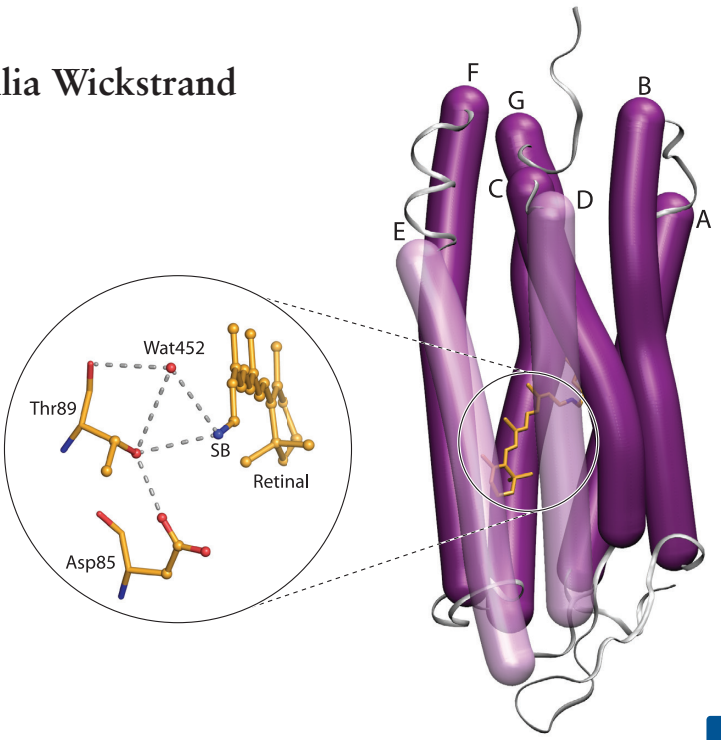
PH.D. THESIS

Production and Structural Dynamics of Microbial Rhodopsins

Cecilia Wickstrand



DEPARTMENT OF CHEMISTRY
AND MOLECULAR BIOLOGY

UNIVERSITY OF
GOTHENBURG

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE NATURAL SCIENCES

Production and Structural Dynamics of Microbial Rhodopsins

Cecilia Wickstrand

UNIVERSITY OF GOTHENBURG

Department of Chemistry and Molecular Biology

Gothenburg, 2019

Thesis for the degree of Doctor of Philosophy
in the Natural Sciences

Production and Structural Dynamics of Microbial Rhodopsins

Cecilia Wickstrand

Cover: Bacteriorhodopsin model from time-resolved serial femtosecond crystallography at the time 760 ns after light-activation, illustrating the hydrogen bonding network before the primary proton transfer reaction

ISBN 978-91-7833-492-6 (PRINT)
ISBN 978-91-7833-493-3 (PDF)
Available online at http://hdl.handle.net/2077/60173

Department of Chemistry and Molecular Biology
Division of Biochemistry and Structural Biology
University of Gothenburg
SE-405 30 Göteborg, Sweden

Printed by BrandFactory
Göteborg, Sweden, 2019

Till Henrik, Miranda och Vilma

Abstract

Rhodopsins are a family of membrane proteins that are found in a wide range of organisms and provide them with the ability to sense and react to light. When light is absorbed by these proteins, structural changes occur that initiate the pumping of protons, opening of a channel or signal transfer. The rhodopsins have different functions but they are structurally similar. This makes the protein family interesting from the perspective of structural biology, a field that aims to link function of proteins to structure and dynamics.

The structure of a protein may be obtained with X-ray crystallography. It requires a pure protein sample that can form protein crystals, data collection at an X-ray source and analysis of data for structural modelling and interpretation. In the last ten years, the development of X-ray free electron lasers (XFELs) generating very intense and short X-ray pulses has made it possible to capture structural changes in real-time with a time resolution of femtoseconds. This technique, called time-resolved serial femtosecond crystallography (TR-SFX), is particularly suited for studying structural dynamics of light-activated proteins.

The first part of this thesis is about production of channelrhodopsins. Channelrhodopsins are light-gated ion channels that we aim to produce for a future TR-SFX experiment. We describe protein production in the yeast *Pichia pastoris* and strategies to handle glycosylations of the proteins. We also establish protocols for purification and screen for microcrystals in the lipidic cubic phase.

The second and major part of this thesis is about structural dynamics of bacteriorhodopsin, a light-driven proton pump and by far the most studied microbial rhodopsin. We perform a TR-SFX study on bacteriorhodopsin that reveal structural changes in the time window from nanoseconds to milliseconds. In addition, we develop tools for comparison and analysis of structures and difference Fourier electron density maps from intermediate trapping studies and TR-SFX. These analyses give new insights in the mechanism of proton pumping.

Sammanfattning

Rhodopsiner är en familj av membranproteiner som finns i många olika organismer, från bakterier och alger till människor, och ger dem förmågan att uppfatta och reagera på ljus. När ljus absorberas av rhodopsinerna sker strukturella förändringar som leder till att protoner pumpas, att en kanal öppnas i proteinet eller att en signal skickas. Trots att rhodopsinerna har olika funktioner är de väldigt lika varandra i sin uppbyggnad. Detta gör proteinfamiljen intressant ur ett strukturbiologiskt perspektiv, eftersom strukturbiologi handlar om att koppla proteiners funktion till struktur och dynamik.

Strukturen hos ett protein kan bestämmas med hjälp av röntgenkristallografi. Det kräver ett rent proteinprov som kan bilda proteinkristaller, datainsamling med hjälp av röntgenstrålning och analys av data för strukturell modellering och tolkning. Under de senaste tio åren har en ny sorts röntgenkälla utvecklats, så kallade X-ray free electrons lasers (XFELs). En XFEL genererar mycket intensiva och korta röntgenpulser och det har gjort det möjligt att fånga strukturella förändringar i realtid med en tidsupplösning på femtosekunder. Denna teknik som kallas för tidsupplöst seriell femtosekunds-kristallografi (TR-SFX) är speciellt lämpad för att studera strukturdynamik hos ljusaktiverade proteiner.

Den första delen i denna avhandling handlar om att producera kanalrhodopsiner, ett slags ljus-styrda jonkanaler, för en framtida TR-SFX-studie. Vi utvecklar produktion av kanalrhodopsiner i jästen *Pichia pastoris* och protokoll för proteinrening med syfte att få stora mängder rent och homogent protein för kristallisering.

Den andra och största delen av avhandlingen handlar om strukturell dynamik hos bakterierhodopsin, en ljusdriven protonpump som har studerats under årtionden. Vi genomför en TR-SFX-studie på bakterierhodopsin som visar strukturella förändringar i ett tidsspann från nanosekunder till millisekunder. Vi utvecklar också nya verktyg för analys av strukturell information. Med hjälp av dessa verktyg kan vi tydligare än tidigare beskriva de strukturförändringar som sker när proteinet aktiveras av ljus. Detta ger nya insikter om den mekanism som gör att bakterierhodopsin pumpar protoner.

Publications

Paper I **Cecilia Wickstrand**, Giorgia Ortolani, Tristan Wiessalla, Ferran Riera Boronat, Richard Neutze, Kristina Hedfalk

"Production and purification of channelrhodopsins expressed in Pichia Pastoris"

Manuscript (2019)

Paper II **Cecilia Wickstrand**, Robert Dods, Antoine Royant, Richard Neutze

"Bacteriorhodopsin: Would the real structural intermediates please stand up?"

Biochimica et Biophysica Acta ***1850*** (3): 536-553 (2015)

Paper III Przemyslaw Nogly, Daniel James, Dingjie Wang, Thomas A. White, Nadia Zatsepin, Anastasya Shilova, Garrett Nelson, Haiguang Liu, Linda Johansson, Michael Heymann, Kathrin Jaeger, Markus Metz, **Cecilia Wickstrand**, Wenting Wu, Petra Båth, Peter Berntsen, Dominik Oberthuer, Valerie Panneels, Vadim Cherezov, Henry Chapman, Gebhard Schertler, Richard Neutze, John Spence, Isabel Moraes, Manfred Burghammer, Joerg Standfuss, Uwe Weierstall

"Lipidic cubic phase serial millisecond crystallography using synchrotron radiation"

Journal of the International Union of Crystallography **2** (2): 168-176 (2015)

Paper IV Eriko Nango, Antoine Royant, Minoru Kubo, Takanori Nakane, **Cecilia Wickstrand**, Tetsunari Kimura, Tomoyuki Tanaka, Kensuke Tono, Changyong Song, Rie Tanaka, Toshi Arima, Ayumi Yamashita, Jun Kobayashi, Toshiaki Hosaka, Eiichi Mizohata, Przemyslaw Nogly, Michihiro Sugahara, Daewoong Nam, Takashi Nomura, Tatsuro Shimamura, Dohyun Im, Takaaki Fujiwara, Yasuaki Yamanaka, Byeonghyun Jeon, Tomohiro Nishizawa, Kazumasa Oda, Masahiro Fukuda, Rebecka Andersson, Petra Båth, Robert Dods, Jan Davidsson, Shigeru Matsuoka, Satoshi Kawatake, Michio Murata, Osamu Nureki, Shigeki Owada, Takashi Kameshima, Takaki Hatsui, Yasumasa Joti, Gebhard Schertler, Makina Yabashi, Ana-Nicoleta Bondar, Jörg Standfuss, Richard Neutze, So Iwata
"A three-dimensional movie of structural changes in bacteriorhodopsin"
Science **354** (6319): 1552-1557 (2016)

Paper V **Cecilia Wickstrand**, Gergely Katona, Richard Neutze
"A tool for visualizing protein motions in time-resolved crystallography"
Manuscript (2019)

Paper VI **Cecilia Wickstrand**, Przemyslaw Nogly, Eriko Nango, So Iwata, Jörg Standfuss, Richard Neutze
"Bacteriorhodopsin: Structural insights revealed using X-Ray lasers and synchrotron radiation"
Annual Review of Biochemistry **88** (2019)
Published online ahead of print in June 2019

Additional publications

These publications I have co-authored but they are not part of the thesis.

Linda C. Johansson, David Arnlund, Gergely Katona, Thomas A. White, Anton Barty, Daniel P. DePonte, Robert L. Shoeman, **Cecilia Wickstrand**, Amit Sharma, Garth J. Williams, Andrew Aquila, Michael J. Bogan, Carl Caleman, Jan Davidsson, R Bruce Doak, Matthias Frank, Raimund Fromme, Lorenzo Galli, Ingo Grotjohann, Mark S. Hunter, Stephan Kassemeyer, Richard A. Kirian, Christopher Kupitz, Mengning Liang, Lukas Lomb, Erik Malmerberg, Andrew V. Martin, Marc Messerschmidt, Karol Nass, Lars Redecke, M Marvin Seibert, Jennie Sjöhamn, Jan Steinbrener, Francesco Stellato, Dingjie Wang, Weixaio Y. Wahlgren, Uwe Weierstall, Sebastian Westenhoff, Nadia A. Zatsepin, Sébastien Boutet, John C.H. Spence, Ilme Schlichting, Henry N. Chapman, Petra Fromme, Richard Neutze
"Structure of a photosynthetic reaction centre determined by serial femtosecond crystallography"
Nature Communications **4**: 2911 (2013)

David Arnlund, Linda C Johansson, **Cecilia Wickstrand**, Anton Barty, Garth J Williams, Erik Malmerberg, Jan Davidsson, Despina Milathianaki, Daniel P DePonte, Robert L Shoeman, Dingjie Wang, Daniel James, Gergely Katona, Sebastian Westenhoff, Thomas A White, Andrew Aquila, Sadia Bari, Peter Berntsen, Mike Bogan, Tim Brandt van Driel, R Bruce Doak, Kasper Skov Kjær, Matthias Frank, Raimund Fromme, Ingo Grotjohann, Robert Henning, Mark S Hunter, Richard A Kirian, Irina Kosheleva, Christopher Kupitz, Mengning Liang, Andrew V Martin, Martin Meedom Nielsen, Marc Messerschmidt, M Marvin Seibert, Jennie Sjöhamn, Francesco Stellato, Uwe Weierstall, Nadia A Zatsepin, John C H Spence, Petra Fromme, Ilme Schlichting, Sébastien Boutet, Gerrit Groenhof, Henry N Chapman, Richard Neutze
"Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser"
Nature Methods **11**: 923–926 (2014)

Erik Malmerberg, Petra H. M. Bovee-Geurts, Gergely Katona, Xavier Deupi, David Arnlund, **Cecilia Wickstrand**, Linda C. Johansson, Sebastian Westenhoff, Elena Nazarenko, Gebhard F. X. Schertler, Andreas Menzel, Willem J. de Grip, Richard Neutze
"Conformational activation of visual rhodopsin in native disc membranes"
Science Signalling **8** (367): ra26 (2015)

Robert Dods, Petra Båth, David Arnlund, Kenneth R. Beyerlein, Garrett Nelson, Mengling Liang, Rajiv Harimoorthy, Peter Berntsen, Erik Malmerberg, Linda Johansson, Rebecka Andersson, Robert Bosman, Sergio Carbajo, Elin Claesson, Chelsie E. Conrad, Peter Dahl, Greger Hammarin, Mark S. Hunter, Chufeng Li, Stella Lisova, Despina Milathianaki, Joseph Robinson, Cecilia Safari, Amit Sharma, Garth Williams, **Cecilia Wickstrand**, Oleksandr Yefanov, Jan Davidsson, Daniel P. DePonte, Anton Barty, Gisela Brändén, Richard Neutze
"From macrocrystals to microcrystals: A strategy for membrane protein serial crystallography"
Structure **25** (9):1461-1468 (2017)

Contribution

Paper I: I started the project and planned a majority of the experiments. I performed a large extent of the work in expression and purification and instructed students helping me. I was part of setting up crystal screens. I wrote the manuscript and made all figures.

Paper II: I developed the tools for comparison of structures and performed most of the analysis. I created almost all figures and was involved in writing the manuscript.

Paper III: I analysed the structures, performed the hierarchical clustering and draw the corresponding figure.

Paper IV: I was preparing samples for injection during the beamtime. I analysed structures and difference electron density maps and wrote scripts for linear decomposition. I draw most of the figures and contributed to writing of the manuscript.

Paper V: I developed the analysis tool for comparison of electron density maps and structures, made all figures and contributed in the writing process.

Paper VI: I made all figures and was involved in writing of the manuscript.

Abbreviations

bR	bacteriorhodopsin
C1C2	chimaera of ChR1 and ChR2
ChR	channelrhodopsin
CHS	cholesteryl hemisuccinate
DDM	n-dodecyl-β-D-maltoside
EM	electron microscopy
Endo H	endoglycosidase H
HEK-cells	human embryonic kidney cells
IMAC	immobilized metal ion affinity chromatography
kDa	kilo Dalton (mass unit for molecules)
LCP	lipidic cubic phase
MS	mass spectrometry
NMR	nuclear magnetic resonance
PDB	Protein Data Bank (online database for protein structures)
PNGase F	peptide N-glycosidase F
RMSD	root mean square deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SFX	serial femtosecond crystallography
SMX	serial millisecond crystallography
SVD	singular value decomposition
TEV	tobacco etch virus
TR-SFX	time-resolved serial femtosecond crystallography
TR-SMX	time-resolved serial millisecond crystallography
Å	Ångström (unit of distance, corresponding to 0.1 nm)

Contents

CHAPTER 1. INTRODUCTION

Every living cell is enclosed by a membrane that protects the cell from the surrounding environment. The membrane consists of two layers of fat-like molecules, called lipids, with polar head-groups facing the inside and outside of the membrane and with hydrophobic tails towards the middle (Figure 1). The membrane creates a selective barrier, that is permeable and semi-permeable to non-polar and smaller polar substances, such as water, but restricts ions and larger hydrophilic substances to pass.

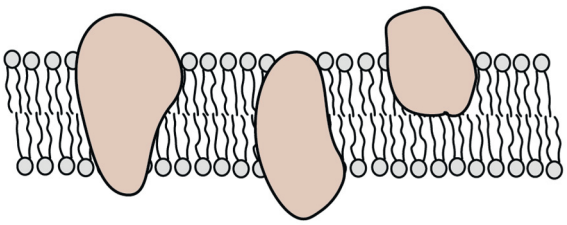


Figure 1. Schematic drawing illustrating a section of a membrane with the lipid bilayer, two integral membrane proteins and one peripheral membrane protein.

The membrane is also crowded by proteins, both integral membrane proteins spanning the whole membrane and peripheral proteins that are anchored to the membrane. They allow cells to communicate, transport nutrients and waste across the membrane and perform reactions. Even though membranes are a small fraction of the total cell volume it has been estimated that as much as 20-30% of most genomes encodes membrane proteins [1]. This reflects the importance of membrane proteins to the cell.

1.1. Transport proteins

Transport proteins are classified by the International Union for Biochemistry and Molecular Biology according to the type of transport they perform [2]. The largest three classes are involved in passive transport and primary or secondary active transport processes. Channels and pores are passive transporters that facilitate diffusion through the membrane along a concentration gradient. Channels are gated and open temporarily when receiving a chemical or physical stimulus. These stimuli include ligand binding, depolarization of the membrane, mechanical changes in the membrane or

light. Pumps are primary active transporters that transport substances against a concentration gradient by utilizing a primary energy source, such as hydrolysis of adenosine triphosphate (ATP) or absorption of light. Finally, secondary active transporters utilize proton or ionic gradients to drive the transport of a substance against a concentration gradient. Examples of this class are symporters, that transport the substrate in the same direction as the energizing co-transported substance, and antiporters, that perform transport and co-transport in opposite directions.

1.2. Rhodopsins

Rhodopsins are a superfamily of membrane proteins consisting of two groups; microbial (type I) and animal (type II) rhodopsins (reviewed in [3]). Despite their lack of sequence homology, the two groups of rhodopsins share the same architecture. They consist of seven transmembrane helices (here referred to as A-G) with the N-terminus facing the outside of the cell and the C-terminus on the cytoplasmic side (Figure 2). A light-absorbing chromophore, retinal, is covalently linked to a lysine residue in helix G, forming a Schiff base (SB). The retinal isomerizes upon light absorption, which initiates conformational changes in the protein once the protein is activated by light.

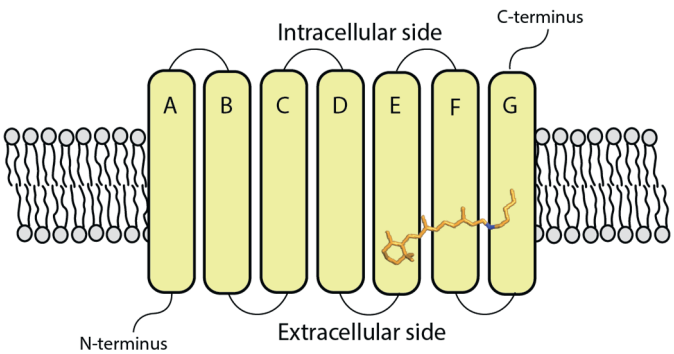


Figure 2. The overall structure of rhodopsins.

Animal rhodopsins are a class of G-protein coupled receptor proteins (GPCRs) found in animals, including for example visual rhodopsins that are expressed in the rods of the eye and are essential in vision of vertebrates [3].

Microbial rhodopsins are found in bacteria, archaea and lower eukaryotes (reviewed in [4]). Bacteriorhodopsin (bR), an outward-directed proton pump, was the first of this type to be identified. It was isolated from *Halobacterium salinarum* in 1971 [5] and is by far the most studied microbial rhodopsin. The inward-directed chloride pump halorhodopsin [6] and the signalling proteins sensory rhodopsin I and II [7,8] were later identified in the same organism.

Now almost half a century has passed and several thousand microbial rhodopsins are known to exist, including additional bacteriorhodopsins, halorhodopsins and sensory rhodopsins [4]. Around year 2000 two new classes, proteorhodopsins (PRs) and channelrhodopsins (ChRs), were identified. PRs are found in marine microorganisms and are the most abundant microbial rhodopsins (reviewed in [9]). They have been identified as proton pumps similar to bR but may also have other functions. It is debated whether these retinal-based proteins can provide an alternative system to photosynthesis [10] but so far only a few studies can directly link PRs to increased growth under light conditions [11]. Channelrhodopsins are light-gated ion channels found in algae (reviewed in [12,13]).

1.3. Bacteriorhodopsin

Source and function

bR was discovered in the cell membrane of *H. salinarum* [5], an halophilic archaea that live in environments with extreme salt concentrations [14]. bR is the only protein present in this membrane and responsible for its purple colour, which is the reason that the membrane itself is termed the purple membrane. Soon after discovery of the rhodopsin-like protein, bR was proposed to be an outward-directed proton pump that build up a proton gradient driving ATP synthesis and transport reactions in the archaea [15].

Structures

The first structural model of bR that was obtained with electron crystallography in 1975 to 7 Å resolution [16] was also the first model of an integral membrane protein. A high-resolution electron diffraction structure was published in 1990 [17] and since then 126 structures of bR have been deposited in the protein data bank (PDB) [18]. This set includes resting state structures at increasing resolution, obtained with different techniques and crystallization methods. It also includes models of different intermediate states, with structural changes induced by light, mutations, variations in pH

etc. Before 2015 all structures were collected at cryogenic temperature. In the last years studies using X-ray lasers have revolutionized the field by revealing time-resolved sequences of structural changes (**Paper IV** and [19]).

Photocycle and major steps in proton pumping

bR absorbs light at 570 nm due to the retinal chromophore that isomerizes from an *all-trans* to a *13-cis* conformation (Figure 3A). This initializes a series of proton transfer reactions leading to net transport of one proton from the cytoplasm to the extracellular space. bR changes its absorbance during the reaction cycle and its intermediate states – termed K, L, M, N and O – can be followed spectrophotometrically [20,21]. The so called photocycle of bR (Figure 3B) is now well established.

The intermediate structures in combination with spectral data and mutational studies have revealed the major structural changes and key residues in proton pumping (Figure 3C). In the K-state, that is reached within picoseconds after light-activation, retinal has completed its isomerization. Disordering of an adjacent water molecule interrupts a hydrogen bonding network connecting the SB and the primary proton acceptor Asp85. This leads to structural changes close to the retinal and formation of the L-state within microseconds. The proton affinity of the SB is reduced and the affinity of Asp85 is increased. After approximately 40 μs the proton transfer (Figure 3C, step 1) occurs in the L-to-M transition. This is followed by release of another proton from a group of residues including Glu194 and Glu204 to the extracellular side (step 2). The early M-state (referred to as M_1) initiates rearrangements on the cytoplasmic side of the protein, leading to the late M state (M_2). The final three intermediates, M_2, N and O, last for about 5 ms each. The M_2-N transition corresponds to reprotonation of the SB from Asp96 (step 3). The latter residue is in turn reprotonated (step 4). In the N-O transition the retinal isomerizes back to its *all-trans* conformation. Finally, Asp85 is deprotonated and the proton is transferred to the proton release group at the extracellular surface (step 5). This completes the photocycle of bR [21,22].

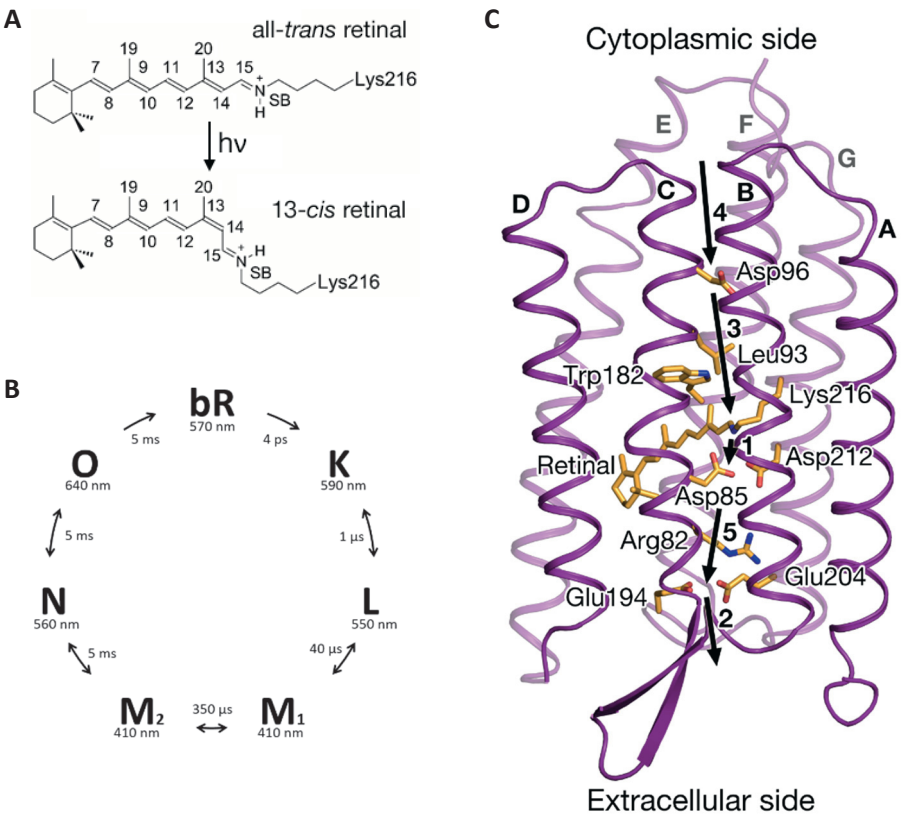


Figure 3. Intermediates and major steps in the photocycle of bR.
A) Isomerization of retinal from the *all-trans* to the *13-cis* conformation.
B) Photocycle of bR.
C) Structure of bR showing key residues. Arrows indicate steps in proton transfer.

1.4. Channelrhodopsins

Source and function

The first two ChRs to be discovered were the cation channels ChR1 [23] and ChR2 [24] from *Chlamydomonas reinhardtii*. Both the proteins have been shown to permeate Na^+, K^+ and Ca^{2+}, even though ChR1 was initially believed to be a proton-specific channel [25]. *C. reinhardtii* is a unicellular and motile alga belonging to the chlorophyta, that are often referred to as green algae. It uses two flagella to swim relative to light in a process called phototaxis, that allows the alga to find optimal conditions for photosynthesis [26]. ChR1 and ChR2 are involved in regulation of this process [23,24]. Many chlorophyte ChRs have now been identified as well as two new groups of ChRs; anion ChRs and cryptophyte cation ChRs. These groups are found in cryptophyte algae (or

cryptomonads), which are another linage of unicellular algae that may be blue-green, red or brown [27]. Anion ChRs are closely related to their chlorophyte homologs but are permeable to anions [4]. The cryptophyte cation ChRs transport cations but are more similar to bacteriorhodopsin than to other ChRs [28].

Optogenetics and engineered ChRs

The discovery of ChRs made it possible to trigger depolarization of membranes with light [29]. This started a new field in neuroscience called optogenetics and it is the reason why ChRs has gained remarkable attention in the last 14 years [30]. In optogenetics, ChRs and similar proteins are expressed in neuronal cell cultures or in living organisms such as drosophila [31,32], rodents [33] and non-human primates [34], for the study of everything from regulation of firing on the cellular level to behaviour, learning and complex neuronal networks [35]. There have been lots of efforts to develop new variants of ChRs, especially focused on ChR2, that widen the spectrum of possible applications in neuroscience. There are now variants with for example faster kinetics, increased photocurrents, improved light-sensitivity, changed ion selectivity and shifted absorption spectra. There are also step-function channelrhodopsins that have a slow photocycle where closing can be controlled by illumination at a different wavelength [13]. The work on engineered ChR variants has contributed to the understanding of how ChRs function and which residues that are important for channel gating, kinetics etc.

Photocycle

The photocycle of ChR2 is more complex than the photocycle of bR. Several models have been suggested to explain the order of intermediate states [13,36]. The models consist of at least two branched cycles, each with unique resting and open states. The channel is opening within 100 µs after excitation and closing within 100 ms. However, the total length of the photocycle until full recovery of the resting state is in the range of 20 s [37].

Structures

There are nine structures of ChRs in the Protein Data Bank [18]. Four of them are related to *C. reinhardtii*; two models of a chimera of ChR1 and ChR2 (called C1C2) [38,39] and two models of ChR2 [40]. The rest of the models include one of a ChR called Chrimson from *Chlamydomonas noctigama* [41], one of a natural anion ChR [42] and three of designed anion channels [43].

ChR1 and ChR2 consist of 712 and 737 amino acids, each divided into two domains. The N-terminal channel domain consists of approximately 300 residues and is functional on its own [23,24]. In most studies the protein is truncated to only include the channel domain. The C-terminal intracellular domain may be involved in localization or signalling in the alga [44].

The crystal structures of C1C2 (consisting of helices A-E from ChR1 and helices F-G from ChR2) and wild-type ChR2, reveals high similarity to bR (Figure 4). There are three main differences; a prolonged N-terminal part that contributes to dimer formation, an extension of helix G on the cytoplasmic side and a larger cavity on the extracellular side due to outward tilting of helices A and B. In C1C2 there is an electronegative cavity between helices A, B, C and G from the SB towards the extracellular side. On the cytoplasmic side the path for cations is closed by two constriction sites [38]. The central pore in the wildtype structure of ChR2 is more blocked than in C1C2, due to three constriction sites referred to as the intracellular, central and extracellular gates [40].

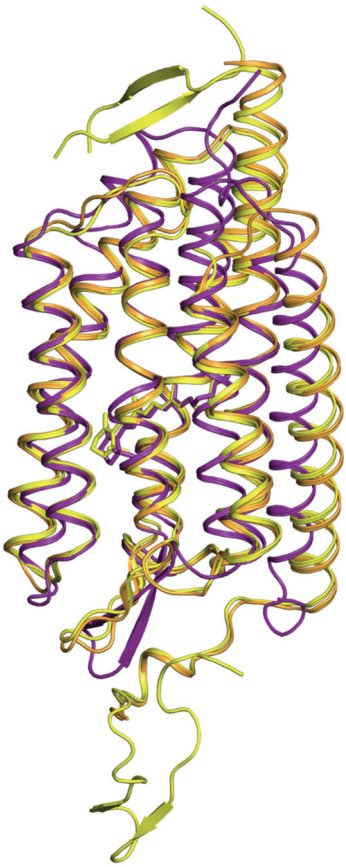


Figure 4. Overlay of C1C2 (yellow) [38], ChR2 (orange) [40] and bR (purple) [Paper IV].

ChR2 has been shown to also be a leaky proton pump [45] and as in bR the isomerization of the retinal initiates a set of proton transfer reactions. The crystal structures of C1C2 and ChR2 [38,40] provide insight about the channel and potential residues involved in proton transfer. Details and how opening and closing of gates are coordinated in space and time require further investigation.

1.5. Scope of this thesis

This thesis is focusing on two microbial rhodopsins. The first part is about channelrhodopsins from *C. reinhardtii* (**Paper I**) and the second and main part is on bacteriorhodopsin from *H. salinarum* (**Paper II-VI**).

Part 1 – ChR

Paper I is describing production of ChRs in the yeast *Pichia pastoris* aiming at a pure and homogenous sample for microcrystallization. It covers construct design, over-production in fermentors, solubilization, purification, analysis of protein quality and initial crystallization. The recombinant proteins are glycosylated, which is an obstacle in crystallography. Deglycosylation assays and design of mutants are applied to produce a homogenous protein.

Part 2 – bR

Paper II is an analytical review about bR, focusing on all the structures that had been published until 2014. New tools based on internal distance matrices and hierarchal clustering are developed for objective comparison of structures. Deposited X-ray diffraction datasets are also reanalyzed for calculation and comparison of difference electron density maps. The internal distance matrix tool is used in **Paper III** to compare a room temperature bR structure from serial millisecond crystallography (SMX) to earlier structures from experiments performed at cryogenic temperature.

Paper IV presents the results of the first time-resolved serial femtosecond crystallography (TR-SFX) experiment on bR**.** We create a three-dimensional movie of structural changes from nanoseconds to microseconds revealing new mechanistic details in proton pumping.

Paper V describes a new tool for analyzing sets of difference electron density maps. Two time-resolved data sets on bR are analyzed, in the range from nanoseconds to milliseconds (**Paper IV**) and on the sub-picosecond time scale [19]. This tool provides additional insight into how conformational changes are correlated with time, and reveals ultrafast dynamics within the first picosecond of photoactivation of bR.

Paper VI is a review on the mechanism of proton pumping in bR. In the paper, recent TR-SFX studies are compared to previous intermediate trapping studies at synchrotrons. The focus is on new insights from time-resolved diffraction studies at X-ray free electron lasers (XFELs).

CHAPTER 2. METHODS IN STRUCTURAL BIOLOGY

The main objective in structural biology is to model proteins and other biological macromolecules and link the structural information to function and dynamics. Rhodopsins are popular as model proteins in this field for two main reasons. Firstly, the proteins are activated with light and this makes it easier to study the dynamics of the proteins in action. Secondly, rhodopsins have different functions and thus serve as models for several types of membrane proteins. Since the rhodopsins are structurally similar this makes it interesting to study how small differences in structure affect function.

Protein models may be obtained by X-ray crystallography, nuclear magnetic resonance (NMR) or electron microscopy (EM) methods. Currently 90% of the >138 000 protein structures published in Protein Data Bank (PDB) are X-ray structures, 8% origin from NMR and 1.5% from EM [18]. All three techniques are under constant development and the border of what is possible is extending. In X-ray crystallography, a relatively new methodology is time-resolved serial femtosecond crystallography (TR-SFX) that allows studies of protein dynamics at room-temperature. This chapter will describe the process going from a gene or protein of interest to an X-ray structure (Figure 5), with a special focus on membrane proteins, time-resolved studies and techniques that have been used in this thesis.

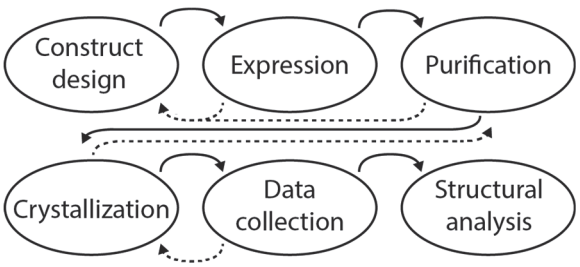


Figure 5. Overview of the process towards an X-ray crystallography structure. The process is often non-linear; dashed arrows indicate parts that are commonly iterated.

The initial task of a structural study is to achieve relatively large amounts of functional protein for structural analysis. The protein may in some cases be purified from its natural source, but this requires a high expression level and that the original organism can be either obtained in large quantities or produced inside a lab. If not, and that is often the case, the protein has to be

produced through heterologous expression in a host organism. In this thesis, bR is obtained from the natural source and ChRs are produced through heterologous expression in *Pichia pastoris*.

2.1. Codon optimization and construct design

Most projects in structural biology start with a gene that encodes the protein of interest. Before expression in a host organism of choice, the gene may be codon optimized. In addition, the gene can be modified in many ways, which is called construct design.

The original DNA sequence is adapted to the source organism through evolution and even though the genetic code is almost universal the frequency of codons (and the frequency of the corresponding tRNA anticodons) differ between organisms. When a gene is codon optimized rare codons of the host are exchanged to preferred ones in order to speed up the translation from mRNA to protein. Often codon optimization also involves adjusting the GC-content of the DNA and avoiding repeats in the mRNA, that could cause loop and hairpin formations [46]. Overall the goal is to stabilize the mRNA template and aid the translation process, which in theory should increase the chance to reach high expression levels of the desired protein.

There are many options in construct design and often several constructs have to be evaluated throughout a project. The gene may be truncated and mutations can be introduced to make the protein more stable or for the removal of glycosylation sites. The next option is to add sequences to the gene, that are encoding for purification tags and fusion proteins to aid protein purification or analysis. It is also possible to include cleavage sites for the removal of tags and fusion proteins during the purification process. When the gene has been designed it is commonly ordered from a gene synthesis company and subcloned into a plasmid that is suitable for expression in the host organism of choice.

2.2. Expression

Expression is a complex process that for membrane proteins include transcription, translation, insertion into the membrane, folding and (in eukaryotic systems) post-translational modifications. Similarity between the source organism and the host increases the chance of a properly folded and

functional protein in high yields. This often makes *E. coli* a standard choice for bacterial proteins, but for eukaryotic proteins there are many options. Yeasts such as *P. pastoris* and *Saccharomyces cerevisiae* are cheap and easy to cultivate in large scale. It is also fast to produce over-expressing strains. Insect cells and mammalian cells are often considered for human proteins. They require expensive complex media and it takes longer time to get stable cell lines, but transient expression methods may be an alternative. The choice of a host is also about personal preferences and trial and error, since it is not possible to predict the outcome [47,48].

P. pastoris is a well-established expression system that has been successful for the production of many membrane proteins [49]. It is a methylotrophic yeast that can use methanol in low concentrations as a carbon and energy source. This capability is due to two alcohol oxidases, AOX1 and AOX2. The promotor of AOX1 is very strong and inducible with methanol, which is utilized for high level expression of heterologous proteins [50]. *P. pastoris* is transformed with a plasmid that contains the gene of interest, the promotor and a selection marker, for example an antibiotic resistance gene. The DNA is stably incorporated into the genome by homologous recombination, occasionally in multiple copies. Multiple insertion may increase expression levels and thus, screening of high-expressing clones can improve protein yields [51]. *P. pastoris* can be grown in shake flask cultures or in fermentors/bioreactors for protein production. Initially, cells are grown on glycerol to high cell densities before expression is induced with methanol.

2.3. Solubilization and purification

Crystallization of a protein typically requires a very homogenous and pure protein solution, free from other proteins, as well as a stable protein that does not aggregate at high concentrations. Since integral membrane proteins are embedded in the membrane they have to be extracted from the cell membrane into solution before the purification can start. This involves breaking the cells, separating the membrane from the rest of the cell content and solubilisation of the protein with detergents.

Detergents are water-soluble surfactants with hydrophilic head-groups and hydrophobic tails. Above a critical concentration (termed the critical micelle concentration, CMC) the monomers aggregate into micelles with their tails facing the inside, forming a hydrophobic interior. Micelles can protect the

hydrophobic parts of a membrane protein in a similar way as the lipid bilayer and bring the protein into solution. When solubilizing membrane proteins, a detergent of choice is slowly added. Initially, the detergent is incorporated into to the lipid bilayer but as the concentration increases the membrane is eventually disrupted and protein-lipid-detergent complexes are formed. At higher concentration more of the lipids are replaced and proteins may be in micelles essentially without lipids [52].

There are a variety of detergents to choose from, with different shapes and charges. Often 10-20 detergents are screened for efficiency towards the target protein in order to optimize the solubilisation step with respect to yield and protein stability. The detergent should be efficient but mild so that the protein is kept in its native fold when extracted from the membrane. This requires the detergent to be long enough to cover hydrophobic parts of the protein. However, for subsequent crystallization it should not be too long, since a large detergent could cover hydrophilic parts of the protein and prevent crystal contacts [53]. It is usually possible to change detergent during purification if the optimal detergent for solubilisation is not the best choice for crystallization.

After solubilisation the protein of interest is separated from other proteins by chromatography. Depending on the chromatographic method of choice proteins are separated by affinity, charge or size. Recombinant proteins are often designed to have a purification tag with affinity towards a certain ligand. A frequently used tag is the polyhistidine tag, commonly six to eight residues long, that allows the protein to bind to Ni^{2+} or Co^{2+} in so called immobilized metal ion affinity chromatography (IMAC).

In the purification of ChRs we have used an IMAC-resin consisting of nickel immobilized to agarose beads. The purification consists of three steps; binding, wash and elution. First, the protein mixture is incubated with the resin for a given time, from one hour up to over night, to allow the tagged protein to bind. Then the resin is washed several times to reduce the amount of non-specifically bound proteins. Finally, the tagged protein is eluted with a buffer containing a high concentration of imidazole that compete for the binding to the Ni^{2+} ions and causes release of the protein to be purified. Normally, if the production level of the target protein is rather high, the protein is relatively pure after IMAC purification, since all natural proteins of the host lack the synthetic tag.

Size exclusion chromatography (SEC) or gel filtration is often used before crystallization to further increase purity and homogeneity of the sample. It is a method for separation of proteins by size. The gel filtration column contains a porous resin that let smaller proteins interact with the matrix whereas large proteins are excluded from the pores and move faster through the column. The chromatogram shows if the sample is pure and homogenous and thus SEC is also a method for quality control and size estimation of the protein before crystallization.

2.4. Crystallization

Crystals

A crystal is a solid consisting of atoms, ions or molecules organized in a regular pattern. The pattern is built of small identical units, referred to as unit cells, that are repeated by translational symmetry in three dimensions. Thus, the unit cell is a building block representing the whole crystal. It forms a grid called the crystal lattice, which is defined by the lengths of the individual axes (a, b, c) and the angles (α, β, γ) between them.

The unit cell in turn may consist of either a single atom/ion/molecule or several that are symmetrically organized within the cell. There is a finite number of variants of crystal packings that can occur. These 230 variants are called space groups and are defined by the arrangement of unit cells relative to one other and the type of symmetry within the unit cell, including rotation, translation and mirroring. Proteins are chiral molecules and since these cannot have mirror symmetry the number of space groups available to proteins are reduced to 65. The space groups are denoted with a capital letter that represent the type of lattice and a number that describes the combination of symmetry operations [54]. As an example, the most common space group for bR is $P6_3$. It has a hexagonal lattice and the protein molecules are organized within the unit cell by a so called screw axis, which is a combination of translational and rotational symmetry operations.

Crystal formation

A protein crystal is held together by weak (non-covalent) interactions. A large extent of the crystal consists of solvent, and the crystal contacts that form the interface between neighbouring protein molecules involve only a limited part of the protein surface. This makes the protein crystal more fragile than for

example a densely packed salt crystal. It also makes the process of crystal formation more complex.

Most proteins crystallize under very specific conditions and thus extensive crystal screening procedures are required to find these conditions. In crystal screening, a number of parameters are varied in a controlled manner, including the concentration of the protein, the buffer compositions, pH, temperature etc. A precipitant such as a salt or polyethylene glycol (PEG) is often added in a range of concentrations to aid crystal formation. A great help in the screening process is pre-made commercial screens that contain various concentrations of buffers, precipitants and other additives.

One of the most popular crystallization methods is vapour diffusion. A small amount of the concentrated protein solution is mixed with a solution that contains one or more precipitants. The protein-precipitant droplet is then enclosed in a chamber with a reservoir containing the precipitant solution. The crystal formation process can be described in a phase diagram (Figure 6). Initially, water will evaporate from the droplet since the concentration of the precipitant is higher in the reservoir (arrow 1). Eventually this leads to oversaturation of the protein and under perfect conditions, cores of protein crystals (nuclei) start forming. The so called nucleation reduces the concentration of free protein in the droplet (arrow 2) and this eventually prevents formation of more nuclei. If the protein concentration reaches the metastable zone in the phase diagram, the existent nuclei continue growing and crystals are formed. If any part of this process goes wrong, the drops remain unaffected or the protein precipitates.

Crystallization in lipidic cubic phase

Membrane proteins may sometimes be crystallized with vapour diffusion methods directly after solubilisation and purification. An alternative is crystallization in lipidic cubic phase (LCP). In this method the protein solution is mixed with a lipid (often monoolein) to give a homogenous, transparent and viscous phase. The LCP creates an environment that resembles the membrane and may aid formation of crystals. A popular method for crystallization in LCP is batch crystallization, when the LCP is overlayed with a precipitant solution and then sealed in a plate or tube for crystals to form.

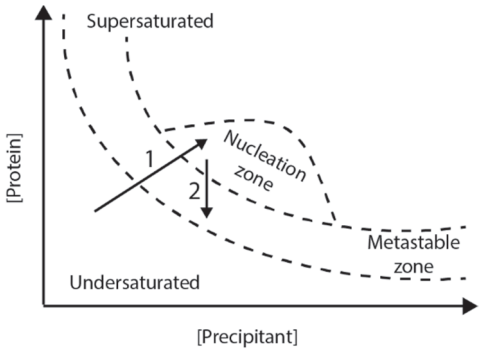


Figure 6. A crystallization phase diagram. The arrows indicate a path for formation of large crystals in a vapour diffusion experiment.

2.5. X-ray diffraction and data collection

X-ray diffraction

Proteins are in the order of 10 nm (or 100 Å) in size. This makes it impossible to study proteins in normal microscopes. Instead X-rays may be used, that have a much shorter wavelength than visual light (approx. 0.01-0.1 nm) and thus can give atomic resolution. When X-rays hit a molecule the electromagnetic waves are scattered by the electrons in the sample. The scattering of a single molecule is very weak, but a crystal with regularly organized molecules increases the signal.

The scattering from a crystal gives a unique diffraction pattern with distinct spots, so called reflections. The reflections occur when X-rays hit different unit cells and scatter in phase. Bragg's law (Equation 1) describes the geometric requirements for this.

$$n \cdot \lambda = 2d \sin\theta \qquad (1)$$

It states that when parallel X-rays are scattered from equal planes separated with a distance *d,* constructive interference occurs at the angle θ only if the difference in path length (*2d sin* θ, as shown in Figure 7) between the X-rays correspond to an integer *n* times the wavelength λ. Rotation of a crystal relative to the beam changes which planes that give reflections.

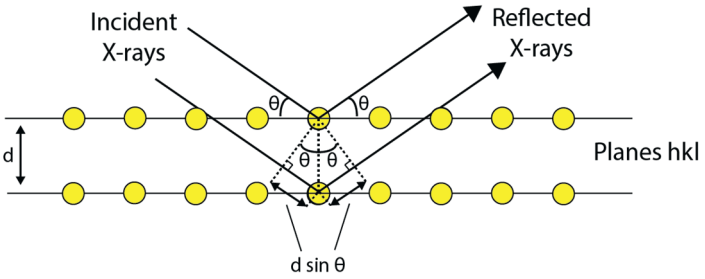


Figure 7. The geometry described in Bragg's law.

Data collection from a single crystal

In a traditional X-ray experiment (Figure 8A) a single crystal is placed in the X-ray beam and diffraction patterns are collected with an X-ray detector. During data collection the crystal is rotated to allow collection of as much information as possible. However, repeated exposure to X-rays will damage the crystal and cooling to cryogenic temperature is necessary to reduce these effects. This type of experiment is normally performed at synchrotrons or in a research lab equipped with a smaller, less powerful X-ray source.

Serial crystallography

In a serial crystallography experiment (Figure 8B) thousands of microcrystals are continuously injected across a highly focused X-ray beam. Thus, every diffraction pattern is the result of a single shot on a fresh crystal in a random orientation. When performed at an XFEL, the intense X-rays will entirely destroy the crystal but since it diffracts before it is destroyed [55] the data will be free of X-ray damage. At synchrotron sources the crystal will also suffer damage, but the source is less intense and less focused than at an XFEL and will not necessarily destroy the crystal. In both cases, the data is collected with no or little radiation damage and can thus be performed at room temperature. Around 10 000 diffraction images have to be collected for a complete data set.

Time-resolved serial crystallography is a method to study protein dynamics. It is mainly used for naturally light-driven proteins, but there are efforts to adapt the technique for other types of systems. The set-up is identical to a serial crystallography experiment, except that a laser pulse is applied to trigger conformational changes just before the X-ray pulse. Changes in the

timing between the pump laser and the probe X-ray pulse allow collection of structural information from different time-points in the photocycle.

Synchrotron based time-resolved serial crystallography is called serial millisecond crystallography (SMX) since the time-resolution at a synchrotron is a few milliseconds. XFELs provide shorter and more intense X-ray pulses that give a time-resolution in the femtosecond range. Thus, when experiments are performed at XFELs the methodology is called serial femtosecond crystallography (SFX).

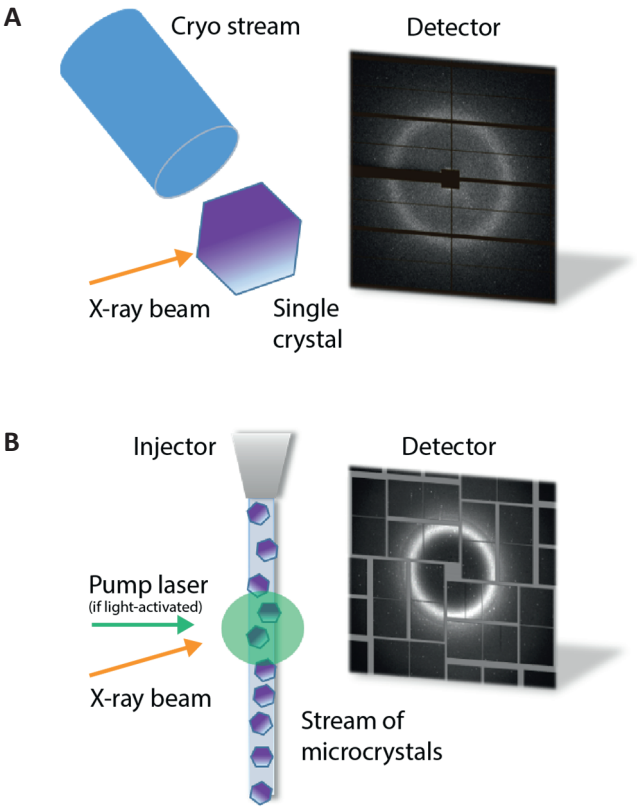


Figure 8. Comparison of single-crystal and serial crystallography set-ups.
A) A traditional single crystal experiment.
B) A (time-resolved) serial crystallography experiment.

2.6. Structure determination

A single data set contains thousands of diffraction patterns. The positions of the reflections in each image are determined by the crystal lattice, the space group and the rotation of the crystal. The amplitudes and phases of the reflections depend solely on the molecule itself. Processing of the diffraction data consists of a number of steps:

1. Determination of unit cell parameters
2. Indexing and integration of reflections
3. Scaling and merging of intensities
4. Phasing
5. Calculation of an electron density map
6. Structural modelling and refinement

The first three steps are performed to identify all reflections and measure their amplitudes. The next step, phasing, is more complicated since phases cannot be measured. There are several methods to solve this so called "phase problem". Molecular replacement is the most straight-forward method but it requires that a model of a related protein exists. Because the target protein will be structurally similar to the related model it will also have similar phases. In molecular replacement the related structure is placed within the unit cell and the optimal translation and rotation of the molecule is found. Then phases are calculated from the model to be used as a first estimation of phases.

Calculations of amplitudes and phases give wave functions, that are referred to as structure factors. The structure factor F_{hkl}, with amplitude $|F_{hkl}|$ and phase α_{akl}, represents the wave that caused the reflection with index *hkl*. Structure factors are used in the calculation of electron density maps. The electron density ρ at a position a, b, c in the unit cell with volume V is obtained by Fourier transform of the structure factors (Equation 2).

$$\rho(a,b,c) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} |F_{hkl}|\, e^{-2\pi i(ha+kb+lc)+i\alpha_{hkl}} \qquad (2)$$

Once the Fourier electron density map has been obtained, a model of the protein can be built within the electron density. The initial model gives new phases that can be used for calculation of a slightly better electron density map and a new round of modelling. Repeated cycles of refinement may lead

to further improvement of the model. During the refinement process, the R-factors R_{work} and R_{free} are calculated (Equation 3), that measure the difference between observed amplitudes and amplitudes calculated based on the model.

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|} \quad (3)$$

R_{work} is calculated on the data that is used for refinement. R_{free} is instead calculated on a small fraction of the data that is excluded in refinement and reveals if the data is over-fitted. If both the parameters decrease during refinement it is a good indication that the process goes in the right direction. The final model has to be examined so that torsion angles, bond lengths etc. make physical sense [54].

After structure determination it is time to analyse the new model. This may for example include structural comparison with previous models or identification of new interactions or conformations that may be critical in understanding the dynamics or function of the protein. This analysis is the final goal of a structural biology project.

CHAPTER 3. CHANNELRHODOPSIN

The main goal of the channelrhodopsin project is to study the structural dynamics of an ion-channel with TR-SFX at an XFEL. This requires relatively large amounts of protein for micro-crystallization, at least 30 mg, and thus the initial task was to set up reliable methods for high level production and crystallization. When the project started the first ChR crystal structure of the chimaera C1C2 had been published by Kato *et al.* [38]. We aimed to build on these achievements and recover microcrystals by essentially following their protocol. However, the crystal structure of C1C2 was from protein produced in insect cells and since we wanted to have a large quantity of protein and had good experience of protein production in yeast, we instead decided to express C1C2, and eventually also ChR2, in *Pichia pastoris*. When both the proteins were shown to be glycosylated in the host we evaluated enzymatic removal of glycosylations and expression of non-glycosylated mutants that could give a more homogenous sample for crystallization. This work is described in **Paper I**. The project is still ongoing and we have just identified leads from initial crystallization screens.

3.1. Production and purification of ChRs for crystallization (Paper I)

The focus of **Paper I** is to achieve large amounts of homogenous and functional protein for crystallization. Handling and understanding of glycosylations in ChRs was part of the work required towards reaching this goal. This section starts with some theory about glycosylations before results from production, purification and analysis of ChRs are discussed.

N-glycosylations

Glycosylations are oligosaccharides that are attached to membrane proteins and secretory proteins in eukaryotes. N-linked glycosylations (or N-glycosylations) are linked to the side-chain nitrogen of asparagine residues that occur in specific amino acid sequences; the glycosylations motifs NXT or NXS, where X is any amino acid except proline. The N-glycosylation precursor molecule is attached to the nascent protein in the endoplasmatic reticulum (ER) and further processed throughout the ER-golgi pathway. The processing gives three different types of glycosylations: high-mannose, hybrid and

complex glycans [56]. *P. pastoris* is only capable of producing N-glycosylations of the high-mannose type [57].

Since the N-glycosylation process is complex and irregular in its nature, it causes heterogeneity within a protein population. Both the number of potential glycosylation sites that are glycosylated and the structure of each individual N-glycan may differ from molecule to molecule. In addition, oligosaccharides are large and flexible molecules. This makes glycoproteins problematic in X-ray based structural projects since it reduces the chance of crystal formation. There are several options to tackle the glycosylation problem, including introduction of mutations that remove N-glycosylation sites, expression in the presence of N-glycosylation inhibitors and enzymatic deglycosylation.

Enzymatic removal of N-glycosylations may be performed with peptide N-glycosidase F (PNGase F) or endoglycosidase H (Endo H). PNGase F removes the complete N-glycosylation of all types of N-glycosylations. Endo H cleaves off most of the N-glycosylation but leaves the innermost N-acetylglucose-amine residue. This enzyme requires the N-glycosylations to be of the high-mannose or hybrid type [58]. The cheaper variant Endo H_f (NEB) has the same functionality as Endo H as it is a recombinant fusion of Endo H and maltose binding protein [59].

Overview of the production and purification of ChRs

In **Paper I**, we express C1C2, ChR2 and mutants of these in *P. pastoris*. All the constructs are similar, with a C-terminal truncation after residue 309 in ChR2, a TEV cleavage site and a His_8-tag. The production was started in a small scale to show that the proteins were expressed before up-scaling to optimized growth in 2L fermentor-cultures. This scale gives enough material for purification, analysis of protein quality and initial crystal screening.

Our purification protocol is partially based on the protocol previously developed by Kato *et al.* [38], which includes solubilization in 2.5% n-dodecyl-β-D-maltoside (DDM) and 0.5% cholesteryl hemisuccinate (CHS) followed by purification using IMAC and gel filtration. However, a few adaptions of the protocol were necessary to optimize the purification of the yeast-derived protein:

- Addition of membrane washing steps to improve efficiency of the IMAC purification.

- Incubation with retinal before solubilization to increase the binding of retinal, as verified by an increase of the specific absorbance at 470 nm (at least for C1C2).
- Optimization of the solubilization of ChR2, since it was not efficiently solubilized in DDM/CHS. A solubilization screen indicated that this was mainly due to the buffer composition. Therefore, the initial Tris-buffer was replaced with a potassium phosphate buffer.

The purity and homogeneity of the proteins were analyzed with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The quality (or potential functionality) of the proteins was assessed as the ratio between the absorbance at 280 and 470 nm and measured with spectrophotometry.

Heterogeneity and glycosylation of C1C2 and ChR2

The samples of C1C2 and ChR2 appear pure on SDS-PAGE after the IMAC purification (Figure 9AB, lane 1). However, both the proteins give two separate bands on the gels, which indicated that the proteins could be glycosylated. Treatment with the deglycosylating enzymes Endo H_f and PNGase F (Figure 9AB, lanes 2-3) leads to shift of the bands and confirms that the proteins are N-glycosylated.

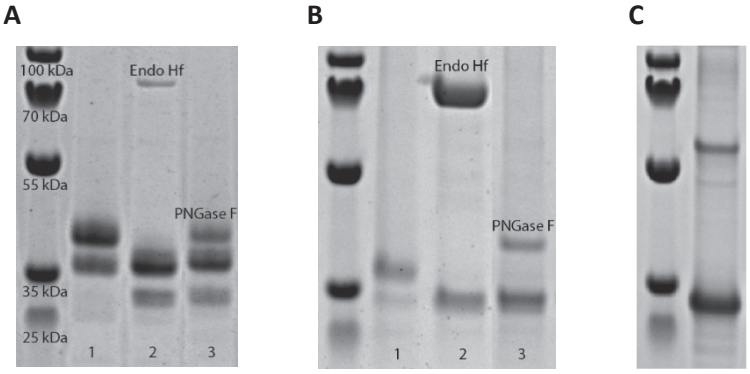


Figure 9. SDS-PAGE on samples purified using IMAC purification, before (lane 1) and after treatment with Endo H_f (lane 2) or PNGase F (lane 3).
A) C1C2. B) ChR2. C) ChR2-N24Q (untreated).

C1C2 exists in two forms also after deglycosylation, even if the amounts of enzymes or incubation times are increased. Initially, this was interpreted to be due to incomplete deglycosylation, which may occur if some glycosylation sites in the native protein are inaccessible to the enzymes. It is not possible

to fully determine from SDS-PAGE which apparent molecular weight that correspond to the non-glycosylated, full-length protein, since membrane proteins may migrate differently than a standard soluble protein due to anomalous binding of SDS [60]. This complicates the interpretation of the results from SDS-PAGE. The two bands of the untreated protein were excised from the gel, treated with trypsin and analyzed with liquid chromatography coupled to mass spectrometry (LC-MS). The mass corresponding to a peptide with residue numbers 2-45 was detected from the higher band but not from the lower band. This indicates that the band with a lower apparent molecular weight may have a shorter N-terminus. According to Uniprot [61] ChR1 has a N-terminal signal peptide of 22 residues. Thus, a theory for explaining the heterogeneity of C1C2 could be partial cleavage of the signal peptide in *P. pastoris*. It is also possible that some other type of post-translational modification differs between the two bands. One method to investigate this would be to see if homogeneity can be achieved through expression of an N-terminally truncated construct.

ChR2 has no natural signal peptide, according to Uniprot [61], and deglycosylation of this protein gives a single band for both the enzymes. The successful deglycosylation results in a homogenous protein for crystallization, but it also adds an extra step in the purification protocol.

Design and evaluation of glycosylation mutants

To overcome the limitations of glycosylations we also evaluated a number of C1C2 and ChR2 mutants, where the potential glycosylation sites (four in C1C2 and two in ChR2) were removed by amino acid substitution.

Two mutants of C1C2 were expressed and purified in larger scale: the quadruple-mutant C1C2-N61Q/N82Q/N153Q/N176Q and the triple-mutant C1C2-N61Q/T155A/N176Q. Analysis with SDS-PAGE showed that both the mutant constructs yielded two bands at the same apparent molecular weights as seen for enzymatically treated C1C2. Since the quadruple-mutant lack all N-glycosylation motifs, this analysis confirms that the deglycosylation of C1C2 is complete and that the heterogeneity has other reasons (as discussed above). The purified samples of the C1C2 mutants appeared uncolored and analysis with spectrophotometry confirmed complete loss of the specific absorbance. Thus, the efforts to produce mutant forms of C1C2 were unsuccessful.

Two mutants of ChR2 were designed: ChR2-N24Q/N137Q and ChR2-N24Q. In a small-scale analysis of expression levels by Western blotting procedures it was shown that both mutants yielded a single band with the same apparent molecular weight as seen for deglycosylated ChR2. This implied that a single mutation was enough to give a non-glycosylated protein and, consequently, only ChR2-N24Q was further evaluated. Expression and purification of ChR2-N24Q gives a non-glycosylated protein that appears homogenous (Figure 9C) and has a strong yellow colour (Figure 10). Thus, the evaluation of ChR2 mutants was successful and ChR2-N24Q became the main candidate for crystallization.

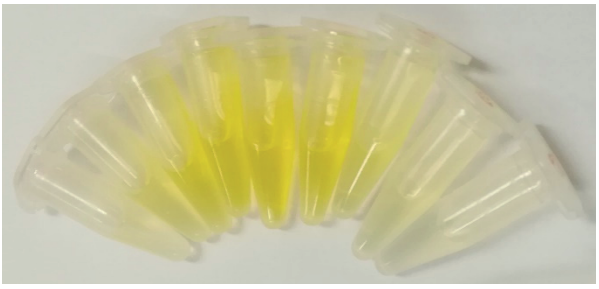


Figure 10. The colour of ChR2-N24Q indicates that the protein is functional. This is the elution profile from IMAC purification.

Absorbance and yields of C1C2, ChR2 and ChR2-N24Q

All-trans retinal was supplemented in the cultivation medium so that it could bind to the opsins during expression. Despite this, the A_{280}/A_{470}-ratio of C1C2 was high in the first purifications, indicating incomplete binding of retinal. Incubation of homogenized membranes with retinal before solubilization decreases the ratio from 3.6 to 2.3 (Figure 11A), as measured after the IMAC-purification step. This shows that C1C2 is able to bind retinal in the folded state.

The retinal incubation step was also included in every purification of ChR2 and ChR2-N24Q. Regardless of this, the proteins give an A_{280}/A_{470}-ratio close to 4 after the IMAC-purification step (as shown for ChR2 in Figure 11A). Whether this is due to limited affinity for the retinal or partial misfolding, it appears to have some effect on the size or shape of the proteins, since gel filtration partially separates the absorbing and non-absorbing forms. Selection of fractions with a high specific absorbance results in an A_{280}/A_{470}-ratio close to

2.1 (Figure 11B). This is the same value as achieved by Krause *et al.* [62] (though given as A_{280}/A_{480}), which was estimated to correspond to 80% active protein.

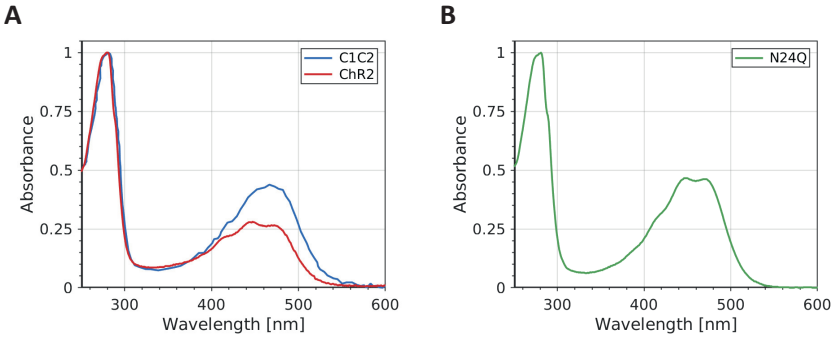


Figure 11. Absorbance spectra
A) Absorbance spectra of C1C2 and ChR2 after IMAC-purification.
B) Absorbance spectrum of ChR2-N24Q after gel filtration.

We estimate the yields of C1C2 after the initial IMAC step to be around 17 mg per liter of culture. However, this protein has little relevance for our purposes since it remains heterogeneous after purification. In contrast, both wild-type ChR2 and ChR2-N24Q can be purified to homogeneity. The yields of deglycosylated ChR2 and ChR2-N24Q are estimated to 3 mg/L after gel filtration with subsequent selection of the better fractions (with a low A_{280}/A_{470}-ratio). The yield is sufficient for a TR-SFX study on microcrystals in LCP. As a comparison, two previous TR-SFX studies on bR crystallized in LCP (**Paper IV** and [19]) required 30-80 mg protein. This amount would correspond to 10-30 L of cultures, which is fully realistic when scaling up to 10 L cultures.

Crystallization of wild-type ChR2 and ChR2-N24Q

Purified samples of deglycosylated ChR2 and ChR2-N24Q were concentrated to 10-30 mg/ml and mixed in a 2:3 ratio with monoolein in connected syringes. This gives a yellow, transparent LCP that appears to be stable for weeks (Figure 12A). The LCP was used for setting up small scale screens with the Mosquito robot (TTP Labtech). A number of commercial screens, several especially designed for membrane proteins, as well as screens around the literature conditions were set up, on material with and without previous removal of the His_8-tag using TEV protease.

Within the last weeks, we have set up a few screens on wild-type ChR2 after treatment with PNGase F. Some promising leads were found and could be reproduced. Elongated crystals with a length of 30-60 µm (Figure 12B) form within four days. There are also star-shaped crystals in another condition. The current plan is to optimize around the identified conditions, scale up the crystallization and send samples for screening at a synchrotron or an XFEL, to see if the crystals diffract.

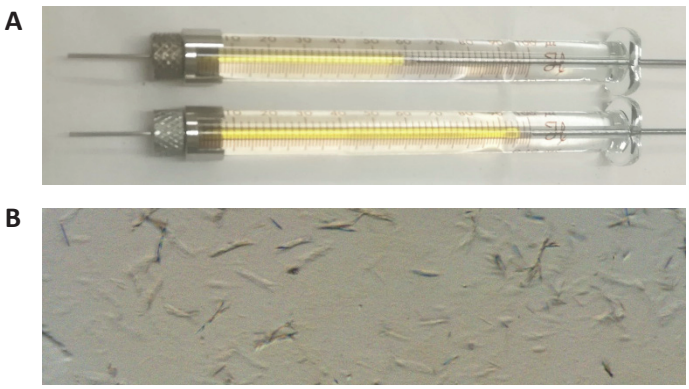


Figure 12. LCP and crystallization
A) ChR2-N24Q in LCP, at two different concentrations.
B) Crystals of deglycosylated wild-type ChR2.

Glycosylations in C1C2 and ChR2

We have recently analysed C1C2 and ChR2 with LC-MS to identify glycoforms of the proteins (see Supplementary material of **Paper I**). The results show that C1C2 is likely non-glycosylated at N86 and N153. In addition, the site at N174 is on the intracellular side and should not be glycosylated. This suggests that a single-mutant, C1C2-N61Q, could be a candidate for crystallization, possibly in combination with a truncated N-terminus. The results of LC-MS on ChR2 confirms that N24 is glycosylated. The relative distribution of different high-mannose glycans was estimated and glycopeptides with 6 to 23 mannose residues were detected.

Previous studies on ChR2 indicate that glycosylations are crucial for expression and functionality. The protein has been reported to give non-detectable yields in *E. coli*, possibly due to lack of post-translational modification [63]. When the fusion protein ChR2-YFP expressed in HEK-cells is deglycosylated with PNGase F it leads to complete loss of specific

absorbance. In addition, expression in the presence of the N-glycosylation inhibitor tunicamycin causes a ten-fold reduction in light-activated photocurrents [64]. Furthermore, an attempt to create a non-glycosylated form of the functional mutant ChR2-H134R by introduction of the S26D mutation (that alters the N24-G-S26 glycosylation site) gives extremely low yields in *P. pastoris* [65].

In contrast, we show that deglycosylation of ChR2 as well as expression of the non-glycosylated mutant ChR2-N24Q give retained specific absorbance at 470 nm and high yields of protein. This implies that the N24 glycosylation site is less crucial for expression and functionality than has been previously suggested, at least when the protein is expressed in *P. pastoris.*

Future work

The work presented here give two options for crystallization; deglycosylation of wild-type ChR2 and expression of ChR2-N24Q. Our focus has been on ChR2-N24Q since it does not require deglycosylation as part of the purification and we still aim to improve production and purification of this construct. However, the recent microcrystals of wild-type ChR2 have changed our perspective. We will investigate whether these crystals diffract and we aim to further optimize the conditions for crystallization. Hopefully, these efforts will lead towards a TR-SFX experiment.

In this study, spectrophotometry has been used to assess protein quality. Absorbance spectra of C1C2, ChR2 and ChR2-N24Q show the expected specific absorbance peak at 470 nm. Thus, the retinal is bound and absorbing and it is very likely that the proteins are functional. We aim to use time-resolved spectrophotometry to see if light-activation leads to the formation of expected intermediates in the photocycle, since this would be a stronger indication of the proteins being functional.

It would also be interesting to further study the role of glycosylations, with for example time-resolved spectrophotometry and differential scanning calorimetry. Changes in intermediate absorbance and temperature stability would reveal whether glycosylations affect function or stability of the protein.

CHAPTER 4. BACTERIORHODOPSIN

The bacteriorhodopsin project is mainly about analysis of structural dynamics. We start off reviewing the field by exploring the wealth of previously published structures and data. This reveals a unified picture on the structural mechanism behind light-induced proton pumping in bR. Then the focus shifts towards time-resolved experiments. The result of the first TR-SFX experiment on bR is a sequence of structural changes, like a movie that show the true light-induced response as well as correlation in time. The data from this and a later TR-SFX study [19] are largely in agreement with previous intermediate trapping results but add new details about proton transfer and ultrafast movements. An important part in my work was the development of new tools for comparison and interpretation of structures and difference Fourier electron density maps, that help us interpret the data and reveal new insight in the mechanism of proton pumping.

4.1. Objective comparison of deposited structures and X-ray diffraction data (Paper II-III)

bR is one of the most studied membrane proteins in many fields of biophysics, including structural biology. Until 2014 more than 90 structures had been published from X-ray crystallography, NMR and EM studies. This large set of structural information has had a major contribution to understanding of the mechanism behind proton pumping but it has also been a source for contradictions and debates. Since publishing of scientific data requires novelty and a selling story, there is a risk that differences are highlighted even when the overall data are in agreement with previous results. We aimed to reanalyze structural data on bR and objectively compare both structures and difference density maps, to see if the whole set of data reveal a different perspective than the individual structure papers. We were particularly interested in comparing all structures proposed to represent intermediates and see if subsets of these showed consistent conformational changes. This required the development of a new tool for structural comparison that we based on internal distance matrices. The results of the analysis are presented in an analytical review (**Paper II**). We also use this tool when analyzing the first room temperature structure of bR (**Paper III**).

Structural comparison

The function of a protein is closely related to its structure. When the structure of a novel protein is solved it may be compared to all known structures and in case there is another protein with a similar fold it may provide information about the function. As the number of structures published in the Protein Data Bank has increased over the last 30 years [18], there is a growing need for structural comparison. This has resulted in development of tools for structural alignment (for example DALI [66] and FAST [67]) and databases that classify proteins by fold and sequence similarity (including CATH [68], SCOP [69] and FSSP [70], reviewed in [71]). However, all these tools are developed for structural comparison of two or multiple proteins and not for analysis of a set of very similar models of a single protein.

Structural comparison of alternative structures can be done visually after alignment. How the alignment is performed – on the overall structure or locally – and where in the protein differences are expected may change the interpretation. Manual inspection of multiple alternative models is time-consuming and the complexity of the problem makes it even harder to be objective. An automatic procedure for structural comparison is faster and can be set up to be objective. The implementation requires a score and a method to represent structural differences.

The similarity of any two structures can then be scored as the root mean square deviation (RMSD) calculated over a subset of atoms (Equation 4)

$$RMSD = \sqrt{\frac{1}{n}\sum_{i}^{n} d_i^2} \qquad (4)$$

where n is the number of atoms in the subset and d_i is the distance between the atomic pair i. However, since the method of alignment changes the relative atomic positions, RMSD is dependent on how the alignment is performed. Another drawback with the use of RMSD as a score is that it contains information about the amplitudes of displacements but not the directions. An alternative and alignment-independent method is to score similarity of structures based upon differences in their internal interatomic distances. The differences between two structures are represented in a difference distance matrix which contains information about both amplitudes

and directions of displacement. The programs ESCET [72] and STRuster [73] are examples of programs utilizing internal distance matrices.

ESCET calculates and plots difference internal distance matrices on the C_α atoms of two structures. For crystallographic structures it also estimates errors in atomic coordinates and plots an alternative error-scaled difference distance matrix. The estimated error that is calculated for each pair of atoms is based on the "diffraction-component precision index" [72]. It is an empirical formula described by Cruickshank [74], that includes for example the completeness, the average B-factor and the maximum resolution of a structure.

STRuster is a program that clusters alternative structural models based on difference distanse matrices on C_α atoms. Two thresholds are applied before clustering. Firstly, a threshold on the maximum distance, which reduces the effect of long-distant conformational changes. Secondly, the minimum difference in internal distance is set and values above this threshold are set to 1. The sum over the resulting matrix is then used as a score in the clustering [73].

Hierarchical clustering of bR resting state structures

We began by analysing all resting state structures of wildtype bR. We wanted to use difference distance matrices on C_α atoms in the analysis, as it is independent on alignment. We also decided to develop a new tool since it would give us more freedom in the analysis. An internal distance matrix ($[C_\alpha]$) of each structure was calculated and the matrices were pair-wise subtracted to give difference distance matrices for all pairs (exemplified in Figure 13).

The score S_{ij} for structures *i* and *j* was calculated as the average of the absolute values in the difference distance matrix (Equation 5).

$$S_{ij} = Average(\left|[C_\alpha]_i - [C_\alpha]_j\right|) \qquad (5)$$

If some residues were missing in either of the structures in a pair, the score was calculated only from the residues present in both. Loop regions were excluded from the analysis because their flexibility would add differences that are less relevant for proton pumping than changes within the protein core. Finally, the structures were sequentially clustered with their nearest neighbour (corresponding to a low S_{ij} score) and the resulting hierarchical tree was plotted as a dendrogram. All of this was implemented in Matlab [75].

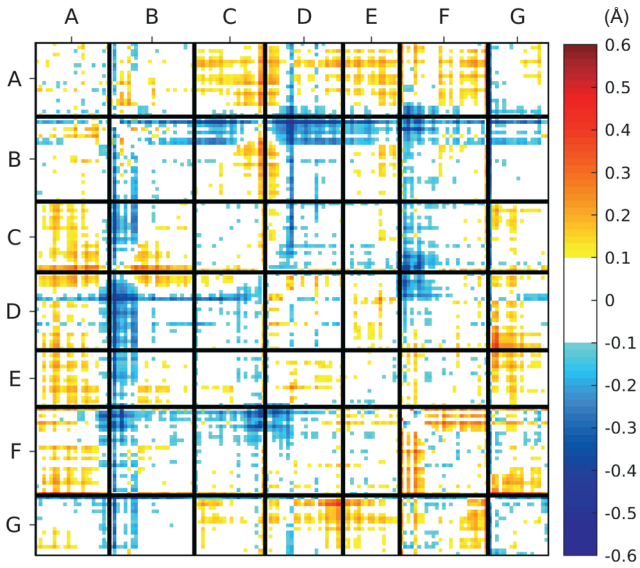
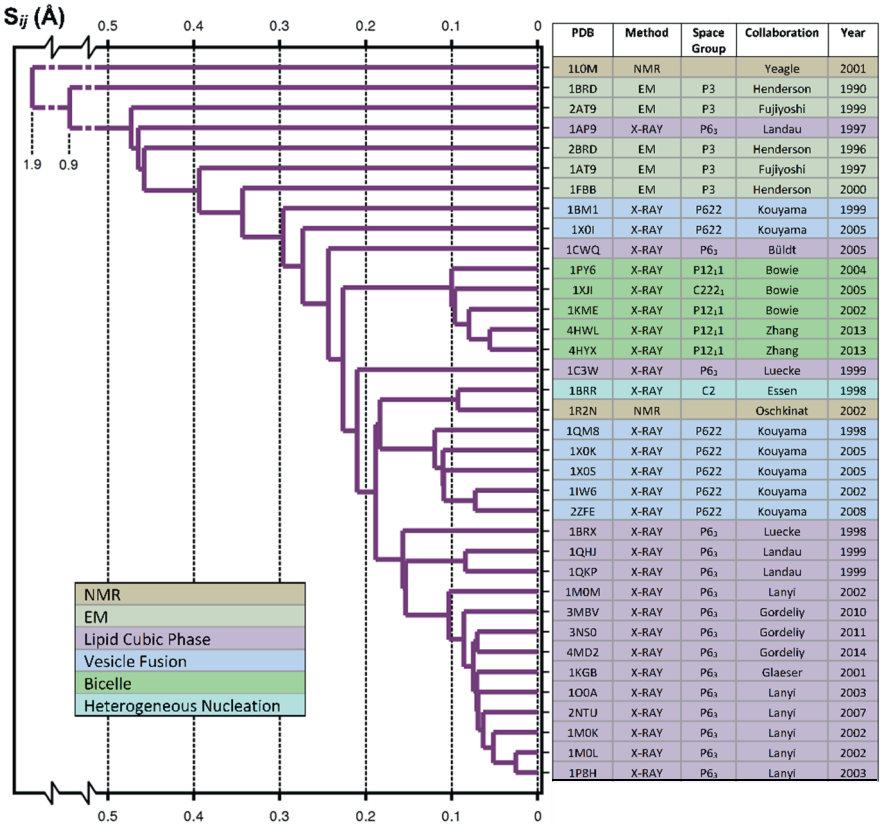


Figure 13. A difference internal distance matrix of two resting state structures, 2NTU [76] – 1QHJ [77]. In 2NTU helix A is more distant from the other helices and the intracellular part of helix B is closer than in 1QHJ. The differences are small, at maximum 0.55 Å.

Initially, difference distance matrices were error-scaled as in ESCET [72]. However, this error-scaling is not applicable for NMR and EM structures and even some crystal structures lack values for one or more of the required parameters. Another problem is that error-scaling seems to change the clustering score in an unpredictable way. We decided to abandon error-weighting to let close to identical structures cluster independent of whether or not some of the parameters for error weighting were slightly deviant. The resulting hierarchical clustering tree analysis (Figure 14) shows that resting state structures of bR mainly cluster by crystallographic space groups. This may not be surprising since crystal contacts will affect protein conformation and they differ between the crystal forms. Structures from the same collaboration (often using similar conditions for crystallization) correlate with space groups, which is why structures are also clustering by collaborations. The initial model used for structural refinement and the details in the crystallization protocol may also contribute to the clustering. Almost all X-ray diffraction structures are found in the lower part of the dendrogram and are thus more similar to each other than to all EM and most NMR structures. Altogether, the clustering of structures reflects the type of experiment and the influence of crystal packing rather than structural variances due to noise.



PDB	Method	Space Group	Collaboration	Year
1L0M	NMR		Yeagle	2001
1BRD	EM	P3	Henderson	1990
2AT9	EM	P3	Fujiyoshi	1999
1AP9	X-RAY	$P6_3$	Landau	1997
2BRD	EM	P3	Henderson	1996
1AT9	EM	P3	Fujiyoshi	1997
1FBB	EM	P3	Henderson	2000
1BM1	X-RAY	P622	Kouyama	1999
1X0I	X-RAY	P622	Kouyama	2005
1CWQ	X-RAY	$P6_3$	Büldt	2005
1PY6	X-RAY	$P12_11$	Bowie	2004
1XJI	X-RAY	$C222_1$	Bowie	2005
1KME	X-RAY	$P12_11$	Bowie	2002
4HWL	X-RAY	$P12_11$	Zhang	2013
4HYX	X-RAY	$P12_11$	Zhang	2013
1C3W	X-RAY	$P6_3$	Luecke	1999
1BRR	X-RAY	C2	Essen	1998
1R2N	NMR		Oschkinat	2002
1QM8	X-RAY	P622	Kouyama	1998
1X0K	X-RAY	P622	Kouyama	2005
1X0S	X-RAY	P622	Kouyama	2005
1IW6	X-RAY	P622	Kouyama	2002
2ZFE	X-RAY	P622	Kouyama	2008
1BRX	X-RAY	$P6_3$	Luecke	1998
1QHJ	X-RAY	$P6_3$	Landau	1999
1QKP	X-RAY	$P6_3$	Landau	1999
1M0M	X-RAY	$P6_3$	Lanyi	2002
3MBV	X-RAY	$P6_3$	Gordeliy	2010
3NS0	X-RAY	$P6_3$	Gordeliy	2011
4MD2	X-RAY	$P6_3$	Gordeliy	2014
1KGB	X-RAY	$P6_3$	Glaeser	2001
1O0A	X-RAY	$P6_3$	Lanyi	2003
2NTU	X-RAY	$P6_3$	Lanyi	2007
1M0K	X-RAY	$P6_3$	Lanyi	2002
1M0L	X-RAY	$P6_3$	Lanyi	2002
1P8H	X-RAY	$P6_3$	Lanyi	2003

Figure 14. Hierarchical analysis clustering tree of wild-type bR resting state structures.

Hierarchical clustering of structural perturbations

Investigation of light driven structural changes can be achieved through intermediate trapping. In a trapping study, crystals are either illuminated and frozen before data collection or illuminated during data collection. Variations in the trapping protocol, such as the wavelength of illumination, the temperature and the time delay before flash-freezing, determine the conformation that is captured. Structural changes that are believed to mimic features in the photocycle may also be captured by introduction of mutations or through physical pertubations such as changes in pH.

There have been many efforts to capture intermediate states of bR. At the time of **Paper II** 32 structures with structural perturbations had been published. To cluster this set by the same method as the ground states would result in a dendrogram similar to before, with a large influence of the crystal

packing. We were more interested in comparing the type of movements that occurred in the structures. Thus, we adapted the clustering scheme to capture this. The movement was interpreted as the change in internal distance matrix of the excited (intermediate) state and the corresponding resting state (often referred to in the same paper), giving one difference distance matrix per structural pair *i* (Equation 6).

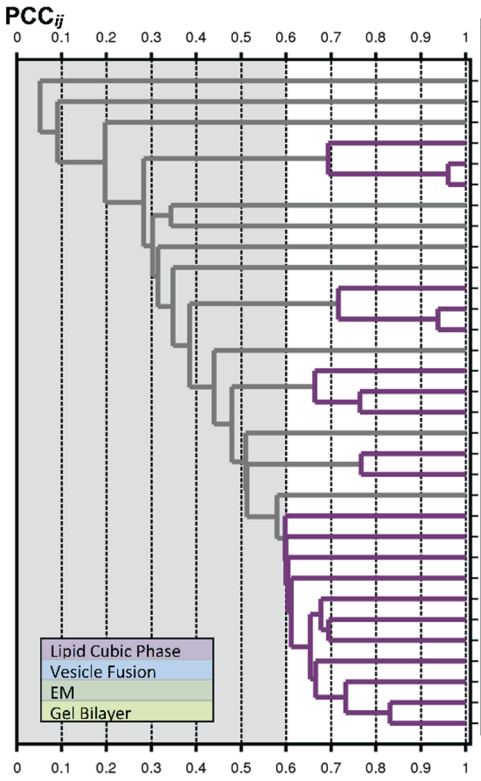
$$\Delta[C_\alpha]_i = [C_\alpha]_{i,excited} - [C_\alpha]_{i,ground} \quad (6)$$

The Pearson correlation coefficient of all pairs of difference distance matrices *i* and *j* were calculated (Equation 7) and used as a score for hierarchical clustering by the nearest neighbour algorithm as before.

$$Pearson\ clustering\ score = PCC\left(\Delta[C_\alpha]_i, \Delta[C_\alpha]_j\right) \quad (7)$$

A score close to 1 shows that two difference distance matrices are highly correlated. The score reflects the type of movement that occurs (the residues that are involved and the direction) but the value is not affected by the overall amplitude of the movement.

The largest cluster in the clustering tree (Figure 15) includes 11 structures of L, M and O states from all major groups that have contributed to the field [76,78-86]. The movement that they represent can be examined with single value decomposition (SVD) analysis. The principle SVD component (Figure 16) reveals two distinct interhelical movements; the extracellular part of helix C moving towards helix G and the cytoplasmic part of helix F moving away from helices A, B and G. This set of light-induced structures display conformational changes that are replicated in many studies published by several independent research groups using different crystal forms. We therefore conclude that this set likely represents real light-induced structural changes that occur in the photocycle of bR.



PDB	Ground state	Notes *K, L, M, N & O are illuminated*	Collaboration	Year
4MD1	4MD2	Orange species	Gordeliy	2014
1M0K	1M0K	K	Lanyi	2002
1IXF	1IW6	K	Kouyama	2002
1X0I	1IW6	Acid blue (pH 2)	Kouyama	2005
1JV6	1BRX	D85S/F219L	Luecke	2001
1JV7	1BRX	D85S	Luecke	2001
1X0K	1IW6	Alkaline (pH 10)	Kouyama	2005
2ZFE	1IW6	Xenon-bound	Kouyama	2008
3NSB	3NS0	X-ray damaged	Gordeliy	2011
1X0S	1IW6	Dark adapted	Kouyama	2005
1M0M	1M0M	M1	Lanyi	2002
1QKO	1QKP	K	Landau	1999
1QKP	1QKP	K	Landau	1999
1UCQ	1IW6	L	Kouyama	2004
1P8U	1P8U	N (V49A)	Lanyi	2003
1O0A	1O0A	L	Lanyi	2003
1P8H	1P8H	M1	Lanyi	2003
1KG9	1KGB	Annealed	Glaeser	2001
1FBK	1FBB	D96G/F171C/F219L	Henderson	2000
4FPD	1KGB	D96G/F171C/F219L	Glaeser	2013
1C8S	1C8R	M (D96N)	Luecke	1999
2NTW	2NTU	L	Lanyi	2007
1CWQ	1CWQ	M	Büldt	2000
2I20	2I1X	M (D96A)	Lanyi	2006
1F4Z	1F50	M (E204Q)	Luecke	2000
1KG8	1KGB	M	Glaeser	2001
1E0P	1QHJ	K:L:M (1:3:1)	Landau	2000
1VJM	1QHJ	L	Landau	2004
3VI0	3VHZ	O (L93A)	Kouyama	2012
1DZE	1IW6	M	Kouyama	2004
1IW9	1IW6	M	Kouyama	2004
2ZZL	1IW6	M	Kouyama	2009

Figure 15. Hierarchical clustering tree of structural perturbations in bR induced by light, mutations, pH, xenon, thermal annealing, dark adaption and radiation damage.

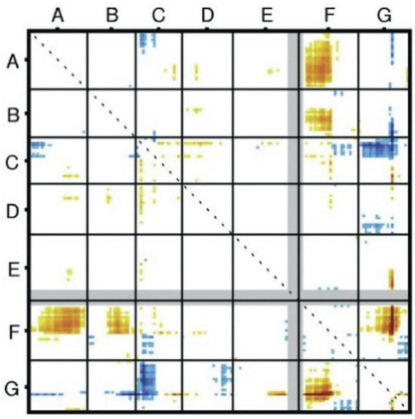


Figure 16. The principal SVD component of 11 difference internal distance matrices describes the structural perturbations that occur in the main cluster (at the bottom of Figure 15). In the intermediate structures the intracellular part of helix F is moving away from helices A, B and G and the extracellular part of helix C is moving towards helix G.

Difference Fourier analysis of light-induced structural changes

When a new X-ray structure is published it is commonly presented with figures of the structure and a corresponding electron density map that support and explain structural findings. The point of view in a figure, which atoms that are included and the sigma level of the map are set by the author. This makes it hard for other researchers to get an objective view of the data, unless crystallographic observations or electron density maps are available. Depositing of crystallographic observations is mandatory for X-ray structures submitted to the PDB from February 1, 2008 [87]. Many of the bR structures were deposited earlier and this explains why structure factor files are only accessible for eight out of 19 intermediate states. This set contains one K-state [88], four L-states [76,78,82,89], two M-states [80,85] and one O-state [86].

We recalculated difference Fourier electron density maps for these data sets. The observed structure factor amplitudes of the light state ($F_{obs,\ light}$) were subtracted by those from the dark data set ($F_{obs,\ dark}$) and the phases were retrieved from the dark state model. Experimental difference amplitudes were error-weighted before calculation of the map according to a method founded on Bayesian statistics [90], termed q-weighting. It is based on *A priori* knowledge about the reflections, for example that stronger reflections have larger errors, and takes this information into account for better estimation of structure-factor amplitude differences. This reduces noise and improves the quality of the retrieved difference density map.

Recalculation of all electron difference density maps with the same approach and comparison of maps from the same angle reveal high similarity between datasets. An example is the L-state structure of Lanyi *et al*. [76] compared to the L-state (with some K and M contamination) of Royant *et al*. [82] (Figure 17). Almost all features are existent in both the maps but the former model captures less of the structural perturbations that are indicated. This shows that contradictions between individual publications may sometimes be due to modelling and interpretation rather than differences in the collected data. The reanalysed data reveal a surprisingly unified picture of the key conformational changes in light-activated bR. Structural insights from this analysis are further discussed in section 4.4.

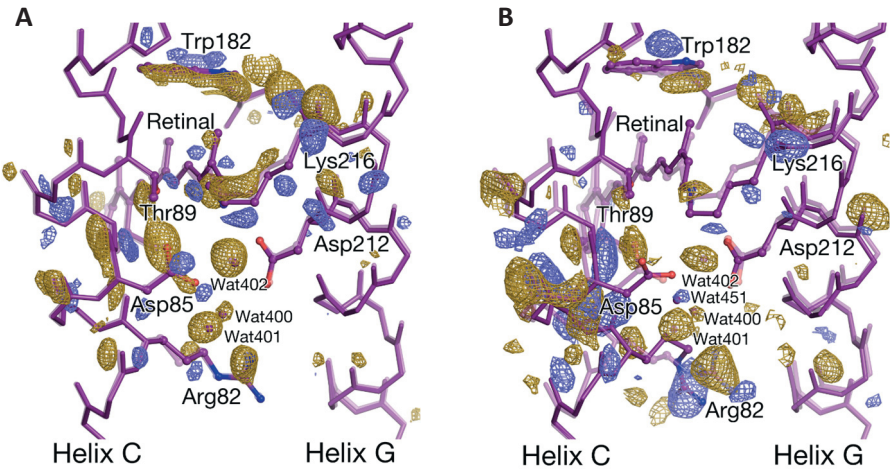


Figure 17. Two L-states and their corresponding recalculated difference density maps that show high similarity. The resting states (transparent) are shown for comparison.
A) An L-state of Lanyi et al. (2NTW-2NTU) [76]. The map is contoured at 4σ.
B) An L-state of Royant et al. (1E0P-1QHJ) [82]. The map is contoured at 3.2σ.

Analysis of a room-temperature structure of bR from SMX data collection

After publication of the analytical review (**Paper II**), the first room temperature structure of bR was solved through a serial millisecond crystallography (SMX) experiment on microcrystals in lipidic cubic phase (**Paper III**). This was a state-of-the-art experiment, showing the possibilities of room-temperature SMX data collection at a synchrotron microfocus beamline. The data collection is similar to an SFX experiment at an XFEL (as described in **Paper IV**), but the aim of adapting the technique to a synchrotron beamline is that synchrotrons are commonly available whereas XFELs have very limited access. The LCP with bR microcrystals was injected to the beam with an LCP injector [91] at a slow speed, around 0.1 mm/s, giving an efficient use of sample. Despite the low speed, a significantly smaller dose of radiation per crystal is achieved than in a single crystal experiment using synchrotron radiation. X-ray exposure causes both global and specific damages in proteins. Disulfide bridges and the acidic residues aspartate and glutamate are especially sensitive, even at cryogenic temperatures [92]. Signs of radiation damage in bR include cleavage of the carboxylate head groups of Asp 85 [93,94] and Asp212 [93]. The damage is detected as strong negative difference Fourier electron densities when comparing maps obtained later

and earlier during data collection on a single crystal. In the room temperature TR-SMX structure of bR there are no such signs of radiation damage and this confirms that the dose per crystal is low.

The SMX structure of bR was compared to a conventional cryo structure that was collected on one of the larger bR crystals grown in the same LCP setup as the microcrystals for SMX. The two new structures, both in the space group $P6_3$, were analysed with the hierarchical clustering tool together with all the previous wild-type bR structures.

The hierarchical tree shows that the cryo structure cluster within the large cluster with space group $P6_3$, in agreement with previous clustering results (Figure 14). In contrast, the room temperature structure ends up in the middle of the tree (clustering at a score slightly higher than 1C3W). The reason for this is that helices D, E and F are slightly further away from A and B in the room-temperature structure, as is seen in the difference distance matrix of the SMX and cryo structures (Figure 18). This is corresponding to a small compression of the cryo structure which we interpret as an effect of the cryo-cooling. The SMX setup that is described in the paper is useful for screening purposes and time-resolved experiments that does not require the brilliance or time resolution of an XFEL.

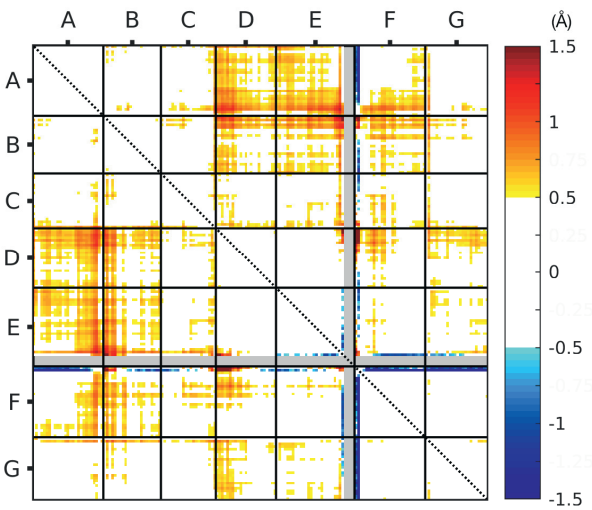


Figure 18. Difference internal distance matrix of the SMX room-temperature structure and the cryo-temperature structure (SMX-cryo).

4.2. Time-resolved serial femtosecond crystallography from nanoseconds to milliseconds: Background and analysis of time-scales in the data (Paper IV)

The crystallography structures of bR on cryo-trapped intermediates and mutants (reviewed in **Paper II**) revealed insight into the mechanism of proton pumping. However, differences between models of the same intermediate state and potential effects of x-ray induced radiation damage raised many questions about structural details and timing of events. With the development of XFELs it is now possible to collect diffraction data free from radiation damage on microcrystals at room temperature. For light-activated proteins the shift from synchrotrons to XFELs also enables real-time studies of structural dynamics through TR-SFX experiments. The development and applications of this new technique are reviewed in **Paper VI**.

bR was one of the first proteins to be studied with TR-SFX. **Paper IV** describes an experiment that allowed us to generate a 3D-movie of structural changes in bR, showing in detail the initial cascade of motions that leads to active transport of protons against a concentration gradient. We elucidate the mechanism of pumping in the time span from nanoseconds to milliseconds after activation of the photocycle.

The focus of this subchapter is on the background to the experiment and on the analysis of time-scales in the data. Further analysis of electron density maps is described in subchapter 4.3 and the final structural results are presented and discussed in subchapter 4.4.

Requirements for TR-SFX experiments at XFELs

The development of the first X-ray free electron lasers [95,96] increased the brilliance of X-rays with nine orders of magnitude compared to synchrotron facilitates. This makes it possible to collect high resolution data on micro-sized crystals. The X-ray pulse duration was also significantly shortened, down to the range of a few tens of femtoseconds, allowing for increased time-resolution. A few additional developments in sample handling, detector performance and data processing were necessary to allow TR-SFX experiments at XFELs. New sample handling systems such as the gas virtual dynamic nozzle device [97] and the LCP-injector [98] allow continuous delivery of new microcrystal into the beam; crystals in liquid samples are sprayed into the beam as a jet, whereas LCP is squeezed out as a strain. New

X-ray detectors have faster read-out [99,100] and new data-processing software packages [101,102] detect when crystals are hit, process large numbers of diffraction patterns from randomly oriented crystals and handle variances in intensities due to large intensity fluctuations between XFEL pulses.

Time-resolved studies on bR

The first SFX study on bR [103] was a proof-of-principle experiment, demonstrating that a new LCP injector device could be used at XFELs for collection of time-resolved structural data. A single time-delay of 1 ms between the pump laser and the X-ray probe resulted in a structure resembling previous M-states. This laid the foundation for our TR-SFX study (**Paper IV**), and a later study spanning from sub-picoseconds to milliseconds [19].

Project overview

A TR-SFX experiment on a protein is complicated and requires expertize in many different areas. Before the experiment large amounts of protein have to be purified and crystallized, and a set-up including a sample injector and a laser has to be coupled to the beamline. During the beamtime samples are loaded, all instruments have to be operated and diffraction patterns are pre-processed and analyzed online. After the experiment enormous amounts of data are further processed for calculation of improved difference density maps, structural modelling and interpretation of results. This explains why TR-SFX experiments often are performed in large collaborations.

The TR-SFX study on bR was initiated by a Japanese group lead by Eriko Nango in the lab of So Iwata. Our group from University of Gothenburg was involved in planning the strategy for data collection and we went to the experiment at SACLA [96], the Japanese XFEL. My main focus was on the later part of the analysis, to extract information from difference density maps, structural models and spectral measurements to explain the structural dynamics of bR.

The bR sample was isolated from its natural source, the purple membrane of *H. salinarum,* and was crystallized in LCP. It was injected with an LCP injector [91] placed above the XFEL beam. A pump laser beam was positioned to overlap with the X-ray beam on the sample stream. The timing between the pump laser and the X-ray probe was controlled for capturing diffraction at different times in the early part of the bR photocycle. Diffraction patterns for the dark state and 13 different time points were collected, evenly spaced on

a logarithmic scale over the time interval from 16 ns to 1.725 ms. Difference electron density maps were calculated and structures of the dark state and 13 illuminated states were modelled with resolution of 2.2-2.0 Å.

Spectral decomposition

Spectroscopic measurements were performed on slurries of bR microcrystals to confirm that bR in the crystal form was active and had a photocycle similar to bR in solution. The absorbance data were analysed with spectral decomposition, in part based on a routine for linear decomposition of time-resolved wide-angle X-ray scattering data [104]. The early photocycle was treated as a first-order reaction with two visible species (A, B) and an invisible end state (C) (Equation 8). How the populations (or concentrations) of the species change with time *t* depend on the rate constants k_1 and k_2 (Equation 9-11).

$$A \overset{k_1}{\rightarrow} B \overset{k_2}{\rightarrow} C \tag{8}$$

$$[A] = e^{-k_1 t} \tag{9}$$

$$[B] = \frac{k_1}{k_2 - k_1} \left(e^{-k_1 t} - e^{-k_2 t}\right) \tag{10}$$

$$[C] = 1 - [A] - [B] \tag{11}$$

The algorithm for spectral decomposition optimizes the values of the two time constants. First, an initial guess of the time constants is given and the populations for each time point are calculated as in Equation 9-10. The measured absorbance spectrum *Abs* at each time point is a function of the basis absorbance spectra *S* of the present species and the populations *P* (Equation 12).

$$\boldsymbol{Abs}(\lambda, t) = \boldsymbol{S}(\lambda) * \boldsymbol{P}(t) \tag{12}$$

The basis spectra can then be estimated from this equation system with least squares methods. The time constants are fitted to minimize the difference between the data, Abs_{data}, and the theoretical absorbance data, Abs_{theory}, that is reconstituted from the basis spectra and the populations. The best fit of the data yields the final populations and basis spectra.

A problem with the absorbance measurements was that the light pulse for measuring the absorbance consisted of a peak with 0.6 µs full width at half

maximum (FWHM) with a 35 µs tail. Since 40% of the light came from this tail it reduced the temporal resolution of the measurement. To partially compensate for this the spectral decomposition routine was set up to be iterative. Before each iteration except the first, the influence of the tail on the absorbance was estimated from the basis spectra of the previous run. The part of the absorbance origin from the tail was then subtracted from the experimental data and the corrected data was used for another round of spectral decomposition, until the solution converged.

The change in difference absorbance with time (Figure 19A) show an increase around 410 nm and a depletion in the range from 450 to 650 nm before returning back to absorbance of the resting state. Spectral decomposition with the iterative routine described above give two basis spectra (Figure 19B) and the corresponding populations of the first and second species (Figure 19C). The first basis spectrum has the characteristics of an L-state with some contribution from a K-intermediate. The second basis spectrum has a positive peak centered at 410 nm and a depletion at 550 nm as would be expected of the M-intermediate in which the retinal is deprotonated. This implies that the second state corresponds to the M-intermediate. The estimated growth of the M-state population with time aid interpretation of the time-resolved difference density maps.

Analysis of difference density maps with linear decomposition

When interpreting the 13 difference density maps we started by identifying strong negative or positive features, with an amplitude larger than ±2σ (where σ is the root-mean square electron density over the unit cell). The selected features were found around the active site and along helices. The aim was to determine the time when a feature appeared or disappeared and how its amplitude changed with time. Initially each feature was assigned exact coordinates in the map and the amplitude at that positon was extracted, but this was not plausible since the “center” of a negative or positive difference density feature (corresponding to maxima or minima in the map) moves slightly from map to map. Instead the amplitudes of the features were read manually from visual inspection of each map. This is a tedious work and there is a risk that relevant movements in parts of the protein are overseen. This is the reason why a new tool for interpreting set of electron density maps was developed later (**Paper V**). Two maps (Δt = 95.2 µs and 657 µs) were noisier and the last map (Δt = 1.725 ms) had higher quality than the others, so that the amplitudes were lower or higher relative to the

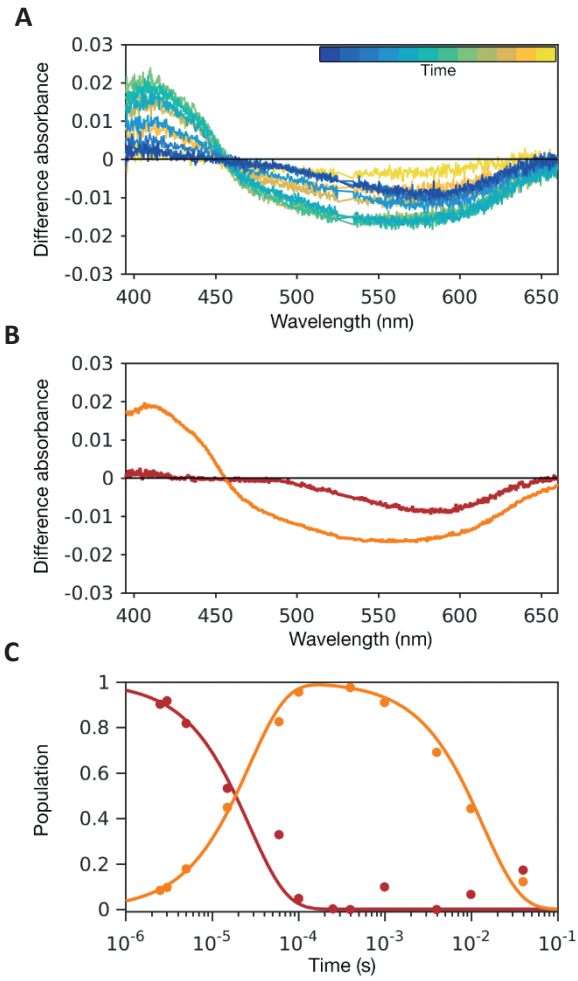


Figure 19. Spectral decomposition of time-resolved difference absorption spectra recorded from slurries of microcrystals.

A) Difference absorbance spectra recorded at twelve time-points (from blue to yellow). The spectra have been numerically corrected to compensate for an extended tail in the time-dependence of the microsecond flash-lamp used for these measurements.

B) Basis spectra from the spectral decomposition for the first (red) and second (orange) components. The first basis spectrum appears to be a spectrum from a mixture of the K and L intermediates, the second is similar to the spectrum of the M intermediate state.

C) Time dependence of the first (red) and second (orange) components from the spectral decomposition. The dots are the amplitudes that give an optimal linear combination of the basis spectra when fitted to the measured data. Thus, they indicate how well the theoretical curves fit the data for each measured time-point.

rest of the maps. This discrepancy was handled by rescaling of the amplitudes with the factors 2, 1.5 and 0.8. These scaling factors were extracted from SVD analysis assuming that the amplitude of the principle SVD component should increase linearly with the logarithm of time.

The data on difference density features and their amplitudes was used for linear decomposition similar to the spectral decomposition described above. Instead of an absorbance spectrum, with amplitudes at different wavelengths changing with time, the input data consisted of difference Fourier electron density peak amplitudes at different features in the map and how these changed with time (Figure 20A). The linear decomposition give three basis spectra (Figure *20*B) and shows how these are related to time (Figure *20*C). The first component is associated with amplitudes on the retinal and Wat402, the second involve Helix F and Helix G and the third show when amplitudes appear around additional water molecules and helix C. The last component is almost spot on the time scale for the raise of the M-state that was retrieved through spectral decomposition. This shows that the large structural changes in Helix C, modelled as an inward movement, are correlated to the proton transfer event that occur in the L-to-M transition. The largest growth of the M-state population occurs in the time range from approximately 5 to 40 μs.

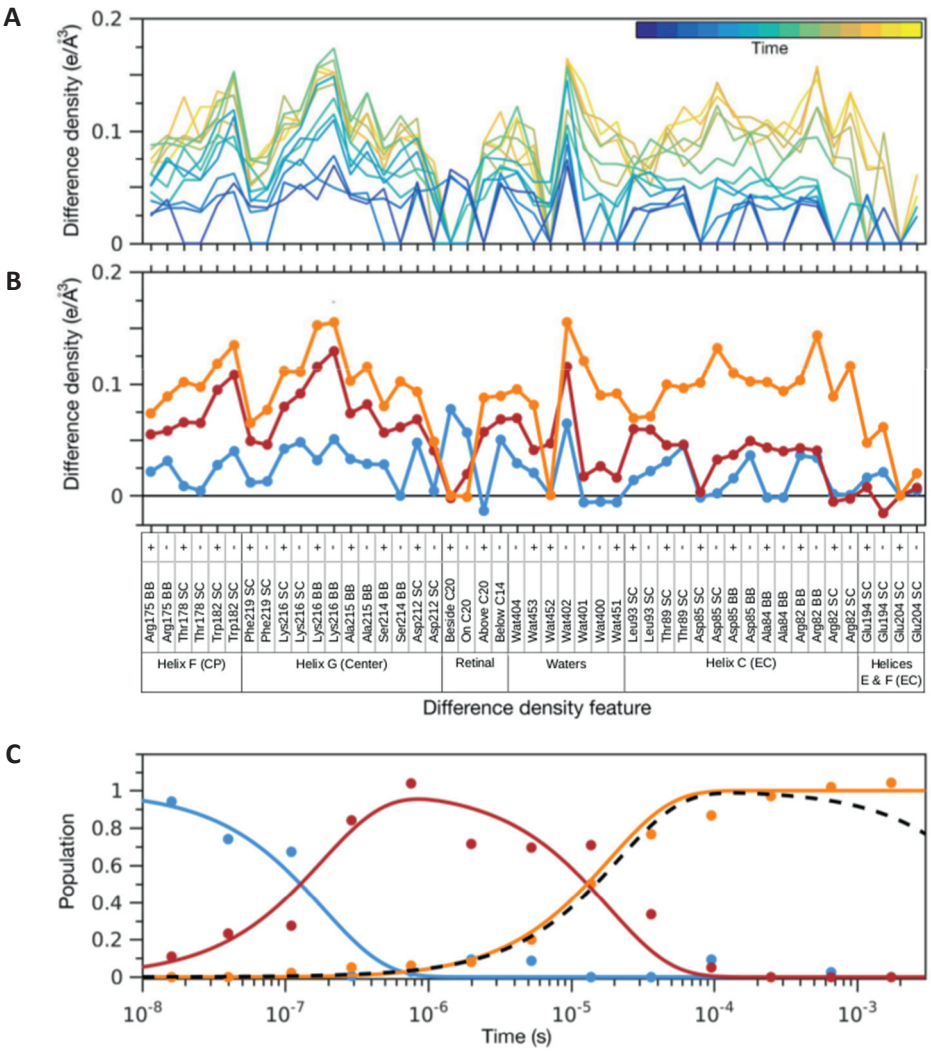


Figure 20. Linear decomposition of time-resolved difference Fourier electron density amplitudes.

A) Absolute difference amplitudes for selected features in the difference maps and how they change with time (from blue to yellow). This is a graphical representation of the time-resolved data used for linear decomposition.

B) Basis spectra from linear decomposition of the data in (A) assuming three components.

C) Relative populations of the three basis spectra in (B). The dashed line that overlaps with the third (orange) component indicates the growth and decay of the M-state population from spectral decomposition.

4.3. A new tool to analyze difference density maps (Paper V)

Time-resolved crystallography experiments enable a detailed view of structural changes occurring on time-scales from femtoseconds to milliseconds. The data is collected at multiple time-points and give a large set of closely related difference Fourier electron density maps. It is not straight-forward to visually analyze sets of maps because conformational changes may occur in several parts of the protein and the maps may also differ in quality. Modelling of structural intermediates may aid interpretation, but there is a risk that important features of the maps are not correctly captured in the models.

In **Paper V** we describe a new tool for analysis of sets of difference Fourier electron density maps. This tool was developed and validated on TR-SFX data on bR. The tool aids identification of reproducible electron density features and comparison of experimental maps. We also calculate theoretical difference density maps from models of intermediates and show that the tool enables validation of structural models towards experimental data.

An algorithm for visualizing three-dimensional maps in one-dimension

Three-dimensional difference density maps contain large amounts of information. We simplify the interpretation by converting this information into one-dimensional functions capturing positive and negative difference density amplitudes along the atom trace.

In the algorithm, each map is first converted to Cartesian coordinates using the program maprot in the CCP4 suit [105,106]. The map is then read into Matlab for further processing. A sphere, typically with a 2 Å radius, is placed inside the map at every atomic position of the resting state structure (as shown in Figure 21). Next, difference density amplitudes at grid positions inside this sphere are calculated through interpolation and values under a certain threshold (typically 3σ) are set to zero. Finally, mean positive and mean negative amplitudes are calculated separately over each sphere. This results in dual amplitude functions along the atom trace that can be plotted as a single one-dimensional graph per map.

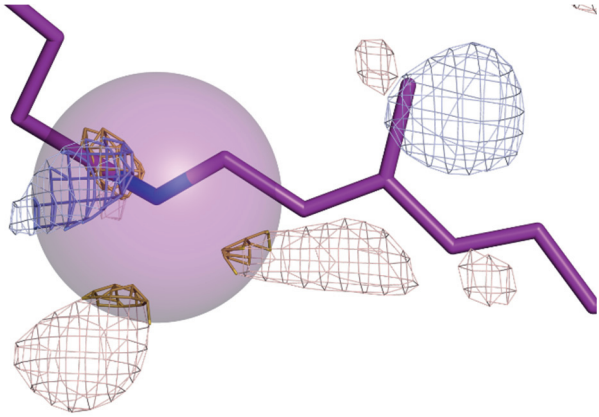


Figure 21. Sphere of radius 2 Å around the SB nitrogen atom to illustrate the selection of difference Fourier electron density amplitudes inside the sphere (marked as a thicker mesh).

Visualization of structural changes in bR in one and three dimensions

The power of our tool can be demonstrated with TR-SFX data from bR. We first visualize how electron density amplitudes change around the retinal in the time window from 16 ns to 1.725 ms (Figure 22A). The one-dimensional representation shows how small difference densities immediately appear around the SB, Lys216 and Trp182 to thereafter increase in amplitude with time. Around 5.25 ms after activation the motion starts spreading to helix C. The amplitudes around Asp82, Asp85 and Thr89 are increasing from this time point to the last (except in the 95.2 µs map due to lower quality). As expected, these findings are comparable to what is seen in the corresponding difference density maps (Figure 22B). The one-dimensional representation makes it easier to identify exactly when and where structural changes occur. Thus, it is a good complement to figures of three-dimensional maps that better illustrate the direction and extent of a conformational change.

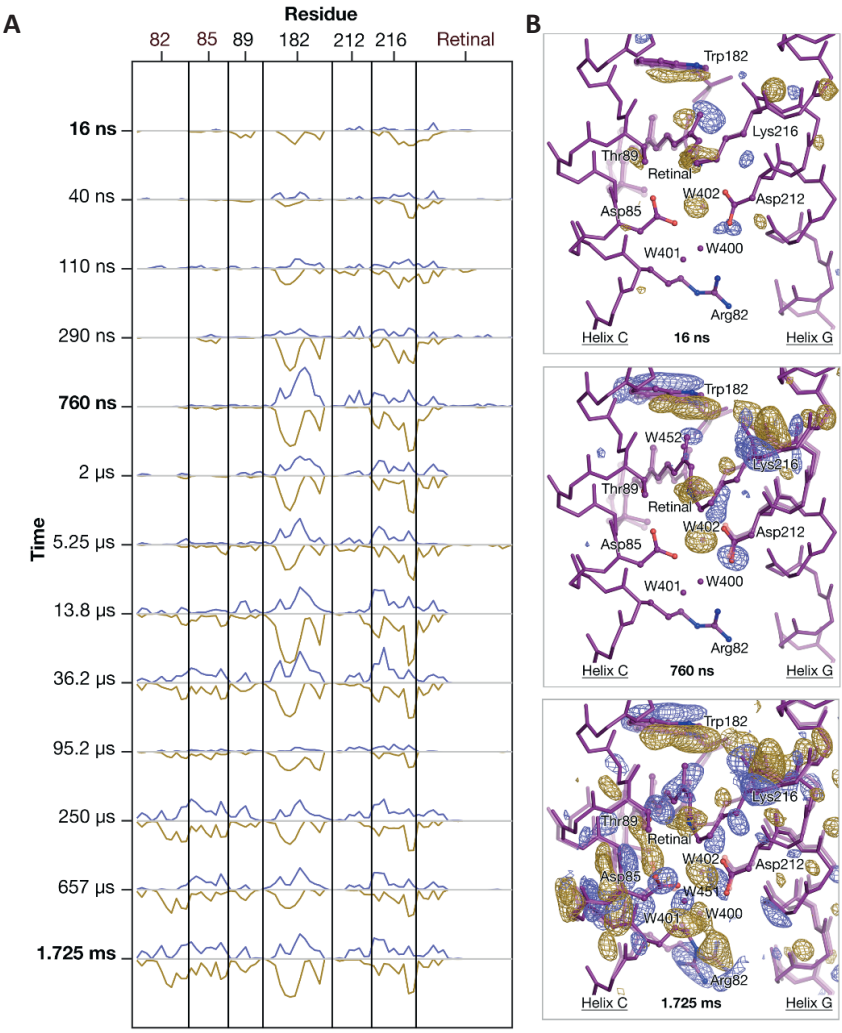


Figure 22. Representation of difference Fourier electron difference densities from the nanosecond to millisecond TR-SFX experiments on bR (**Paper IV**) with the dual amplitude function.

A) One-dimensional representation of difference electron densities of residues around the active site at different times. Mean positive (blue) and mean negative (yellow) amplitudes were calculated within a sphere of 2 Å radius at each atom, with a threshold on amplitudes set to 3σ.

B) Difference Fourier electron density maps for the time points 16 ns, 760 ns and 1.725 ms, contoured at 3.5, 4 and 3.5σ.

Reanalysis of difference density maps from the sub-picosecond TR-SFX data

When the sub-picosecond data of bR [19] was to be published I got a request to analyse difference density amplitudes in the maps using linear decomposition (as was performed on the previous TR-SFX data in **Paper IV**). Amplitudes of a number of difference density features, mainly around the retinal, were manually read-out in coot for the analysis. However, the linear decomposition did not give any results since there was no clear time progression in the given amplitudes. As a consequence, this analysis was not included in the paper. Now when having developed this new map tool we get a second chance to analyse the data.

When difference density amplitudes around all protein atoms are plotted in the one-dimensional representation (Figure 23), paired positive and negative densities are seen in parts of helices C, F and G. The amplitudes appear not to change with time, except within a small region of helix G (Leu206, Leu207 and Phe208) that seems to oscillate within the time window. There are also paired positive and negative densities around helix B that appear within 300 fs from light-activation and are damped around the time 800 fs. These oscillatory motions are also found in the maps but were overseen in the manual analysis. This reveals that the one-dimensional representation aids interpretation of maps and that bR has a more complex structural response to photoactivation than previously suggested. However, some types of conformational changes, like stepwise rotation around bonds in the retinal, are not highlighted by the one-dimensional representation. Such movements that have a clear direction but little effect on amplitudes are better visualized in the difference density maps.

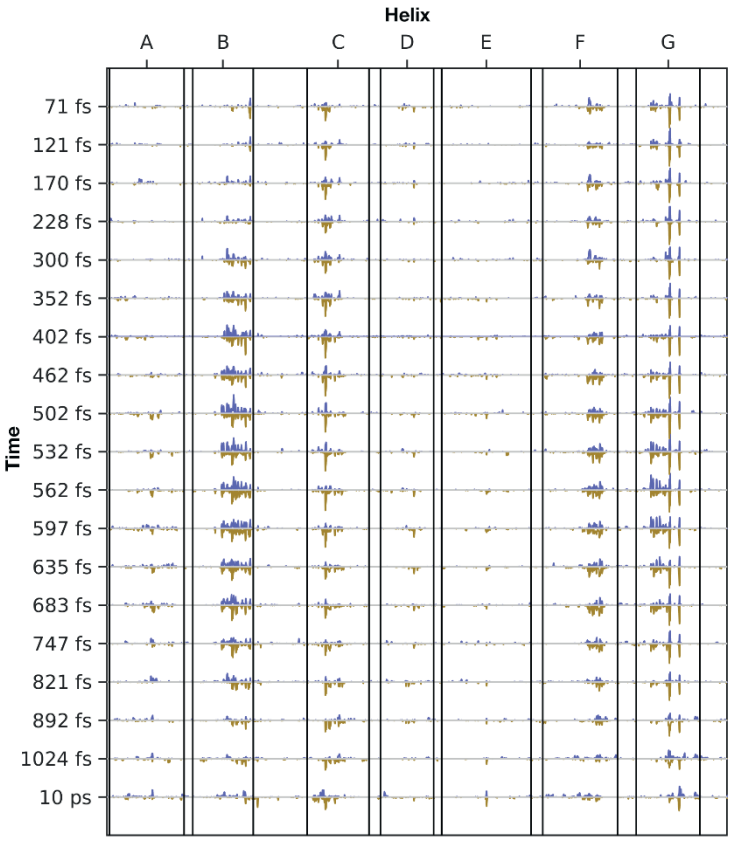


Figure 23. Representation of difference Fourier electron difference densities from the sub-picosecond TR-SFX experiments on bR [19] with the dual amplitude function. All protein atoms are represented along the x-axis in the same order as in the PDB-file. Mean positive (blue) and mean negative (yellow) amplitudes have been calculated within a sphere of 2 Å radius, with a threshold on amplitudes set to 3σ.

Correlation of maps

The map tool can be used to find similarities between maps through calculation of Pearson correlations. Initially, correlation scores were calculated on all grid-points within a sphere of an atom, but this resulted in a score that was sensitive to variations and hard to interpret. Instead, we decided to calculate Pearson correlations on the dual amplitude functions. This gives a single score for any two maps and a diagonal matrix when a set of maps are pair-wise compared.

A Pearson correlation matrix was calculated on the nanosecond-to-millisecond data (Figure 24A). It shows that the first difference density maps are least correlated to the rest of the maps. This is likely due to initial changes occurring only in a small region of bR. There are two blocks in the matrix that are highly correlated. The boundaries of these blocks indicate the times when larger structural changes occur. The same correlation analysis was performed on theoretical difference density maps, that were calculated from the resting state model and the thirteen intermediate models. The theoretical Pearson correlation matrix (Figure 24B) is similar to experimental matrix. However, the blocks appear clearer and correlation scores are higher, since the theoretical maps contain less noise and fewer details [88].

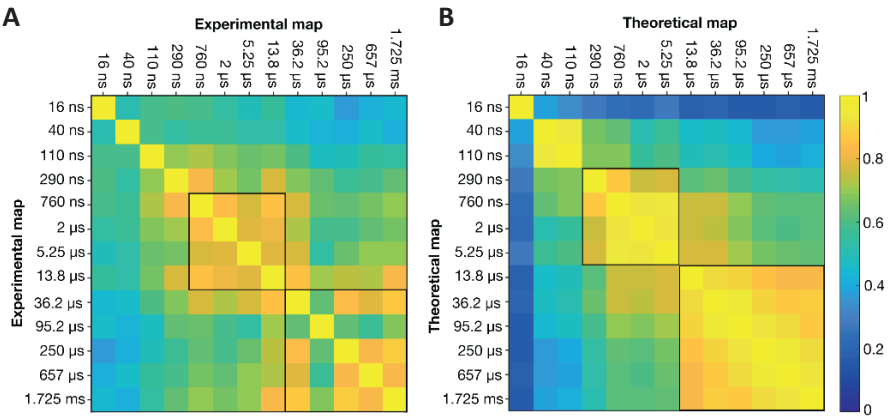


Figure 24. Pearson correlation between maps from the ns-to-ms TR-SFX experiment on bR (**Paper IV**), calculated from the dual amplitude functions. The black squares mark regions with higher correlation.
A) Correlation of experimental maps. The 95.2 µs map has a lower quality, which explains the lack of correlation between this map and the adjacent maps.
B) Correlation of theoretical maps, calculated based on the published structures.

Validation of models

Another use of the map tool is validation of structural models. If theoretical difference maps are compared to experimental maps it is possible to see how well the models capture the information in the data. An overall picture may be drawn from a Pearson correlation matrix of all pairs of experimental and theoretical maps. However, a much higher level of detail is given from the one-dimensional representation. When the thirteen theoretical maps of bR are plotted next to the experimental maps (Figure 25) a high level of agreement is seen. Zooming of the plots makes it possible to pin-point single

atoms that have different amplitudes between the sets. This type of analysis may guide modelling of TR-SFX data.

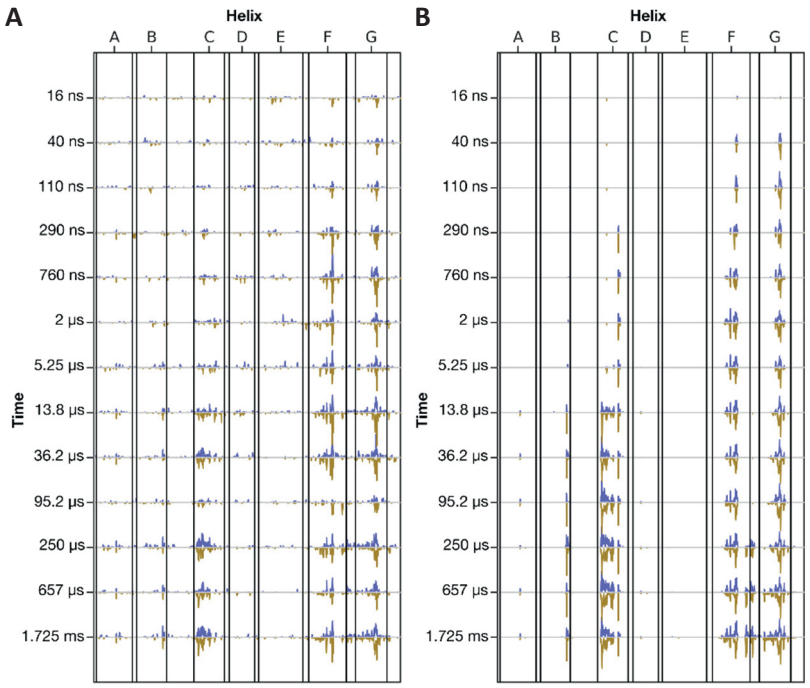


Figure 25. Dual amplitude functions calculated on experimental and theoretical difference Fourier electron density maps. All protein atoms are represented along the x-axis in the same order as in the PDB-file. Mean positive (blue) and mean negative (yellow) amplitudes have been calculated within a sphere of 2 Å radius, with a threshold on amplitudes set to 3σ for the experimental maps. The theoretical maps have a different scale and as a consequence, the threshold cannot be defined in terms of σ. Instead, the threshold for these maps were set by visual comparison between (B) and (A).
A) Experimental dual amplitude functions.
B) Theoretical dual amplitude functions.

4.4. The mechanism of proton pumping: Structural insights revealed from previous intermediate trapping crystallography and recent TR-SFX studies (Paper II, IV-VI)

In this thesis I have presented work on bR that contributes to a more detailed picture on the mechanism of proton pumping. Here I will start describing the findings from the analysis of intermediate trapping studies in **Paper II**. Then I will present the results from TR-SFX, both from the sub-picosecond study [19] and the nanosecond to millisecond study of **Paper IV**. Some results were interpreted with the map tool of **Paper V**.

In **Paper VI** we review the mechanism of proton pumping in bR with a focus on insights revealed from TR-SFX at XFELs. When we analyse the difference electron density maps and structures from these TR-SFX studies we find that they are in most details very similar to the previous results from intermediate trapping studies. However, the major advantages with the TR-SFX data are that the time scale is absolute and that structural changes are due to light-activation, practically without artefacts of radiation damage, mutations or physical perturbations. This means that the TR-SFX data reveal for the first time what is truly happening in bR in real time.

Structural results from intermediate trapping studies

As described in the introduction, decades of research in spectrophotometry and structural biology have resulted in the well-established photocycle of bR and high understanding of major proton transfer reactions as well as structural intermediates (Figure 3) [20-22]. However, there have been many controversies around the interpretation of results from intermediate trapping studies. In **Paper II** we reanalyze previously published structures and data and we find that the controversies are exaggerated. Only a limited set of structures are published together with crystallographic data, but recalculated difference density maps from this set provide a relatively unified picture of light-induced conformational changes in bR.

In the K-state, conformational changes occur in a small region around the retinal and Lys216 (Figure 26A). The strongest difference density is seen on the adjacent water molecule (Wat402) [88]. It has been observed that radiation damage gives a similar difference density pattern on Wat402, as an effect of cleavage of the adjacent aspartate carboxylate groups [93,94]. Due to this, Wat402 is retained in the K-state of Matsui *et al.* [94]. Schobert *et al*.

[107] also have Wat402 in the K-state model, partly due to FTIR spectroscopy indicating that the hydrogen bond to the SB is unchanged. Unfortunately, none of these structures are published with the crystallographic data.

In the L-, M- and O-states, several structural perturbations are reproducibly captured to various extent in most of the recalculated maps (as exemplified in Figure 26B); Two additional water molecules are disordering (Wat400 and Wat401), a new water molecule (Wat451) is ordering between Asp85 and Asp212, a hydrogen bond from Thr89 to Asp85 is breaking, Trp182 is moving towards the cytoplasm, the side-chain of Arg82 is twisting towards the extracellular media and the extracellular part of helix C is moving towards helix G.

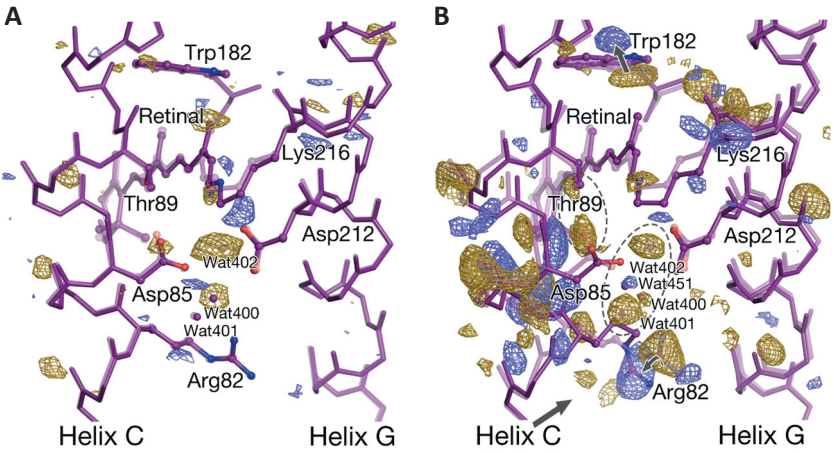


Figure 26. Recalculated difference Fourier electron density maps (**Paper II**), the intermediate structure (solid) and the resting state (transparent, as 1QHJ [77]).
A) A K-intermediate (1QKP) [88], with the map contoured at ± 3σ.
B) An L-intermediate from a 1:3:1 mixture of K:L:M (1E0P) [82], with the map countered at ± 3.2σ.

SVD-analysis of the main cluster from hierarchical clustering of structural perturbations (Figure 16) shows that L-, M- and O-states from several groups [76,78-86] represent an inward movement of the extracellular part of helix C towards helix G, as is seen in the L-state map above. In addition, the intracellular part of helix F is moving away from helices A, B and G.

When low-temperature and TR-SFX studies are compared in **Paper VI** we find that debated differences in trapping results in many cases describe different

real-time events. Thus, the combined set of low-temperature structures does not only describe major structural changes but also provides a more detailed picture than expected.

Early retinal isomerization

The earliest structural changes that occur in the bR photocycle and have been studied with TR-SFX are the ultrafast motions associated with the initial retinal isomerization. This is the focus of the sub-picosecond TR-SFX study [19]. Already within 450 fs from light-activation, the initially planar retinal starts sampling possible conformations. An interesting feature in the earliest difference density map (Figure 27) is a paired negative and positive difference density on Wat402 showing that the water is moving away from the SB. It is argued that the contact with the water molecule helps guiding the retinal isomerization to give a higher quantum yield of the *13-cis* conformer. The retinal isomerization occurs simultaneous with an ultrafast breathing-like motion of Asp85, Asp212, Tyr157, Wat400 and Wat401 away from the retinal and inwards again. The retinal fully adopts the *13-cis* conformation within 10 ps.

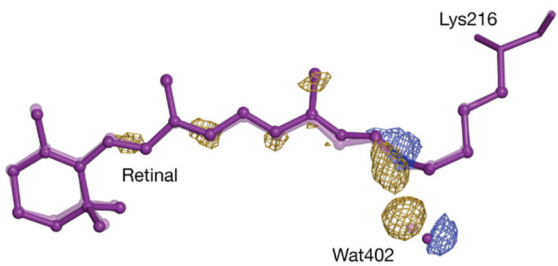


Figure 27. Early retinal isomerization in the time window Δt = 49 – 406 fs from TR-SFX data [19]. The F_{obs}^{light} - F_{obs}^{dark} difference Fourier electron density map indicate twisting of the retinal close to the SB, changed electron configurations along conjugated double bonds and a movement of Wat402 away from the retinal. Yellow represents negative difference electron density and blue positive. The electron density is contoured at ± 4.6σ.

Our reanalysis of the data reveals that the structural response to photo-activation is more complex than previously suggested. The representation of the difference density maps with the dual amplitude function of **Paper V** (Figure 23) shows how the motion that starts in the retinal is immediately affecting residues in helices C, F and G. The motion is then within 200-400 fs spread towards the more distant helices B and A with a speed of

approximately 30 Å/ps, which corresponds to the speed of sound in proteins as estimated from molecular dynamics simulations [108-110]. This ultrafast oscillatory motion is a direct effect of energy dissipating from the active center.

Early structural changes – the K and L intermediates

The nanosecond to millisecond TR-SFX study (**Paper IV**) covers the time-span of the three spectral intermediates K, L and M. The difference density map at Δt = 16 ns (Figure 28A), corresponding to the K-state intermediate, shows difference density amplitudes around the retinal C20 atom and the SB nitrogen. The retinal is initially in a twisted conformation that becomes planar within 290 ns. The isomerization creates a small cavity that allow Wat402 to start disordering. This is seen as a negative feature on Wat402 that increase in amplitude with time throughout the time window of the experiment, indicating that the mobility of the water increases further. Since there is no radiation damage in the TR-SFX data there are no doubt that the disordering is a light-induced change. The disordering breaks the initial hydrogen bonding network between the primary proton donor-acceptor pair; the protonated SB and Arg85.

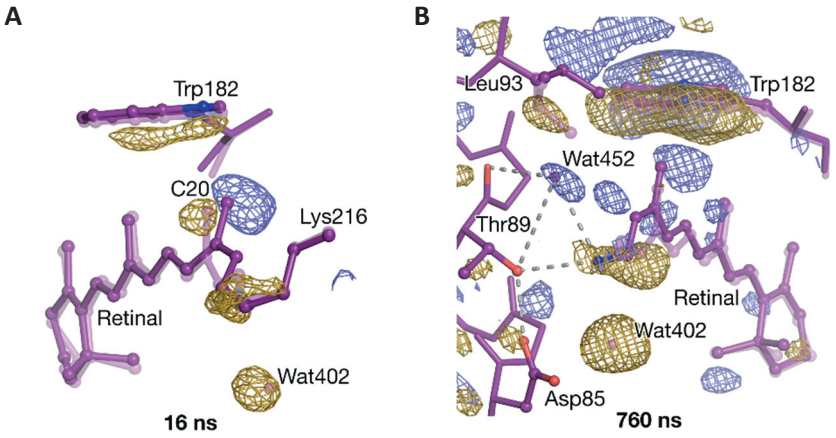


Figure 28. Structures and difference Fourier electron density maps of K and L intermediates. The intermediate state is drawn in solid purple, the resting state is transparent. Blue indicates positive and yellow negative difference density.
A) 16 ns, contoured at ± 3.7σ.
B) 760 ns, contoured at ± 3.1σ.

In the time interval from 40 ns to 13.8 µs conformational changes occur on the cytoplasmic side of the retinal. This is exemplified with the difference density map at Δt = 760 ns (Figure 28B) that we interpret as close to a pure L-state. Paired negative and positive difference densities around Leu93 and Trp182 indicate that the two side-chains shift away from the retinal, leaving an empty pocket. A positive difference feature appears transiently in this pocket, modelled as Wat452. The water creates a transient hydrogen bonding network from the SB to Asp85 via Thr89. Wat452 has earlier been modelled in a single trapped L-intermediate structure [89]. A water at this position is in agreement with observations from Fourier transform infrared (FTIR) spectroscopy that predict a transient hydrogen bond to the SB in the L-state [111,112].

Primary proton transfer – the transition from L to M

As described above, disordering of Wat402 in the K-state and ordering of Wat452 in the L-state change the hydrogen bonding network around the retinal. This in turn leads to additional rearrangement of waters and conformational changes in the extracellular part of the protein, for example seen at the time 13.8 µs (Figure 29).

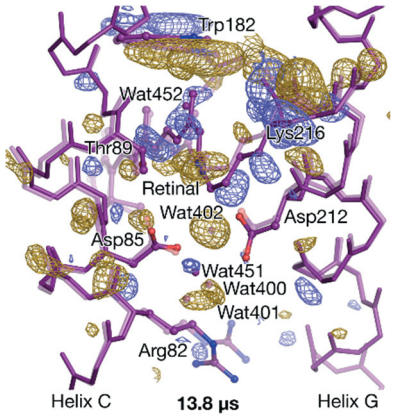


Figure 29. Structure and difference Fourier electron density maps at 13.8 µs, contoured at ± 3.7σ. The intermediate state is drawn in solid, the resting state is transparent. Blue indicates positive and yellow negative difference density.

As also known from previous trapping studies, Wat400 and Wat401 are displaced and another water molecule, Wat451, is ordering between Asp85 and Asp212. On the same time-scale the extracellular part of helix C is

bending towards helix G. We argue that this is the rate limiting step that is needed for transfer of the proton from the SB to Asp85 via Thr89 and Wat452. Computer simulations confirm that the proton transfer can occur via a water at this position [113]. After proton transfer the retinal is straightened and the distance to Asp85 is increased. Asp85 is interacting with Wat451 and the hydrogen bond to Thr89 is broken. The breakage is seen as a negative electron density feature already after 5.25 μs that increase in amplitude until the time 250 μs. This completely breaks the reverse path from Asp85 to the SB and prevents backward proton transfer from the M intermediate, ensuring efficient pumping of protons towards the outside of the cell.

The role of water molecules in proton pumping

Proton transport through bR is a stepwise reaction involving many key residues. The transport pathway from the cytoplasmic side include at least Asp96, the SB itself, Thr89, Asp85, Glu194 and Glu204. However, not all of the residues that are involved are situated close enough to allow direct passage of protons between them. When going through the different intermediate states of bR it is striking that specific water molecules have such a prominent role throughout the cycle (see insets in Figure 30). In the resting state Wat402, Wat400 and Wat401 stabilize the primary proton donor-acceptor pair, forcing the proton to remain strongly bound to the SB (inset 1). Wat402 has the role to guide the early retinal isomerisation towards formation of the K intermediate (inset 2). When Wat402 later gets disordered the initial hydrogen bonding network is broken and a new water, Wat452, is bound between Thr89 and the SB, contributing to a proton transfer pathway to Asp85 in the L intermediate (inset 3). Finally, Wat451 is ordering between Asp85, Arg82 and Asp212, contributing to the breakage of the hydrogen bond between Asp85 and Thr89 in the M intermediate (inset 4).

Helical movements

The sub-picosecond data reveal ultrafast movements of helices C, F and G that spread mainly to the extracellular side of helix B (Figure 23). All these initial movements appear damped to various extent within 10 ps. Later structural perturbations that occur in helices C, F and G are seen in the data from intermediate trapping, but the time scale and extent of these movements are better described in the ns-to-ms TR-SFX data (also shown in Figure 30; note that movements are four-fold exaggerated for visibility).

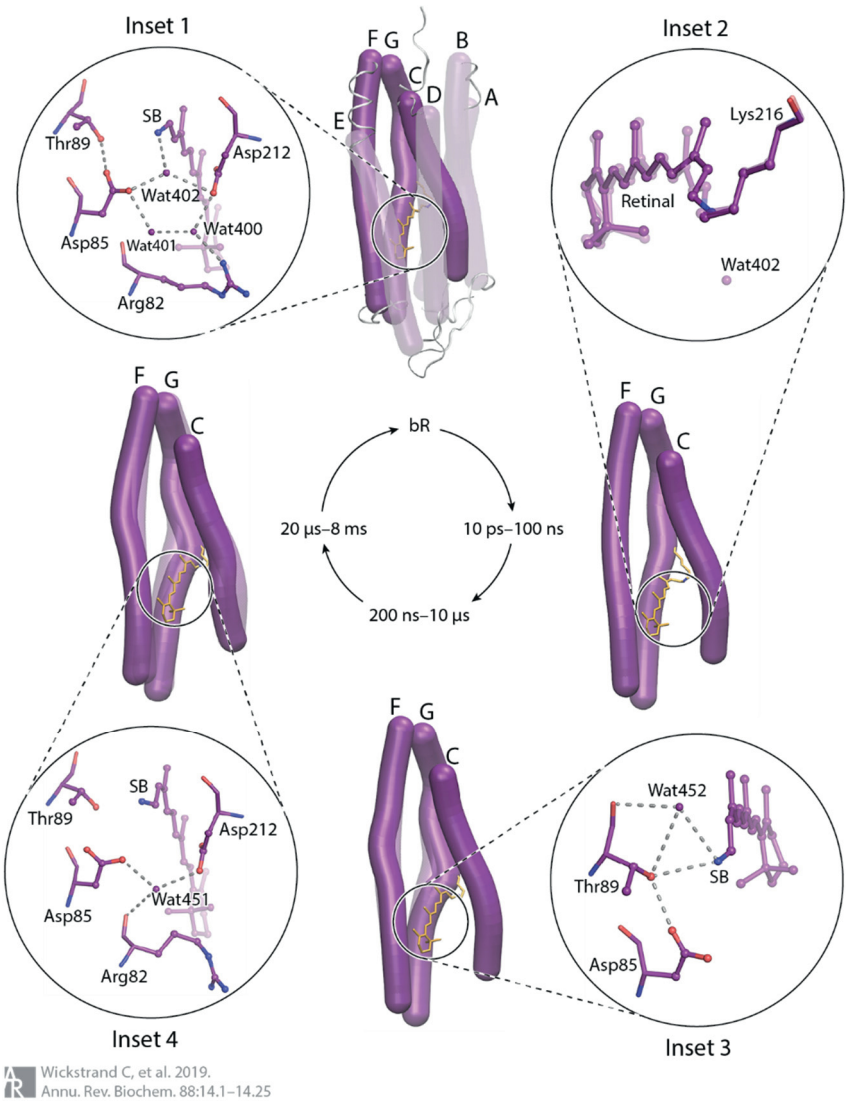


Figure 30. Schematic of the bR photocycle, showing helix movements and rearrangements of water molecules close to the retinal (insets). Major conformational changes occur in transmembrane helices C, F and G, that are superimposed upon their resting state conformation (shown in transparent). Helices A, B, D and E are only shown in the resting state. Helix movements are exaggerated four-fold to make them more visible. The structures were solved in the TR-SFX study in **Paper IV**.

The analysis in **Paper V** reveals that movements close to the retinal in helix G and in the cytoplasmic part of helix F are growing in amplitude in the time span from 110 to 760 ns. The inward flex of the extracellular part of helix C starts after approximately 2 µs and the amplitude is increasing until around 13.8-36.2 µs, which is in the time window when the primary proton transfer occurs. Overall, the helical movements are small; the maximum change in distance between C_α atoms of the 1.725 ms intermediate and the resting state is 1.5 Å (Figure 31). It has been predicted from structures of a triple-mutant (D96G/F171C/F219L) that large movements of helices E and F occur in the M-state [114,115]. These movements are seen only to a small extent in the TR-SFX studies or in the M-state models from intermediate trapping. One explanation could be that movements in the $P6_3$ crystal form are limited by crystal contacts between the EF-loop and the C-terminus (Paper IV, Figure S7). However, since the photocycle appear unaffected by these crystal contacts, large movements in helices E and F are not critical for the mechanism of proton pumping.

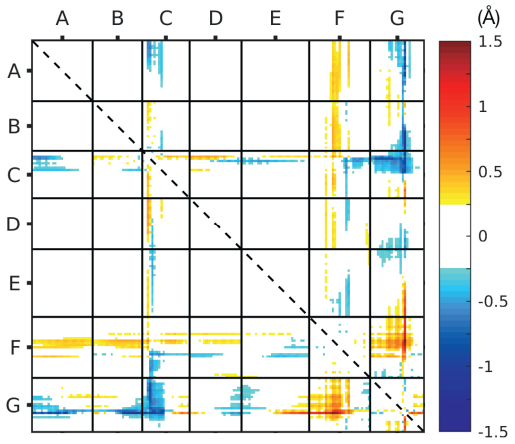


Figure 31. Difference internal distance matrix on C_α atoms at the time 1.725 ms (5B6Z-5B6V, **Paper IV**).

CHAPTER 5. CONCLUSIONS

5.1. Summary of the thesis

This thesis on chlorophyte channelrhodopsins and bacteriorhodopsin explores two different parts that are equally important in a TR-SFX study. The chapter on ChRs (**Paper I**) describes the work that proceeds data collection, including constructs design, protein expression, purification and crystallization. The chapter on bR (**Paper II-VI**) is focused on analysis and interpretation of structural models and difference Fourier electron density maps, the work that is performed after collection and processing of data.

In **Paper I,** we establish production and purification of ChRs in *P. pastoris* for a future TR-SFX experiment. We show that the channelrhodopsin chimaera C1C2 and wild-type ChR2 are glycosylated and handle this problem by enzymatic deglycosylation as well as expression of mutant forms. High yields of homogenous proteins are achieved through either deglycosylation of wild-type ChR2 or through expression of the mutant ChR2-N24Q. Initial LCP crystallization screens have been set up and we now have potential microcrystals of deglycosylated ChR2, that we will evaluate and optimize further.

Paper II gives a historical background on structural studies of bR. We review all published structures solved by X-ray crystallography at cryogenic temperature, by NMR or EM. To aid understanding of this large set of structures we also develop a clustering tool based on internal distance matrices. Clustering of resting states using this tool leads to a natural sorting of the models, mainly by crystal forms. Clustering of intermediate states let us extract models that show the same reoccurring conformational changes. This together with reanalysis of difference Fourier electron density maps reveal a high level of agreement between results from different research groups.

Paper III describes the first SMX-experiment on bR performed at a synchrotron. A room temperature structure is determined that show high similarity with previous resting state structures. This experiment was followed by the first time-resolved SFX-experiment on bR at an XFEL, which is described in **Paper IV**. Thirteen intermediate structures covering the time from 16 ns to 1.725 ms after light-activation are modelled and reveal new

insights in the proton pumping mechanism. For the first time, we can present the actual order of events and when they occur in real time. Rearrangements around the retinal in the K- and L-states reveal a novel path for the initial proton transfer reaction from the SB to Asp85 *via* Thr89. This together with a slow inward movement of helix C explain why the proton transfer event does not occur until the L-to-M transition.

In **Paper V,** we develop a tool for analyzing sets of difference Fourier electron density maps from TR-SFX experiments. The tool extracts information about positive and negative difference density amplitudes around every atomic position and represent these amplitudes as a trace of positive and negative values in a single one-dimensional plot per map. This simple representation aids identification of relevant density features and makes it easy to follow progression in time. It may be used for comparison of maps and validation of structural models via calculation of theoretical maps. Visualization and analysis of the sub-picosecond TR-SFX data [19] reveal temporary ultrafast helical movements that have not been identified before.

Paper VI is a review about TR-SFX experiments at XFELs with a focus on bR. Results from the two TR-SFX studies on bR (**Paper IV** and [19]) are compared to previous results from X-ray crystallography studies at cryogenic temperature. Again there is a high level of agreement but we find the TR-SFX technique superior for several reasons. It gives information about truly light-induced structural changes, no radiation damage, additional details and the actual order of events in real time.

5.2. Concluding remarks and future perspectives

Bacteriorhodopsin has been extensively studied for almost half a century and an essential part of this research has been within the field of structural biology. It has revealed details about retinal isomerization, reorganization of water molecules, helical movements and side-chain conformational changes that proceed and steer the proton transfer events. The research that is presented in this thesis adds an additional dimension to the picture; the dimension of time.

Channelrhodopsins are relatively new targets in structural biology research. The work described in this thesis brings ChR2 from *C. reinhardtii* closer

towards a TR-SFX experiment, that could hopefully explain how a protein so similar to bR can function as a cation channel.

Most TR-SFX studies have so far been performed on naturally light-driven systems but we see a rapid progress in the field and believe that new triggering techniques and increased XFEL availability will lead to an explosion of new insights in structural dynamics. This will put an increased demand on the analysis of large sets of time-resolved data. The analysis tools that we have implemented contribute to this development and to improved interpretation of protein structural dynamics.

Acknowledgements

When I started as a PhD student I could not at all imagine what the next eight years would bring. Over such a long time lots have changed, including projects, moods, people around me, my family and, not at least, myself. However, one thing has remained the same; I have never been alone on this journey.

First of all, I would like to thank **Richard Neutze**, my supervisor. You are inspiring, always full of ideas and probably the most optimistic person I have met. Maybe that explains why you keep achieving what you aim for. Thank you for letting me be a part of all this. You have been a mentor to me and I have learned a lot through all our discussions. I have always felt that you believe in me and that you respect me, also when we sometimes have had different views. I really appreciate all your support towards reaching my scientific goals, but I would also like to thank you for being open with that life is not only about science.

I also want to thank **Kristina Hedfalk**, my co-supervisor. Thank you for all your support in the channelrhodopsin project. Our discussions on cloning, expression and purification really helped me forward. **Gisela Brändén**, you came in at a later stage of the project. Thanks for supporting Giorgia and me when setting up crystal screens. I also thank **Martin Billeter** for being my examiner.

Several persons have contributed to the channelrhodopsin project. **Jennie Sjöhamn**, thank you for teaching me everything about Pichia and for your patience when answering my never-ending number of questions when I started working in the lab. **Tristan Wiessalla** and **Ferran Riera Boronat**, as bachelor students you came into the lab full of enthusiasm and really helped me a lot in cloning, expression and purification. I enjoyed introducing you to the field and having you in my team. **Giorgia Ortolani**, I am happy that you want to take the channelrhodopsin project further. Thank you for purifying and setting up crystal screens. Now we are getting somewhere. I wish you luck and hope that you let me know whenever you need me for discussing the project.

In the bacteriorhodopsin project I would like to acknowledge all our collaborators in the TR-SFX study and in the related analysis of data. I appreciate that I was invited to be part of this work. Especially I would like to

thank **Eriko Nango**, **Takanori Nakane**, **So Iwata**, **Przemyslaw Nogly** and **Jörg Standfuss**. Some persons were more involved in the development of tools and reanalysis of data; thanks to **Robert Dods**, **Antoine Royant** and **Gergely Katona**.

We are many groups at the division of Biochemistry and Structural Biology in the Lundberg laboratory. What is unique with this environment is that there are almost no boundaries between the groups and I am happy to have had so many friends around me. Thus, I would like to thank all my colleagues throughout the years. You have come running immediately whenever I needed help; discussed and tried to solve problems I have had; had pep-talks with me; and we have had the most wonderful, awkward, serious or completely unserious conversations during lunch- and fika-times.

Rebecka and **Petra**, I have had lots of fun with you and also learned so much and had support from you. **Rob**, thanks for lots of discussions and that you share my interest for rhodopsins and data analysis. **Erik**, **David**, **Maria**, **Johan** and **Anna L**, thanks for working together even though our work was not included in this theses. **Anna F** and **Viktor**, I have enjoyed sharing the office with you. To all the rest of **the old Lundberg crowd**, I have not mentioned all of you but I will remember many of you as special and I will miss you. To all **the new-comers**, sorry that my last year was so intense that I did not really have the chance to get to know you all, but I wish you good luck with your projects.

Tack till **min familj**, för att ni alltid finns där och stöttar mig. **Henrik**, tack att du fortfarande är samma fantastiska människa som när jag föll för dig för nästan 18 år sedan. Utan dig hade jag inte kunnat åstadkomma något i livet. Tack att du ger mig så mycket kärlek. Jag hoppas att det nu ska bli mer tid för oss igen, och för våra döttrar och den lille som ska komma. **Miranda** och **Vilma**, ni är det bästa som har hänt mig. Tack att ni varje dag påminner mig om vad som är viktigast i livet.

References

1 Wallin, E. & von Heijne, G., *Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms*. Protein Sci. **7** (4): 1029-1038 (1998).

2 Busch, W. & Saier, M. H., *The IUBMB-endorsed transporter classification system*. Mol. Biotechnol. **27** (3): 253-262 (2004).

3 Ernst, O. P. *et al.*, *Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms*. Chem Rev **114** (1): 126-163 (2014).

4 Govorunova, E. G., Sineshchekov, O. A., Li, H. & Spudich, J. L., *Microbial Rhodopsins: Diversity, Mechanisms, and Optogenetic Applications*. Annu. Rev. Biochem. **86**: 845-872 (2017).

5 Oesterhelt, D. & Stoeckenius, W., *Rhodopsin-like protein from the purple membrane of Halobacterium halobium*. Nat New Biol **233** (39): 149-152 (1971).

6 Lindley, E. V. & Macdonald, R. E., *A Second Mechanism for Sodium Extrusion in Halobacterium-Halobium: a Light-Driven Sodium-Pump*. Biochem. Biophys. Res. Commun. **88** (2): 491-499 (1979).

7 Spudich, E. N., Sundberg, S. A., Manor, D. & Spudich, J. L., *Properties of a second sensory receptor protein in Halobacterium halobium phototaxis*. Proteins **1** (3): 239-246 (1986).

8 Spudich, J. L. & Bogomolni, R. A., *Mechanism of Color Discrimination by a Bacterial Sensory Rhodopsin*. Nature **312** (5994): 509-513 (1984).

9 Pinhassi, J., DeLong, E. F., Beja, O., Gonzalez, J. M. & Pedros-Alio, C., *Marine Bacterial and Archaeal Ion-Pumping Rhodopsins: Genetic Diversity, Physiology, and Ecology*. Microbiol. Mol. Biol. Rev. **80** (4): 929-954. (2016).

10 Fuhrman, J. A., Schwalbach, M. S. & Stingl, U., *Proteorhodopsins: an array of physiological roles?* Nat. Rev. Microbiol. **6** (6): 488-494 (2008).

11 Gomez-Consarnau, L. *et al.*, *Proteorhodopsin light-enhanced growth linked to vitamin-B1 acquisition in marine Flavobacteria*. ISME J **10** (5): 1102-1112 (2016).

12 Lorenz-Fonfria, V. A. & Heberle, J., *Channelrhodopsin unchained: structure and mechanism of a light-gated cation channel*. Biochim. Biophys. Acta **1837** (5): 626-642 (2014).

13 Schneider, F., Grimm, C. & Hegemann, P., *Biophysics of Channelrhodopsin*. Annu Rev Biophys **44**: 167-186 (2015).

14 Soppa, J., *From genomes to function: haloarchaea as model organisms*. Microbiology-Sgm **152**: 585-590 (2006).
15 Oesterhelt, D. & Stoeckenius, W., *Functions of a new photoreceptor membrane*. Proc Natl Acad Sci U S A **70** (10): 2853-2857 (1973).
16 Henderson, R. & Unwin, P. N., *Three-dimensional model of purple membrane obtained by electron microscopy*. Nature **257** (5521): 28-32 (1975).
17 Henderson, R. *et al.*, *Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy*. J. Mol. Biol. **213** (4): 899-929 (1990).
18 Berman, H. M. *et al.*, *The Protein Data Bank*. Nucleic Acids Res. **28** (1): 235-242 (2000).
19 Nogly, P. *et al.*, *Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser*. Science **361**: 6398 (2018).
20 Haupts, U., Tittor, J. & Oesterhelt, D., *Closing in on bacteriorhodopsin: Progress in understanding the molecule*. Annu. Rev. Biophys. Biomol. Struct. **28**: 367-399 (1999).
21 Lanyi, J. K., *Mechanism of ion transport across membranes - Bacteriorhodopsin as a prototype for proton pumps*. J. Biol. Chem. **272** (50): 31209-31212 (1997).
22 Neutze, R. *et al.*, *Bacteriorhodopsin: a high-resolution structural view of vectorial proton transport*. BBA Biomembranes **1565** (2): 144-167 (2002).
23 Nagel, G. *et al.*, *Channelrhodopsin-1: a light-gated proton channel in green algae*. Science **296** (5577): 2395-2398 (2002).
24 Nagel, G. *et al.*, *Channelrhodopsin-2, a directly light-gated cation-selective membrane channel*. Proc Natl Acad Sci U S A **100** (24): 13940-13945 (2003).
25 Berthold, P. *et al.*, *Channelrhodopsin-1 initiates phototaxis and photophobic responses in Chlamydomonas by immediate light-induced depolarization*. Plant Cell **20** (6): 1665-1677 (2008).
26 Erickson, E., Wakao, S. & Niyogi, K. K., *Light stress and photoprotection in Chlamydomonas reinhardtii*. Plant J. **82** (3): 449-465 (2015).
27 Kim, J. I. *et al.*, *Evolutionary Dynamics of Cryptophyte Plastid Genomes*. Genome Biol Evol **9** (7): 1859-1872 (2017).
28 Sineshchekov, O. A., Govorunova, E. G., Li, H. & Spudich, J. L., *Bacteriorhodopsin-like channelrhodopsins: Alternative mechanism for control of cation conductance*. Proc Natl Acad Sci U S A **114** (45): E9512-E9519 (2017).

29 Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K., *Millisecond-timescale, genetically targeted optical control of neural activity*. Nat. Neurosci. **8** (9): 1263-1268 (2005).
30 Deisseroth, K., *Optogenetics: 10 years of microbial opsins in neuroscience*. Nat. Neurosci. **18** (9): 1213-1225 (2015).
31 Schroll, C. *et al.*, *Light-induced activation of distinct modulatory neurons triggers appetitive or aversive learning in Drosophila larvae*. Curr. Biol. **16** (17): 1741-1747 (2006).
32 Simpson, J. H. & Looger, L. L., *Functional Imaging and Optogenetics in Drosophila*. Genetics **208** (4): 1291-1309 (2018).
33 Aravanis, A. M. *et al.*, *An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology*. J Neural Eng **4** (3): S143-156 (2007).
34 Galvan, A., Caiola, M. J. & Albaugh, D. L., *Advances in optogenetic and chemogenetic methods to study brain circuits in non-human primates*. J. Neural Transm. **125** (3): 547-563 (2018).
35 Chen, I. W., Papagiakoumou, E. & Emiliani, V., *Towards circuit optogenetics*. Curr. Opin. Neurobiol. **50**: 179-189 (2018).
36 Saita, M. *et al.*, *Photoexcitation of the P4(480) State Induces a Secondary Photocycle That Potentially Desensitizes Channelrhodopsin-2*. J. Am. Chem. Soc. **140** (31): 9899-9903 (2018).
37 Bamann, C., Kirsch, T., Nagel, G. & Bamberg, E., *Spectral characteristics of the photocycle of channelrhodopsin-2 and its implication for channel function*. J. Mol. Biol. **375** (3): 686-694 (2008).
38 Kato, H. E. *et al.*, *Crystal structure of the channelrhodopsin light-gated cation channel*. Nature **482** (7385): 369-374 (2012).
39 Kato, H. E. *et al.*, *Atomistic design of microbial opsin-based blue-shifted optogenetics tools*. Nat. Commun. **6**: 7177 (2015).
40 Volkov, O. *et al.*, *Structural insights into ion conduction by channelrhodopsin 2*. Science **358**: 6366 (2017).
41 Oda, K. *et al.*, *Crystal structure of the red light-activated channelrhodopsin Chrimson*. Nat. Commun. **9** (1): 3949 (2018).
42 Li, H. *et al.*, *Crystal structure of a natural light-gated anion channelrhodopsin*. Elife **8**: e41741 (2019).
43 Kato, H. E. *et al.*, *Structural mechanisms of selectivity and gating in anion channelrhodopsins*. Nature **561** (7723): 349-354 (2018).
44 Mittelmeier, T. M., Boyd, J. S., Lamb, M. R. & Dieckmann, C. L., *Asymmetric properties of the Chlamydomonas reinhardtii cytoskeleton direct rhodopsin photoreceptor localization*. J. Cell Biol. **193** (4): 741-753 (2011).

45 Feldbauer, K. *et al.*, *Channelrhodopsin-2 is a leaky proton pump*. Proc. Natl. Acad. Sci. USA **106** (30): 12317-12322 (2009).

46 Elena, C., Ravasi, P., Castelli, M. E., Peiru, S. & Menzella, H. G., *Expression of codon optimized genes in microbial systems: current industrial applications and perspectives*. Front Microbiol **5**: 21 (2014).

47 Lyons, J. A., Shahsavar, A., Paulsen, P. A., Pedersen, B. P. & Nissen, P., *Expression strategies for structural studies of eukaryotic membrane proteins*. Curr. Opin. Struct. Biol. **38**: 137-144 (2016).

48 Tate, C. G. *et al.*, *Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter*. BBA Biomembranes **1610** (1): 141-153 (2003).

49 Hedfalk, K., *Further advances in the production of membrane proteins in Pichia pastoris*. Bioengineered **4** (6): 363-367 (2013).

50 Juturu, V. & Wu, J. C., *Heterologous Protein Expression in Pichia pastoris: Latest Research Progress and Applications*. ChemBioChem **19** (1): 7-21 (2018).

51 Öberg, F., Sjöhamn, J., Conner, M. T., Bill, R. M. & Hedfalk, K., *Improving recombinant eukaryotic membrane protein yields in Pichia pastoris: the importance of codon optimization and clone selection*. Mol. Membr. Biol. **28** (6): 398-411 (2011).

52 Luckey, M. *Tools for studying membrane components: detergents and model systems*. in Membrane structural biology with biochemical and biophysical foundations Ch. 3: 42-48 (Cambridge University Press, 2008).

53 Prive, G. G., *Detergents for the stabilization and crystallization of membrane proteins*. Methods **41** (4): 388-397 (2007).

54 Rhodes, G. *Crystallography Made Crystal Clear*. 3 edn: (Academic Press, 2006).

55 Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J., *Potential for biomolecular imaging with femtosecond X-ray pulses*. Nature **406** (6797): 752-757 (2000).

56 Stanley P, T. N., Aebi M. *N-Glycans*. in Essentials of Glycobiology (ed Cummings RD Varki A, Esko JD, et al.,) Ch. 9, (Cold Spring Harbor Laboratory Press, 2017).

57 Bretthauer, R. K. & Castellino, F. J., *Glycosylation of Pichia pastoris-derived proteins*. Biotechnol. Appl. Biochem. **30**: 193-200 (1999).

58 Freeze, H. H., *Endoglycosidase and glycoamidase release of N-linked oligosaccharides*. Curr Protoc Mol Biol: Ch 17: 13A (2001).

59 New England Biolabs. *Endo Hf* retrieved from <https://international.neb.com/products/p0703-endo-hf#Product%20Information>, 17 April, 2019.

60 Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M., *Detergent binding explains anomalous SDS-PAGE migration of membrane proteins*. Proc. Natl. Acad. Sci. USA **106** (6): 1760-1765 (2009).

61 Bateman, A. *et al.*, *UniProt: a worldwide hub of protein knowledge*. Nucleic Acids Res. **47** (D1): D506-D515 (2019).

62 Krause, N. *Structural rearrangements upon opening of Channelrhodopsin-2 (PhD thesis)*, Freie Universität Berlin, (2016).

63 Hou, S. Y. *Homology Cloning, Heterologous Expression and Characterization of a New Channelrhodopsin (PhD thesis)*, Graduate School of Biomedical Sciences, (2012).

64 Kirsch, T. *Funktionelle Expression von Channelrhodopsin 2 (ChR2) in der methylotrophen Hefe Pichia pastoris und biophysikalische Charakterisierung (PhD thesis)*, Goethe-Universität, Frankfurt am Main, (2007).

65 Bruun, S. *Resonance Raman spectroscopy on microbial rhodopsins (PhD thesis)*, die Technische Universität Berlin, (2013).

66 Holm, L. & Sander, C., *Dali - a Network Tool for Protein-Structure Comparison*. Trends Biochem. Sci. **20** (11): 478-480 (1995).

67 Zhu, J. H. & Weng, Z. P., *FAST: A novel protein structure alignment algorithm*. Proteins: Struct. Funct. Bioinform. **58** (3): 618-627 (2005).

68 Orengo, C. A. *et al.*, *CATH - a hierarchic classification of protein domain structures*. Structure **5** (8): 1093-1108 (1997).

69 Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C., *Scop - a Structural Classification of Proteins Database for the Investigation of Sequences and Structures*. J. Mol. Biol. **247** (4): 536-540 (1995).

70 Holm, L., Ouzounis, C., Sander, C., Tuparev, G. & Vriend, G., *A Database of Protein-Structure Families with Common Folding Motifs*. Protein Sci. **1** (12): 1691-1698 (1992).

71 Hadley, C. & Jones, D. T., *A systematic comparison of protein structure classifications: SCOP, CATH and FSSP*. Structure with Folding & Design **7** (9): 1099-1112 (1999).

72 Schneider, T. R., *Objective comparison of protein structures: error-scaled difference distance matrices*. Acta Crystallogr D Biol Crystallogr **56** (Pt 6): 714-721 (2000).

73 Domingues, F. S., Rahnenfuhrer, J. & Lengauer, T., *Automated clustering of ensembles of alternative models in protein structure databases*. Protein Eng. Des. Sel. **17** (6): 537-543 (2004).

74 Cruickshank, D. W., *Remarks about protein structure precision*. Acta Crystallogr D Biol Crystallogr **55** (Pt 3): 583-601 (1999).

75 *MATLAB, The Mathworks, Inc., Natick, Massachusetts, United States,* <http://www.mathworks.se/products/matlab/>.

76 Lanyi, J. K. & Schobert, B., *Structural changes in the L photointermediate of bacteriorhodopsin*. J. Mol. Biol. **365** (5): 1379-1392 (2007).

77 Belrhali, H. *et al.*, *Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 A resolution*. Structure **7** (8): 909-917 (1999).

78 Edman, K. *et al.*, *Deformation of helix C in the low temperature L-intermediate of bacteriorhodopsin*. J. Biol. Chem. **279** (3): 2147-2158 (2004).

79 Facciotti, M. T. *et al.*, *Structure of an early intermediate in the M-state phase of the bacteriorhodopsin photocycle*. Biophys. J. **81** (6): 3442-3455 (2001).

80 Lanyi, J. K. & Schobert, B., *Propagating structural perturbation inside bacteriorhodopsin: crystal structures of the M state and the D96A and T46V mutants*. Biochemistry **45** (39): 12003-12010 (2006).

81 Luecke, H. *et al.*, *Coupling photoisomerization of retinal to directional transport in bacteriorhodopsin*. J. Mol. Biol. **300** (5): 1237-1255 (2000).

82 Royant, A. *et al.*, *Helix deformation is coupled to vectorial proton transport in the photocycle of bacteriorhodopsin*. Nature **406** (6796): 645-648 (2000).

83 Sass, H. J. *et al.*, *Structural alterations for proton translocation in the M state of wild-type bacteriorhodopsin*. Nature **406** (6796): 649-653 (2000).

84 Takeda, K. *et al.*, *Crystal structure of the M intermediate of bacteriorhodopsin: allosteric structural changes mediated by sliding movement of a transmembrane helix*. J. Mol. Biol. **341** (4): 1023-1037 (2004).

85 Yamamoto, M., Hayakawa, N., Murakami, M. & Kouyama, T., *Crystal Structures of Different Substates of Bacteriorhodopsin's M Intermediate at Various pH Levels*. J. Mol. Biol. **393** (3): 559-573 (2009).

86 Zhang, J. *et al.*, *Crystal structure of the O intermediate of the Leu93-->Ala mutant of bacteriorhodopsin*. Proteins **80** (10): 2384-2396 (2012).

87 RSCB PDB. *Newsletter, number 36* (2008), retrieved from <https://cdn.rcsb.org/rcsb-pdb/general_information/news_publications/newsletters/2007q4/deposition.html>, 19 March, 2019.

88 Edman, K. *et al.*, *High-resolution X-ray structure of an early intermediate in the bacteriorhodopsin photocycle*. Nature **401** (6755): 822-826 (1999).

89 Kouyama, T., Nishikawa, T., Tokuhisa, T. & Okumura, H., *Crystal structure of the L intermediate of bacteriorhodopsin: Evidence for vertical translocation of a water molecule during the proton pumping cycle*. J. Mol. Biol. **335** (2): 531-546 (2004).

90 Ursby T, B. D., *Improved Estimation of Structure-Factor Difference Amplitudes from Poorly Accurate Data*. Acta Cryst. **A53**: 564-575 (1997).

91 Weierstall, U. *et al.*, *Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography*. Nat. Commun. **5**: 12314 (2014).

92 Weik, M. *et al.*, *Specific chemical and structural damage to proteins produced by synchrotron radiation*. Proc Natl Acad Sci U S A **97** (2): 623-628 (2000).

93 Borshchevskiy, V. I., Round, E. S., Popov, A. N., Buldt, G. & Gordeliy, V. I., *X-ray-Radiation-Induced Changes in Bacteriorhodopsin Structure*. J. Mol. Biol. **409** (5): 813-825 (2011).

94 Matsui, Y. *et al.*, *Specific damage induced by X-ray radiation and structural changes in the primary photoreaction of bacteriorhodopsin*. J. Mol. Biol. **324** (3): 469-481 (2002).

95 Emma, P. *et al.*, *First lasing and operation of an angstrom-wavelength free-electron laser*. Nat. Photonics. **4** (9): 641-647 (2010).

96 Tono, K. *et al.*, *Beamline, experimental stations and photon beam diagnostics for the hard x-ray free electron laser of SACLA*. New J Phys. **15** (8): 083035 (2013).

97 DePonte, D. P. *et al.*, *Gas dynamic virtual nozzle for generation of microscopic droplet streams*. J Phys D Appl Phys. **41** (19): 5505 (2008).

98 Weierstall, U., Spence, J. C. H. & Doak, R. B., *Injector for scattering measurements on fully solvated biospecies*. Rev Sci Instrum **83** (3): 035108 (2012).

99 Hart, P. *et al.*, *The CSPAD megapixel x-ray camera at LCLS*. X-Ray Free-Electron Lasers: Beam Diagnostics, Beamline Instrumentation, and Applications **8504**: 85040C (2012).

100 Struder, L. *et al.*, *Large-format, high-speed, X-ray pnCCDs combined with electron and ion imaging spectrometers in a multipurpose chamber for experiments at 4th generation light sources*. Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment **614** (3): 483-496 (2010).

101 Barty, A. *et al.*, *Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data*. J. Appl. Crystallogr. **47**: 1118-1131 (2014).
102 White, T. A. *et al.*, *CrystFEL: a software suite for snapshot serial crystallography*. J. Appl. Crystallogr. **45**: 335-341 (2012).
103 Nogly, P. *et al.*, *Lipidic cubic phase injector is a viable crystal delivery system for time-resolved serial crystallography*. Nat. Commun. **7**: 12314 (2016).
104 Andersson, M. *et al.*, *Structural Dynamics of Light-Driven Proton Pumps*. Structure **17** (9): 1265-1275 (2009).
105 Winn, M. D. *et al.*, *Overview of the CCP4 suite and current developments*. Acta Crystallogr D Struct Biol. **67**: 235-242 (2011).
106 Stein, P. E. *et al.*, *The Crystal-Structure of Pertussis Toxin*. Structure **2** (1): 45-57 (1994).
107 Schobert, B., Cupp-Vickery, J., Hornak, V., Smith, S. O. & Lanyi, J. K., *Crystallographic structure of the K intermediate of bacteriorhodopsin: Conservation of free energy after photoisomerization of the retinal*. J. Mol. Biol. **321** (4): 715-726 (2002).
108 Xu, Z. P. & Buehler, M. J., *Mechanical energy transfer and dissipation in fibrous beta-sheet-rich proteins*. Physical Review E **81** (6): 061910 (2010).
109 Young, H. T., Edwards, S. A. & Grater, F., *How Fast Does a Signal Propagate through Proteins?* Plos One **8** (6): e64746 (2013).
110 Yu, X. & Leitner, D. M., *Vibrational energy transfer and heat conduction in a protein*. J. Phys. Chem. B **107** (7): 1698-1707 (2003).
111 Maeda, A. *et al.*, *Interaction of internal water molecules with the Schiff base in the L intermediate of the bacteriorhodopsin photocycle*. Biochemistry **41** (11): 3803-3809 (2002).
112 Maeda, A., Tomson, F. L., Gennis, R. B., Balashov, S. P. & Ebrey, T. G., *Water molecule rearrangements around Leu93 and Trp182 in the formation of the L intermediate in bacteriorhodopsin's photocycle*. Biochemistry **42** (9): 2535-2541 (2003).
113 Bondar, A. N., Baudry, J., Suhai, S., Fischer, S. & Smith, J. C., *Key Role of Active-Site Water Molecules in Bacteriorhodopsin Proton-Transfer Reactions*. J. Phys. Chem. B **112** (47): 14729-14741 (2008).
114 Wang, T. *et al.*, *Deprotonation of D96 in Bacteriorhodopsin Opens the Proton Uptake Pathway*. Structure **21** (2): 290-297 (2013).
115 Subramaniam, S. & Henderson, R., *Molecular mechanism of vectorial proton translocation by bacteriorhodopsin*. Nature **406** (6796): 653-657 (2000).

ISBN 978-91-7833-492-6 (TRYCKT)
ISBN 978-91-7833-493-3 (PDF)