is a weighted sum of the individual background-corrected pixel counts (rather than a simple unweighted sum, as is the case for summation integration).

For very weak reflections, where all the weights w_i are approximately the same, the variance in I_P is given by

$$\sigma_{I_p}^2 = \sum \mathrm{Var}(\rho_i - a p_i - b q_i - c) P_i^2 (\sum P_i / \sum P_i^2)^2 \tag{64}$$

Assuring a flat background and very weak intensity, from Poisson statistics

$$\mathrm{Var}(\rho_i - a p_i - b q_i - c) \simeq G \rho_i \tag{65}$$

and, as ρ_i has approximately the same value ρ for all pixels,

$$\sigma_{I_p}^2 = G\rho \sum P_i^2 (\sum P_i / \sum P_i^2)^2 = G\rho [(\sum P_i)^2] / \sum P_i^2 \tag{66}$$

The variance in the summation-integration intensity is simply

$$\sigma_{I_s}^2 = Gm\rho \tag{67}$$

The ratio of the variances is thus

$$\sigma_{I_s}^2 / \sigma_{I_p}^2 = m \sum P_i^2 / (\sum P_i)^2 \tag{68}$$

For a typical spot profile, the right-hand side (which depends only on the shape of the standard profile) has a value of 2, showing that profile fitting can reduce the standard error in the integrated intensity by a factor of $2^{1/2}$.

Profile-fitted intensities has also some other benefits.If adjacent spots are not fully resolved, there will be a systematic error in the integrated intensity which will be largest for weak spots which are adjacent to very strong spots. However, the profile-fitted intensity will be affected less than the summation-integration intensity because the peripheral pixels (where the influence of neighboring spots is greatest) are down-weighted relative to the central pixels (where the neighbors will have least influence). Further steps can be taken to minimize the errors caused by overlapping spots. Firstly, when forming the standard profiles, reflections are only included if they are significantly stronger than their nearest neighbors. This will minimize the errors in the standard profiles. Secondly, when evaluating the profile-fitted intensity of a particular reflection, pixels can be omitted if they are adjacent to a pixel which is part of a neighboring spot (rather than having to be part of that spot). Because of the limited dynamic range of current detectors, it is common for many low-resolution spots to contain saturated pixels. Providing the saturation level of the detector is known, such pixels can simply be excluded from the profile fitting, allowing a reasonable estimate of the true intensity (except when the majority of the pixels are saturated). All the described procedures to evaluate the intensities of reflections and their standard deviations are implemented in the programm MOSFLM written by Andrew Leslie 2002 [69]

3.6 Scaling of the data

All the integrated intensities should be put on the common scale. So the list of structure amplitudes $|\mathbf{F}|$ for a symmetry-unique set of the reflections can be derived. This includes the placing reflection on the same scale, rejection of outliers, analysis for systematic errors, averaging redundant measurements of intensity and estimating the standard deviations, computing $|F|$ and σ_F from intensities (Evans et al. 1993 [70]).

There are several scale factors which modifies the intensities of the reflections. Lorentz factor - the relative speed of a reflection through the Ewald sphere, can be calculated from the diffraction geometry. Polarization - correction for polarization of the diffracted beam by the crystal itself, and for the polarization of the incident beam. In the case of the synchrotron radiation polarization of the beam needs to be measured. The beam intensity could change (mainly at synchrotron). Change of the illuminated volume during the crystal rotation if the illuminated volume is smaller than the crystal, absorption which is pronounced for the longer wavelength, radiation damage are the known scale factors.

In order for the scale factor to be determined, repeated measurements of the equivalent reflections need to be brought together: this include observations from different runs or different crystals. A scale factor $k_i(h)$ is required for each observation $I_i(h)$ (the ith measurement of reflection $\mathbf{h}$) which will put it on the same scale as all the others. This scale factor $k_i(h)$ will be a function of other parameters. The usual practice is to determine scales from the internal redundancy of the data. The scale factors are determined by minimizing the function

$$\Phi = \sum_{\mathbf{h}} \sum_{i} w_i(h) \left(\frac{I_i(h) - \langle I(h) \rangle}{k_i(h)^2} \right) \tag{69}$$

with respect to $k_i(h)$. In this equation $I_i(h)$ is the ith measurement of reflection h, $k_i(h)$ is the scale factor belonging to that observation, $w_i(h)$ is the weight for that observation, $w_i(h) = 1/\sigma^2_{I_i}(h)$, $\langle I(h) \rangle$ is the current best estimate of the intensity

$$\langle I(h) \rangle = \frac{\sum_i w_i(h) I_i(h)/k_i(h)}{\sum_i w_i(h)/k_i^2(h)} \tag{70}$$

Different forms of $k_i(h)$ can be introduced. The simplest case is when there is one scale factor per 'batch', where a batch is for instance one image or film in a 'rotation' dataset. Whatever the form of $k_i(h)$, it is important that the scale factors are well defined by the redundancy of the data. In the simple batch scaling case, this means that each batch should overlap with a reasonable number of other batches, and not just its neighbors. Suppose the scale factor $k_i(h)$ which puts the observation on a common scale is determined, then the mean intensity can be estimated

$$\langle I(h)\rangle = \frac{\sum_i u_i(h)k_i(h)I_i(h)}{\sum_i u_i(h)} \tag{71}$$

where $u_i(h)$ is a weight. It is usual to weight observations by their estimated variances i.e. $u_i(h) = 1/(k_i(h)\sigma_i(h))^2 = w_i(h)/k_i(h)^2$.

3.7 Crystallographic refinement

The refinement is a chemically constrained or restrained nonlinear optimization of a target function, which usually measures the agreement between observed diffraction data and data computed from an atomic model (Eyck et al. 2001 [71]). The goal of the refinement is to optimize simultaneously the agreement of the atomic model with observed diffraction data and with a priori chemical information. It is a search for a global minimum of the target

$$E = E_{chem} + w_{X-ray}E_{X-ray} \tag{72}$$

as a function of the parameters of atomic model, in particular atomic coordinates. E_{chem} comprises empirical information about chemical interactions; it is a function of all atomic positions, describing covalent (bond length, bond angles, torsion angles, chiral centers and planarity of aromatic rings) and non-bonded (intramolecular as well as intermolecular) interactions. E_{X-ray} is related to the difference between observed and calculated data, and w_{X-ray} is a weight appropriately chosen to balance the gradients (with respect to atomic parameters) arising from the two terms.

The traditional form of E_{X-ray} consists of the crystallographic residual, E^{LSQ}, defined as the sum over the squared differences between the observed $|\mathbf{F}_o|$ and calculated $|\mathbf{F}_c|$ structure factor amplitudes for a particular atomic model:

$$E_{X-ray} = E^{LSQ} = \sum_{hkl}(|\mathbf{F}_o| - k|\mathbf{F}_c|)^2 \tag{73}$$

where k is a relative scale factor.

3.7.1 Least squares method

Least squares is the simplest statistical method used in macromolecular refinement. The least-squares residual function is

$$f(\mathbf{p}) = \sum_{i}^{alldata} \left(Q_o(i) - Q_c(i,\mathbf{p})\right)^2 / \sigma_o(i)^2 \tag{74}$$

where $Q_o(i)$ and $\sigma_o(i)$ are the value and the standard deviation for the observation number i. $Q_c(i,\mathbf{p})$ is the model's prediction for the observation number i using the set of model parameters $\mathbf{p}$. One varies the parameters of the model to find a set that gives the lowest sum of deviants. The values of the parameters found by minimizing this function are those that have the smallest

individual standard deviation or the smallest probable error. The assumptions of least squares are that the errors in the observations obey a normal distribution with completely known ('observed') variances and that, given perfect observations and the best parameters, the model would predict the observations perfectly.

3.7.2 Maximum likelihood method

Minimization of E^{LSQ} can produce improvement in the atomic model, but it can also accumulate systematic errors in the model by fitting noise of diffraction data. The least squares residual is a limiting case of the more general maximum-likelihood theory. The goal of the maximum-likelihood method is to determine the likelihood of the model, given estimates of the model's errors and those of the measured intensities ([72]).

The starting point for the maximum-likelihood formulation of crystallographic refinement is the Sim distribution ([66]), i.e. the gaussian conditional probability distribution of the 'true' structure factors, $\mathbf{F}$ given a partial model with structure factors $\mathbf{F}_c$ and the model's error:

$$P_a(\mathbf{F};\mathbf{F}_c) = (1/\pi\epsilon\sigma_\Delta^2)exp[-(\mathbf{F}-D\mathbf{F}_c)^2/\epsilon\sigma_\Delta^2] \tag{75}$$

where σ_Δ is a parameter that incorporates the effect of the fraction of the asymmetric unit that is missing from the model and errors in the partial structure. Assuming the Wilson distribution of intensities

$$\sigma_{Delta}^2 = \langle|\mathbf{F}_o|^2\rangle - D^2\langle|\mathbf{F}_c|^2\rangle \tag{76}$$

where D is a factor that takes into account model error. Taking measurement errors into account requires multiplication of the equation with an appropriate probability distribution of the observed structure factor amplitudes $|\mathbf{F}_o|$ around the 'true' structure factor amplitudes $|\mathbf{F}|$,

$$P_{meas}(|\mathbf{F}_o|;|\mathbf{F}|) \tag{77}$$

Prior knowledge of the phases of the structure factors can be incorporated by multiplying equation with a phase probability distribution

$$P_{phase}(\phi) \tag{78}$$

The unknown variables $|F|$ and ϕ in equations have to be eliminated by integration in order to obtain the conditional probability distribution of the observed structure factor amplitudes, given a partial model with errors, the amplitude measurements errors and prior phase information:

$$P_a(|\mathbf{F}_o|;\mathbf{F}_c) = (1/\pi\epsilon\sigma_\Delta^2) \times$$
$$\int d\phi d|\mathbf{F}||\mathbf{F}|P_{meas(|\mathbf{F}_o|;|\mathbf{F}|)}P_{phase}(\phi)exp\{-\langle|\mathbf{F}|exp(i\phi) - D\mathbf{F}_o\rangle^2/\epsilon\sigma_\Delta^2\}. \tag{79}$$

The likelihood L of the model is defined as a joint probability distribution of the structure factors of all reflections in the working set. Assuming independent and uncorrelated structure factors, L is simply the product of the distributions in equation (79) for all reflections. Instead of maximizing the likelihood, it is more common to minimize the negative logarithm of the likelihood,

$$E_{X-ray} = L = -\sum_{hkl} log\langle P_a(|\mathbf{F}_o|;\mathbf{F}_c)\rangle \tag{80}$$

Of the many methods of minimizing functions, the simplest methods to describe are the search methods. The common motif of search methods is that they each have some means of selecting which combination of parameters to test and simply keep track of the best one found so far. One can systematically sample all combinations or randomly pick values for the parameters. Among the methods of minimizing function gradient-descent method and simulating annealing are the most widely used in macromolecular crystallography.

3.7.3 Simulated annealing

Annealing denotes the physical process wherein a solid is heated until all particles randomly arrange themselves in a liquid phase and is then cooled slowly so that all particles arrange themselves in a lowest energy state. By formally defining the target E to be the equivalent of the potential energy of the system, one can simulate such an annealing process (Brunger et al. 2001 [73]). There is no guarantee that simulated annealing will find the global minimum. However, compared to conjugate-gradient minimization, where search directions must follow the gradient, simulated annealing achieves better solutions by allowing motions against the gradient. The likelihood of uphill motion against the gradient is determined by a control parameter referred to as temperature. It should be noted that simulated annealing temperature normally has no physical meaning and merely determines the likelihood of overcoming barriers of the target function E.

The simulated-annealing algorithm requires a mechanism to create a Boltzmann distribution at a given temperature T, and an annealing schedule, that is a sequence of temperatures $T_1 \geq T_2 \geq ...$ at which the Boltzmann distribution is computed. Implementations differ in the way they generate a transition, or move from one set of parameters to another that is consistent with a Boltzmann distribution at a given temperature. The two most widely used methods are Metropolis Monte Carlo and molecular dynamics simulations. For X-ray crystallographic refinement, molecular dynamics has proven extremely successful because it limits the search to physically reasonable 'moves'.

A suitably chosen set of atomic parameters can be viewed as generalized coordinates that are propagated in time by the classical equations of motion. If generalized coordinates represent x,y,z positions of the atoms of a molecule, the classical equations of motion reduce to the Newton second law:

$$m_i \frac{\delta^2 \mathbf{r}_i}{\delta t^2} = - \bigtriangledown_i E \tag{81}$$

The solution of the partial differential equations can be achieved numerically using finite-difference methods. Initial velocities for the integration of the equation 81 are usually assigned randomly from a Maxwell distribution at the appropriate temperature.

Another parametrization can be chosen. If the bond length and bond angles are defined correctly then they could be fixed and only torsion angles will have the degree of freedom. In torsion-angle space molecular dynamics can be performed.

Simulated annealing requires the control of the temperature during molecular dynamics. The current temperature of the simulation T_{curr} is computed from the kinetic energy

$$E_{kin} = \sum_{i}^{n} 1/2 m_i \left(\frac{\delta r_i}{\delta t} \right)^2 \tag{82}$$

of the molecular-dynamics simulation,

$$T_{curr} = 2E_{kin}/3nk_B \tag{83}$$

Here n is number of atoms, m_i is the mass of the atom, k_B is Boltzmann constant. One commonly used approach to control the temperature of the simulation consists of coupling the equations of motion to a heat bath through a 'friction' term. Another approach is to rescale periodically the velocities in order to match T_{curr} with the target temperature.

The simulated-annealing temperature needs to be high enough to allow conformational transitions, but not so high that the model moves too far away from the correct structure. Higher

temperatures are attainable using torsion-angle molecular dynamics. The duration of the annealing schedule is another parameter. Too short a protocol does not allow sufficient sampling of conformational space.

Simulated annealing has a much larger radius of convergence than conjugate gradient minimization. It must be able to find a lower minimum of the target E than the local minimum found by simply moving along the negative gradient of E. For simulated annealing initial temperature must be large enough to overcome smaller barriers, but low enough to ensure that the system will not escape the global minimum if it manages to arrive there

3.7.4 Automated refinement protocol

In number of cases when the data obtained experimentally has a good quality significant progress in the structure determination could be done in the automated way. The method was introduced which is different from the conventional crystallographic refinement. Conventional crystallographic refinement optimizes the parameters of the model with the fixed atoms supplied by the user to fit both the experimental data and set of a priori stereochemical observations. The automated refinement procedure (ARP) can remove the atoms and build the new atoms in the electron density maps. ARP is a combination of least-square refinement with automatic updating of the model on the basis of the calculated Difference Fourier syntheses (Lamzin et al 2001 [74]).

Atom rejection in ARP is primarily based on the interpolated $2mF_o - \delta F_c$ or $3F_o - 2F_c$ electron density at its atomic centre and the agreement of the atomic density distribution with a target shape. Applied together, these criteria offer powerful means of identifying incorrectly placed atoms. Atoms addition uses the difference $mF_o - \delta F_c$ or $F_o - F_c$ Fourier synthesis. The map grid point with the highest electron density satisfying the defined distance constraints is selected as a new atom, grid points within a defined radius around this atom are rejected and the next highest grid point is selected. This is iterated until the desired number of new atoms is found and reciprocal space minimization is used to optimize the new atomic parameters. Geometrical constraints based on a priory chemical knowledge of the distances between covalently linked carbon, nitrogen and oxygen atoms and hydrogen-bonded atoms are applied in rejection and addition of atoms.

Real space refinement based on density shape analysis around an atom can be used for the definition of the optimum atomic position. The procedure of real-space refinement is coupled to least-squares or maximum-likelihood optimization of the model's parameters against the X-ray data. There are two different modes of ARP. In the restrained mode, all atoms in reciprocal-space refinement are treated as free atoms with unknown connectivity and are refined against the experimental data alone. This mode has a higher radius of convergence but needs high-resolution diffraction data to perform effectively. In the restrained mode, a model or a hybrid model is required, i.e. the atoms must belong to groups of known stereochemistry. This stereochemical information, in the form of restraints, can be utilized during the reciprocal-space minimization, allowing it to proceed with less data, presuming that the connectivity of the input atoms is basically correct.

The main problem in automatically reconstructing a protein model from electron density maps is in achieving an initial tracing of the polypeptide chain. The method utilizes the fact that all residues that comprise a protein have chemically identical main-chain fragments which are close to planar and that the protein is composed of linear non-branching polypeptide chain. Usual result of the tracing is a set of the several main chain fragments. Residues are differentiated only as glycine, alanine, serine and valine, and complete side chains are not build at the refinement stage. For every polypeptide fragment, a side-chain type can be assigned with a

defined probability, using connectivity criteria from the free atom model. After all the refinement is done the attempt to build the complete side chains using the protein sequence is undertaken. For the success of the ARP model building and refinement X-ray data should be complete as possible especially in the low resolution range (5 Å and lower). If the low resolution strong data are systematically incomplete, the density map even in the case of a good model may be discontinuous. This discontinuity could lead to a slow convergence and even non-interpretable maps. The highest resolution should be at least 2.6 Å to be able to utilize the atomicity during the protein model update. The number of X-ray reflections should be at least six to eight times higher than the number of atoms in the model.

3.8 Protein structure validation

As deriving the atomic models from experimental data involves sophisticated optimization (refinement) procedures, the resulting models are prone to errors which fall into two broad categories: systematic errors caused by biases during the structure determination and refinement procedures, and random errors which affect the precision of the models. With the rapid growth in the number of structures of macromolecules the availability of objective criteria and methods for evaluating the quality of these structures has become a very important requirement. A variety of validation procedures have been proposed. The procedures involve two main approaches. One approach comprises procedures that validate the geometric and conformational parameters of the final model. This is done by measuring the extent to which the parameters deviate from standard values, derived from crystals of small molecules or from a set of high quality structures of other macromolecules (Engh et al. 2001 [75]). The second and most important approach comprises procedures that taka into account experimental data and evaluate the agreement of the atomic model with this data. The most commonly cited measures of agreement between the model as a whole and the data are the R factor and the 'free R factor'.

Validation of stereochemical parameters. Validation of the covalent geometry of the atomic model involves comparing the bond distances and angles of the macromolecule against standard values. The standard values derived from CSD (Cambridge Structural Database) and used as restraints in crystallographic refinement programs. As a result the bond distances and angles of the final model usually agree well with their standard values. For proteins, the most commonly used standard values are those compiled by Engh & Hubber [75].

Protein structure validation package such as PROCHECK ([76]) flag all bond distances and angles that deviate significantly from the database-derived reference values. This includes analysis of the deviations from planarity in aromatic rings and planar side-chain groups. The output of PROCHECK comprises a number of plots, together with detailed residue-by-residue listings of secondary-structure assignment, non-bonded interactions between different pairs of residues, main chain bond length and bond angles, and peptide bond planarity. The program also displays main-chain dihedral angles (ϕ and ψ) as a two-dimensional Ramachandran plot. The Ramachandran plot classifies each residue in one of three categories: 'allowed' conformations, 'partially allowed' conformations, which give rise to modestly unfavorable repulsion between non-bonded atoms, and which can be overcome by attractive effects such as hydrogen bonds; and 'disallowed' interactions which gives highly unfavorable non-bonded interatomic distances. The Ramachandran plot can identify unacceptable clusters of $\phi - \psi$ angles, revealing possible errors made during model building and refinement. As opposed to covalent bond angles and bond lengths, the main-chain dihedral angles are not usually restrained during X-ray refinement and therefore can be used to validate the structural model independently. In practice, the Ramachandran plot is one of the simplest, most sensitive tools for assessing the quality of a protein model. Other stereochemical parameters, such as side-chain torsion angles (χ_1, χ_2, χ_3, etc), the peptide bond torsion (ω), the C^α tetrahedral distortion, disulfide bond geometry and stereochemistry are evaluated. An evaluation of the backbone hydrogen-bonding energy is also performed, by comparison with distributions computes from high-resolution protein structures.

Validation of a model versus experimental data. The most important measure of the quality of a given atomic model is its agreement with the experimental data. This type of validation is geared towards detecting systematic errors, which determine the overall accuracy of the model, and random errors, which affect the precision of the model. The most commonly used measures of the agreement between the atomic coordinates and the X-ray data are the classical R factor and the 'free R factor' (R_{free}). The latter is based on standard statistical cross-validation techniques and is therefore less amenable to manipulation, such as leaving out the weak data or

overfitting the data with too many parameters.

$$R = \sum w||F_o| - k|F_c||/\sum |F_o| \tag{84}$$

Conventional R value is meaningful if the number of experimental observations and restraints greatly exceeds the number of model parameters.

The definition of R_{free} is identical to that of the conventional R factor, except that the R_{free} is calculated for a small subset of reflections that are not used in the refinement of the model. Free R factor, therefore, measures how well the model predicts experimental observations that are not used to fit a model (cross-validation).

The fit of the model to the data can also be assessed in the real space, which has the advantage that it can be performed for arbitrary sets of atoms. The real space R value which measures the similarity of the map calculated directly from the model ρ_c and one which incorporates experimental data ρ_o is

$$R = \sum |\rho_o - \rho_c|/\sum |\rho_o + \rho_c| \tag{85}$$

where the sum extend over all grid points in the map that surround the selected set of atoms. The real space fit can also be expressed as a correlation coefficient, which has the advantage that no scaling of two densities is necessary.

Coordinate error estimate In principle, upon convergence of a least-square refinement, the variances and covariences of the model parameters may be obtained through inversion of the least-squares full matrix (Cruickshank et al. 2001 [77]). In practice, however, this is seldom performed as the matrix inversion requires enormous computational resources.

Instead, the elegant method of Luzatti has been used to estimate average coordinate errors of the macromolecular models (Luzatti [78]). A Luzatti plot is a plot of R factor versus $2\sin\theta/\lambda$ and a comparison with a theoretical curves is used to estimate the average positional error. Read (refer) estimated coordinate error from σ_A plots.

4 Results

4.1 Overexpression, extraction and purification of pSRII

For pSRII-HtrII complex as well as for BR crystallization one need to have an amount of the protein which is measured in 100-th of mg. In the case of Bacteriorhodopsin protein is produced in Halobacteria, the yield of protein is quite high because of original high level of BR production in Halobacteria. To get pSRII (or transducer) from *N.Pharaonis* demands a lot of time and efforts because of the low original yield of the protein, complexity of the extraction and purification protocol and low growth speed of the bacteria. The protein therefore was overexpressed in *E.Coli*. Overexpression of pSRII in *E.Coli* strain $BL21-CodonPlus(DE3)$ results in substantially high yield of the protein. In comparison to BR which in *E.Coli* forms inclusion bodies pSRII is overexpressed in a functional form, including bound all-trans retinal and is incorporated into the cytoplasmic membrane of *E.Coli*. Photocycles of the protein overexpressed in *E.Coli* and protein grown up in wild strain *N.Pharaonis* are identical.

Expression of the gene, located at pET27bmod plasmid in the absence of induction is controlled by lac-repressor and gene is not toxic for *E.Coli*. After 4 hours of growth in incubator starting from the moment of the expression induction with IPTG and addition of the retinal to the media (at O.D. = 16), cells with the concentration reaching 15 g/l are isolated and frozen. This cells have bright orange color indicating the high content of the pSRII.

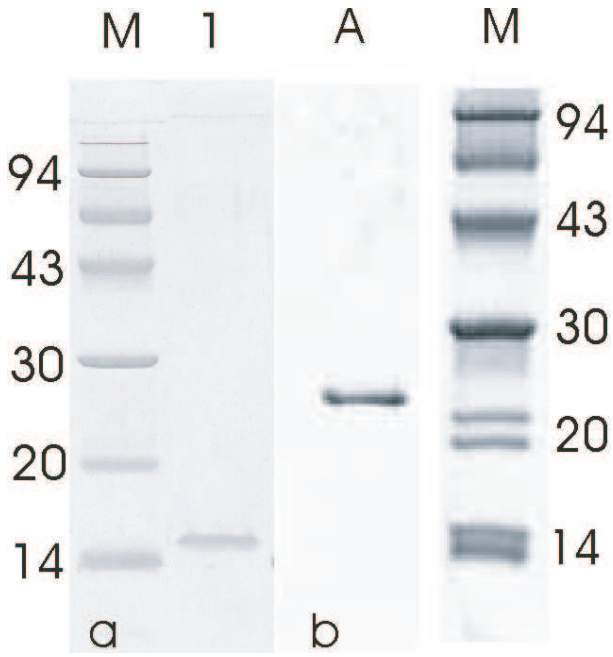

Figure 28: Gel-electrophoresis of purified proteins. Line M - protein marker. (A) Line 1 - HtrII after purification. (B) Line A - pSRII after purification

Extraction and purification of pSRII was performed in trials dealing with low amount of the cells at a time: 60-80 g. Attempts to increase the amount of the cells used for one trial was not successful. The reasons for that is not yet clear. Final outcome of the pSRII is 0.4-0.5 mg from 1 g of the cells. Results of the SDS polyacrylamide gel electrophoresis are shown in Figure 28. The quality of the protein can be further confirmed by the absorption spectra in visual range by comparing the absorption at 280 nm and 500 nm. As a rule the ratio of two absorptions

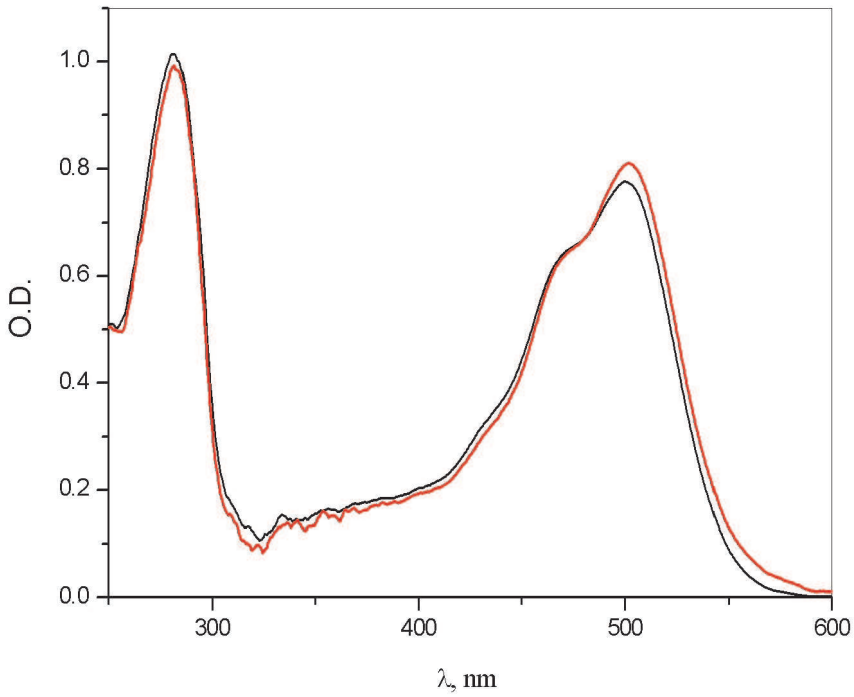

Figure 29: Absorption spectra of pSRII and pSRII-HtrII complex. Black - absorption spectra of pSRII alone, red - absorption spectra of the reconstituted into polar lipids liposomes pSRII-HtrII complex

is in the range 1.2-1.4 (protein with the ratio more than 1.5 was not used for crystallization). Characteristic absorption spectra of the purified protein is shown in Figure 29.

4.2 Overexpression, extraction and purification of the transducer

The work on yielding the pure transducer was performed at Max Plank Institute Dortmund. The same system as for pSRII was used for overexpression of the transducer. The final outcome of the purified transducer was 0.3 mg from 1 liter of the growth media. Results of the gel-electrophoresis are shown in Figure 28. (line 1).

4.3 Reconstitution of pSRII and HtrII

To obtain the complex of the proteins they was reconstituted into the bilayer made of polar lipids of the purple membranes. During the reconstitution of the proteins into the bilayer they form the complexes dimer (stoichiometry 2:2). The complexes (stoichiometry 1:1) are stable after resolubilization with OG. Lipidic bilayer works as a filter for the protein as only functional protein is reconstituted into the bilayer. Further extraction of the membranes with reconstituted protein leads to a better quality of the protein. This can be proved from the absorption spectra. The ratio of the absorption peak maxima is correspondingly reduced after reconstitution procedure. The outcome of the complex is 50-60% of the initial amount of the protein.

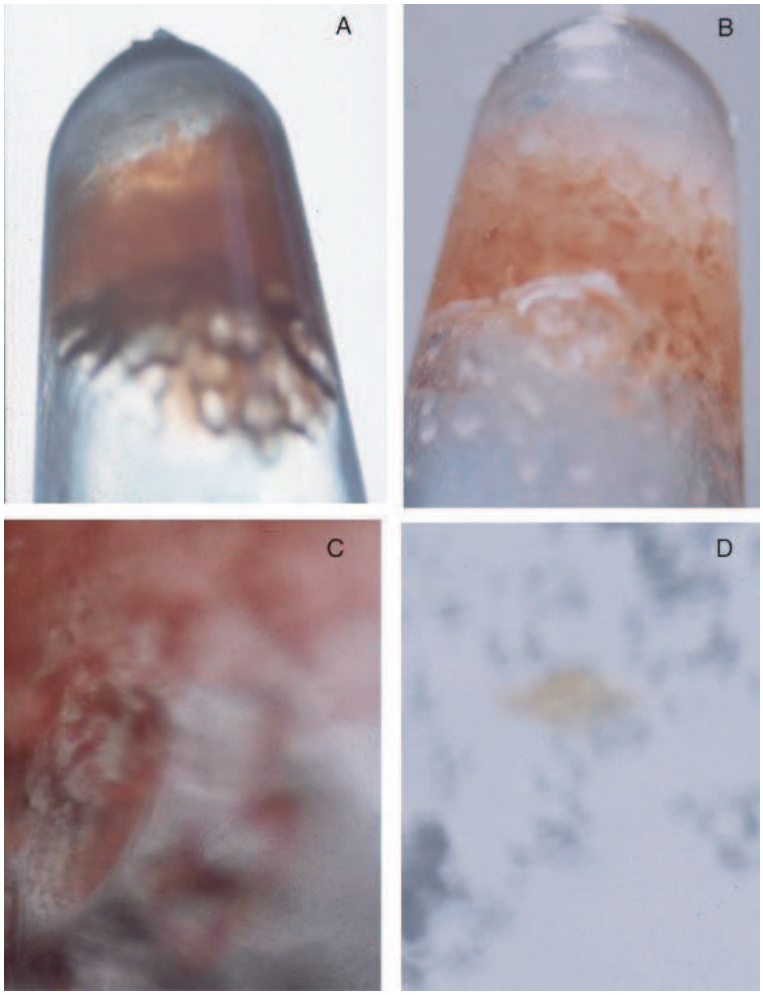

Figure 30: Photos of the crystallization probe for pSRII-HtrII complex

4.4 Crystals of the pSRII-HtrII complex. Crystallization conditions screening

For both truncated transducer crystallization of the complex was successful. Initially crystals of the complex were obtained in the cubic phase formed by monooleyn. Initial conditions for the crystal growth are 1:1 (w/w) MO and crystallization buffer - 150 mM NaCl, 25 mM Na/Na Pi, pH=8.0. Growth of the crystals has happened only in probes with high concentration of the protein in the buffer 25-30 mg/ml (Figure 30). Attempts to increase the volume of the buffer with a low concentration of the protein in crystallization probe was not successful.

First crystals are formed in 3-5 days. This crystals are orthorhombic, 3-4 weeks after crystallization initiation these crystals reach the size of 100-130 μm with the thickness of 5-10 μm. Crystals can grow at various precipitant concentrations between 0.5 and 2.0 M Na/K Pi, pH=5.6. From the amount of the precipitant the time of the crystal growth depends. At low concentration of the precipitant salt crystals were growing slowly but the number of the crystals were bigger than in the case of high precipitant concentration.

Among the variations of the parameters of crystallization particularly fruitful was increase of the crystallization buffer content in the crystallization probe. At the optimized proportion 2:1 of

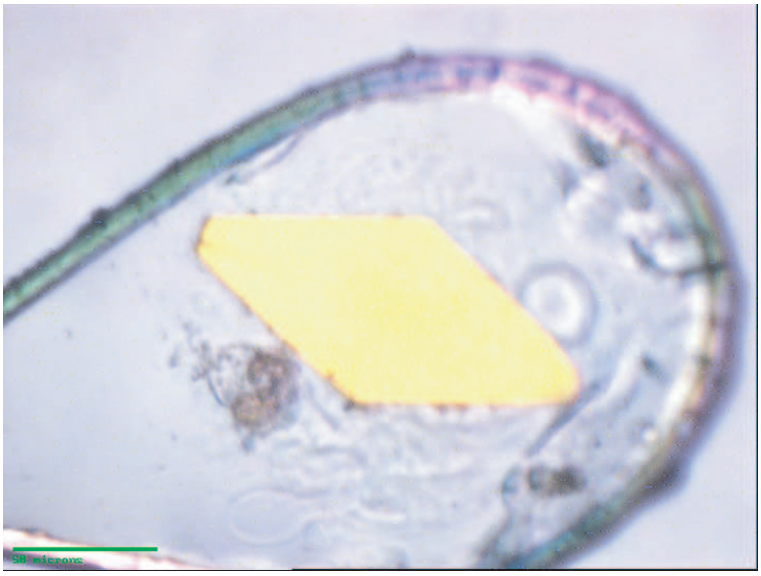

Figure 31: pSRII-HtrII complex crystal frozen in the liquid nitrogen in the cryoloop before the X-ray data collection experiment

the aqueous to lipidic parts in the crystallization probe were obtained more and better crystals (concentration should be high 25-30 mg/ml). Moderate variation of the buffer content around the above conditions does not seem to be crucial for the crystal growth (quantity of NaCl, Na/Na Pi and/or Na/K Pi) also the difference in the amount of the detergent in the crystallization buffer (0.8-1.2 w/v %).

Crystals were not always growing at the same experimental conditions, e.g. the reproducibility of the crystal growth was moderate. The reasons for that are not absolutely clear.

Later also two other monoglycerids were used for crystallization. For these trials the proportion between the aquous and lipidic parts was fixed to 2:1. In both monoglycerids the crystals grew. But in monopalmitoyl crystals were small in number and size, reproducibility of the result was low. Crystallization in monovaccenin has brought the improvements of the crystals. Particularly the size of the crystal has increased up to 500-600 μm.

4.5 Cubic lipidic phase dissolution

Obtained crystals were successfully detached from the cubic lipidic phase after being subjected to the buffer 3M Na/Na Pi, pH=5.1, 0.1% OG (1.2 ml of the buffer to one crystallization probe). Dissolution of the cubic lipidic phase took 2-3 days. Crystal were stable in this buffer during 2-3 weeks without decrease of their size and also decrease of the diffraction quality.

4.6 Diffraction of the crystals

Quality of the crystals were assessed by the laboratory X-ray diffractometer. After 1-1.5 hours of exposition to X-ray separate diffraction peaks were observed some of them was reaching to 8-9 Å resolution. During the crystal growth optimization both resolution and quality of the reflections were improved. The resolution achieved 1.8-2.0 Å.

Preliminary selected crystals (at laboratory X-ray diffractometer) was finally studied at synchrotron beamlines in ESRF. Example of the diffraction picture at synchrotron is in Figure

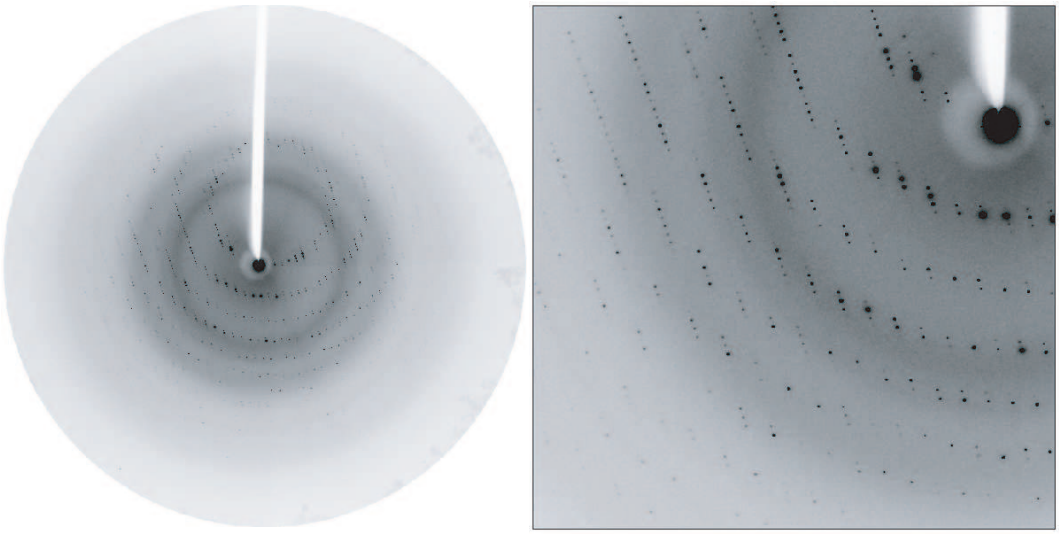

Figure 32: Diffraction pattern to 1.8 Å resolution recorded from pSRII-HtrII complex crystals.

32.

4.7 Analysis of the diffraction of the crystals of pSRII-HtrII complex

An example of a diffraction image obtained with the rotation method is presented in Figure 32. This images were recorded from the crystals of pSRII-HtrII(1-114) complex. Autoindexing procedure (Rossmann et al. 2001 [79]) showed that crystals have a space group $P2_12_12$ and the cell dimensions $a = 124.59$ Å, $b = 46.96$ Å, $c = 53.84$ Å. Crystal-to-detector distance was 115 mm, wavelength 0.931 Å. Resolution defined by the utmost weak reflections is 1.9 Å. Oscillation range is 1^0. Crystal mosaicity is 0.6^0. Several lunes could be seen in the figure. They are well separated on the first image. Within the lunes diffraction spots are arranged in lines with spacings which corresponds to a dimension 124.6 Å and b dimension 47.0 Å. At high resolution there are deviations from linear arrangement what can be demonstrated with the curvature of Ewald sphere which is preserved in projection of the reflections to the detector plane. This image is collected with the c-axis 53.8 Å being parallel to the incident X-ray beam. The longest axis of the crystal is oriented along the rotation axis. The diffraction patterns for the orientations separated by 90^0 do not look very different. Only the spacing in one direction is different. The borders of the lunes are not sharp. Reflections positioned on the border are fading out being only partial reflections. As this fading out zone is a substantial part of the lune the mosaicity has a significant value. In this case it is more than a half of the rotation range. Data collection over a rotation range of 90^0 with the starting rotation angle corresponding to c-axis being parallel to the incident X-ray direction allows to collect almost complete dataset of intensities with a minimal blind region.

In the case of data collection for the crystal of complex to obtain the maximum resolution which could for some crystals achieve a value of 1.7 Å, one should increase the exposure time up to 12 seconds per image to get sufficient contrast of the highest resolution reflections to the background. Under this data collection conditions overloading of the strong reflections in the

low resolution zone is unavoidable. One should decrease the time acquisition up to the factor of 10 to obtain the intensities of those reflections.

4.8 Example of the data integration for pSRII-HtrII complex crystals with MOSFLM

MOSFLM [69] starts with several user supplied parameters as a crystal-to-detector distance and positions of the incident beam on the detector, wavelength and the type of the detector. Several diffraction images types are supported. In an automated fashion reflections are selected on the diffraction image. According to the position of the reflection peak on the detector plane and the intensity of the reflection, they are passed to the indexing procedure. The reflection pattern could be analyzed by the means of several algorithms which as an output present the possible space groups, crystal cell dimensions and a calculated penalty value. The penalty value is defined according to the correspondence of the positions of the observed and predicted diffraction pattern for each suggested space group. The autoindexing algorithm is described in [79]. For the precise determination of all the cell parameters diffraction spots from images with a big separation in rotation angle should be considered. As a next step the space group is selected from the list of the space groups and the cell parameters are refined with the chosen symmetry imposed. After the refinement corrected values describing the geometry, which were supplied in the beginning, are updated. The orientation of the crystal is presented as the orientation matrix. An estimate of the mosaic spread could be obtained. If necessary this estimate of the mosaicity can be corrected manually. With the given value of the mosaic spread the cell parameters could be further refined in the Post-refinement procedure. Post-refinement uses the distribution of the intensity of partially recorded reflections over the images on which the partial is recorded (the previous limit of two images no longer applies) to refine cell parameters, orientation and mosaic spread. It has the distinct advantage that the derived cell parameters are entirely independent of all detector parameters (crystal to detector distance and detector orientation) and distortions which, if inaccurate, can lead to significant errors in the cell parameters derived from autoindexing.

4.9 Attempts to elucidate the structure of the pSRII-HtrII (1-159) complex

4.9.1 Two crystal types

Obtained ground state structure of the pSRII complex with the transducer of the length 114 amino acids has left an important region of the transducer out of the investigation. This region, known as a linker domain seems to be important for the interaction of the receptor with its transducer.

Attempts were undertaken to investigate the structure of the complex of receptor molecule with longer transducer including the linker domain. The length were suggested on the basis of the observations on the proton translocation in the presence of the transducer. Those observations shown that the first 157 residues of the N-terminal part of the transducer are enough to sustain the responses to the photon absorption by the receptor as in the case of the native system with the complete transducer (personal communications, J. Klare).

Additional crystallization trials were made which resulted in the appearance of the crystals of the extended complex. This crystals were like the first reported crystals of the shorter transducer complex in orthorhombic space group with their size approaching 200-250 μm. This size of the crystals was sufficient for X-ray diffraction experiments.

The first diffraction data collection from these crystals has shown that the cell parameters of the crystals indexed in space group $P2_12_12$ were surprisingly the same as those of the shorter transducer. This was unexpected since 45 new residues were introduced into the structure. The cell parameters of the short transducer complex were $a = 124.30$ Å, $b = 46.96$ Å, $c = 53.84$ Å against the cell parameters $a = 124.53$ Å, $b = 46.80$ Å, $c = 53.76$ Å of the long transducer complex.

Also crystals of another type were found. These crystals looks similar to the first type with the rhombic shape and are of the same size. But the indexing in MOSFLM has shown that one of the cell dimensions which corresponds to the membrane layers thickness was double that of the first type crystals and the space group was like before orthorhombic. The size of this dimension was still precisely double size of the first type crystals.

Several datasets were collected and examined for each kind of the crystals and in all cases there was no significant difference in the cell dimensions.

The data from the crystals were integrated and scaled as it was described in Chapter 3. The mosaicity values for this crystals appeared to be in the range $0.5 - 0.7^0$. On the basis of the R_{merge} factors, completeness and ratio of mean intensity to its sigma final resolution was established. For several datasets this final resolution reached 1.7-1.9 Å, though in average it was in the range of 1.9-2.2 Å. The information about the integrated and scaled datasets is given in Table 2.

Table 2.

Dataset name	Resolution	R_{merge} (last shell)	I/σ (last shell)	Completeness	Number of reflections	B factor
SHL3	2.2	0.333	2.2	99.4	18985	23.3
SHL29	1.9	0.339	2.2	99.8	25605	19.4
SHL8	2.1	0.296	2.6	98.4	18617	22.7
SHL19	2.3	0.396	1.9	98.1	14287	23.2
SHL6	1.92	0.382	1.6	93.8	23819	22.61
SHL5	1.7	0.465	1.6	90.2	32085	21.7
SHL13	2.0	0.235	2.7	99.0	19759	22.1
SHL16	2.15	0.436	1.7	99.7	17735	21.7
SHL12	2.2	0.380	2.0	99.4	16492	23.4

last resolution shell is of 0.05 Å width

4.9.2 Initially phased electron density maps

To obtain the initial phases the method of molecular replacement was used. In this case a very good initial model for the molecular replacement is available. This is the model of the shorter transducer complex (PDB entry 1H2S). To avoid the bias model was reduced to include only the receptor and only the main chain atoms. All the side chains were reset to Ala or Gly. The retinal was also not included into the starting model.

In the next round the position of this starting model was searched by molecular replacement (MOLREP, Vagin et al. 1997 [80]). Simple rotation search was made with resolution limits that included all the obtained data to orient the molecule of the reduced receptor in the unit cell of $P2_12_12$ symmetry. The unit cell symmetry implies one independent molecule per unit cell.

The rotation function showed an appreciable contrast of the first significant solutions over the background tale. The proper rotation can be unambiguously selected, and the molecule can so be oriented in the unit cell prior to the translation search. The translation search is performed by simple scanning of the crystal cell asymmetric unit and again a clear unique solution arises. This solution is characterized with a R-factor = 0.477 and a quite high correlation coefficient CC = 0.527.

The positioned model allows to calculate the first set of phases. The first electron density map is obtained by the combination of the observed structure factor amplitudes with molecular replacement phases.

The inspection of electron density map shows if the quality of amplitudes and phases is good enough to show the electron density of the side chains. The side chains of the receptor could be easily guessed from the electron density map. During the first observation of the electron density maps almost all the side chains of the receptor can be assigned. Moreover for any side chain the correct rotamer can be chosen and the torsion angles can be further adjusted. After the first round of model building all the side chains of the receptor can be introduced into the model except some flexible side chains which are located on the surface of the receptor and also 15 residues of the C-terminus of the receptor can not be defined from the so obtained molecular replacement electron density maps.

The building procedure in this case can be passed to the Automatic refinement protocol (ARP) form ARP/wARP program suite [74]. All the characteristics of the data are satisfactory or even very good therefore a successful automated building by ARP could be expected. Parameters of the data are defined in CCP4 graphical interface (CCP4, 1994 [81]) these include the information about the crystal, resolution range of the data, resolution range used for the scaling of experimental and calculated data.

A bulk solvent correction model was invoked. This model is used in all procedures, including all the possible refinement procedures and map calculations. It defines the region of the unit cell occupied by the solvent. This region is defined as complimentary to the region occupied with the model. In most cases for successful bulk solvent correction it is sufficient to define the protein region as the region occupied by the current model. One can define a mask - the surface which will set explicitly the structured region. The mask is determined by default as a surface at a fixed distance from a model atoms. Within the so defined region the density of the solvent is put to be at constant level. Fourier transformation of this density defines the set of structure factors F_{bulk} which represent the scattering from the solvent. In the next step this structure factors F_{bulk} are scaled to the observed amplitudes F_{obs} and calculated from model structure factors F_{calc}. During this scaling not only the overall scale is defined but also B-factor for solvent.

For updating the atoms two thresholds are defined with which the electron density maps are scanned. By default these levels are 3.2σ for building the new atoms in density and 0.9σ for removal of the atoms from weak density.

The set of amplitudes for cross-validation including 5 per cent of the total structure factor amplitudes set is defined therefore the R_{free} can be monitored during the refinement. For refinement 10 cycles of REFMAC maximum likelihood based refinement ([72]) is performed between each cycle of the autobuilding. Altogether 100-150 cycles are performed.

The behavior of the R-factor and R_{free} is shown in Figure 33. It drops down significantly from the value of 0.45 to value of 0.20 before the first cycle of autobuilding is performed for first 20 cycles of the automated atom update joined with the refinement. During the rest cycles R-factor drops not significantly but it shows still a gradual improvement. Cycles of autobuilding due to the restrictions introduced by the geometry constraints always increase the R-factor by a step which decreases upon the convergence of the refinement. As a result the ARP/wARP built the completed receptor model with partially defined transducer and a significant number

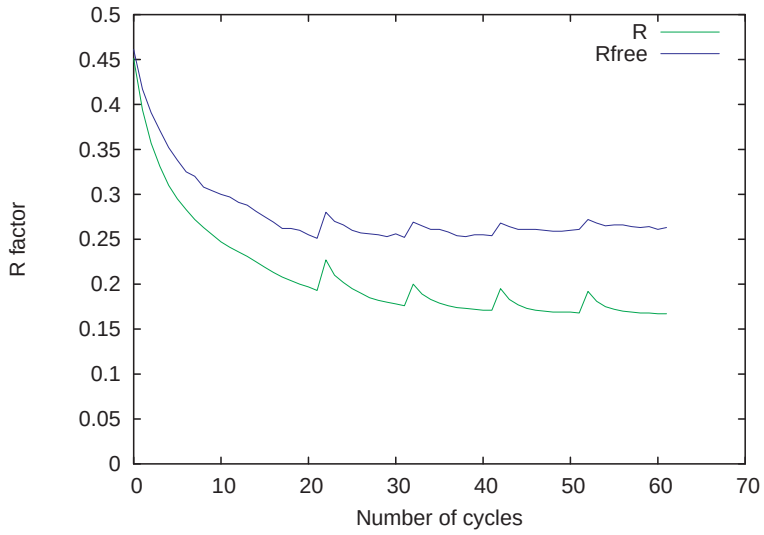

Figure 33: Refinement R factors monitored during the refinement run following ARP protocol. Green - conventional R factor, blue - free R factor

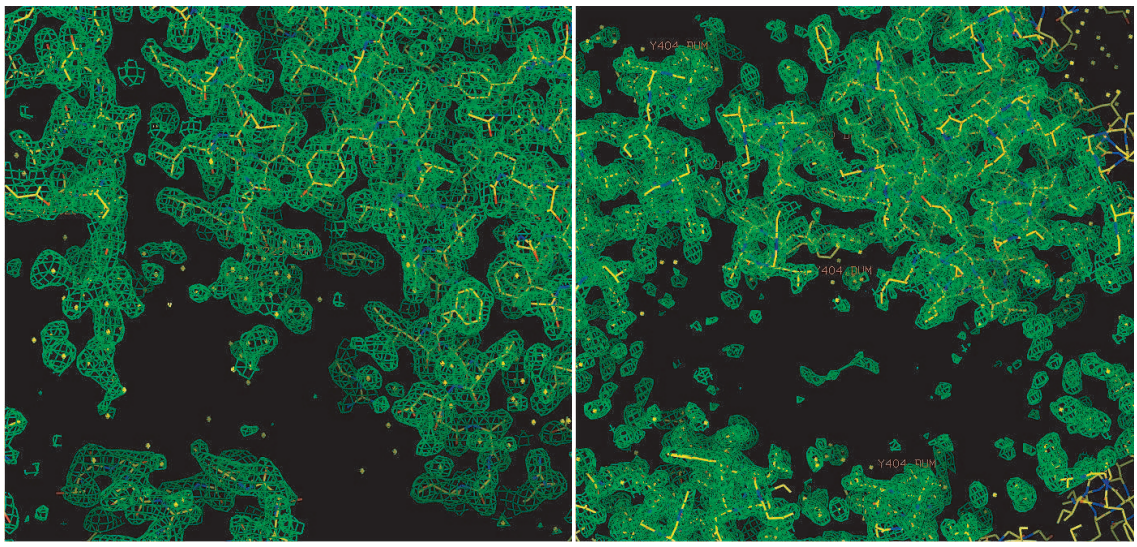

Figure 34: View of the ARP defined model with $2F_o - F_c$ electron density maps countoured at 1σ obtained after ARP procedure.

of dummy atoms (Figure 34). The final R-factor is 0.17 and R_{free} is 0.25.

Based on the updated model a new electron density map is calculated. This map has a much better appearance, especially in the receptor region, but also in the transducer region there are certain improvements (Figure 34). The main chain of the residues 27 - 76 of the transducer is traced by ARP. The side chains are only build partially with Ser, Val and Ala residues substituted instead of those in the protein sequence. The ends of helices TM1: 23 - 26, and TM2: 77 - 82 (of the known part of the transducer) are not modelled, though the densities are defined for this part and the pattern of the densities is helical. Unfortunately the quality of these densities is not good enough to evaluate the positions of atoms stereochemically bound in automatical way.

Instead, a number of dummy atoms covering these densities are assigned to density.

In the next step the residues of the receptor are examined one by one and where it is necessary the corrections are made. This procedure is done in O (Kleywegt et al. 2001 [82]). This is the visualization framework for macromolecular crystallography that allows the user all necessary manipulations of the model that are encountered during model rebuilding.

The choice of the correct rotamer from the database of rotamers existing in O can be done. These rotamers can be further corrected in their torsion angles. The atoms can be shifted if necessary as rigid groups. Geometry optimization refinement is used after each rebuilding manipulation to bring the geometry of concerned atomic groups to the correct one. Finally real space refinement is available to fit the rebuilt model optimally to the electron density map in real space.

Most residues defined by ARP have correct conformation and there is no need for significant corrections except for several residues which have poorly defined density. These residues are often modelled with side chain atoms in the wrong region of the density because dummy atoms were assigned at their place. The proper modelling seems not to work because of the slightly incorrect density distribution for the side chain. Such residues are corrected to be in the density and dummy atoms are removed. After such a round of corrections the new complete model of the receptor is obtained that still lacks several residues of the C-terminal end.

In the case of the transducer the situation is different. In most cases the residues have to be buildt into the model or already existing Ser,Val and Ala residues have to be mutated to the proper one. Then the proper rotamer for each newly introduced residue should be defined. For this rotamer the torsion angles are corrected to match better the residue into electron density maps and in the end the real space refinement is performed which via the correlations between the observed electron density map and the extrapolated for the model electron density map defines the improved conformation. For the residues which were totally absent in the electron density maps first the backbone should be positioned in the density. In the beginning the helical fragment of the 5 residues was introduced into the model. This fragment of an ideal helix is properly positioned in the electron density and joined with the rest of the model. Then the already described procedure for the side chain building is applied. In the end the model of the transducer including residues 23 - 82 is completed.

4.9.3 Searching for the missing part of the transducer

Unfortunately the standard procedure of molecular replacement with existing model of the receptor still left almost one forth of the protein undetermined. This includes 15 amino acids of the C-terminus of the receptor, 23 amino acids of the N-terminus of the transducer and residues 83-159 of the transducer. Instead, a number of dummy atoms were introduced by ARP which represented the uninterpreted density stretches. Mostly these dummy atoms were located around the receptor molecule. They were arranged in a number of direct stretches parallel to the membrane normal. Almost no density was found in the middle region between two complexes related by a unit cell translation.

Taking into account that all the missing residues are located close to each other in space it is possible to conclude that their absence in the electron density maps is due to the lack of information carried in the current set of phases about this part of the space. An alternative explanation would be the global disordering of this part of the protein. In any case phases know only about a part of the molecule which was used for their calculation and the information which is in the structure factors amplitudes is not strong enough to reproduce the missing part in the electron density maps in a way that allows a confident interpretation. In Figure 35 the packing for the receptor/transducer is shown. It is clear that the missing parts of the proteins should

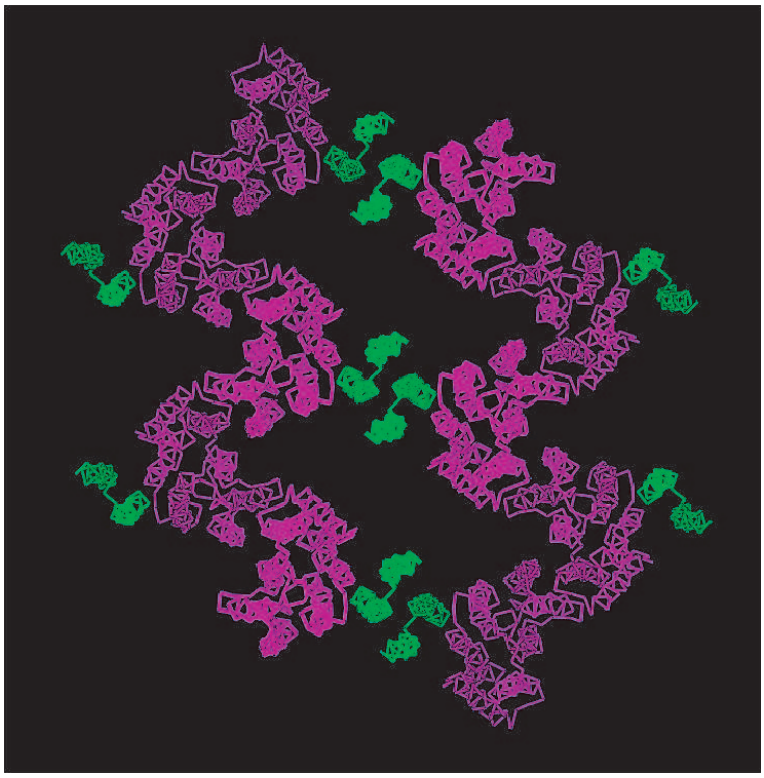

Figure 35: Crystal packing of the receptor-transducer complex (defined part). Violet - ribbon model of receptor molecules, green - ribbon model of transducer molecules

be located in the empty region between two complexes which as it is seen from Figure 34 do not contain sufficient electron density. This empty region contain 30 % of the unit cell volume. Judged by the number of unresolved residues 15 % are expected to be occupied with the missing part of the protein and only 15 % are left for the solvent. This consideration shows that the protein is actually tightly packed in the crystal unit cell and not so much conformational freedom is left for unresolved part.

Further implications arise from the consideration of the crystal packing. In the homodimer of two complexes they are interacting within four helical bundle TM1, TM2, TM1' and TM2'. As the model of the cytoplasmic part of the chemotaxis receptor demands two transducers should interact with each other almost along of their complete length forming a coiled-coil motive of four helices. This picture is not acceptable taking into account the molecular packing within the crystal and the crystal symmetry $P2_12_12$. The thickness of one molecular layer corresponds to the cell dimension c=53.84 Å. The length of the known part of the TM2 is ≈ 47 Å and only gap of 7 Å is left along the transmembrane helices as the next layer transducer is positioned beneath the current layer. The transducer can not be extended completely in the direction of the crystal c-axis within one layer of the crystal and hence it has to go to the adjacent layer or fold back in the same layer in the direction opposite to TM2 direction

One could suggest that the missing residues of the two transducers of the 2:2 complex dimer interact with each other in the linker domain with further tight coiled structure in the cytoplasmic part of the transducers. But these residues can not be close enough to each other to interact if they would originate from the molecules of the same 2:2 complex dimer, because they are separately placed in different free channels of the crystal. Exactly two transducers are located

in each compartment but they are not from the same 2:2 complex dimer, they are from the 1:1 dimers which has no direct contacts in the known part of the structure. These transducers can interact with each other being positioned symmetrically to preserve the $P2_12_12$ symmetry in the so far empty channel of the crystal.

To proceed in this situation with the structure solution the starting model for the missing part of the transducer should be introduced and should be positioned in the free compartment. Such starting model is not available. The missing sequence is expected to be to high extent α-helical. Part of the sequence is known to form the coiled-coil structure with itself, this sequence is manifested with the heptade motive $a-b-c-d-e-f-g$, with hydrophobic residues at first and forth positions, e.g. a and d ([41]). In the missing part of the transducer there is indeed such a sequence. It starts from residue Phe124 and lasts till the end of the truncated transducer. Taking into account these features one can proceed with the modelling of the missing part of the transducer.

Attempts were made to position the polyalanine helical fragments of 7-11 residues in the electron density map. It is important to notice that the map itself has revealed stretches of density at the 0.8 rms. These strechess repeated with a spacing close to a helical turn and were oriented in the proper way to represent a helix parallel to the crystallographic c-direction. On the other hand these stretches of density could also belong to lipids which surround the protein.

The thorough analysis of the electron density maps allowed to position the individual helical fragment. After positioning the electron density maps were recalculated using phases determined from the extended model. These new electron density maps presented more solid density for the new fragment, but the rms level of the map was lower than it should be for well defined atoms. Also there were some gaps in the density left. On the basis of this density the orientation 'up-down' of the helical fragment had been fixed. This was done by averaging the density by the rotation around the axis of the helix. Averaged density can show more clearly the orientation of C_β atoms and therefore allows to distinguish between up and down of the helical fragment. The position of the helical fragments were further corrected in the density map. In reciprocal space the rigid body refinement was used to adjust the position of the helix.

This procedure has shown that in the initial electron density map the helical fragments can be located. These fragments produce maps which are substantially improved for the backbone of the new helix. The question whether this position is true or not should be considered in the context of the new unbiased densities appearing in the map which should account for the side chains of the positioned helical fragment. Unfortunately all attempts to position such helical fragments though have shown electron densities that do not contain significant side chain density. Introduction of each new fragment has improved very little the overall R-factor of the structure. The introduction of several helical fragments at a time also has not resulted in new densities for the side chains. Though in this case the phases should be closer to the true phases than in the case of a single fragment. This way to solve the problem seemed not any longer to be fruitful.

The second type of the crystals was investigated in parallel. They are like the first type. It was not possible to reveal the missing part of the structure. In comparison to the first type these crystals have one dimension of the unit cell two times longer than in the conventional unit cell of the first type crystals. It is the dimension corresponding to the membrane layer thickness that is doubled. The space group for these crystals was also different. The doubled axis was screw-dyad axis. Therefore the space group for the type II crystals is $P2_12_12_1$. This space group actually reveals a very similar packing to that of the space group $P2_12_12$. The unit cell consists now of two layers which are identical to each other. The screw axis symmetry operation brings one layer to the top of the other. But in the comparison to the conventional type I these layers are not exactly upon each other but rather shifted in one direction relative to each other by ≈ 4 Å. The density in the free region of the unit cell was not as much different as for the type I unit

cell. It was not possible to prolong the defined part of the transducer (residues 23-83).

Several datasets of the 'single' and 'doubled' crystals were investigated, but in all cases there was the same negative result. It could be suggested that the region of interest is disordered in the crystal. But in this case disordered density pattern should be different for different crystal. In the same time the pattern of the rest density was very similar not only among the 'single' datasets or 'doubled' datasets but also between the 'single' and the 'doubled' datasets. This argument encouraged the further investigations.

4.9.4 Reconsidering the model of the crystal

The absence of the densities for the residues of the transducer could also arise from a mistaken interpretation of the data. Any data interpretation starts from the assignment of the space group symmetry . Usually the highest possible symmetry group is chosen for data interpretation, but in some cases a lower symmetry space group is correct. Therefore the space group of the crystal had been reconsidered. Consequently the data was reduced in the monoclinic space groups.

There are three possibilities for the monoclinic space group with one of the three orthogonal unit cell axis being the dyad axis at a time. The choice of the space group is made judging on the R_{merge} factor which compares the symmetry related reflections for a given space group. So for orthorhombic space group the intensity of the following reflections is averaged: hkl, $h\bar{k}\bar{l}$, $\bar{h}k\bar{l}$, $\bar{h}\bar{k}l$ and the differences between each of the reflections and average value are summed up. In the monoclinic space group only the intensities of the reflections connected by the dyad rotation symmetry are averaged, e.g. hkl and $h\bar{k}\bar{l}$ for the space group with a-axis as a dyad. The R_{merge} factors for the different choice of the space group are shown in the Table 3. The difference in the R_{merge} is very small.

Table 3.

Space group	R_{merge}	Cell constants
$p2_12_12$	0.075	124.53 46.80 53.76 90^o 90^o 90^o
$p2_1$	0.070	53.80 **46.82** 124.52 90^o 89.98^o 90^o
$p2$	0.069	46.84 **53.82** 124.52 90^o 90.05^o 90^o
$p2_1$	0.064	46.79 **124.45** 53.75 90^o 90^o 90^o
$p1$	0.063	46.82 53.83 124.52 89.94^o 89.93^o 89.89^o

The satisfactory description in the orthorhombic space group of the receptor protein means that if there are deviations from the orthorhombic symmetry then for the receptor this deviations should be small. In the not yet defined region the deviations could be bigger.

The unit cell contains four complex heterodimers. In the case of the orthorhombic symmetry all four are symmetry related and only one species is independent. In the case of the monoclinic symmetry two heterodimers are related to another two heterodimers. Depending on the particular symmetry axis different pairs of the complex could be independent. If a is a monoclinic dyad axis then two heterodimers in the homodimer are different and two homodimers with up and down orientations are equivalent. The same will be in the case if b is a symmetry axis. And in the case when c is a dyad axis two heterodimers in 2:2 complex are equivalent and up and down oriented homodimers are different. From the general consideration there is no reason for up and down oriented homodimers to be different. So the first two cases are favorable. Taking into account the R_{merge} factors a becomes the most prominent candidate to be the symmetry axis.

Table 4.

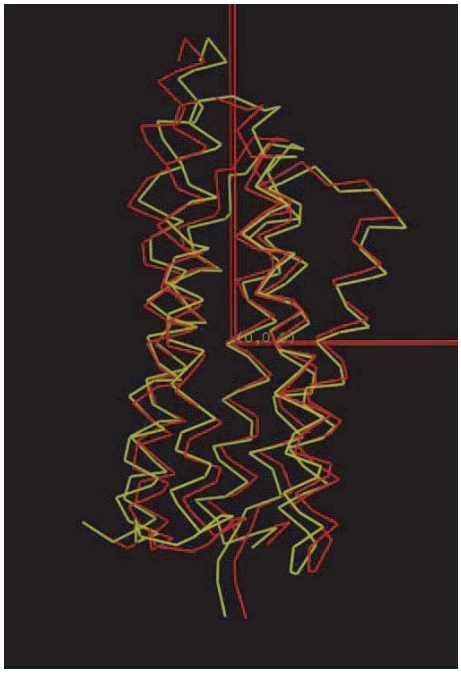

Figure 36: 4 degree difference in rotation angles between two solutions of molecular replacement rotation search for pSRII

Solution	α	β	γ	Rf/σ
1	264.48	4.28	274.48	8.35
2	87.05	4.27	272.06	7.73
3	180.56	0.00	0.00	6.17
4	0.00	0.00	0.00	5.65
5	352.51	51.71	10.16	4.79
6	357.50	76.74	183.45	4.62
7	177.38	90.00	7.83	4.58
8	357.50	55.34	2.14	4.51
9	172.34	51.53	9.74	4.49
10	175.42	17.54	192.68	4.43

Molecular replacement was performed in the monoclinic space groups for the data from the 'single' type crystals. The list of the solutions for the rotation search in the case of the monoclinic space group with $P2_1$ symmetry is shown in the Table 4 (α, β and γ - Euler rotation angles, Rf/σ - rotation function divided by its RMS).

There is still the same solution with the highest score as it was for the orthorhombic space group. But since two independent molecules can be placed in the unit cell and so for the monoclinic space group two solutions should be chosen from the list, the attention should be drawn to the second solution. This solution is nicely distinguished from the following group of solutions with approximately the same score as the first solution . The rotation angles parameterizing this solution are very close to those of the first solution. Namely the difference is only 4 degree rotation around the b-axis. The second solution for the rotation function can be the true solution (Figure 36). Several datasets were reduced in $P2_1$ symmetry group and for each of them the

molecular replacement procedure was performed. In all cases the second solution of the rotation function which is different in 4 degree rotation from the first solution was found. This interesting fact allowed to conclude that the second solution is not an artifact of the molecular replacement method and not the consequence of the random errors.

The second solution was accepted and with a new model for packing in the space group $P2_1$ electron density maps were calculated. The R-factor was increased to 0.45 for the packing according to both solutions. And the striking difference to the electron density maps in the orthorhombic space group was observed. In the region which has had no density before the average level of the density has increased. A new round of density interpretation was started. Density stretches appeared at regular distances which nicely matched the distance corresponding to the step of α-helix. Again helical fragments of 7-11 residues were incorporated into the electron density maps. But in this case it was possible to get a protein backbone trace which would account almost for all the missing transducer part. Helixes were positioned in the initial density map with only some breaks in between. It should be emphasized that the found trace is not the only one possible as a satisfactory density appears at the 0.8 rms level.

The electron density maps were recalculated with inclusion of the newly positioned helical fragments. Not only $2F_o - F_c$ maps were used but also the difference Fourier density maps $F_o - F_c$. Upon the inspection of this new maps the positions of the helices were confirmed or rejected. The modified model for the missing part was further subjected (like it was already described) to rigid body refinement. Again the electron density maps were calculated and new helical fragments were searched for. This procedure were undertaken several times and finally a set of the helices was established which described satisfactorily the initial densities.

The next step would be the building and the assignment of the side chains. But unfortunately the densities were too weak to establish the side chains with any confidence. This situation implies that the ordering of the searched atoms is insufficient or some other facts should be taken into account. The progress which has happened in the model building and description of the electron density maps is shown in Figure 42 (green density). Still the current model of the crystal was not satisfying.

4.9.5 Twinning

Again the crystals were reinvestigated. Twinning was considered as a possible cause for the problems encountered (Parsons et al. 2003 [83]). For the rhodopsins with crystals grown in the cubic lipidic phase the packing of the membrane layers of the protein is essential. The interaction between the membrane layers is not strong as is demonstrated by the mosaicity of the crystals in the direction normal to the layers and also the decay of the diffraction pattern with resolution. Because of this weak interaction and also for some other reasons (Efremov et al. 2004 [84]) in the case of Bacteriorhodopsin crystals were twinned. A similar phenomena is not excluded for the case of the crystals of pSRII-HtrII complex.

A twinned crystal is an aggregate in which different domains are joined together according to a specific symmetry operation: the twin law (Figure 37). The diffraction patterns derived from different domains are rotated, reflected or inverted with respect to each other, depending on the nature of the relationship between the different domains, and weighted according to the quantity of a particular domain present in the crystal. The diffraction pattern measured during data collection is a superposition of all of these. Reflections from different domains may overlap and twinned crystals fall broadly into two categories in which either all reflections or only certain zones of reflections are affected by overlap. In the case of Bacteriorhodopsin the overlap of all reflections was observed.

Merohedral twinning may occur when a unit cell (or a supercell) has higher symmetry than

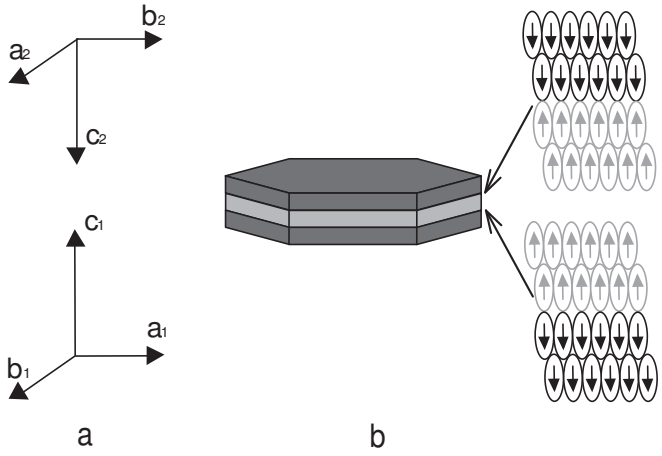

Figure 37: Example of the twinning. Crystal consists of several twin domains which have an identical crystalline structure. From Efremov et al. [84]

implied by the space group of the crystal structure. For example, the case of the symmetry mmm and monoclinic symmetry $P2$ with β angle equal to 90^0. The shape and dimensions of the unit cell in this case are the same for both point symmetry groups, but it is possible to overlay the unit cells in two different ways. This two orientations are related with the symmetry operator which belongs to the mmm symmetry group. Therefore the crystal can be arranged as a combination of both orientations of the monoclinic space group. In this case twinning could happen. Twinning will only be observed if intermolecular interactions across a twin boundary are energetically competitive with those that would have been formed in a single crystal.

The relative orientations of the diffraction patterns from different domains are the same as the relative orientations of the domains.

Twinning is a problem in crystallography because it causes superposition or overlap of reflections which are not related by crystal symmetry (Yeates et al. 1997 [85]). In Figure 38 diffraction patterns are shown for the domains alone and the case of superposition. Since the lattices of the twin domains are identical, the reciprocal lattices overlap, so that each measured diffraction intensity contains contributions from two (or more) twin reflections. Because of this exact overlap, the diffraction pattern from a merohedrally twinned crystal appears 'normal' and the anomaly is therefore easy to overlook; it is impossible to recognize this type of twinning directly from a diffraction image. It is necessary to note that in the case when the domains are equally weighted the diffraction pattern symmetry is higher than the symmetry of the separate domain diffraction pattern.

Assuming that the twin domains are larger than the coherence length of the X-ray beam, they will scatter independently. In this case, the observed intensities can be described by a summation of the intensities of the twin components. Each observed intensity consists of contributions from two reflections related by the twin operation,

$$I_{obs,1} = (1-\alpha)I_1 + \alpha I_2 \tag{86}$$

$$I_{obs,2} = \alpha I_1 + (1-\alpha)I_2 \tag{87}$$

Once the twin operation and α are known, the contributions of each twin can be separated, i.e. the data can be detwinned. Perfect twins, when $\alpha = 0.5$ cannot be detwinned without additional information (reference).

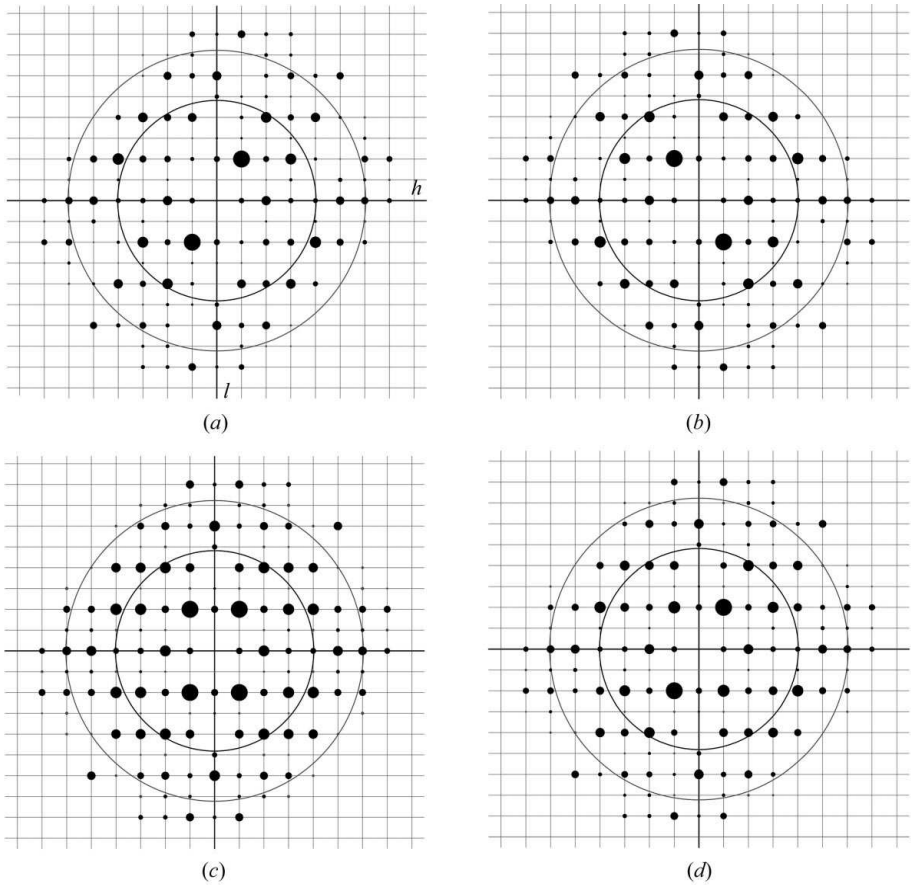

Figure 38: The effect of twinning by a two fold rotation about a-axis on the diffraction pattern of monoclinic crystal. (a) $h0l$-zone from a single crystal, representing diffraction from one domain. (b) same pattern but rotated around a. (c) superposition of a and b simulating the perfect twinning $\alpha = 0.5$ case. (d) superposition of a and b simulating twinning with $\alpha = 0.2$. From Parsons [83]

4.9.6 Detection of the merohedral twinning

Several tests based on various properties of the intensity distributions have been proposed and used in practice.

Moments of intensities distribution One of the consequences of twinning is the smaller fraction of very weak as well as very strong intensities in the entire population of reflections (Stanley et al. 1972 [86]). This is analogous to the difference between diffraction patterns of centrosymmetric and non-centrosymmetric crystals. In a quantitative representation, the variance (and higher moments) of the intensity distribution is smaller for twinned crystals than for single specimens. The Wilson ratio (reference) $< F >^2 / < I >$ is expected to be 0.785 for normal non-centrosymmetric reflections and 0.885 for such reflections from twinned crystals. Because of the same reasons the ratio $< I^2 > / < I >^2$ is equal to 2.0 for non-twinned and 1.5 for perfectly twinned crystals.

Britton plot of negative intensities As pointed out above, the intensities measured from a twinned crystal are combined from two twin domains with weights proportional to their volumes.

From the equations 86, 87 one can get for $\alpha < 0.5$

$$J_1 = [(1-\alpha)I_2 - \alpha I_1]/(1-2\alpha) \tag{88}$$

$$J_2 = [(1-\alpha)I_1 - \alpha I_2]/(1-2\alpha) \tag{89}$$

Britton (1972 [87]) pointed out, that the assumption of a too large value of α results in an estimation of negative true intensities. Based on this principle a practical method for the estimation of the twin fraction α was proposed. The 'Britton plot', giving the number of negative intensity estimations as a function of the assumed value of α in the detwinning procedure, has two linear asymptotes, one for $\alpha < \alpha_{opt}$ and another for $\alpha > \alpha_{opt}$. The point at which these two lines cross gives the estimated value of the twin fraction.

Rees $N(z)$ ***plot*** The argument z is the fraction of the average intensity, calculated in narrow resolution ranges, and N(z) is the fraction of reflections with intensities below this level. Various intensity distribution functions show different N(z) curves, particularly for low values of z. As noted above, the fraction of weak intensities is lower for non-centrosymmetric crystals than for centrosymmetric crystals and is lower still for twinned crystals, which shows up clearly on the N(z) plot. This criterion was proposed and theoretically worked out by Stanley (1972, [86]) and elaborated by Rees (1980, [88]), who gave the general formula for the cumulative intensity distribution for a non-centrosymmetric case with the twin fraction α. As Stanley pointed out, the $N(z)$ curve for twinned crystals is characteristic 'in having an opposite initial curvature'; it has a sigmoidal shape in contrast to an exponential character for normal crystals.

Yeates S(H) plot This test, proposed by Yeates (1988, [85]), is based on the behavior of the ratio of the difference to the sum of intensities of reflections related by the twin operation,

$$H = \frac{|I_1 - I_2|}{I_1 + I_2} \tag{90}$$

The dependence of the cumulative distribution of this parameter, $S(H)$, on the twin fraction is very simple and is linear in H for non-centrosymmetric crystals. The slope of the $S(H)$ plot is $1/(1-2\alpha)$ and depends on the twinning factor more sensitively than the $N(z)$ plot. Apart from the cumulative distribution $S(H)$, the average values $< H >$ and $< H^2 >$ are characteristic for particular twin fractions. For non-centrosymmetric reflections $< H >= 1/2-\alpha$ and $< H^2 >= (1-2\alpha)^2/3$.

4.9.7 Twinning possibility for pSRII-HtrII complex crystals

The lattice of the crystal has the orthorhombic symmetry and as it was suggested before the space group could be only $P2_1$. In this case it is not excluded that the crystals can be composed of domains of $P2_1$ symmetry which are oriented in opposite directions, e.g. crystals could be merohedrally twinned. As it was already noticed due to the arrangement of the protein in membrane-like layers which interact with each other such a type of twinning could be observed. And in the case of Bacteriorhodopsin crystals indeed were twinned.

For the crystals of the complex, if they are twinned the situation would be different from that which was in the BR case. Refinement of the structure in $P2_12_12$ space group has shown that the receptor molecule is nicely described with the orthorhombic symmetry. And only in the second approximation the suggestion was made about the deviations from this symmetry in the form of a slight rotation of 4 degree. That means that for the known part of the structure two domains of $P2_1$ symmetry would be very close to each other in the electron density distribution, e.g. $\rho(r) \approx \rho(Ar)$. This is different from the case of Bacteriorhodopsin were two twinning related domains have significantly different electron density functions. That is why the question of whether crystals of the complex are twinned or not should be considered with special care.

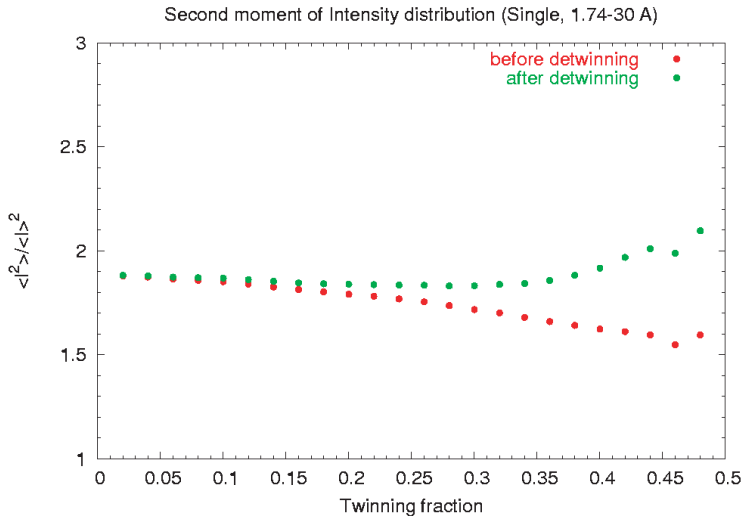

Figure 39: Plot of the second moment of the intensity distribution for the ground state data collected from pSRII-HtrII complex crystals

The analysis of the R_{merge} for the different possible space groups has shown that the differences between the monoclinic space groups and the orthorhombic are very small. This suggests the orthorhombic space group as a true symmetry of the crystal. One way to reconsider the model of the crystal was to suppose that the symmetry is monoclinic and to postulate the presence of twinning. In this case the monoclinic domains can mimic the apparent orthorhombic symmetry because of the twinning if the twin fraction is close to 0.5. In this case the deviation from 0.5 can explain the lower R_{merge} when data is indexed in monoclinic space group.

This model looks like a rather complicated presentation of the crystal but since it could resolve the problem of the missing part it is worth to consider this model. Some prove should be given to the presence of the twinning.

In Figure 39 the second moment of the intensity distribution is shown for the whole data range and only the high resolution data, together with a correlation coefficient of twin-related intensity. As a twin law the rotation operator of 180^0 around the b-axis of the crystal is considered. From these plots indeed it can be inferred that the data could be merohedrally twinned as two curves - one for original data and another for the detwinned data according to equations 88, 89 - are separated for higher twin fractions. This would not happen if the data would be purely orthorhombic, because the twin related reflections in this case should be equal by symmetry. The twinning factor as it was already suggested from the R_{merge} differences should be close to 0.50. Also the correlation coefficient curves show the drop of the correlation coefficient after the detwinning how it should be expected in the cases of merohedral twinning. There is no substantial difference between data for the whole resolution range and only the high resolution range.

The Britton plot is shown in Figure 40. It also suggests that the data could be twinned with twinning fraction more that 0.4.

Unfortunately the Yeates statistic and Britton plot can not distinguish the true symmetry case from the symmetry introduced by twinning at high twin ratio.

Yeates statistic still can be useful for the determination whether the data are twinned or not. The H value being the contrast between the measured twin related intensities, e.g. $(I_2 - I_1)/(I_2 + I_1)$ due to the equations 88, 89 is also the contrast between the real intensities multiplied by a

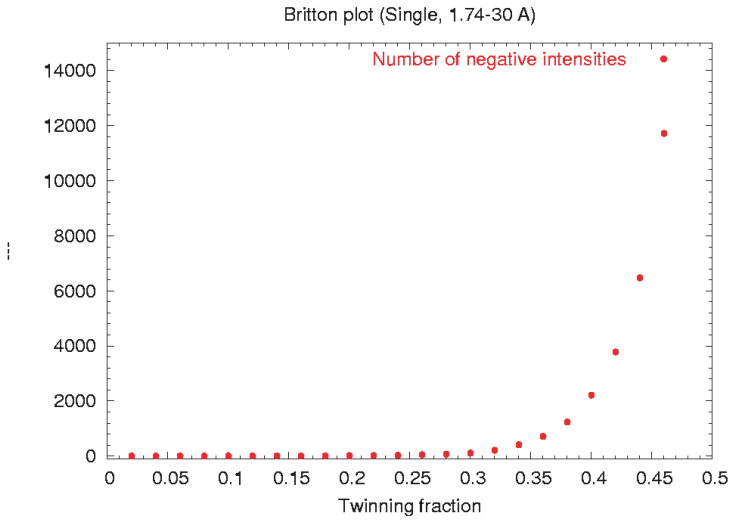

Figure 40: Britton plot for the ground state data collected from pSRII-HtrII complex crystals

factor of $(1-2\alpha)$. When the density distribution of two unit cells is very similar, the difference in intensities of the same reflections is almost absent at low resolution and it becomes more prominent at high resolution where the small differences in a structure start to play a role. The same will be the case for two twin domains when there is only small asymmetry which makes them different. The low resolution reflection will almost be the same as in the case when the twin operation belongs to the true symmetry and the high resolution reflections will show some deviations. Therefore the difference between low resolution and high resolution should rise. In Figure 41 the Yeates cumulative distribution plotted for the whole resolution and only the high resolution ranges is shown. From this plots one can identify the relative shift of the curves for whole resolution and high resolution only. This shift is in line with the suggested model of the crystal and the expected density distribution, e.g. the density within the twin domain deviates not significantly from the higher $P2_12_12$ symmetry. This fact can be considered to be another confirmation of the suggested model. On the other hand this difference can arise also from simple statistical error in the observations of the structure factor amplitudes. The statistical error also grows with resolution, higher resolution reflections has more differences due to the error and higher contrast value. Unfortunately the Yeates statistic can not distinguish between this cases, but if the statistical error effect is diminished for example by selection of only confidently measured reflections, then the difference between resolution ranges becomes an indicator of the twinning. In Figure 41 plot reflections are selected with the ratio of I/σ more than 3.

4.9.8 Electron density map for the twinning case

The introduced model of the crystal with the twinning has to be furthermore approved because there is no significant evidence from the statistic on the twin intensities. From this statistic the high twin ratio of 0.45-0.5 is expected. The crucial argument for this model will be the final impact of the twinning on the electron density maps.

In the case of the perfect twinning with the twin fraction of 0.50 percent a special type of electron density maps is calculated. The twinned structure factor amplitudes are first corrected

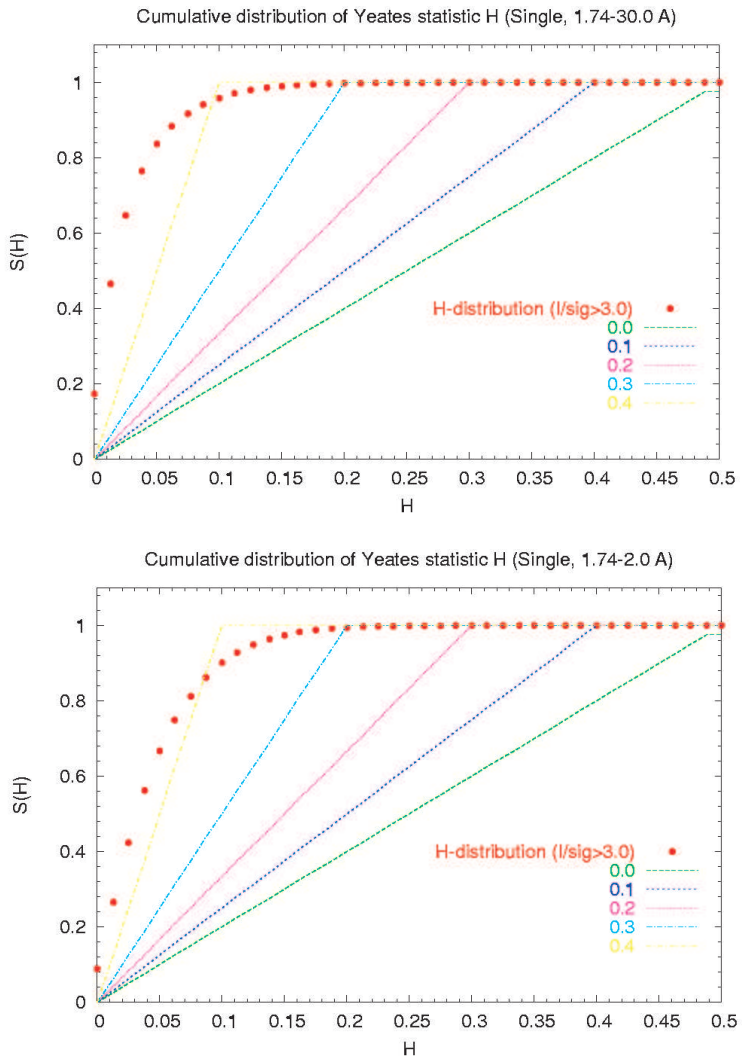

Figure 41: Yeates distribution for the ground state data collected from pSRII-HtrII complex crystals

and then combined with the phases obtained for a given model. In the case of perfect twinning two equations 86, 87 become similar and

$$I_{obs}(\mathbf{h_1}) = I_{obs}(\mathbf{h_2}) = I(\mathbf{h_1}) + I(\mathbf{h_2})/2 \tag{91}$$

Additional equations come from the model amplitudes

$$I(\mathbf{h_1}) \approx I_{calc}(\mathbf{h_1}) \tag{92}$$

$$I(\mathbf{h_2}) \approx I_{calc}(\mathbf{h_2}) \tag{93}$$

for the model closely resembling the actual structure. Thus

$$I(\mathbf{h_1}) = 2I_{obs}(\mathbf{h_1}) - I_{calc}(\mathbf{h_2}) \tag{94}$$

$$I(\mathbf{h_2}) = 2I_{obs}(\mathbf{h_1}) - I_{calc}(\mathbf{h_1}) \tag{95}$$

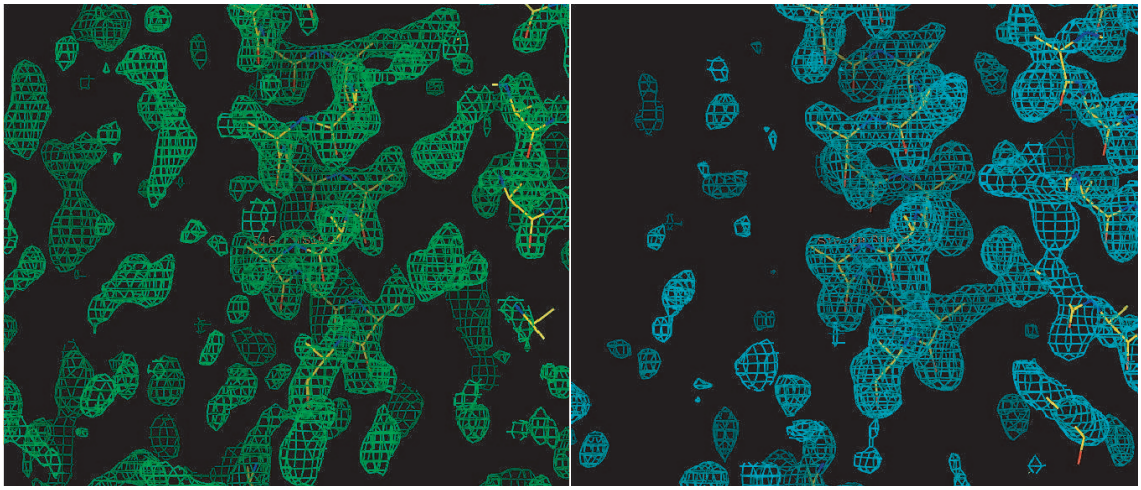

Figure 42: $2F_o - F_c$ electron density maps for located polyalanine helical fragments. Green - electron density map for the case of crystal with two orientations of pSRII following molecular replacement solutions. Blue - electron density map for the twinned model of the crystal. Both maps are contoured at 0.8σ

Two estimations of the intensity are given for each domain. Least squares minimization reduces this equation pairs to the arithmetic mean of the two estimates,

$$I_{detwin}(\mathbf{h_1}) = \left(I_{obs}(\mathbf{h_1}) + I_{calc}(\mathbf{h_1}) - I_{calc}(\mathbf{h_2})\right)/2 \tag{96}$$

$$I_{detwin}(\mathbf{h_2}) = \left(I_{obs}(\mathbf{h_1}) + I_{calc}(\mathbf{h_2}) - I_{calc}(\mathbf{h_1})\right)/2 \tag{97}$$

From these estimates of detwinned structure factor amplitudes can be obtained and upon the combination with the phases calculated from the model electron density maps in the case of twinning can be calculated.

The electron density maps taking into account twinning has significantly changed in quality. The most significant improvement were observed for the receptor molecule placed according to the second solution of the molecular replacement in the unit cell. Without taking into account twinning the densities for the receptor were not satisfactory in the peripheral part of the model were the displacement upon the 4 degree rotation is appreciable. The 'twinned' maps have revealed a better agreement of the densities with the model of the rotated receptor molecule. Another important improvements were observed for the polyalanine alpha-helical fragments. The example of such an improvement is shown in Figure 42. On the 'twinned' maps the backbone fragments of the polyalanine helices representing the missing parts of the transducer are almost of the same good quality with the electron densities of the well-defined receptor backbone. There is also an improvement in the densities which should be attributed to the side chains. There is significantly better connectivity with the densities of the expected side chains and the backbone densities. But still this densities are evident only for 1-2 residues from ten. This is not enough to proceed with the determination of the structure. So the introduction of the twinning model of the crystal has brought improvements but in spite of this it is still not possible to determine the side chains of the polyalanine helices correctly.

After the analysis of the electron density maps the model for the folding of the missing part of the transducer is proposed. The hypothetical folding is shown in Figure 43.

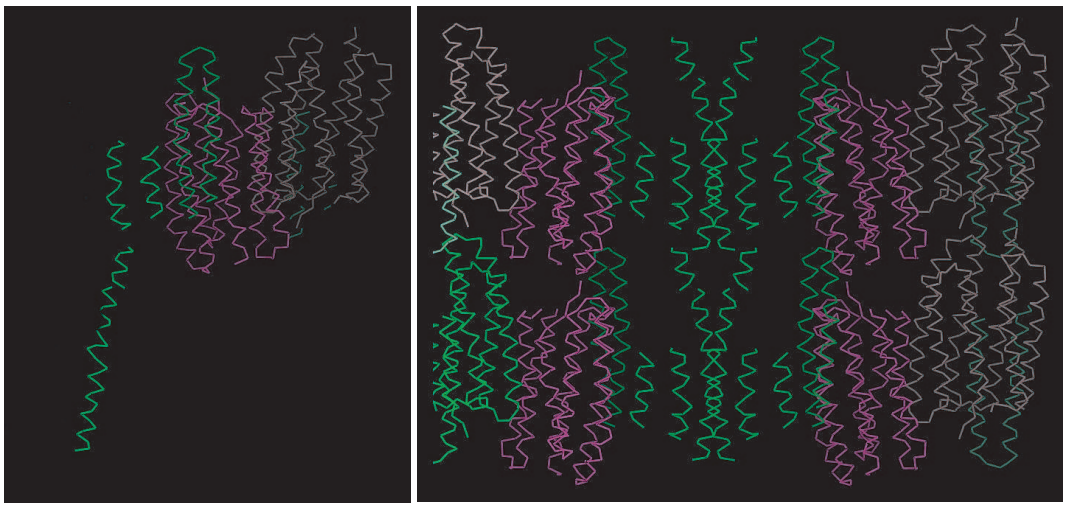

Figure 43: Putative trace model of the transducer including several helical fragments derived for the twinned crystal model

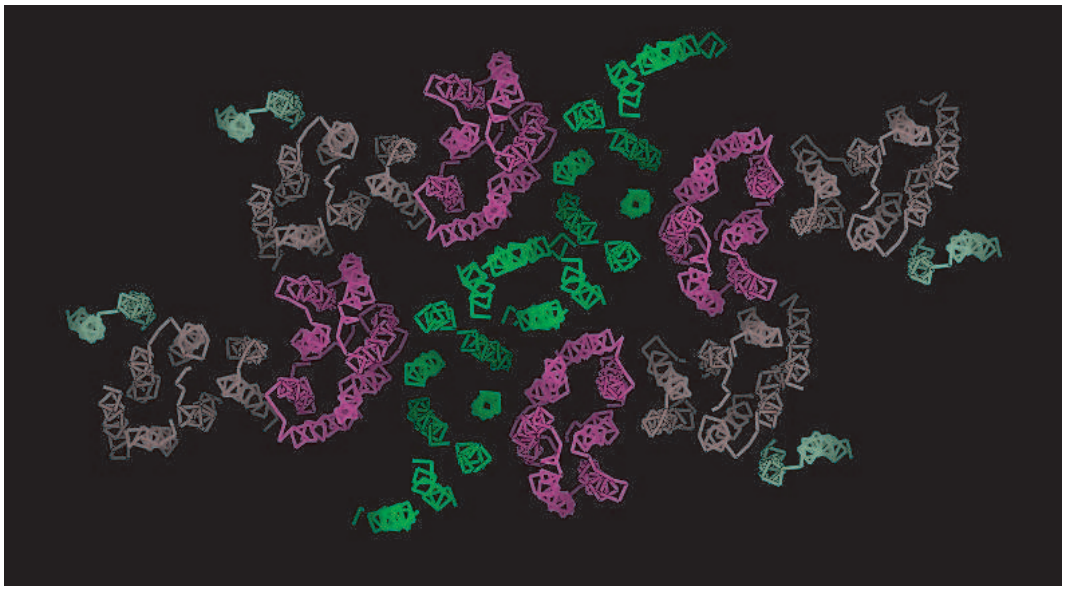

Figure 44: Single crystal layer consisting of homodimers of pSRII-HtrII complex interacting via transducers. Receptors of interacting dimers are shown in magenta and corresponding transducer in green.

The presented folding can nicely account for the preservation of the lattice parameters after introduction of the longer transducer to the crystal. The space along the crystallographic c axis is enough to accommodate the hydrophilic part of the transducer.

An interesting consequence of the transducer-transducer interaction is the formation of the two dimensional lattice of the molecules (Figure 44). Such a lattice was already suggested in the discussions ([55]). It's role is to amplify the signal of one particular receptor/transducer complex.

4.10 Isomorphous replacement

The method of isomorphous replacement was tried for the solution of the ground state structure of the receptor complex with the long transducer HtrII(1-159). Namely after the application of the molecular replacement method which allowed to obtain the initial phases for the structure factors the structure of the part of the transducer starting from the residue 82 remained unknown. The new independent source of phases could improve the situation. The use of the isomorphous replacement method was suggested to gain further improvement in structure solution.

The heavy atom compound potassium tetrachlorplatinum(II) was chosen as it reacts most commonly with methionine and cystine residues in acidic and neutral environment. Three structured methionine residues are located on the surface of the receptor pSRII and are likely to be potential binding sites. The sequence of the unknown part of the transducer structure also contains several methionines. Locating these methionine residues would allow to get an idea about the trace of the missing part and allow to create initial models which could be used further to progress in phasing.

Selected crystals of pSRII-HtrII(1-159) complex were soaked in K_2PtCl_4 containing buffer. Two different approaches were used. Slow soaking in low molarity buffer during 1-3 days, and quick soaking in high molarity Pt-buffer during 15-90 minutes. Only very moderate screening of soaking conditions was undertaken.

Several complete datasets with resolution varying from 4 to 2.5 Å were collected at the X13 beamline, DESY (Hamburg) and the ID29 beamline, ESRF (Grenoble, France). For MAD experiment datasets from the same crystal were collected at three wavelengths corresponding to L3 anomalous absorption edge of Pt: energy absorption peak wavelength, inflection of the peak wavelength and remote from the peak wavelength. Preliminary scan to find the exact energy value corresponding to the absorption edge was recorded with the X-ray fluorescence detector. The following wavelength were chosen during the data collection for MAD experiment: $\lambda_{peak} = 1.$; $\lambda_{inflection} = 1.$; $\lambda_{remote} = 1.$

A good agreement of the cell parameters was observed between the datasets of the heavy atom derivatives. The maximum difference in the cell parameters was found to be 0.2%. Data analysis showed that for isomorphous replacement derivative data scales to the native data without the significant increase of the R-factor which is defined as follows:

$$R = \frac{\sum_{hkl} F^{iso}_{hkl} - F^{nat}_{hkl}}{\sum F^{nat}_{hkl}} \tag{98}$$

This means that both the native and the derivative crystals are good isomorphs of each other. In the same time the value of the R-factor clearly shows that there is a systematic difference between these two datasets which indicates the presence of the bound Pt in the derivative crystals. This was encouraging starting point for the search for the location of the heavy atoms.

The typical curve for the R-factor after the scaling of the derivative to the native data is shown in Figure 45. The scaling is isotropic so that the reflections chosen around the principal directions (h00) or (0k0) or (00l) shows the same pattern of the scaling R-factor. The resolution to which the datasets are isomorphous reaches the diffraction limit of the data.

All attempts to locate the heavy atom positions in the crystal were done via the analysis of the difference Patterson maps. The proper native set was selected. The differences between the scaled F_{PH} and F_P were used as the amplitudes for the isomorphous difference Patterson map calculation.

The heavy-atom-heavy-atom vector peaks should be the strong peaks among all the peaks of the difference Patterson map denoted as (u, v, w). This peaks was searched on so called Harker sections of the Patterson map. As it was pointed out by Harker the symmetry operations of

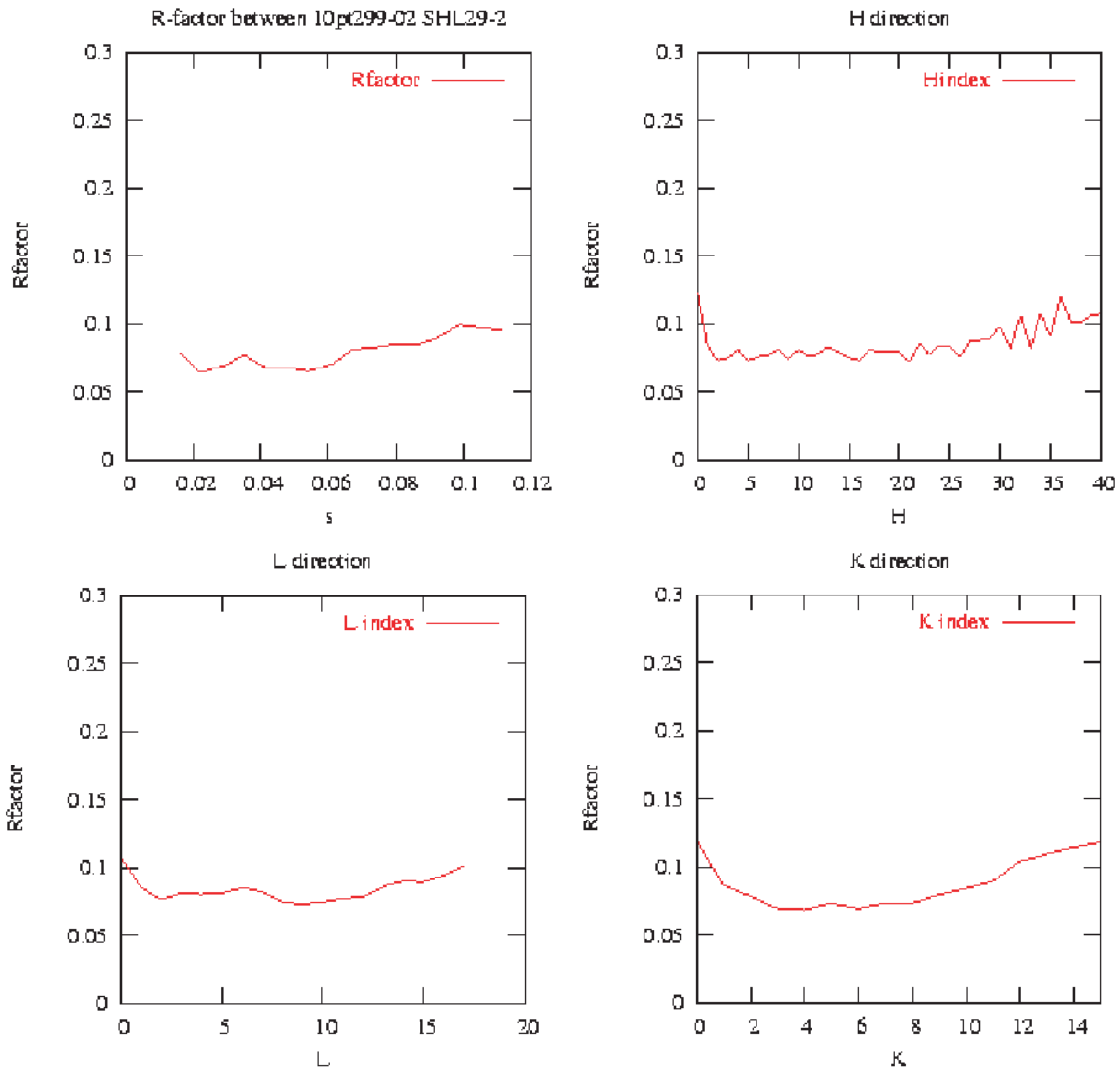

Figure 45: Scaling of the derivative data to native data. Overall R factor and R factor for the selected axial direction (h00), (0k0) or (00l) are presented

the crystal define the planes where the heavy-atom-heavy-atom vector peaks should be located. For the case of crystals in the space group $P2_12_12$ there are 4 symmetry operations relating coordinates X, Y, Z in the crystal with symmetry equivalent coordinates: 1.X,Y,Z 2.-X,-Y,Z 3.1/2-X,1/2+Y,-Z 4.1/2+X,1/2-Y,-Z From this symmetry relations one could define the family of vectors connecting the symmetry equivalent points. Among these vectors are those relating the one heavy atom position with the other heavy atom position within a unit cell. By calculating the differences between the coordinates listed above one can get the equations of three planes which are known as Harker sections. In our case the cross-vectors families are $(2X, 2Y, 0)$, $(1/2{-}2X, 1/2, -2Z)$ and $(1/2, 1/2{-}2Y, -2Z)$ and the corresponding Harker sections are $u = 1/2$, $v = 1/2$, $w = 0$. This three sections should be searched for the prominent peaks which will

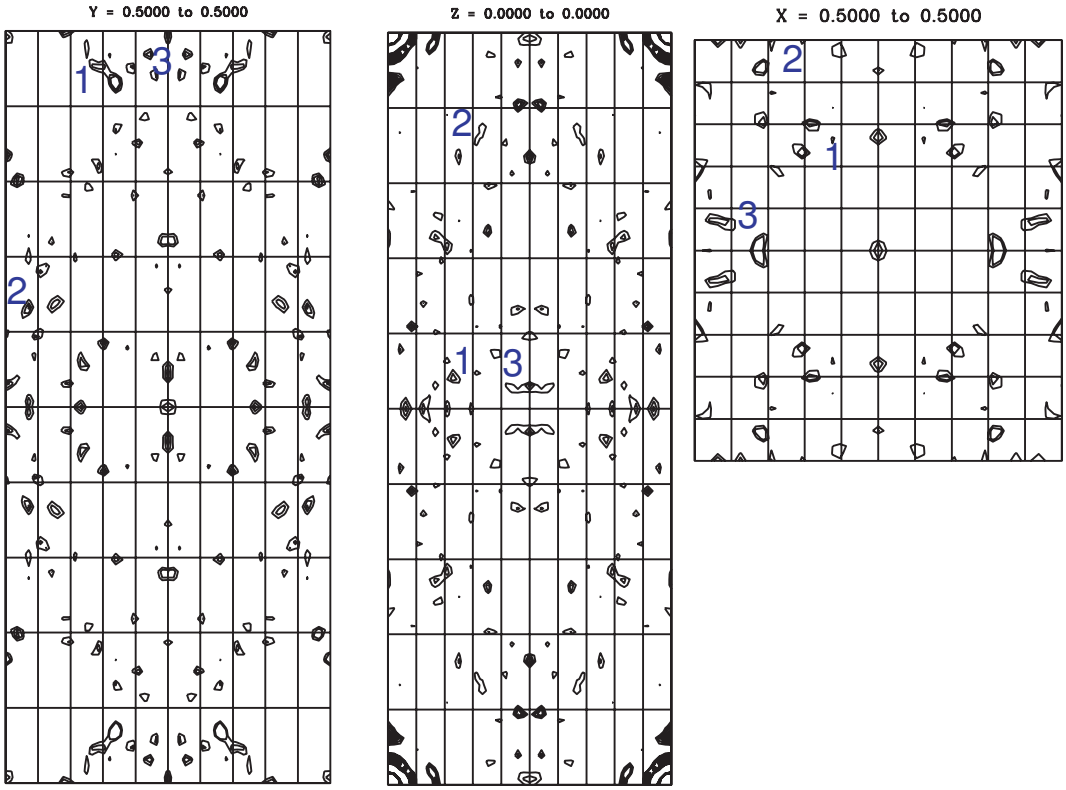

Figure 46: Harker sections plot for the isomorphous Patterson map calculated for $F_{PH}-F_P$

represent the heavy atom site. If the peak can be located in all the Harker sections then this peak is considered for the further refinement. From the (u,v,w) coordinates of the peak one can define directly the (X,Y,Z) location of the heavy atom site. Three Harker sections for one of the datasets are shown in Figure 46, several possible solution are chosen.

To look for the possible solutions is also possible in the automatic way which is implemented in the program RSPS from the CCP4 program suite ([81]). RSPS is a grid search program in a Patterson vector space. Searches are carried out by assigning trial positions on a grid covering the asymmetric unit of the crystal, and then computing a score for each trial position, based on the Patterson densities at the positions corresponding to the predicted vectors for each position. The approach consists of the scanning of the volume of the unit cell (usually one asymmetric unit). In single and more site searches, the scan parameter is the coordinates of a heavy atom position. Test points assigned on a grid within this volume. For each test point, the values of the Patterson function at the predicted Patterson vector positions are collected and summed. The map resulting from a search will contain peaks at the positions of possible heavy atom (or molecule) sites and at positions related to these by a shift in origin or by inversion. In the case of several heavy atoms present the cross-vectors between them are taken into account. And the current heavy atom is scanned with the fixed positions of already located atoms.

Previously obtained solutions separately or combined are subjected to vector-space refinement. This refinement minimizes the squared difference between the observed and calculated heavy atom difference Patterson function values. The minimization in vector space is only done for sample points where the calculated Patterson density is significantly positive, in other words near the interatomic vector peaks. For vector-space refinement the theoretical convergence radius is the apparent atomic diameter, which is about $dmin/\sqrt{2}$. The vector space refinement is

implemented in the program VECREF from the CCP4 program suite ([81]). The relative occupancies for the sites with the supplied positions are defined during the vector-space refinement though their estimates are not accurate so far.

Altogether two possible solutions were identified from the automatic search combined with the analysis of the Harker sections. Initial occupancies estimated by relating the observed and calculated Patterson were 0.25 for the first solution and 0.12 for the second solution. Obtained information about the sites were used to calculate the initial phases. The refinement of the heavy atom parameters were made in reciprocal space within the program MLPHARE from the CCP4 program suite ([81]). Refinement tends to minimize the lack of closure sum for all the reflections. In MLPHARE this refinement is performed via the maximum likelihood target. The most probable phases are determined and the corresponding figures of merit are calculated. When the individual solutions were supplied to the refinement procedure using only centric reflections the calculated phases has shown mostly the figures of merit between the 0.2 and 0.3 for the best solution. The scattering of the heavy atom was $\approx 1/20$ of the total protein scattering. Unfortunately the occupancies after several refinement cycles has dropped down though the found sites were not totally rejected. The Cullis R-factor $R_{Cullis} = 0.95$ and the *Phasing Power* $= 0.42$. Both this indicators have not satisfactory values showing that the signal from the heavy atom does not exceed the lack of closure and so do not allow a good determination of phases. Situation has not changed dramatically after two sites were refined simultaneously.

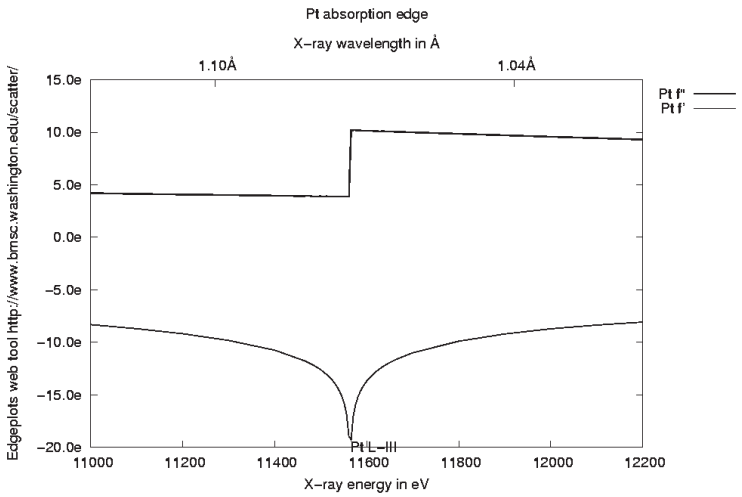

Figure 47: Platinum L-III edge. F' and F'' dependence on the energy of X-ray photons is represented.

4.11 Anomalous scattering

Next the anomalous scattering data were analyzed. Three complete datasets at three wavelength were measured. These wavelength are correspondingly the peak point, inflection point and the point remote from the peak on the Platinum LIII absorption edge. The theoretical curve for F' and F'' is shown in Figure 47. From this curve the following differences in the scattering factors in comparison to the scattering F^0 were introduced (Table 5).

Table 5

	f'	f"
inflection	-19e	4e
peak	-17e	10e
remote	-13e	10e

These differences are expected to generate the differences in the data. Real anomalous differences were in the range of 5% of the reflection intensity magnitude. This is considered to be a good starting point for data treatment.

All the data from anomalous scattering experiment were scaled in the same run with SCALA. Anomalous differences between data measured at different wavelength has shown the correlation of ~ 0.4 between pairs of datasets. For the dispersive difference with inflection point taken as a reference dispersive differences $F'' - F'$ and $F_{remote} - F'$ has shown a correlation coefficient of ~ 0.8. This correlation coefficients are expected to be high as each correlated difference is proportional to the differences listed in Table 5.

The Patterson maps based on dispersive differences between the datasets were calculated as well as Patterson maps with anomalous differences as coefficients. This maps were analyzed together with the previously reported isomorphous difference Patterson maps. Harker sections for difference Patterson map is shown in Figure 48. Two solutions already confirmed from the isomorphous difference Patterson maps were present in all maps. These solutions was once again analyzed taking the anomalous data into account. Occupancy and coordinates defining

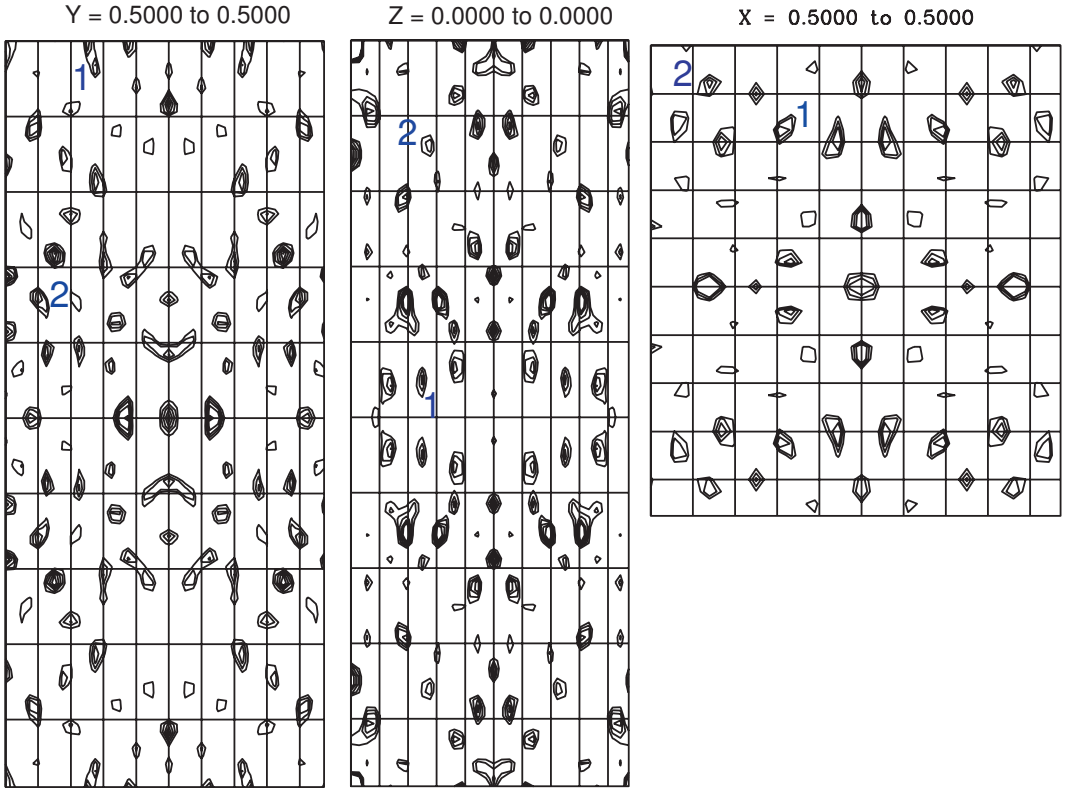

Figure 48: Harker sections plot for the anomalous difference Patterson map calculated for $F_+ - F_-$

the position of the heavy atom were subjected to refinement within MLPHARE. For the best solution after 10 cycles of the refinement only 1/3 of the data has had figures of merit in the range 0.2-0.35 while the rest of the data has had very low figures of merit what implies a very weakly defined phases. The occupancy of the solution was reduced dramatically and was now less than 5%. Cullis R-factor was very close to 1 what means that lack of closure error is dominating the signal. Therefore the solution was not suitable to calculate the phases.

After the complete analysis one can conclude that occupancy of the heavy atom site which was located is very low. Phases calculated using the anomalous signal from the low occupancy site were not good enough to contribute to the structure solution.

It is interesting to notice that maps calculated using obtained weak phases were looking reasonable. Fourier synthesis with the best phases revealed a correct recognizable pattern of helices oriented in the same direction along crystallographic dyad Z-axis (membrane normal axis). This arrangement was already anticipated during the attempts to find the missing part of the transducer using as a starting point molecular replacement phases. Unfortunately the quality of the maps was not good enough to start the map interpretation in this case. A more significant anomalous signal is needed.

4.12 Ground state structure of the Bacteriorhodopsin

4.12.1 Data reduction

Data reduction for Bacteriorhodopsin data was done essentially in the same way as for pSRII-HtrII complex data. Bacteriorhodopsin crystals were frozen with a nitrogen cryostream at 100 K and for the ground state data illuminated with the red light laser with $\lambda = 685$ nm to eliminate the accumulation of the K state. Typical dataset is collected in two passes for low and high resolution data. Wavelength is 0.934 Å, oscillation angle is 0.4^o for high resolution set, oscillation range is 90^o and exposition time per image is 5 sec, for low resolution dataset the exposition time is 2 sec at attenuated beam intensity.

Crystal of Bacteriorhodopsin achieving size of 400-500 μm has shown an excellent diffraction. The resolution for the data was better than 1.5 Å for most of the crystal. The best resolution obtained was 1.35 Å. The dataset collected with this resolution was used for the determination of the ground state structure. The dataset statistic is shown in Table 6.

Table 6

Resolution	R_{merge} (1.35-1.30 Å)	I/σ (1.35-1.30 Å)	Completeness	Number of reflections	B factor
1.35 Å	0.331	1.8	96.1%	48537	21.3

Crystals of the Bacteriorhodopsin belong to $P6_3$ space group with typical cell parameters $a = b = 60.75$ Å $c = 110.44$ Å. After evaluation of the dataset in MOSFLM the mosaicity of the crystal was estimated as 0.45^o. Both low and high resolution datasets was integrated with profile fitted intensities and consequently merged and scaled in SCALA as it was already described.

4.12.2 Twinning of Bacteriorhodopsin crystals

Bacteriorhodopsin crystals are subjected to the merohedral twinning. In Figure 37 twinned Bacteriorhodopsin crystal is schematically shown. It is supposed to consist preferentially of two domains. As it is shown in Figure 37 in the case of twinning two possibilities of the domain surfaces contacts can be observed. From the previously reported crystal structures of the ground state in $P6_3$ symmetry, only one possible interaction is expected to be strong so defining the nature of the twinning (for the details see Efremov et al. [84])

Twinning fraction varies significantly from crystal to crystal from almost zero twinning fraction to 0.5 in the case of the perfect twinning. According to this, crystals were subjected to the selection before the data collection at the synchrotron. Twinning was checked according to already described diagnostics including intensity statistic tests, Britton plot and Yeates distribution. For the crystal used to collect the ground state data twinning fraction was estimated as 0.02. Respectively, detwinning corrections were made for this dataset.

4.12.3 Structure solution

Molecular replacement method was used to obtain the initial phases. Starting model was that of Luecke et al., solved at 1.55 Å resolution. To eliminate the model bias all side chains of the starting model were reduced to alanine side chain. MOLREP was used to position starting model in the unit cell of the crystal. Unique distinct solution was found after rotational search and subsequent translation search with correlation coefficient CC = 0.680.

Electron density maps with phases calculated from the polyalanine model was at 1.35 Å resolution of excellent quality which made it possible to identify almost all the side chains. To

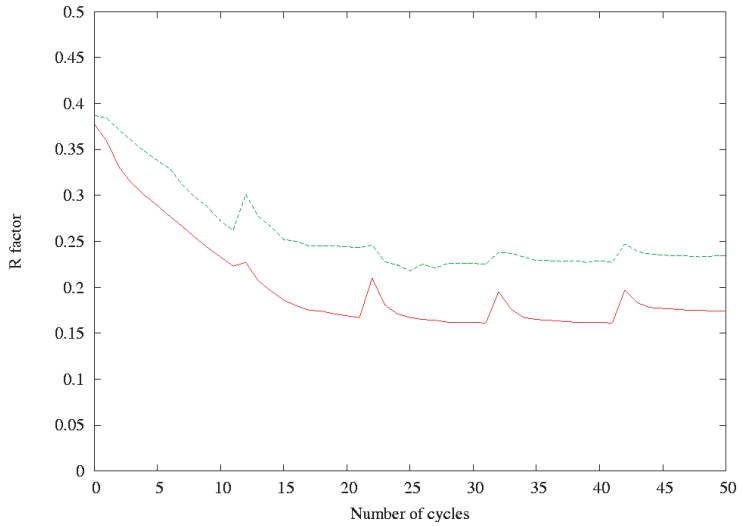

Figure 49: R factors monitored during the refinement run following ARP protocol. Red - conventional R factor, green - free R factor

accomplish the building of the side chains the ARP procedure was used. The crystallographic R factor drops significantly during the first 10 cycles of refinement as it is shown in Figure 49. During the next 40 cycles of the refinement there is further not significant improvement of the R factor. After the first autobuilding run model is to 95% completed. The ARP procedure finishes with a complete model of the Bacteriorhodopsin excluding loop EF which is not well defined. This is probably a consequence of disorder or the presence of more than one conformation of this loop. Retinal was modelled on the basis of dummy atoms positioned by ARP.

Further work on the model includes some minor corrections made in O after examining individual residues of the protein. At this stage also water molecules were picked up from the electron density map by judicious consideration of the possible hydrogen bonding. This step is followed by extensive refinement cycles in CNS by simulated annealing and energy minimization methods. After refinement of the positional coordinates B-factors were refined.

The final model was checked for the correct geometry in PROCHECK. The refinement statistic and the deviations from ideal geometry are listed in Table 7.

Table 7

R_{work}	R_{free}	Average B factor, $\AA^2$	Bond length dev., Å	Angle dev., ^{o}
0.226	0.231	17.8	0.010	1.3

4.13 Determination of the intermediate state structure

4.13.1 Accumulation of the intermediate state in the crystal

Intermediate states were trapped in the crystals of pSRII-HtrII complex and Bacteriorhodopsin.

pSRII-HtrII

The rise of the M state happens in tenth of microseconds and its decay in hundred of milliseconds synchronously with rise of the O state. This long life-time of photocycle suggests that it is possible to accumulate a high percentage of the M state even at room temperature under intensive illumination with blue light. Indeed illumination of the crystal with a 488 nm laser (with power $\approx 1mW/mm^2$) at room temperature result in about 50% of M state and about 15% of O state. The amount of accumulated intermediate is constant in the temperature range between room temperature and -15^oC (personal communications, Efremov).

The trapping was performed at cryo temperature. The procedure of trapping was as following: flash frozen in a nitrogen cryostream crystal was heated to room temperature, illuminated with blue laser light during several seconds and flash frozen to 100K under constant illumination. This procedure results in about 60% of the trapped M state and 15% of the O state.

For the K intermediate trapping, the crystal was illuminated with a green laser at 473 nm (with power $\approx 1mW/mm^2$) during 30-40 seconds at 100K. The amount of the K intermediate which was accumulated by this procedure was about 50%.

Bacteriorhodopsin

For accumulation of the late M state in the crystal it was heated to room temperature by blocking the nitrogen cryo stream. Crystal was illuminated with a green light (33 mW, 514 nm) for 1 s, then cooled by unblocking the cryo stream while the green light was still on. One second after recooling started, the illumination was turned off.

4.13.2 Data acquisition and reduction

Several datasets were collected on the beamlines ID14-1(2), ID13 (ESRF, France) from the crystals which were according to the procedure described above transferred to the intermediate K or M states.

pSRII-HtrII

For data collection were selected similar to those used for the protein complex ground state crystals, Each dataset was collected in two passes: low and high resolution pass - in order to collect the data of maximum possible resolution (for high resolution data collection the time of acquisition of a single diffraction image has been long enough to produce a number of overloaded reflections at low resolution). In the beginning only two images separated in 90 degrees of the rotation around the spindle axis were collected. This images were used for autoindexing (space group and orientation determination). The space group was the same as for the ground state $P2_12_12$. For the orthorhombic space group $P2_12_12$ only 90 degree of the rotation is needed to record the complete dataset and so with the known orientation of the crystal the rotation range of 90 degree was selected. To obtain the data to 2 Å resolution the oscillation angle of 1 degree was chosen, at the mosaicity 0.4-0.5 degree this oscillation angle allows to collect the data without overlap of the reflections.

When it was possible the illuminated and non illuminated datasets were collected from the same crystal. This can be realized with big enough crystals used for data collection. If the size of the crystal was more than two times bigger than the biggest size of the X-ray beam, then two datasets could be collected from different parts of the crystal. Acquisition of the illuminated and non illuminated data from the same crystal is preferred because of a more accurate scaling of the data what is especially useful for the determination of the structural differences between the intermediate and ground states.

The time of data acquisition was adjusted by the observation of the diffraction resolution decay during the data collection from the typical crystal. Images of the same crystal orientation were collected before the dataset collection and after. This images were carefully compared. The signal to background ratio was estimated for the same reflection. If this ratio became lower after the dataset collection then this could be attributed only to the radiation damage of the crystal. The time of data acquisition was reduced until the images before and after data collection were not possible to distinguish. The typical time of data acquisition was $\approx$ 10 seconds at a flux counts of 10^{10} per second. This restriction of measurement time did not result in a decrease of resolution for the collected data.

Data integration was performed in MOSFLM. First data was indexed in space group $P2_12_12$ and initial lattice parameters were defined to be $a = 124.53$ Å, $b = 46.80$ Å, $c = 53.76$ Å. The mosaicity was estimated by the matching of the predicted diffraction pattern with the observed pattern. After that, the cell parameters and initial offsets values describing the geometry of the experiment were defined more precisely during the postrefinement procedure which operates with data selected from two images remote in rotation angle ϕ. These parameters were imputed to the integration run of the complete dataset. The following offset parameters were monitored during the data integration: camera constants CCX, CCY (reflect the accuracy of the supplied direct beam coordinates), CCOMEGA accounts for errors in orientation around the direct beam, crystal missetting angles PHIX, PHIY, PHIZ for deviations from the crystal orientation matrix, Deviations from normal incidence on the detector TILT, TWIST, TILT is a rotation about a horizontal axis, TWIST about a vertical axis. There should not be any substantial changes in this parameters during the data integration image by image. The crystal-to-detector distance should change continuously. For each new image these parameters were updated during the refinement which uses the adjustment of the position of the central reflections of the image to the predicted with the fixed unit cell parameters. Also the mosaicity of the crystal was refined from image to image. This characteristic should not show significant variations during the data integration.

The area of the detector were divided to 25 squares, for each square a reflection profile on the basis of sequential 5 images was estimated. Intensity of the individual reflections were fitted to the standard profile and when there were significant deviations reflections were discarded. All this procedures were performed in an automated way, but the results was extensively observed and when it was necessary the corrections to the initially supplied parameters for data integration which includes the raster shape, range of the resolution for refinement, use of postrefinement, etc. were made. Integration was performed for several initial mosaicity values and the optimal mosaicity was chosen, which do not changes during the refinement from image to image. The integrated intensities and estimates of their standard deviations were stored in the MTZ format file (CCP4 [81]).

Integrated data were further scaled in SCALA from the CCP4 program suite ([81]). The data from both passes (low and high resolution pass) were scaled together. The average intensity level for two passes was very different and initial prescaling by a constant scale factor was done. 7-10 cycles of the refinement of the scale factors were performed until the scaling has converged. The scales were allowed to vary continuously across the detector (resolution bins were defined) and

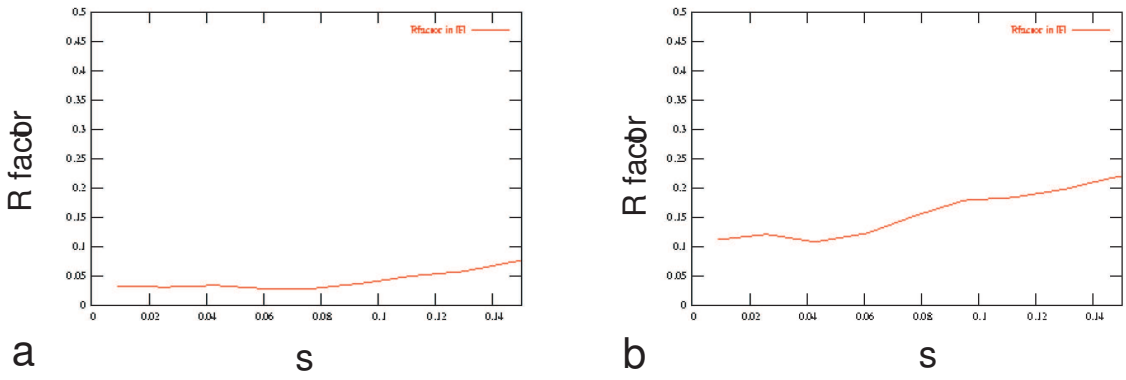

Figure 50: R factor calculated between A. the not iluminated data from different crystals and B. illuminated and not illuminated data from the same crystal

from image to image. The isotropic B-factor model was introduced to reflect the decrease of the diffraction with diffraction angle. The scaling cycles of scale factor optimization were performed as described (Chapter 3). Final analysis run on the scaled data and rejection of outliers were performed.

Analysis of the ratio of the average intensity to its standard deviation together with R_{sym} value allowed to establish the maximum resolution limit where the data were truncated. For the resolution limit determination the values of R_{sym} of 0.3-0.4 were considered and ratio of the intensity to its σ 2.0 in the last resolution shell was chosen. In Table 8 the typical parameters of the dataset are shown. Among them are completeness of the data, ratio of the intensity to its standard deviation, R_{sym} and R_{merge}.

Table 8

Dataset name	Resolution	R_{merge} (last shell)	I/σ (last shell)	Completeness	Number of reflections	B factor
SHL19M	2.65	0.449	1.7	86.2	8055	42.0
SHL29M	2.2	0.413	1.8	97.9	16637	25.5
SHL8M	2.25	0.416	1.8	99.8	15667	31.0
SHL6M	2.11	0.223	3.0	77.7	15787	14.5
SHL11M	2.25	0.416	1.8	99.8	15667	31.0
SHL15M	2.33	0.306	2.5	93.4	13208	34.0
SHL17M	2.36	0.167	4.6	96.0	13086	29.8
SHL12M	2.15	0.353	2.2	99.7	17899	25.7
SHL10M	2.0	0.395	2.0	99.7	16642	32.3

last resolution shell is of 0.05 Å width

In the end the standard deviations of the reflections were corrected.

As it could be expected integrated data for the ground state show a good agreement with each other as it is indicated by the R factor calculated between the corresponding reflection intensities of two datasets. Between the illuminated and the ground state data agreement is not as good, indicating the presence of the structural differences of the protein within the crystal which rise upon illumination. R factor between the datasets are shown in Figure 50.

Bacteriorhodopsin

M state was trapped in the crystals according to the procedure described. Data was collected under similar to the ground state data collection conditions. After evaluation of the dataset with the best obtained resolution in MOSFLM and SCALA following statistic emerged.

Table 9

Resolution	R_{merge} (1.50-1.45 Å)	I/σ (1.50-1.45 Å)	Completeness	Number of reflections	B factor
1.50 Å	0.489	1.8	93.7%	37707	25.6

As for the ground state merohedral twinning diagnostic was performed by conventional Yeates statistic, Britton plot and second moment of intensity distribution. Twinning ratio was estimated to be 0.13. Program DETWIN (CCP4) was used to detwin the data with provided estimated twinning fraction.

4.13.3 Structure determination of the intermediate state

The integrated data was subject to treatment aimed to elucidate the final structure of the intermediate state. The used procedure consists of three main steps: observation of the difference Fourier maps, estimation of the content of the intermediate in the crystal and refinement of the model representing the intermediate state.

4.13.4 Difference Fourier map

The first step, being simple in realization and giving the direct evidence about the structure of the intermediate state is the inspection of difference Fourier density maps. At this step there is no need to know what is the content of the intermediate in the crystal. The difference maps present the difference between the two states, scattering by the ground state which was not transferred to the intermediate state cancels out.

The difference Fourier density maps between two datasets (illuminated data and non illuminated data) is calculated as follows: the difference $\Delta F = |F^{il}| - |F^{g}|$ is build and combined with the phases determined from the model of the ground state. Such a map represents a difference between the illuminated and non-illuminated (ground) states. The desired difference in electron density between the two structures is given by the Fourier density calculated with the difference $\mathbf{F}^{il} - \mathbf{F}^{g}$. Such difference is not available experimentally. Therefore the difference Fourier map which one uses is only an approximation to the difference electron density between two structures. Actually the difference Fourier coefficient $m\Delta Fexp(i\alpha)$ represents only the vector component of the difference between the structure factors in the direction of $\mathbf{F}^{g}$ if the difference is much smaller than F^{g}. The phase for both structure factors is assumed to be the same and the phase of the ground state structure factors is employed.

The difference map is very sensitive tool which allows to draw the most important conclusions about the structure of the intermediate state. Moreover it implies no additional manipulations of the data and is based on directly observed in the experiment structure factors amplitudes.

4.13.5 Refinement scheme

The structural changes were followed by the refinement. The crystal content was modelled by two equivalent conformations of the pSRII-HtrII complex one representing the ground state and another the intermediate state (same for Bacteriorhodopsin). The starting model was refined against the ground state data. For the pSRII-HtrII complex the model included the receptor molecule up to the residue 225 and the transducer molecule consisting of the fragment of residues 21-83. For the Bacteriorhodopsin the model included protein residues 7-235 excluding the region of the EF loop residues 157-162. Occupancies for two models representing the ground state and the intermediate state were assigned. These occupancies were fixed during the refinement runs. The value of occupancies were varied in different runs.

The model representing the Ground state in the crystal was fixed during the refinement. The model representing the intermediate state was refined against the observed data. The calculated structure factors in this case are given with the expression

$$F_{calc} = \alpha F^{i}_{calc} + (1 - \alpha) F^{g}_{calc} \tag{99}$$

where F^{i}_{calc} are calculated structure factors of the intermediate state and F^{g}_{calc} are calculated structure factors of the ground state respectively. During the refinement the difference between F_{calc} and F_{obs} was minimized.

Several different strategies for the refinement were chosen. Namely, the conjugate gradient energy minimization method and the simulated annealing method were used. The methods have

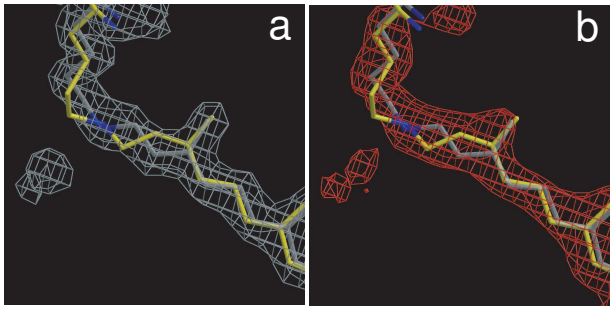

Figure 51: $2F_o - F_c$ electron density and retinal models for pSRII-HtrII complex. (a) Ground state density and retinal models of ground and M state. (b) Composite omit map for illuminated crystal data and retinal models. Grey model - ground state, yellow model - M state.

shown a different efficiency in refinement of the intermediate state structure. The application of the conjugate gradient energy minimization method showed that this method is not able to find a correct solution. The radius of convergence of this refinement method is very small and the found solution corresponds only to the local minimum of the energy landscape. Some parts of the model were nevertheless refined correctly as the agreement with the Fourier difference density maps and the final model has shown.

The method of choice was simulated annealing. Among the different possible protocols of the simulated annealing which differ in temperature control during the annealing so called 'slow cooling' protocol was selected as it is presented in CNS (Crystallography & NMR System, Brunger et al. 1998 [89]). The simulated annealing was performed in several steps. First about 100 cycles of standard conjugate gradient minimization was performed to relieve the initial strain or close nonbonded contacts in the structure that could cause the problems during the molecular dynamics calculations. For the refinement the weight between the pseudo-energy terms standing for the match between calculated structure factors and phased observed amplitudes and stereochemistry of the model was selected. The protein structure first was heated up to 3000 K (or alternatively to 5000 K) by assigning the speeds corresponding to the Maxwellian distribution at this temperature. Molecular dynamics was carried out at this temperature during 0.001 ps. Velocities were rescaled to reduce the temperature by 25 K (or 50K) and again the molecular dynamics was performed at current temperature. In this manner the system was cooled down to 300 K and again the conjugate gradient minimization was performed.

4.13.6 Determination of the intermediate state content in the crystal

Initial refinement by simulated annealing was performed by fixing an arbitrary occupancy for the intermediate state. Models were produced with different occupancies values chosen on a grid with step 0.1. These models were hardly different from each other when the difference in occupancy was 0.1 and the differences in models increased when the difference in occupancy values was 0.2 or more. Refined with simulated annealing models were not yet correct though the conformation of the most of the residues was close to the conformation reported in the final model.

To fix the occupancy value use of the composite omit map was made (Bhat et al. 1998 [90]). This map was constructed according to the following procedure. The asymmetric unit of the unit cell is partitioned into a number of 'phased' volumes and the associated 'neutral' and 'phasing'

volumes. A 'neutral' volume encloses the space in the immediate neighborhood of the 'phased' volume to which it is associated. An asymmetric unit from which the 'neutral' and 'phased' volumes have been excluded is called the 'phasing' volume. 'phased' and 'neutral' volumes are selected as boxes. The dimensions of the 'phased' volume are chosen so as to include a small fraction 5 % of the unit cell at a time. Dimensions of the box defining the 'neutral' volume are chosen on the basis of the maximum resolution of the data. The electron density values within each 'phased' volume are calculated by the Fourier transforms of the observed amplitudes, the phase angles ϕ and the standard weights. The phases are obtained by the inverse Fourier transforms of the electron density values found within the 'phasing' volume that corresponds to the particular 'phased' volume, together with a constant-threshold electron density value everywhere else. This calculation is performed for each 'phased' volume and the density obtained is written down to the final composite omit map. The $2F_o - F_c$ map was calculated.

Composite omit map is the most precise presentation of the true electron density in the crystal because it has suppressed bias originating from the model used for phase calculation. Instead it promotes the information stored in the structure factors amplitudes.

Calculated omit maps represented the features of both ground state and intermediate state in the correct proportion (Figure 51). This was the ground for the approach to estimate the amount of the intermediate in the crystal from data. To get a quantitative estimate of the occupancy of the intermediate state the composite omit maps was compared to the model obtained from the refinement. This comparison was performed by correlating the observed density to the extrapolated model density (atomic densities represented with gaussians). The set of model densities defined by the relative occupancy of the ground and intermediate states is calculated. The model densities of the ground and intermediate state are summed in real space. The correlation of this densities to the composite omit map density is calculated for a certain region of the electron density map, e.g. defined by a certain residue. The region around the residue is selected by the fixed distance cutoff (2.5 Å) around the atoms of the selected residue. Within this region the correlation coefficient was calculated on the selected grid. The correlation coefficient allowed to identify the best fit of the sum of models to the electron density. The fraction which corresponds to the best fit was selected to be the amount of the intermediate in the crystal on the basis of the statistical analysis of the matches for individual residues.

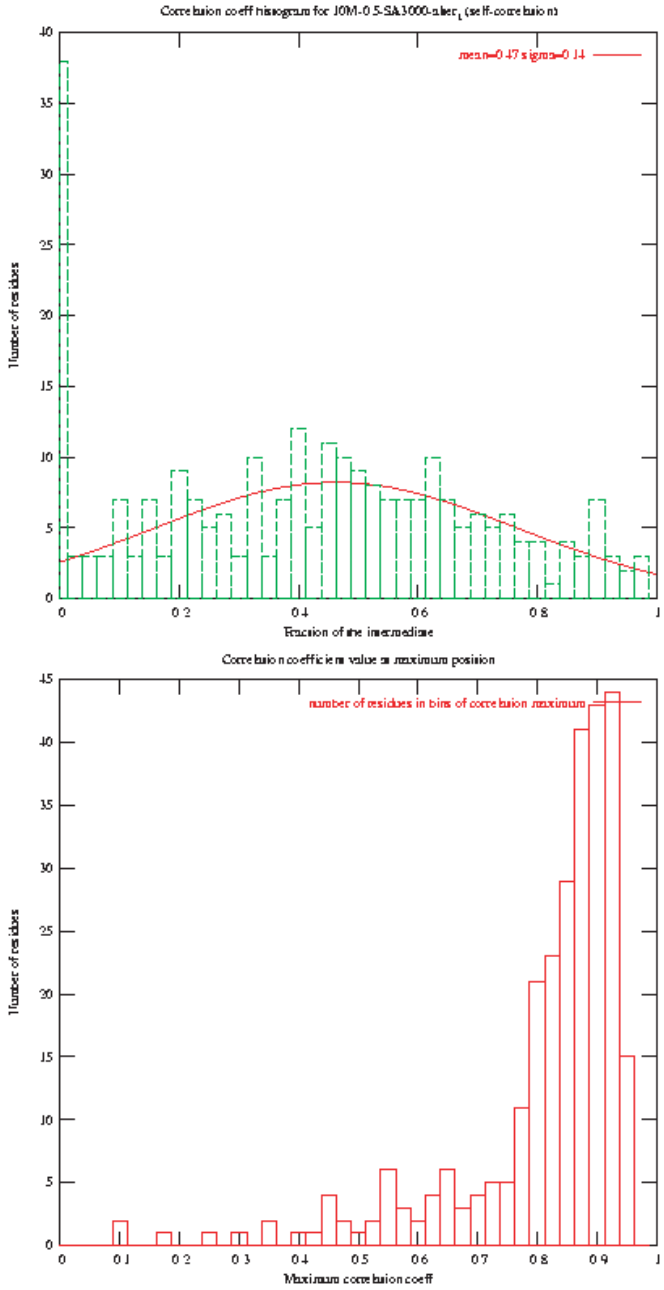

Figure 52: Fraction determination from the correlation coefficient distribution for pSRII-HtrII complex. Number of residues with the maximum correlation corresponding to the occupancy bin is presented. Number of residues with correlation coefficient in the bin is presented.

Some residues which has no significant change in there conformation do not allowed to estimate the fraction of the intermediate. In Figure 52 the correlation coefficient distribution for the individual residues of the pSRII-HtrII complex is shown, with zero value represented those which do not contributed to the fraction determination. Fraction corresponding to the maximum

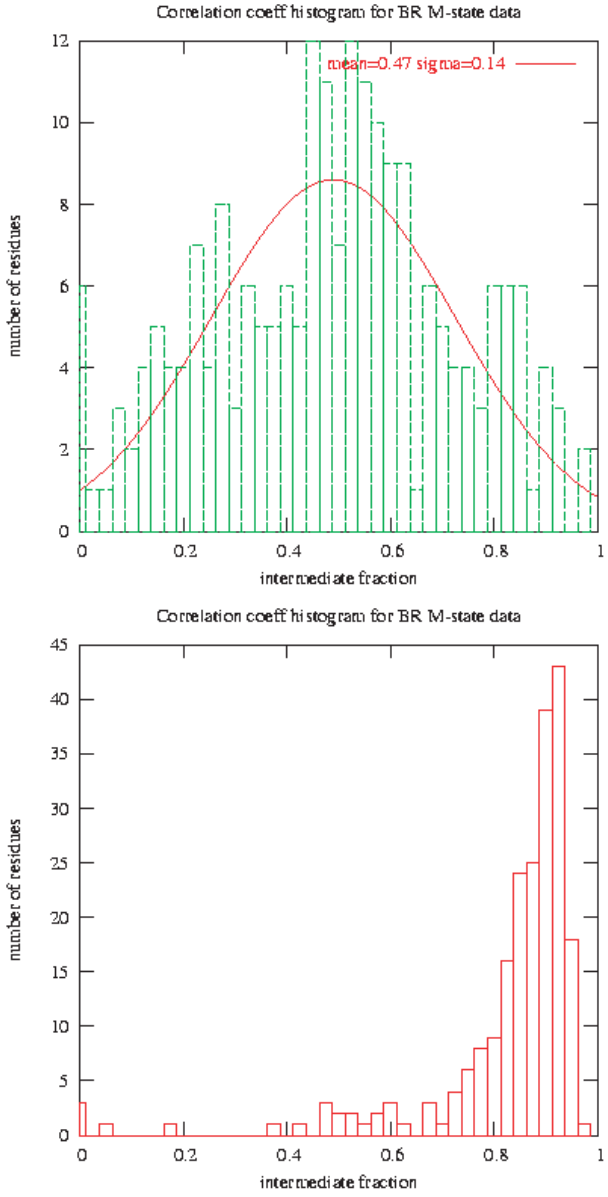

Figure 53: Fraction determination from the correlation coefficient distribution for Bacteriorhodopsin. Number of residues with the maximum correlation corresponding to the occupancy bin is presented. Number of residues with correlation coefficient in the bin is presented.

is 0.50. This fraction presents the amount of the M state in the pSRII-HtrII complex. In Figure 53 analogous graph is presented for Bacteriorhodopsin. Again the fraction corresponding to the M state accumulation in the crystal is 0.5.

4.13.7 Final refinement statistic and structure validation

Model building together with refinement according to the scheme above and analysis of difference electron density maps was done for the intermediate state fraction selected from correlation coefficient distribution histogram.

The refinement data are listed below in Table 10 for pSRII-HtrII complex

Table 10

	Ground state	K state	M state
Resolution, Å	1.9	2.0	2.2
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Cell, Å	124.59 46.96 53.84	124.34 47.03 53.82	124.62 46.90 53.85
Unique reflections	25627	21010	16147
Completeness	99.8%(100%)	95.4% (98.7%)	97.9% (99.4%)
R_{merge}	0.079 (0.393)	0.091 (0.494)	0.085 (0.414)
I/σ	5.6 (2.2)	7.1 (1.8)	6.3 (1.8)
Refinement			
Reflections used	24964	20360	15688
R_{work}	0.218	0.218	0.217
R_{free}	0.246	0.245	0.241
Average B factor, $Å^2$	29.7	27.6	27.9
Bond length dev., Å	0.007	0.009	0.007
Angle dev., ^{o}	1.0	1.2	1.1

last resolution shell has width of 0.05 Å

Refinement data for Bacteriorhodopsin M state are listed in Table 11.

Table 11

R_{work}	R_{free}	Average B factor, $Å^2$	Bond length dev., Å	Angle dev., ^{o}
0.241	0.270	33.1	0.005	0.8

4.13.8 Ensemble of the M state structures of pSRII-HtrII complex

The resolution of the data collected from the crystal defines the success of the structure solution. Especially the need for high resolution arises in the case when two similar models should represent the content of the crystal. Unfortunately refinement procedures could not always elucidate the correct solution for the structure being searched. The energy function which is minimized has a multiple local minima. Methods like simulating annealing allow some of this local minima to be overcome but still they do not always find the single unambiguous solution. Taking into account the known examples of protein structure refinement one can expect at a resolution of ≈ 2 Å the error in the coordinates to be 0.05-0.15 Å which is already comparable to most of the structural differences between the Ground and M state conformations. To elucidate the correct progress of the refinement several data sets from the illuminated crystals trapped in the intermediate state and from the non illuminated crystals were treated in the same way. The obtained models were compared to each other. In both cases one conformer describing the crystal was fixed in the ground state conformation while the other was refined.

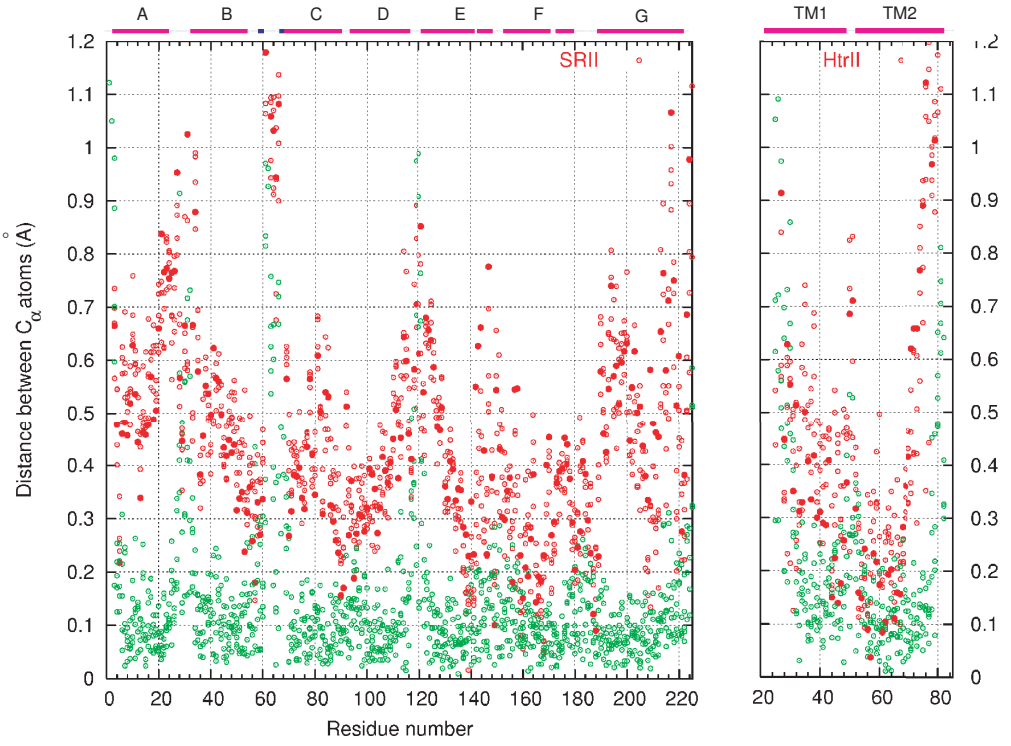

Figure 54: Statistic of distances between C_α atom positions of 2 models representing protein conformations in the crystal. Red - for illuminated crystals trapped in the M state with 50% occupancy of each conformation: fixed ground state conformation and determined by refinement M state conformation (open - for 4 crystals trapped in M, solid for crystals used for derivation of the final M state model). Green - for non illuminated crystals kept in ground state with the same occupancies (open - for 5 crystals kept in ground state)

For the ground state data second refined conformation was almost not different to the first (Figure 54). Difference in C_α atom positions was in average 0.1 Å. And only for the flexible loops this difference was at the significant level. This observation indicates that already in the ground state there are at least two possible conformations for the flexible loops.

Refined M state models for different data has shown an average divergency in the coordinates of 0.15 Å. Conformational changes were clearly distinguishable between ground and M state. Obtained models for the M intermediate nicely reproduce each other.

4.13.9 Double Distance Matrix Plot

Double Distance Matrix Plot (DDMP) is a useful tool to trace structural differences between molecules. Particularly it is very helpful for visualization of the differences between 2 conformations of the same molecule. C_α atoms are chosen in each residue and matrix of all the distances between pairs of C_α atoms is calculated for each conformation. Difference of the 2 matrices,

one for each conformation of the molecule, gives a quantitative information about the structural differences between conformations. This information does not depend on the mutual orientation and position of the two conformations in the space and reflects only relative displacements of the parts of the molecules.

pSRII-HtrII

In Figure 55 the DDMP is presented relating G and M state conformations of the pSRII-HtrII complex. Arrangement of the transmembrane helices interconnected by a flexible loops manifests itself in the plot as a checked pattern defined by the significant differences in distances involving C_α-atoms of the loops. One square of the pattern corresponds to the selection of distances between C_α-atoms where first atom is chosen from one helix and second from the other and so reflects the relative change in position of these helices for two conformations. Main diagonal squares correspond to the distances within the same helix. Changes in these distances between ground state and M state would mean a deformation of the helix. These changes are pronounced for helix A, helix C, helices E and F with connecting loop EF of the receptor and helix TM1 of the transducer.

The most significant changes appear in the regions [140-190]*[190-220] (F and G pair), [150-170]*[0-30] (A and F pair) [70-95]*[190-210] (C and G pair) within the receptor and [327-350]*[140-170] (TM1 and F pair), [0-30]*[140-170] (TM1 and A pair), [350-379]*[200-220] (TM2 and G) , [370-379]*[0-60] (end of TM2 and A,B) within the complex.

Four helices F, G, TM1 and TM2 that define the contact interface between the proteins show most of the structural change with the main difference in the distances between F and G helices. This relative displacement could cause further displacements of the transducer helices which as it is seen from the plot happens so that helix TM1 is moving concertedly with helix G and TM2 moves concertedly with helix F ([327-350]*[170-220] and [350-379]*[150-200])

The other parts of the plot shows that also A, B and G helices move concertedly as well as C, D pair. Among the loops the most pronounced displacement shows loop BC.

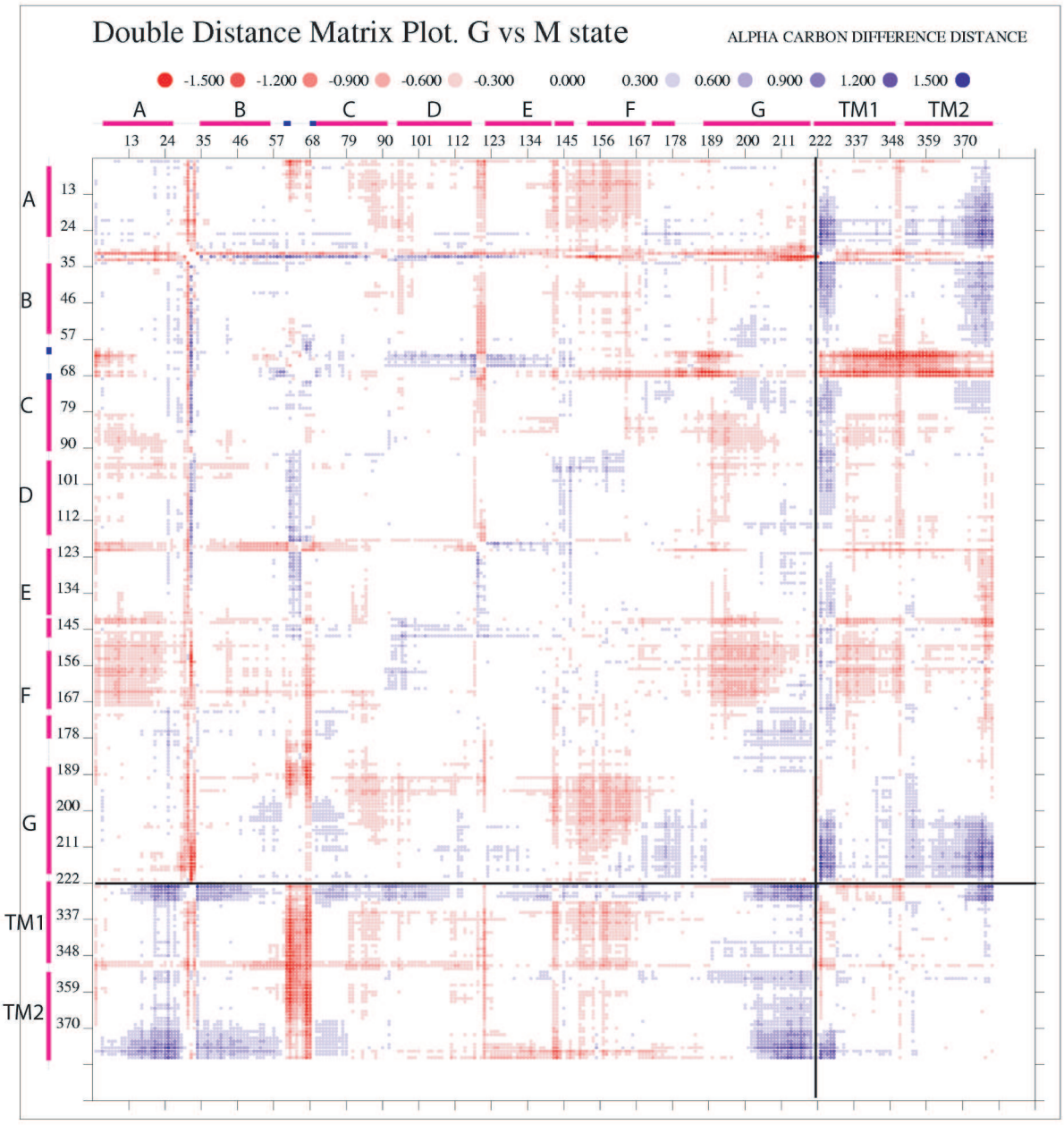

Figure 55: Double distance matrix plot for pSRII-HtrII(1-159) complex conformations of ground and M state. Contoured at the 1.5 Å level in color grade steps of 0.3 Å. Calculated for residues 3-222 of pSRII and 27-79 of HtrII denoted as 327-379 respectively.

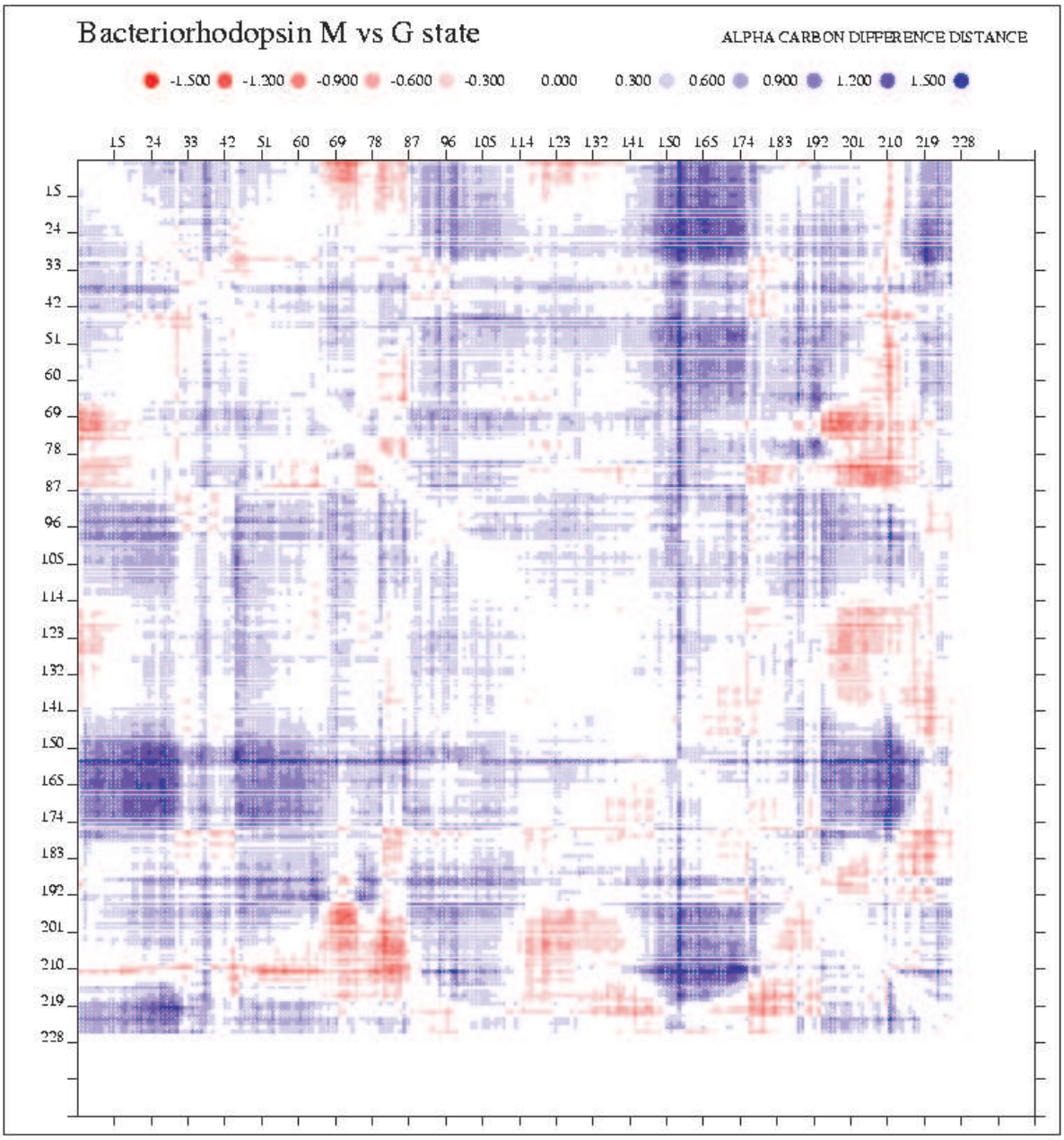

Figure 56: Double distance matrix plot for Bacteriorhodopsin conformations of ground and M state. Contoured at the 1.5 Å level in color grade steps of 0.3 Å.

Bacteriorhodopsin

In Figure 56 DDMP relating M and Ground state of Bacteriorhodopsin is presented. Two main diagonal squares corresponding to helices C [65-95]*[65-95] and G [190-228]*[190-228] show higher differences indicating deformation of these helices in the M state.

The most significant changes appear in the regions [140-170]*[190-228] (F and G pair), [140-170]*[7-33] (F and A pair), [140-170]*[40-70] (F and B pair) and [85-115]*[7-33] (C and A pair). Taking into account that the displacement of the helix A relative to B is moderate changes imply

that helix F moves relative to A, B pair and G. The amplitude of this movement reaches 1.5 Å. Another movement which could be deduced is the relative displacement of helix C which has lower amplitude.

5 Disscusion

5.1 Structure of pSRII-HtrII complex intermediates

Structural changes connecting ground and M state start with a photoisomerization of the retinal, a process which happens in the picosecond time range. The immediate protein response to this event happens in the same time domain and is evident in the K state conformation of the protein.

5.1.1 Structural changes in the K state

To accommodate the new conformation of the retinal within a nanosecond time domain structural changes in the protein start to propagate in the direction of the extracellular site.

The change in the geometry of the retinal causes a significant rearrangement of the Lys205 side chain. While the peptide atoms of the Lys205 show almost no change in the K state compared to the ground state, atoms Cϵ and Cδ of the Lysine side chain are shifted by approximately $1-1.2$ Å from their initial positions so that the distance to the ground state position of water molecule W1 is shortened from 3.3 Å to 2.5 Å.

Water molecule W1 which bridges the Schiff Base nitrogen with the oxygens of the carboxylic groups of Asp75 and Asp201 is next to be involved in the cascade of changes launched by the retinal photoisomerization. The pentagon of hydrogen bonds including two other water molecules W2 and W3 and oxygen atoms of aspartic acids is formed in the ground state (Figure 57a). This hydrogen bond pattern can not exist any longer because the water W1 is sterically driven away from its original location. A new conformation of the Lysine side chain and a new position of the Schiff base nitrogen excludes the hydrogen bonds of the water molecule W1 and destabilizes its position, so that it is not any more crystallographically defined in the K state. According to the direction of the displacement of the Lysine side chain this water is expected to be still located in the vicinity of Asp75.

Temperature factors of the ground and K state for the Schiff base nitrogen ($B_G = 14$ Å^2 to $B_K = 32$ Å^2) indicate the increased amplitude of motion in the K state and imply an increased number of allowed conformations of the Schiff Base nitrogen.

The charge separation between the protonated Schiff base nitrogen and Asp75-O_δ increases from 4.2 Å to 4.9 Å in the K state. Asp75 in the K state adopts a new conformation. In the K-state the carboxylic group of Asp75 is reorients to approximately 90^o from its ground state conformation and therefore the hydrogen bond to Thr79 is lost. Also the hydrogen bond between Asp75 and water molecule W3 is lost and the carboxylic group has now an increased number of allowed conformations. This is indicated by the change in the temperature factor from 18 Å^2 in the ground state to 31 Å^2 in the K state. Water molecule W3 is no longer present in the electron density maps for K state. Destabilization of this water molecule positioned in 6 Å distance from the Schiff Base goes in line with the new orientation of the Asp75. These changes in the environment of Asp75 imply an increase of the apparent pK_a as compared to the ground state.

Two water molecules within a 9 Å distance from the Schiff Base acquire new locations in the K state. Water molecule W2 within a 6 Å sphere in the extracellular vicinity of the retinal occupies a new position which is 1.5 Å closer to the Schiff Base which is possible due to the new location of disordered water W1. Water molecule W2 which was initially hydrogen bonded to water W3 and Asp201 in the K state has a new hydrogen bonding environment which includes Trp76 and Asp201. The new location of this water molecule is still in the vicinity of Asp201. The symmetry which was initially between Asp75 and Asp201 within the pentagon of hydrogen bonds is significantly broken after these rearrangements of water molecules. From the Asp75 side

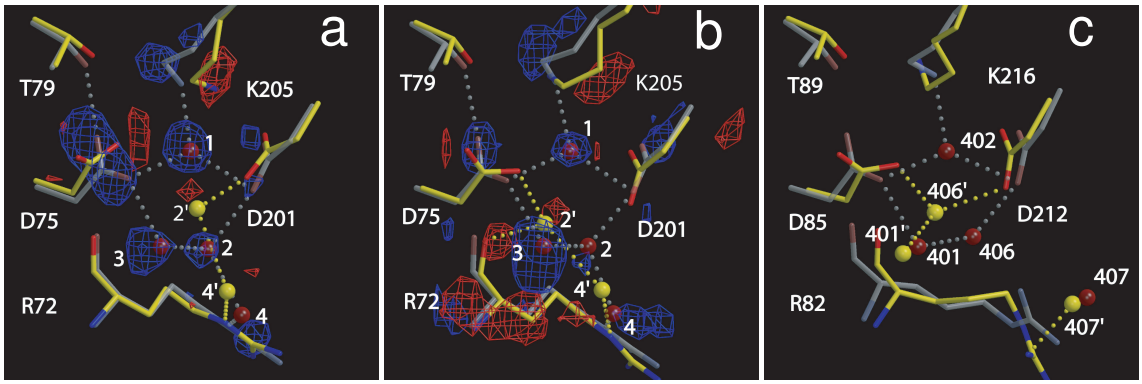

Figure 57: Extracellular vicinity of the retinal Schiff Base. Hydrogen bond pattern including 4 water molecules, Lys205, aspartic acids Asp75 and Asp201, Arg72 and Thr79 side chains: ground state (grey), intermediate states (yellow), negative (blue) and positive (red) electron densities, water molecules of ground state red and intermediate state yellow balls, hydrogen bonds in dotted lines in grey for ground state and in yellow for M state. (a) K state of pSRII-HtrII(1-114) complex with electron density maps contoured at 4.5σ and -4.5σ, respectively. (b) M state of pSRII-HtrII(1-159) complex with electron density maps contoured at 3σ and -3σ, respectively. (c) ground and M state conformations of Bacteriorhodopsin

disposition of the elements has significantly changed. In the same time Asp201 shows almost no rearrangement. The hydrogen bonds of Asp201 still exist, although water molecules W2 and W4 shift towards the retinal by ≈ 1.5 Å in K state. The distance of Asp201-O_δ is reduced from 4.9 to 2.7 Å due to motion of the Schiff base. The hydrogen bond structure (Asp201-O_δ - W2 - W4 -Arg72-N_ϵ) is retained. Therefore a significant shift of the apparent pK_a for this residue is not expected.

Together with water W2 another water molecule W4 hydrogen (bonded to water W2 and W5 in the ground state) moved towards the Schiff base. Water molecule W4 keeps a hydrogen bond with water W2, but now is not connected to water W5. Though this water is still located within a hydrogen bond distance, the angle including this two water molecules and carbonyl oxygen of Val194, which is a hydrogen bond partner for water W5 is not favorable for permanent hydrogen bonding. Water W4 has in the K state also one connection less with Arg72 guanidinium group.

Slight displacements of Arg72 and Asp193 are in the utmost part of the region to which structural changes were spread from the isomerized retinal in the extracellular direction within a picosecond time domain. The conformation of Arg72 coincides with that of Arg82 in the Bacteriorhodopsin M state.

There are no significant changes in the other half of the protein happening in the nanosecond time domain.

5.1.2 Structural changes in the late M state

Structural changes are developing further in the time range of hundred milliseconds. They lead to the signalling late M state conformation of the protein.

In K to M state transition the proton translocation from the Schiff base to the aspartic acid Asp75 happens. pK_a values of the Schiff base and Asp75 are adjusted to new values. A new conformation of the carboxylic group of Asp75 and a new location of water molecule 1 brings $pK_a$of Schiff base and Asp75 closer for proton transfer to happen. But since their is still a distance of almost 4 Å between the Schiff base nitrogen and Asp75 the rise of the M state is delayed to the hundred millisecond time scale. Probably pK_a adjustment is influenced also by additional displacements of other water molecules hydrating Asp75 and Asp201 observed between K and M state.

Displacement of water molecule 1 and reorientation of the carboxylic group of Asp75 together with the loss of all hydrogen bonds by Asp75 allows for a part of helix C which starts at Pro81 and includes Pro71 to be located closer to the Schiff Base in the M state (Figure 57b). The average observed local displacement of this part of helix C in the late M state conformation is 0.4 Å and is directed towards the water channel. Such displacement is inline with the presence of a proline residue located in the middle of helix C. Pro81 introduces an additional degree of freedom in the conformational space of the helix, being its hinge residue, and defines a local flexibility of this helix.

Several conformations of Asp75 carboxylic group are observed (in the M state it has somewhat different orientation from the K state). Observation of van der Waals contacts in K and M state reveal a presence of the free space close to Asp75. This space can accommodate conformations defined by Pro81 for the extracellular part of helix C and allows for the increased mobility of Asp75 in the M state. In a signalling state this space could be occupied because of the absence of the structured hydrogen bond network between the Schiff base - water molecule 1 - Asp75 - water molecule 3 and Thr79. Difference densities clearly display changes in the position of the backbone in the M state which were absent in the K state. Attractive interaction of charges located at the Schiff base and Asp75 bring the residues 72-76 closer to the Schiff base. Because of thermal fluctuations a distance between the Schiff base and Asp75 could be shorten to the extent at which presumably the direct proton transfer could occur within a time range of hundred milliseconds. The increased temperature factors for the side chain atoms of Asp75 in M state underlie the feasibility of such a mechanism of proton transfer. Apart from the direct proton transfer there are also other plausible ways of proton translocation which could be suggested. Those would involve Thr79, Asp201 and water molecules surrounding Asp75.

The local bend of helix C brings also the backbone of Arg72 0.5 Å closer to the extracellular water channel. Water molecule 2 becomes further displaced in the M state. In the new position this water molecule is strongly bound as it has all for possible hydrogen bonds. In the M state this water molecule is clearly decoupled from Asp201. Also water molecule 4 is additionally displaced during K to M state transition. With the relocations of these two water molecule in the M state less connections between helices C and G are left.

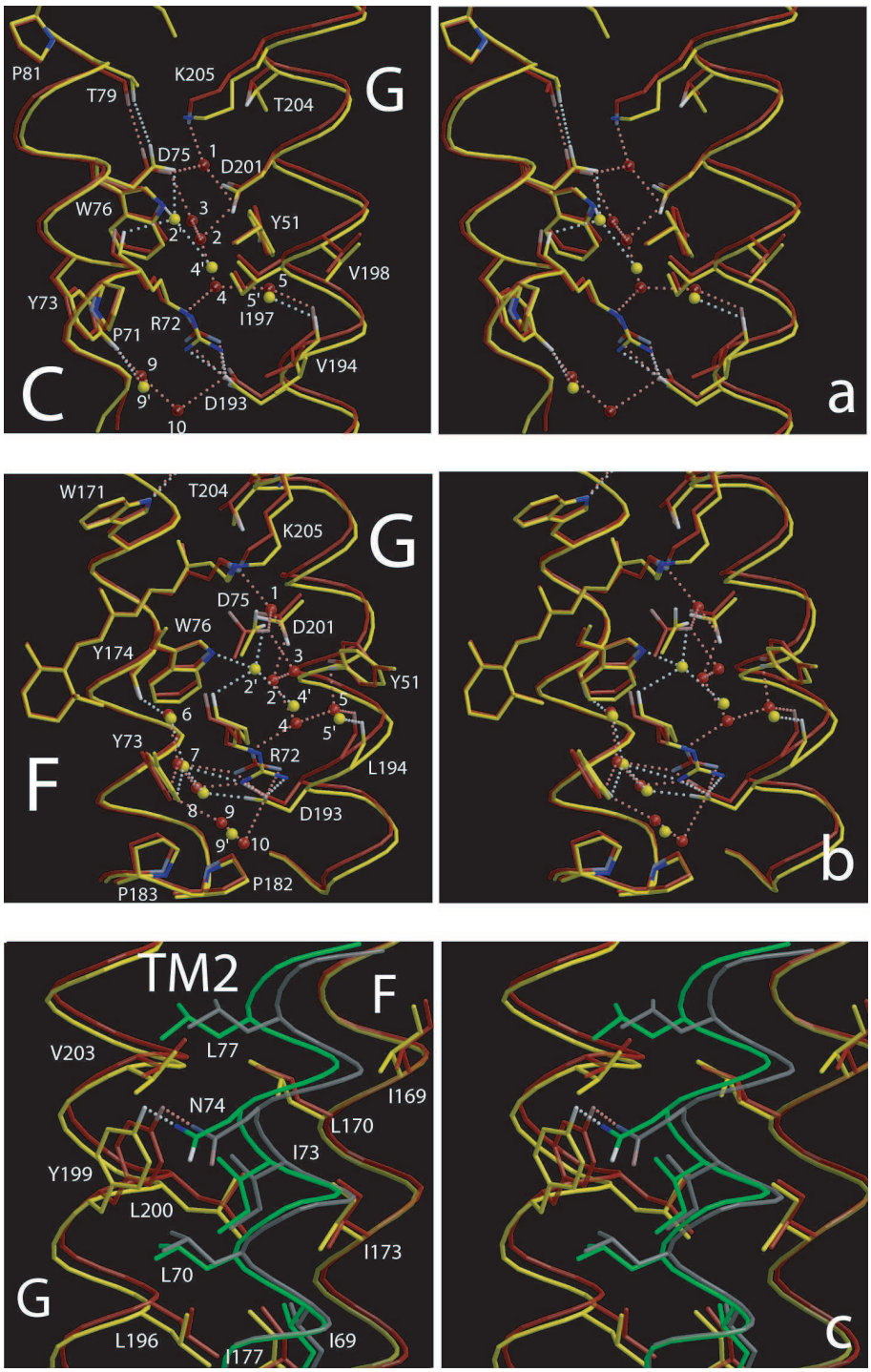

Figure 58: Changes in helical structure and interactions between ground state (red) and M state (yellow) including water molecules as red and yellow balls, respectively; hydrogen bonded network in red for ground and in white for M state; a. helices C and G; b. helices F and G; c. helices F and G, TM2 (transducer helices grey for ground state and green for late M state)

In addition to the changes described so far in the extracellular part of the protein relative displacements of the secondary structure elements of the protein happen upon which the signaling M state conformation is completed (Figure 58). The rearrangement of a tertiary structure are governed by the traits of the contacts between the transmembrane helices which could be considered to present the elementary units involved in a relative motion. The atomic details of the structure provide a key for the understanding of how helical arrangement acts to form a signalling state.

Considering the arrangement of the transmembrane helices of the receptor one could define within it domains consisting of several helices which due to the strength of their mutual interaction would displace cooperatively without a significant change of their individual conformation when the external force is exerted. Interactions between the helices are defined by the interhelical packing angle and the contact surface area.

The geometry of the ground state conformation defines the modes of motion which could occur. Hydrogen bond restraints and the interdigitation of the residue side chains could suggest the directions of possible relative motions of helices. Trajectory of motion should be consistent with van der Waals interactions and the hydrogen bonds.

The motion which could happen within the transmembrane helices arrangement is a shear type motion. This is a sliding type motion which maintains a well-packed interface. This motion occurs without the appreciable main chain conformational change and is parallel to the plane of the interface. There is no change in the side chain rotamers and translation which could reach $\simeq 2$ Å in maximum, happens due to rotations of the side chain with change in torsion angles less than 15^o.

Consideration of each pair of the contacting helices within a receptor shows that one can divide pSRII to two parts underlying the domains defined by a helix coupling: one including helices A,B and G and another including helices C,D,E and F. This domain definition is extracted from the DDMP for pSRII-HtrII. This matrix reveals that the receptor is divided to domains which could be predicted on the basis of interhelical interactions.

Helixes A,B and G are positioned relatively close to each other and have the surface contacts along of their complete length. The number of the hydrogen bonds defines the interaction between the helices. These hydrogen bond connections between two side chains are formed directly or via water molecules. The latter way to connect the helices is very common. Water molecules fulfil the hydrogen bonding ability of the polar groups which due to the geometry have a distant location and can not have a direct hydrogen bond.

Within A, B and G triad there are 4 hydrogen bonds directly connecting the side chains of the one helix with the side chains of the other and 3 connections involving the water molecules between the side chains of the different helices. Also a number of the well packed interdigitating bulky residues contributes to the rigidity of the domain. The observed relative motion of helix A with respect to B and G is indeed quite small so that this helices are part of a single domain.

In contrast to the interactions within the ABG helical bundle, helices B and C are not connected by hydrogen bonds and the interaction purely originates from the van der Waals contacts of the hydrophobic side chains. Predominantly Leu and Ile residues are located on the contact interface. The presence of especially these residues could account for the pronounced relative shear motion. For the Leu and Ile side chains moderate changes in the values of the torsion angles allow to the atoms located on the surface to displace up to 1-2 Å from their original location.

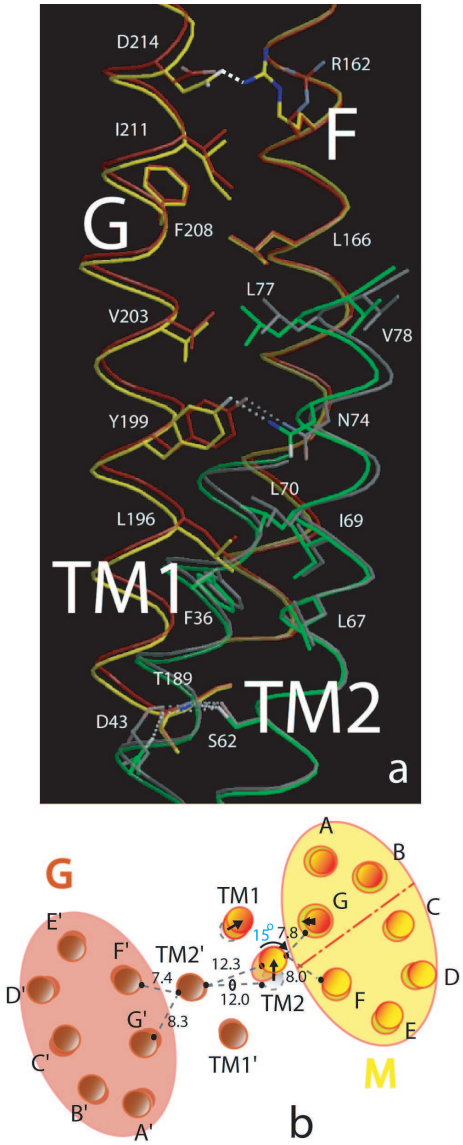

Figure 59: Overall views on helices F, G, TM1 and TM2 in ground and late M state. a. Structural changes along this helices; b. schematic picture of helical displacements viewed from the cytoplasmic surface, ground state complex in red and M state complex in yellow, black numbers gives the distances between C_α atoms of the following amino acids in ground and in M state: F(Leu170)-TM2(Leu77): ground state - 7.4 Å, M state - 8.0 Å, G(Val203)-TM2(Leu77): ground state - 8.3 Å, M state - 7.8 Å, TM2(Val78)-TM2'(Val78): ground state - 12.0 Å, M state - 12.3 Å. The arrows on TM1 and TM2 depict displacements parallel to the membrane plane in M state, the arrow on helix G depicts its downward movement. A clockwise rotation of 15^o is observed for TM2 and an in-plane displacement of 0.9 Å(black arrow) near the cytoplasmic membrane surface in M state with respect to the ground state. A red line in the receptor (colored in yellow) divides the molecule into the two functionally important domains ABG and CDEF

Four helices C, D, E and F are arranged in the three dimensional space so that they could be viewed as two intercalating V-like pairs CE and DF oriented up and down respectively. The angle of the mutual orientation around the normal to the membrane axis is 90^o and the overall shape of the domain is cylindrical. Such an arrangement with multiple hydrogen bonds between the helices in the intersection point of the wings defining V and number of the short loops interconnecting the helices apply a significant restriction to the motion freedom of the individual helices. Namely, it defines a rigidity of the whole arrangement. And indeed double distances between this helices are quite small and assignment of the second domain CDEF is valid.

The interface of helices F and G, as in the case of the BC pair, enables a shear motion of an appreciable amplitude. Altogether 7 Leu or Ile residues of helix F face 4 Leu or Ile residues of helix G. For this pair one direct hydrogen bond connection in the middle of the helices and two connections involving water molecules imply a restriction for the motional freedom.

The observed structural changes are in line with the above consideration of the ground state structure. In the course of the motion two domains clearly show up. The relative displacement of these domains describes almost all the rearrangement of the proteins tertiary structure. Helixes A, B and G being the first domain nicely conserve their relative packing in the M state with only a minor displacement of helix G relative to A and B in the region of the retinal. This domain moves concertedly from helices C,D,E and F. The component of the ABG domain displacement in the membrane plane is in average of 0.2 Å magnitude. One can consider this to be a sliding motion of the CDEF domain which occurs exactly along the retinal in the direction from the Lys205 of helix G to the β-ionone ring of the retinal. Such motion presumably allow the cis retinal to remove the strain implied by the tight packing of the residues of the retinal pocket which is present in the K state.

Observed relative displacements of the helix G with respect to helix F is of ≈ 0.6 Å magnitude in the direction towards the extracellular side and is attained via changes of torsion angles of Leu and Ile side chains in the interhelical surface. It presents the main movement occurring between the helices.

In comparison to the late M state of Bacteriorhodopsin where helix F exerted larger outward bending motion helix F does not show any significant changes in the late M state of pSRII-HtrII complex. A small exception is seen for the EF loop which is slightly shifted relative to the ground state conformation in the M state. Two reasons for that can be underlined: the larger EF loop of Bacteriorhodopsin compared to the short loop of pSRII and/or packing interactions with the adjacent layer in the crystal. The outward movement of helix F in Bacteriorhodopsin opens the cytoplasmic domain for the penetration of water molecules which are necessary for the reprotonation of the Schiff base and Asp96. In difference to this, pSRII when complexed with its transducer is not pumping protons and therefore an opening of the cytoplasmic domain is not required.

The motion of helix F is restricted in the crystal due to the crystal packing. The EF loop of pSRII from one crystal layer interacts closely with the BC loop of an adjacent crystal layer (Figure 60). This interaction can reduce the in-plane component of the helix F displacement.

The relative sliding of the domains occurs with changes of the torsion angles of several Leu and Ile residues in BC and FG helix pair interfaces. The shear motion is especially pronounced in the case of the F and G helix pair. The interface between the helices F and G seems to be designed to accommodate the most significant conformational change occurring in the tertiary structure of the receptor - the relative displacement of these helices.

The TM2 helix of the transducer is docked to the receptor exactly in between helices F and G. These helices create a groove on the surface which is occupied by the side chains of the TM2. Complementarity of TM2 and receptor segment surfaces is very high. An optimized docking by three hydrogen bonds, which involve residues Tyr199 and Thr189 of the receptor stabilize the

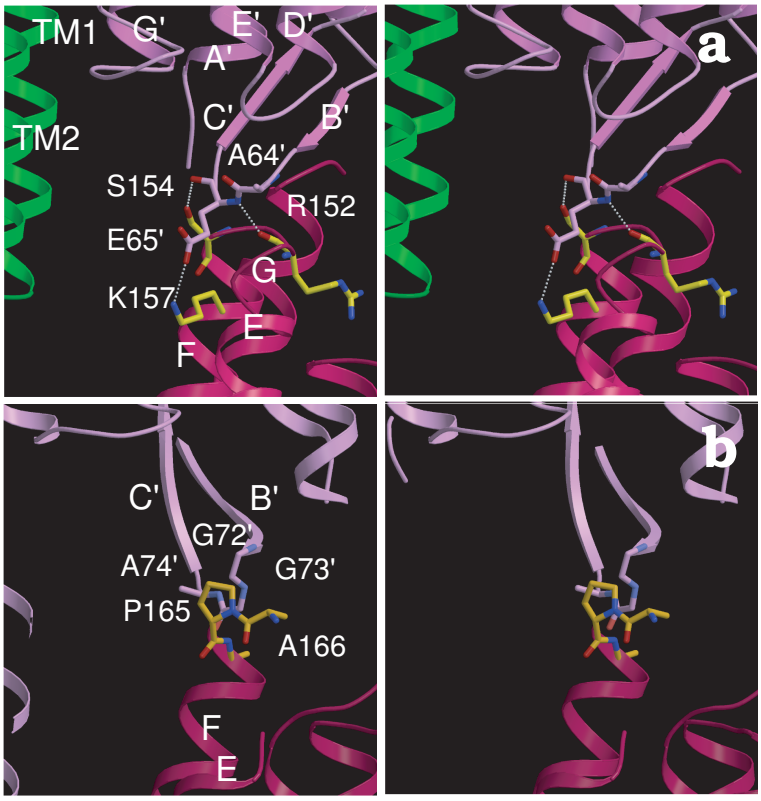

Figure 60: Stereo view of the crystal layer interactions. (a) Interaction between adjacent 'membrane' layers in pSRII-HtrII crystals of $p2_12_12$ symmetry involving 3 hydrogen bonds and a number of van der Waals interactions between the BC loop from one layer and the EF loop from the other. (b) Interaction between adjacent 'membrane' layers in Bacteriorhodopsin crystals of $p6_3$ symmetry involving van der Waals contacts between Bacteriorhodopsin BC loop from one layer and EF loop from the other.

protein complex.

All hydrogen bonds observed for the ground state between receptor and transducer are conserved in the late M state. The X-ray structure shows the constant hydrogen bond length between Tyr199 (helix G) and Asn74 (helix TM2) and a parallel shift of both amino acids of 0.6 Å to the extracellular side. The hydrogen bond orientation is preserved.

Apart from a possible contribution of electrostatic interactions to the signal transfer which are effective also over larger distances, signal transfer by the short range interactions takes place in the interface between helices F, G and TM2. In the signalling state surface of helices F and G towards TM2 is changed, thus TM2 helix of the transducer is displaced from its original position to fit the requirements of the new surface. This displacement can be described as a tilting motion. In addition TM2 in the upper part (residues 76-79) rotates to $\approx 15^o$ around its axis clockwise as seen from cytoplasmic side. All the hydrogen bonds between the receptor and transducer are conserved for this motion in late M state. This rotation of TM2 agrees with the values deduced from EPR-measurements, but it significantly exceeds the values observed for the other helices

which do not exceed 4^o.

Distance changes between helices F and TM2 are quite small with exception of the cytoplasmic side where a distance increase of 0.6 Å is observed. This change is solely produced by the in-plane movement of TM2. Double distances of helices G and TM2 are larger so that the helix G plays an active role in changing the interaction surface faced by TM2 and in this was conducts the signal from the receptor to the transducer. The overall signal is a clockwise rotation of TM2 by 15^o (observed from the cytoplasmic side) and a tilt of TM2 giving an in-plane displacement of 0.9 Å at the cytoplasmic side with the hinge at Ser62 which is hydrogen bonded to Thr189 (helix G).

5.2 Structure of the M state of Bacteriorhodopsin

In response to the absorption of a photon, the retinal and its nearest environment are changed during intermediates K and L to allow the proton transfer from the Schiff base to Asp85. In the late M state there are further changes which make possible the uptake of the proton from the cytoplasmic side. Retinal and Lys216 undergo further structural rearrangements. Changes in torsion angles of the retinal compared to its L state conformation (Royant et al. [32]) happen in the late M state as a new volume from the cytoplasmic side is now available for the retinal atoms. Retinal atoms C11,C12, C13, C14 and C15 are additionally displaced in the direction of the membrane normal to the cytoplasmic side and as a consequence the dihedral and bond angles acquire values expected for the relaxed 13-cis retinal. The Schiff base nitrogen is oriented to the cytoplasmic side. The carbonyl oxygen of Lys216 is oriented in the plane perpendicular to helix G like in the L state. The side chain of the Lys216 has the same conformation as in the L state ([32]).

The central water cluster together with oxygens of Asp85 and Asp212 (pentagon of hydrogen bonds in the ground state) undergo rearrangements (Figure 61a). This region is arranged in the same way in both Bacteriorhodopsin and Sensory rhodopsin II in the ground state. The conservation of the structure of this region for two proteins suggests that similar mechanism should transfer the proton from the Schiff base to Asp85 (Asp75 for pSRII).

Indeed the late M state structure of this region for Bacteriorhodopsin is almost identical to that of pSRII in the late M state. Water molecule W1 is disordered and water molecule W2 occupies a new position between aspartic acids. But in the new location water molecule W2 still connects Asp85 and Asp212 in Bacteriorhodopsin M state. This connection is lost in Sensory rhodopsin II late M state. Asp85 has changed its orientation in comparison to the ground state conformation by rotation to approximately 90^o (similar to pSRII) (Figure 61b).

From both structures a uniform mechanism of the initial proton transfer emerges. Namely after isomerization of the retinal Lys216 (205 for pSRII) changes its conformation so that $C\delta$ and $C\epsilon$ atoms are displaced to the new position approximately 1 Å away from the original location. In the new position they collide with water molecule W1 and consequently this water molecule becomes disordered because it can not form any longer hydrogen bonds to aspartic acids and the Schiff base nitrogen. The negatively charged Asp85 (75, pSRII) attracted by the positive charge of the Schiff base not shielded any longer by water molecule W1 reorients its side chain. The hydrogen bond of Asp85 (75, pSRII) to Thr89 (79, pSRII) is lost. The new environment of Asp85(75, pSRII) makes it ready to accept the proton from the Schiff base by adjustment of its pK_a.

The proton translocation to the Asp85 changes the electrostatic field inside the protein. This happens in line with the reorientation of the Arg82 side chain. In M state this residue points down to the extracellular surface of the protein (Figure 62c). Together with this residue surrounding water molecules change their location. It should be noticed that there is no such change in the Sensory rhodopsin II case, where Arg72 (analogous to Arg82) already in the ground state points down to the extracellular surface. This difference has consequences for the proton release mechanism.

Glu194 and Glu204 are residues included in the proton release group . These groups together with a surrounding water cluster share the proton which should be released in M state (Gerwert et al. [91]) (Figure 62d). In ground state these residues face each other. In M state the conformation of the release group is changed. Both glutamic acids reorient in the late M state so that the distance between them increases (Figure 62e). This new conformation of the release group accounts for a changed pK_a of the glutamic acids. Eventually a proton is transmitted to the extracellular bulk water.

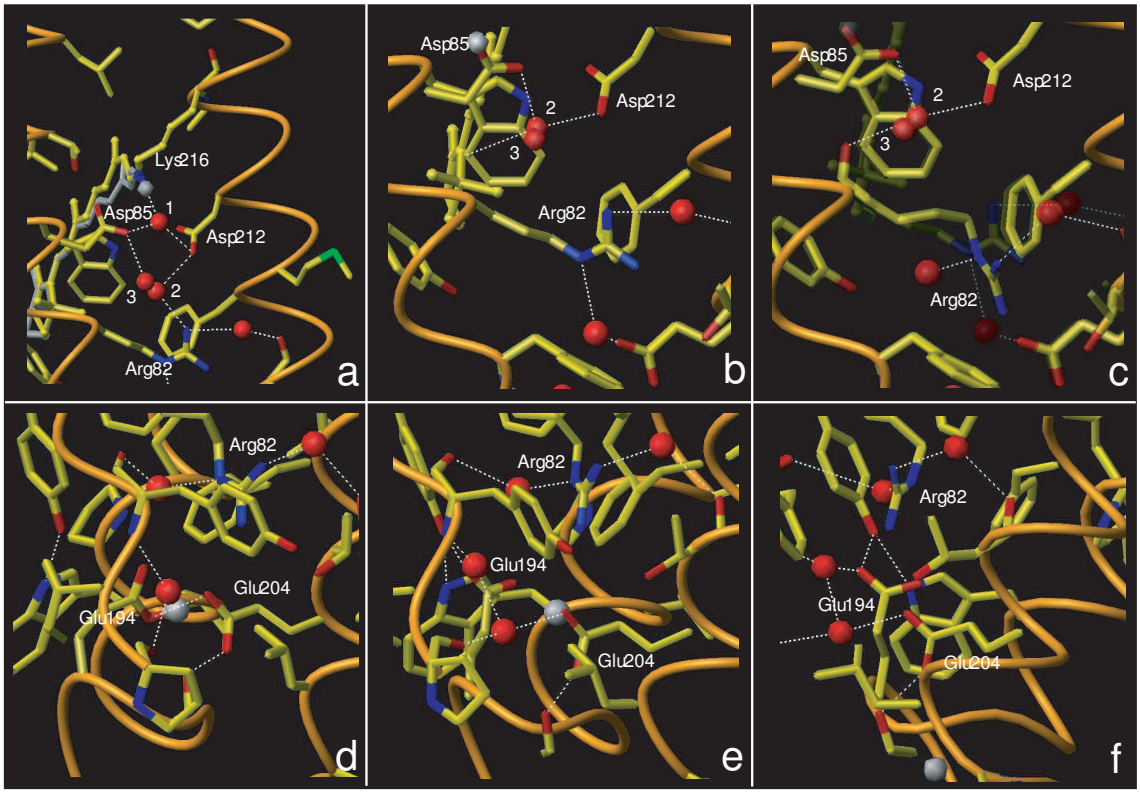

Figure 61: Developing changes leading to proton release in extracellular part of BR in late M state. a. Ground state, b. Initial proton transfer, c. Reorientation of Arg82, d. Proton release group, e. Change in proton release group, f. proton release to extracellular medium

Significant rearrangements of the tertiary structure happen in the late M state. As in the sensory rhodopsin case helix G is displaced towards the extracellular side. The magnitude of the displacement is 0.6 Å. Helix F is tilted in its cytoplasmic part with a maximum displacement of about 1.3 Å occurring between Leu174 and Asn176. A similar motion of helix F is not observed in the pSRII/Transducer complex.

The changes described so far are in a good agreement with the previously reported structure of the late M state of wild type Bacteriorhodopsin. (Sass et al. [5]). But improved resolution and higher occupancy of the M state as well as lower twinning of the crystal used for data collection resolve some new features. Among the most important new features is the more complete network of water molecules between Asp96 (residue which donates a proton to the Schiff base during M to N transition) and the Schiff base (Figure 63a,b). An additional water molecule located close to the Schiff base extends the previously observed (in Sass structure) water network. There is still a distance of 5.1 Å between this water molecule and the Schiff base, therefore the pathway is not yet active. The N state structure from Lanyi et al. [28] reveals the water molecule close to the Schiff base which form the network that provides connection between Asp96 and the Schiff base.

A displacement of helix F opens the cytoplasmic channel and additional water molecules can penetrate the channel and consequently assist to the reprotonation of Asp96 (Grudinin et al. [92]). None of those water molecules is crystallographically observed in the new structure of the late M state of Bacteriorhodopsin.

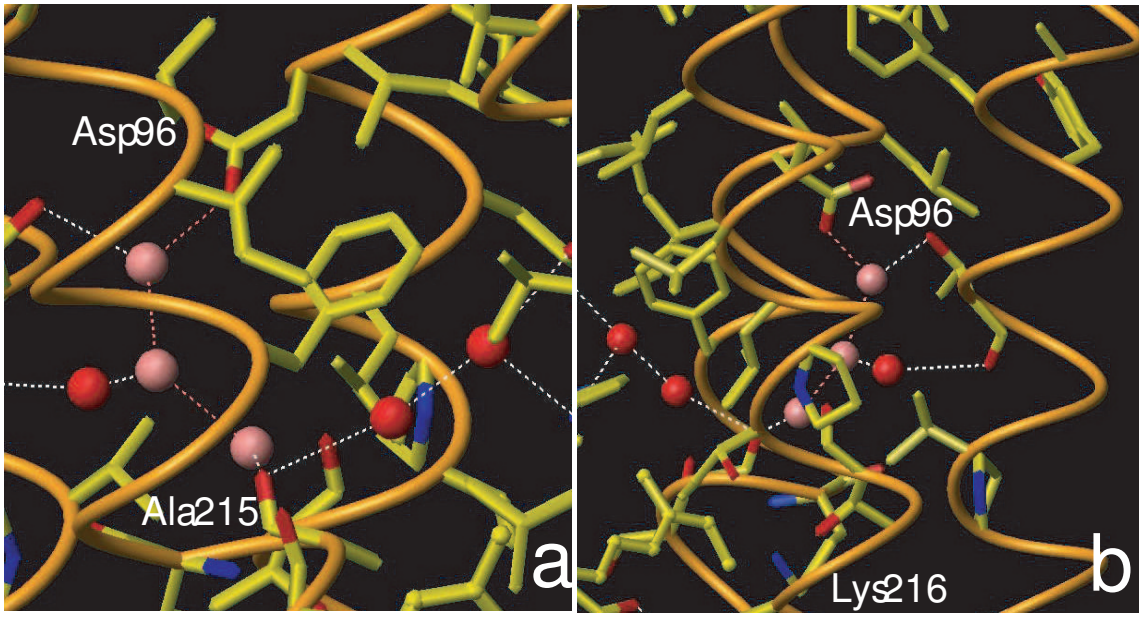

Figure 62: Proton translocation pathway between Asp96 and Schiff base in late M state. Pink are water molecule chain appearing in the M state.

5.3 Relation between ion pumping and signal transduction

Structures of the late M state of Bacteriorhodopsin and the Sensory rhodopsin II/Transducer complex could provide an answer to the question of why highly homologous proteins with similar geometry could perform such different functions as ion pumping and signal transduction.

Clearly similar events happen in the extracellular vicinity of the retinal in Bacteriorhodopsin and the pSRII/Transducer complex in the first half of the photocycle as it is evident from the comparison of the late M state structures. This justifies the suggestion that the initial proton transfer from the Schiff base to the corresponding aspartic acid is governed by the same mechanism. Also this could explain why the lifetimes of the early intermediates are very similar in the photocycles of the two proteins.

Since in the extracellular part there are differences with respect to important proton release residues, proton release to the extracellular media proceeds for these proteins in different ways. In sensory rhodopsin II the Arg72 side chain in M state as well as in ground state points to the extracellular surface whereas in Bacteriorhodopsin this residue reorients to that conformation only in the M state. Also the Glu194 residue is exchanged to Pro183 in the Sensory rhodopsin II. These facts could account for the difference in proton release.

Proton uptake is different for these proteins as well. In sensory rhodopsin II Phe86 is at the position equivalent to the position of Asp96 in Bacteriorhodopsin. This alone is believed to influence significantly the proton uptake. In addition the opening of the cytoplasmic channel due to the helix F tilt is not observed for Sensory rhodopsin II/Transducer complex which implies a lower possibility for reprotonation of the Schiff base from the cytoplasmic side. Absence of the tilt of helix F in the pSRII/Transducer photocycle could account for the blocking of the proton transport activity of sensory rhodopsin II in complex with the transducer. The presence of the transducer could hinder the possible helix F tilt since the TM2 helix of the transducer is in the way of this movement.

Sensory rhodopsin II uses the relative displacement of helixes F and G to relay a signal to

the transducer. These displacements are correlated with the conformation of the water molecule cluster near retinal. Minor changes in the conformation or type of residues found at homologous places of the 7 transmembrane protein structure could result in a quite different evolution of the structures of the water molecule clusters during the photocycle which would produce the different extent of the tertiary structure rearrangement. For example optimization of the residues surrounding water molecules could provide a hydrogen bond environment which will trap this water molecule at the certain position and so modify the structure of the water molecule network.

5.4 Conclusions

The archaeal rhodopsins are the most understood proteins among seven-helix receptors with respect to structural information from X-ray crystallography. High-resolution structures have already been obtained for three of the proteins within this family: Halorhodopsin, Bacteriorhodopsin and Sensory rhodopsin II (SRII). Recent progress concerns the nature of the interaction between archaeal sensory rhodopsins and their integral membrane transducer proteins during the signalling process, and the relationship between ion transport and sensory signalling by this family of proteins.

Receptor activation and conformational coupling with the transducer is understood at the atomic level for Sensory rhodopsin II and its cognate transducer complex from *N. pharaonis*. The complex of pSRII and HtrII(1-114) was crystallized and its structure was solved by X-ray at 1.94 Å resolution. This significant structural result provided a picture of how Sensory rhodopsin II is associated with its transducer within a dimer complex, and it opened the broad possibilities for further studies of how a light driven structural signal can propagate from sensor to transducer.

In this thesis the structure of the ground state, the early K state and the signalling M state of Sensory rhodopsin II and its cognate transducer as well as the structure of the ground state and the late M state of Bacteriorhodopsin were investigated by means of X-ray crystallography. Crystals which provided the resolution of 1.9 Å for ground state, 2.0 Å for K state and 2.2 Å for M state of Sensory rhodopsin II/Transducer complex and 1.35 Å for ground state, 1.5 Å for late M state of Bacteriorhodopsin were grown in the lipidic cubic phase. Structures of these photocycle states of archaerhodopsins brought a wide insight into how the seven helical arrangement and highly homologous structure of retinal pocket conducts two different functions: sensory signalling and ion pumping.

In the case of sensory signalling light activated signal transfer from receptor to transducer starts from the retinal isomerization, which creates local changes in a region near the retinal and modifies hydrogen bonding network mediated by a central water molecule cluster and residues in the retinal vicinity (K state). During the transition to the signalling state (late M) these disturbances increase so that in late M helices C and G become strongly decoupled as fewer hydrogen bonds connect them. These alterations in the hydrogen bond network changes the pK_a values of Schiff base and proton acceptor group Asp75, thereby enabling proton transfer which produces a redistribution of charges (M state). Thus in the late M state tertiary structural changes occur. The receptor behaves as consisting of two domains including helices ABG and CDEF. These domains move relative to each other creating the strongest change in the interface of helices F and G. Signal transfer to the transducer takes place within the interface of receptor helices F and G and transducer helix TM2. The main effect on TM2 is a clockwise rotation of about 15° and a tilt with the hinge at Ser62 which amounts to 0.9 Å at the cytoplasmic side.

In the case of ion pumping isomerization of the retinal induced by light perturbs the central water cluster structure and initiates in the M state the initial proton transfer from Schiff base to Asp85. This charge rearrangement along with hydrogen bonding network changes is similar to that happening in Sensory rhodopsin II. But as a consequence the Arg82 residue placed in the extracellular part of the ion pumping channel reorients in Bacteriorhodopsin which leads to the proton release to the extracellular medium by a group of charged residues, probably together with a water cluster, located on the surface of the protein. In contrast to Sensory rhodopsin II helices C and G still have hydrogen bond connection and in the cytoplasmic part the channel opens by tilting of helix F and a water molecule network builds up to conduct the proton from the cytoplasmic surface to the retinal Schiff Base via Asp96 residue. The structure of Bacteriorhodopsin late M state at 1.5 Å resolution reveals how water network connecting Asp96 and retinal Schiff base is extended relative to previous studies [5], [28].

Current work provided an important insight in the functioning of two membrane localized machines but still a number of questions has to be addressed in future investigations on Bacteriorhodopsin and Sensory rhodopsin II/Transducer complex. Particularly high resolution atomic structures of other photocycle states will accomplish our understanding of these proteins. In the case of Sensory rhodopsin II/Transducer complex a structure of the complex with a longer transducer including the linker domain in the ground and signalling state have to be resolved in order to provide an answer to the question of how the signal is transmitted to the kinase protein located at a distance of 260 Å from the receptor. The putative folding model of the ground state complex with longer transducer is defined in this work. Better data is needed to yield the complete structure of the linker domain.

It would be interesting to investigate the structure of photocycle states of Sensory rhodopsin II/Transducer complex from *H.salinarum*. In this complex the transducer HtrII plays a dual role as it participates both in photo- and chemotaxis. Also the structure of Sensory rhodopsin I/Transducer complex would contribute to our view on light induced transmembrane signalling.

References

[1] Spudich JL, Yang CS, Jung KH, Spudich EN. Retinylidene proteins: structures and functions from archaea to humans. Annu Rev Cell Dev Biol. 2000;16:365-92. Review.

[2] F.T.Robb, A.R.Place Archaea. A laboratory manual. 1995

[3] Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R. Electron-crystallographic refinement of the structure of Bacteriorhodopsin. J Mol Biol. 1996 Jun 14;259(3):393-421.

[4] Pebay-Peyroula E, Rummel G, Rosenbusch JP, Landau EM. X-ray structure of Bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. Science. 1997 Sep 12;277(5332):1676-81.

[5] Sass HJ, Buldt G, Gessenich R, Hehn D, Neff D, Schlesinger R, Berendzen J, Ormos P. Structural alterations for proton translocation in the M state of wild-type Bacteriorhodopsin. Nature. 2000 Aug 10;406(6796):649-53.

[6] Luecke H, Richter HT, Lanyi JK. Proton transfer pathways in Bacteriorhodopsin at 2.3 angstrom resolution. Science. 1998 Jun 19;280(5371):1934-7.

[7] Luecke H, Schobert B, Richter HT, Cartailler JP, Lanyi JK. Structure of Bacteriorhodopsin at 1.55 A resolution. J Mol Biol. 1999 Aug 27;291(4):899-911.

[8] Kolbe M, Besir H, Essen LO, Oesterhelt D. Structure of the light-driven chloride pump Halorhodopsin at 1.8 A resolution. Science. 2000 May 26;288(5470):1390-6.

[9] Kunji ER, Spudich EN, Grisshammer R, Henderson R, Spudich JL. Electron crystallographic analysis of two-dimensional crystals of Sensory rhodopsin II: a 6.9 A projection structure. J Mol Biol. 2001 Apr 27;308(2):279-93.

[10] Royant A, Nollert P, Edman K, Neutze R, Landau EM, Pebay-Peyroula E, Navarro J. X-ray structure of Sensory rhodopsin II at 2.1-A resolution. Proc Natl Acad Sci U S A. 2001 Aug 28;98(18):10131-6.

[11] Schertler GF, Hargrave PA. Projection structure of frog rhodopsin in two crystal forms. Proc Natl Acad Sci U S A. 1995 Dec 5;92(25):11578-82.

[12] Davies A, Schertler GF, Gowen BE, Saibil HR. Projection structure of an invertebrate rhodopsin. J Struct Biol. 1996 Jul-Aug;117(1):36-44.

[13] Krebs A, Villa C, Edwards PC, Schertler GF. Characterisation of an improved two-dimensional p22121 crystal from bovine rhodopsin. J Mol Biol. 1998 Oct 9;282(5):991-1003.

[14] Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. Science. 2000 Aug 4;289(5480):739-45.

[15] Alam M, Oesterhelt D. Morphology, function and isolation of halobacterial flagella. J Mol Biol. 176, 459-475

[16] Hildebrand E, Schimz A. Integration of photosensory signals in Halobacterium halobium. J Bacteriol. 1986 Jul;167(1):305-11.

[17] Alam M, Lebert M, Oesterhelt D. Methyl-accepting taxis proteins in *H. halobium*. 1989. EMBO J., 8, 631-639

[18] Falke JJ, Bass RB, Butler SL, Chervitz SA, Danielson MA. The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. Annu Rev Cell Dev Biol. 1997;13:457-512. Review.

[19] Oesterhelt D. The structure and mechanism of the family of retinal proteins from halophilic archaea. Curr Opin Struct Biol. 1998 Aug;8(4):489-500. Review.

[20] Stoeckenius W, Oesterhelt D. Nat New Biol. 233, 1971, 149

[21] Racker E, Stoeckenius W. Reconstitution of purple membrane vesicles catalyzing light-driven proton uptake and adenosine triphosphate formation. J Biol Chem. 1974 Jan 25;249(2):662-3.

[22] Casadio R, Gutowitz H, Mowery P, Taylor M, Stoeckenius W. Light-dark adaptation of Bacteriorhodopsin in triton-treated purple membrane. Biochim Biophys Acta. 1980 Mar 7;590(1):13-23.

[23] Henderson R, Unwin PN. Three-dimensional model of purple membrane obtained by electron microscopy. Nature. 1975 Sep 4;257(5521):28-32.

[24] Nollert P, Qiu H, Caffrey M, Rosenbusch JP, Landau EM. Molecular mechanism for the crystallization of Bacteriorhodopsin in lipidic cubic phases. FEBS Lett. 2001 Aug 31;504(3):179-86. Review.

[25] Edman K, Nollert P, Royant A, Belrhali H, Pebay-Peyroula E, Hajdu J, Neutze R, Landau EM. High-resolution X-ray structure of an early intermediate in the Bacteriorhodopsin photocycle. Nature. 1999 Oct 21;401(6755):822-6.

[26] Royant A, Edman K, Ursby T, Pebay-Peyroula E, Landau EM, Neutze R. Helix deformation is coupled to vectorial proton transport in the photocycle of Bacteriorhodopsin. Nature. 2000 Aug 10;406(6796):645-8.

[27] Schobert B, Cupp-Vickery J, Hornak V, Smith S, Lanyi J. Crystallographic structure of the K intermediate of Bacteriorhodopsin: conservation of free energy after photoisomerization of the retinal. J Mol Biol. 2002 Aug 23;321(4):715-26.

[28] Schobert B, Brown LS, Lanyi JK. Crystallographic structures of the M and N intermediates of Bacteriorhodopsin: assembly of a hydrogen-bonded chain of water molecules between Asp-96 and the retinal Schiff base. J Mol Biol. 2003 Jul 11;330(3):553-70.

[29] Rouhani S, Cartailler JP, Facciotti MT, Walian P, Needleman R, Lanyi JK, Glaeser RM, Luecke H. Crystal structure of the D85S mutant of Bacteriorhodopsin: model of an O-like photocycle intermediate. J Mol Biol. 2001 Oct 26;313(3):615-28.

[30] Heberle J, Fitter J, Sass HJ, Buldt G. Bacteriorhodopsin: the functional details of a molecular machine are being resolved Biophys Chem. 2000 Jul 15;85(2-3):229-48.

[31] Haupts U, Tittor J, Bamberg E, Oesterhelt D. General concept for ion translocation by halobacterial retinal proteins: the isomerization/switch/transfer (IST) model. Biochemistry. 1997 Jan 7;36(1):2-7.

[32] Edman K, Royant A, Larsson G, Jacobson F, Taylor T, van der Spoel D, Landau EM, Pebay-Peyroula E, Neutze R. Deformation of helix C in the low temperature L-intermediate of Bacteriorhodopsin. J Biol Chem. 2004 Jan 16;279(3):2147-58.

[33] Tomioka H, Sasabe H. Isolation of photochemically active archaebacterial photoreceptor, pharaonis phoborhodopsin from Natronobacterium pharaonis. Biochim Biophys Acta. 1995 Mar 22;1234(2):261-7.

[34] Shimono K, Iwamoto M, Sumi M, Kamo N. Functional expression of pharaonis phoborhodopsin in Escherichia coli. FEBS Lett. 1997 Dec 22;420(1):54-6.

[35] Chizhov I, Schmies G, Seidel R, Sydor JR, Luttenberg B, Engelhard M. The photophobic receptor from Natronobacterium pharaonis: temperature and pH dependencies of the photocycle of Sensory rhodopsin II. Biophys J. 1998 Aug;75(2):999-1009.

[36] Sudo Y, Iwamoto M, Shimono K, Sumi M, Kamo N. Photo-induced proton transport of pharaonis phoborhodopsin (Sensory rhodopsin II) is ceased by association with the transducer. Biophys J. 2001 Feb;80(2):916-22.

[37] Schmies G, Engelhard M, Wood PG, Nagel G, Bamberg E. Electrophysiological characterization of specific interactions between bacterial sensory rhodopsins and their transducers. Proc Natl Acad Sci U S A. 2001 Feb 13;98(4):1555-9.

[38] Schmies G, Luttenberg B, Chizhov I, Engelhard M, Becker A, Bamberg E. Sensory rhodopsin II from the haloalkaliphilic natronobacterium pharaonis: light-activated proton transfer reactions. Biophys J. 2000 Feb;78(2):967-76.

[39] Kandori H, Shimono K, Sudo Y, Iwamoto M, Shichida Y, Kamo N. Structural changes of pharaonis phoborhodopsin upon photoisomerization of the retinal chromophore: infrared spectral comparison with Bacteriorhodopsin. Biochemistry. 2001 Aug 7;40(31):9238-46.

[40] Sudo Y, Iwamoto M, Shimono K, Kamo N. Role of charged residues of pharaonis phoborhodopsin (Sensory rhodopsin II) in its interaction with the transducer protein. Biochemistry. 2004 Nov 2;43(43):13748-54.

[41] Le Moual H, Koshland DE Jr. Molecular evolution of the C-terminal cytoplasmic domain of a superfamily of bacterial receptors involved in taxis. J Mol Biol. 1996 Aug 30;261(4):568-85.

[42] Stoddard BL, Biemann HP, Koshland DE Jr. Receptors and transmembrane signaling. Cold Spring Harb Symp Quant Biol. 1992;57:1-15. Review

[43] Kim KK, Yokota H, Kim SH. Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. Nature. 1999 Aug 19;400(6746):787-92.

[44] Jung KH, Spudich EN, Trivedi VD, Spudich JL. An archaeal photosignal-transducing module mediates phototaxis in Escherichia coli. J Bacteriol. 2001 Nov;183(21):6365-71.

[45] Zhang XN, Zhu J, Spudich JL. The specificity of interaction of archaeal transducers with their cognate sensory rhodopsins is determined by their transmembrane helices. Proc Natl Acad Sci U S A. 1999 Feb 2;96(3):857-62.

[46] Hippler-Mreyen S, Klare JP, Wegener AA, Seidel R, Herrmann C, Schmies G, Nagel G, Bamberg E, Engelhard M. Probing the Sensory rhodopsin II binding domain of its cognate transducer by calorimetry and electrophysiology. J Mol Biol. 2003 Jul 25;330(5):1203-13.

[47] Gordeliy VI, Labahn J, Moukhametzianov R, Efremov R, Granzin J, Schlesinger R, Buldt G, Savopol T, Scheidig AJ, Klare JP, Engelhard M. Molecular basis of transmembrane

signalling by Sensory rhodopsin II-transducer complex.

[48] Zhang W, Brooun A, Mueller MM, Alam M. The primary structures of the Archaeon Halobacterium salinarium blue light receptor sensory rhodopsin II and its transducer, a methyl-accepting protein. Proc Natl Acad Sci U S A. 1996 Aug 6;93(16):8230-5.

[49] Hou S, Brooun A, Yu HS, Freitas T, Alam M. Sensory rhodopsin II transducer HtrII is also responsible for serine chemotaxis in the archaeon Halobacterium salinarum. J Bacteriol. 1998 Mar;180(6):1600-2.

[50] Yeh JI, Biemann HP, Prive GG, Pandit J, Koshland DE Jr, Kim SH. High-resolution structures of the ligand binding domain of the wild-type bacterial aspartate receptor. J Mol Biol. 1996 Sep 20;262(2):186-201.

[51] Ihara K, Umemura T, Katagiri I, Kitajima-Ihara T, Sugiyama Y, Kimura Y, Mukohata Y. Evolution of the archaeal rhodopsins: evolution rate changes by gene duplication and functional differentiation. J Mol Biol. 1999 Jan 8;285(1):163-74.

[52] Klare JP, Bordignon E, Engelhard M, Steinhoff HJ. Sensory rhodopsin II and Bacteriorhodopsin: light activated helix F movement. Photochem Photobiol Sci. 2004 Jun;3(6):543-7.

[53] Wegener AA, Chizhov I, Engelhard M, Steinhoff HJ. Time-resolved detection of transient movement of helix F in spin-labelled pharaonis sensory rhodopsin II. J Mol Biol. 2000 Aug 25;301(4):881-91.

[54] Ottemann KM, Xiao W, Shin YK, Koshland DE Jr. A piston model for transmembrane signaling of the aspartate receptor. Science. 1999 Sep 10;285(5434):1751-4.

[55] Kim SH, Wang W, Kim KK. Dynamic and clustering model of bacterial chemotaxis receptors: structural basis for signaling and high sensitivity. Proc Natl Acad Sci U S A. 2002 Sep 3;99(18):11611-5.

[56] Bailon, Pascal. Affinity chromatography: methods and protocols. Totowa, NJ: Humana Press. 2000

[57] Rigaud JL, Pitard B, Levy D. Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. Biochim Biophys Acta. 1995 Oct 10;1231(3):223-46. Review.

[58] Landau EM, Rosenbusch JP. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. Proc Natl Acad Sci U S A. 1996 Dec 10;93(25):14532-5.

[59] Misquitta LV, Misquitta Y, Cherezov V, Slattery O, Mohan JM, Hart D, Zhalnina M, Cramer WA, Caffrey M. Membrane protein crystallization in lipidic mesophases with tailored bilayers. Structure (Camb). 2004 Dec;12(12):2113-24.

[60] Giacovazzo, Carmelo, Monaco. Fundamentals of crystallography. Oxford : Oxford University Pr. 1992

[61] Taylor G. The phase problem. Acta Crystallogr D Biol Crystallogr. 2003 Nov;59(Pt 11):1881-90. Review.

[62] Navaza J. Rotation functions. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[63] Tong L. Translation functions. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[64] Carvin D, Blundell T. Isomorphous replacement. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[65] Matthews BW. Heavy-atom location and phase determination with single wavelength diffraction data. International tables for crystallography. F. Crystallography of biological macromolecules. / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[66] Blundell TL, Johnson LN. Protein crystallography. New York : Academic Press. 1976

[67] Lindley PF. Macromolecular crystallography with a third-generation synchrotron source. Acta Cryst. 1999. D55, 1654-1662

[68] Leslie AGW. Integration of macromolecular diffraction data. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[69] Leslie AGW. Recent changes to the MOSFLM package for processing film and image data. CCP4 ESF-EACMB Newsletter Protein Crystallography 26 (2002).

[70] Evans PR. Data reduction. Proceedings of CCP4 Study Weekend, 1993, on Data Collection & Processing. 114-122

[71] Ten Eyck LF. Introduction to refinement. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[72] N.S. Pannu, G.N. Murshudov, E.J. Dodson, R.J. Read. 1998. Incorporation of prior phase information strengthens maximum likelihood structure refinement." Acta Cryst. D54, 1285-1294.

[73] Brunger AT, Adams PD. Enhanced macromolecular refinement by simulated annealing. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[74] Lamzin VS, Perrakis A, Wilson KS. The ARP/wARP suite for automated construction and refinement of protein models. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[75] Engh RA, Huber R. Structure quality and target parameters. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[76] Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst., 26, 283-291.

[77] Cruickshank DWJ. Coordinate uncertainty. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[78] Luzatti, V. (1952). Acta Cryst. 5, 802-810

[79] Rossmann MG., Automatic indexing of oscillation images. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[80] Vagin A, Teplyakov A. MOLREP: an Automated Program for Molecular Replacement. J. Appl. Cryst. (1997). 30, 1022-1025

[81] Collaborative computational project The CCP4 suite: programs for protein crystallography. Acta Crystallogr D54, 760-763. 1994.

[82] Kleywegt GJ, Jones TA. Atound O. International tables for crystallography. F. Crystallography of biological macromolecules / edited by Rossmann MG. Dordrecht : Kluwer Academic Publ. , 2001

[83] Parsons S. Introduction to twinning. Acta Cryst. (2003). D59, 1995-2003

[84] Efremov R, Moukhametzianov R, Buldt G, Gordeliy V. Physical detwinning of hemihedrally twinned hexagonal crystals of Bacteriorhodopsin. Biophys J. 2004 Nov;87(5):3608-13.

[85] Yeates TO. Detecting and Overcoming Crystal Twinning. 1997 Methods in Enzymology 276, 344-358.

[86] Stanley E. The identification of twins from intensity statistics. J. Appl. Cryst. (1972). 5, 191-194

[87] Britton D. Estimation of twinning parameter for twins with exactly superimposed reciprocal lattices. Acta Cryst. (1972). A28, 296-297

[88] Rees DC. The influence of Twinning by Merohedry on Intensity Statistics. Acta Cryst. 1980 A36, 578-581.

[89] Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. 1998 Sep 1;54 (Pt 5):905-21.

[90] Bhat TN. Calculation of an OMIT map. J. Appl. Cryst. (1988). 21, 279-281

[91] Spassov VZ, Luecke H, Gerwert K, Bashford D. pK(a) Calculations suggest storage of an excess proton in a hydrogen-bonded water network in bacteriorhodopsin. J Mol Biol. 2001 Sep 7;312(1):203-19.

[92] Grudinin S, Buldt G, Gordeliy V, Baumgaertner A. Water molecules and hydrogen-bonded networks in bacteriorhodopsin–molecular dynamics simulations of the ground state and the M-intermediate. Biophys J. 2005 May;88(5):3252-61.

Acknoledgements

- I would like to express my grate gratitude to my supervisor, Prof. G.Büldt, for guiding my doctoral work, giving me the opportunity to make my doctoral research in ISR and his kindness.
- I am grateful to Dr Valentin Ivanovich Gordeliy my co-supervisor, for intellectual support and sharing his life experience.
- Special thanks are addressed to guest Prof. Martin Engelhard and Johann Klare, who introduce me to the sensory signalling field, for the distribution of knowledge during scientific discussions and substantial help in the scientific project.
- Sincere thanks are due to Dr. J. Granzin for introducing me to protein crystallography field and graphical art in crystallography and to Dr. J. Labahn for reading of my thesis, and many useful suggestions, scientific disscusions and introduction in the art of scientific writing he has made.
- I deeply thank Dr Rouslan Efremov for performing spectroscopic measurements without which part of this work have not been feasible.
- Special thanks are due to Ramona Schlesinger for help with protein preparation, carefulness in the laboratory and her kindness.
- I really pay respect to Dr. J. Fitter and Dr. J.Heberle for their scientific ideas and discussions.
- Sincere thanks are due to Oliver Weiergräber and Sabina Pohlmeyer for their support
- Sincere thanks are due to Oliver Weiergräber and Sascha Lehmann for computer administration.
- I admire the support from Kristian Bäcken, Ilona Ritter and Birgit Gehrmann in all the different aspects of my life in the laboratory.
- I keep in mind the grate gratitude to Marita Bergs, Wolfgang Bergs, Brigitte Neifer, Gerd Ransone, Werner Ullrich, Willy Pelzer, Willy Lange and all of their family members, for their love, friendship, support and all personal cares during my stay in Germany.
- Sincere thanks are due to Ken and Aki Ataka for their kindness.
- Sincere thanks are due to Achmed Husainovich Islamov, Aleksander Ivanovich Kuklin, Vasiliy Matveevich Smeyan, Alexander Nikiforovich Ozerin and Aziz Mansurovich Muzafarov.
- Sincere thanks are due to my old and new friends from Moscow Institute of Physics and Technology (MIPT) as well as rector, decans and all my teachers.
- Sincere thanks are due to Alexander Yaroschuk, Olga and Ivan Budyak, Sergei Grudinin, Britta Person, Darko Skegro, Jan Kubicek, Dominick Immeln, Teresa Bednarzt and Tilman Kottke
- I would like to express to my grate gratitude to my admirable grandparents, beloved parents, all the family members and all teachers, who gave me physical and mental power, nice education, bearing good moral and right attitude since my childhood and to succeed this study in Germany.
- Finally, I would like to thank Deutsche Forschung Gemeinschaft for supporting this study.