UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Physique pour les sciences du vivant**

Arrêté ministériel : 7 août 2006

Présentée par

Ivan GUSHCHIN

Thèse dirigée par **Prof. Dr. Valentin GORDELIY**
codirigée par **Dr. Sergei GRUDININ**

préparée au sein du **Laboratoire des Transporteurs Membranaires de l'Institut de Biologie Structurale**
dans **l'École Doctorale de Physique**

Études structurales des rhodopsines microbiennes et des autres protéines membranaires au moyen de la cristallographie aux rayons X et de la modélisation informatique

Thèse soutenue publiquement le **5 Septembre 2014**,
devant le jury composé de :

Prof. Dr. Valentin GORDELIY **Directeur de thèse**
Directeur de Recherche CEA, Institut de Biologie Structurale, Grenoble, France

Dr. Sergei GRUDININ **Co-encadrant de thèse**
Chargé de Recherche CNRS, Inria, Grenoble, France

Prof. Dr. Georg BÜLDT **Rapporteur**
Head of the Laboratory of Prospective Studies of Membrane Proteins, Moscow Institute of Physics and Technology, Moscow, Russia

Prof. Dr. Dieter WILLBOLD **Rapporteur**
Director of the Institute of Complex Systems-6, Structural Biochemistry, Research Center Jülich, Jülich, Germany

Prof. Dr. Martin ENGELHARD **Examinateur**
Research group leader at the Max Planck Institute of Molecular Physiology, Dortmund, Germany

Prof. Dr. Eva PEBAY-PEYROULA **President du jury**
Directrice de l'Institut de Biologie Structurale, Grenoble, France

Université Joseph Fourier / Université Pierre Mendès France / Université Stendhal / Université de Savoie / Grenoble INP

Contents

Abbreviations

A.f., Archaeoglobus fulgidus

E. coli, Escherichia coli

H.m., Haloarcula marismortui

H.s., Halobacterium salinarum

N.p., Natronomonas pharaonis

BR, bacteriorhodopsin

ChR, channel rhodopsin

ESR, *Exiguobacterium sibiricum* rhodopsin

GR, *Gloeobacter* rhodopsin

HR, halorhodopsin

KR2, *Krokinobacter* rhodopsin 2

PR, proteorhodopsin

SR, sensory rhodopsin

XR, xanthorhodopsin

HAMP domain, histidine kinases, adenylyl cyclases, methyl-accepting chemotaxis proteins and phosphatases domain

IPCT, inositol-phosphate cytidylyltransferase

DIPPS, di-inositol-phosphate-phosphate synthase

MD, molecular dynamics

MR, molecular replacement

RMSD, root mean square deviation

UV, ultra-violet (light)

This page is intentionally left blank

Abstract

Every living cell on Earth is surrounded by a lipid membrane. Proteins residing in the membrane perform a variety of functions crucial for the cell's survival. Among them are active and passive transport in and out of the cell, signaling and reaction catalysis.

One of the largest membrane protein families are microbial rhodopsins, which utilize light energy for their function. Members of this family count among them light-driven proton, cation and anion pumps, light-gated ion channels and photoreceptors. While the basic aspects of their functioning have been known for some time, there is a plenty of unanswered questions. In this dissertation, several structures of microbial rhodopsins (among them the first proteorhodopsin structure and the first light-driven sodium pump structure) are presented and analyzed. The structures open the way for understanding the similarities and differences between the various microbial rhodopsins and for exploiting this understanding to create better microbial rhodopsin-based instruments for biological applications, for example, in the field of optogenetics.

While the first part of this work deals with the novel structures of microbial rhodopsins, the second part presents the simulation approach for understanding the sensory rhodopsin-based signaling in phototaxis. The HAMP domains of the sensory rhodopsin transducer protein are studied by means of molecular dynamics, and it is demonstrated that the simulations may be used for building and validating the atomic structures of signaling domains, as well as for understanding the signaling-associated conformational changes, initiated by light-driven sensory rhodopsin transformations.

The third and the last part describes the work on the *Archaeoglobus fulgidus* IPCT-DIPPS proteins, an enzyme catalyzing two consecutive steps of di-inositol-phosphate biosynthesis. The determined structure may serve as a model for understanding the catalytic mechanism of CDP-alcohol transferases, a large family of proteins counting thousands of members, among which are five human proteins that catalyze the major steps of lipid biosynthesis. The structure was also used to predict the binding sites of the ligands at the enzyme active site and to propose the mechanism of catalytic action.

To sum up, this dissertation presents the structural studies of various membrane proteins by means of X-ray crystallography and modeling that advance our understanding of fundamental and practical aspects of membrane protein functioning.

The text of the manuscript is partially based on the published works[1–4]:

Section 3.2 is an adaptation of the article "Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria" (Proceedings of the National Academy of Sciences of the USA, 2013)[1].

Section 3.4 is an adaptation of the article "Active State of Sensory Rhodopsin II: Structural Determinants for Signal Transfer and Proton Pumping" (Journal of Molecular Biology, 2011)[3].

Section 3.5 is an adaptation of the article "Role of the HAMP Domain Region of Sensory Rhodopsin Transducers in Signal Transduction" (Biochemistry, 2011)[4].

Section 3.6 is an adaptation of the article "Two Distinct States of the HAMP Domain from Sensory Rhodopsin Transducer Observed in Unbiased Molecular Dynamics Simulations" (PLoS ONE, 2013)[2].

Finally, section 3.7 is partially adapted from the manuscript "X-ray Structure of a CDP-alcohol phosphatidyltransferase membrane enzyme and insights into its catalytic mechanism" (Nature Communications, accepted).

Résumé

Chaque cellule vivante sur notre Terre est entourée d'une membrane lipidique. Les protéines résidant dans la membrane exécutent multitude de fonctions essentielles pour la survivance de la cellule. Parmi eux sont le transport actif et passif dans et hors de la cellule, la signalisation et la catalyse des réactions.

Une des plus grandes familles de protéines membranaires sont rhodopsins microbiennes, qui utilisent l'énergie de la lumière pour leur fonction. Les membres de cette famille comptent parmi eux les pompes de protons, cations et anions, entraînée par l'illumination, les canaux ioniques activés par l'illumination et, finalement, photorécepteurs. Bien que les aspects fondamentaux de leur fonctionnement ont été connus depuis un certain temps, il ya une abondance de questions sans réponse. Dans cette thèse, plusieurs structures de rhodopsines microbiennes (y compris la première structure de protéorhodopsine et la première structure de la pompe à sodium) sont présentés et analysés. Les structures ouvrent la voie pour comprendre les similitudes et les différences entre les différents rhodopsines microbiennes et pour exploiter cette connaissance pour créer de meilleurs instruments à base de rhodopsines microbiennes pour des applications biologiques, par exemple, dans le domaine de optogenetics.

Alors que la première partie de ce travail porte sur les nouvelles structures de rhodopsines microbiennes, la deuxième partie présente l'approche de simulation pour comprendre la signalisation en fonction des rhodopsines sensorielles dans phototaxie. Les domaines HAMP des protéines transductrices des signals des rhodopsines sensorielles sont étudiés au moyen de la dynamique moléculaire, et il est démontré que les simulations peuvent être utilisés pour la construction et la validation des structures atomiques des domaines de signalisation, ainsi que pour la compréhension des changements conformationnels associée à signalisation, initié par les transformations des rhodopsine sensorielles.

La troisième et la dernière partie décrit le travail sur la protéine IPCT-DIPPS de *Archaeoglobus fulgidus*, une enzyme catalysant deux étapes consécutives de di-inositol-phosphate biosynthèse. La structure résolue peut servir de modèle pour comprendre le mécanisme catalytique de transférases CDP-alcool, une grande famille de protéines comptant des milliers de membres, parmi lesquels sont cinq protéines humaines, qui catalysent les étapes majeures de la biosynthèse des lipides. La structure a également été utilisé pour prédire

les sites de liaison des ligands sur le site actif de l'enzyme et pour proposer le mécanisme d'action catalytique.

Pour résumer, cette thèse présente les études structurales de diverses protéines membranaires par la cristallographie aux rayons X et la modélisation qui font progresser notre compréhension des aspects fondamentaux et pratiques de fonctionnement des protéines membranaires.

1 Introduction

1.1 Microbial rhodopsins

1.1.1 Description of the family

Microbial rhodopsins are a large family of photoactive membrane proteins present in bacteria, archaea, eukaryota and viruses[5]. The proteins are composed of seven transmembrane α-helices, labeled A to G, three intracellular and three extracellular loops (Figure 1.1.1a). Among them, the B-C loop is usually the longest one and its conformation is varied the most among the different microbial rhodopsins. The unifying feature of microbial rhodopsins is the cofactor retinal (Figure 1.1.1b) that is covalently bound via the Schiff base to the side chain of a lysine residue situated in the middle of the helix G. The retinal imparts the microbial rhodopsins the ability to utilize light for their function.

1.1.2 Bacteriorhodopsin: historical perspective

The first member of the microbial rhodopsins family, and the most studied to date is the light-driven proton pump bacteriorhodopsin from *Halobacterium salinarum*[6]. It was discovered in 1971 as the major constituent of the *H.s.* purple membrane[7]. Due to its high abundance in the *H.s.* membrane and ease of purification, BR can be produced in large quantities. This property made it the model membrane protein, extensively used for the development of numerous biophysical techniques. One of the first applications of bacteriorhodopsin was in providing the evidence for the Mitchell's "chemiosmotic theory", according to which the proton gradient across the cell membrane can be utilized for ATP production[8]. Later, bacteriorhodopsin served as a model system for development of the electron microscopy techniques, which provided the first glimpse into its structure at the resolution of 7 Å[9]. After this, the resolution was gradually improved to almost-atomic[10,11].

At the same time, bacteriorhodopsin was extensively studied using different optical techniques such as UV-Visual[12,13], FTIR[14] and resonance Raman[15] spectroscopy. It was determined that upon absorption of a light photon the visual light absorption spectrum of bacteriorhodopsin undergoes a series of transformations, called photocycle (Figure 1.1.2a). Based on the absorption spectra, presence of several interconverting quasi-stable states, called K, L, M, N and O (Figure 1.1.2b), was postulated. In the M-state the retinal is in the 13-*cis* state and deprotonated, and, consequently, the M-state spectrum differs the most from the

ground state spectrum. Apart from the main photocycle, there are also states P and Q that can be induced by second photon absorption. Finally, in the absence of illumination, thermal isomerization of the retinal from the all-*trans* to 13-*cis* 15-*syn* conformation is possible that gives rise to a so-called "dark" photocycle, which doesn't result in a proton translocation.

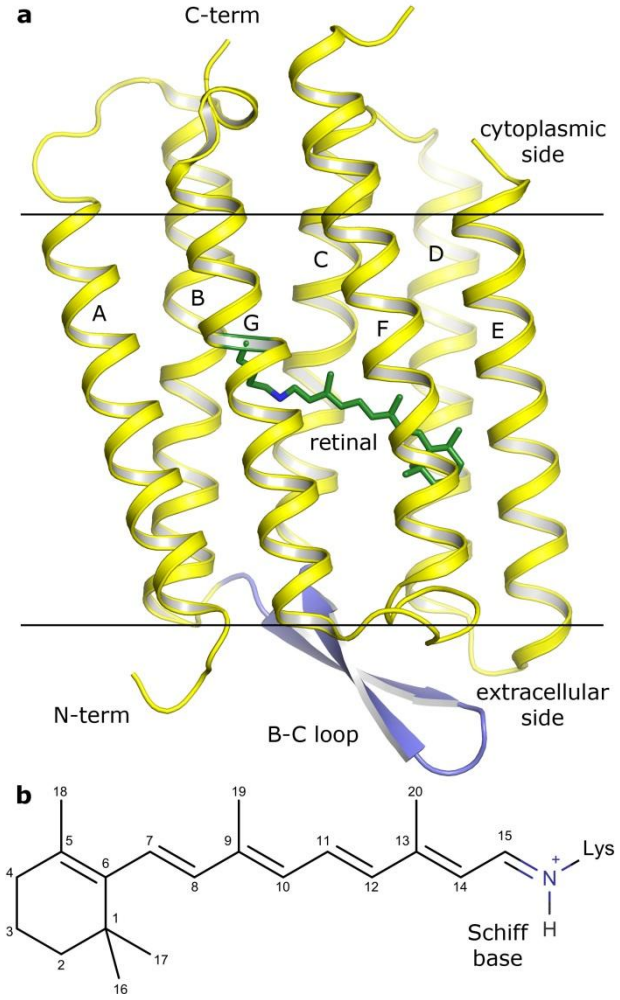

Figure 1.1.1. General features of the microbial rhodopsin structures. a) Microbial rhodopsin fold, exemplified by *Halobacterium salinarum* bacteriorhodopsin[16]. There are 7 transmembrane helices marked A to G. The B-C loop (blue) is elongated and caps the extracellular side of the protein. Membrane borders calculated using the PPM server[17] are shown with the black lines. b) Chemical structure of the all-*trans* retinal covalently linked to a lysine side chain. There are 20 carbon atoms in the retinal. Usually, the isomerization occurs around the C_{13}-C_{14} and C_{15}-N double bonds. The Schiff base bears a positive charge when protonated.

Although bacteriorhodopsin did not become the first membrane protein, for which X-ray crystallographic structure could be obtained (it was the *Rhodopseudomonas viridis* photosynthetic reaction canter in 1985[18]), experiments with it gave rise to a plethora of nontraditional crystallization techniques, where different lipidic phases are utilized as a

crystallization media. Among them are crystallization in the lipidic cubic phase[19,20], crystallization from bicelles[21] and crystallization from vesicles[22]. Initially, these crystallization methods were considered too difficult for routine application, however, advent of automatization[23] and successful determination of first non-rhodopsin GPCR structure[24] resulted in widespread adoption of *in meso* crystallization. Meanwhile crystallization from bicelles is less known, it also was used successfully for proteins other than microbial rhodopsins such as β_2 adrenergic G-protein-coupled receptor[25] and respiratory complex I[26]. Finally, crystallization from vesicles was used to obtain not-twinned crystals of BR[22] and crystals of trimeric ar-2[27].

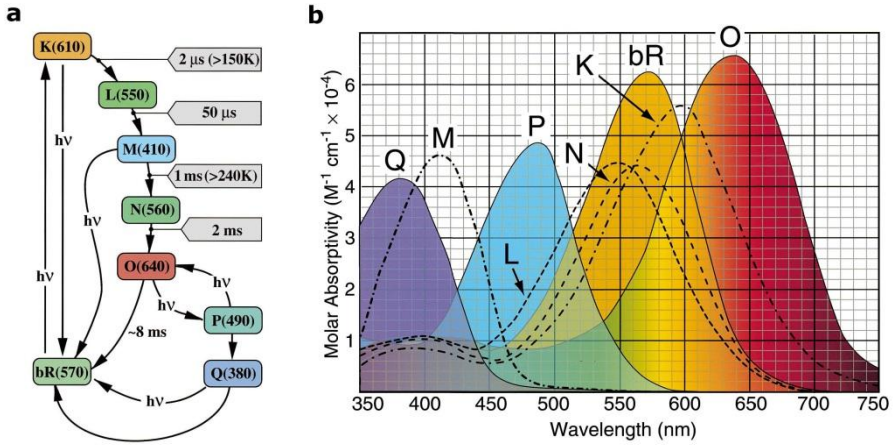

Figure 1.1.2 Bacteriorhodopsin photocycle (adapted from Birge *et al.*[28]). a) The main bacteriorhodopsin photocycle starting from the light adapted state. Additional transitions are possible upon absorption of a second photon. b) Absorption spectra of the ground and intermediate states.

Other structural techniques have also been applied to bacteriorhodopsin. Magic angle spinning NMR studies of ^{13}C labeled BR were used to get insight into protonation states of important aspartic acid residues[29]. Neutron and X-ray diffraction experiments on the normal and deuterated BR samples provided the information about the motions of BR helices during the photocycle[30,31]. Spin label attachment at the specific BR sites allowed probing of the microenvironment by the electron spin resonance technique[32]. Finally, various atomic force microscopy (AFM) experiments were performed on BR. While simple AFM experiments corroborated the information about the BR packing in the purple membrane[33], advanced high-speed AFM allowed visualization of BR motions in the membrane and assembly of the highly-ordered hexagonal lattice[34]. Combination of AFM and single-molecule force spectroscopy revealed the unfolding pathways of bacteriorhodopsin[35].

1.1.3 Bacteriorhodopsin: structure and function

The aforementioned studies resulted in extensive characterization of bacteriorhodopsin and its mechanism of function[6]. The proton translocation pathway consists of three major checkpoints: the proton uptake site, the Schiff base region and the proton release region (Figure 1.1.3).

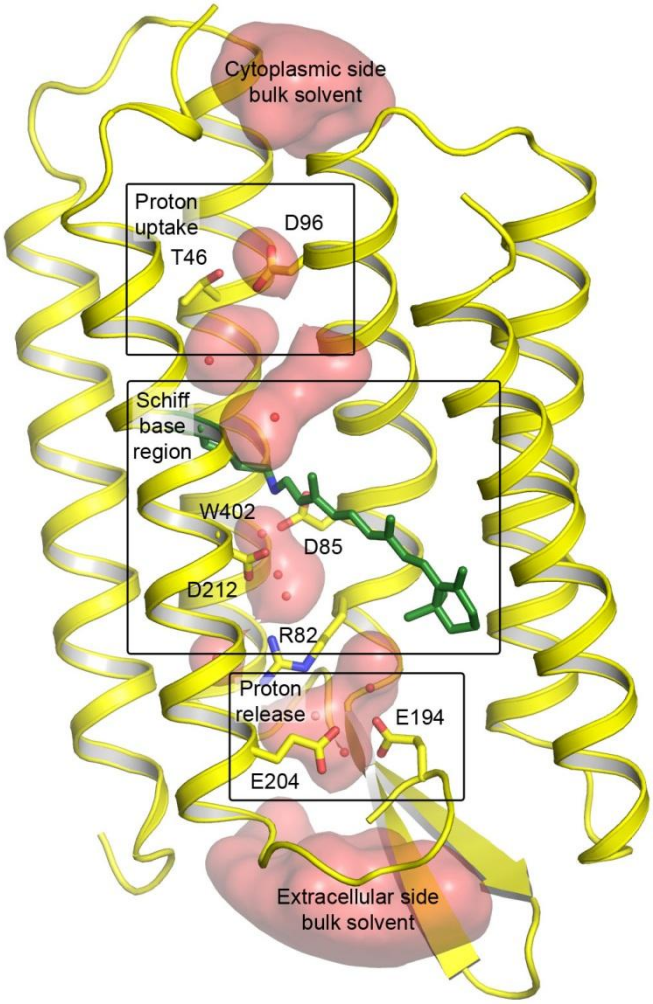

Figure 1.1.3. Structure of bacteriorhodopsin (PDB ID 1C3W). The internal cavities are represented by the red surfaces. The crystallographically observed water molecules are shown as the red spheres. Helix F is not shown.

There is a general agreement on the mechanism of the proton pumping by bacteriorhodopsin. In the ground state the proton is stored at the retinal Schiff base that makes a hydrogen bond to a proximal water molecule W402. W402 is also coordinated by the aspartates 85 and 212. The W402 cavity also contains two other water molecules. The cavity is blocked by the retinal at one side and by the Arg-82 side chain at the other.

Absorption of the light photon by the retinal leads to isomerization of the retinal from the all-*trans* to 13-*cis* conformation. In this new position, the protonated state of the Schiff base is energetically unfavorable, and the proton transitions to the proton acceptor residue Asp-85.

After that, the proton needs to pass to the proton release group, consisting of the glutamates 194 and 204. However, this is impossible because of the Arg-82 side chain. In the M-state this side chain turns away from the Schiff base.

As a next step the retinal is reprotonated with the proton stored at the proton uptake site on the Asp-96 side chain. Consequently, Asp-96 needs to be reprotonated from the cytoplasm, which can be possible due to a transient water molecule wire[36] or unlatching of the cytoplasmic side[37]. Reprotonated retinal pushes the Asp-85 proton to Glu-194/204, where it displaces another proton occupying that site, which goes in its turn into the bulk solvent.

Although these general points are agreed upon by most of the researchers, the details of the proton translocation mechanism and the associated conformational changes in particular, are often disputed. For example, there are considerable differences in the crystallographic structures of the intermediate states published by different research groups[38]. These differences might be a consequence of crystal twinning[39] or unusual radiation sensitivity and consequent damage of BR active site[40]. However, not only particular details but also the amplitudes of structural changes are questioned, as the electron microscopy studies, where the protein motions are supposedly less restrained by the environment, provide a picture of the larger-amplitude motions which are not observed in any of the crystallographic structures[41]. All these discrepancies are yet to be resolved.

1.1.4 Microbial rhodopsins

While *Halobacterium salinarum* bacteriorhodopsin is the most studied microbial rhodopsin, many other proteins of this family have been identified. Soon after the discovery of BR, a light-driven chloride pump was identified in the membrane of *H.s.*, dubbed halorhodopsin[42,43]. Although no other ion pumping proteins were detected in *H. salinarum*, two light sensors with different functions were found: sensory rhodopsin I, responsible for both attractant and repellent responses[44–47], and sensory rhodopsin II, responsible for the repellent response and also known as phoborhodopsin[48–51]. Similar proteins were also found in other haloalkaliphilic archaea[52].

Advent of metagenomics and sequencing of environmental samples that contained uncultivated bacteria marked a new era of microbial rhodopsin research. In the year of 2000, a

bacterial rhodopsin gene, coding for a functional protein, was discovered in a genomic fragment from a marine γ-proteobacteria, collected in Californian Monterey Bay[53]. Soon, it was demonstrated that the protein, dubbed proteorhodopsin, is indeed present in oceanic bacteria[54]. It also appeared that the proteins from this microbial rhodopsin subfamily were spectrally tuned to different habitats – that is, they absorbed light at different wavelengths in accordance with illumination available in the environment[54]. Since then, the proteorhodopsin genes were found to be distributed among divergent bacterial taxa[55,56], even as peculiar as Siberian permafrost bacteria[57], and pump not only protons but also the sodium and chloride ions[58,59].

Finally, presence of microbial rhodopsins is not limited to bacteria and archaea as they were also found, for example, in algae[60] and giant viruses infecting unicellular eukaryotes[61]. Algal channelrhodopsins, which open a cation channel upon illumination, became of extreme importance in the recently developed field of optogenetics[62] and were extensively modified to achieve some useful properties such as anion conduction[63–65].

To sum up, microbial rhodopsins performing a large variety of functions were identified. Among them are light-driven proton pumps, such as bacteriorhodopsin from the archaeon *Halobacterium salinarum*[7] or proteorhodopsins from marine bacteria[53,56], anion pumps such as halorhodopsin[66,67] or *Nonlabens marinus* rhodopsin 3[59], cation pumps[58,68], light-gated ion channels (channelrhodopsins)[69–71], and, finally, light-regulated sensory proteins, such as archaeal sensory rhodopsins[72] that conduct the signal through a chemoreceptor-like transducer protein, bacterial *Anabaena* sensory rhodopsin that utilizes a soluble second messenger[73], or algal histidine kinase rhodopsins[74,75].

1.1.5 Structures and relationship between microbial rhodopsins

In the recent years, atomic structures of various microbial rhodopsins have been elucidated. Although the majority of them come from X-ray crystallography, some structures were determined using solution[76] or solid state[77] NMR. Comparison of the available structures is presented in the Figure 1.1.4 and the corresponding information is summarized in the Table 1.1.1 (the structures presented in this work were not included). In the situations where several structures are available, only the one corresponding to the ground state and the best resolution is shown. The phylogenetic tree showing the relations between the important microbial rhodopsins is presented in the Figure 1.1.5.

Table 1.1.1. Known structures of microbial rhodopsins.

Name of the protein	Abbreviation	PDB ID	Resolution, Å	Reference
Bacteriorhodopsin	BR, HsBR	1C3W	1.55	Luecke et al.[16]
Archaerhodopsin 1	aR-1	1UAZ	3.4	Enami et al.[78]
Archaerhodopsin 2	aR-2	2EI4	2.1	Yoshimura and Kouyama[27]
Deltarhodopsin 3	dR-3	4FBZ	2.7	Zhang et al.[79]
Halorhodopsin	HR, HsHR	1E12	1.8	Kolbe et al.[80]
Natronomonas pharaonis halorhodopsin	NpHR	3A7K	2.0	Kouyama et al. [81]
Natronomonas pharaonis sensory rhodopsin II	NpSRII	1H68	2.1	Royant et al.[82]
Anabaena sensory rhodopsin	ASR	1XIO	2.0	Vogeley et al.[73]
Acetabularia rhodopsin 2	AR2	3AM6	3.2	Wada et al.[83]
Xanthorhodopsin	XR	3DDL	1.9	Luecke et al.[84]
Channelrhodopsin	ChR	3UG9	2.3	Kato et al.[85]
Proteorhodopsin from Med12	PR	4JQ6	2.3	Ran et al.[86]
Proteorhodopsin from HOT75	PR	4KLY	2.7	Ran et al.[86]

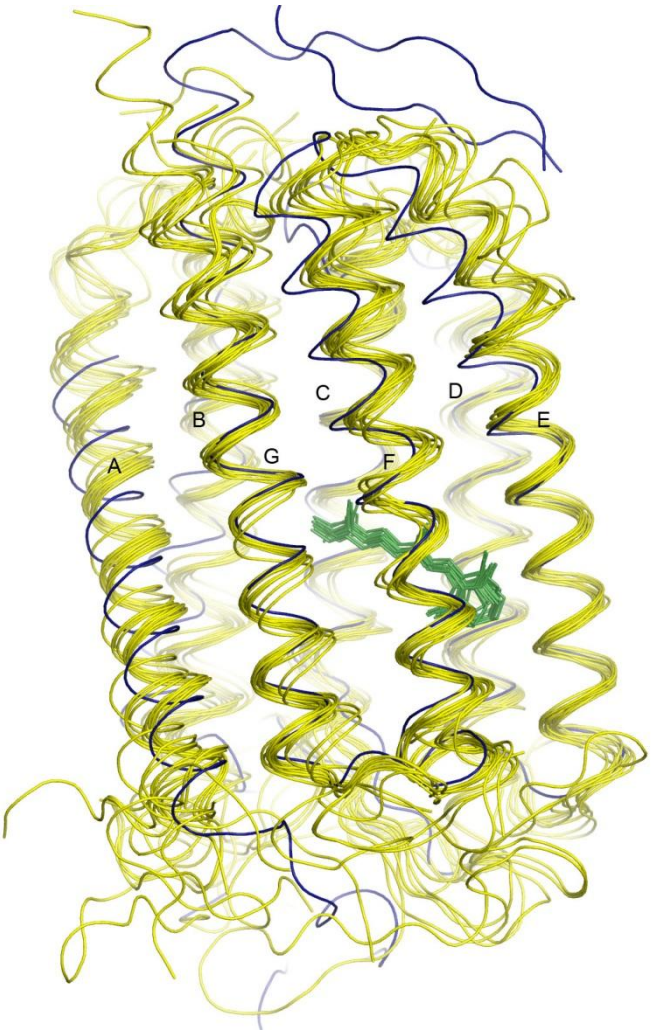

Figure 1.1.4 Superposition of the known microbial rhodopsin structures. The retinal is shown in green. Channelrhodopsin, whose structure differs the most from other proteins, is shown in blue (structure of the soluble parts is not shown). Most variation is observed in the exposed loop regions and, among the transmembrane helices, in the position of the helix A.

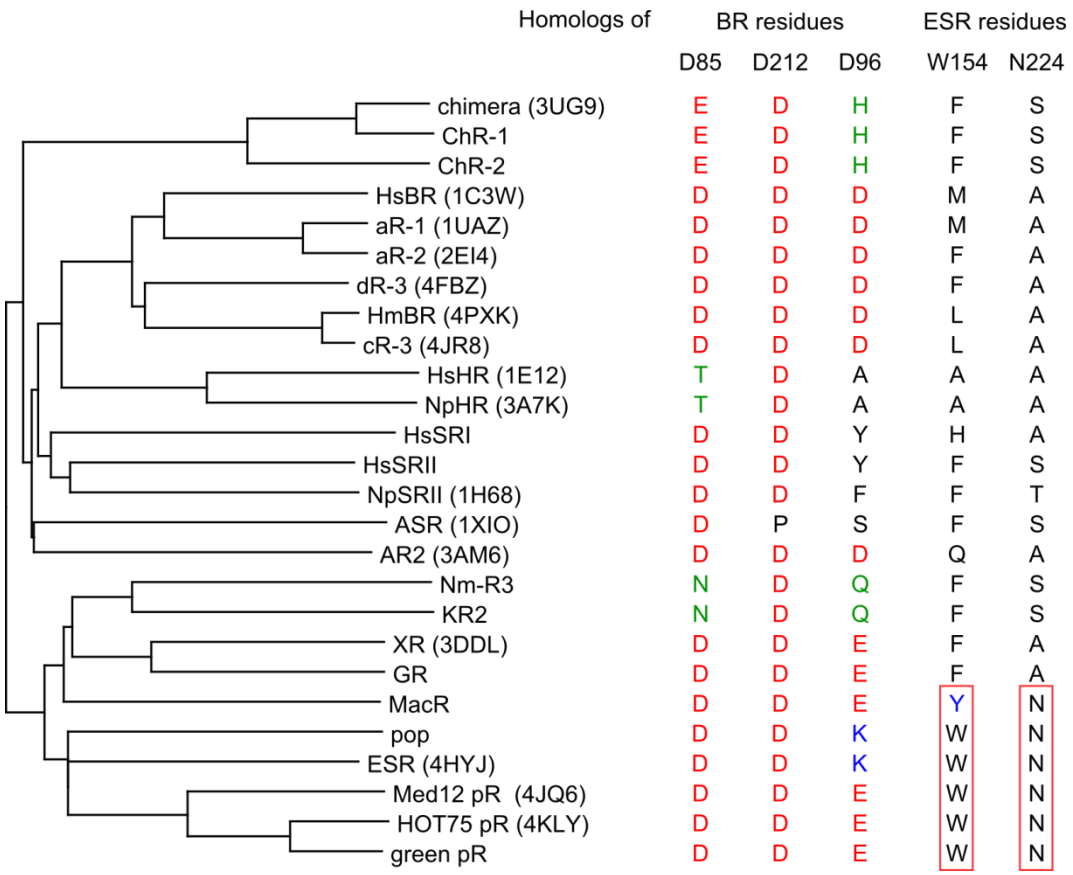

Figure 1.1.5. Phylogenetic tree of microbial rhodopsins. PDB IDs are shown in parentheses where available. The residues at the positions of BR's proton path-forming D85, D212 and D96, and at the positions of ESR's helix F kink-forming residues W154 and N224 (Section 3.2.1 and Figure 3.2.2) are shown. pop is a proteorhodopsin gene from the recently described uncultured marine Euryarchaeota[87]. MacR (marine Actinobacteria rhodopsin) is a protein from ultra-small marine Actinobacteria[88]. GR is *Gloeobacter* rhodopsin.

1.1.6 *Haloarcula marismortui* bacteriorhodopsin I

Historically, microbial rhodopsins performing four different physiological functions have been described in archaea. These are proton pumping, chloride pumping, and mediation of light-attractant and/or repellent functions. Correspondingly, four or less microbial rhodopsins are observed in a single species – for example, four in *Halobacterium salinarum* and three in *Natronomonas pharaonis*.

Recently, an unusual six-rhodopsin system of *Haloarcula marismortui* has been described by Fu and coworkers[89]. The system comprises two proton pumps (called *Hm*BRI and *Hm*BRII and possessing 50% sequence identity), a chloride pump, sensory rhodopsins I and II and an additional sensory rhodopsin of a novel type. Interestingly, absorption maximum of both *Hm*BRI and *Hm*BRII is at 552 nm. It is not clear, why two proton pumps absorbing at the same wavelength are needed, however it is possible that the proteins are optimized for functioning at different pH[90].

*Hm*BRI could be expressed in *Escherichia coli* with a high expression level, and the D94N mutation allowed to raise it to 70 mg per 1 liter of culture[91]. Such a high yield allowed using *Hm*BRI as an expression tag for efficient production of other membrane proteins[91]. As *E. coli* production allows for easy genetic manipulation and rapid production of mutants, *Hm*BRI could become a convenient model for the proton transport mechanism studies as well as for protein engineering for optogenetics needs[62].

Section 3.1 of this work describes the crystal structure of the *Hm*BRI D94N mutant. The structure is interesting as, despite the expression in *Escherichia coli*, the protein still assembles as a trimer, and the *E. coli* lipids occupy the positions very similar to that of archaeal *H. salinarum* lipids in *Hs*BR structures.

1.1.7 *Exiguobacterium sibiricum* rhodopsin

Permafrost soil is a soil that remains at or below the freezing point of water for extended periods of time. High numbers of microorganisms such as fungi, yeast, algae, actinomycetes, and bacteria have been discovered in the Siberian permafrost samples. Recently, one particular member of this unique microbial community, called *Exiguobacterium sibiricum*, has been isolated from 3 million years-old samples and extensively characterized[92]. *E.s.* was found to be adapted to the harsh conditions of the permafrost as it can grow at the temperatures as low as -5º C (the full range of growth temperatures is -5º C to +39º C)[93]. At the same time, this adaptation appears to be constitutive as little differential gene expression is observed between 4º C and 28º C[93].

Interestingly, sequencing of the *E.s.* genome has revealed presence of potential proteorhodopsin gene[93]. This protein, called *E.s.* rhodopsin (ESR), could be expressed in *E. coli*, bound the all-*trans* retinal and displayed an absorption maximum at 534 nm[57]. ESR, incorporated in vesicles, could efficiently transport the protons across the membrane under light illumination[57].

Compared to other microbial rhodopsins, ESR revealed two notable features. First, its proton uptake residue is a lysine, which can only be positively ionized, while all the other proton pumps studied to date possess a negatively ionizable amino acid (Asp or Glu) at that position[94]. This fact bears qualitative consequences for the proton pump function, as the proton uptake group is neutral in the ground state, thus it is protonated in the classical proton pumps and deprotonated in ESR. Recently, several other proteorhodopsins also having a lysine residue at the proton uptake have been found in marine uncultured Euryarchaeota[87] (*pop* gene, Figure 1.1.5). The second unique quality of ESR is its broad proton pumping pH range - above the pH of 4.5[95]. This feature was ascribed to the retinal-proximal pair of residues Asp-85 and His-57[95], which are, however, also observed in xanthorhodopsin and proteorhodopsins.

ESR presented an interesting target for a study as, first, it is an unusual protein from a psychrophilic bacterium, and second, no atomic three-dimensional proteorhodopsin structure has been known at the beginning of the study. Its structure is described in the Section 3.2.

1.1.8 Light-driven sodium pump KR2

While microbial rhodopsins acting as anion or proton pumps were known for a long time, the light-driven sodium pumps remained elusive. This elusiveness could even be rationalized by speculating that the positively charged protonated retinal Schiff base displaces a proton during its isomerization, and can displace an anion, pulling it by Coulombic attraction, but cannot move a cation as there is no suitable interaction.

Recently, a novel class of microbial rhodopsins, called NQ-rhodopsins, was described[68]. The name of the proteins comes from the fact that the retinal counterion position (D85 in bacteriorhodopsin) is occupied by asparagine in this family, and proton uptake (D96 in bacteriorhodopsin) is occupied by glutamine (Figure 1.1.5). Despite these seemingly destructive substitutions the proteins were still able to pump protons[68].

Soon thereafter, Inoue and colleagues[58] studied the rhodopsins 1 and 2 of the bacterium *Krokinobacter eikastus* (Figure 1.1.6a) and showed that, while KR1 is a typical proton-pumping proteorhodopsin, KR2, a member of the NQ-rhodopsins subfamily, is a light-driven sodium pump, which can pump protons in absence of sodium. In both cases, the photocycle consists of several bacteriorhodopsin-like intermediates (K, L, M and O, Figure 1.1.6b), although there are slight differences probably explained by the fact that KR2 pumps preferably sodium and in its absence the turnover is slower and the pump is less efficient. The protein was also found to be very selective, pumping the Li^+ and Na^+ ions but not K^+, Rb^+ or Cs^+.

The unique properties of KR2 could potentially be very useful in the field of optogenetics[62], and knowledge of its structure could open the way to rationalization of its pumping and generation of, for example, more effective or potassium-selective pumps. Section 3.7 of this work describes the structure of KR2.

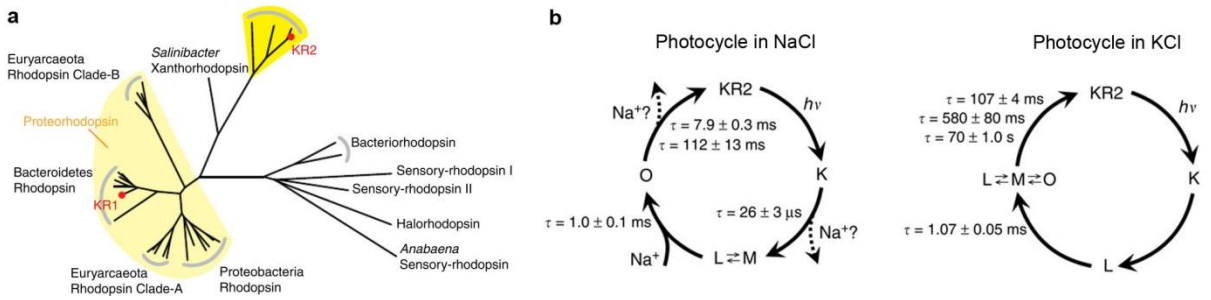

Figure 1.1.6 Position of Krokinobacter rhodopsins in the phylogenetic tree (a) and KR2 photocycles in NaCl and KCl (b)[58].

1.2 Photo- and chemotaxis

1.2.1 Bacterial and archaeal taxis

Taxis is the movement of an organism in response to environmental stimuli such as light or chemical gradients. Some of the examples are aerotaxis (response to oxygen), chemotaxis (to chemicals), phototaxis (to light) and thermotaxis (to temperature).

Bacteria and archaea live in highly variable and often dangerous environments, and their survival and proliferation depends heavily on their ability to move towards energy sources and to avoid hazardous conditions such as UV-illumination. In this chapter, bacterial and archaeal chemotaxis and phototaxis that are related by common mechanisms of action are described.

1.2.2 Principles of microbial locomotion

Microorganisms such as bacteria and archaea move with a help of flagella – lash-like appendages that protrude from the cell body and whose motions are controlled by the organisms (Figure 1.2.1). Although the structures of bacterial[96,97] and archaeal[98,99] flagella are different, the principles of the flagella-based motion of these organisms are the same. In both cases there is a filament anchored to the membrane (Figure 1.2.1d) that can rotate either clock-wise (CW) or counter-clock-wise (CCW). CCW rotation of the flagellum results in the cell moving straight, while CW rotation leads to tumbling and change of direction of the straight motion.

Rational movements of the bacterium or archaeon in response to environmental stimuli are a consequence of the behavior called biased random walk: during the straight runs the cell measures the changes in the environment; if the changes are positive, the motion continues; if not, the cell stops and starts to tumble – with a high probability, the new orientation will be in the opposite direction to previous motion and probably towards the positive stimuli.

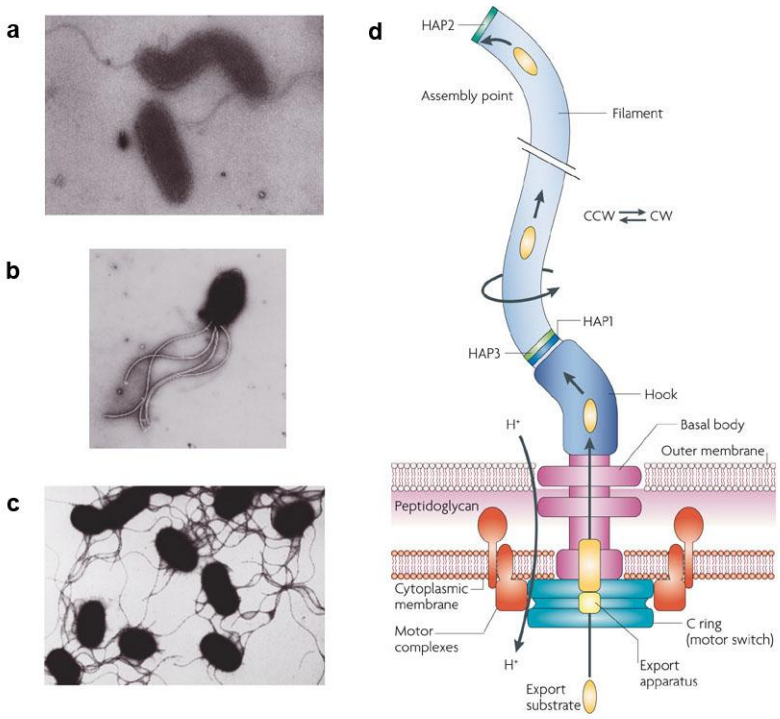

Figure 1.2.1. Bacterial flagella. a-c) Electron micrograph images illustrating the different types of flagellar arrangement in bacteria[96]. Some bacteria, such as *Vibrio cholerae*, *Pseudomonas aeruginosa* or *Idiomarina loihiensis*, have a single flagellum and called monotrichous (a). Many bacteria have numerous flagella and, if these are co-located on the surface of the cell to form a tuft, the bacterium is lophotrichous (b, *Vibrio fischeri*). For spirochaetes, such as species of *Borrelia* (c), a specialized set of flagella are located in the periplasmic space, the rotation of which causes the entire bacterium to move forward in a corkscrew-like motion. d) Model of the bacterial flagellum structure and assembly[97,100]. The bacterial flagellum consists of three parts: the filament, the hook and the complex basal body. The locations and roles of many of its components are known. The proton gradient across the cytoplasmic membrane drives rotation of the rotor and the attached filament. HAP is a hook-associated protein.

1.2.3 Regulation of flagella by chemo- and photoreceptors

The cell controls the rotation of the flagella by means of its chemotaxis and/or phototaxis system. The systems are very similar, both for phototaxis and chemotaxis, and bacteria and archaea. In most cases, the signal is detected by a transmembrane receptor (although some receptors are soluble). The receptor controls the activity of histidine kinase CheA, attached to it via the adapter protein CheW. Attractant signal results in less intensive phosphorylation of CheY, which in its turn results in CCW rotation of the flagellum (Figure 1.2.3). An increased sensitivity of the chemo- or phototaxis signal is attained by methylation of the receptor by the

protein CheR (which sensitizes the receptors) and demethylation by CheB (which has an opposite effect). Finally, the protein CheZ serves to dephosphorylate CheY.

The described proteins, present, for example, in *Escherichia coli*, form the minimal version of the functioning yet highly sensitive taxis system. In other bacteria, the supplementary proteins such as CheC and CheD can also be present, as well as multiple variants of CheY, CheW and other proteins[101].

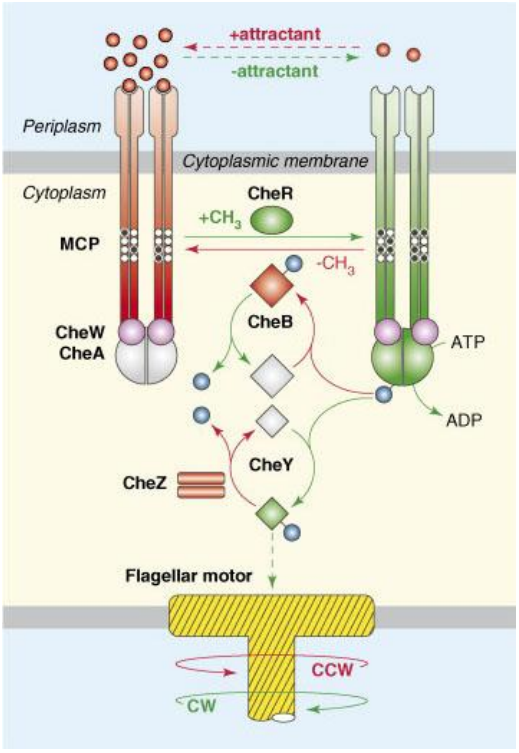

Figure 1.2.2. The chemotaxis signaling pathway in *E. coli* (from Hazelbauer *et al.*[102]). Components and reactions in red promote counter clockwise (CCW) flagellar rotation; those in green promote clockwise (CW) flagellar rotation. Components in gray represent inactive forms. Solid lines represent enzymatic reactions; broken lines indicate binding interactions. CheA-derived phosphoryl groups are shown as blue spheres. Receptor modification sites are shown as white (unmethylated) and black (methylated) circles.

1.2.4 Spatial organization of chemotaxis system

The large complexes comprising the chemoreceptors and proteins CheA, CheW and CheZ are usually found close to the cell pole of bacteria[103,104] and can be easily visualized by electron tomography[105–107] (Figure 1.2.3a) or light microscopy[104,108]. There, the elongated chemoreceptor proteins form hexagonally ordered arrays (Figure 1.2.3b), which are extremely stable *in vitro*[109]. The degree of ordering of the chemoreceptors within the arrays appears to depend on the environmental conditions as well as on the presence of attractants or repellents[110].

While the nature and stoichiometry of the proteins forming the chemoreceptor arrays were known for a long time[102,104,111], their exact arrangement remained unknown. The cryo-electron tomography studies showed that inside the arrays, the inherently dimeric chemoreceptors form trimers (trimers-of-dimers, Figure 1.2.3d-g), with the kinase CheA and accessory protein CheW residing at the receptor tips (Figure 1.2.3d). Soon thereafter, the experiments with nanodisc-reconstituted chemoreceptors demonstrated that two chemoreceptor trimer-of-dimers, together with two CheW molecules and CheA dimer form the core unit of the chemotaxis signaling complexes[112].

1.2.5 CheA and CheW proteins

The histidine kinase CheA in its long form ($CheA_L$) consists of five domains of known structure[113]: histidine phosphotransfer domain (P1), CheY/CheB binding domain (P2), dimerization domain (P3), ATP-binding phosphotransfer domain (P4) and CheW-like receptor binding domain (P5, Figure 1.2.4). The shorter form $CheA_S$ lacks the most of its P1 domain and thus cannot be phosphorylated. The most important for understanding chemoreceptor cluster assembly are the crystallographic structures of P3, P4 and P5 domains[114] and of the complex of CheA P4 and P5 domains with CheW[115]. Besides the crystallographic CheA-bound structure, there are structures of the CheW protein, determined by NMR spectroscopy[116,117] and by X-ray crystallography[118].

Although the details of the interaction of CheA and CheW with chemoreceptors were not elucidated for a long time, the recently determined crystallographic structure of the chemoreceptor-CheA-CheW complex together with electron-cryotomographic studies provided a clear picture of the chemoreceptor array organization[119,120].

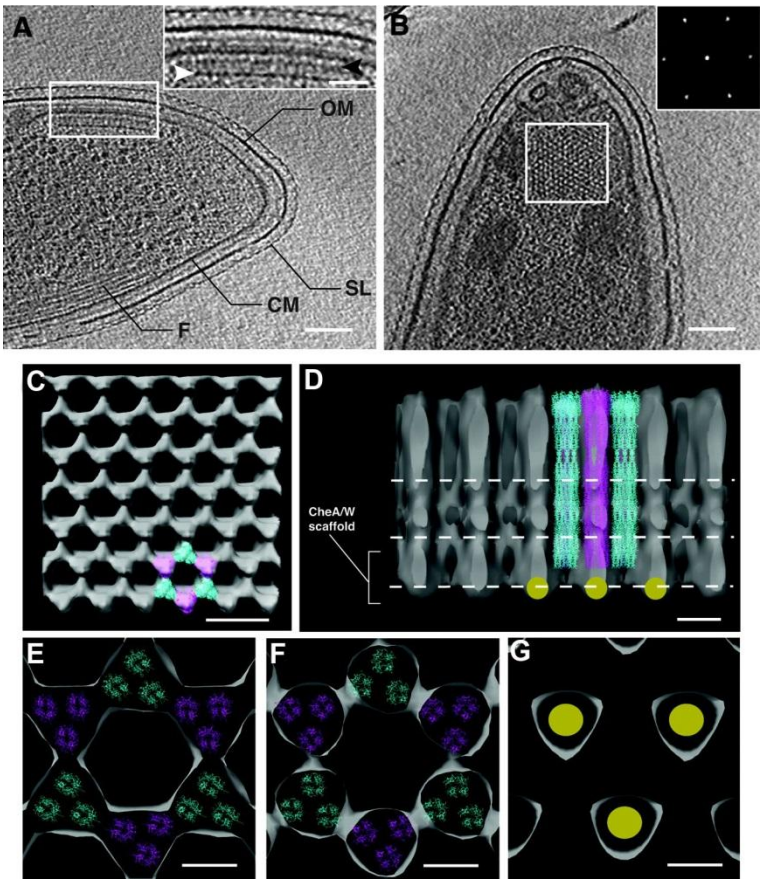

Figure 1.2.3. Spatial organization of chemoreceptor arrays (*Caulobacter crescentus*, adapted from Khursigara *et al.*[106]). a-b) Direct visualization of *Caulobacter* chemoreceptor arrays. a) Tomographic slice of the polar region of an intact Caulobacter swarmer cell demonstrating a continuous surface layer (SL), outer membrane (OM), cytoplasmic membrane (CM), and cytoplasmic filaments (F). The inset shows an expanded view of the chemoreceptors array and signaling scaffold (white arrowhead) and proposed CheR/B interaction sites (black arrowhead). b) In the face-on orientation, chemoreceptor arrays are characterized by a distinct, partially ordered pattern ~32 nm below the cytoplasmic membrane that corresponds to the signaling scaffold. (Inset) The power spectrum of this region clearly demonstrates the ~12-nm hexagonal spacing. The scale bars in panels a and b are 100 nm; the panel a inset scale bar is 50 nm. c-g) Averaged packing arrangement of trimeric receptors in *Caulobacter* chemoreceptor arrays. c) Averaged density map of a chemoreceptor array demonstrating idealized hexagonal packing, with trimeric chemoreceptors model structures (cyan and magenta) docked into a single hexagonal unit. d) One-half of the hexagonal unit described in panel c, with continuous receptor densities emanating from the signaling scaffold. The gold circles represent the extra density attributed to molecules within the signaling scaffold. The three white dashed lines represent the heights at which cross-sectional cuts were made in the density map to obtain panels e through g, from top to bottom, respectively. e, f) Trimer-of-dimer organization of chemoreceptors. There are multiple interactions between adjacent trimers. g) Cross-sectional view from below the level of chemoreceptor shows the extra density directly below every other receptor trimer (magenta). The scale bars in panels A and B are 20 and 10 nm, respectively, and the scale bars in panels C to E are 5 nm.

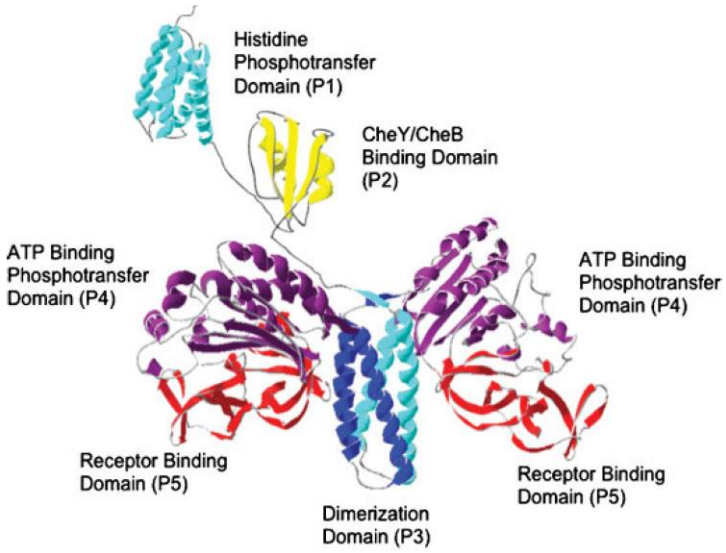

Figure 1.2.4. Molecular model of the histidine protein kinase CheA[113]. The histidine phosphotransfer domain (P1) and the response regulator CheY/CheB-binding domain (P2) are depicted as monomers connected to one another and the remainder of CheA via flexible linkers. The dimerization domain (P3), ATP-binding phosphotransfer domain (P4), and the receptor-binding domain (P5) are all depicted within a CheA dimer.

1.2.6 Architecture of chemoreceptors and sensory rhodopsin-transducer complexes

Overall, the primary receptors of the chemo- and phototaxis systems have a very similar architecture (Figure 1.2.5). In the chemotaxis system, it is a homodimer of chemoreceptor proteins. In the phototaxis system, the photosensor and signal transducer are separate polypeptide chains that interact via their transmembrane domains and dimerize via transducers.

In the chemotaxis system, the signal is generated by the extracellular domains of chemoreceptors, and then transmitted inside the cell by the transmembrane helices. In the phototaxis system, the signal (illumination) is detected by the photosensor sensory rhodopsin, a member of microbial rhodopsin family. The sensory rhodopsin passes the signal to the transmembrane part of the transducer protein. In the cytoplasm, in both systems the signal passes through the HAMP domain to the kinase control module, which, in its turn, regulates the activity of the kinase CheA.

In the following sections, structures and conformational changes of taxis systems constituents are discussed.

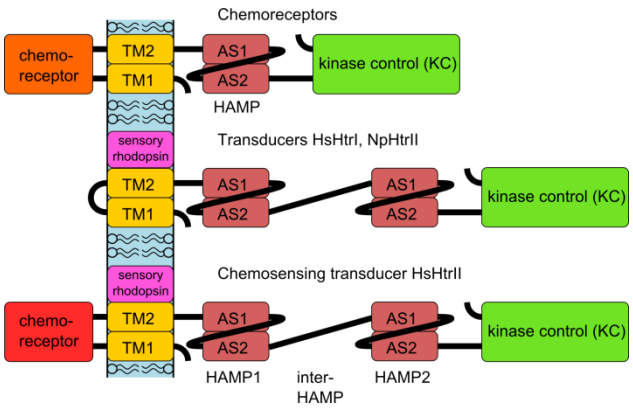

Figure 1.2.5. Domain architecture of bacterial chemoreceptors (top), of the phototactic signal transducers HsHtrI and NpHtrII (middle), and of the transducer and chemoreceptor HsHtrII (bottom). Both chemoreceptors and sensory rhodopsin transducers are dimeric, but only one protomer is shown for clarity. TM1 and TM2 are the transmembrane helices, AS1 and AS2 are the helices of the HAMP domain. The cell membrane is shown in blue.

1.2.7 Signal generation in chemoreceptors

At the moment, structures of several chemoreceptor domains belonging to different families are known. However, for most of them only the apo- or holo- structure is known, which is not sufficient for speculations about the signaling mechanism. The most structurally characterized domain, and the first one for which the atomic structure has been determined, is the aspartate sensing domain of *Escherichia coli* aspartate receptor Tar. The structures of the domain in apo- and holo- forms[121] led to the first model of the signal transduction through the membrane[122] (Figure 1.2.6).

Briefly, in the absence of ligand the domain is a symmetric dimer. Each protomer consists of 4 α-helices, where the first one continues right after the first transmembrane helix TM1, and the forth one is followed by the second transmembrane helix TM2 (Figures 1.2.5, 1.2.6a). Binding of the ligand at the dimerization interface of the protomers leads to displacement of the helix α4 of one of the protomers. This displacement is then passed to TM2 and inside the cell. This piston-like motion of TM2[122] is also observed in chemoreceptor kinases of two-component signaling systems[123].

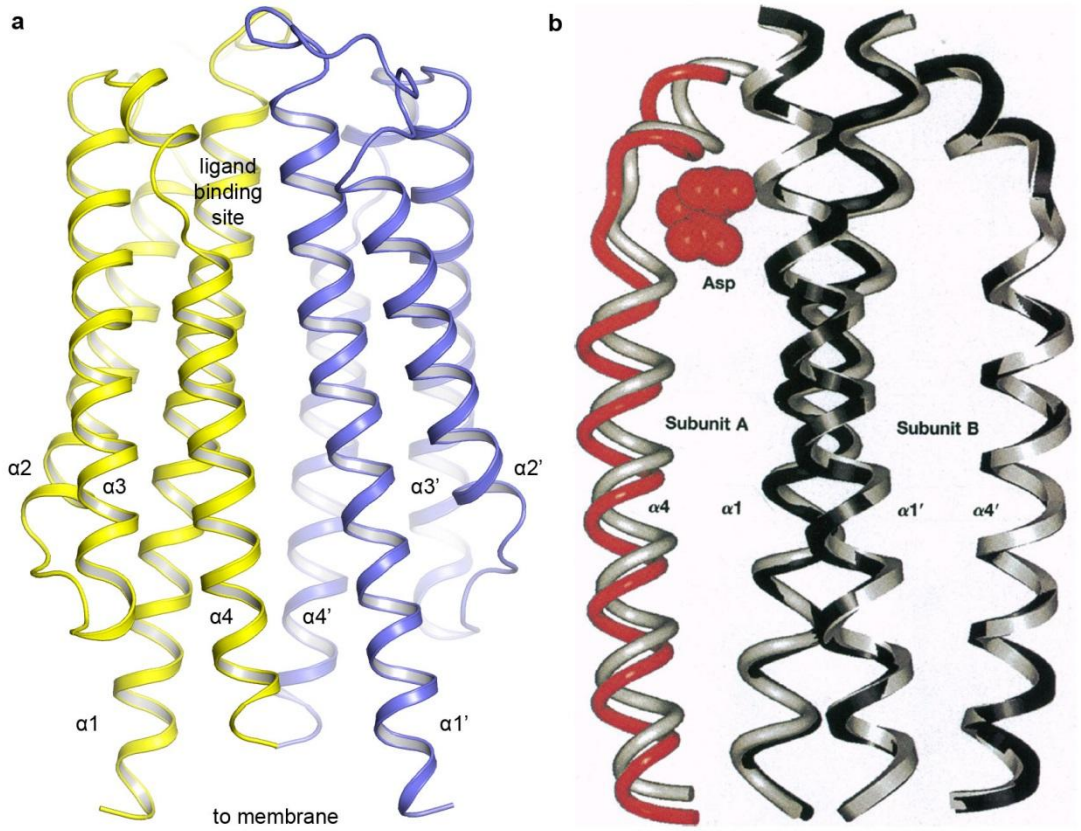

Figure 1.2.6. Structure and aspartate-induced conformational changes in the Tar periplasmic ligand-binding domain. a) Structure of the periplasmic ligand-binding domain (PDB ID 1VLS[124]). In the absence of aspartate, the protein is symmetric dimer. The protomers are shown in yellow and blue, each protomer consists of 4 α-helices, marked α1-α4 (prime denotes the second protomer). b) Aspartate-induced conformational changes in the Tar periplasmic ligand-binding domain[122]. Shown is a superposition of the crystal structures[121] for the apo (grey) and aspartate-occupied (black) ligand-binding domain, depicting the single bound aspartate molecule and the periplasmic regions of the four membrane-spanning helices (residues 44-75 of helices αl/TMI and αl'/TM1'; residues 146-175 of helices α4/TM2 and α4'/TM2'). Upon aspartate binding, the α4/TM2 helix of subunit A is observed to translate 1.6 ± 0.2 Å downward (or toward the cytoplasm in the intact receptor) and to tilt 5º, yielding the new position highlighted in red. By contrast, the remaining three transmembrane helices of subunits A (cylindrical ribbon) and B (square ribbon) are relatively stationary, exhibiting aspartate-induced translational and angular displacements less than 0.5 Å and 1º in magnitude, respectively.

1.2.8 Signal generation in the phototaxis system

The primary light receptors of the phototaxis system are sensory rhodopsins. While sensory rhodopsin I serves for both attractant response to 565 nm light and repellent to 370 nm light, sensory rhodopsin II (sometimes called phoborhodopsin) generates only repellent signal to illumination with 480 nm light. Much more structural information is known for *Natronomonas pharaonis* SRII than for other sensory rhodopsins and, consequently, we will focus on NpSRII.

In the cellular membrane NpSRII is bound to its cognate transducer NpHtrII in a 2:2 complex[125,126] (Figures 1.2.5 and 1.2.7a). Absorption of a light photon and consequent trans-cis isomerisation of the retinylidene chromophore of NpSRII initiates the photocycle. As a consequence the protein passes through several intermediates until it reaches back to the original state in about 1 second. The functionally important conformational step occurs during the M1→M2 transition which leads to the signaling state[127,128]. During this transition the signal is transferred through the interface comprising helices F and G to NpHtrII, which results in a rotary motion of transmembrane helix TM2[126,129] (Figure 1.2.7b). This activation occurs at the level of the membrane. How this signal is then transmitted to signal domain thereby modulating the activity of the histidine kinase CheA and triggering the cytoplasmic two component signaling cascade is not known yet [102,125,129,130].

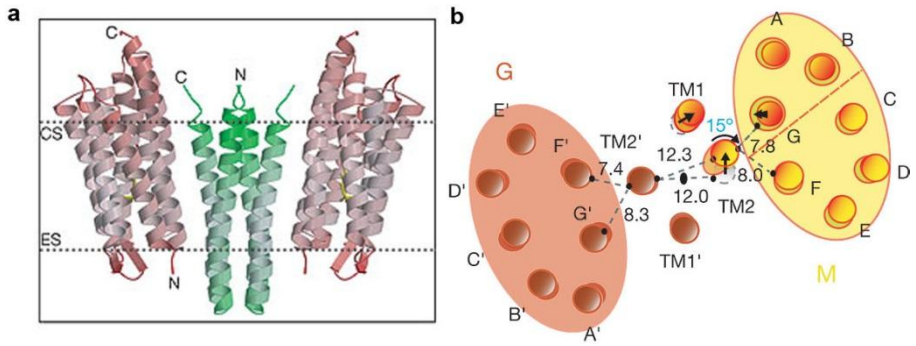

Figure 1.2.7. Structure of the sensory rhodopsin-transducer complex and conformational changes during the photocycle. a) Structure of 2:2 NpSRII:NpHtrII complex in the membrane[125]. The sensory rhodopsin is shown in red and the transducer in green. b) Conformational changes in the NpSRII-NpHtrII complex during the photocycle[129].

Published X-ray and NMR structures of NpSRII[76,82,131,132] coincide within experimental errors and strongly resemble those obtained for bacteriorhodopsin[16]. The conformation of the backbone and that of conserved side-chains in the retinal-binding pocket are more or less

identical. Moreover, the structure of the receptor[76,82,131,132] is almost the same as compared to the NpSRII/NpHtrII complex[125].

These observations pose an interesting question. How does nature fine-tune common scaffolds to engender two completely different functions: sensor and ion pump? From previous work it is evident that NpSRII as well as SRI is capable of pumping protons, albeit with poor efficiency[133–136]. Similarly, bacteriorhodopsin can be converted into a sensor by just three mutations[137]. On the other hand, despite great efforts it has not been possible to change NpSRII into an efficient proton pump[138]. Apparently, the requirements for an effective ion pump are much more demanding than those for a functional sensor.

A second observation relates to the inhibition of the proton pump on transducer binding NpHtrII to its cognate receptor NpSRII[139,140]. It has been discussed that in the 2:2 complex the cytoplasmic channel cannot open sufficiently, thereby altering the proton uptake kinetics such that the reprotonation of the Schiff base is faster from the extracellular side. This kinetic correlation would result in a futile proton cycle (reviewed in Sasaki and Spudich[141]). Certainly, this proposal can only be verified by a comparison of NpSRII active states with and without bound transducer NpHtrII.

Crystal structures are available for NpSRII ground state[82,131,132] and for its K-intermediate, which is formed at room temperature in the nanosecond range[142], as well as for the NpSRII/HtrII complex, for which data of the ground state and its K- and M-intermediate (the active state) are available[125,129]. Despite several attempts the structures of the late (active) states of NpSRII were not obtained yet. The reason for this failure has been related to large conformational changes upon the light-activated M-formation, which lead to severe disturbance of the crystal packing and consequently a substantial decrease in resolution. A similar difficulties arise in experiments with the intermediate states of visual pigment rhodopsin[143,144]. Section 3.4 of this work describes the process of determining the structure of the NpSRII in the active state and the structure itself.

1.2.9 Signal transduction by the HAMP domain.

HAMP domain is a ubiquitous signaling module found in **h**istidine kinases, **a**denylyl cyclases, **m**ethyl-accepting chemotaxis proteins and **p**hosphatases (recently reviewed by Parkinson[145]). The domain was first identified as an amphipathic linker between the transmembrane helices and the signal output domain[146,147]. Much later, the atomic structure of the HAMP domain part of the thermophile Archaeoglobus fulgidus putative protein Af1503 was determined by

NMR[148] (Figure 1.2.8). The structure revealed that the HAMP domain is organized as a symmetric homodimeric parallel coiled coil. Each protomer has two α-helices, AS1 and AS2, connected by a flexible linker segment. Later, a similar structure was observed in a crystallographic structure of three consecutive HAMP domains from the Aer2 protein of Pseudomonas aeruginosa[149]. The arrangement was also verified by biochemical and biophysical methods for a number of other proteins – chemoreceptors Tar and Tsr[150–153], aerotaxis protein Aer[154], phototaxis signal transducer HtrII[155–157] and sensory histidine kinases EnvZ and NarX[158,159].

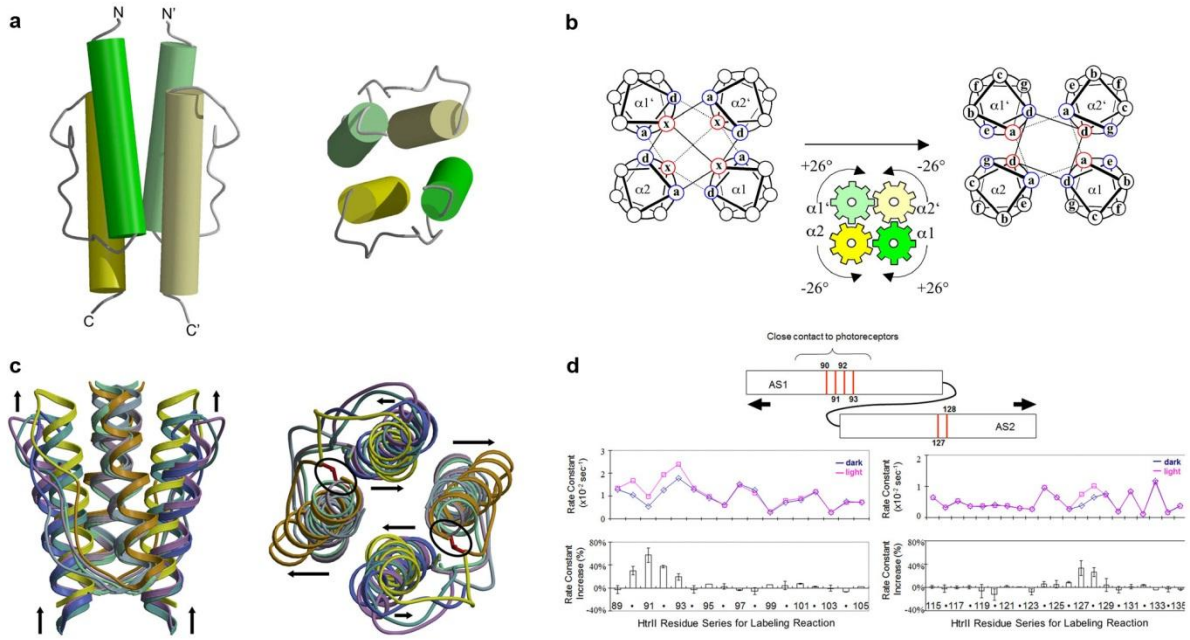

Figure 1.2.8. Structure of the HAMP domain and different models of its conformational changes. a) Structure of the Af1503 HAMP domain[148]. The domain is a symmetrical homodimer/4-helical parallel coiled coil. b) Gear-box model of Hulko *et al.*[148] Schematic representation of complementary x-da packing (left) versus knobs-into-holes packing (right). The two packing modes can be interconverted by rotating adjacent helices by 26 in opposite directions, as illustrated by the cogwheel diagram. c) Model of Airola et al.[149], based on the structures of the three HAMP domains from the soluble chemoreceptor Aer2. d) Model of Wang et al.[160], based on the cysteine accessibility scanning of the NpHtrII first HAMP domain. Upon illumination, accessibility of the residues at the HAMP domain ends increases, which implies that the helices move in opposite directions.

Currently, there are several models of signal transduction through the HAMP domain[145]. The gearbox model posits that the HAMP domain helices switch between the orthodox a-d packing and the unusual x-da packing[148,161,162] (Figure 1.2.8b). Atomic structures of HAMP domain-DHp phosphotransfer domain fusions show that the rotation of the HAMP domain'

helices results in rotation of adjacent helices of DHp[161,162]. This mechanism explains the signal transduction in receptor histidine kinases, but it is not clear whether it is the case for chemo- and photoreceptors. Alternatively, experimental data reveal that the signal input in chemoreceptors and NarX is a piston-like motion of the transmembrane helix, to which the HAMP domain is connected[122,123,163] (Figure 1.2.6b). The HAMP domain itself may switch between two conformations[145,149] (Figure 1.2.8c). The output was proposed to be coded by the dynamic properties – looser or tighter packing of the HAMP domain' helices[145,150,151,164].

As for phototactic signal transducers, it was first proposed that the HAMP domain of NpHtrII transduces the signal via switching between a compact and a highly dynamic states[156,157]. Later, the fluorescent labeling studies revealed that the helices AS1 and AS2 move in opposite directions during signal transduction[160] (Figure 1.2.8d). Molecular modeling and NMR studies have shown that the NpHtrII HAMP domains have the same fold as the HAMP domains for which the structure is known[155,165].

Recently, several groups have studied the properties of chemo- and phototaxis proteins by means of modeling. Models of the NpHtrII HAMP1 as well as the HAMP domain region were built by Nishikata et al.[165]. Nishikata et al. have also studied the dynamics of the NpSRII-NpHtrII complex in the ground and the M states by means of molecular dynamics[166]. Signal transduction via the transmembrane part of chemoreceptors Tar[167,168] and sensor kinase PhoQ[169] was studied extensively by different groups. Finally, Hall et al. have generated a model of the entire chemoreceptor Tsr and of the trimer-of-dimers of these chemoreceptors that has shown how the small structural changes may be propagated across the system[170].

Despite this extensive work on the HAMP domains, little information existed about the inter-HAMP region of sensory rhodopsin transducers (Figure 1.2.5). It was proposed that it might act as a mechanical joint[171], however no details were described. The model of this region and its implications are presented in the Section 3.5 of this work. After the model of the HAMP-domain region was established, the question about the conformational changes, accompanying NpHtrII HAMP domain signaling, remained. The Section 3.6 describes the analysis of the NpHtrII HAMP1 motions in molecular dynamics simulations and consequent elucidation of two distinct states, displaying the features similar to those observed in experiments (Figure 1.2.8c-d).

1.3 CDP-OH transferases and IPCT-DIPPS

1.3.1 CDP-OH transferases

CDP-alcohol phosphatidyltransferases (CDP-OH transferases) are a large family of membrane proteins that catalyze the displacement of CMP from a CDP-alcohol by a second alcohol with formation of a phosphodiester bond and concomitant breaking of a phosphoride anhydride bond[172]. Currently (checked on 15/05/2014) there are 11822 CDP-OH transferase sequences in the Pfam database[172], 27917 sequences in InterPro[173], and 25077 sequences in Prosite[174].

The majority of the CDP-OH transferase proteins are involved in lipid biosynthesis. For example, there are five such enzymes in mammals, including humans, that participate in the *de novo* biosynthesis pathways of all the major glycerophospholipids in mammalian cells[175] (Fig. 1.3.1).

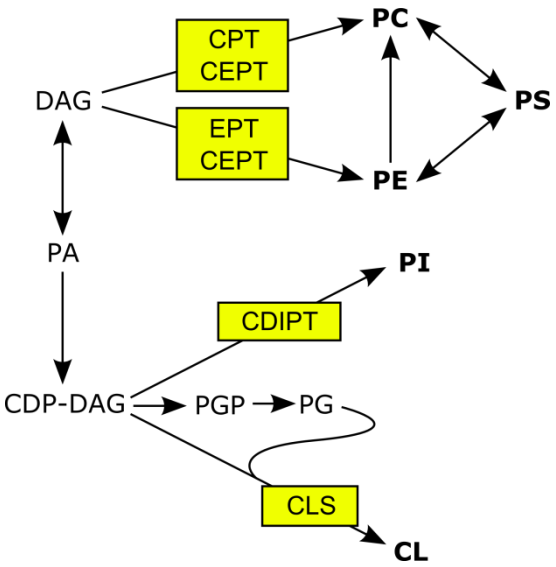

Figure 1.3.1. Pathways of glycerophospholipid biosynthesis in mammalian cells[175]. The biosynthesis starts either from diacylglycerol (DAG) or phosphatidic acid (PA), and results in production of phosphatidylcholine (PC), phosphatidyletanolamine (PE), phosphatidylserine (PS), phosphatidylinositol (PI) and cardiolipin (CL)[175]. The intermediate metabolites are CDP-diacylglycerol (CDP-DAG), phosphatidylglycerol phosphate (PGP) and phosphatidylglycerol. The CDP-alcohol phosphatidyltransferase family enzymes in mammals are highlighted, CPT1 is cholinephosphotransferase 1, CEPT1 is choline/ethanolaminephosphotransferase 1, EPT1 is ethanolaminephosphotransferase 1, CDIPT1 is phosphatidylinositolsynthase 1 and CLS1 is cardiolipinsynthase 1.

As it can be seen, the ligands in the reactions catalyzed by CDP-OH transferases are often very different. Consequently, there is a high level of diversity among the CDP-OH transferase sequences, which is highlighted in the sequence logo[176,177] of the family (Figure 1.3.2). Overall, the consensus motif of the family is $D(x)_2DG(x)_2AR(x)_{7\text{-}12}G(x)_3D(x)_3D$.

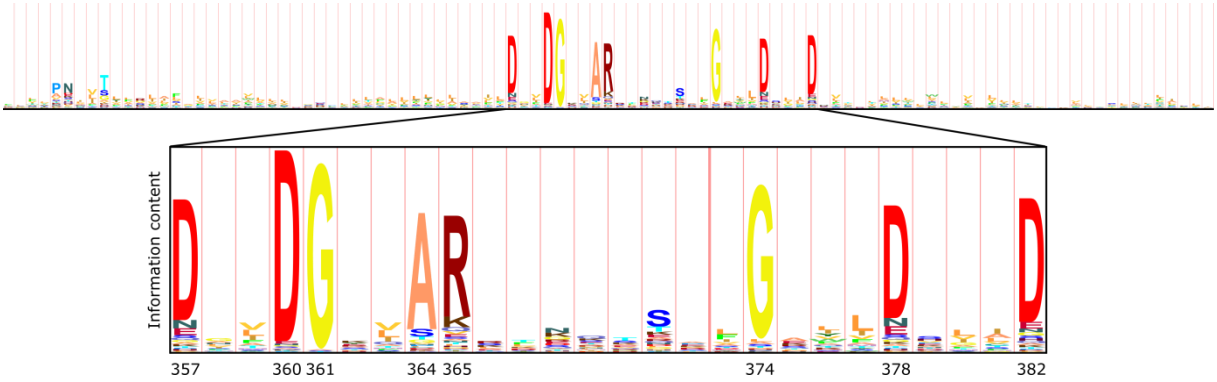

Figure 1.3.2. Sequence logo[176,177] of the CDP-OH transferase domain. The Hidden Markov Model used in preparation of this figure is based on the alignment of 124 representative Pfam protein family sequences as seeds[172].

1.3.2 Di-*myo*-inositol phosphate and the IPCT-DIPPS enzymes

Di-*myo*-inositol phosphate (DIP) is one of the major osmoprotecting molecules in a number of hyperthermophilic species of archaea and bacteria[178]. DIP level is also highly increased in some species at supraoptimal growth temperatures[179,180]. Based on this, it was proposed that DIP, besides being an osmoprotectant, might also protect the cellular macromolecules against the harmful effects of high temperature.

Recently, it was clarified that DIP is synthesized from glucose-6-phosphate in 4 steps (Figure 1.3.3)[181,182]. First, glucose-6-phosphate is converted to L-*myo*-inositol-1-phosphate by NAD^+-dependent inositol phosphate synthase (IPS). Then, the inositol-1-phosphate is activated to CDP-inositol via the activity of CTP:inositol-1-phosphate cytidylyltransferase (IPCT). In the next step, CDP-inositol is condensed with inositol-1-phosphate to DIP-phosphate (DIPP) by DIPP synthase (DIPPS). Finally, DIPP is dephosphorylated to DIP by the action of inositol monophosphatase (IMP).

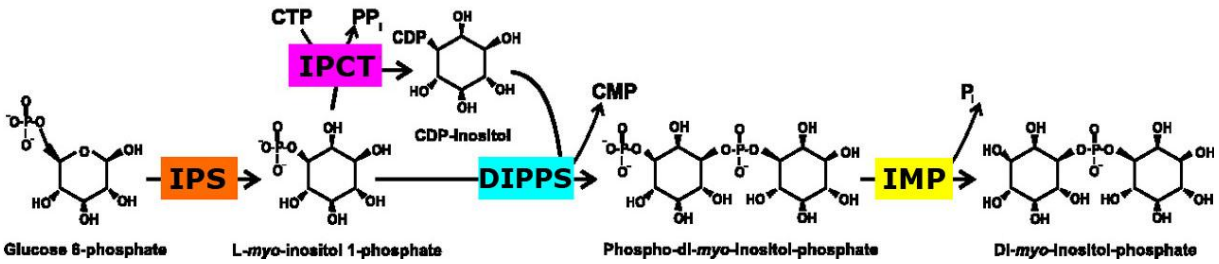

Figure 1.3.3. DIP synthesis pathway. Adapted from Rodionov *et al.*[181]

Meanwhile two of the enzymes in the DIP synthesis pathway were known for a long time[183,184], the two others (IPCT and DIPPS) have only recently been discovered[181,185]. It was shown that in some organisms they are conducted by two parts of a bifunctional enzyme IPCT-DIPPS, meanwhile in others the same domains are coded in separate genes. The DIPPS enzymes and the DIPPS domains of the bifunctional enzymes are the members of the CDP-OH transferase family (Section 1.3.1).

One of the best studied IPCT-DIPPS proteins is the enzyme from the hyperthermophilic archaeon *Archaeoglobus fulgidus*[185]. It was shown that its cytidylyltransferase domain is absolutely specific for CTP and l-*myo*-inositol-1-phosphate and that the DIPP synthase domain used only l-*myo*-inositol-1-phosphate as an alcohol acceptor, but could recognize CDP-glycerol, as well as CDP-l-*myo*-inositol and CDP-d-*myo*-inositol, as alcohol donors.

Crystallographic structure of the IPCT domain was recently determined at a resolution of 1.9 Å[186]. The protein revealed a Rossman-like fold characteristic for nucleotidyltransferase family enzymes (Figure 1.3.4). The domain has a central mixed β-sheet with six parallel β-strands and one antiparallel. These are surrounded by six α-helices. The domain could not be co-crystallized with ligands to determine the active site and ligand binding sites. However, based on the structures of homologous enzymes, it was possible to predict the binding sites of CDP and inositol-1-phosphate (Figure 1.3.4).

Structure of the DIPPS domain remained unknown, and its determination was one of the goals of the present work, as well as analysis and prediction of the ligand binding site (Section 3.7).

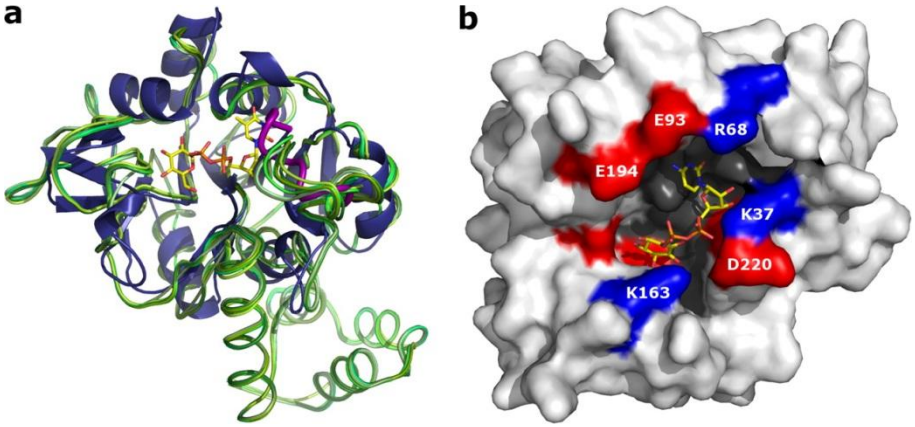

Figure 1.3.4. Active site of the *Archaeoglobus fulgidus* IPCT-DIPPS IPCT domain[186]. a) Superposition of the IPCT (cartoon in blue) with several structures of homologous protein RmlA. RmlA is in the apo form (PDB code 1FZW; ribbon in dark green), and in complex with different ligands (PDB codes 1FXO, 1G0R, 1G2V, 1G1L, 1G23, and 1G3L; ribbons in different shades of green). dTDP-d-glucose (from 1G1L) is shown in sticks. b) IPCT active site pocket with fitted CDP-inositol; residues at the entrance are labeled. The figure is taken from Brito *et al.*[186]

2 Materials and methods

2.1 Protein expression and purification

2.1.1 Expression and purification of HmBRI

Haloarcula marismortui bacterio-opsin gene (*bop*, UniProt ID: Q5UXY6) with D94N mutation was synthesized *de novo*. The nucleotide sequence was optimized for *E. coli* expression using the GeneOptimizer™ software (Life Technologies, USA). The gene was introduced into the pSCodon1.2 expression vector (Staby™Codon T7, Eurogentec, Belgium) via *Nde*I and *Xho*I restriction sites. Consequently, the expressed construct harbored an additional C-terminal tag with a sequence LEHHHHHH. For protein expression, the E. coli strain SE1 cells (Staby™Codon T7, Eurogentec, Belgium) were transformed with the pSC-o*Hm*BRI-His_6 plasmid. The cells were grown at 37º C in shaking baffled flasks in an auto-inducing medium ZYP-5052[187] containing 100 mg/L ampicillin. After the glucose level in the growing bacterial culture dropped below 10 mg/L, the incubation temperature was reduced to 20º C and incubation continued overnight. Collected cells were disrupted using the M-110P Lab Homogenizer (Microfluidics) at 25000 psi in a buffer containing 20 mM Tris-HCl pH 8.0, 5% glycerol and 50 mg/L DNase (Sigma-Aldrich, USA). Membrane fraction of cell lysate was obtained by ultracentrifugation at 90000 g for 1 h at 4º C. The pellets were resuspended in a buffer containing 20 mM Tris-HCl pH 8.0, 0.1 M NaCl and 1% DDM (Anatrace, Affymetrix, USA). All-trans-retinal (Sigma-Aldrich, USA) was added to 10 µM and immediate red shift of the solution color was observed. The mixture was left overnight for solubilization. Insoluble fraction part was removed by ultracentrifugation at 90000 g for 1 h at 4º C. The supernatant was loaded on Ni-NTA column (Qiagen, Germany) and the His-tagged protein was eluted in a buffer containing 20 mM Tris-HCl pH 7.5, 0.1 M NaCl, 50 mM EDTA, 0.01% DDM. The eluate was subjected to size-exclusion chromatography (125 ml Superdex 200 PG, GE Healthcare Life Sciences, USA) in a buffer containing 50 mM NaH_2PO_4/Na_2HPO_4 pH 7.5, 0.1 M NaCl, 0.01% DDM. Protein-containing colored fractions were collected and concentrated to 40 mg/ml for crystallization.

2.1.2 Expression and purification of ESR

ESR with a C-terminal hexahistidine tag was expressed in *Escherichia coli* strain Rosetta2(DE3)pLysS and purified as described[57]. ESR was growing in fermenter for 3 days at

30 C in 2xZYM5052 autoinduction medium[187] with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol. After obtaining a membrane fraction ESR was extracted by incubation in buffer 50 mM Tris 1% n-dodecyl-β-D-maltopyranoside (DDM), 10 mM imidazole pH 8,0 overnight at +4C. The solubilized membrane fraction was purified on Ni-sepharose 6 Fast Flow (GE Healthcare) column, washed with buffer 1 50 mM sodium phosphate, 500 mM NaCl, 2 M urea, 0,1% DDM, 10 mM imidazole pH 8,0 and buffer 2 50 mM sodium phosphate, 200 mM NaCl, 0,1% DDM, 20 mM imidazole pH 8,0 and eluted with buffer 50 mM sodium phosphate, 200 mM NaCl, 0,1% DDM, 0,01% Na azide, 300 mM imidazole pH 7,4. After removing imidazole the subsequent sample concentration was performed on Amicon ultrafiltration device with a regenerated cellulose membrane (10 kDa MWCO). The final concentration of ESR was up to 2 mg/ml.

2.1.3 Expression and purification of KR2

< Not shown as the results have not been published yet >

2.1.4 Expression and purification of NpSRII

Protein preparation and crystallization were performed as described previously[188,189]. The NpSRII gene was cloned into a pET27bmod expression vector with a C-terminal 7x-His tag. The protein was expressed in E. coli strain BL21 (DE3), and purified as described[188,189]. After removal of imidazole by DEAE-chromatography, NpSRII-His was reconstituted into purple membrane lipids (protein to lipid ratio 1:35). After filtration, the reconstituted protein was pelleted by centrifugation at 100,000 g. For resolubilisation the samples were resuspended in a buffer containing 2% n-octyl-β-D-glucopyranoside (OG) and shaken for 16 h at 4°C in the dark. The resolubilized protein was isolated by centrifugation at 100,000 g.

2.1.5 Expression and purification of IPCT-DIPPS

The *Archaeoglobus fulgidus* AF0263 gene encoding for IPCT/DIPPS was cloned into the expression vector pET52b(+) (Novagen) using SacI and XmaI restriction sites, with N-terminal Strep tag II and C-terminal His_{10} tag. The resulting plasmid was verified by DNA sequencing. The expression vector was transformed into *Escherichia coli* C43 DE3. Cells were cultivated in terrific broth medium at 37°C in a 30 l fermenter with pH controlled at 7.4, and induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside when the OD_{600} reached 1.0. The temperature was subsequently lowered to 30°C. Cells were harvested 6 h following induction and resuspended in lysis buffer (50 mM sodium citrate, 50 mM sodium phosphate,

200 mM NaCl, 5% glycerol, 100 mM sucrose, 1 mM ethylenediaminetetraacetic acid, 0.5 mM phenylmethanesulfonylfluoride, 2 mM β-mercaptoethanol and protease inhibitors cocktail tablet SIGMAFAST; pH 7.5). Cells were disrupted using a constant flow cell disrupter, and cell debris and inclusion bodies removed by centrifugation at 10,000*g* for 30 min. *E. coli* membranes were isolated by ultracentrifugation at 200,000*g* for 2 h. The protein was extracted from the membranes by solubilization in 2% Triton X-100, 20% glycerol, 1 M NaCl, 50 mM phosphate buffer, 50 mM imidazole, 100 mM sucrose, 2 mM β-mercaptoethanol and 1 mM $MgCl_2$ (final pH 7.4 at 277 K) for 2 h. Insoluble fraction was separated by 1 h centrifugation at 100,000*g*. The supernatant was loaded onto HisTrap HP column (GE Healthcare). The loaded sample was washed with buffer A (0.1% Triton X-100, 50 mM phosphate buffer, 50 mM imidazole, 10% glycerol, 1 M NaCl; pH 7.4), followed by an additional wash with buffer A supplemented with 110 mM imidazole. The protein was eluted with buffer B containing 0.1% Triton X-100, 50 mM phosphate buffer, 500 mM imidazole, 10% glycerol, 300 mM NaCl, 1 mM $MgCl_2$; pH 7.4. Immediately after elution the sample was supplemented with 1 mM β-mercaptoethanol and concentrated to ~20 mg/ml in centrifugal filter unit with membrane of 100 kDa molecular weight cut-off.

2.2 Crystallization details

All the proteins were crystallized using the *in meso*[19,190] approach, similarly to the sensory rhodopsin-transducer complex[125]. The solubilized protein in the crystallization buffer was added to the monooleoyl-formed lipidic phase (Nu-Chek Prep, USA) and covered with a precipitant solution. The crystals were grown 22º C.

The best HmBRI crystals were obtained using the protein concentration of 20 mg/mL and the 0.1M Tris pH 8.8 and 2.6M Ammonium Sulfate precipitant solution from the Qiagen Cubic Phase I screening kit (Qiagen, Germany). Regular shaped hexagonal crystals reaching 100 µm in size appeared in approximately 4 weeks.

The best ESR crystals were obtained using the protein concentration of 35 mg/ml and the 0.1 M Sodium acetate, 0.2 M Sodium malonate, 12% (w/v) PEG 3350 pH 4.6 precipitant solution (Qiagen, Germany).

The best NpSRII crystals were obtained using the protein concentration of 20 mg/ml and the 1 M Na/K-PO_4 pH 5.1 salt with 0.3 M of trehalose.

Finally, the best IPCT-DIPPS crystals were obtained using the protein concentration of 20 mg/ml and the 4-12% PEG 550-6000, 0.2-2 M sodium malonate, pH 5.8-8.2 precipitant solution (multiple conditions resulted in highly diffracting crystals).

2.3 Data collection and processing

The X-ray diffraction data for all the presented crystals were collected at the European Synchrotron Radiation Facility (ESRF) as follows:

Protein	Wavelength (Å)	Beamline(s)
HmBRI	0.976	ID23-1
ESR	0.934 and 0.976	ID14-1 and ID23-1
KR2	0.976	ID29
NpSRII	0.934	ID14-1 and ID23-1
IPCT-DIPPS	0.8726	ID23-1 and ID23-2

The diffraction data were usually integrated using either MOSFLM[191] or XDS[192]. The diffraction data were analyzed using the POINTLESS software[193] and scaled using the SCALA software from the CCP4 program suite[194].

2.4 Crystallographic structure determination

For all the presented structures, the initial phases were obtained by the molecular replacement (MR) method using MOLREP[195]. Then, the initial MR model was subjected to the automatic model building software ARP/wARP (ref) and PHENIX Autobuild (ref). After that the model was iteratively refined using REFMAC5[196] and Coot[197]. The final steps of the refinement were usually conducted using the PHENIX software[198].

2.4.1 HmBRI structure determination

For HmBRI, the space group was determined to be P321. The crystals were twinned, with the twin fraction in the range 0–25%. The crystal with the best diffraction was not twinned and was used for structure determination. The final data were collected in two wedges from different parts of the same crystal. The diffraction was slightly anisotropic (I/σI of ~2.5 at 2.5 Å along the directions *a* and *b*, I/σI of ~2.0 for all the reflections in the range 2.5÷2.65 Å along the direction *c*), but the data were treated isotropically. The structure was solved using the *Hm*BRI homology model built with RaptorX[199]. The R-factors were ~45% immediately after the MR step and ~40% after 4 cycles of automatic refinement with REFMAC5[196]. There

is one monomer in the asymmetric unit in this space group. The initial MR model was then iteratively refined using REFMAC5[196] and Coot[197]. The final steps of the refinement were conducted using the PHENIX software[198].

2.4.2 ESR structure determination

For ESR, the molecular replacement solution was found in the spacegroup P321. There are two monomers in the asymmetric unit in this solution. Although the overall configuration of the proteins in the crystal is close to the higher-order symmetry $P6_322$, multiple analyses show that this is not correct. Typical R_{merge} values are 1.5 times higher if the data are integrated in $P6_322$, and systematic absences, characteristic of $P6_322$, are not observed. Also, the molecular replacement solution could not be found in $P6_322$. In P321, similar solutions were found using either XR coordinates (PDB ID 3DDL[84]) or the homology model of ESR, based on PDB ID 3DDL. The homology model created using the SWISS-MODEL server[200] was eventually used that resulted in the R-factors of ~45% after the MR step.

The initial MR model was iteratively refined using intensity-based twin refinement in REFMAC5[196] and Coot[197]. During the starting stage of the refinement, the medium NCS restraints were applied to the monomers. After most of the model was built, release of the NCS restraints led to the improvement of the R and R_{free} factors.

The initial atomic model was obtained using the 2.9 Å data from the twinned crystal (twin fraction ~40%) and was later refined against the 2.3 Å data from the crystal that was not twinned. Along with the protein structure, positions of 33 water molecules and 30 lipid tail fragments were determined. The final refinement, including the TLS treatment of the B-factors, was conducted using the PHENIX software suite[198].

2.4.3 KR2 structure determination

< Not shown as the results have not been published yet >

2.4.4 NpSRII structure determination

Structure of the NpSRII in the ground state was obtained by iterative refinement of the PDB ID 1H68 model[82]. The obtained structure was used as a reference for the active state structure determination.

The active state diffraction data were usually anisotropic. Consequently, to retain the meaningful data, the resolution was cut off anisotropically at 2.5/2.9/2.6 Å using the anisotropy server[202].

The q-weighted difference density maps built with CNS[203] were used for preliminary analysis of the structural changes[204]. Most differences were located in the retinal-binding pocket and along the F and G helices. Isomerization and deviation of the retinal out of the F-helix, as well as displacement of W171 are clearly seen for all datasets. Changes for different crystals were similar.

Analysis of the active state structure began with the rigid body refinement of the ground state structure that was minimized in CNS[203] in order to equilibrate the model with CNS parameter set. After that a part of the protein model, corresponding to the ground state was fixed, and the other part, corresponding to the intermediate was subjected to simulated annealing. For the annealed part of the model, a less ordered loop and terminal residues (1-2, 28-32 and 217-219) were removed, along with surface water molecules and small lipid fragments. Internal water molecules, expected to be present but not clearly seen at present resolution, were harmonically restrained to positions in the ground-state structure. After simulated annealing procedure conformations of some residues were manually changed towards ideal geometry.

Different active state occupancies were tested to determine the correct one that was consequently used to obtain the final structure (Section 3.4.4).

2.4.5 IPCT-DIPPS structure determination

For structure determination, the diffraction data from two IPCT-DIPPS crystals were merged. The merging is justified as the two crystals had very similar unit cell parameters (41.35/107.66/124.25 and 41.39/107.77/124.01 Å correspondingly), and the resolution of the resulting data was better by 0.2 Å.

The initial phases were obtained via Molecular Replacement with MOLREP[42] using the crystallographic structure of the soluble domain (PDB ID 2XME[186]) as a starting model. The electron density maps revealed that the soluble domains form layers with a separation of ~ 70 Å that is enough for the membrane-like layer containing transmembrane domains. Additional search for alpha-helical fragments resulted in establishment of position of five transmembrane alpha-helices. Refinement of the fragments' structure with Refmac5[43] and PHENIX[44] resulted in improvement of the electron density maps, and following identification of the sixth transmembrane helix and 2 alpha-helices parallel to the membrane plane. Eventually, identities of the side-chains were determined. The final R and R_{free} factors are 24.3% and 30.0% correspondingly. There are 96.3% Ramachandran-favored, 3.2% allowed and 0.5% not allowed residues in the final model.

2.5 Details of the modeling studies: simulations of the NpHtrII HAMP domain region

2.5.1 Initial data and models

All sequences and domain assignments were taken from reviewed UniProt database (16) records P33741 (HsHtrI), Q9HP81 (HsHtrII) and P42259 (NpHtrII). Initial model for NpHtrII α-helical fragment 135-150 was taken from the PDB entry 2RM8 (9). HAMP domain model for homology modeling was taken from the PDB entry 2ASW (8). Initial models of α-helices for which no structure was present were built using PyMOL (DeLano Scientific, Palo Alto, CA, USA.). For homology modeling of the NpHtrII HAMP domains MODELLER (17) was used. Some energy minimizations and structure manipulations in the dihedral subspace were performed using SAMSON symmetry module (18).

2.5.2 Molecular dynamics simulations

All molecular dynamics simulations were conducted using NAMD2 program (19) with CHARMM27 parameters (20). Preparations of initial models and data analysis using our own tcl-scripts were performed in VMD (21). As a first step, hydrogen positions of a model in a vacuum were minimized for 50 standard NAMD minimization steps. This was followed by addition of a water box with a padding of 8 Å. Sodium and chloride ions were added in total concentrations of 0.5, 1, 2 or 4 M, accordingly, for different simulations, in such amounts that the system had zero total charge. Simulation in the water box was conducted with periodic boundary conditions. The water box was minimized for 50 steps and then was subjected to 1 ps of molecular dynamics, with protein atoms frozen. After that the whole system was minimized for 1000 steps and then heated to 310 K in steps of 10 K and 0.1 ps per step. Integration time step was 2 fs, SHAKE algorithm was used to keep bonds between hydrogens and heavy atoms rigid. The temperature of 310 K was maintained with Langevin thermostat with the damping coefficient of 5 ps^{-1}. The pressure of 1 bar was maintained with Langevin piston with parameters as follows: period of 100 fs, decay of 50 fs. For electrostatics calculations particle-mesh Ewald method was used.

2.5.3 Listing of the simulated systems

Model name (numbers of residues included in the simulation are in parenthesis)	Preparation of initial model	Simulation Time	Structure stability	Residue packing conservation
1. NpHtrII inter-HAMP region (135-153)	Two ideal helices of the corresponding aminoacid sequence were built and then manually positioned to face each other with hydrophobic groove and approximately zero longitudinal shift. All residues are in default protonation states at neutral pH.	10 ns	See Figure 2.5.1 for the RMSD data. The system samples a variety of conformational states and longitudinal shifts of up to 2 Å in both directions. This is extensively discussed in the main text of the article.	The α-helical structure is not changed during the whole simulation. Hydrophobic grooves of the coiling helices remain turned to each other. Thus the packing is maintained.
2. NpHtrII inter-HAMP region (135-153) PMF calculation	Resulting model of the simulation 1 was taken.	For each of the 4 tried molarities (0.5, 1.0, 2.0 and 4.0 M) 7 trajectories of 20 ns	The same as for the model 1.	The same as for the model 1.
3. HsHtrII inter-HAMP region (356-400)	Two ideal helices of the corresponding aminoacid sequence were built and then manually positioned to face each other with hydrophobic groove and approximately zero longitudinal shift. All residues are in default protonation states at neutral pH.	25 ns	See Figure 2.5.2 for the RMSD data for the longest trajectory. Ideal helices, prepared at the beginning, connect with their hydrophobic grooves and bend slightly in approximately 1 ns, with the development of longitudinal shift.	The same as for the model 1.
4. HsHtrI inter-HAMP region (106-150)	The same as for the model 3.	28 ns	The same as for the model 3, although the structure is somewhat more flexible. See Figure 2.5.3 for the RMSD data.	The same as for the model 1.
5. Af1503 HAMP domain (276-331)	NMR structure (PDB 2ASW) was taken as a starting model, with all residues in default protonation states at neutral pH.	4 ns	See Figure 2.5.4 for the RMSD data.	Core residues RMSD values (calculated for the heavy atoms) do not generally exceed 2 Å and thus, their positions are not substantially changing during the course of the simulation, and the packing is maintained.

6. Af1503 expanded HAMP domain (276-338)	Resulting model of the simulation 5 was extended with the residues 332-338, manually built in the ideal helical conformation to continue AS2 helices.	8 ns	See Figure 2.5.5 for the RMSD data.	Core residues RMSD values (calculated for the heavy atoms) do not generally exceed 2 Å and thus, their positions are not substantially changing during the course of the simulation, and the packing is maintained.
7. NpHtrII first HAMP domain (85-134)	Homology model was built using MODELLER, based on the NMR structure of Af1503 HAMP domain (PDB 2ASW). All residues are in default protonation states at neutral pH.	4 ns	See Figure 2.5.6 for the RMSD data.	Core residues RMSD values (calculated for the heavy atoms) do not generally exceed 2 Å and thus, their positions are not substantially changing during the course of the simulation, and the packing is maintained.
8. NpHtrII second HAMP domain (155-210)	The same as for the model 7.	4 ns	See Figure 2.5.7 for the RMSD data.	Core residues RMSD values (calculated for the heavy atoms) do not generally exceed 2 Å and thus, their positions are not substantially changing during the course of the simulation, and the packing is maintained.
9. NpHtrII first HAMP domain and inter-HAMP fragment (85-153)	Resulting models of the simulations 1 and 7 were combined and minimized. All residues are in default protonation states at neutral pH.	10 ns	No analysis was made, as this structure is a part of the model 10 and analyzed there.	No analysis was made, as this structure is a part of the model 10 and analyzed there.
10. NpHtrII HAMP domain region – both HAMP domains and the inter-HAMP fragment (85-210)	Resulting models of the simulations 1, 7, and 8 were combined and minimized. All residues are in default protonation states at neutral pH. Model was symmetrized prior to the simulation.	60 ns	The system is quite flexible, which is extensively discussed in the main text of the article. Nonetheless, both HAMP domains and the inter-HAMP region retain their general fold. See Figure 2.5.8 for the RMSD data of the first HAMP domain, Figure 2.5.9 for the RMSD data of the inter-HAMP region and Figure 2.5.10 for the RMSD data of the second HAMP domain.	See Figures 2.5.8-10 for the RMSD data on the HAMP domain residues and Figures 3.5.7-8 for the structures and double-distance matrices. Core residues RMSD values (calculated for the heavy atoms) do not generally exceed 2 Å and thus, their positions are not substantially changing during the course of the simulation, and the packing is maintained.

2.5.4 Evaluation of structure stability of the inter-HAMP and HAMP domain regions in MD simulations

The stability was evaluated using three different measures, presented in three separate graphs in the following figures.

The left graph shows the unaveraged protein backbone atoms' RMSD as a function of time, with the last trajectory's snapshot as a reference. It is seen that the starting equilibration is followed by normalization, during which the RMSD value is practically unchanged and corresponds to the thermal fluctuations. The magnitude of the thermal fluctuations is as expected for typical soluble proteins.

The central graph shows, which RMSD is to be expected for the snapshots separated by the time τ, and thus reflects the dynamical properties of the corresponding construct. It is calculated as follows:

$$RMSD(\tau) = \frac{1}{N}\sum_{i=0}^{N} RMSD(i, i+n), \text{ where } n = \frac{\tau}{timestep} \text{ and } N = \frac{total\ time}{timestep} - n,$$

where *RMSD(i, i+n)* denotes RMSD between the frames *i* and *i+n*. The time step used in the calculation is 50 ps. RMSD values are calculated for the backbone atoms, and the alignment was also performed using the backbone atoms. RMSD values at the maximum time separations (last 10-20% of the trajectory) may be misleading, as they were obtained by averaging of few data points. These values are also typically higher compared to the rest of the graph, because they represent mostly the difference between the initial unrelaxed structure and the final equilibrated one.

The right graph shows the RMSD for each residue, measured to the last trajectory's snapshot as a reference and averaged over the whole trajectory. The alignment of the corresponding structure snapshots was done using all the protein's backbone atoms. Data for both protomers are shown, with RMSD values for all heavy atoms in magenta and blue, and backbone atoms in green and yellow. Regions corresponding to the residues in alpha-helical structure are shaded gray. Large RMSD values (>4 Å) for some sidechains usually correspond to exposed arginines.

Following remarks should also be made:

a) the alternate pattern of the all-atoms RMSD (blue and magenta) corresponds to the buried (lesser mobility) and exposed (higher mobility) sidechains. All the backbone atoms have similar RMSD.

b) the mobility of residues in the HAMP domain linker is increased compared to the helical regions.

c) the ends of the polypeptide chain have an increased mobility due to the truncation artifacts. In the full-length native constructs these residues are expected to be more confined.

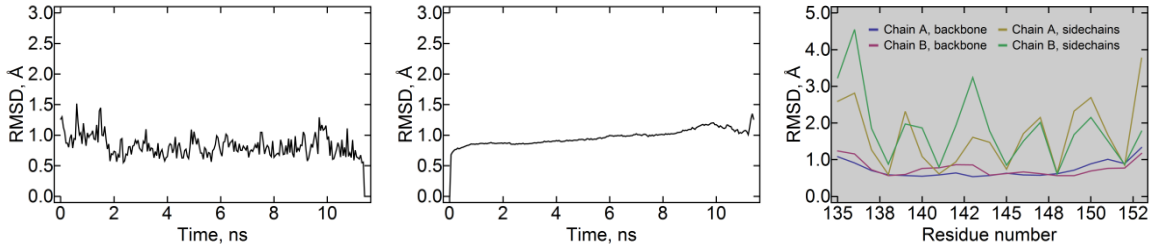

Figure 2.5.1. Structure stability of the NpHtrII inter-HAMP region (model 1 in section 2.5.3).

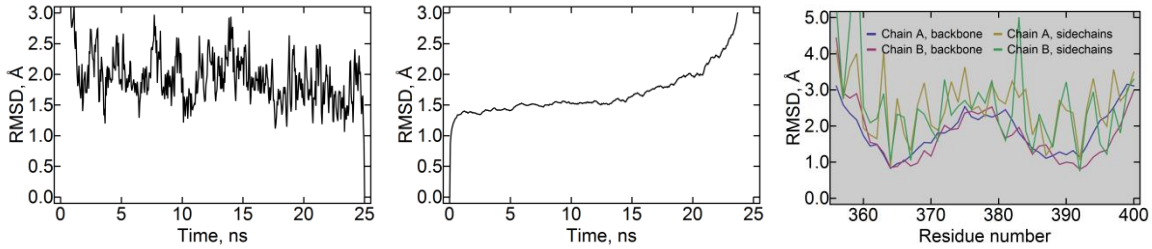

Figure 2.5.2. Structure stability of the HsHtrII inter-HAMP region (model 3 in section 2.5.3).

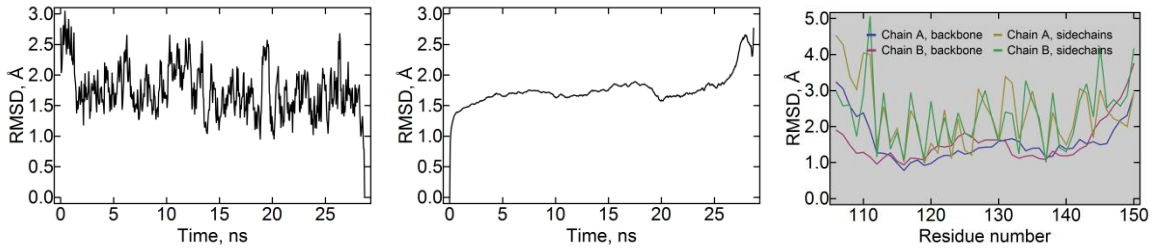

Figure 2.5.3. Structure stability of the HsHtrI inter-HAMP region (model 4 in section 2.5.3).

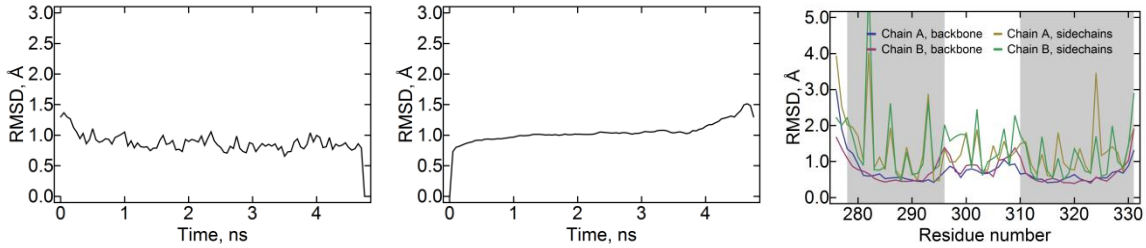

Figure 2.5.4. Structure stability of the Af1503 HAMP domain (model 5 in section 2.5.3).

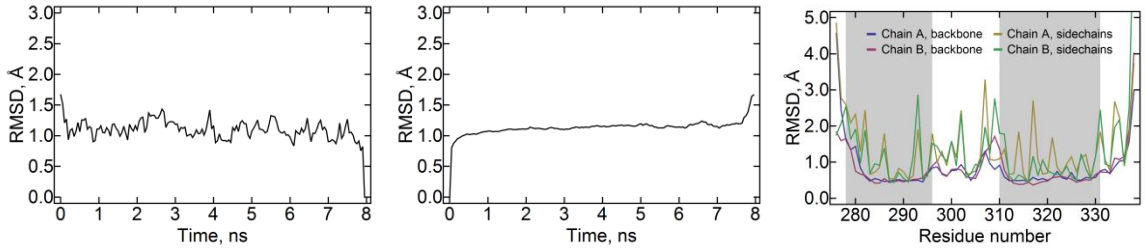

Figure 2.5.5. Structure stability of the Af1503 extended HAMP domain (model 6 in section 2.5.3).

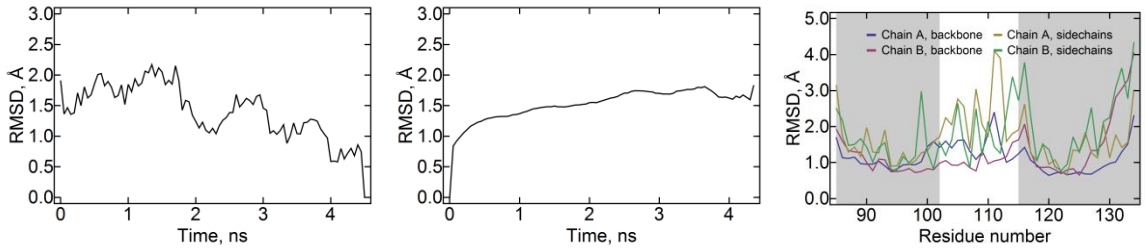

Figure 2.5.6. Structure stability of the NpHtrII HAMP1 domain simulated alone (model 7 in section 2.5.3).

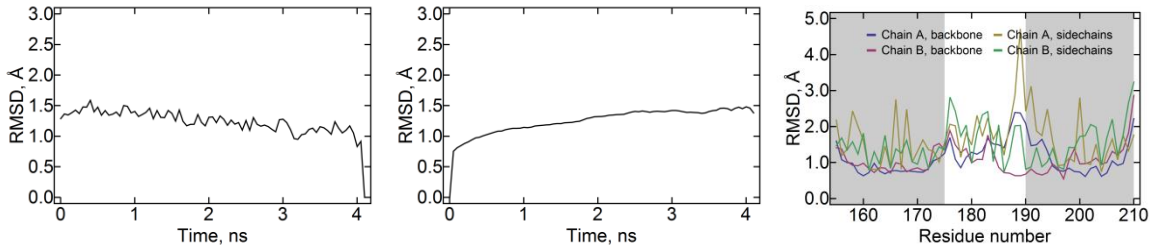

Figure 2.5.7. Structure stability of the NpHtrII HAMP2 domain simulated alone (model 8 in section 2.5.3).

Figures 2.5.8-10 refer to the stability of the parts of the HAMP domain region simulated as a whole (model 10 in Table S1). The structure was initially prepared as symmetric to avoid a bias in the simulation. The asymmetry has developed during the course of the modeling.

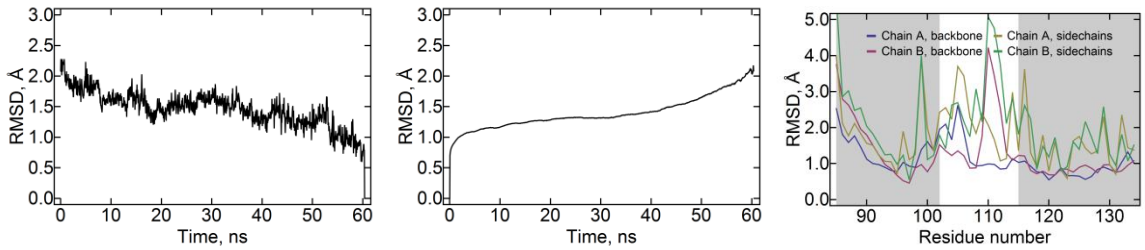

Figure 2.5.8. Structure stability of the NpHtrII HAMP1 domain as a part of the whole HAMP domain region (model 10 in section 2.5.3).

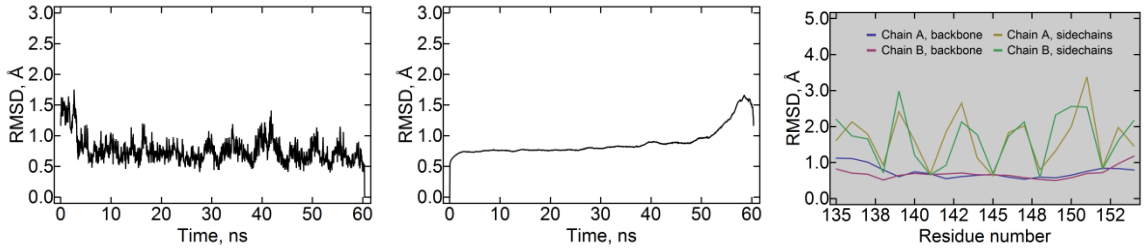

Figure 2.5.9. Structure stability of the NpHtrII inter-HAMP region as a part of the whole HAMP domain region (model 10 in section 2.5.3).

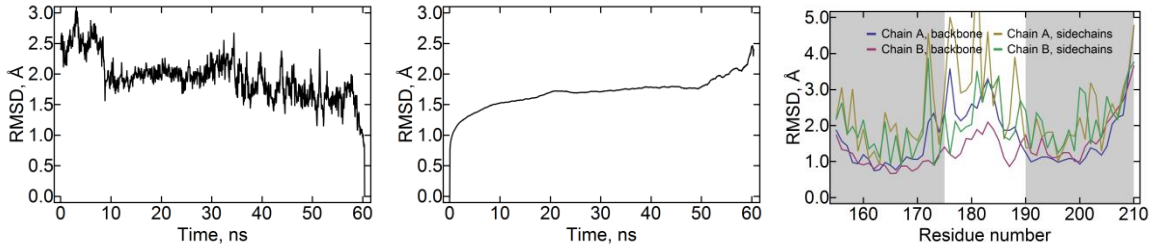

Figure 2.5.10. Structure stability of the NpHtrII HAMP2 domain as a part of the whole HAMP domain region (model 10 in section 2.5.3).

2.5.3 Longitudinal shift calculations

For each frame of the MD trajectory calculations were conducted as follows. First, direction of each helix was determined as the main axis of its inertia tensor, along which the moment of inertia is minimal. Second, the coiled coil axis was calculated as an average of directions of the two individual helices. And finally, the vector difference in positions of centers of mass of respective helices was projected on the coiled coil axis to obtain the relative shift of helices. For these calculations, only positions of the backbone atoms were used.

2.5.4 Free energy calculations

To determine the free energy of a system as a function of longitudinal shift we used the umbrella sampling technique (22) followed by either weighted histogram analysis method (WHAM) (23) or multiple Bennett acceptance ratio method (MBAR) (24). We used WHAM implementation by Alan Grossfield (http://membrane.urmc.rochester.edu/Software/WHAM/WHAM.html, version 2.04) and MBAR implementation by John Chodera and Michael Shirts (https://simtk.org/home/pymbar).

Sampling positions along the longitudinal shift were -2.1, -1.4, -0.7, 0, 0.7, 1.4, 2.1 Å with force constants of 10 kcal/(mol·Å) for each simulation. Shift values larger than 2.4 Å were not sampled as we believe that they are not realized in properly folded proteins. Free energy values for such shifts would also be affected by erroneous contributions from artificial exposition of hydrophobic surfaces at the ends of the inter-HAMP model. The system was modeled for at least 20 ns around each position. As the protomers are absolutely identical, the resulting free energy should also be symmetric as a function of the longitudinal shift. Thus, convergence was postulated if the energy difference between symmetrical positions (shifts of the same absolute value, but different signs) was less than 0.2 kcal/mol.

2.5.5 HAMP domain axis calculation

To calculate the HAMP domain axis, the directions of helices AS1 and AS2 of both protomers were averaged.

2.5.6 Protonation state of the ionizable residues

Based on the data describing the environmental pH (25) as well as the optimal pH for enzymes from halophilic archaea (26), we expect the inner-cell pH to be in the range of 7.0-9.0. To make the correct choice of protonation states of ionizable residues, their pK_a values

were predicted in corresponding environments. Predictions were performed with MCCE2 multiconformation continuum electrostatics package (27).

In the NpHtrII HAMP domain region model (number 10 in Section 2.5.3), 98 residues in both protomers were treated as ionizable. Multiple conformers for these residues were built with the total number of conformers 5247. A conformer is defined as a heavy atoms rotamer of a certain side chain, in which protons are added, their positions optimized, and a certain ionization state is chosen. Many conformers were included into the prediction because conformer sampling improves the match between experiment and calculation for individual residues and dramatically diminishes the dependence on the starting structure (26). For our system, no residues were found to have the pK_a value in the range 7-11. Therefore, all residues in the initial models of this simulation were left in their default protonation states.

We also studied the possibility of protonation-deprotonation events during the simulation of the NpHtrII inter-HAMP region. pKa values of the key rechargeable residues (D137, R142, D144) were predicted along the PMF trajectories. These residues, along with the non-rechargeable Q149, were subjected to extensive multiconformer trials, corresponding to different rotamers and ionization states. Conformers for other residues did not include side chain rotamers but only different ionization states. The total number of conformers was about 200-300 for different snapshots of the trajectory. Calculations were performed for the simulation at molarity 0,5 M, because at higher molarities pK_a shifts caused by solvation contributions would be even weaker, as follows from the Debye – Huckel theory of solvation. The total length of this simulation was 140 ns, snapshots were taken every 200 ps. For both D137 and D144 peaks of their distributions were shifted from the pK_a value in solution 3.71 not more than by 0.1, with standard deviations of about 0.65 for D144, and 0.42 for D137. In case of R142 the peak of its pKa distribution is shifted from the pK_a value in solution 12.1 to 13.1 with the standard deviation of 0.4. No points were observed in the range of 7-11 pH. Therefore, the probability for the key residues to switch its default protonation state is negligibly small at physiological conditions (pH 7.0 – 9.0).

2.6 Details of the modeling studies: simulations of the NpHtrII first HAMP domain

2.6.1 Initial model preparation

In all the simulations we have studied the first HAMP domain of *Natronomonas pharaonis* HtrII (residues 85-133). The initial model was obtained by automated homology modeling procedure by SWISS-MODEL server[201] using a HAMP domain from putative protein Af1503, PDB ID 2ASW[148] as a template, similarly to what was described previously[4,205]. The Af1503 HAMP domain has currently the closest sequence to the NpHtrII HAMP1 among the HAMP domains of known structure. For the helices AS1 and AS2, backbone structure was not changed, whereas the side chains were mutated to the correct ones. The inter-helical linker of NpHtrII HAMP1 is shorter by one residue than that of Af1503 HAMP, and thus it could not be modeled without introducing structural perturbations. However, the modeled conformation is highly similar to the one observed in previous simulations[4,205], with hydrophobic residues V107 and L109 facing towards the hydrophobic core of the coiled coil. The modeled protomers were aligned by least-squares method to the dimeric HAMP domain structure PDB ID 2ASW. The resulting homodimer was perfectly symmetrical and was used as a starting structure in molecular dynamics simulations.

2.6.2 Molecular dynamics simulations

Molecular dynamics simulations were conducted using the GROMACS software version 4.5.3[206] with forcefields CHARMM22[207] with CMAP correction[208] and AMBER ff99-SB-ILDN[209,210] (Section 2.6.3). The TIP3P model water model was used. All the simulations were run identically. First of all, the initial model was solvated in a water box of approximately 3400 water molecules with NaCl concentration of 1 M. This salt concentration reflects the high salinity of the environment *Natronomonas pharaonis* lives in. The box size was chosen so that the minimal distance between the atoms of periodic images of the protein was 1.6 nm. The resulting system was minimized with a steepest descent algorithm using a tolerance of 1,000 kJ mol^{-1} nm^{-1} and a step size of 0.01 nm. After that, the solvent was equilibrated for 10 ps at a constant volume (NVT ensemble) and then for 10 ps at a constant pressure (NPT ensemble). The protein atoms were harmonically restrained during the solvent equilibration. The starting frames for the simulations #3 and #4 were chosen randomly from those of the simulation #1 with a needed HAMP domain state and are shown in Figure 3.6.3a. These

simulations were preceded by several (less than 10) energy minimization steps to allow for the forcefield difference. The step size was 2 fs. The Coulombic interactions, van der Waals interactions and the short-range neighborlist were cut off at 1 nm. Electrostatics was calculated using a fourth-order particle mesh Ewald method[206] with a Fourier spacing of 0.16 nm. The temperature was kept at 310 K using the modified Berendsen thermostat[211] with a time constant of 1 ps. The pressure was kept equal to 1 bar using the Parrinello-Rahman scheme[212] with a time constant of 1 ps and a compressibility of 4.5×10^{-5} bar^{-1}.

To restrain alpha-helicity, harmonic potentials were applied to the φ and ψ dihedral backbone angles. The average values of -63.8° and -41.1° correspondingly were used[213]. The rigidity constant was chosen so that the 10°-deviation corresponded to the energy penalty of 1 kT (at 310 K). For the production runs, 4 residues at the N-terminus and 4 residues at the C-terminus were restrained. Dihedral backbone angles of other AS1 or AS2 residues were not restrained.

2.6.3 Listing of the simulated systems

Simulation #	Starting coordinates	Forcefield	Number of trajectories and their length	Average RMSD of backbone atoms
1	Symmetrical homology model based on the Af1503 HAMP domain	CHARMM22 with CMAP correction	10 × ~205 ns	1.1 Å
2	Symmetrical homology model based on the Af1503 HAMP domain	AMBER ff99-SB-ILDN	10 × ~205 ns	1.3 Å
3	"Resting state" conformations from different trajectories of simulation #1	CHARMM22 with CMAP correction	5 × ~205 ns	1.2 Å
4	"Active state" conformations from different trajectories of simulation #1	AMBER ff99-SB-ILDN	3 × ~205 ns	1.4 Å

2.6.4 Principal components analysis

Principal components analysis[214] (PCA) was conducted using the tools g_covar and g_anaeig of the GROMACS suite[206]. For the analysis, coordinates of heavy backbone atoms (N, C, C_α, O) of the HAMP domain alpha-helices were used. Atoms of the linker were not included in the analysis as the linker was flexible and its conformation correlated weakly with the state of the alpha-helices. The principal component 1 was determined for the concatenated trajectory that includes all the simulations and was used throughout the analysis.

2.7 Details of the IPCT-DIPPS modeling studies

2.7.1 Modeling of the soluble domain dimerization

The model of the IPCT-DIPPS dimer with both the soluble and transmembrane domains dimerized was prepared as follows. First, the symmetry axes for the transmembrane domain (present data) and the soluble domain (PDB ID 2XME[186]) were determined using a VMD[215] script as principal axes (eigenvectors of the inertia matrices). Then, the domain homodimers were aligned so that their symmetry axes coincided. As the model is hypothetical and serves to illustrate the point that both domains can be dimerized at the same time, there is a considerable leeway in positioning the structures. Consequently, the position there the active sites of the domains are almost across each other was chosen (Figure 3.7.6c). However, such rearrangement of the soluble domain relative to the transmembrane domain is not possible without the displacement of the connector helix 1, as the loop between the soluble domain and connector helix 1 is very short. So, the position of the connector helix 1 was remodeled. Displacement of the connector helix 1 is possible as the loop between the connector helices 1 and 2 is disordered in the crystallographic structure, and of sufficient length for large motions of the connector helix 1.

2.7.2 Docking of the IPCT-DIPPS ligands

Autodock Vina software[216] was used for molecular docking along with the Autodock/Vina plugin for PyMOL to set-up the simulations. Autodock Vina uses as default the Iterated Local Search global optimization algorithm for an efficient global search. For the local optimization, Autodock Vina uses a quasi-Newton, Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. The docking was performed within the axis-aligned bounding box in the cytoplasmic part of the transmembrane domain that contains all three binding pockets visualized using HOLLOW[21] (Figure 3.7.8a). The volume of the search space was about 20,000 A^3. During the docking, ligands were flexible, i.e., all non-aromatic bonds were rotatable. The magnesium ion was included in the calculations. First, the smaller compound inositol-1-phosphate was docked around pocket 2. During this simulation, four amino acid residues flanking the binding pocket, Asp-296, Ser-300, Arg-305, and Arg-365, were allowed to change their conformations. Then, after inositol-1-phosphate was fixed in the best position, CMP docking simulations were run with a single flexible residue, Arg-365.

The scoring function used by Autodock Vina is given as a weighted sum of steric interactions, hydrophobic interactions, and, where applicable, hydrogen bonding[216]. Because this software energy (scoring) function, similarly to others used in computational docking, is usually not enough to discriminate the correct docking pose from other poses[217], the 100 top-ranking possible ligand positions were generated by the Vina software, and the best one was chosen according to the biochemical rationales.

3 Results and discussion

3.1 Crystal structure of *Escherichia coli*-expressed *Haloarcula marismortui* bacteriorhodopsin I in the trimeric form

3.1.1 Crystallization and crystal packing of HmBRI

The HmBRI protein was crystallized using the in meso approach[19,190] similarly to the previous work[1,125]. Two types of crystals appeared in the same probes: regularly shaped hexagonal crystals and bunches of needles (Figure 3.1.1a), both approaching 100 μm in the largest dimension. Absorption maximum of HmBRI in the hexagonal crystals is at ~567 nm. While the hexagonally shaped crystals produced good diffraction patterns up to the resolution of 2.3 Å (Figure 3.1.1b, Table 3.1.1), the needles produced overlapping diffraction patterns up to the resolution of 3.5 Å, which could not be indexed (Figure 3.1.1c).

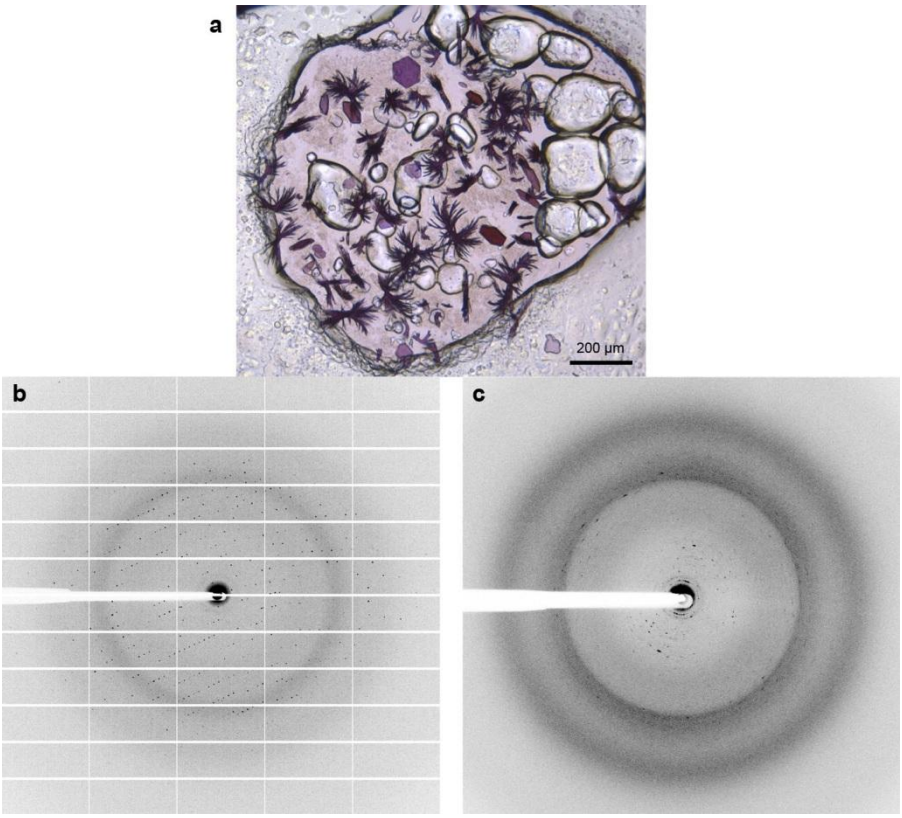

Figure 3.1.1. *Hm*BRI crystals and diffraction. a) *Hm*BRI crystals in the crystallization probe. Two types of crystals are observed: regular shaped hexagonal crystals and bunches of needles. (b) Example of the diffraction from the hexagonal crystals. Resolution at the detector edge is 2.4 Å. (c) Example of the diffraction from needle-shaped crystals. Resolution at the detector edge is 2.5 Å. The diffraction patterns are smeared and observed up to the resolution of 3.5 Å.

The spacegroup of the hexagonally shaped crystals was determined to be P321, cell parameters a, b = 102.9 Å, c = 60.0 Å. The protein is in the same arrangement as archaerhodopsin-2 in the trimeric form (PDB ID: 2EI4[27], space group P321, cell parameters a, b = 98.2 Å, c = 56.2 Å). Typically for in meso crystallization, HmBRI crystals are organized as stacks of the membrane-like layers (Figure 3.1.2). The unit cell spans one such layer, and contains two HmBRI trimers, oriented in opposite directions (Figure 3.1.2a and b).

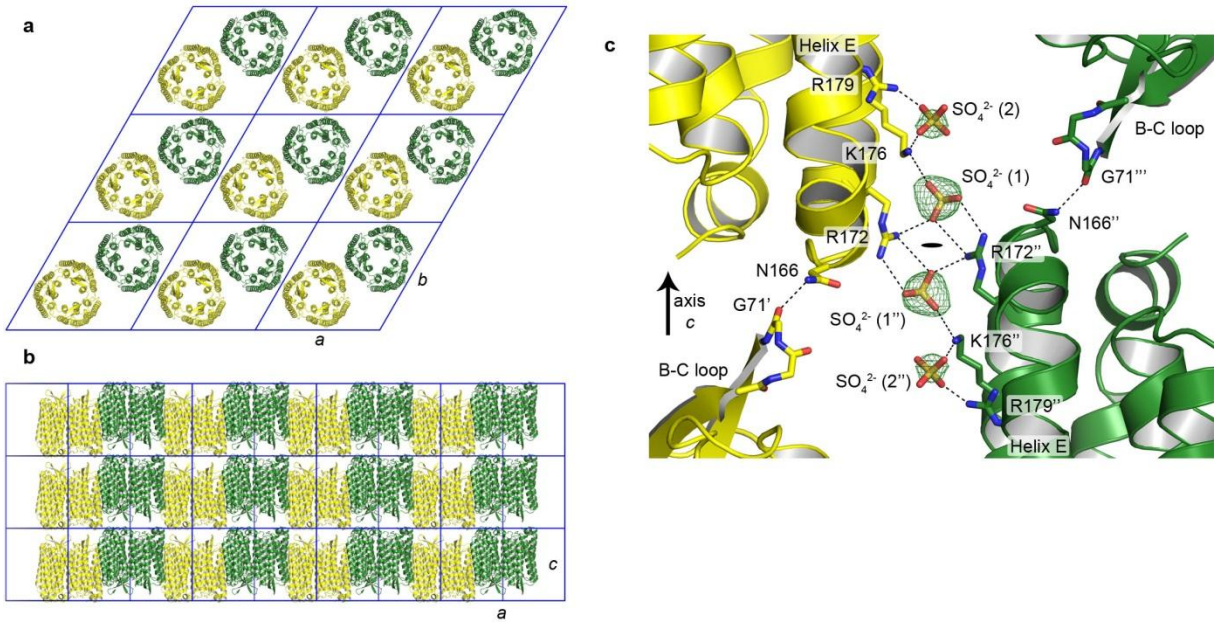

Figure 3.1.2. *Hm*BRI packing in hexagonal crystals. (a) Packing in the membrane plane. There are two trimers in the unit cell (shown in yellow and green) that are oriented in opposite directions. (b) Packing of the membrane-like layers. (c) Crystal contacts between the layers. The contacts rely on the ionic interactions between the SO_4^{2-} ions and the R172 and K176 side-chains, and on the hydrogen bond between the N166 side-chain of one protomer and the G71 backbone oxygen of another. A second putative SO_4^{2-} ion is observed close to the first one that interacts with K176 and R179 side-chains. Potential interactions are marked by dashed lines. F_o-F_c difference electron densities before the insertion of the ions into the model are shown at the level of 3σ. The symbols ', '' and ''' denote different crystallographic symmetry-related HmBRI molecules. The black oval denotes the crystallographic symmetry rotation axis (C_2).

The inter-layer crystal contacts rely on the ionic interactions between the sulfate ions and the side-chains of the helix F lysine 176 and arginine 179 of the opposing HmBRI molecules (Figure 3.1.2c). The second partially occupied sulfate ion binding site is observed close to the first one. However, the second sulfate does not participate in inter-molecular contacts and is bound to the residues K176 and R179 from the same HmBRI molecule. A sulfate ion at the similar position is observed in the structure of deltarhodopsin-3[79], where it is bound by R160

of the helix E and R174 of the helix F. The other inter-layer contact in HmBRI crystals is formed by the hydrogen bond between the G71 oxygen of one HmBRI molecule and N166 amine nitrogen of another (Figure 3.1.2c).

The intra-layer contacts between the HmBRI trimers are mediated mostly by the ordered hydrophobic tail groups of the lipid molecules, and by the two symmetric hydrogen bonds between the amine nitrogen of N180 from the helix F of one protomer and the backbone carboxyl of L202 from the F-G loop of another.

3.1.2 Structure of *Hm*BRI

Similarly to other microbial rhodopsins, *Hm*BRI is organized as a bundle of seven transmembrane helices (A-G) connected by relatively short loops, with the exception of an elongated beta hairpin-forming B-C loop and the extended D-E loop (Figure 3.1.3a). The root mean square deviation (RMSD) of the backbone heavy atoms positions between the HmBRI and HsBR (PDB ID: 1C3W) structures is ~0.6 Å. In the native HmBRI, all the important proton translocation pathway residues, corresponding to D96, D85 and D212, R82, E194 and E204 in HsBR, are conserved. For the crystallization trials, we have used the D94N mutant of HmBRI (which has higher expression level), corresponding to the D96N mutant of HsBR. This mutation is not expected to introduce the structural changes in the protein ground state other than in the vicinity of D94[6,218]. Consequently, we will start with the discussion of the retinal binding pocket and the proton release group of HmBRI.

At the core of the HmBRI retinal binding pocket is the cofactor retinal in the all-trans conformation (Figure 3.1.3b). Bound to the Schiff base is the water molecule W402, coordinated by the aspartates 83 and 216. This arrangement, observed in bacteriorhodopsin-like proton pumps[6,16] as well as in proteorhodopsins[1,86] appears to be critical for the proton pumping.

The proton release group configuration in HmBRI is similar to that in the highest-resolution ground state HsBR structure (PDB ID: 1C3W), as the $R80_{Hm}$ side-chain (corresponds to $R82_{Hs}$) is turned towards the Schiff base, and the $E198_{Hm}$ ($E194_{Hs}$) and $E208_{Hm}$ ($E204_{Hs}$) side-chains are close to each other (Figure 3.1.3c and d). Positions of the water molecules, such as 401, 403, 404, 405 and 407, are similar to those in HsBR (Figure 3.1.3d). However, the extracellular part of the HmBRI's helix G appears to be slightly unlatched, which results in its outward displacement. As a consequence, there is more water-accessible space in the HmBRI proton release region, and two more water molecules are observed (Figure 3.1.3c and d). It is

not clear how this alteration could affect the proton release mechanism[6], which possibly involves formation of Zundel cation[219,220].

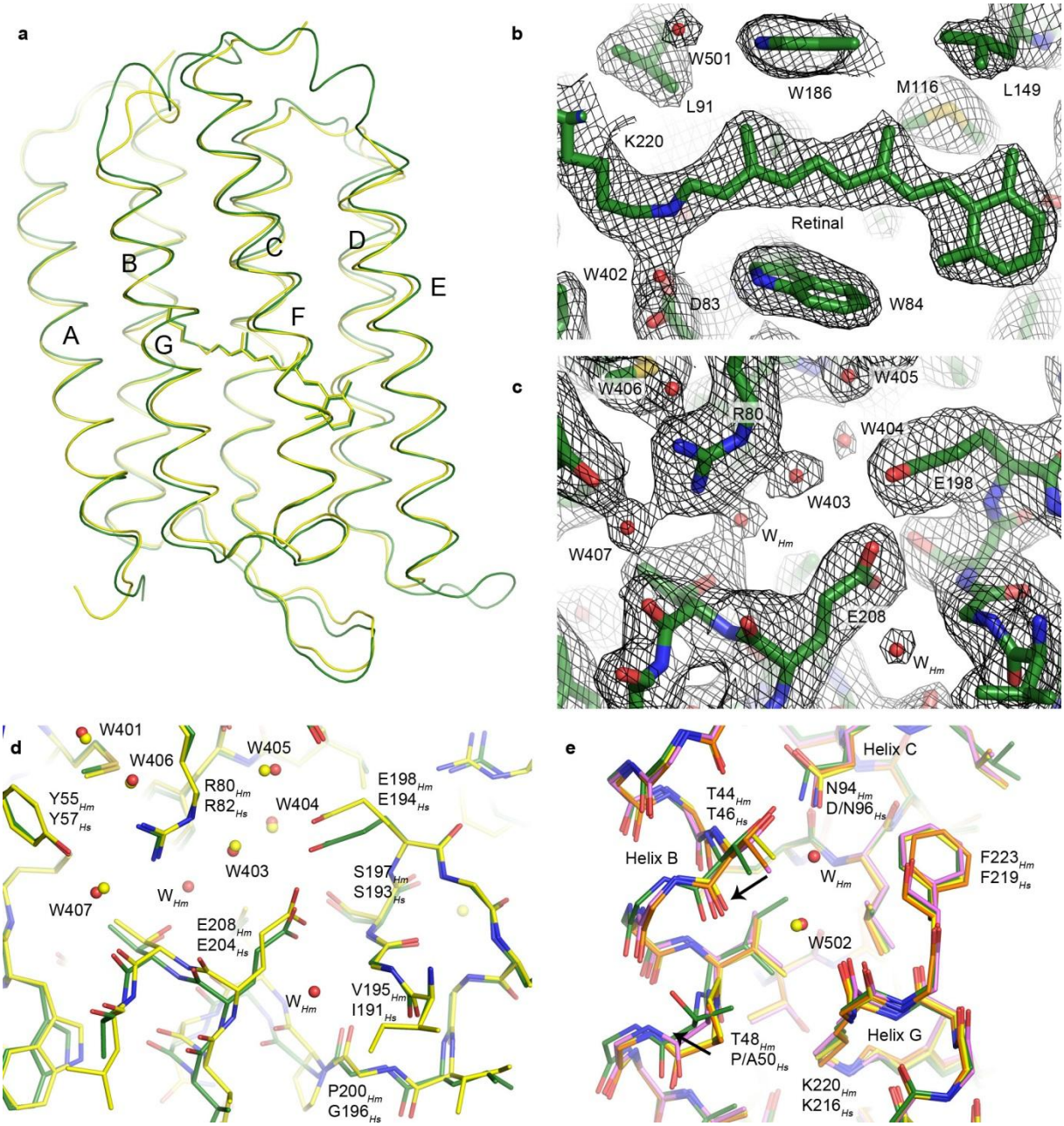

Figure 3.1.3. Crystallographic structure of the *Hm*BRI D94N mutant. (a) Comparison of the *Hm*BRI backbone structure (green) with that of *Hs*BR[16] (yellow). (b) $2F_o$-F_c electron density maps in the vicinity of the retinal. The maps are contoured at the level of 1.5σ. (c) $2F_o$-F_c electron density maps in the proton release region. The maps are contoured at the level of 1.2σ. (d) Comparison of the *Hm*BRI proton release group (green) with that of *Hs*BR[16] (yellow). Overall, the conformations of the side-chains and positions of water molecules are very similar. However, the water accessible space is larger in *Hm*BRI, and additional water molecules are observed (W_{Hm}). One of the reasons for this difference might be the substitution of *Hs*BR' proline 200 with the glycine 204 in *Hm*BRI, that allows unlatching of the extracellular part of the helix G. (e)

Comparison of the *Hm*BRI proton donor region with that of wild-type *Hs*BR[16] (green) and its D96N[218] (orange) and P50A[221] (magenta) mutants. It appears that in *Hm*BRI the effects of the D94N mutation and the $P50_{Hs} \rightarrow T48_{Hm}$ substitution combine and result in a larger displacement of the helix B relative to the helices C and G (black arrows), as similar displacements are present in the P50A and D96N mutants of HsBR, albeit with a smaller amplitude.

Finally, we analyze the HmBRI proton donor region and the effects of the introduced HmBRI mutation D94N and the $P50_{Hs} \rightarrow T48_{Hm}$ substitution (Figure 3.1.3e). Structures of the D96N and P50A mutants of HsBR have been determined before[218,221] and allow us to compare the effect of double "mutation" in HmBRI. Structure of the HsBR D96N mutant reveals the outward motion of T46 side-chain, which is hydrogen-bonded to this residue 96. At the same time, positions of the residues close to T46 in sequence, such as P50, remain unchanged (Figure 3.1.3e). In the P50A HsBR mutant, on the contrary, the backbone around T46 is relatively unchanged, meanwhile the mutated A50 is moved away from the helix C. The HmBRI D94N structure reveals the most pronounced motion of the helix B relative to the helix C, which seems to be explained by the combination of the effects of the mutation D94N and substitution $P50_{Hs} \rightarrow T48_{Hm}$. It must be noted that, while the structures of ar-1 and ar-2 are closer to HsBR in this region, dr-3 also has $P50_{Hs} \rightarrow$ T substitution. Its structure is intermediate between HmBRI and HsBR P50A mutant[79,221].

3.1.3 Novel inter-helical hydrogen bonds in HmBRI

There are several local alterations present in the HmBRI structure relative to the structure of HsBR and other bacteriorhodopsin-like proton pumps. First, T124 of HmBRI helix D makes a hydrogen bond to the indole nitrogen of the helix F tryptophan W193. Although W193 that interacts with the retinal β-ionone ring is extremely conserved among archaeal rhodopsins, the threonine amino acid is observed only in HmBRI.

Then, there are two additional inter-helical hydrogen bonds in the hydrophobic regions of HmBRI, which are introduced by coupled substitutions of amino acids.

The first coupled substitution, $M56_{Hs} \rightarrow N54_{Hm}$ and $A84_{Hs} \rightarrow T82_{Hm}$ results in introduction of the hydrogen bond between the helices B and C (Figure 3.1.4a). Interestingly, the intra-helical hydrogen bond between the I52 backbone oxygen and T55 side-chain that existed in HsBR is replaced with the hydrogen bond between the homologous I50 backbone oxygen and N54 side-chain amine in HmBRI. None of these bonds is present in ar-1, ar-2 or dr-3.

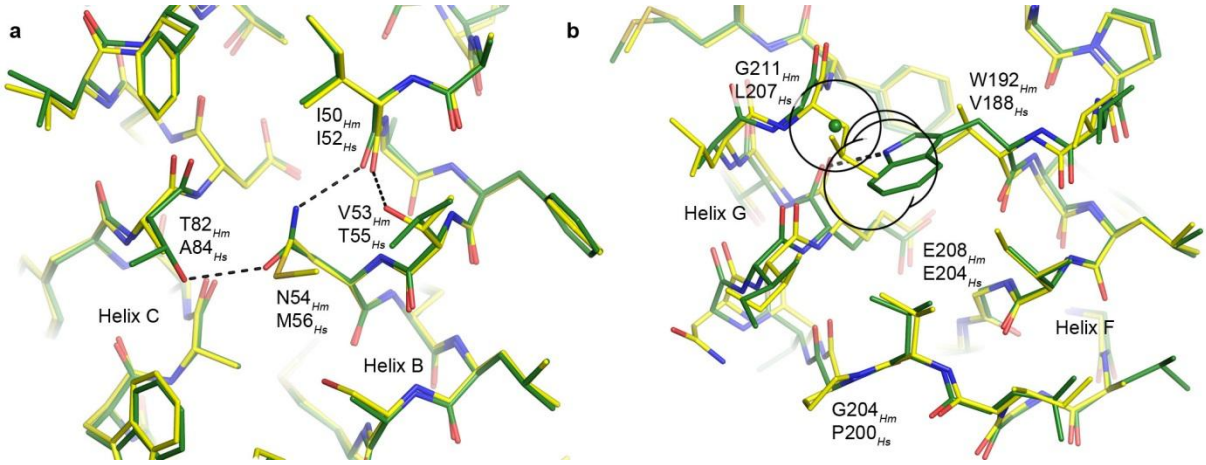

Figure 3.1.4. The inter-helical hydrogen bonds in *Hm*BRI introduced by coupled substitutions. In each case, the structure alignment was done locally to emphasize the local effects. a) The coupled substitution $M56_{Hs} \rightarrow N54_{Hm}$, $A84_{Hs} \rightarrow T82_{Hm}$ results in introduction of the hydrogen bond between the helices B and C. Interestingly, the intra-helical hydrogen bond between the I52 backbone oxygen and T55 is replaced with the hydrogen bond between the homologous I50 backbone oxygen and N54 side-chain amine. b) Coupled substitution $V188_{Hs} \rightarrow W192_{Hm}$, $L207_{Hs} \rightarrow G211_{Hm}$ results in introduction of the hydrogen bond between the helices F and G (E208 backbone oxygen and W192 indole nitrogen). Interestingly, the glycine is the only possible amino acid at the position 211, as the C_β atom of any other amino acid would create a steric conflict with W192 side-chain. The hypothetical position of the residue 211 C_β atom is shown by the green sphere and its van der Waals radius, as well as those of proximal W192 heavy atoms, are shown as black circles. Introduction of the bulky tryptophan at the position 192 might be another reason for the divergence of the helices F and G in *Hm*BRI, and, as a consequence, the bigger water-accessible volume of the proton release region.

The second coupled substitution $V188_{Hs} \rightarrow W192_{Hm}$ and $L207_{Hs} \rightarrow G211_{Hm}$ results in introduction of the hydrogen bond between the helices F and G (namely, between the G208 backbone oxygen and W192 indole nitrogen; Figure 3.1.4b). Similar tryptophan nitrogen – backbone oxygen hydrogen bonds have already been observed in retinylidene proteins, however, always in special cases. The W138-P186 bond in HsBR (W142-P190 in HmBRI) links the helices E and F, moved away from each other by the intercalating retinal. The similar bond in proteorhodopsins (residues W154-I187 in ESR) brings the cytoplasmic end of the helix F closer to the helix E, but causes a transformation of the α-helix to a complex conformation including a π-helix like element[1]. In HmBRI, glycine is the only possible amino acid at the position 211, as the C_β atom of any other amino acid would create a steric conflict with the W192 side-chain (Figure 3.1.4b). Thus, HmBRI structure presents an example of the inter-helical hydrogen bond involving tryptophan that does not result in disruption of normal

α-helical structure. Interestingly, the distance between the helices F and G is also elevated in ar-1 and ar-2, the reason for which is most probably the bigger-volume leucine amino acid at the place of HsBR's valine. In dr-3, despite an additional aminoacid in the F-G loop, the distance between the helices F and G is the same as in HsBR.

Although the stabilization of membrane proteins due to the hydrogen bonds in the transmembrane region might be moderate[222], presence of three additional bonds at the same time should undoubtedly stabilize HmBRI and might be a reason for its high E. coli expression level[91].

3.1.4 HmBRI trimers and its D-E loop

Similarly to other archaeal proton pumps[10,20,78,79] and despite being heterologously expressed in a bacterium Escherichia coli, HmBRI assembles as a trimer in crystal (Figure 3.1.5a). The RMSD between the backbone heavy atoms positions in the HmBRI and HsBR trimers is ~1.0 Å. This value reflexes the slightly different orientation of the HmBRI protomers in crystals, where the extracellular sides of the trimers align well, but the cytoplasmic sides are slightly displaced (Figure 3.1.5a). Such trimeric organization of HmBRI is similar to that of deltarhodopsin-3[79], as the RMSD of atomic positions is ~0.45 Å when either monomers or trimers are compared.

However, there is an important detail that separates HmBRI from the others bacteriorhodopsin-like proton pumps (Figure 3.1.5b). The HmBRI's D-E loop is 6 residues longer than the loops of other trimeric proton pumps and makes contacts with the adjacent protomer (Figure 3.1.5b and c). At the core of the loop are the buried E137 side-chain and three structural water molecules, participating in numerous hydrogen bonds (Figure 3.1.5c). The loop expands the trimerization interface and might stabilize the trimeric assembly.

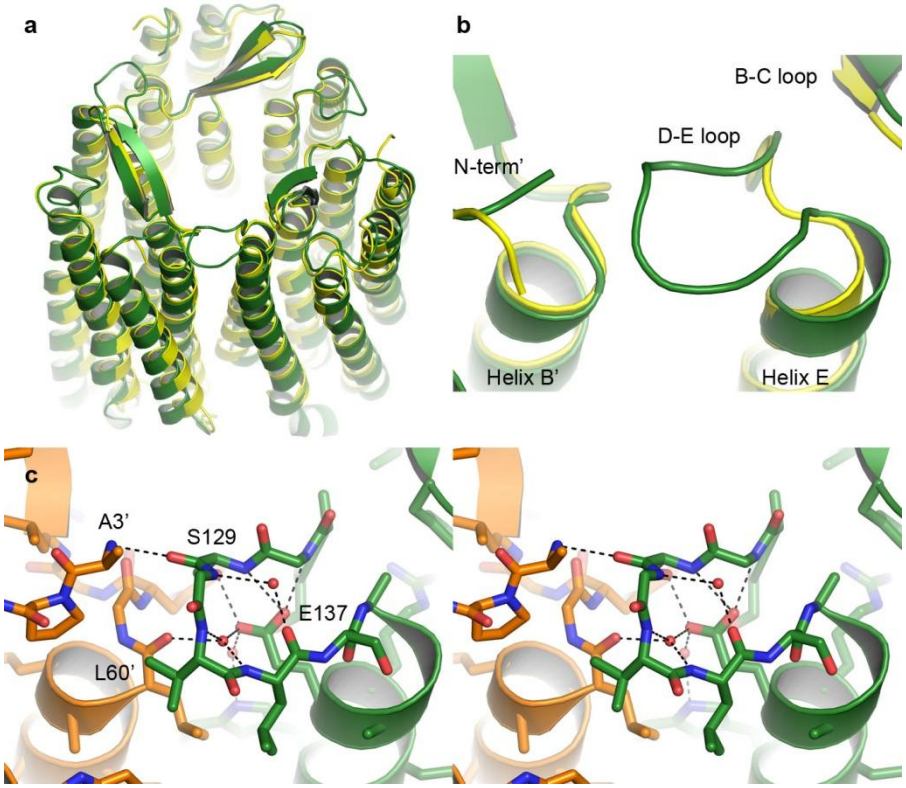

Figure 3.1.5. Structure of the *Hm*BRI trimer and its D-E loop. a) Comparison of the *Hm*BRI trimer structure (green) with that of *Hs*BR[16] (yellow). *Hm*BRI trimer aligns well in the extracellular region, but the protomers are slightly rotated at the cytoplasmic side. b) Magnification of the D-E loop. Unlike in other trimerizing retinylidene proteins, in *Hm*BRI the loop is extended and makes contact to the adjacent protomer. c) Wall-eyed stereogram of the *Hm*BRI D-E loop. The adjacent protomer is shown in orange. Three structural water molecules are observed that stabilize the loop structure.

3.1.5 Lipid molecules around HmBRI

Interestingly, all the retinylidene proteins, for which the trimeric assembly was observed in crystal, were purified from natural or natural-like sources[16,22,78–81,223,224]. The HmBRI crystals, presented here, were grown with the protein expressed in Escherichia coli, a bacterium, although the protein itself originates from the archaeon Haloarcula marismortui. The ordered lipid tails, belonging either to the E. coli lipids or the in meso crystallization matrix lipid monooleoyl, are observed at the same positions as the native lipids in other structures (Figure 3.1.6). First, there are three paired hydrophobic tails at the extracellular side of the intra-trimer cavity of HmBRI (Figure 3.1.6a), where binding of highly specific S-TGA-1 is observed in HsBR crystals[22,223,224]. Second, there are ordered hydrophobic tails on the outer surface of the trimer (Figure 3.1.6b and c), whose position is also very similar to that in the crystals of HsBR and other proteins. It is remarkable that the lipid binding mode is conserved

not only across the different proteins of the family but even across the different kinds of the hydrophobic tail moieties of the lipid molecules (branched isoprenoid chains in archaea and straight fatty acid chains in bacteria). Thus, stronger ability to assemble into a trimer and bind the host lipids might be another reason for the efficient expression of HmBRI in E. coli cells.

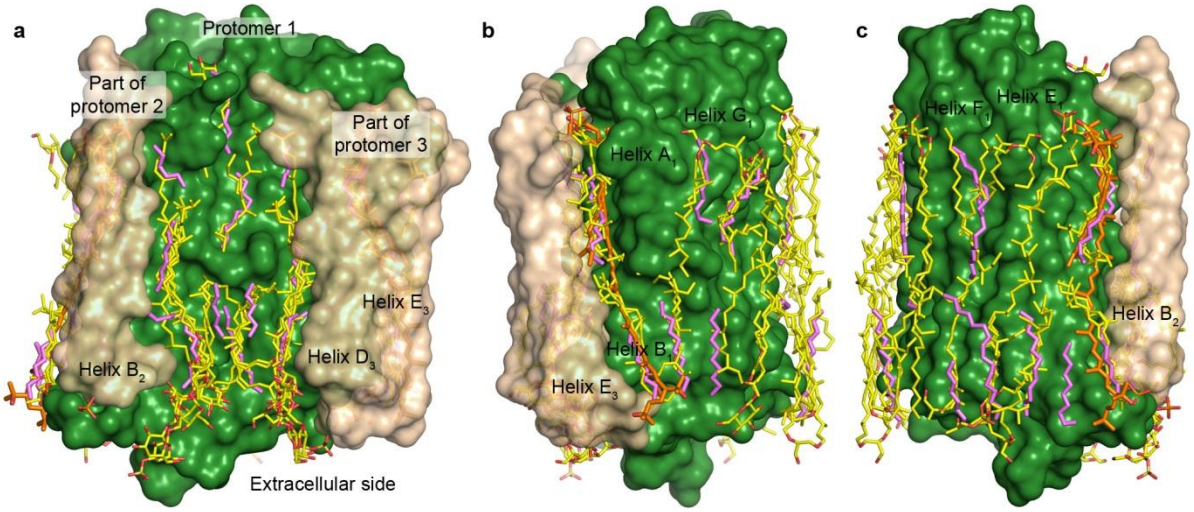

Figure 3.1.6. Comparison of the ordered lipidic tails observed in *Hm*BRI structure with those observed in the structures of trimeric *Hs*BR[16,22,223,224], ar-2[27] and dr-3[79]. The *Hm*BRI surface is shown in green, parts of the adjacent protomers are in beige, *Hm*BRI lipids are in magenta, the other lipids are in yellow and bacterioruberin molecules observed in aR-2 and dR-3 structures are in orange. a) Lipids inside the trimer. b) Lipids close to the helices A, B and G. c) Lipids close to the helices E and F. To obtain the positions of the lipidic moieties observed in the structures of the other proteins, their trimeric assemblies were aligned first to *Hm*BRI trimer.

3.1.6 Conclusions

In this section, the structure of the D94N mutant of Haloarcula marismortui bacteriorhodopsin I at the resolution of 2.5 Å was presented. The protein is trimeric in crystal. Its retinal-binding pocket and proton acceptor site are similar to those of other archaeal proton pumps, but its proton release region is extended and contains additional water molecules. There are three novel inter-helical hydrogen bonds and a unique extended loop between the helices D and E that makes contacts with the adjacent protomer and might stabilize the trimer. Many lipidic hydrophobic tail groups are discernible in the membrane region, and their position is similar to that of archaeal lipids in the crystals of other proton pumps, isolated from native or native-like sources. All these features might explain the high level of HmBRI E. coli expression and establish the protein as a novel model for the microbial rhodopsin proton pumping studies.

3.2 Crystal structure of *Exiguobacterium sibiricum* rhodopsin

In this section, a high resolution crystallographic structure of ESR is presented. It is shown that the structure is very unusual compared to the structures of BR and xanthorhodopsin (XR)[84], the closest ESR homologue and the only proton-pumping bacterial rhodopsin for which the crystallographic structure is known. The ESR structure possesses a proteorhodopsins-specific feature and may serve as a model for the protein family.

3.2.1 Crystallographic structure of ESR and its characteristic features

The *in meso*-grown[19,190] ESR crystals diffracted to 2.3 Å (Table 3.2.1). There are two ESR monomers with an almost identical structure in the crystallographic asymmetric unit. The most extended B-C loop could be built in one chain but not in the other (Figure 3.2.1).

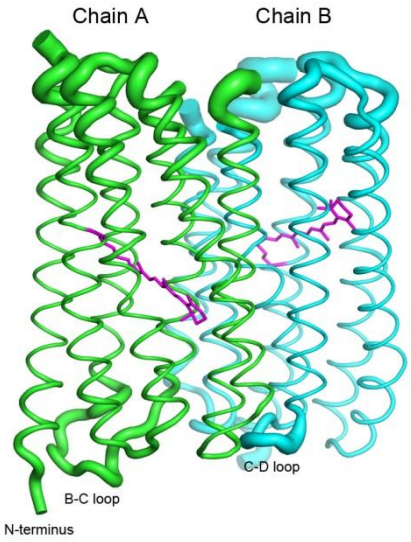

Figure 3.2.1. Distribution of the B-factors of the two ESR molecules in asymmetric unit. The molecules are oriented in opposite directions. The upper side is more disordered than the lower side. As a consequence, the B-C loop can be built for the chain A, and not for the chain B, and the C-D loop can be built for the chain B, and not for the chain A.

Comparison of the structure with the known structures of retinylidene proteins reveals that both the overall fold (Figure 3.2.2) and the Schiff base-proximal region (Figure 3.2.3) are very well conserved. The helix G π-bulge distortion observed in the other proteins is also present in ESR. However, ESR possesses another feature that is unique to the protein (Figure 3.2.2b). Namely, the α-helical structure in the middle part of the helix F is severely disrupted. First, there is a π-bulge-like distortion near the carbonyl oxygen of the residue 185 (hydrogen bonds O_{185}-N_{190} and O_{186}-N_{191}). Then, next to it, are 3_{10}-helix-like hydrogen bonds O_{189}-N_{192},

O_{190}-N_{193} and O_{191}-N_{194}. Due to this concomitant occurrence of the π- and 3_{10}-like structures the overall register of the helix is unchanged. The structural element is stabilized by hydrogen bonds between the backbone and neighboring side chains (O_{187}-Trp-154 and O_{190}-Asn-224). The residues Trp-154 and Asn-224 are present in all the proteorhodopsins but are absent in other proteins (Figure 3.2.2c). Based on this, we conclude that helix F should adopt the same structure in proteorhodopsins and this unusual structure discriminates them from the other retinylidene proteins, including XR. This notion is supported by the recently released crystallographic structure of blue light-absorbing proteorhodopsin from Med12 (PDB ID 4JQ6, Figure 3.2.4, the description was later published by Ran *et al.*[86]).

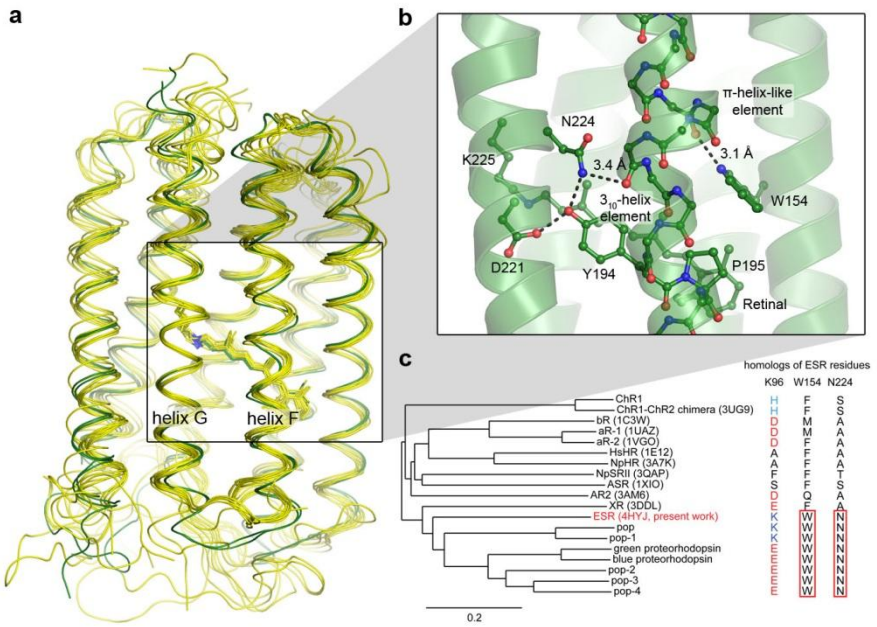

Figure 3.2.2. Comparison of ESR with the retinylidene proteins of known structure. a) Overlay of the known retinylidene protein structures. ESR is shown in green and the other proteins are in yellow. Most variation occurs in the helix A position. ESR differs markedly from the other proteins by disrupted α-helical structure in the middle part of the helix F. The helices A, E, F and G are signed. b) π-bulge and 3_{10}-helix-like structures in the ESR's helix F. The disruption of normal α-helical structure is stabilized by the hydrogen bonds between N224 and the carbonyl oxygen of G190, and between W154 and the carbonyl oxygen of I187. Meanwhile the proline 195 is conserved in other proteins, disruption of α-helical structure in the region of residues 187-191 is unique to ESR. c) Phylogenetic tree showing the relations between the retinylidene proteins of known structure, ESR and other proteorhodopsins. Residues W154 and N224 are unique to the proteorhodopsin family. Lysine at the proton donor position is also observed in the Clade B proteorhodopsins (genes *pop* and *pop*-1). The multiple sequence alignment and the phylogenetic tree were generated using ClustalX 2.1[225], and the tree was drawn using the TreeDyn web server[226]. The proteins used for comparison are channelrhodopsin (ChR), PDB ID 3UG9[85], bacteriorhodopsin (bR), PDB ID 1C3W[16], archaerhodopsins 1 and 2 (aR-1 and aR-2), PDB IDs

1UAZ and 1VGO[78], *Halobacterium salinarum* halorhodopsin (HsHR), PDB ID 1E12[80], *Natronomonas pharaonis* halorhodopsin (NpHR), PDB ID 3A7K[81], *Natronomonas pharaonis* sensory rhodopsin II (NpSRII), PDB ID 3QAP[3], *Anabaena* sensory rhodopsin (ASR), PDB ID 1XIO[73], *Acetabularia* rhodopsin 2 (AR2), PDB ID 3AM6[83], and xanthorhodopsin (XR), PDB ID 3DDL[84].

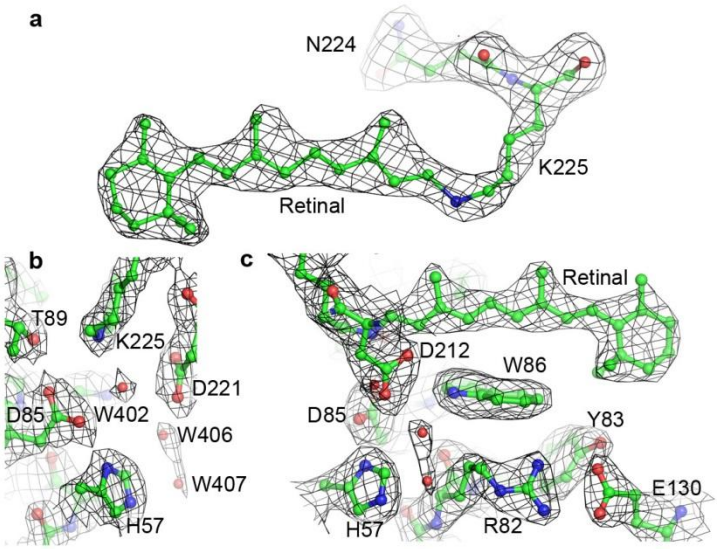

Figure 3.2.3. Structure of the ESR retinal and Schiff base-proximal region with electron density maps. a) The retinal in the all-*trans* conformation. b) View along the retinal. c) View perpendicular to the retinal. Aspartates 85 and 221 are traditionally close to the Schiff base, with a water molecule (W402) in between. There is a histidine residue (H57) close to D85. There are additional electron densities close to is that can be treated as partially ordered water molecules (W406 and W407). The arginine 82 side chain points away from the Schiff base and forms hydrogen bonds with E130. The electron densities are contoured at the level of 1.6σ.

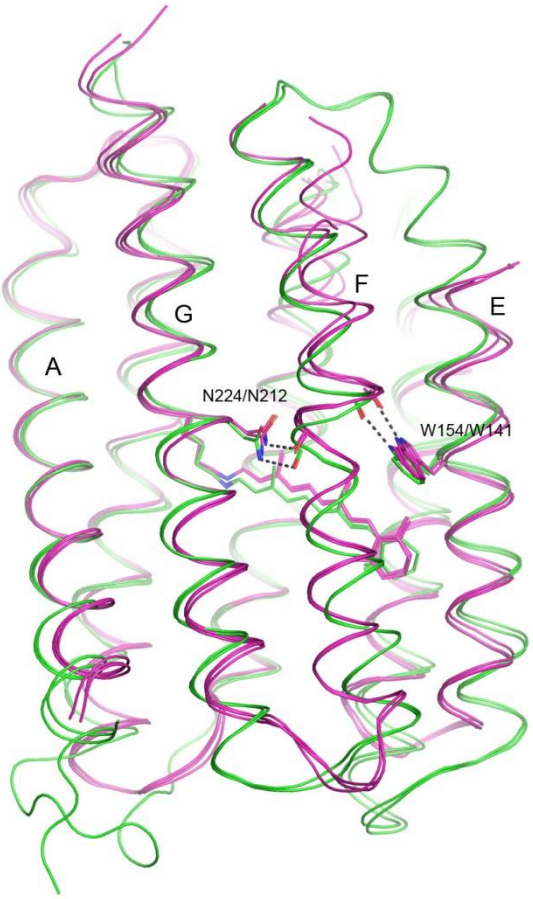

Figure 3.2.4. Comparison of the ESR and blue light-absorbing proteorhodopsin from Med12 (PDB ID 4JQ6) structures. Two ESR monomers and three BLAPR monomers, present in the asymmetric units, are shown. The ESR structure is in green and the BLAPR structure is in magenta. Positions of all the helices except F are similar. Despite the difference in position, the helices F of both proteins reveal the same deviations from the ideal α-helical geometry that are stabilized in both cases by the hydrogen bonds with neighboring aminoacids W154/W141 and N224/N212 in ESR and BLAPR correspondingly. These residues are shown in sticks representation.

To study the structure of the helix F in detail, we analyzed the distances between the backbone nitrogens and oxygens as well as the angles between the C=O and O···H bonds, separated by 3, 4 and 5 residues (Figure 3.2.5). As a guide for the hydrogen bond annotations, idealized 3_{10}-, α- and π-helix structures, based on the Grishaev and Bax data[227], were used (Figure 3.2.6). The analysis clearly shows the presence of 3_{10}- and π-helix hydrogen bonds in the ESR helix F.

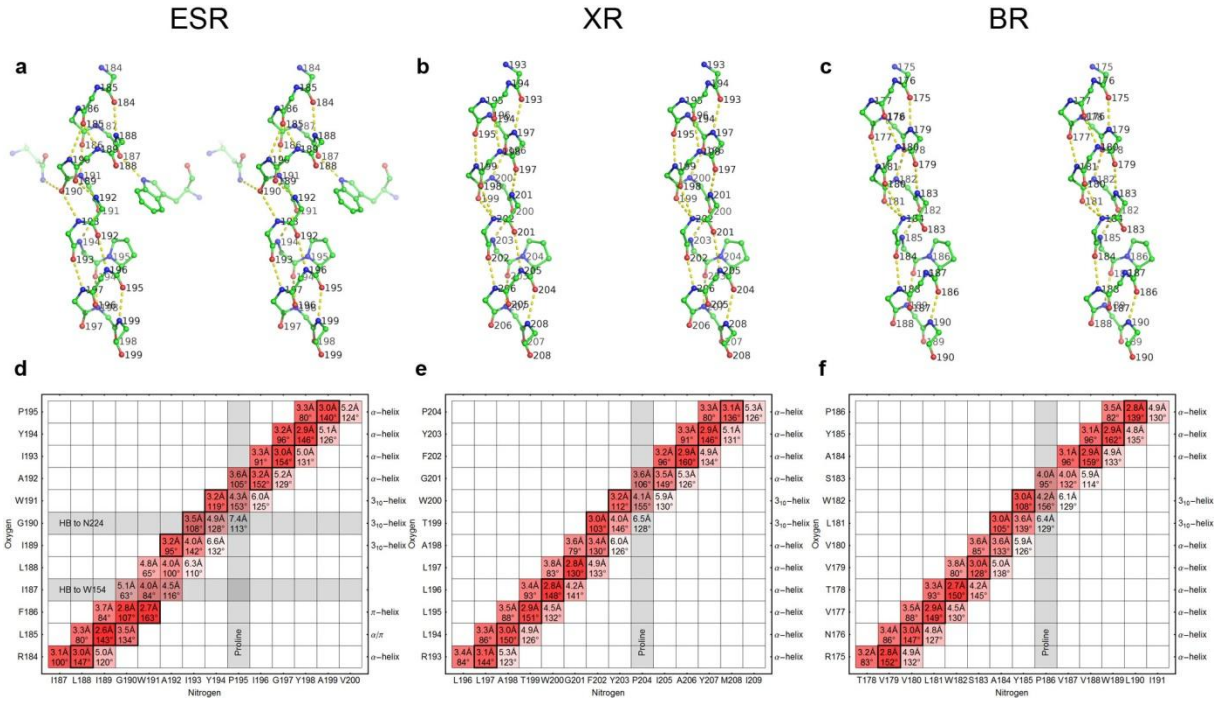

Figure 3.2.5. Analysis of the hydrogen bonds in the helix F in ESR, XR and BR. a,b,c) Cross-eyed stereographic images of the helix F and the residues stabilizing the secondary structure. d,e,f) Distances between the backbone O and N atoms and the angles between the C=O and O···H bonds, separated by 3, 4 and 5 residues. The values are averaged over the monomers present in the asymmetric unit. The cells corresponding to the putative hydrogen bonds are emphasized. The assigned secondary structure is listed at the right side of the table.

Role of the disruption of the helix F α-helical structure in proteorhodopsins is not clear at the moment. Both 3_{10}- and π-structures have been implicated to possess functional roles other than simply structural in other membrane and soluble proteins[228–230]. One possibility is that the novel structure facilitates the "unlatching" of the cytoplasmic site, similarly to what was proposed in bacteriorhodopsin photocycle[37]. As for the stabilizer residues, meanwhile the role of the Trp-154 position has not been highlighted previously, Asn-224, the residue preceding the retinal-binding lysine, is known to be important for the function of retinylidene proteins. For example, its mutation A215T (a single mutation in the retinal binding pocket; two other mutations needed for efficient signal transducer binding) is enough to convert the proton pump bacteriorhodopsin into a photosensor[137].

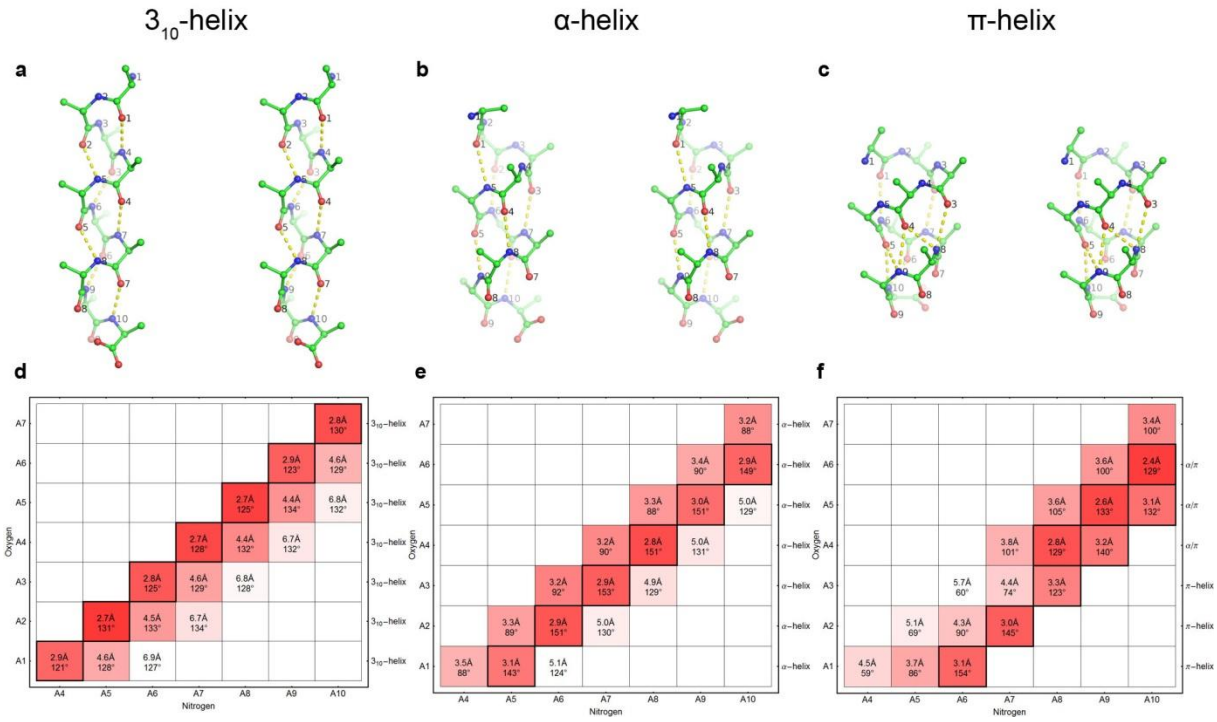

Figure 3.2.6. Analysis of the hydrogen bonds of the "ideal geometry" polyalanine 3_{10}-, α- and *π*-helices. a,b,c) Cross-eyed stereographic images of the helix F and the residues stabilizing the secondary structure. d,e,f) Distances between the backbone O and N atoms and the angles between the C=O and O···H bonds, separated by 3, 4 and 5 residues. The cells corresponding to the putative hydrogen bonds are emphasized. The assigned secondary structure is listed at the right side of the table. The atomic coordinates were taken from Justin Lorieau website (http://www.lorieau.com/methods-and-reference/biophysics-and-biochemistry/9-model-protein-helix-structures) and are based on the data obtained from the high-resolution crystallographic structures[227]. The average values for the i→i+3, i→i+4 and i→i+5 bonds and angles are 2.78 Å, 4.51 Å, and 6.78 Å and 126º, 131º and 131º.

Another difference between ESR and the proteins of known structure is the position of the loop connecting the helices B and C, that interacts excessively with the helix A in ESR and XR, and interacts with the helices D and E in BR (Figure 3.2.7).

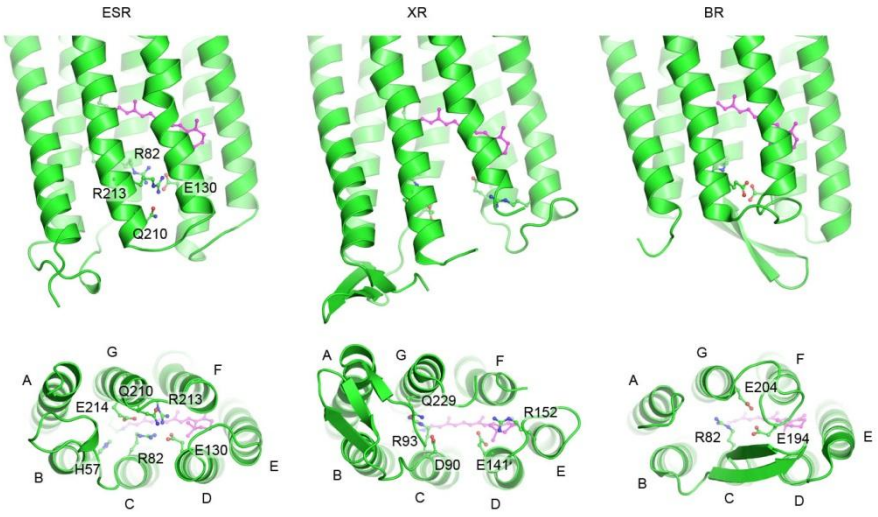

Figure 3.2.7. Comparison of the proton release sides of ESR, XR and BR. The most pronounced difference in the backbone conformation is the position of the B-C loop. In ESR, the loop is elongated, directed towards the helix A and does not adopt a secondary structure. In XR, the loop forms a beta hairpin and directed towards the helix A. In BR, the loop also forms a beta hairpin but directed towards the helices D and E. Another striking difference is in orientation of the R82 side chain. In ESR, the arginine 82 points away from the retinal Schiff base and is a part of a cluster of ionizable and polar side chains comprising residues R82, E130, R213, E214 and Q210. In XR, R93 (analogue of ESR's R82) is closer to the Schiff base and forms a hydrogen bond with Q229. In BR, R82's conformation is similar to that in XR.

Photosynthetic organisms often increase the efficiency of light harvesting by utilization of antennae. Archaeal proton pumps BR and archaerhodopsin function without any such additional mechanisms. However, it was shown recently that a light-harvesting antenna can function also in a retinal protein. Xanthorhodopsin, a retinylidene protein of the extreme halophilic bacterium *Salinibacter ruber* isolated from salt-crystallizer ponds, contains a single energy-donor carotenoid, salinixanthin, as an additional chromophore [84,231]. The salinixanthin binds to the outer surface of the helix F at the protein–lipid boundary. The keto-ring of salinixanthin interacts with the residues at the extracellular ends of the helices E and F and the β-ionone ring of the retinal. The binding pocket of the keto-ring is formed by Leu-148, Gly-156, Phe-157, Thr-160, Met-208, and Met-211 and the retinal β-ionone ring [84]. An intriguing question is whether other retinal proteins might also have an antenna. As it follows from our

data, the residues that are characteristic for xanthorhodopsin and *Gloeobacter* rhodopsin, another carotenoid-binding proton pump, are not conserved in ESR. Many of the ESR residues have larger side chains than the corresponding residues of XR. As a consequence, binding of the carotenoid or some other molecule in the retinal proximity is impossible (Figure 3.2.8).

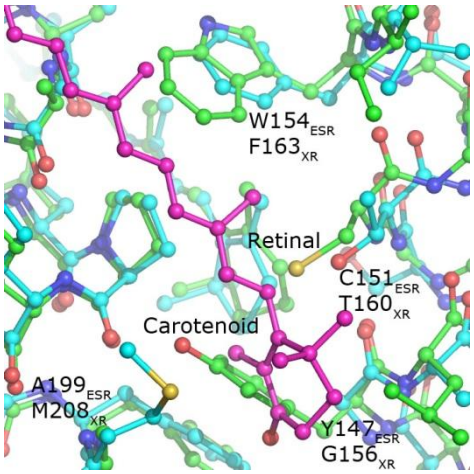

Figure 3.2.8. Comparison of the XR carotenoid-binding site with the same region of ESR. ESR is shown in green, XR in blue and the carotenoid salinixanthin in magenta. This binding mode of the carotenoid is impossible in ESR. Most importantly, there is a direct access to the retinal β-ionone ring in XR that is made possible by Gly-156. In ESR the retinal is covered by the bulky Tyr-147. Other notable residues with larger sizes are Trp-154 of ESR (compared to Phe-163 of XR) and Cys-151 of ESR (compared to Thr-160 of XR). Also, in the XR structure the carotenoid interacts with Met-208, which is mutated to alanine in ESR.

3.2.2 Structure of the retinal binding pocket, proton release and proton uptake groups.

The retinal binding pocket of ESR is presented in Figure 3.2.3 and compared to those of other proteins in Figure 3.2.9. The retinal is covalently linked to lysine 225 via a Schiff base, similarly to XR and BR. The electron densities around the retinal definitely show that the chromophore is in all-trans conformation (Figure 3.2.3a). The protonated Schiff base of the retinal points towards the extracellular part of the protein and, like in BR and XR, donates a hydrogen bond to a key water molecule W402. In its turn, W402 donates hydrogen bonds to 2 anionic residues, Asp-85 and Asp-221 (Asp-85 and Asp-212 in BR correspondingly) [16,224] (Figures 3.2.3 and 3.2.9). This arrangement stabilizes the positive charge of the Schiff base and is conserved in all the currently known proton pumps (and in most other retinylidene proteins). The carboxylate of Asp-85 in ESR is oriented in a similar way to Asp-85 of BR and

unlike its homolog Asp-96 in xanthorhodopsin, which is considerably rotated [84]. One more water molecule W406 is observed in the retinal binding pocket of ESR, where it occupies a similar position to W406 of BR (Figures 3.2.3 and 3.2.9). Meanwhile in BR W406 forms a hydrogen bond to Arg-82 [16,224], in ESR it is close to other water molecules directly facing the bulk solvent. In the structure of xanthorhodopsin only the water molecule W402 was found [84].

A remarkable difference of the bacterial proteins ESR and XR from archaeal BR is the presence of the histidine residue not far from the retinal (Figure 3.2.3). Like its homologue His-62 in XR, ESR's His-57 forms a tight hydrogen bond to the proton acceptor Asp-85. The observed bond lengths in the two monomers of 2.58 Å and 2.66 Å are somewhat larger than those of xanthorhodopsin, 2.55 Å and 2.42 Å. Although in both cases these distances may be indicative of an unusual bonding, for example low-barrier hydrogen bond[232–234], higher resolution structures are needed to test this hypothesis.

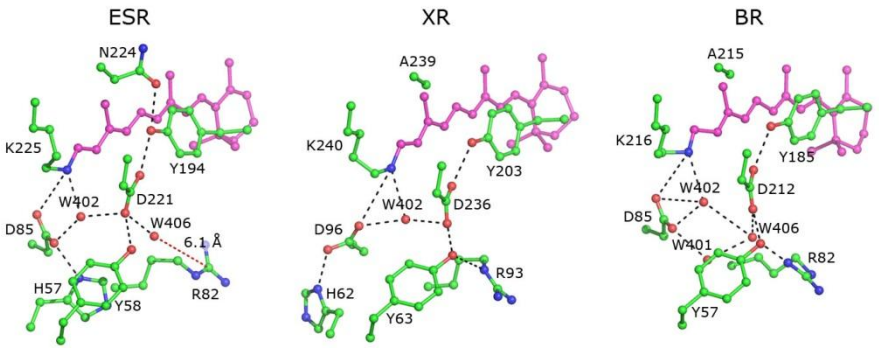

Figure 3.2.9. Comparison of the ESR, XR and BR retinal binding pockets in the ground state. Both ESR and XR differ from BR by presence of the histidine residue not far from the retinal. ESR's retinal pocket also possesses other differences. First and probably most important, its histidine is turned towards the arginine 82, which in its turn is removed from the retinal, similarly to the bacteriorhodopsin's M state. Second, in ESR the residue preceding the retinal-bound lysine is asparagine, as opposed to alanine in BR, XR and most of other proton-pumping bacterial rhodopsins.

Orientations of Asp-His pairs are completely different in ESR and XR (Figures 3.2.3 and 3.2.9). In ESR, His-57 is rotated towards Arg-82 and is immersed in a cavity of a size sufficient for a water molecule from the bulk to come in contact with it. The electron densities in the cavity are indicative of several partially ordered water molecules (Figure 3.2.3), however their positions differ in the two ESR molecules. The Asp-His pair is the only suitable proton acceptor group in the retinal vicinity. The Asp-His coupling explains the experimental observation that the pKa of His-Asp group is by several units higher than pKa of Asp-85 in

BR[95]. Lower affinity to the proton may explain the longer time needed for acceptor protonation[95]. At the same time, as His-57 is almost in direct contact with the bulk solvent, it itself or together with Asp-85 could serve as a proton release group. In this respect XR differs from ESR, as in the former the analogous histidine 62 points away from Asp-236 (analogue of Asp-212 of BR), and does not participate in the hydrogen bond connection of the Schiff base to the solvent. Arginine 93 of XR separates the Schiff base from the bulk solvent similarly to BR. There is no histidine residue in the vicinity of the retinal in BR, the arginine 82 separates the Schiff base from the bulk solvent, and the proton release group is believed to consist of glutamates 194 and 204 and the adjacent water molecules[6,16,224].

ESR's Arg-82 points away from the retinal (Figure 3.2.3), similarly to the bacteriorhodopsin's M state [6], and occupies the outermost position compared to the other proteins (Figure 3.2.10). It is believed that Arg-82 plays a key role in proton release in BR[6,16,224]. In the BR photocycle, protonation of Asp-85 from the Schiff base upon L to M state transition results in a flip of Arg-82 from the retinal to extracellular direction (outward from the retinal). However, already in the ground state of ESR, Arg-82 points to the extracellular part of the protein. There, it forms a hydrogen bond to Glu-130 (Figure 3.2.3b) and is a part of a larger cluster of ionizable and polar side chains (Figure 3.2.7). This conformation is expected to remain unperturbed during the whole photocycle and therefore, role of Arg-82 in the ESR's proton release mechanism can be not as important as it is in BR. The data described here provide an explanation of a recent observation that Asp-85 and Arg-82 are not coupled in ESR (as the former does not influence pKa of the latter, unlike in BR) [95].

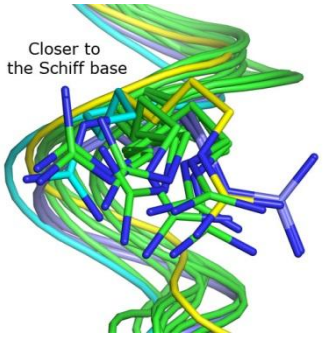

Figure 3.2.10. Conformations of the Arg-82 homologs' side chains in the retinylidene proteins of known structure. The figure is a magnification of a part of the Figure 3.2.2. ESR is shown in dark blue, XR in light blue, ChR in yellow and the other proteins in green. In ESR, Arg-82 occupies the outermost position among the presented proteins.

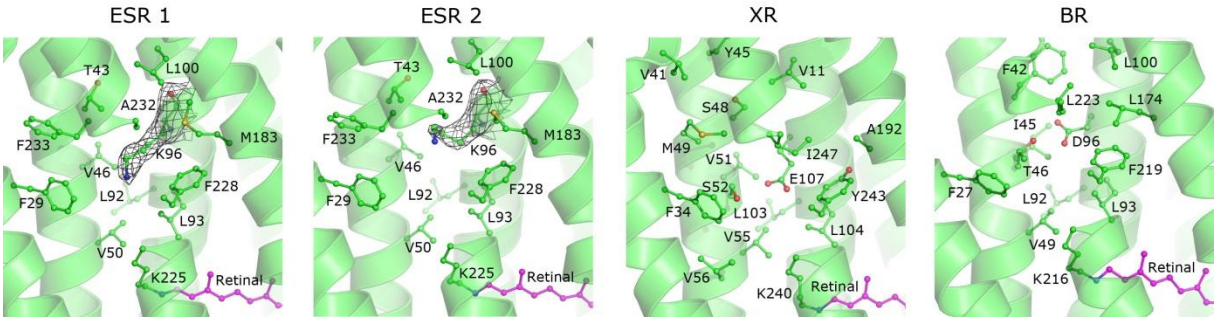

Figure 3.2.11. Comparison of the ESR, XR and BR proton uptake regions. The conformations differ in the two observed ESR molecules. Contrary to BR, XR and other light-driven proton pumps of known structure, ESR has a lysine (Lys-96) residue at the proton uptake, whose electron densities are shown at the level of 1.4σ. The cavity around Lys-96 is mostly surrounded by hydrophobic aminoacids and may accommodate at least three water molecules. The only polar residue, Thr-43, separates the cavity from the bulk solvent. On the contrary, the proton uptake residues Glu-107 of XR and Asp-96 of BR are far from the bulk solvent, being separated from it by Ser-48 and Tyr-45 in XR and Phe-42 in BR. Lys-96 side chain is ordered in one ESR molecule and partially disordered in the other. Additional positive electron densities around Lys-96 (not shown) are observed in the difference maps that are probably related to mobile water molecules.

The ESR's proton uptake region is also unique among the proton pumps of known structure from both archaea and bacteria. There is the cationic aminoacid lysine 96 at its core, contrary to the anionic amino acids Asp-96 and Glu-107 in BR and XR respectively (Figures 3.2.2c and 3.2.11). The same aminoacid is found in the proton uptake region of recently identified proteorhodopsins from marine uncultured MG-II Euryarchaeota[87] (Figure 3.2.2c). The Lys-96 side chain is to some extent mobile as its positions differ in the two observed ESR molecules, and the electron densities are weak in one of them (Figure 3.2.11). This side chain flexibility might be required for the structural rearrangements needed for the reprotonation. However, it could also be a result of the overall lesser quality of the proton uptake part of one of the two molecules in crystallographic asymmetric unit (Figure 3.2.1). The cavity around Lys-96 is mostly surrounded by hydrophobic aminoacids and may accommodate at least three water molecules. Residual electron densities are seen in this region although they are not sufficient for a reliable placement of these water molecules. Next, the cavity is very close to the bulk solvent as it is separated from it only by a polar side chain of Thr-43. This is a remarkable difference from BR and XR, where there is a large gap between the proton uptake residue and the bulk solvent (Figures 3.2.11-12). Proximity of the bulk solvent can result in an easy direct access of the protons from the cytoplasm to Lys-96 through a water 'funnel' and might explain why proton uptake precedes proton release in ESR. The small hydrophobic gate

between K96 and the Schiff base, created by the leucine 93 side chain, is similar to that of XR and BR. We suggest that upon isomerization of the retinal Leu-93 may change its conformation and allow a faster Schiff base reprotonation.

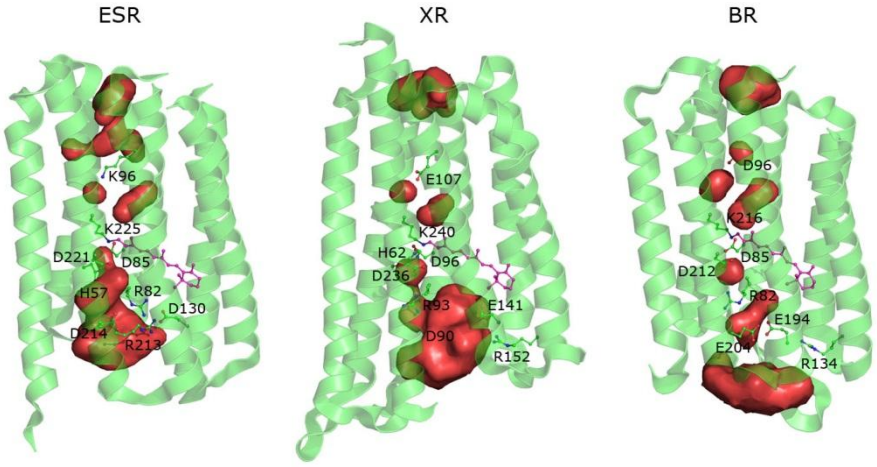

Figure 3.2.12. The cavities on the putative proton path in ESR, XR and BR that may be occupied by water molecules. The upmost and the lowest cavities on the figure face the bulk solvent. The important ionizable and charged residues are shown explicitly. In the proton uptake region, ESR's Lys-96 is separated only by the Thr-43 side chain from the solvent, meanwhile there is a large gap in XR and BR between the bulk and Glu-107 and Asp-96 correspondingly. There is enough space for at least 3 water molecules around Lys-96 in ESR. In ESR, the histidine residue of a putative proton release group Asp-221/His-57 is immersed in a cavity of a size sufficient for a water molecule from the bulk to come in contact with it. This continuous cavity contains the ordered water molecules 402 and 406, and transitions into the bulk. In XR the release group is shielded by Arg-93. In BR, the proton is released from the completely different group, a pair of glutamates (Glu-194 and Glu-204) that are separated by Ser-193 side chain from the bulk. The cavities are determined as a composition of the crystallographically recognizable water molecules and the space determined using the Hollow software[235] with a 1.4 Å probe and 0.2 Å grid spacing.

Why do ESR and some other proteorhodopsins possess a cationic aminoacid at the proton uptake? Exiguobacterium sibiricum inhabits permafrost soil which is characterized by a wide range of pH (5.6-7.8). Therefore, the function of ESR must be adapted to this variation. Indeed, light-induced proton transport by ESR-containing E. coli cells and proteoliposomes is observed in a broad pH range of 4.5-8.5[95]. It was shown that the pair Asp-85/His-57 allows the protein to stabilize Asp85 in the unprotonated state in a wide range of pH which is necessary to keep the proton pump functional[95]. However, in proteorhodopsin, where the proton acceptor pair is the same, direction of the proton flow may be reversed at acidic pH[236]. Therefore, a priori Asp-85/His-57 cannot be considered as sufficient and universal stabilizer

of proton pumping in a wide range of pH. One could speculate that Lys-96 with its reduced pKa is a key element of stabilization of proton pumping. However, it is not clear how to explain a recent discovery that proteorhodopsins corresponding to the genes revealed in the metagenomic sequences of uncultured marine group II Euryarchaeota contain a lysine residue in the proton donor position, similarly to ESR[87]. This question about the proton pumping by ESR requires further investigation.

3.2.3 Proposed model of the photocycle.

The structural and spectroscopic studies of the intermediate states are necessary for a complete description of the mechanism of vectorial proton translocation in ESR and explanation of the role of Lys-96 and the Asp-85/His-57 couple. Nevertheless, the outline of the ESR photocycle has been obtained recently by Balashov *et al.* (Figure 3.2.13) and thus, using the structure, we can propose the major steps of the proton pumping by the protein. First, upon isomerization of the retinal the Schiff base flips from the extracellular to the cytoplasmic part of the retinal pocket. There, position of the charged group in the mostly hydrophobic environment is energetically unfavorable, and the Schiff base must be deprotonated as a result. The only primary proton acceptor candidate is the residue Asp-85[95], which is coupled strongly to His-57 (Figures 3.2.3). Next, the Schiff base must be reprotonated. Indirect experimental evidence shows that the reprotonation proceeds via Lys-96 as it is the only ionizable residue in that region and the M state decay is significantly slowed down in the K96A mutant[95]. It can be expected that Lys-96 is deprotonated in the ground state, as it is immersed in a mostly hydrophobic cavity, and it was shown that in such cases the lysine side chain pKa may be as low as 5.3[237]. It means that Lys-96 must be reprotonated following the Schiff base deprotonation, and only then it can pass the proton to the Schiff base. The reprotonation step is probably facilitated by presence of the continuous connection to the bulk solvent (Figure 3.2.12). Finally, following the experimental observations[95], release of the proton to the bulk solvent occurs after the Schiff base reprotonation.

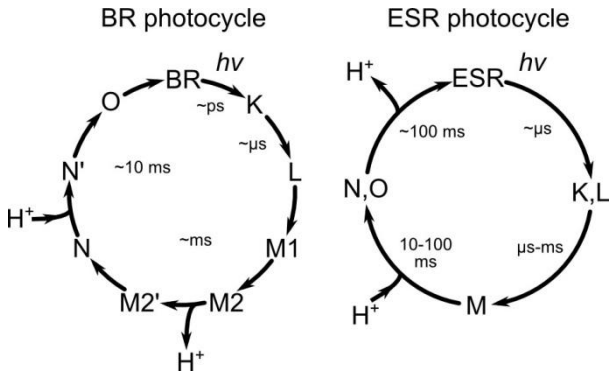

Figure 3.2.13. The models of the BR and ESR photocycles[6,95]**.** Some of the transitions are reversible. The timescales are very approximate as they depend strongly on the conditions such as pH, temperature and others.

3.3 Crystal structure of the light-driven proton pump KR2

< Not shown as the results have not been published yet >

3.4 Determination of the *Natronomonas pharaonis* sensory rhodopsin II active state structure

Sensory rhodopsin is the primary light receptor of the archaeal phototaxis system, and knowledge of its conformation in the resting and signaling states is the first prerequisite for understanding the signaling mechanism. The sensory rhodopsin, whose structure is characterized best, is the *Natronomonas pharaonis* sensory rhodopsin II (*Np*SRII). Crystallographic structures of resting-state *Np*SRII in absence of its cognate transducer *Np*HtrII have been determined in 2001 by Royant *et al.*[82] (up to a resolution of 2.1 Å) and Luecke *et al.*[131](up to a resolution of 2.4 Å). In 2002, structures of sole *Np*SRII in the early intermediate state[142] and resting-state *Np*SRII in complex with *Np*HtrII[125] have been determined. Although the structures of *Np*SRII in complex with *Np*HtrII in the early K and signaling M states have been determined in 2006 by Moukhametzianov *et al.*[129], structure of the sole NpSRII in the active (signaling) state was elusive. Here, the problems associated with solving the NpSRII active state structure are discussed, as well as the structure itself. The section 3.4.1 describes the high-quality crystallographic structure of the ground state NpSRII. The section 3.4.2 describes the details of calculations of the intermediate state occupancies in the illuminated crystal. Finally, the sections 3.4.3 and 3.4.4 describe the active state structure and discuss its implications for signaling and proton transfer.

3.4.1 High quality structure of *Np*SRII in the ground state

Calculations of the partially-occupied intermediate state structures benefit from the ground state models of the best quality. One of the properties determining the quality of a model is its resolution. Another property is whether the model includes all the atoms that are discernible in the electron density maps. Although the membrane protein crystals often feature ordered lipid molecules, no lipids or their parts are present in the previously published ground state models (PDB ID 1H68[82] and 1JGJ[131]). In this section, the model of the NpSRII in the ground state refined to a resolution of 1.9 Å that includes 31 hydrocarbon chains is presented and discussed.

The model contains residues from 1 to 219 with poorly ordered A-B and C-D loop regions. Similarity between the new and published X-ray structures is very high. RMSD of coordinates equals 0.31 Å for 1H68[82], 0.43 Å for 1JGJ[131] and 0.27 Å for the structure reported in the paper of Reshetnyak et al.[132] (the position of A-B loop was omitted from calculation; RMSD

was calculated for the backbone atoms). Differences are observed in the position of the C-D loop (residues 93 to 97) which in our structure is displaced by about 1.5 Å as compared to the published structures. D193 position is closer to R72, than in 1JGJ (for 0.8 Å) as well as in 1H68 (for 1.8 Å). We do not observe water molecules 2017 and 2026 of 1H68, but identify three additional water molecules: near the Y174 carboxyl group, near Y73 hydroxyl group and near D193. Also some small amount (occupancy < 35%) of water molecules at position 2013 of 1H68 may be replaced by chloride ions, as it is in 1H68, since it would better fit the electron density. The cytoplasmic loop F-G, comprising residues 175-195, is ~0.5 Å closer to the extracellular channel in our model. Finally, in 1JGJ a number of residues have a different rotameric conformation, than in our or 1H68 structures. These differences may be ascribed either to different crystallization conditions and/or to a better resolution of our structure and seem not to be important for protein function.

As for the recently published NMR structure 2KSY[76], the differences are much larger. Substantially, they include displacements of up to 1 Å of residues known to be important for signal transduction (Y199, Y174 and T204), and changes in their hydrogen bonds. These differences may probably be attributed to the lack of proper restraints in NMR structure determination, the absence of water molecules and retinal proton in NMR model, or to the fact that the protein was in detergent micelles and not in the lipid bilayer.

We also observe a strong similarity between the ground-state structures of NpSRII alone and in complex with NpHtrII. RMSD of coordinates are 0.53 Å for 1H2S[125] and 0.45 Å for the structure in $I2_12_12_1$ space group of Ishchenko *et al.*, to be published.. There are only slight differences in poorly ordered loop regions which include also the position of the C-D loop. An important observation concerns D193 which occupies the same position in the structure of the complex and the present structure, contrary to that in the structures 1H68 and 1JGJ. This data emphasizes the importance of identical crystallization conditions.

Additionally, it was possible to assign 31 hydrocarbon chains (Figure 3.4.1). Two of them may unambiguously be identified as archaeal phytanyl chains by their electron densities. Both eubacterial and archaeal lipids may be present in the crystal as NpSRII was expressed in *E. coli* and subsequently reconstituted into *H. salinarum* polar lipids. Additionally one molecule of octyl-β-D-glucopyranoside could be identified. Inclusion of the lipid molecules into the model turned out to be necessary for obtaining the active state structure of NpSRII, because it resulted in more consistent electron density maps as well as better R-factors.

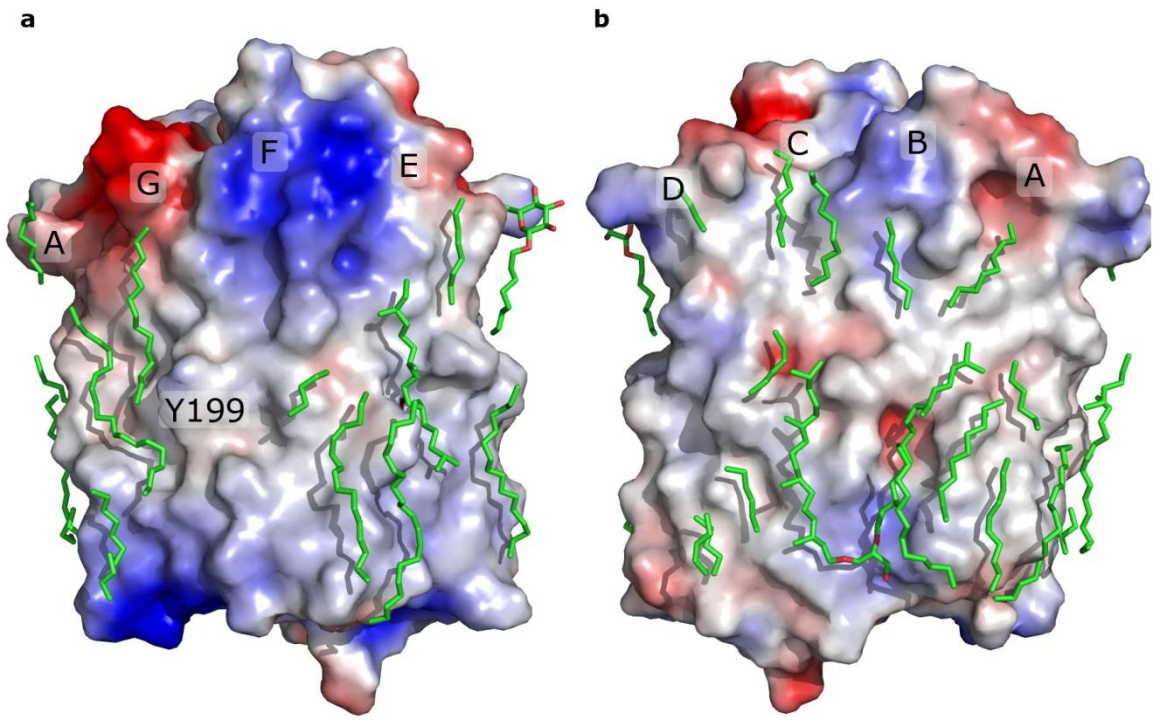

Figure 3.4.1. Lipid tails in the ground state model of *Np*SRII. Lipid tails are shown in green sticks representations. Surface potential was calculated using ABPS[204]. Blue color corresponds to potential value of +3 kT/e and higher, and red color to -3 kT/e and lower. Lipid bilayer was not included in the calculation. *Np*SRII helices are marked by corresponding letters. Tyrosine Y199 and the region of positive potential near the helix F correspond to the binding site of *Np*HtrII transducer in *Np*SRII-*Np*HtrII complex.

3.4.2 Procedure of the active state trapping

The crystallized NpSRII was converted into the active state by illumination with a 488 nm laser. In various experiments, the illumination time was constant and the laser power was varied. The amount of the protein converted to the active state increased with increasing the laser power.

At the same time, it was observed that upon trapping of the active state in crystals the diffraction quality worsens (in anisotropic fashion). The largest loss in resolution of 1 Å occurs along the b-axis, which is perpendicular to the membrane plane. At the same time we observed large changes in the unit cell dimensions. The most pronounced change was again in b-axis direction, with the cell dimension increasing from 128 Å up to 131 Å. These effects were stronger with the higher intensity of the light used for trapping of the active state.

Thus, the effects of stronger illumination were both positive (more protein was converted to the active state) and negative (the loss of diffraction resolution was stronger), and the optimal intermediate state trapping conditions had to be established.

3.4.3 Spectroscopic characterization of the intermediate state occupancy

The analysis of the absorption spectra of NpSRII in crystal is based on the model of Chizhov *et al.*[241] Meanwhile the ground state and the O-state are fitted satisfactorily, the M-state absorption of NpSRII in crystal (present data) differs from that of NpSRII reconstituted in purple membrane lipids[241] (Figure 3.4.2a). Change in the peak width does not fully accounts for the observed differences (Figure 3.4.2b), as there is an additional shoulder on the left peak. Therefore, the M-state absorption was modeled by three unskewed gaussians (Figure 3.4.2c). The resulting fit is good given the data accuracy.

Use of the modified parameters complicates estimation of the occupancy of the intermediates. In our analysis we employed two assumptions. First, we assumed that the extinction coefficients of NpSRII in crystal are the same as in purple membrane lipids[241]. Second, for the M-state, where the different parameter set was used, peak extinctions were normalized so that the total area of the peaks, corresponding to the M-state, would be the same as in Chizhov *et al*. [241] Despite the differences in the form of the M-state peak used for fitting, all the estimates agree well with each other. The results are presented in the Table 3.4.1.

Illumination of the crystals (excitation wavelength 488 nm) with the laser power 0.33 W for 2 s (the cryostream is blocked for that period) results in accumulation of approximately 35% M-state and 25% O-state. Further increase in the laser power slightly raises the amounts of the accumulated active states, to the maximum of ~45% M-state and 30% O-state. The laser power of 0.33 W appears to be the optimal as the resolution loss along the b-axis does not exceed 1 Å, meanwhile the trapped amount of the active state allows for the structure determination. Using such an approach, a number of diffraction datasets were obtained with different NpSRII crystals. Resolution limits for the best crystal were 2.5/2.9/2.6 Å for *a*, *b* and *c*-axes correspondingly. For that crystal, amount of the active state determined crystallographically was found to be approximately 45% (Section 3.4.4).

Table 3.4.1. Estimates of the active state occupancy in *Np*SRII crystals with different assumptions.

Laser power, mW	Occupancy (G, M, O and A stand for ground, M, O and total active state correspondingly), %											
	Parameters as in Chizhov *et al.*				Parameters as in Chizhov *et al.* with width of the M state peaks adjusted to match the spectra				Parameters as in Chizhov *et al.* with absorbance of the M state being modelled by three unskewed gaussians.			
	G	M	O	A	G	M	O	A	G	M	O	A
0	88	0	12	12	88	0	12	12	88	0	12	12
0.33	40	34	26	60	42	33	25	58	43	31	27	58
0.7	34	40	26	66	37	38	25	63	38	36	26	62
1.22	27	50	23	73	28	50	22	72	29	46	25	71
2	25	52	24	75	27	50	23	73	27	46	26	73
3	29	44	27	71	31	43	26	69	31	45	24	69

3.4.4 Determination of the intermediate occupancy by crystallographic methods

The intermediate occupancy was estimated as follows. Superposition of a fixed and a refinable ground-state structure with occupancies varying in the range 0.25–0.9 were refined by simulating annealing (using 3,000 K and a slow cooling protocol in CNS[203]). For the refinable part of the model, a less ordered loop and terminal residues (1-2, 28-32 and 217-219) were removed, along with surface water molecules and small lipid fragments. Internal water molecules, which are expected to be present but are not clearly seen at the present resolution, were harmonically restrained to their positions in the ground-state structure.

The R factor is clearly improved by raising the intermediate occupancy up to 50%. No improvement is seen with the further increase in occupancy (Figure 3.4.3a). The R_{free} in its turn has a clear minimum at the intermediate occupancy values from 40% to 55% (Figure 3.4.3b). Therefore, it can be concluded that the real intermediate state occupancy lies in the range from 40% to 55%.

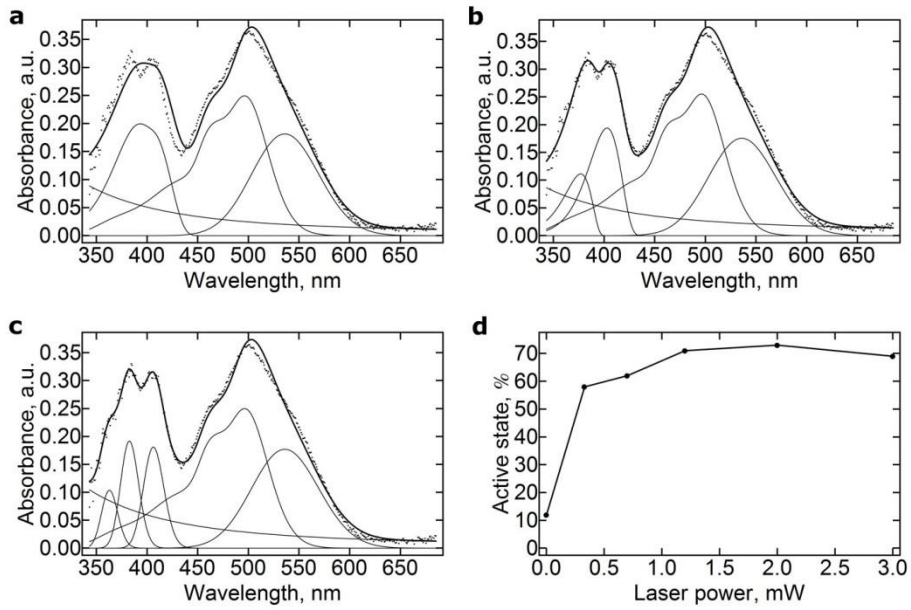

Figure 3.4.2. Examples of fitting of the absorbance by NpSRII crystals with the accumulated active (M and O) states by different functions for the M-state. a) Parameters for the all states are as in Chizhov *et al*. b) Parameters are as in Chizhov *et al*. with the width of the M-state peaks adjusted to match the spectrum. c) Parameters are as in Chizhov *et al*. with the absorbance of the M-state being modeled by three unskewed gaussians. In all the panels the experimental data are shown by the dots, the global fit by the bold line, and its components (background, the ground state, the M-state and the O-state) by the thin lines. The spectrum corresponds to 0.33 W laser power. d) Dependence of the amount of the accumulated active state on power of the laser used for illuminating the crystals (data from the M-state fit by three peaks are shown).

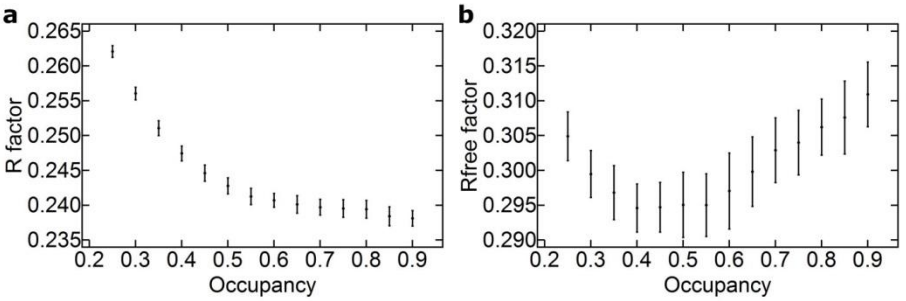

Figure 3.4.3. Determination of intermediate state occupancy by crystallographic methods. a) Dependence of the crystallographic R factor on the supposed amount of accumulated active state for the diffraction data used for structure determination (laser power 0.33 mW). b) Dependence of the crystallographic Rfree factor on the supposed amount of accumulated active state for the same data. a) and b) result in estimation of the accumulated active state in range from 40% to 55%, which is in line with spectroscopic data, given the errors of the methods. Values and error bars correspond to means and standard deviations, obtained from 100 simulated annealing runs per one occupancy value.

3.4.5 Heterogeneity of the *Np*SRII active state

The procedure employed in this work did not allow trapping pure M state. The next intermediate in the photocycle, the O-intermediate could have been accumulated under these conditions as well. Indeed, the spectroscopic data show that the amount of the O-state reaches up to two thirds of that of the M-state (Table 3.4.1 and Figure 3.4.2). Independent refinement of the M and O structures is impossible due to low data-to-parameter ratio, which is slightly over 1 when only one intermediate structure is refined. Nevertheless independent data from the literature indicate that the structure of the M-state closely resembles that of the O-state. First, the photophobic signal is relayed to the transducer in both the M and O states[242,243]. EPR experiments also show that reordering of helix F and TM2 occurs, believed to represent the signal, occurs concomitantly with O-decay[126]. FTIR studies[244–246] have shown that differences between M- and O-states are much smaller than between M and the ground state. Transient-grating analysis also shows little volume changes upon M to O transition compared to transition from ground to M intermediate[247]. The main difference concerns the back isomerization of the retinal to the all-trans conformation[244–246]. Global similarity of the late intermediate structures was also proposed for BR[248]. From these considerations, it can be assumed that structures of the M- and O- intermediates are quite similar at present resolution. It is therefore reasonable, to treat M- and O-state as one entity. This assumption is also justified by the fact that the occupancy of the active state obtained by crystallographic means is in good agreement with the results of the spectroscopy, as M- and O- state combined which amounts to about 50%. Finally, it should be noted, that in the case of the *Np*SRII/*Np*HtrII complex the trapped active state not only contained the M-intermediate but also some amounts of the O-intermediate. Nonetheless no large unaccounted difference electron densities were seen, when the data were treated with the only one model of the active state of protein. The resolution of the structure was 2.0 Å and thus the large structural deviation in the O state, would it be present, would have been detected[129].

3.4.6 Conformational changes in the NpSRII active state

As it could be predicted from the loss of diffraction quality during the trapping, we observe large conformational changes in the NpSRII active state (Figure 3.4.4). The structural changes reported below are reproducible for 3 different illuminated crystals displaying the highest resolution, for which refinement of the active state was possible. The changes are in accordance with difference Fourier electron density maps between the active and the ground

states. Thereafter we discuss the changes as they are observed in the structure of highest resolution. Alignment of structures for the comparison was done by the non-hydrogen atoms for the whole protein.

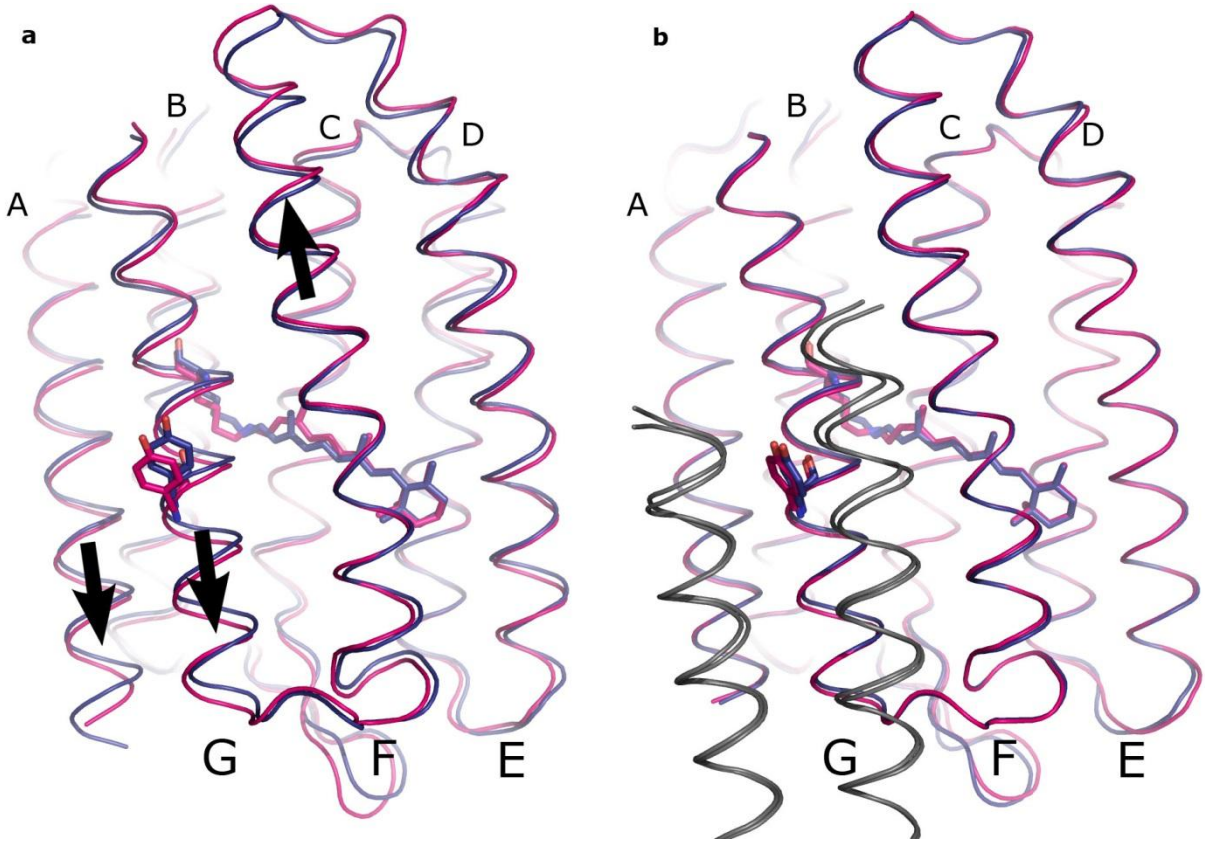

Figure 3.4.4. Overall structural rearrangements in *Np*SRII upon transition to the active state. a) Structural rearrangements in sole *Np*SRII. b) Structural rearrangements in *Np*SRII in complex with *Np*HtrII[129] (PDB ID 2F95). The ground state is shown in blue and the active state is in magenta. *Np*HtrII in b) is shown in grey for both states. Lysine 205, retinal and tyrosine 199 are shown in the balls-and-sticks representation. *Np*SRII helices are marked by corresponding letters (the larger letters correspond to the front helices, and the smaller letters correspond to the back helices). Arrows show movements of the corresponding helices. Alignment is done by the *Np*SRII backbone atoms.

Large conformational changes are observed at the transducer-binding interface comprising helices F and G. Helix F moves towards the cytoplasm by about 0.3 Å, whereas helix G moves in opposite directions towards extracellular side by 0.7 Å. Helix A moves together with helix G, though with a twice as lower amplitude. The extracellular F-G loop moves out of the membrane plane. These movements are similar to those observed for the NpSRII-NpHtrII complex, but display larger amplitudes (Figure 3.4.4).

Movement of the helix G, with Y199, which forms a crystal contact in case of NpSRII, seems to be the main cause for the increase of crystal b-dimension, as the increase, 3 Å, corresponds well to 4 times by 0.7 Å, which follows from packing geometry.

On the cytoplasmic surface, the E-F loop changes its conformation and becomes more compressed. The movement of the C-D loop is questionable, since this loop is poorly ordered, and its position differs as exemplified above in the known ground-state structures.

The alterations on the surface are paralleled by changes in the interior of light activated NpSRII. The difference density maps clearly show that 13-cis retinal is present in the crystals which amounts to about 35% of the M-intermediate as determined spectroscopically (see above). Interestingly, the position of T204 relative to Y174 and helix F does not change (Figure 3.4.5). Unfortunately, the structure resolution does not allow for the analysis of the changes in position of water molecule, bonding to W171 and T204 backbone. W171, forming a steric conflict with the retinal C_{13}-methyl group, is pushed away from retinal by ~0.6 Å. W76 is displaced towards helix F in the membrane plane.

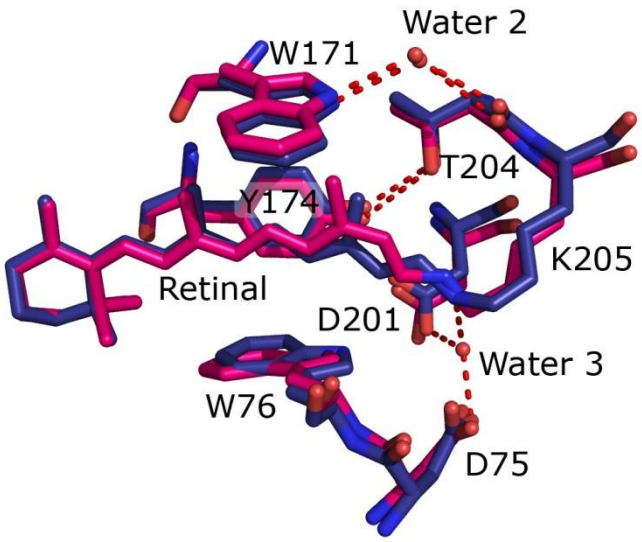

Figure 3.4.5. Comparison of the retinal environment in the NpSRII ground state (blue) and active state (magenta) structures. Note that the active state is a mix of the M and O states, and retinal is shown only as it would be in the M state in 13-*cis* conformation. Positions of other residues represent the averages of those in the M and O states and reflect the general conformational changes in this region. Hydrogen bonds are shown by dashed lines. Water molecule 3 disappears in the active state.

Electron densities are observed only for few highly-ordered water molecules in the extracellular channel. Nevertheless, the absence of water molecule adjacent to the Schiff base

(analogous to W402 of BR) is clearly seen, as there are no $2F_o$-F_c densities in that region and, in addition, inclusion of this water molecule into the model results in pronounced negative difference electron density maps F_o-F_c. Disappearance of this water molecule is the result of the Schiff base deprotonation and was observed in the *Np*SRII/*Np*HtrII complex[129] and in BR[249] as well.

Table 3.4.2. Diffraction data collection and refinement statistics

	NpSRII G-state	NpSRII active state
Data collection		
Space group	$C222_1$	$C222_1$
Cell dimensions		
a, *b*, *c* (Å)	86.99, 128.09, 50.63	86.44, 131.09, 51.23
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	28-1.9 (2.0-1.9)*	43-2.6 (2.74-2.6)*
R_{merge} (%)	7.0 (39.3)*	11.7 (83.7)*
$I / \sigma I$	14.4 (3.8)*	6.5 (1.5)*
Completeness (%)	96.9 (97.4)*	92.7 (86.1)*
Redundancy	3.8 (3.8)*	3.4 (3.1)*
Refinement		
Resolution (Å)	22-1.9	20 - 2.6, 2.9, 2.6**
No. reflections	20672 (1107***)	7633 (401***)
R_{work} / R_{free} (%)	15.34 / 17.31	25.12 / 27.58
No. atoms		
Protein	1694****	3184 (1552*****)
Ligand/ion	375	726 (352*****)
Water	73	88 (15*****)
B-factors		
Protein	11.09	18.39 (18.08*****)
Ligand/ion	58.70	54.30 (54.27*****)
Water	27.28	22.66 (27.15*****)
R.m.s. deviations		
Bond lengths (Å)	0.013	0.012
Bond angles (°)	1.248	1.393

*Values in parentheses are for highest-resolution shell

** For M-state structure anisotropic resolution limits were used

*** Number of reflection not used for refinement (free reflections)

**** Including double conformers

***** Number of atoms whose position was refined (atoms of active-state model)

3.4.7 Functionally important conformational changes in NpSRII

The conformational changes observed for free NpSRII and NpSRII in complex with its transducer are quite similar in their nature (Figure 3.4.4). Major movements are observed for helices F and G which have been correlated with signal propagation from the receptor to the transducer[126,129,250,251]. Additionally, an inward movement of the extracellular part of helix C is observed. Although these conformational changes are detected in both cases, in the case of NpSRII alone they have almost twice as large amplitudes (Figure 3.4.4). Smaller changes in the complexed NpSRII cannot be attributed to crystal packing, as it is looser in the crystals of NpSRII/NpHtrII, especially in the $I2_12_12_1$ space group (Ishchenko *et al.*, to be published). Thus it means that the transducer opposes such structural rearrangements in the receptor. As this opposition leads to conformational changes in transducer itself, it constitutes a signal transfer.

In addition to the helix movements, we observe conformational changes in the E-F loop, which are not present in the NpSRII/NpHtrII complex. They may result from the crystal packing (E-F loops of adjacent proteins form a crystal contact), or may be a genuine conformational change in the active state of uncomplexed NpSRII. It has been shown by solid state NMR spectroscopy, fluorescence and EPR spectroscopy that the E-F loop is interacting tightly with the transducer[155,156,252]. However, it is clear that the E-F loop is not crucial for signal transduction, as NpSRII E-F loop deletion mutants as well as BR triple mutant, which has a small disordered E-F loop, are all capable of signaling[137,253]. The remaining question is whether the E-F loop enhances the signal transduction from NpSRII to NpHtrII or just serves for NpSRII stabilization.

So far we have discussed the large conformational changes between ground and active states of NpSRII. Unfortunately, due to the resolution of diffraction data, we are not able to observe neither the minute changes in the structure nor movements of the water molecules in the cytoplasmic channel. Nonetheless we are able to make a conclusion, that there are no changes in the uncomplexed NpSRII active state in the relative positions of the residues T204 and Y174 (Figure 3.4.5), which are proposed to be important for the signal generation in NpSRII/NpHtrII complex[254,255]. Probably, this reflects the spectroscopically observed absence of the alteration in T204-Y174 bonding in uncomplexed NpSRII[256,257].

3.4.8 Implications for NpSRII proton pumping

Like BR, NpSRII is able to pump protons across the membrane, though less effectively. On formation of the NpSRII/NpHtrII complex this function is inhibited[139,140]. This inhibition may originate from kinetics alteration, from structural changes prohibiting the pumping, or from combination of these factors. Here, we analyze the implications of structural changes.

NpSRII active state transition results in disappearance of the water molecule, which makes a hydrogen bond to the retinal in the ground state (Figure 3.4.5). The same is observed in BR M state[258], as well as in NpSRII/NpHtrII complex M state[129]. There are also conformational changes of up to 1 Å in the cytoplasmic channel, probably associated with the charge redistribution. Slight outward tilt of the G helix, combined with E-F loop changes, may facilitate the formation of water chain, which is believed to be needed for the retinal reprotonation and thus for completing the pumping cycle. These changes are not observed in the M state of complexed NpSRII, and may account for the NpHtrII proton pumping inhibition.

There are two ways in which NpHtrII may affect NpSRII dynamics. The first possibility is that NpHtrII may directly influence NpSRII conformational changes. The other possibility is as follows. The hydrophobic/hydrophilic boundary of NpSRII is uneven in the region of the helices F and G, as seen from the surface potential distribution (Figure 3.4.1). Thus, the natural tendency of a lipid bilayer to be flat would result in energetic penalty for the NpSRII ground state. In the active state the mismatch is lowered, and thus the active state is preferred from the bilayer curvature energy point of view. NpHtrII shields the helices F and G from the lipid bilayer and thus reduces its influence on the NpSRII structure. In that case the energy of the active state is higher than in the absence of NpHtrII, and thus the conformational changes are lower.

3.5 Modeling of the Natronomonas pharaonis HtrII HAMP domain region

Sensory rhodopsin transducers are multidomain proteins that pass the signal from the sensory rhodopsin to the kinase CheA. The domains common to all sensory rhodopsin are the transmembrane part, responsible for interaction with the sensory rhodopsin, HAMP domains 1 and 2, separated by the inter-HAMP region, and the elongated kinase control module (Figure 1.2.5). Structural models of all these domains except the inter-HAMP region are established: there are the crystallographic structure of the *Np*HtrII transmembrane part[125], NMR and crystallographic structures of various HAMP domains[148,149], and crystallographic structure of the kinase control module[259].

Structure of the *Np*HtrII proteolytic fragment (residues 100-159) that includes the inter-HAMP region was studied by Hayashi *et al.* by means of NMR[171]. It was found that the inter-HAMP region residues 135-150 adopted the α-helical structure, and the other residues that correspond to the HAMP domains were disordered. The inter-HAMP region environment was non-native. No clear dimerization, expected in the correctly folded protein, was observed. Consequently, the atomic model of the *Np*HtrII inter-HAMP region was not established.

In this section, the model of the inter-HAMP region of *Np*HtrII and other proteins (*Hs*HtrI, *Hs*HtrII) is proposed. Stability of the model in different simulation conditions is tested and its properties are studied extensively.

3.5.1 Model of the inter-HAMP region of SR transducers

First, the PSIPRED protein structure prediction server[260] was used to analyze the sequences of studied proteins. For all of them, the α-helical structure is predicted with a high confidence for sequence fragments that consecutively include the helix AS2 of the first HAMP domain, the inter-HAMP region and the helix AS1 of the second HAMP domain. The same fragments are predicted to be in a coiled coil by the coiled coil prediction server[261] (Figure 3.5.1).

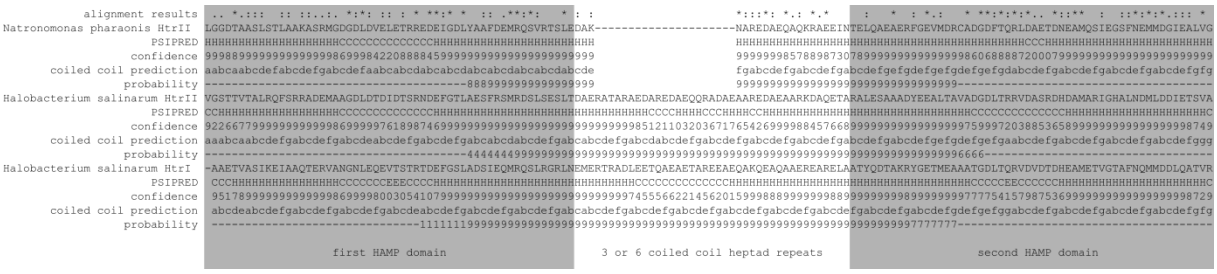

Figure 3.5.1. Alignment of *Np*HtrII, *Hs*HtrI and *Hs*HtrII sequences with prediction results. Predicted HAMP domains are shaded grey. The regions comprising AS2 of the first HAMP domain, inter-HAMP and AS1 of the second HAMP domain are predicted with a high probability to be in the coiled coil conformation.

All inter-HAMP regions studied here have a pronounced continuous hydrophobic groove when modeled as an α-helix (see Figure 3.5.2a for NpHtrII). Interestingly, this groove is composed mostly of alanines. Two identical fragments, modeled as ideal α-helices in PyMOL and located to have the interface at the hydrophobic groove, form a coiled coil structure in time less than 100 ps for the short inter-HAMP region 135-153 of NpHtrII (Figure 3.5.2b), and less than 1 ns for the longer inter-HAMP regions 356-400 of HsHtrII (Figure 3.5.2c) and 106-150 of HsHtrI. After the coiled coil formation, the structure remains stable during the whole length of the simulation (see Table 2.5.3 and Figures 2.5.1-3 for additional information). In the longer inter-HAMP regions, deviations from the ideal α-helical geometry are clearly seen, as is typical for coiled coils. The helices are bent around each other, with a total twist of 90° (Figure 3.5.2c). The twist is defined as the angle between projections of the lines, connecting the alpha-helices' starts and ends, on the plane perpendicular to the coiled coil axis (Figure 3.5.2d).

For all the modeled inter-HAMP regions, the homodimer adopts the asymmetric conformation, with a longitudinal shift of one protomer relative to the other of about 1 Å. The reason for this is most clear in the case of NpHtrII. Charged residues framing the hydrophobic core are distributed in such a way that the shifted conformation is preferred (Figure 3.5.2b). Namely, the positively charged R142' may be electrostatically attracted either by the negatively charged D137 or by the negatively charged D144. Simulations show that the conformation when R142' attracted by both D137 and D144 is unstable. Attraction of R142' of one protomer to D137 of the other leads to a longitudinal displacement, so that R142 comes closer to D144', and vice versa. The resulting shift is additionally stabilized by short-living bonds between Q149' and D144. Designations without prime correspond to one protomer, and those with prime correspond to the other protomer.

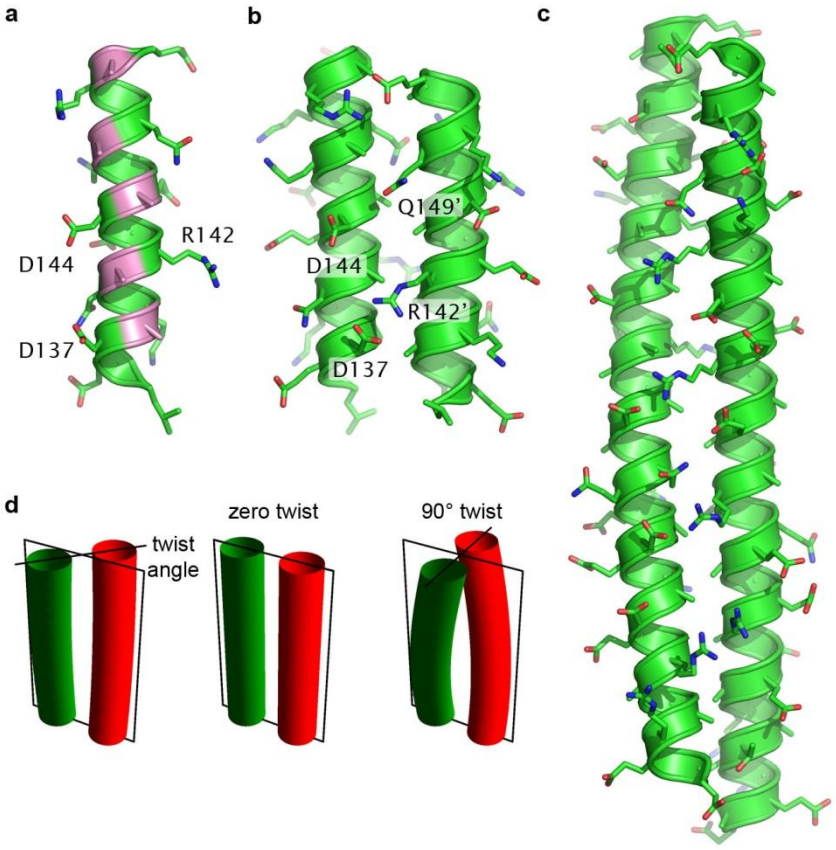

Figure 3.5.2. Proposed models for the inter-HAMP region. a) Hydrophobic groove of the NpHtrII inter-HAMP 135-153 fragment. Alanines are highlighted in pink. b) Structure of a homodimer of the two NpHtrII inter-HAMP 135-153 regions. Note the asymmetric position of protomers and the differences in ionic bonds formed on each side of the structure. c) Structure of a homodimer of the two HsHtrII inter-HAMP 356-400 regions. Positions of the protomers are also asymmetric. The structure presented in a) is determined by NMR[171] (PDB ID 2RM8), b) and c) were built manually and subjected to molecular dynamics. d) Definition of the twist angle in the inter-HAMP region. The twist is defined as the angle between projections of the lines that connect the alpha-helices' starts and ends, on the plane perpendicular to the coiled coil axis. Examples of the zero and 90º twists are given.

For a quantitative description of the inter-HAMP domain we calculated free-energy profiles for the longitudinal shift using umbrella sampling and weighted histogram analysis method. For these calculations, the longitudinal shifts are biased towards some particular values, and their distributions are analyzed to devise the profile. As the exact ion molarity inside the *N.p.* cells is not known, the simulations were conducted at 0.5, 1, 2 and 4 M (Figure 3.5.3).

Despite different molarity of the ions, the free energy profiles are similar (Figure 3.5.4). Energy minima are observed at about ±1.3 Å. The barrier height is slightly different for

different molarities, namely 1.5 kcal/mol for 0.5 M, 1.4 kcal/mol for 1 M and 2 M and 1.2 kcal/mol for 4 M. These differences fall within the expected error of the method.

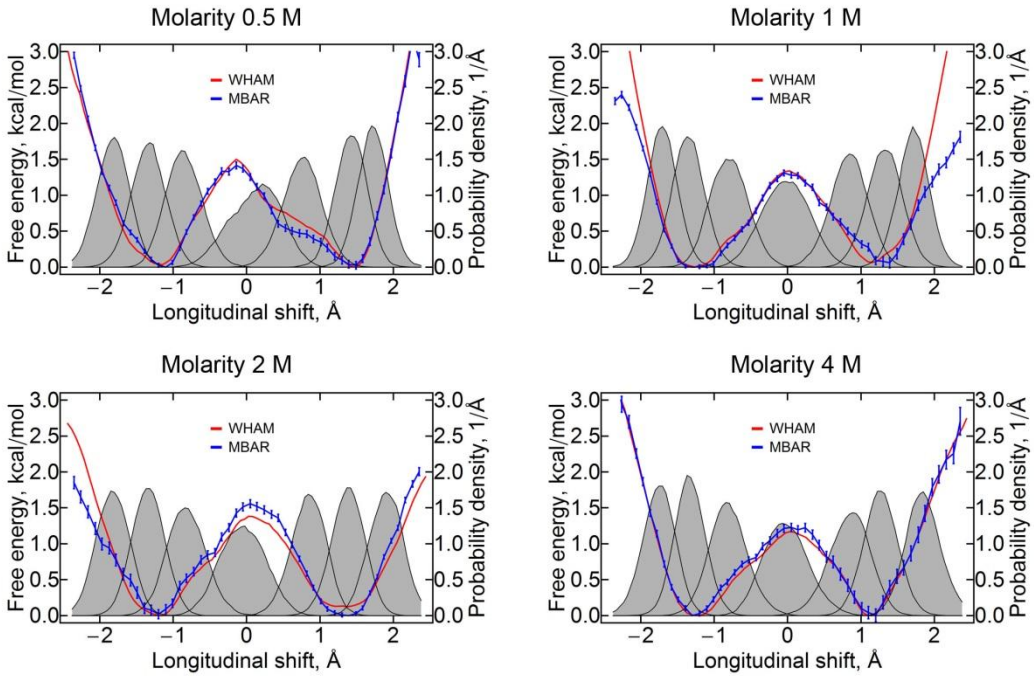

Figure 3.5.3. Potentials of mean force (PMFs) of the NpHtrII inter-HAMP domain at different salt concentrations. PMFs were calculated using weighted histogram analysis method (WHAM)[262] and multiple Bennett acceptance ratio (MBAR) method[262]. Results obtained by different methods correspond well. Umbrella sampling distributions are shown in grey. Both methods give similar results. MBAR also calculates the statistical error.

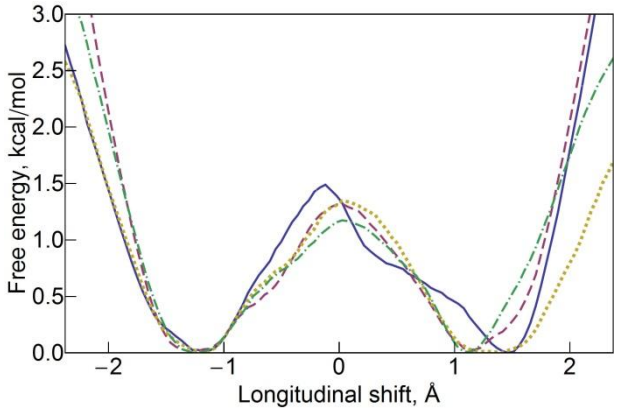

Figure 3.5.4. Free energy profiles for the relative longitudinal shift of the protomers in the inter-HAMP region (*Np*HtrII residues 135-153) at different molarities: 0.5 M (blue, solid), 1 M (magenta, dashed), 2 M (yellow, dotted) and 4 M (green, dot-dashed). Note that the minima are at about ±1.3 Å and the barrier height is 1.2-1.5 kcal/mol.

We believe that the observed features are not artifacts of the calculation, as the standard forcefield was used (CHARMM27). There are no regions with a high induced electronic polarization in the inter-HAMP region, so effects of different forcefields on the obtained PMF profile were not addressed, as is sometimes done for free energy calculations of ion conduction through channels or for calculations of small ligand binding affinities. Weak variability of the free energy profile within the wide ion concentrations range also reflects robustness of the observed features.

To check for possible systematic errors introduced by WHAM, and to estimate statistical errors, we recalculated the free-energy profiles using multiple Bennett acceptance ratio method (MBAR)[262]. This method gave similar results (Figure 3.5.3).

Therefore we conclude that the observed asymmetry is not a calculation artifact.

3.5.2 MD simulations of Af1503 and *Np*HtrII HAMP domains

In order to check the correspondence of the NMR HAMP domain structure[171] to those of Htrs we modeled both HAMP domains of *Np*HtrII (residues 85 to 134 and 155 to 210 accordingly), along with the original HAMP domain of the putative protein Af1503 of *Archaeoglobus fulgidus*. Initial models generated by MODELLER were equilibrated for 4 ns and appeared to be very stable. NMR model (PDB 2ASW) of Af1503 HAMP domain, containing residues 276 to 331, was later expanded by residues 332-338 (aminoacid codes SLEEALK), modeled as an α-helical continuation of AS2. These last seven residues of Af1503 were not included in the construct for NMR structure determination, but could somehow influence the structure of the HAMP domain. The resulting expanded structure was also stable for 8 ns. Average RMSD value for the backbone atoms during the simulations was 1.5 Å for the HAMP1 of NpHtrII, 1.7 Å for the HAMP2 of NpHtrII, 1.4 Å for Af1503 276-331 fragment and 1.6 Å for Af1503 276-331 fragment. A more extensive stability analysis is given in section 2.5.x. Ribbon diagrams of the resulting structures are presented in Figures 3.5.5a and 3.5.5b.

3.5.3 MD simulations of the NpHtrII HAMP domain region

To check whether the structure of the HAMP domain is affected by the inter-HAMP region or vice versa, we performed molecular dynamics studies for two constructs. HAMP1 with the inter-HAMP region of NpHtrII (residues 85 to 153) were simulated for 10 ns, and HAMP1, the inter-HAMP and HAMP2 of NpHtrII (residues 85 to 210) were simulated for 60 ns.

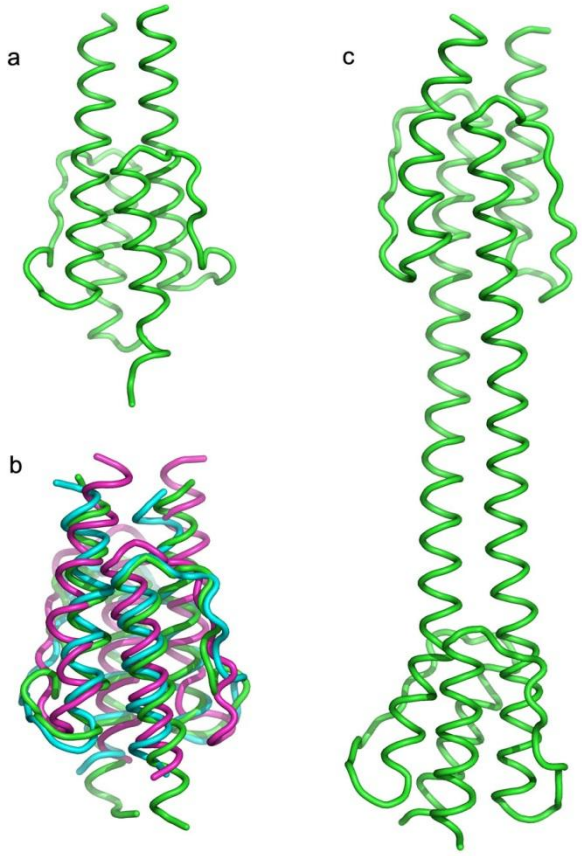

Figure 3.5.5. Modeled constructs. a) Ribbon diagram of the Af1503 HAMP domain, expanded by the 7 terminal amino-acids, after 8 ns of molecular dynamics. b) Overlay of the initial NMR structure of the Af1503 HAMP domain (in green) with homology-modeled structures of the first (blue) and second (purple) NpHtrII HAMP domains. c) Proposed model for the NpHtrII HAMP domain region (residues 85-210), including both HAMP domains and the inter-HAMP region. Note the asymmetry between the protomers and the tilt of the HAMP domains' axes relative to the axis of the inter-HAMP region.

Though large structural movements are seen, they do not change the architecture of these constructs. Root mean square deviations for the whole NpHtrII HAMP domain region are presented in Figure 3.5.6.

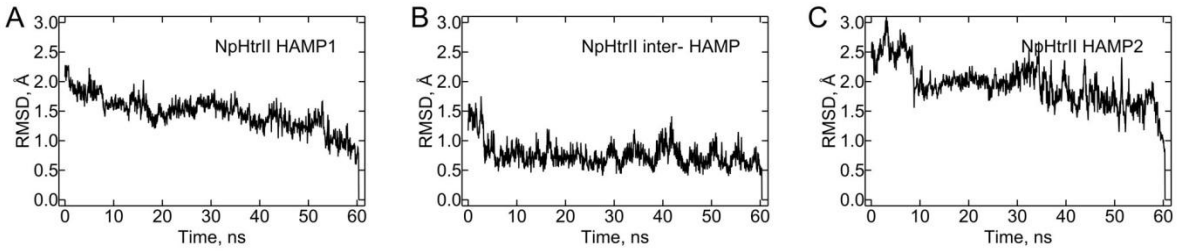

Figure 3.5.6. RMSD of the backbone atoms for different parts of the NpHtrII HAMP domain region, simulated as a whole. The last trajectory snapshot was used as a reference.

Additional extensive stability analysis for this construct is given in Sections 2.5.x and 2.5.y. Despite initial symmetrical conformation of the system, a relative longitudinal shift of corresponding α-helices develops after a few nanoseconds, and persists for the whole length of the simulation. As opposed to the zero twist angle of the inter-HAMP region alone, it equals to approximately 45° in the presence of the HAMP domains. All the domains retain their general conformations (Figures 3.5.7-8). However, the shift imposed by the inter-HAMP region results in a reorientation of corresponding helices and renders HAMPs highly asymmetric (Figure 3.5.5c, 3.5.7-8). The relative position of the two HAMP domains changes considerably, though directions of the main axes of both HAMP domains coincide well with each other. The angle between HAMP1 and HAMP2 axes is 9±5°, as opposed to 13±4° for the angle between the HAMP1 and the inter-HAMP and 16±6° for the angle between the HAMP2 and the inter-HAMP.

To quantify the asymmetry in the HAMP domains and the inter-HAMP region, double distance matrices were calculated (Figures 3.5.7b, 3.5.8d and 3.5.9d). There, the pictured value is distance(i, j) - distance(i' ,j') for the distances within one protomer, and distance(i ,j') - distance(i', j) for the distances between residues from different protomers. Distance(i, j) stands for the distance between the centers of masses of residues i and j correspondingly. Indices without prime correspond to the first protomer, and those with prime correspond to the second protomer. AS1 and AS2 correspond to the first and the second alpha-helices of the HAMP domain.

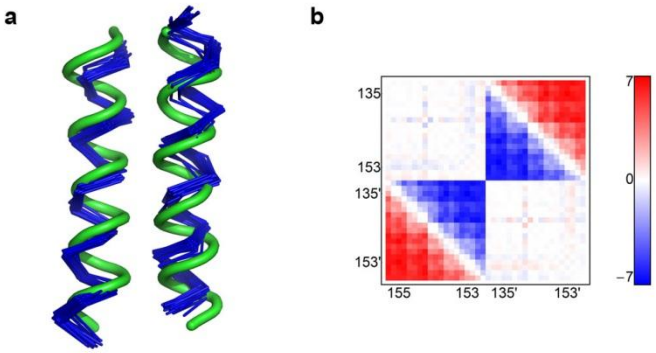

Figure 3.5.7. Packing changes in the inter-HAMP region of NpHtrII simulated as a part of a larger construct (HAMP1 - inter-HAMP - HAMP2). a) Superposition of the symmetrical structure (green tube) and the trajectory snapshots, taken each 2 ns (blue ribbon). b) Double distance matrix.

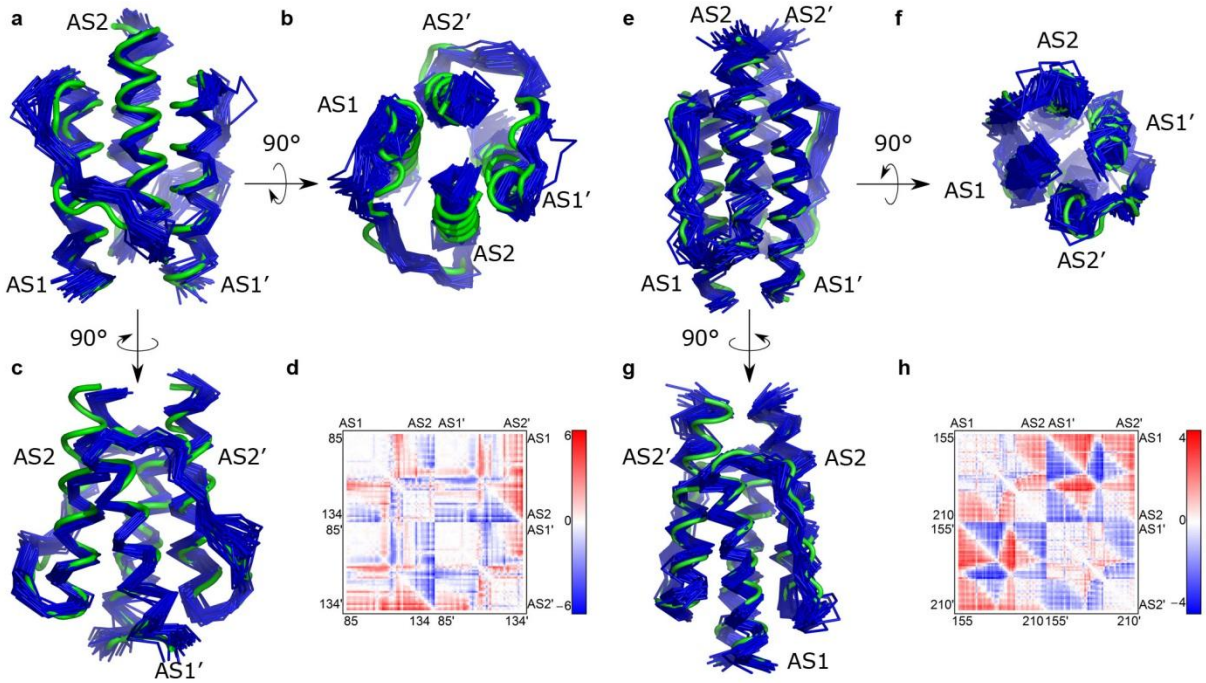

Figure 3.5.8. Packing changes in the NpHtrII HAMP domains simulated as a part of a larger construct (HAMP1 - inter-HAMP - HAMP2). The data related to the first HAMP domain is shown to the left, and data related to the second HAMP domain is to the right. a,b,c,e,f,g) Superposition of the symmetrical structure (green tube) and the trajectory snapshots, taken each 2 ns (blue ribbon). d,h) Double distance matrices. Both domains possess a prominent asymmetry in the structure, especially in the region adjacent to the inter-HAMP region. The changes include mostly the relative sliding of the helices AS2 and AS2' for HAMP1 and the helices AS1 and AS1' for HAMP2. Changes in the HAMP domain hydrophobic cores are minimal and the structures show little variation with time. For the first HAMP domain the AS1-AS2 linker is disturbed, as R112 and R113 sidechains form different ionic bonds and the hydrophobic L105 is exposed.

As one can suppose that the transmembrane helices TM2 and TM2' may continue in the HAMP1 helices AS1 and AS1' without a kink (see a hypothetical model on Figure 3.5.9), we considered it interesting to analyze the motion of HAMP2 relative to the membrane plane and to the transmembrane segment of NpHtrII. For this purpose we calculated the probability density of the projection of the HAMP2 center of mass (COM) on the membrane plane relative to the HAMP1 COM (Figure 3.5.10, including a symmetrical image that corresponds to the longitudinal shift of the opposite value). Probability density distribution function was determined as the time which the projection of the second HAMP domain' center of mass onto the membrane plane spent in a given 1 Å × 1 Å square, divided by the total simulation time. The distance between the projections of centers of mass of the two HAMP domains is 8.6±3.5 Å. This value is approximately equal to what follows from simple geometric

considerations: the length of the inter-HAMP region is approximately 50 Å, the mean distance between centers of the helices is 9 Å, the mean longitudinal shift value is 1.3 Å and thus the in-plane displacement is about 7 Å (Figure 3.5.11c).

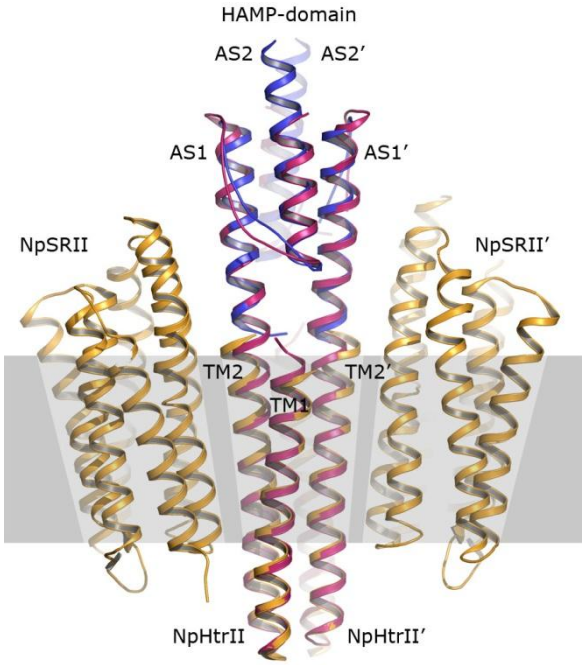

Figure 3.5.9. A hypothetical model of the junction between the HAMP domain and the trans-membrane segment of NpHtrII (in magenta), aligned to the known structures. Crystallographic structure of the NpSRII with NpHtrII transmembrane part (residues 23-82, PDB 1H2S) is shown in yellow, NMR structure of the Af1503 HAMP domain (PDB 2ASW) is shown in blue. Labels are shown for the NpHtrII trans-membrane helices TM1 and TM2 and the HAMP domain helices AS1 and AS2. Approximate position of the cell membrane is shaded gray. This model illustrates the theoretical possibility of TM2 and AS1 to form a continuous helix without a kink. The model of the junction was prepared by minimization of the crystallographic transmembrane part attached to homology modeled structure of the HAMP1 domain using SAMSON program with CHARMM19 forcefield. The region of the junction was subjected to alpha-helical restraints. No steric clashes are observed in the model. No studies were conducted on the stability of such structure.

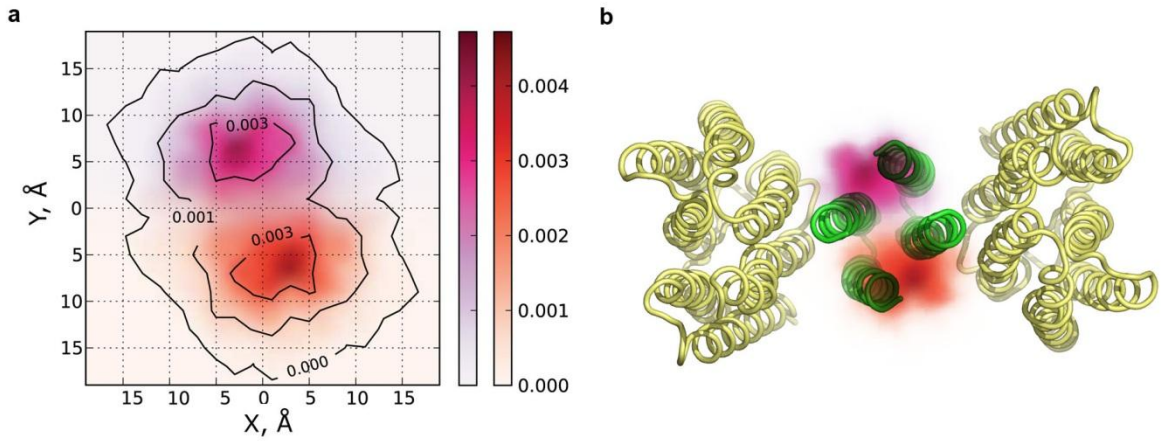

Figure 3.5.10. Displacement of the second HAMP domain in the membrane plane. a) Computed probability of finding the projection of the NpHtrII second HAMP domain center of mass on the membrane plane at the corresponding X, Y coordinates, in $1/Å^2$ units. View from the cytoplasm. X axis is directed along the centers of mass of the HAMP1 AS1 helices, Y axis is perpendicular to it, and both axes lie in the membrane plane. The initial structure was symmetric, and the tilt and the longitudinal shift in the inter-HAMP region have developed later. Two regions with the highest probability correspond to the two possible signs of longitudinal shifts and are colored red and magenta, correspondingly, to guide the eye. b) The same probability density distribution laid over the structure of NpSRII-NpHtrII 2:2 complex in a proper scale (PDB ID 1H2S).

3.5.4 Discussion

The model of the inter-HAMP region of sensory rhodopsin transducers has been described in the previous sections. This model is similar to the signaling helix (S Helix)[263] in that it continues the helix AS2 of the HAMP domain, however, there are also some differences. First, at its C-terminus, the inter-HAMP region also ends with a HAMP domain helix (AS1). Second, *a* and *d* coiled coil positions are occupied by alanines in the inter-HAMP region, as opposed to bulky amino-acids of the S Helix. This may facilitate longitudinal motions of protomers along each other.

Molecular dynamics simulations show that a symmetric conformation of the inter-HAMP domain is unfavorable. A longitudinal shift between the inter-HAMP protomers is observed. Since they are identical, a longitudinal shift of the same value but of the opposite sign is also possible. It means that the inter-HAMP is bistable. The energetic barrier between the two states is high enough to prevent frequent switching by thermal fluctuations, but it is much lower than 60 kcal/mol – the average energy of the photons, absorbed by sensory rhodopsins.

The asymmetric conformation of the inter-HAMP region persists when the HAMP domains are also included in the simulation. Thus, this could be biologically relevant. It is important to note that the asymmetry of the inter-HAMP region is coupled with the asymmetry of the HAMP domains.

The established properties of the inter-HAMP region may also have direct implications for understanding of the signal propagation through the cytoplasmic part of the transducers. As the inter-HAMP α-helices constitute a coiled coil and are always turned to each other by their hydrophobic groove, signaling through the changes in the inter-helical angle as supposed in (8) is questionable. We suggest that the probable mechanism of the signal transduction may involve either switching of the inter-HAMP region or some rigid body motion of the cytoplasmic part without any changes in the inter-HAMP state (Figure 3.5.11). Switching might result in a large structural rearrangement (Figure 3.5.11a), and thus could be prohibited by tight hexagonal packing of KC fragments. Signal encoding by a displacement of the whole cytoplasmic part of the transducer, i.e. by a position of the HAMP2 relative to the HAMP1, in its turn would be affected by very large fluctuations (Figure 3.5.10).

As we have noted, asymmetry of the inter-HAMP is enforced by electrostatic interactions of oppositely charged side-chains of corresponding residues. Flexibility of those side chains allows some longitudinal displacements (up to the zero shift) without breaking the formed electrostatic bonds. Thus, the evolution of the system with time may be different depending on which bonds are formed, that is, the history of the system. This means that the inter-HAMP region is in effect a multistate switch.

It is worth to mention that a study of the HAMP domain alone would not provide sufficient information about the signal transduction.

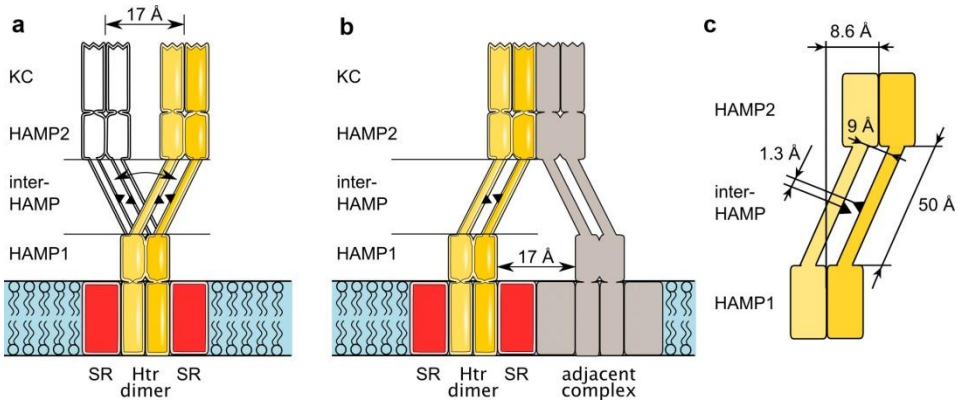

Figure 3.5.11. Two possible roles of the HAMP domain region of phototactic signal transducers. Model of the sensory rhodopsin – transducer 2:2 complex is in yellow, with the protomers shown in different shades. Equivalent positions are marked on the inter-HAMP regions of the two protomers as black triangles to facilitate the perception of the longitudinal shift. a) Change of the longitudinal shift sign results in a distinct conformation, in which the kinase control module (KC) is displaced on average by 8.6 Å × 2 ≈ 17 Å along the membrane plane (second conformation is in white). Thus, the signal may be transduced through the inter-HAMP switching. b) Asymmetry of the HAMP domain region may facilitate the contacts with the adjacent transducers or receptors. Otherwise, these contacts would be impaired by bulky sensory rhodopsins, residing in the membrane. c) Simple geometrical model of the conversion of the inter-HAMP longitudinal shift in the displacement of the transducer's cytoplasmic part. Data presented are the mean values. 8.6 Å is the mean displacement of the second HAMP domain relative to the first one in the membrane plane. 50 Å is the approximate length of the inter-HAMP region. 9 Å is the approximate distance between the axes of the inter-HAMP helices. 1.3 Å is the mean longitudinal shift in the inter-HAMP region.

3.6 Two Distinct States of the HAMP Domain from Sensory Rhodopsin Transducer Observed in Unbiased Molecular Dynamics Simulations

After the model of the HAMP domain region of sensory rhodopsin transducers was established, which means that the structures of all transducer modules became known, the obvious next step was to start studying the signal transduction by the transducer proteins. Meanwhile there is a lot of experimental information on signal generation in the transmembrane region, transduction of the signal to and through the HAMP domain is much less studied. In this section, MD simulations of the first HAMP domain of NpHtrII are presented that shed light on its mechanism of function.

3.6.1 MD simulations of NpHtrII HAMP1.

We have performed the molecular dynamics study of the HAMP domain from halobacterial phototactic signal transducer NpHtrII, the first one of the two HAMP domains present in the protein. The simulations consist of 28 trajectories, each lasting more than 205 ns. The total length of the simulations is more than 6.0 μs. Details of the simulations are presented in the Table 3.6.1. Overall, the domain was not changing its fold during the simulations. The average RMSD of the backbone atoms N, C, Cα, O for the whole simulation is 1.35±0.33 Å. The observed structure of the domain was quite similar to that reported previously[4,205].

Table 3.6.1. Details of the performed simulations of the first HAMP domain of NpHtrII.

Simulation #	Starting coordinates	Forcefield	Number of trajectories and their length	Average RMSD of backbone atoms
1	Symmetrical homology model based on the Af1503 HAMP domain	CHARMM22 with CMAP correction	10 × ~205 ns	1.1 Å
2	Symmetrical homology model based on the Af1503 HAMP domain	AMBER ff99-SB-ILDN	10 × ~205 ns	1.3 Å
3	"Resting state" conformations from different trajectories of simulation #1	CHARMM22 with CMAP correction	5 × ~205 ns	1.2 Å
4	"Active state" conformations from different trajectories of simulation #1	AMBER ff99-SB-ILDN	3 × ~205 ns	1.4 Å

We started the simulations from the homology model based on the NMR structure of the Af1503 HAMP domain. Preliminary results have shown that for optimal simulation several factors have to be taken into account. First, the CHARMM22-CMAP and AMBER ff99-SB-ILDN forcefields bias the structure in different ways and thus using only one forcefield is not reliable. We conducted the simulations using both forcefields. Second, the trajectories are highly sensitive to the starting structure, which is the equilibrated homology model. As the temperature and pressure equilibration process is non-deterministic, we calculated several trajectories for each forcefield with the model independently equilibrated each time. This resulted in a good sampling of the HAMP domain conformational space as each starting structure had a RMSD of backbone atoms' positions of ~0.3 Å relative to the energy-minimized structure and of ~0.5 Å relative to the other starting structures. Finally, the N- and C-termini of the HAMP domain were unfolding on the scale of 20-100 ns. As we expect this to be a result of non-native truncation of the model, the α-helical structure of 4 residues from each terminus was restrained.

3.6.2 Two states of the NpHtrII HAMP domain

Initial simulations #1 and #2 (Table 3.6.1) with the forcefields CHARMM22-CMAP and AMBER ff99-SB-ILDN revealed significant motions of the HAMP domain's helices (data for the CHARMM forcefield is presented in Figure 3.6.1a; data for the Amber forcefield are similar), despite the relatively low overall RMSD values (Figure 3.6.1b; data for the AMBER forcefield are similar). Two metastable states were discernible by visual examination. To obtain quantitative measures, we employed the principal components analysis[214]. The analysis revealed that the motions are dominated by the first principal component (PC1, 35% of total variation, meanwhile the second and third components account for 8.5% and 7% correspondingly), for which the distribution of the projections is bimodal; the distribution is unimodal for other principal components (Figure 3.6.2).

There were differences between the simulations performed using the CHARMM and the AMBER forcefield (Figure 3.6.3a,b). Meanwhile in the AMBER simulation the projections on the first principal component group around the value of ~1.3, the projections in the CHARMM simulation follow a bimodal distribution with centers at ~0.2 and -2.1. AMBER favors the values in the range 1÷2 and CHARMM favors the values in the range -3÷-1. In some of the CHARMM trajectories, the domain immediately switched to the negative

projections (trajectories 4, 5 and 8), meanwhile in others it stayed close to the initial structure for some time (trajectories 3, 6 and 7). Some trajectories revisited the starting state.

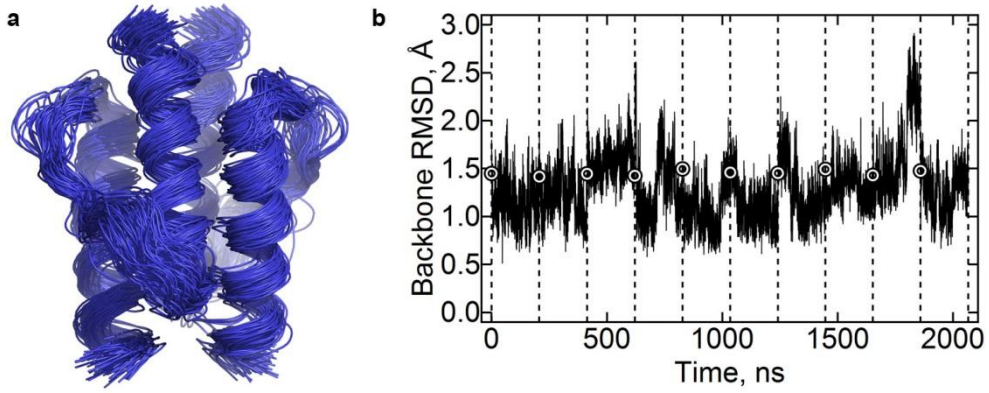

Figure 3.6.1. Stability of the HAMP domain in molecular dynamics simulations. a) Superposition of the conformations observed in the simulations with CHARMM22-CMAP forcefield. The snapshots taken each 20 ns are shown. The domain remains stable during the simulations. b) Root-mean square deviation of the backbone atom positions relative to the average observed in the simulations with CHARMM22-CMAP forcefield. The mean value is 1.3 Å. The data for ten trajectories are shown consecutively; they are separated by the dashed lines. The circles denote the value at the start of each trajectory.

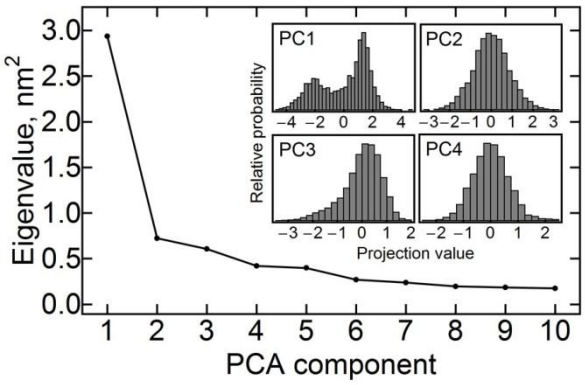

Figure 3.6.2. Principal components analysis of the calculated trajectories. The covariation matrix is dominated by the first principal component (PC1) that accounts for 35% of the matrix trace. Histograms of the trajectory projections on the corresponding principal components are shown in the inset. For PC1, the distribution is bimodal, and for the other principal components it is unimodal. From that, we conclude that PC1 corresponds to transitions between two distinct states, and the other components reflect thermal fluctuations.

We were interested to determine whether there are really two distinctive states (with the average projections of 0.2÷1.3 and -2.1) or this is a result of a forcefield bias. To that end, we have performed additional simulations #3 and #4, with the structures from the simulation #1

taken as starting poses (Figure 3.6.3a). In the CHARMM forcefield simulations, the starting structures had initial projection values in the range 0÷1, and in the AMBER forcefield simulations, the starting structures had initial projection values in the range -4÷-2 (Figure 3.6.3c,d, Table 3.6.1). In both simulations, the domain spent at least some time in the starting state. Transitions from the starting state to the forcefield-preferred state were observed in both simulations and occurred in a switch-like fashion (Figure 3.6.3c,d). Thus, we conclude that each forcefield recognizes two distinct states. At the same time, CHARMM favors the state with the average projection of -2.1 meanwhile AMBER clearly favors the other state with projections in the range of 0.2÷1.3.

The eigenvector corresponding to PC1 characterizes the structural details of the transition between the two discovered states. To determine it in the best way possible we have applied the analysis to the concatenated trajectory that includes all the simulations. The same PC1 was used to obtain all the results presented here (Figures 3.6.2-7). The PC1 determined from the concatenated trajectory has no analogs in the simulation #2, as the second state and the state transition are not present in the latter. In the other simulations, the principal component 1 is highly similar to the one determined from the concatenated trajectory, with the dot products of their normalized eigenvectors being 0.93, 0.77 and 0.95 for simulations #1, #3 and #4 correspondingly. Dot products of PC1 eigenvectors of simulations #1 and #3 as compared to #4 are 0.91 and 0.74.

It could be useful to estimate the free energy difference and the value of the energy barrier between the two states. However, the results obtained with the AMBER and CHARMM forcefields differ not only in the magnitude of the energy difference but also in its sign. Inclusion of the adjacent NpHtrII domains into the model might also change the energetics of the transition from one state to another. Thus, we do not assess the aforementioned energetic properties as the results of such assessment may be misleading. The presented simulations may serve as an example that, despite considerable progress in forcefield development[264,265], a choice of a forcefield can still affect the results, not only quantitatively but also qualitatively.

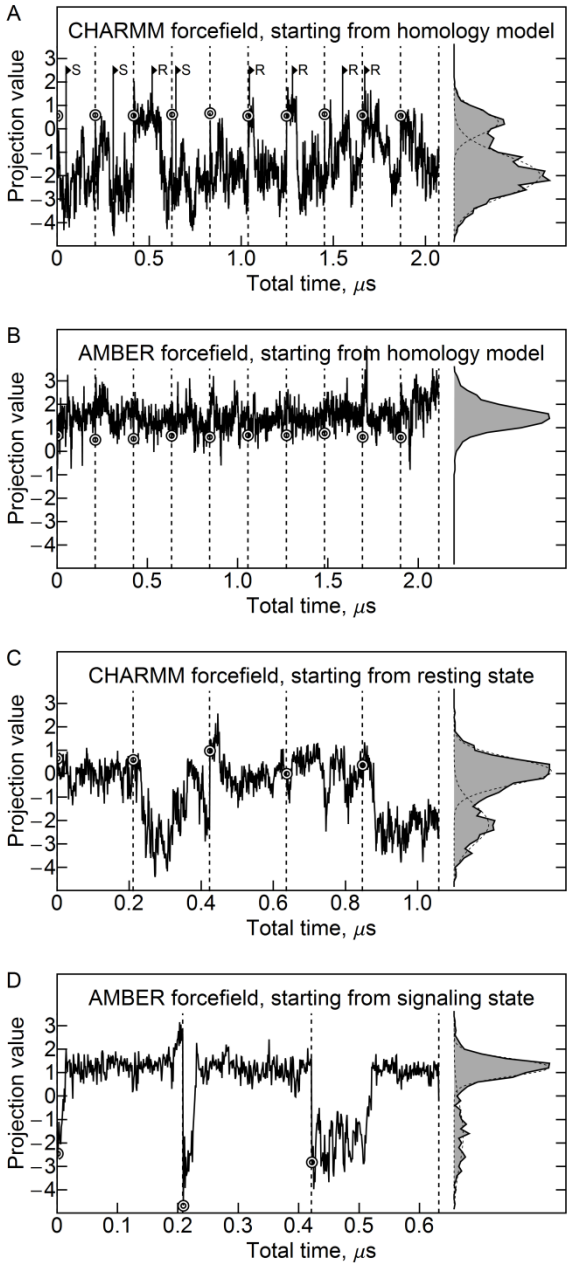

Figure 3.6.3. Projections of the trajectories on the first principal component. Panels A, B, C and D show the data for the simulations #1, #2, #3 and #4 correspondingly. The projections follow bimodal distribution with centers at approximately 0.2 and -2.1 for the CHARMM simulations and 1.3 and -1.7 for the AMBER simulations (fitted by Gaussian distributions). Higher projection values correspond to the resting state and lower values correspond to the signaling state. The CHARMM forcefield favors the signaling state, and the AMBER forcefield favors the resting state. The data for different trajectories are shown consecutively and are separated by the dashed lines. The circles denote the value at the start of each trajectory. The R (resting) and S (signaling) signs at the panel A denote the starting frames for the simulations #3 and #4.

3.6.3 Structural analysis of the two observed states

After establishment of the two states of the NpHtrII HAMP domain we proceed to the structural analysis. The most notable conformational rearrangement associated with the first principal component is the relative longitudinal displacement of the HAMP domain helices AS1 and AS2 (Figure 3.6.4). Such longitudinal displacement of the NpHtrII HAMP1 helices was recently observed experimentally[160]: the HAMP domain was more compact in the resting state and more elongated in the signaling state. Given this, we propose that the observed two states correspond to the resting (projection values 0÷2) and the signaling (projection values -3÷-1) states.

Another notable feature is the change in the cross-section shape of the HAMP domain (Figure 3.6.4d,e,f). In the resting state, the helices are arranged rectangularly, and in the signaling state they are arranged rhombically. This is similar to the compact and loose conformations of the HAMP domains observed in the crystallographic structures of the Aer2 aerotaxis transducer[145,149].

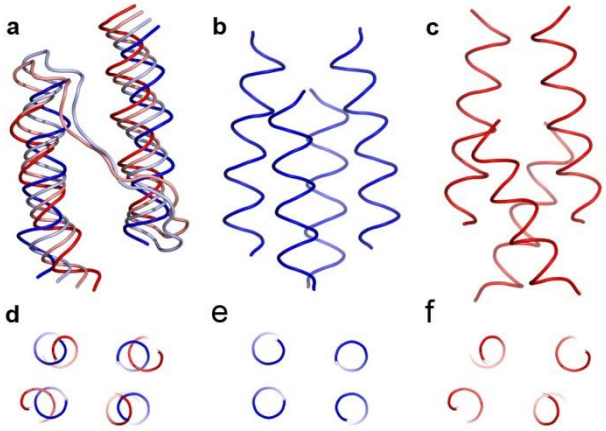

Figure 3.6.4. Conformational changes associated with the first principal component. a,b,c) Side view of the HAMP domain. d,e,f) View along the HAMP domain axis. The extreme projections are shown in blue (resting state) and red (signaling state), and the state averages are shown in light blue and pink correspondingly (A). In the resting state (B, E), the alpha-helices AS1 and AS2 are more parallel and the HAMP domain cross-section is rectangular. In the signaling state (C, F), AS1 and AS2 are displaced in opposite directions along the HAMP domain axis, are no longer parallel and the cross-section is rhombic. The extreme projections are shown as they illustrate best the conformational changes involved.

The conformations corresponding to the two states are almost symmetrical (Figure 3.6.4b,c). However, thorough analysis has revealed that in the resting state the positions of the core

hydrophobic residues F124 are not symmetric. Thus, the symmetry of the resting state structure presented in Figure 3.6.4 is a result of mixing of two asymmetric structures.

Asymmetry has already been noted in other simulations of the HAMP-domain containing systems[4]. It also has been noted that the NpHtrII HAMP1 possesses bulky hydrophobic aminoacids at its core[205] and the bulkiness of the insidious residues is crucial for the HAMP domain function[148,161,162]. To determine the correct asymmetric resting state, we proceeded to analyze the conformations of the F124 pair in different states of the HAMP domain.

3.6.4 Conformation of the F124 pair at the HAMP domain core

Pair of the F124 residues, residing at the HAMP domain core, adopts several conformations: symmetric S in the signaling state, asymmetric R and R_{sym} in the resting state and, finally, transitional R_{trans} on the way from R to R_{sym} and back (Figure 3.6.5). We analyzed the distributions of the χ_1 and χ_2 dihedral angles of F124 sidechain (Figure 3.6.6). In accord with the observations, the χ_1 angles may be in the (*gauche*$^-$, *gauche*$^-$), (*gauche*$^-$, *trans*) or (*trans*, *gauche*$^-$) conformation (Figure 3.6.6b). However, analysis of the χ_1 angles alone does not allow us to differentiate between the Rtrans and S state of the F124 pair. Plotting the χ_2 and χ_2' sum as a function of χ_1 and χ_1' sum allows to do that, as the Rtrans and S differ by 90º rotation of the benzene rings of both F124 residues (Figure 3.6.6c). Consequently, one can classify the state of the HAMP domain by looking at the sum of all the dihedral angles χ_1, χ_1', χ_2 and χ_2': the values in the range 15º÷125º correspond to the resting state, meanwhile the values in the range -215º÷15º reflect the signaling state.

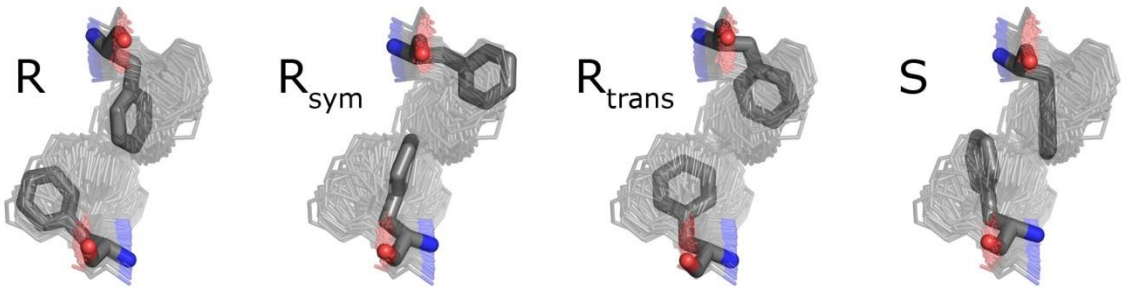

Figure 3.6.5. Configurations of F124 pair in the resting and signaling states. Representatives of each configuration are shown. For reference, the structures are highlighted in the ensemble of snapshots taken each 10 ns and aligned by the C, Ca, N and O atoms of F124. R and R_{sym} correspond to the resting state and are basically the same state as they are related by 180º rotation around the HAMP domain axis. R_{trans} is a transitional conformation between R and R_{sym}, observed transiently in AMBER simulations. S is the conformation observed in the signaling state.

It is interesting to compare the classification by F124 dihedral angles with the classification by the whole domain backbone conformation (value of PC1 projection), as both numbers can be determined for each trajectory snapshot. The analysis shows that the correspondence is clear and unambiguous (Figure 3.6.6d). We use this fact to determine the structure of the resting state without averaging over R and R_{sym}.

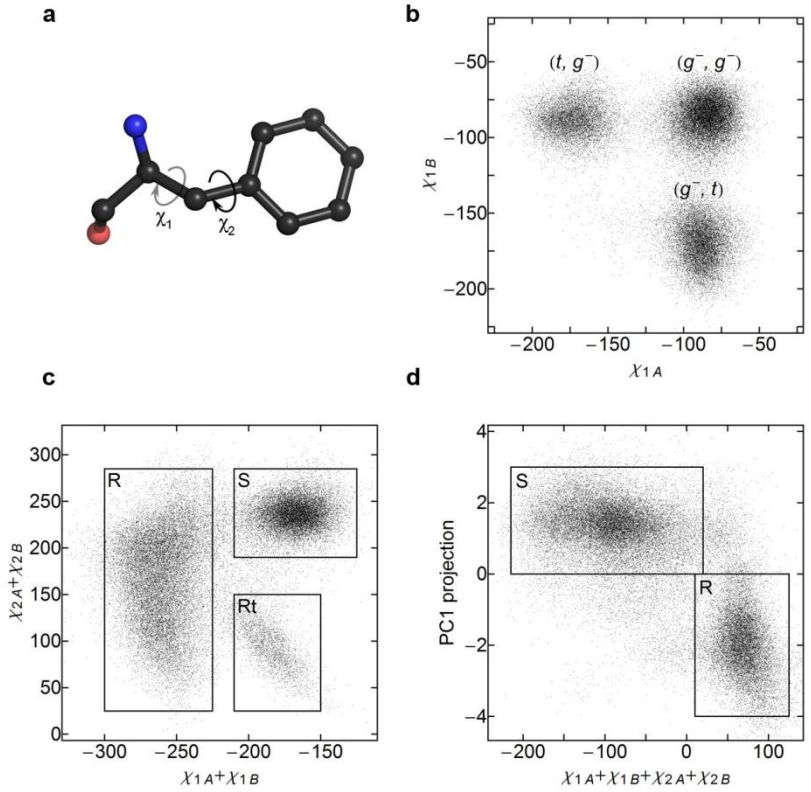

Figure 3.6.6. Use of F124 χ_1 and χ_2 dihedral angles to classify the HAMP domain state. Subscripts A and B denote the protomers of the dimeric HAMP domain. a) Definition of the χ_1 and χ_2 dihedral angles of phenylalanine sidechain. b) Distribution of the χ_1 angles during the simulations. *t* stands for the *trans* conformation, and g^- for the *gauche*$^-$ conformation. (*t*, *t*) conformation is not observed. c) Plot of the sum of the χ_2 angles as a function of the sum of the χ_1 angles. There are two substates corresponding to χ_1 (g^-, g^-) conformation (χ_1 sum of ~-170º). The state with the χ_2 sum of ~240º corresponds to the signaling state, meanwhile the state with the sum of ~100º is transient and visited during transitions from the χ_1 (*t*, g^-) conformation to (g^-, *t*). d) Relation between the F124 conformations, represented by the sum of side chain dihedral angles, and the whole domain backbone conformation, represented by the value of the projection on the first principal component. Unambiguous correspondence may be established between these two values.

3.6.5 Comparison of the NpHtrII HAMP1 resting and signaling states.

The most significant difference between the resting and the signaling states of the NpHtrII HAMP1 domain is the longitudinal shift of the helices AS1 and AS1' relative to the helices AS2 and AS2' (Figure 3.6.7) upon the transition. In the signaling state, the HAMP domain backbone and its hydrophobic core are symmetric. In the resting state, the symmetry of both the backbone and the core breaks down. Position of the helices AS2 remains roughly symmetric, meanwhile AS1 and AS1' shift longitudinally, but for a different distance. Following the coiled-coil terminology, F124 of the helix AS2 forms a „knob" in the „hole" made of the AS1 residues L93, A97 and M100 of the same protomer, and thus its conformation is linked to the relative shift of AS1 and AS2 (Figure 3.6.7).

Interestingly, the residue F124 resides in the same coiled coil layer as the residue A291 of the Af1503 HAMP domain. Mutations of A291 to bulkier hydrophobic aminoacids render the domain, which is natively dysfunctional, able to conduct the signal in Taz (Tar-HAMP-EnvZ) chimeras[161] as well as other systems[148,162]. Taken together with the results of our simulations, these data show the role of large aminoacid side chains at the HAMP domain core.

3.6.6 Conclusions

In this chapter, the results of unbiased molecular dynamics simulations of the NpHtrII HAMP1 domain have been presented. The domain adopts two conformations that share many features with the two HAMP domain states observed experimentally. To the best of our knowledge, this is the first example where the atomic structures of two different conformations of the same HAMP domain are presented.

Although the simulations of the HAMP domain without the flanking modules may be misleading[4], and the structure of its N- and C-termini could differ if other NpHtrII domains were included in the simulation, comparison of the results with experimental data is in favor of biological relevance of the two observed states. At the moment, it is not clear how a transition between these states could result in a signal transduction via the HAMP domain. The structures suggest that the possible mechanisms are the overall elongation of the HAMP1 or twisting of its helices, resulting in rotation of the output modules relative to the input modules around the dimer axis.

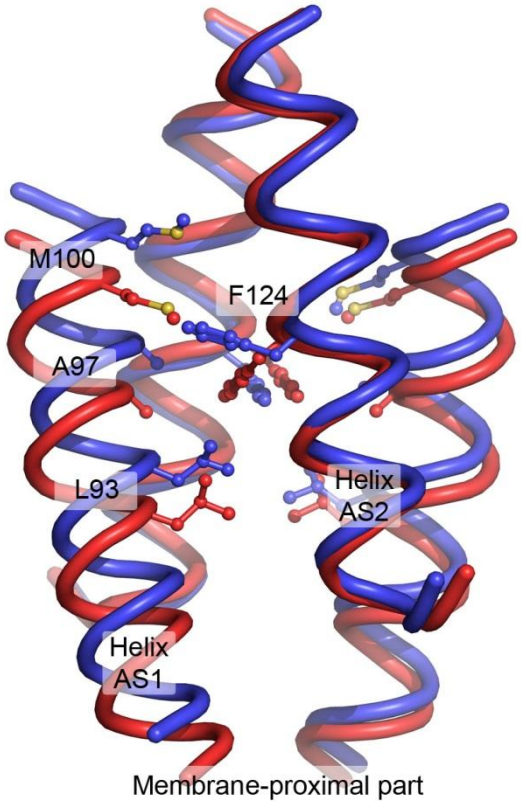

Figure 3.6.7. Comparison of the resting (blue) and the signaling (red) states of the NpHtrII HAMP1 domain obtained in molecular dynamics simulations. Conformation of the F124 pair is closely related to the conformation of the whole HAMP domain. The averaged structures from the CHARMM simulation are shown. For the resting state, only the structures with similar F124 position were chosen and not the symmetrically related ones, that is only (t, g^-) and not (g^-, t). The most notable difference between the observed states is a longitudinal displacement of the AS1 helices relative to the AS2 helices. Following the coiled-coil terminology, F124 from the helix AS2 forms a „knob“ in the „hole“ made of the AS1 residues L93, A97 and M100 of the same protomer. These residues are shown in sticks representation. As a consequence of the interaction, position of AS1 relative to AS2 is closely linked to the F124 rotameric state. Also, in the signaling state, F124 bulges less and the helices are more mobile. The averaged structure of the signaling state is symmetrical. The resting state is inherently asymmetric due to F124, however the backbone atoms' positions in R and Rsym states are almost identical in the AMBER simulations, and are similar in the CHARMM simulations.

3.7 Crystal structure and modeling of the *Archaeoglobus fulgidus* IPCT-DIPPS bifunctional enzyme

3.7.1 Crystallographic data

Structure of the *A.f.* IPCT-DIPPS bifunctional enzyme was solved by molecular replacement using the coordinates of the IPCT domain[186], and refined to a resolution of 2.65 Å. The data statistics are presented in the Table 3.7.1.

Table 3.7.1. IPCT-DIPPS diffraction data collection and refinement statistics.

Data collection	
Space group	$P2_12_12$
Cell dimensions	
a, *b*, *c* (Å)	41.369, 107.584, 123.953
α, *β*, *γ* (°)	90, 90, 90
Resolution (Å)	61.82-2.65 (2.79-2.65)
CC^{*2}	0.991 (0.513)
$I/\sigma I$	7.7 (1.7)
Completeness (%)	99.9 (99.9)
Redundancy	7.2 (5.8)
Refinement	
Resolution (Å)	53.85-2.65
No. reflections	15952
R_{work} / R_{free}	23.6% / 31.1%
Number of atoms	
Protein	3115
Mg^{2+}	1
Lipid fragments	38
Water	8
Average B-factor (Å^2)	
Protein	59.6
Mg^{2+}	65.8
Lipid fragments	73.3
Water	62.1
R.m.s deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.6

The electron density maps are of good quality (Figure 3.7.1a). Most of the side chains in the transmembrane regions were assigned, however, many exposed side chains of the soluble domain and disordered. The reason for this could be the lesser overall ordering of the soluble IPCT domain as compared to the membrane DIPPS one (average B-factors of IPCT and DIPPS backbone atoms are 63.3 and 45.8 $Å^2$, respectively; Figure 3.7.1b).

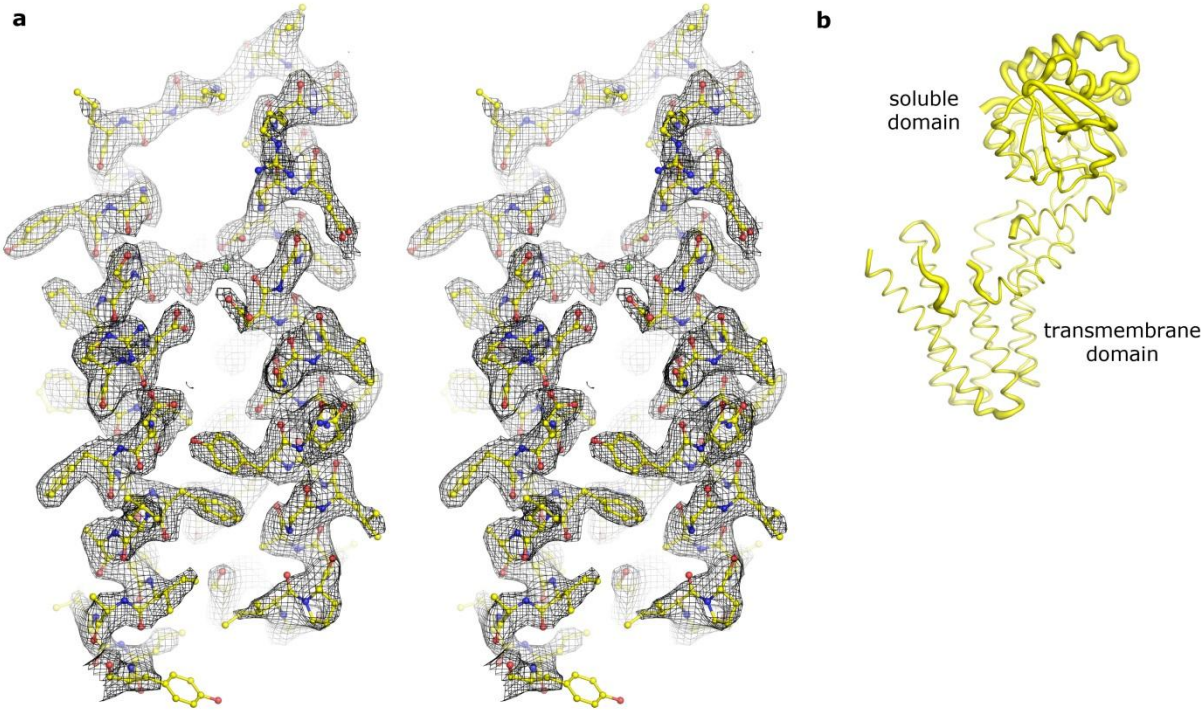

Figure 3.7.1 Sample of the electron density and distribution of the B-factors. a) Stereoscopic image of the electron density (contoured at the level of 1.5σ) around the transmembrane helices 2 and 3. b) Relative distribution of the crystallographic B-factors in the IPCT-DIPPS model. Larger diameter of the tube corresponds to higher B-factors. The soluble domain is less ordered than the transmembrane one as the average B-factor for backbone atoms is 63.3 $Å^2$ for IPCT and 45.8 $Å^2$ for DIPPS, with value ranges 31.7-120.8 and 23.0-99.9, respectively.

As it is typical for *in meso* crystallization, the proteins in the crystal are organized as a stack of membrane layers (type I membrane protein crystals[266]), with interlayer planar contacts relying entirely on the IPCT cytoplasmic domain (Figure 3.7.2).

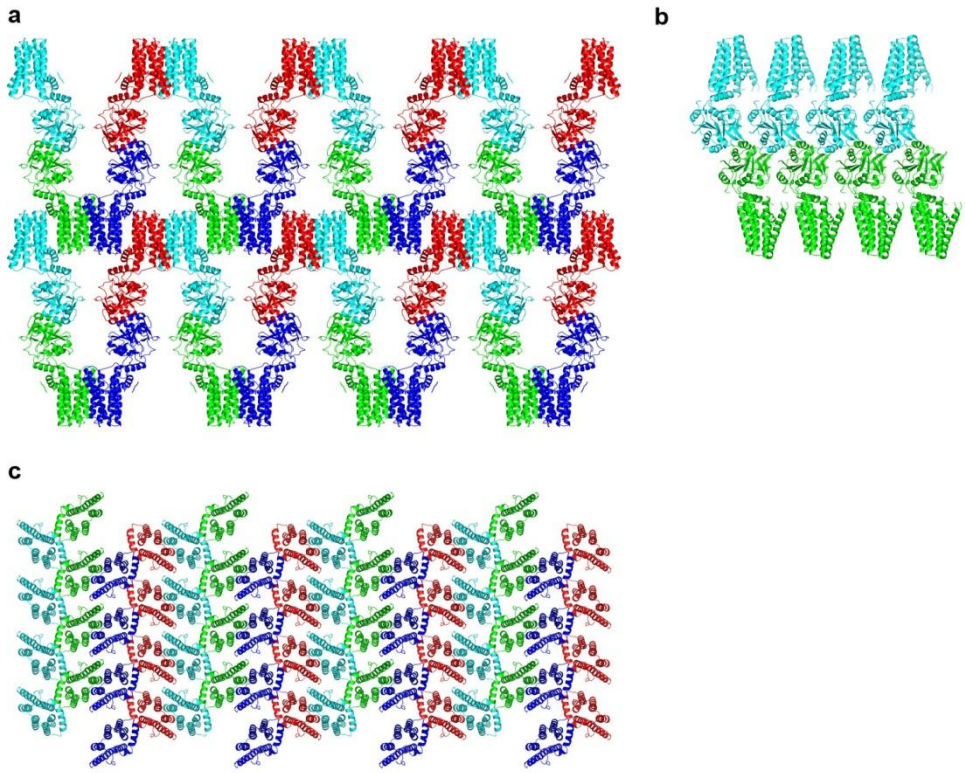

Figure 3.7.2. Crystal packing of IPCT-DIPPS. There are four molecules in the crystallographic unit cell, colored green, light blue, red and dark blue. a) The view along the *a* unit cell dimension (along the membrane plane). b) The view along the *b* unit cell dimension (along the membrane plane). c) The view along the *c* unit cell dimension (perpendicular to the membrane plane).

3.7.2 Description of the *A.f.* IPCT-DIPPS structure

Crystallographic structure reveals the conformations of both the cytoplasmic IPCT domain and transmembrane DIPPS domain. Between them are two amphipathic ("connector") helices that are oriented roughly along the membrane plane. The protein is dimerized via the transmembrane part (Figure 3.7.3).

There are six transmembrane helices in the DIPPS domain (TM1-TM6), connected by two intracellular and three extracellular loops. The loop between the connector helix 2 and TM1 is disordered (residues 292-294 are not observed in the electron density maps), as well as the elongated loop between the helices TM4 and TM5 (residues 430-433 are not observed).

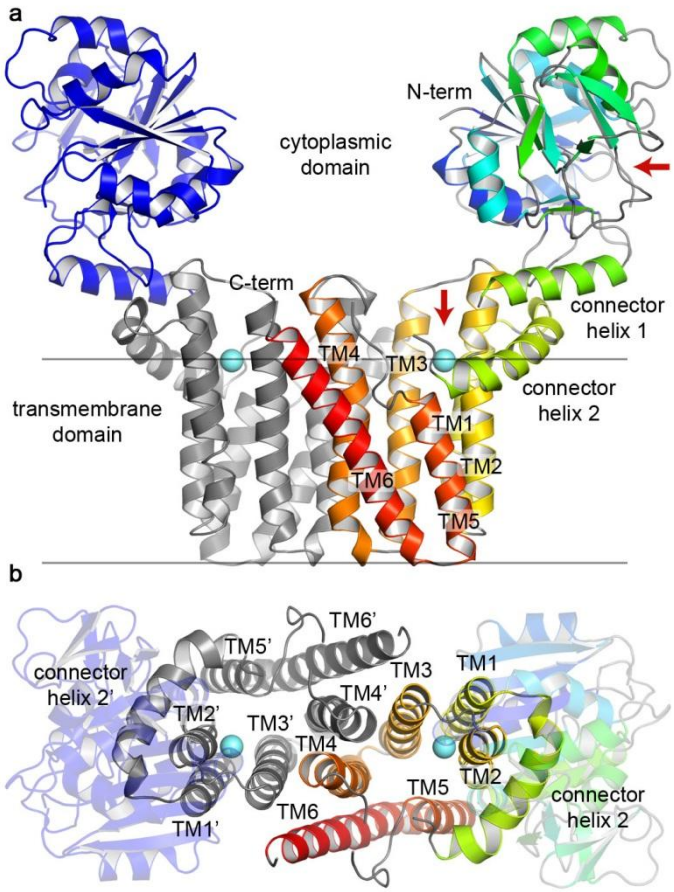

Figure 3.7.3 Overall structure of the IPCT-DIPPS bifunctional enzyme. The protein dimerizes via the transmembrane domain. a) View along the membrane plane. b) View perpendicular to the membrane plane. One protomer is rainbow colored and the other protomer (related by crystallographic symmetry) is in blue for cytoplasmic IPCT domain and grey for membrane DIPPS domain. Red arrows indicate the access to their active sites and putative Mg^{2+} ion is drawn as a cyan sphere. Membrane surface position (grey) was calculated using the PPM server[17].

The DIPPS dimerization interface is symmetric and is formed by the helix TM4 of one protomer protruding between helices TM3' and TM4' of the other protomer and helix TM6 flanking the helix TM3' (Figures 3.7.3b and 3.7.4). The interface is stabilized mostly by hydrophobic contacts and several hydrogen bonds between polar or ionizable residues located at the cytoplasmic side. The total interaction surface area is 1940 $Å^2$.

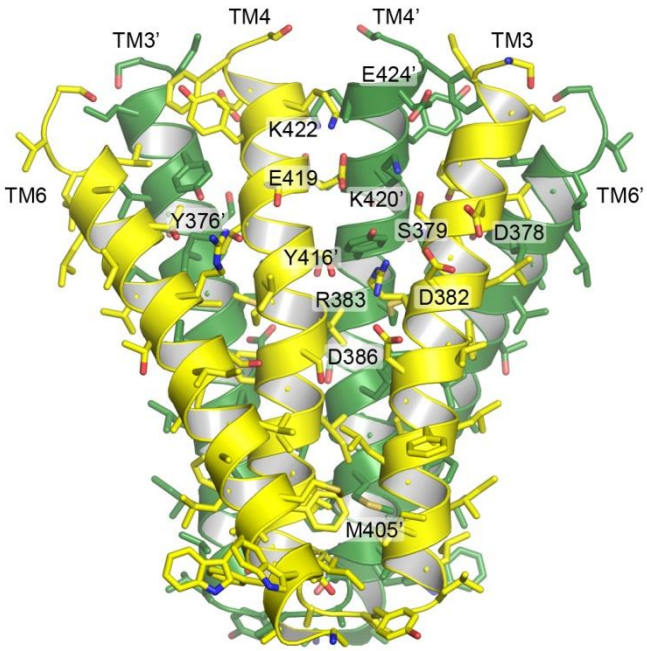

Figure 3.7.4 DIPPS dimerization interface. The residues Y376, M405, Y416, K420 and E424 of each protomer protrude into the opposite protomer.

While the DIPPS domain interacts extensively with the connector helix 2, the connector helix 1 is mostly in contact with the IPCT domain, facing its short proximal hydrophobic β-sheet. The polar and positively charged residues of the connector helix 1 point towards the connector helix 2 and the cytoplasmic side of the membrane. The connector helix 1 can be ascribed to IPCT and the connector helix 2 to DIPPS, based on their distribution in the organisms where IPCT and DIPPS do not belong to the same polypeptide chain (for example, *Thermotoga maritima*). It should also be noted that, although the connector helix 2 is conserved among the DIPPS proteins, it is absent in many other CDP-alcohol phosphatidyltransferases (Figure 3.7.10).

The cytoplasmic IPCT domain binds at the edge of the DIPPS domain (Figure 3.7.3). It is formed of seven-stranded mixed β-sheet flanked by six α-helices. This arrangement is reminiscent of a dinucleotide-binding Rossmann fold[267], typical of nucleotidyltransferases. In addition, there is a sugar-binding region consisting of a two-stranded β-sheet and a 30-residue long stretch. The overall fold of the IPCT domain is very similar to the structure of that domain alone, which was reported previously by Brito *et al.*[186] (root mean square deviation of 0.6 Å for Cα atoms, Figure 3.7.5).

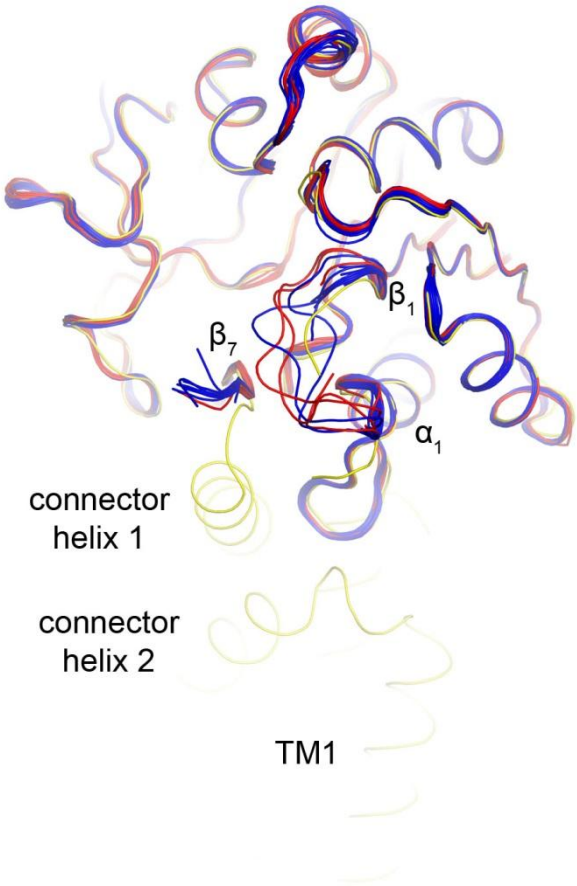

Figure 3.7.5. Structural comparison of the IPCT domain in the bifunctional enzyme IPCT-DIPPS with the truncated sole domain. The IPCT-DIPPS protein is shown in yellow, 12 NCS-related structures from the PDB ID 2XME are shown in blue, 6 NCS-related structures from the PDB ID 2XMH are shown in red. The only notable deviation is in the structure of the very flexible loop between β-strand β_1 and α-helix α_1 (in the front).

Interestingly, the IPCT domains were dimerized in the crystals of the truncated one-domain construct[186] via the same interface that connects the IPCT and DIPPS domains in the full-length protein (Figure 3.7.6a,b). At the same time, the IPCT active site is turned away from DIPPS (Figure 3.7.3). One could imagine that such organization of a bifunctional enzyme that catalyzes two consecutive reactions is ineffective, and thus, it is possible that during the catalysis the IPCT domain turns relative to DIPPS so that its active site faces the DIPPS active site. The modeling shows that if the soluble domains were to dimerize at the same time as the transmembrane domain is dimerized, the active sites of both domains would come much closer to each other (Figure 3.7.6c).

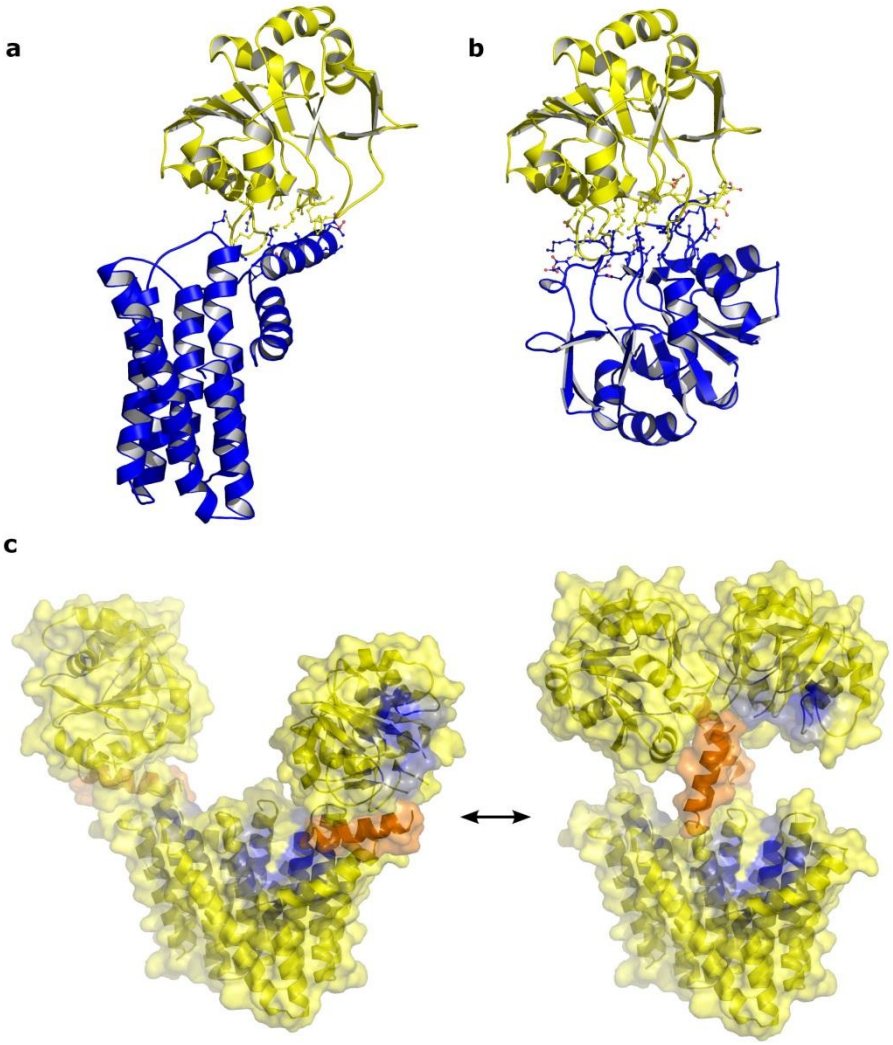

Figure 3.7.6. Interactions of the IPCT domain and a model with both IPCT and DIPPS domains dimerized. a) Interaction of IPCT domain (yellow) with the transmembrane DIPPS domain (blue) in the full-length protein, interface surface area ~460 $Å^2$. b) A similar interface is involved in symmetric homodimerization in the crystal structure of the truncated soluble IPCT, interface surface area ~1070 $Å^2$. The IPCT domain is shown in the same orientation in a) and b). c) Modeling of the IPCT-DIPPS conformational change which would result in dimerization of the soluble IPCT domains while preserving the membrane DIPPS domains dimerization. The two ligand-binding sites of each protomer are highlighted in blue. The ligand-binding sites of the soluble domain (IPCT activity) and the transmembrane domain (DIPPS activity) are oriented in opposite directions in the crystallographic structure (left). This could be unfavorable for the efficiency of the enzyme. Dimerization of the soluble domains via the interface observed in the crystals of the sole soluble domains (right) would bring the ligand-binding sites close to each other. The described conformational change would require concomitant motion of the helix 1 (shown in orange). The dimer is symmetric in the model.

3.7.3 DIPPS active site and ligand docking

Initial examination of the DIPPS structure revealed an interesting feature in between the transmembrane helices TM2 and TM3. The side chains of the aspartates 357, 360, 378 and 382 (all are the residues extremely conserved among CDP-OH transferases, Figure 1.3.2), and the backbone oxygen of Asp-357 all pointed towards the same spot, where a significant unaccounted electron density was observed (Figure 3.7.7a). Such configuration could be explained by presence of a divalent cation in that position. Given that the Mg^{2+} was the only divalent cation present in the crystallization conditions, and it was known that magnesium is required for the catalytic action of DIPPS, the density was assigned to the bound Mg^{2+} ion. The DIPPS activity assays conducted with Mg, Zn, Co, Ni and Ca cations (Nogly *et al.*, Nature Communications, accepted) have corroborated this assignment. Consequently, as magnesium is known to participate in the reactions involving phosphorus moieties, we conclude that the Mg^{2+} ion and the aspartates 357, 360, 378 and 382, belonging to the CDP-OH transferase domain consensus motif, should be a part of the DIPPS catalytic site.

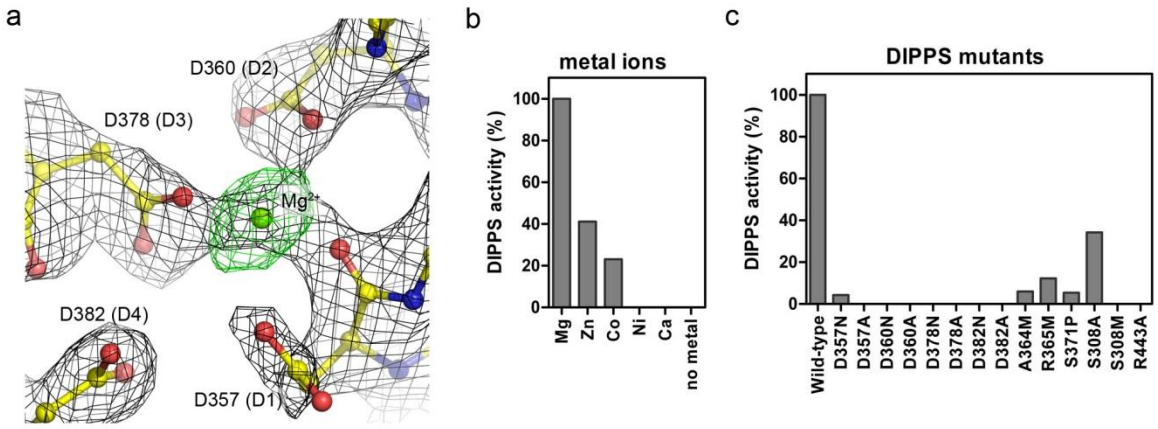

Figure 3.7.7 DIPPS active site and DIPPS activity assays. a) Zoomed view of the electron density map around the conserved aspartate residues. The maps were calculated before insertion of the magnesium ion in the model. The $2F_o$-F_c density is shown in black and contoured at 1.5σ, the F_o-F_c density is shown in green and contoured at 3σ. b, c) Diagram of DIPPS activity in 30 min reactions to assess the effect of divalent metal ions at 10 mM concentration (b) and point mutations (c), where 100% refers to the activity of wild type enzyme in the presence of 10 mM Mg^{2+} (Nogly *et al.*, Nature Communications, accepted).

As a next step, it was interesting to elucidate the binding sites of the DIPPS ligands. Analysis of the cytoplasmic side of DIPPS has revealed three cavities (Figure 3.7.8a), which could potentially bind the ligands, inositol-1-phosphate and CDP-inositol (Figure 1.3.3).

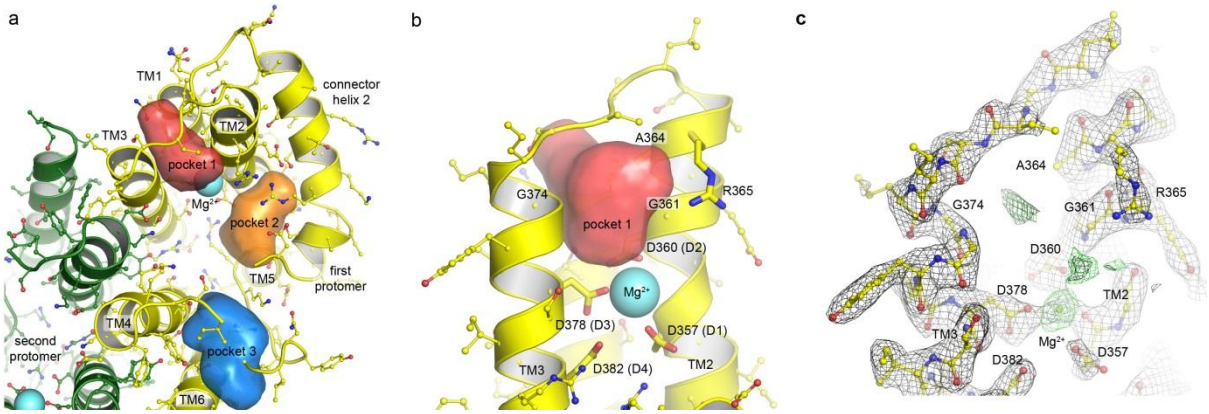

Figure 3.7.8. DIPPS active site. a) Putative ligand-binding pockets on the DIPPS cytoplasmic site. The accessible cavities were calculated using HOLLOW[235]. The pockets are shown in pink (1), orange (2) and blue (3). b) Detailed view of conserved pocket 1. The strictly conserved residues G361, A364 and G374, belonging to the consensus sequence, line the surface of the pocket. c) Unaccounted $2F_o$-F_c and F_o-F_c electron densities inside and near the TM2-TM3 loop. One blob is inside the TM2-TM3 loop and the other is in between the consensus residue R365 and the catalytic magnesium ion.

Pocket 1 is the most interesting pocket, as it is flanked by the loop between TM2 and TM3 and lined by the family consensus residues G361, A364 and G374 (Figure 3.7.8b). Although IPCT/DIPPS protein was crystallized in absence of its substrates, residual $2F_o$-F_c and F_o-F_c electron densities are seen in this region that cannot be accounted for by atoms of the protein (Figure 3.7.8c). Pocket 2 is at the bottom of the hydrophilic cavity formed by strictly conserved residues among DIPPS proteins, namely N304, R305, S308, S354, R443 and D357, situated on the connector helix 2, TM2 and TM5. Pocket 3 is surrounded by helices TM4, TM5, TM6, and loop TM5-TM6. Amino acid residues from this loop are not conserved and the pocket itself is far from the catalytic magnesium ion. Therefore, substrate binding within pocket 3 is unlikely. Since pocket 1 comprises residues from the family consensus region, it can be concluded that it should be the binding site of the cytidine moiety, a common feature among substrates of CDP-OH-phosphotransferases. In contrast, pocket 2 includes connector helix 2 and its sequence conservation is restricted to enzymes assumed to have DIPPS activity, which suggests that it is the binding site of inositol-1-phosphate.

Following these rationales, a ligand docking study, aimed at elucidating the positions of the ligands, was performed. Various ligand binding poses were generated using the Autodock Vina software[216] in the vicinity of the identified pockets. As the Autodock scoring function (binding energy estimate) is not accurate enough to determine the correct binding pose, the final positions were chosen manually. Results of the docking are presented in Figure 3.7.9.

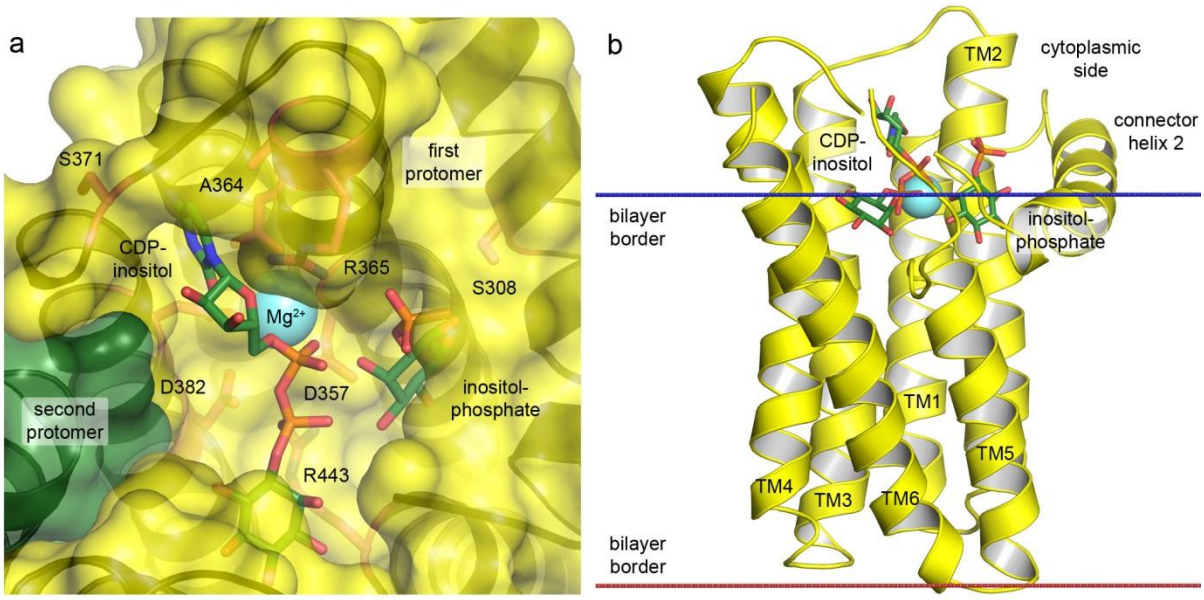

Figure 3.7.9 Model of the inositol-phosphate and CDP-inositol binding to DIPPS domain. a) Inositol-phosphate and CDP-inositol docked into the DIPPS surface cavities. The reacting moieties are conductive to the catalyzed reaction (Figure 1.3.3) and are close to the catalytic magnesium ion. b) Position of the DIPPS substrates relative to the bilayer surface. The CDP moiety is above the bilayer surface, and the inositol group of CDP-inositol and inositol-phosphate are slightly underneath it. The figure shows that the lipophilic chains of diacylglycerol substrate of other CDP-alcohol phosphotransferases can be embedded in the membrane lipidic bilayer. The membrane position was calculated using the PPM server[17].

Results of the ligand binding site assignment have also been tested by point mutagenesis (Figure 3.7.7c) and it was shown that the mutations of the proposed binding site residues, such as S308 and S371 (Figure 3.7.9a) detriments the catalysis.

Finally, the structure of the DIPPS domain could be used for understanding other relevant members of the CDP-alcohol transferase family. Although the sequence conservation is very low, it is still possible to perform a rough assignment of the secondary structures of the corresponding enzymes and align their sequences with that of the *Archaeoglobus fulgidus* IPCT-DIPPS (Figure 3.7.10).

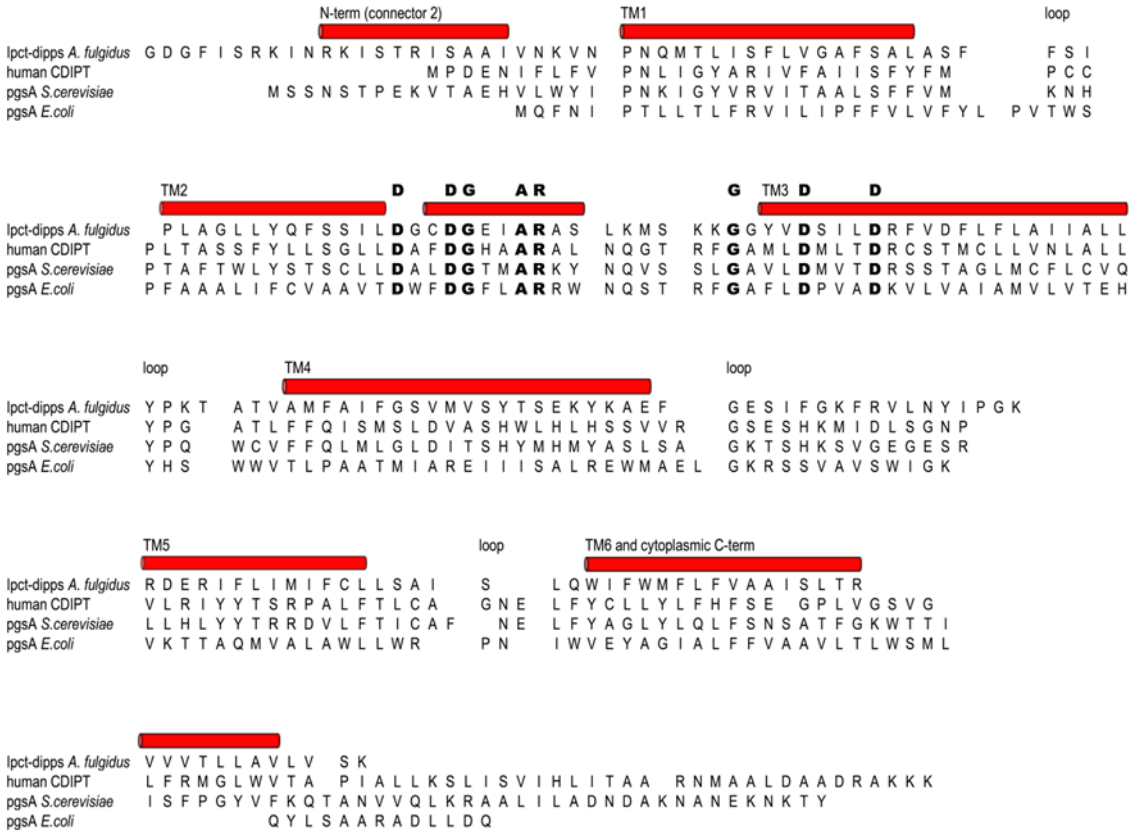

Figure 3.7.10. Manual secondary-structure alignment of the *Archaeoglobus fulgidus* IPCT-DIPPS sequence with the exemplary genes from bacteria, yeast, and human. Shown are the residues 295-490 of IPCT-DIPPS (corresponding to the DIPPS domain), and full-length sequences of human phosphatidylinositol synthase (UniProt reference O14735), *Saccharomyces cerevisiae* phosphatidylinositol synthase (UniProt reference P06197) and *Escherichia coli* phosphatidylglycerophosphate synthase (UniProt reference P0ABF8). The transmembrane helical segments were identified based on the hydropathicity, and the loop segments were identified based on the presence of amino acids with low α-helix formation propensity (proline, glycine, asparagine). The alignment is best in the consensus sequence residues region (TM2-TM3) and especially difficult at the position of TM6.

Future work

In this work, results of the structural studies of several proteins have been presented. However, the work is not over as the structural models are only a basis for future research. As for the microbial rhodopsins, structures of the intermediate states are needed for complete understanding of the mechanism of function, and physiological experiments are needed to elucidate the roles of the discovered structural elements. Also, it would be extremely interesting to understand the principles of microbial rhodopsin design and to craft the proteins with novel functions, such as selective light-driven potassium pumps. As for the chemotaxis and phototaxis systems, having now atomic models of all the domains, one can turn to elucidating the mechanism of signal transduction at the level of a whole protein. Finally, the structure of the IPCT-DIPPS enzyme can now be used to study the functioning of human proteins, such as cardiolipin synthase, an enzyme that catalyzes the last step of the biosynthesis of this complex and beautiful lipid cardiolipin.

Acknowledgements

First of all, I am extremely grateful to my parents, who did everything for me and helped me to pursue my interests in physics and science in general.

Then, I am very thankful to Valentin Gordeliy, who was an excellent supervisor during all these years, provided a lot of help and allowed a great freedom in my scientific projects. I am also thankful to Sergei Grudinin, who was my co-supervisor and taught me everything about modeling biological systems, to Andrii Ishchenko, who has taught me both working in the laboratory and solving protein structures, and to Valentin Borshchevskiy, who refined my crystallography skills and provided great advice about everything. Finally, Ilya Manukhov and Taras Balandin taught me all the tricks of genetic engineering and molecular biology.

One of my warmest thanks goes to Vitaly Polovinkin, who was my great friend and colleague throughout these years. I am also grateful for the help and support to my colleagues from IBS, Ekaterina Round, Pavel Chervakov, Alina Remeeva, Petr Utrobin, Evelina Edelweiss, Georgiy Derevyanko and Florine Dupeux, and to Linda Ponnet, who expertly handled all the administrative matters. I am thankful to Vitaly Shevchenko, who produced many of the proteins presented in this work and with whom we had many interesting discussions, and to other colleagues from the Forschungszentrum Jülich, Chris Baeken, Maria Silacheva, Dmitry Bratanov and Alexander Volkov. I also wish to say thanks to Alexander Popov of the European Synchrotron Radiation Facility, who was very patient during our plentiful experiments, and to my colleagues from MIPT, Alexander Kuklin, Alexey Mishin, Mikhail Sintsov, Kirill Kovalev and Pavel Buslaev!

All of the diffraction experiments presented in this thesis were performed at the European Synchrotron Radiation Facility, and the simulations were done using the JUROPA supercomputer from Forschungszentrum Jülich. Some of the projects presented in this work were conducted in collaboration, and I am grateful to Pavel Kuzmichev, Lada Petrovskaya, and Vladimir Chupin from the Institute of Bioorganic Chemistry, and to Przemyslaw Nogly, Helena Santos and Margarida Archer from the Instituto de Tecnologia Quimica e Biologica.

Last but not least, I wish to say thanks to the reviewers of my work and members of the jury, professors Georg Büldt, Dieter Willbold, Martin Engelhard and Eva Pebay-Peyroula, who contributed greatly to the studies of membrane proteins and especially microbial rhodopsins.

References

1. Gushchin, I. *et al.* Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. *Proc. Natl. Acad. Sci.* **110,** 12631–12636 (2013).
2. Gushchin, I., Gordeliy, V. & Grudinin, S. Two Distinct States of the HAMP Domain from Sensory Rhodopsin Transducer Observed in Unbiased Molecular Dynamics Simulations. *PLoS ONE* **8,** e66917 (2013).
3. Gushchin, I. *et al.* Active State of Sensory Rhodopsin II: Structural Determinants for Signal Transfer and Proton Pumping. *J. Mol. Biol.* **412,** 591–600 (2011).
4. Gushchin, I., Gordeliy, V. I. & Grudinin, S. Role of the HAMP Domain Region of Sensory Rhodopsin Transducers in Signal Transduction. *Biochemistry (Mosc.)* **50,** 574–580 (2011).
5. Grote, M., Engelhard, M. & Hegemann, P. Of ion pumps, sensors and channels — Perspectives on microbial rhodopsins between science and history. *Biochim. Biophys. Acta BBA - Bioenerg.* **1837,** 533–545 (2014).
6. Lanyi, J. K. Bacteriorhodopsin. *Annu. Rev. Physiol.* **66,** 665–688 (2004).
7. Oesterhelt, D. & Stoeckenius, W. Rhodopsin-like Protein from the Purple Membrane of Halobacterium halobium. *Nature* **233,** 149–152 (1971).
8. Racker, E. & Stoeckenius, W. Reconstitution of Purple Membrane Vesicles Catalyzing Light-driven Proton Uptake and Adenosine Triphosphate Formation. *J. Biol. Chem.* **249,** 662–663 (1974).
9. Henderson, R. & Unwin, P. N. T. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257,** 28–32 (1975).
10. Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. Electron-crystallographic Refinement of the Structure of Bacteriorhodopsin. *J. Mol. Biol.* **259,** 393–421 (1996).
11. Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213,** 899–929 (1990).
12. Dencher, N. & Wilms, M. Flash photometric experiments on the photochemical cycle of bacteriorhodopsin. *Biophys. Struct. Mech.* **1,** 259–271 (1975).
13. Lozier, R. H., Bogomolni, R. A. & Stoeckenius, W. Bacteriorhodopsin: a light-driven proton pump in Halobacterium Halobium. *Biophys. J.* **15,** 955–962 (1975).
14. Bagley, K., Dollinger, G., Eisenstein, L., Singh, A. K. & Zimányi, L. Fourier transform infrared difference spectroscopy of bacteriorhodopsin and its photoproducts. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 4972–6 (1982).
15. Braiman, M. & Mathies, R. Resonance Raman spectra of bacteriorhodopsin's primary photoproduct: evidence for a distorted 13-cis retinal chromophore. *Proc. Natl. Acad. Sci.* **79,** 403–407 (1982).
16. Luecke, H., Schobert, B., Richter, H.-T., Cartailler, J.-P. & Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291,** 899–911 (1999).
17. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40,** D370–D376 (2012).

18. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3 Å resolution. *Nature* **318,** 618–624 (1985).

19. Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *Proc. Natl. Acad. Sci.* **93,** 14532–14535 (1996).

20. Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. X-ray Structure of Bacteriorhodopsin at 2.5 Angstroms from Microcrystals Grown in Lipidic Cubic Phases. *Science* **277,** 1676–1681 (1997).

21. Faham, S. & Bowie, J. U. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol.* **316,** 1–6 (2002).

22. Takeda, K. *et al.* A novel three-dimensional crystal of bacteriorhodopsin obtained by successive fusion of the vesicular assemblies. *J. Mol. Biol.* **283,** 463–474 (1998).

23. Cherezov, V., Peddi, A., Muthusubramaniam, L., Zheng, Y. F. & Caffrey, M. A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 1795–1807 (2004).

24. Cherezov, V. *et al.* High-Resolution Crystal Structure of an Engineered Human β2-Adrenergic G Protein–Coupled Receptor. *Science* **318,** 1258–1265 (2007).

25. Rasmussen, S. G. F. *et al.* Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* **450,** 383–387 (2007).

26. Efremov, R. G., Baradaran, R. & Sazanov, L. A. The architecture of respiratory complex I. *Nature* **465,** 441–445 (2010).

27. Yoshimura, K. & Kouyama, T. Structural Role of Bacterioruberin in the Trimeric Structure of Archaerhodopsin-2. *J. Mol. Biol.* **375,** 1267–1281 (2008).

28. Birge, R. R. *et al.* Biomolecular Electronics: Protein-Based Associative Processors and Volumetric Memories. *J. Phys. Chem. B* **103,** 10746–10766 (1999).

29. Engelhard, M. *et al.* Magic angle sample spinning carbon-13 nuclear magnetic resonance of isotopically labeled bacteriorhodopsin. *Biochemistry (Mosc.)* **28,** 3967–3975 (1989).

30. Sass, H. j. *et al.* The tertiary structural changes in bacteriorhodopsin occur between M states: X-ray diffraction and Fourier transform infrared spectroscopy. *EMBO J.* **16,** 1484–1491 (1997).

31. Dencher, N. A., Dresselhaus, D., Zaccai, G. & Büldt, G. Structural changes in bacteriorhodopsin during proton translocation revealed by neutron diffraction. *Proc. Natl. Acad. Sci.* **86,** 7876–7879 (1989).

32. Altenbach, C., Marti, T., Khorana, H. G. & Hubbell, W. L. Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. *Science* **248,** 1088–1092 (1990).

33. Butt, H. J., Downing, K. H. & Hansma, P. K. Imaging the membrane protein bacteriorhodopsin with the atomic force microscope. *Biophys. J.* **58,** 1473–1480 (1990).

34. Shibata, M., Yamashita, H., Uchihashi, T., Kandori, H. & Ando, T. High-speed atomic force microscopy shows dynamic molecular processes in photoactivated bacteriorhodopsin. *Nat. Nanotechnol.* **5,** 208–212 (2010).

35. Oesterhelt, F. *et al.* Unfolding Pathways of Individual Bacteriorhodopsins. *Science* **288,** 143–146 (2000).

36. Grudinin, S., Büldt, G., Gordeliy, V. & Baumgaertner, A. Water Molecules and Hydrogen-Bonded Networks in Bacteriorhodopsin—Molecular Dynamics Simulations of the Ground State and the M-Intermediate. *Biophys. J.* **88,** 3252–3261 (2005).

37. Wang, T. *et al.* Deprotonation of D96 in Bacteriorhodopsin Opens the Proton Uptake Pathway. *Structure* **21,** 290–297 (2013).

38. Hirai, T., Subramaniam, S. & Lanyi, J. K. Structural snapshots of conformational changes in a seven-helix membrane protein: lessons from bacteriorhodopsin. *Curr. Opin. Struct. Biol.* **19,** 433–439 (2009).

39. Borshchevskiy, V., Efremov, R., Moiseeva, E., Büldt, G. & Gordeliy, V. Overcoming merohedral twinning in crystals of bacteriorhodopsin grown in lipidic mesophase. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 26–32 (2010).

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

41. Hirai, T. & Subramaniam, S. Protein Conformational Changes in the Bacteriorhodopsin Photocycle: Comparison of Findings from Electron and X-Ray Crystallographic Analyses. *PLoS ONE* **4,** e5769 (2009).

42. Matsuno-Yagi, A. & Mukohata, Y. Two possible roles of bacteriorhodopsin; a comparative study of strains of Halobacterium halobium differing in pigmentation. *Biochem. Biophys. Res. Commun.* **78,** 237–243 (1977).

43. Schobert, B. & Lanyi, J. K. Halorhodopsin is a light-driven chloride pump. *J. Biol. Chem.* **257,** 10306–10313 (1982).

44. Hildebrand, E. & Dencher, N. Two photosystems controlling behavioural responses of Halobacterium halobium. *Nature* **257,** 46–48 (1975).

45. Spudich, E. N. & Spudich, J. L. Control of transmembrane ion fluxes to select halorhodopsin-deficient and other energy-transduction mutants of Halobacterium halobium. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 4308–4312 (1982).

46. Spudich, J. L. & Bogomolni, R. A. Mechanism of colour discrimination by a bacterial sensory rhodopsin. *Nature* **312,** 509–513 (1984).

47. Bogomolni, R. A. & Spudich, J. L. Identification of a third rhodopsin-like pigment in phototactic Halobacterium halobium. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 6250–6254 (1982).

48. Wolff, E. K., Bogomolni, R. A., Scherrer, P., Hess, B. & Stoeckenius, W. Color discrimination in halobacteria: spectroscopic characterization of a second sensory receptor covering the blue-green region of the spectrum. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 7272–7276 (1986).

49. Tomioka, H., Takahashi, T., Kamo, N. & Kobatake, Y. Flash spectrophotometric identification of a fourth rhodopsin-like pigment in Halobacterium halobium. *Biochem. Biophys. Res. Commun.* **139,** 389–395 (1986).

50. Marwan, W. & Oesterhelt, D. Signal formation in the halobacterial photophobic response mediated by a fourth retinal protein (P480). *J. Mol. Biol.* **195,** 333–342 (1987).

51. Takahashi, T., Tomioka, H., Kamo, N. & Kobatake, Y. A photosystem other than PS370 also mediates the negative phototaxis of< i> Halobacterium halobium</i>. *FEMS Microbiol. Lett.* **28,** 161–164 (1985).

52. Bivin, D. B. & Stoeckenius, W. Photoactive retinal pigments in haloalkaliphilic bacteria. *J. Gen. Microbiol.* **132,** 2167–2177 (1986).

53. Béjà, O. *et al.* Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea. *Science* **289,** 1902–1906 (2000).

54. Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411,** 786–789 (2001).

55. Torre, J. R. de la *et al.* Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc. Natl. Acad. Sci.* **100,** 12830–12835 (2003).

56. Bamann, C., Bamberg, E., Wachtveitl, J. & Glaubitz, C. Proteorhodopsin. *Biochim. Biophys. Acta BBA - Bioenerg.* **1837,** 614–625 (2014).

57. Petrovskaya, L. E. *et al.* Predicted bacteriorhodopsin from Exiguobacterium sibiricum is a functional proton pump. *FEBS Lett.* **584,** 4193–4196 (2010).

58. Inoue, K. *et al.* A light-driven sodium ion pump in marine bacteria. *Nat. Commun.* **4,** 1678 (2013).

59. Yoshizawa, S. *et al.* Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proc. Natl. Acad. Sci.* 201403051 (2014). doi:10.1073/pnas.1403051111

60. Hegemann, P. Algal sensory photoreceptors. *Annu. Rev. Plant Biol.* **59,** 167–189 (2008).

61. Yutin, N. & Koonin, E. V. Proteorhodopsin genes in giant viruses. *Biol. Direct* **7,** 34 (2012).

62. Deisseroth, K. Optogenetics. *Nat. Methods* **8,** 26–29 (2011).

63. Hayashi, S. Silencing Neurons with Light. *Science* **344,** 369–370 (2014).

64. Wietek, J. *et al.* Conversion of Channelrhodopsin into a Light-Gated Chloride Channel. *Science* 1249375 (2014). doi:10.1126/science.1249375

65. Berndt, A., Lee, S. Y., Ramakrishnan, C. & Deisseroth, K. Structure-Guided Transformation of Channelrhodopsin into a Light-Activated Chloride Channel. *Science* **344,** 420–424 (2014).

66. Oesterhelt, D. Structure and Function of Halorhodopsin. *Isr. J. Chem.* **35,** 475–494 (1995).

67. Essen, L.-O. Halorhodopsin: light-driven ion pumping made simple? *Curr. Opin. Struct. Biol.* **12,** 516–522 (2002).

68. Kwon, S.-K. *et al.* Genomic Makeup of the Marine Flavobacterium Nonlabens (Donghaeana) dokdonensis and Identification of a Novel Class of Rhodopsins. *Genome Biol. Evol.* **5,** 187–199 (2013).

69. Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl. Acad. Sci.* **100,** 13940–13945 (2003).

70. Lórenz-Fonfría, V. A. & Heberle, J. Channelrhodopsin unchained: Structure and mechanism of a light-gated cation channel. *Biochim. Biophys. Acta* **1837,** 626–642 (2014).

71. Del Val, C., Royuela-Flor, J., Milenkovic, S. & Bondar, A.-N. Channelrhodopsins: A bioinformatics perspective. *Biochim. Biophys. Acta BBA - Bioenerg.* **1837,** 643–655 (2014).

72. Sasaki, J. & Spudich, J. L. Signal Transfer in Haloarchaeal Sensory Rhodopsin - Transducer Complexes. *Photochem. Photobiol.* **84,** 863–868 (2008).

73. Vogeley, L. *et al.* Anabaena Sensory Rhodopsin: A Photochromic Color Sensor at 2.0 Å. *Science* **306,** 1390–1393 (2004).

74. Kateriya, S., Nagel, G., Bamberg, E. & Hegemann, P. 'Vision' in Single-Celled Algae. *Physiology* **19,** 133–137 (2004).

75. Luck, M. *et al.* A Photochromic Histidine Kinase Rhodopsin (HKR1) That Is Bimodally Switched by Ultraviolet and Blue Light. *J. Biol. Chem.* **287,** 40083–40090 (2012).

76. Gautier, A., Mott, H. R., Bostock, M. J., Kirkpatrick, J. P. & Nietlispach, D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat Struct Mol Biol* **17,** 768–774 (2010).

77. Wang, S. *et al.* Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. *Nat. Methods* **10,** 1007–1012 (2013).

78. Enami, N. *et al.* Crystal Structures of Archaerhodopsin-1 and -2: Common Structural Motif in Archaeal Light-driven Proton Pumps. *J. Mol. Biol.* **358,** 675–685 (2006).

79. Zhang, J. *et al.* Crystal structure of deltarhodopsin-3 from Haloterrigena thermotolerans. *Proteins Struct. Funct. Bioinforma.* **81,** 1585–1592 (2013).

80. Kolbe, M., Besir, H., Essen, L.-O. & Oesterhelt, D. Structure of the Light-Driven Chloride Pump Halorhodopsin at 1.8 Å Resolution. *Science* **288,** 1390–1396 (2000).

81. Kouyama, T. *et al.* Crystal Structure of the Light-Driven Chloride Pump Halorhodopsin from Natronomonas pharaonis. *J. Mol. Biol.* **396,** 564–579 (2010).

82. Royant, A. *et al.* X-ray structure of sensory rhodopsin II at 2.1-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10131–10136 (2001).

83. Wada, T. *et al.* Crystal Structure of the Eukaryotic Light-Driven Proton-Pumping Rhodopsin, Acetabularia Rhodopsin II, from Marine Alga. *J. Mol. Biol.* **411,** 986–998 (2011).

84. Luecke, H. *et al.* Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proc. Natl. Acad. Sci.* **105,** 16561–16565 (2008).

85. Kato, H. E. *et al.* Crystal structure of the channelrhodopsin light-gated cation channel. *Nature* **482,** 369–374 (2012).

86. Ran, T. *et al.* Cross-protomer interaction with the photoactive site in oligomeric proteorhodopsin complexes. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 1965–1980 (2013).

87. Iverson, V. *et al.* Untangling Genomes from Metagenomes: Revealing an Uncultured Class of Marine Euryarchaeota. *Science* **335,** 587–590 (2012).

88. Ghai, R., Mizuno, C. M., Picazo, A., Camacho, A. & Rodriguez-Valera, F. Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci. Rep.* **3,** (2013).

89. Fu, H.-Y. *et al.* A Novel Six-Rhodopsin System in a Single Archaeon. *J. Bacteriol.* **192,** 5866–5873 (2010).

90. Fu, H.-Y., Yi, H.-P., Lu, Y.-H. & Yang, C.-S. Insight into a single halobacterium using a dual-bacteriorhodopsin system with different functionally optimized pH ranges to cope

with periplasmic pH changes associated with continuous light illumination. *Mol. Microbiol.* **88,** 551–561 (2013).

91. Hsu, M.-F. *et al.* Using Haloarcula marismortui Bacteriorhodopsin as a Fusion Tag for Enhancing and Visible Expression of Integral Membrane Proteins in Escherichia coli. *PLoS ONE* **8,** e56363 (2013).

92. Rodrigues, D. F. *et al.* Characterization of Exiguobacterium isolates from the Siberian permafrost. Description of Exiguobacterium sibiricum sp. nov. *Extremophiles* **10,** 285–294 (2006).

93. Rodrigues, D. F. *et al.* Architecture of thermal adaptation in an Exiguobacterium sibiricum strain isolated from 3 million year old permafrost: A genome and transcriptome approach. *BMC Genomics* **9,** 547 (2008).

94. Balashov, S. P. *et al.* Breaking the Carboxyl Rule: Lysine 96 facilitates reprotonation of the Schiff base in the photocycle of a retinal protein from Exiguobacterium sibiricum. *J. Biol. Chem.* **288,** 21254–21265 (2013).

95. Balashov, S. P. *et al.* Aspartate–Histidine Interaction in the Retinal Schiff Base Counterion of the Light-Driven Proton Pump of Exiguobacterium sibiricum. *Biochemistry (Mosc.)* **51,** 5748–5762 (2012).

96. Chevance, F. F. V. & Hughes, K. T. Coordinating assembly of a bacterial macromolecular machine. *Nat. Rev. Microbiol.* **6,** 455–465 (2008).

97. Jarrell, K. F. & McBride, M. J. The surprisingly diverse ways that prokaryotes move. *Nat. Rev. Microbiol.* **6,** 466–476 (2008).

98. Ghosh, A. & Albers, S. Assembly and function of the archaeal flagellum. *Biochem. Soc. Trans.* **39,** 64–69 (2011).

99. Jarrell, K. F. & Albers, S.-V. The archaellum: an old motility structure with a new name. *Trends Microbiol.* **20,** 307–312 (2012).

100. Macnab, R. M. The Bacterial Flagellum: Reversible Rotary Propellor and Type III Export Apparatus. *J. Bacteriol.* **181,** 7149–7153 (1999).

101. Porter, S. L., Wadhams, G. H. & Armitage, J. P. Signal processing in complex chemotaxis pathways. *Nat. Rev. Microbiol.* **9,** 153–165 (2011).

102. Hazelbauer, G. L., Falke, J. J. & Parkinson, J. S. Bacterial chemoreceptors: high-performance signaling in networked arrays. *Trends Biochem. Sci.* **33,** 9–19 (2008).

103. Maddock & Shapiro, L. Polar location of the chemoreceptor complex in the Escherichia coli cell. *Science* **259,** 1717 –1723 (1993).

104. Kentner, D. & Sourjik, V. Spatial organization of the bacterial chemotaxis system. *Curr. Opin. Microbiol.* **9,** 619–624 (2006).

105. Zhang, P., Khursigara, C. M., Hartnell, L. M. & Subramaniam, S. Direct visualization of Escherichia coli chemotaxis receptor arrays using cryo-electron microscopy. *Proc. Natl. Acad. Sci.* **104,** 3777 –3781 (2007).

106. Khursigara, C. M., Wu, X. & Subramaniam, S. Chemoreceptors in Caulobacter crescentus: Trimers of Receptor Dimers in a Partially Ordered Hexagonally Packed Array. *J Bacteriol* **190,** 6805–6810 (2008).

107. Briegel, A. *et al.* Location and architecture of the Caulobacter crescentus chemoreceptor array. *Mol. Microbiol.* **69,** 30–41 (2008).

108. Greenfield, D. *et al.* Self-Organization of the Escherichia coli Chemotaxis Network Imaged with Super-Resolution Light Microscopy. *PLoS Biol* **7,** e1000137 (2009).

109. Erbse, A. H. & Falke, J. J. The Core Signaling Proteins of Bacterial Chemotaxis Assemble To Form an Ultrastable Complex. *Biochemistry (Mosc.)* **48,** 6975–6987 (2009).

110. Khursigara, C. M. *et al.* Lateral density of receptor arrays in the membrane plane influences sensitivity of the E. coli chemotaxis response. *EMBO J* **30,** 1719–1729 (2011).

111. Hazelbauer, G. L. & Lai, W.-C. Bacterial chemoreceptors: providing enhanced features to two-component signaling. *Curr. Opin. Microbiol.* **13,** 124–132 (2010).

112. Li, M. & Hazelbauer, G. L. Core unit of chemotaxis signaling complexes. *Proc. Natl. Acad. Sci.* **108,** 9390 –9395 (2011).

113. Baker, M. D., Wolanin, P. M. & Stock, J. B. Signal transduction in bacterial chemotaxis. *BioEssays* **28,** 9–22 (2006).

114. Bilwes, A. M., Alex, L. A., Crane, B. R. & Simon, M. I. Structure of CheA, a Signal-Transducing Histidine Kinase. *Cell* **96,** 131–141 (1999).

115. Park, S.-Y. *et al.* Reconstruction of the chemotaxis receptor-kinase assembly. *Nat Struct Mol Biol* **13,** 400–407 (2006).

116. Griswold, I. J. *et al.* The solution structure and interactions of CheW from Thermotoga maritima. *Nat Struct Mol Biol* **9,** 121–125 (2002).

117. Li, Y., Hu, Y., Fu, W., Xia, B. & Jin, C. Solution structure of the bacterial chemotaxis adaptor protein CheW from Escherichia coli. *Biochem. Biophys. Res. Commun.* **360,** 863–867 (2007).

118. Bhatnagar, J. *et al.* Structure of the Ternary Complex Formed by a Chemotaxis Receptor Signaling Domain, the CheA Histidine Kinase, and the Coupling Protein CheW As Determined by Pulsed Dipolar ESR Spectroscopy. *Biochemistry (Mosc.)* **49,** 3824–3841 (2010).

119. Li, X. *et al.* The 3.2 Å Resolution Structure of a Receptor:CheA:CheW Signaling Complex Defines Overlapping Binding Sites and Key Residue Interactions within Bacterial Chemosensory Arrays. *Biochemistry (Mosc.)* **52,** 3852–3865 (2013).

120. Briegel, A. *et al.* Bacterial chemoreceptor arrays are hexagonally packed trimers of receptor dimers networked by rings of kinase and coupling proteins. *Proc. Natl. Acad. Sci.* (2012). doi:10.1073/pnas.1115719109

121. Milburn, M. *et al.* Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand. *Science* **254,** 1342 –1347 (1991).

122. Chervitz, S. A. & Falke, J. J. Molecular mechanism of transmembrane signaling by the aspartate receptor: a model. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 2545–2550 (1996).

123. Falke, J. J. & Erbse, A. H. The Piston Rises Again. *Structure* **17,** 1149–1151 (2009).

124. Yeh, J. I. *et al.* High-resolution structures of the ligand binding domain of the wild-type bacterial aspartate receptor. *J. Mol. Biol.* **262,** 186–201 (1996).

125. Gordeliy, V. I. *et al.* Molecular basis of transmembrane signalling by sensory rhodopsin II–transducer complex. *Nature* **419,** 484–487 (2002).

126. Wegener, A. A., Klare, J. P., Engelhard, M. & Steinhoff, H. J. Structural insights into the early steps of receptor-transducer signal transfer in archaeal phototaxis. *EMBO J.* **20,** 5312–5319 (2001).

127. Wegener, A.-A., Chizhov, I., Engelhard, M. & Steinhoff, H.-J. Time-resolved detection of transient movement of helix F in spin-labelled pharaonis sensory rhodopsin II. *J. Mol. Biol.* **301,** 881–891 (2000).

128. Bordignon, E. *et al.* Analysis of light-induced conformational changes of Natronomonas pharaonis sensory rhodopsin II by time resolved electron paramagnetic resonance spectroscopy. *Photochem. Photobiol.* **83,** 263–272 (2007).

129. Moukhametzianov, R. *et al.* Development of the signal in sensory rhodopsin and its transfer to the cognate transducer. *Nature* **440,** 115–119 (2006).

130. Klare, J., Chizhov, I. & Engelhard, M. in *Bioenerg. Energy Conserv. Convers. Eds G Schäfer HS Penefsky* 73–122 (Springer, 2008).

131. Luecke, H., Schobert, B., Lanyi, J. K., Spudich, E. N. & Spudich, J. L. Crystal Structure of Sensory Rhodopsin II at 2.4 Angstroms: Insights into Color Tuning and Transducer Interaction. *Science* **293,** 1499–1503 (2001).

132. Reshetnyak, A. B. *et al.* Comparative analysis of sensory rhodopsin II structures in complex with a transducer and without it. *J. Surf. Investig. X-Ray Synchrotron Neutron Tech.* **2,** 894–899 (2008).

133. Sasaki, J. & Spudich, J. L. Proton Circulation During the Photocycle of Sensory Rhodopsin II. *Biophys. J.* **77,** 2145–2152 (1999).

134. Bogomolni, R. A. *et al.* Removal of transducer HtrI allows electrogenic proton translocation by sensory rhodopsin I. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10188 –10192 (1994).

135. Iwamoto, M., Shimono, K., Sumi, M., Koyama, K. & Kamo, N. Light-Induced Proton Uptake and Release of pharaonis Phoborhodopsin Detected by a Photoelectrochemical Cell. *J. Phys. Chem. B* **103,** 10311–10315 (1999).

136. Schmies, G. *et al.* Sensory rhodopsin II from the haloalkaliphilic natronobacterium pharaonis: light-activated proton transfer reactions. *Biophys. J.* **78,** 967–976 (2000).

137. Sudo, Y. & Spudich, J. L. Three strategically placed hydrogen-bonding residues convert a proton pump into a sensory receptor. *Proc. Natl. Acad. Sci.* **103,** 16129–16134 (2006).

138. Klare, J. P. *et al.* Probing the proton channel and the retinal binding site of Natronobacterium pharaonis sensory rhodopsin II. *Biophys. J.* **82,** 2156–2164 (2002).

139. Schmies, G., Engelhard, M., Wood, P. G., Nagel, G. & Bamberg, E. Electrophysiological characterization of specific interactions between bacterial sensory rhodopsins and their transducers. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 1555–1559 (2001).

140. Sudo, Y., Iwamoto, M., Shimono, K., Sumi, M. & Kamo, N. Photo-Induced Proton Transport of Pharaonis Phoborhodopsin (Sensory Rhodopsin II) Is Ceased by Association with the Transducer. *Biophys. J.* **80,** 916–922 (2001).

141. Sasaki, J. & Spudich, J. L. Proton transport by sensory rhodopsins and its modulation by transducer-binding. *Biochim. Biophys. Acta BBA - Bioenerg.* **1460,** 230–239 (2000).

142. Edman, K. *et al.* Early Structural Rearrangements in the Photocycle of an Integral Membrane Sensory Receptor. *Structure* **10,** 473–482 (2002).

143. Ruprecht, J. J., Mielke, T., Vogel, R., Villa, C. & Schertler, G. F. Electron crystallography reveals the structure of metarhodopsin I. *EMBO J* **23,** 3609–3620 (2004).

144. Choe, H.-W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471,** 651–655 (2011).

145. Parkinson, J. S. Signaling Mechanisms of HAMP Domains in Chemoreceptors and Sensor Kinases. *Annu. Rev. Microbiol.* **64,** 101–122 (2010).

146. Aravind, L. & Ponting, C. P. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol. Lett.* **176,** 111–116 (1999).

147. Williams, S. B. & Stewart, V. MicroReview: Functional similarities among two-component sensors and methyl-accepting chemotaxis proteins suggest a role for linker region amphipathic helices in transmembrane signal transduction. *Mol. Microbiol.* **33,** 1093–1102 (1999).

148. Hulko, M. *et al.* The HAMP Domain Structure Implies Helix Rotation in Transmembrane Signaling. *Cell* **126,** 929–940 (2006).

149. Airola, M. V., Watts, K. J., Bilwes, A. M. & Crane, B. R. Structure of Concatenated HAMP Domains Provides a Mechanism for Signal Transduction. *Structure* **18,** 436–448 (2010).

150. Ames, P., Zhou, Q. & Parkinson, J. S. Mutational Analysis of the Connector Segment in the HAMP Domain of Tsr, the Escherichia coli Serine Chemoreceptor. *J Bacteriol* **190,** 6676–6685 (2008).

151. Zhou, Q., Ames, P. & Parkinson, J. S. Mutational Analyses of HAMP Helices Suggest a Dynamic Bundle Model of Input-Output Signaling in Chemoreceptors. *Mol. Microbiol.* **73,** 801–814 (2009).

152. Butler, S. L. & Falke, J. J. Cysteine and Disulfide Scanning Reveals Two Amphiphilic Helices in the Linker Region of the Aspartate Chemoreceptor. *Biochemistry (Mosc.)* **37,** 10746–10756 (1998).

153. Swain, K. E. & Falke, J. J. Structure of the Conserved HAMP Domain in an Intact, Membrane-Bound Chemoreceptor: A Disulfide Mapping Study. *Biochemistry (Mosc.)* **46,** 13684–13695 (2007).

154. Watts, K. J., Johnson, M. S. & Taylor, B. L. Structure-Function Relationships in the HAMP and Proximal Signaling Domains of the Aerotaxis Receptor Aer. *J. Bacteriol.* **190,** 2118–2127 (2008).

155. Etzkorn, M. *et al.* Complex Formation and Light Activation in Membrane-Embedded Sensory Rhodopsin II as Seen by Solid-State NMR Spectroscopy. *Structure* **18,** 293–300 (2010).

156. Bordignon, E. *et al.* Structural Analysis of a HAMP Domain. *J. Biol. Chem.* **280,** 38767 –38775 (2005).

157. Doebber, M. *et al.* Salt-driven Equilibrium between Two Conformations in the HAMP Domain from Natronomonas pharaonis: The Language of Signal Transfer? *J. Biol. Chem.* **283,** 28691–28701 (2008).

158. Appleman, J. A. & Stewart, V. Mutational Analysis of a Conserved Signal-Transducing Element: the HAMP Linker of the Escherichia coli Nitrate Sensor NarX. *J. Bacteriol.* **185,** 89–97 (2003).

159. Kishii, R., Falzon, L., Yoshida, T., Kobayashi, H. & Inouye, M. Structural and Functional Studies of the HAMP Domain of EnvZ, an Osmosensing Transmembrane Histidine Kinase in Escherichia coli. *J. Biol. Chem.* **282,** 26401–26408 (2007).

160. Wang, J., Sasaki, J., Tsai, A. & Spudich, J. L. HAMP Domain Signal Relay Mechanism in a Sensory Rhodopsin-Transducer Complex. *J. Biol. Chem.* **287,** 21316–21325 (2012).

161. Ferris, H. U. *et al.* Mechanism of Regulation of Receptor Histidine Kinases. *Structure* **20,** 56–66 (2012).

162. Ferris, H. U. *et al.* The Mechanisms of HAMP-Mediated Signaling in Transmembrane Receptors. *Structure* **19,** 378–385 (2011).

163. Cheung, J. & Hendrickson, W. A. Structural Analysis of Ligand Stimulation of the Histidine Kinase NarX. *Structure* **17,** 190–201 (2009).

164. Zhou, Q., Ames, P. & Parkinson, J. S. Biphasic control logic of HAMP domain signalling in the Escherichia coli serine chemoreceptor. *Mol. Microbiol.* **80,** 596–611 (2011).

165. Nishikata, K., Fuchigami, S., Ikeguchi, M. & Kidera, A. Molecular modeling of the HAMP domain of sensory rhodopsin II transducer from Natronomonas pharaonis. *Biophysics* **6,** 27–36 (2010).

166. Nishikata, K., Ikeguchi, M. & Kidera, A. Comparative Simulations of the Ground State and the M-Intermediate State of the Sensory Rhodopsin II–Transducer Complex with a HAMP Domain Model. *Biochemistry (Mosc.)* **51,** 5958–5966 (2012).

167. Park, H., Im, W. & Seok, C. Transmembrane Signaling of Chemotaxis Receptor Tar: Insights from Molecular Dynamics Simulation Studies. *Biophys. J.* **100,** 2955–2963 (2011).

168. Hall, B. A., Armitage, J. P. & Sansom, M. S. P. Transmembrane Helix Dynamics of Bacterial Chemoreceptors Supports a Piston Model of Signalling. *PLoS Comput Biol* **7,** e1002204 (2011).

169. Lemmin, T., Soto, C. S., Clinthorne, G., DeGrado, W. F. & Dal Peraro, M. Assembly of the Transmembrane Domain of E. coli PhoQ Histidine Kinase: Implications for Signal Transduction from Molecular Simulations. *PLoS Comput Biol* **9,** e1002878 (2013).

170. Hall, B. A., Armitage, J. P. & Sansom, M. S. P. Mechanism of Bacterial Signal Transduction Revealed by Molecular Dynamics of Tsr Dimers and Trimers of Dimers in Lipid Vesicles. *PLoS Comput Biol* **8,** e1002685 (2012).

171. Hayashi, K. *et al.* Structural Analysis of the Phototactic Transducer Protein HtrII Linker Region from Natronomonas pharaonis. *Biochemistry (Mosc.)* **46,** 14380–14390 (2007).

172. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32,** D138–141 (2004).

173. Hunter, S. *et al.* InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* **40,** D306–312 (2012).

174. Sigrist, C. J. A. *et al.* New and continuing developments at PROSITE. *Nucleic Acids Res.* **41,** D344–347 (2013).

175. Hermansson, M., Hokynar, K. & Somerharju, P. Mechanisms of glycerophospholipid homeostasis in mammalian cells. *Prog. Lipid Res.* **50,** 240–257 (2011).

176. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).

177. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18,** 6097–6100 (1990).

178. Müller, V., Spanheimer, R. & Santos, H. Stress response by solute accumulation in archaea. *Curr. Opin. Microbiol.* **8,** 729–736 (2005).

179. Ciulla, R. A., Burggraf, S., Stetter, K. O. & Roberts, M. F. Occurrence and Role of Di-myo-Inositol-1,1′-Phosphate in Methanococcus igneus. *Appl. Environ. Microbiol.* **60,** 3660–3664 (1994).

180. Martins, L. O., Carreto, L. S., Costa, M. S. D. & Santos, H. New compatible solutes related to Di-myo-inositol-phosphate in members of the order Thermotogales. *J. Bacteriol.* **178,** 5644–5651 (1996).

181. Rodionov, D. A. *et al.* Genomic identification and in vitro reconstitution of a complete biosynthetic pathway for the osmolyte di-myo-inositol-phosphate. *Proc. Natl. Acad. Sci.* **104,** 4279–4284 (2007).

182. Borges, N. *et al.* Biosynthetic Pathways of Inositol and Glycerol Phosphodiesters Used by the Hyperthermophile Archaeoglobus fulgidus in Stress Adaptation. *J. Bacteriol.* **188,** 8128–8135 (2006).

183. Drobak, B. K. The plant phosphoinositide system. *Biochem. J.* **288,** 697–712 (1992).

184. Majerus, P. W. Inositol Phosphate Biochemistry. *Annu. Rev. Biochem.* **61,** 225–250 (1992).

185. Rodrigues, M. V. *et al.* Bifunctional CTP:Inositol-1-Phosphate Cytidylyltransferase/CDP-Inositol:Inositol-1-Phosphate Transferase, the Key Enzyme for Di-myo-Inositol-Phosphate Synthesis in Several (Hyper)thermophiles. *J. Bacteriol.* **189,** 5405–5412 (2007).

186. Brito, J. A., Borges, N., Vonrhein, C., Santos, H. & Archer, M. Crystal Structure of Archaeoglobus fulgidus CTP:Inositol-1-Phosphate Cytidylyltransferase, a Key Enzyme for Di-myo-Inositol-Phosphate Synthesis in (Hyper)Thermophiles. *J. Bacteriol.* **193,** 2177–2185 (2011).

187. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).

188. Hohenfeld, I. P., Wegener, A. A. & Engelhard, M. Purification of histidine tagged bacteriorhodopsin, pharaonis halorhodopsin and pharaonis sensory rhodopsin II functionally expressed in Escherichia coli. *FEBS Lett.* **442,** 198–202 (1999).

189. Shimono, K., Iwamoto, M., Sumi, M. & Kamo, N. Functional expression of pharaonis phoborhodopsin in Eschericha coli. *FEBS Lett.* **420,** 54–56 (1997).

190. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4,** 706–731 (2009).

191. Leslie, A. G. W. & Powell, H. R. in *Evol. Methods Macromol. Crystallogr.* (Read, R. J. & Sussman, J. L.) 41–51 (Springer Netherlands, 2007). at <http://link.springer.com/chapter/10.1007/978-1-4020-6316-9_4>

192. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 (2010).

193. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 72–82 (2005).

194. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242 (2011).

195. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 22–25 (2009).

196. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 355–367 (2011).

197. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

198. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).

199. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7,** 1511–1522 (2012).

200. Bordoli, L. *et al.* Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **4,** 1–13 (2009).

201. Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L. & Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **37,** D387–D392 (2009).

202. Strong, M. *et al.* Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci.* **103,** 8060 –8065 (2006).

203. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921 (1998).

204. Ursby, T. & Bourgeois, D. Improved Estimation of Structure-Factor Difference Amplitudes from Poorly Accurate Data. *Acta Crystallogr. A* **53,** 564–575 (1997).

205. Nishikata, K., Fuchigami, S., Ikeguchi, M. & Kidera, A. Molecular modeling of the HAMP domain of sensory rhodopsin II transducer from *Natronomonas pharaonis*. *Biophysics* **6,** 27–36 (2010).

206. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4,** 435–447 (2008).

207. MacKerell *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102,** 3586–3616 (1998).

208. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25,** 1400–1415 (2004).

209. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **65,** 712–725 (2006).

210. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **78,** 1950–1958 (2010).

211. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **100,** 020603 (2008).

212. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).

213. Hovmöller, S., Zhou, T. & Ohlson, T. Conformations of amino acids in proteins. *Acta Crystallogr. D Biol. Crystallogr.* **58,** 768–776 (2002).

214. Kitao, A. & Go, N. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* **9,** 164–169 (1999).

215. Humphrey, W., Dalke, A. & Schulten, K. VMD Visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).

216. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31,** 455–461 (2010).

217. Chang, M. W., Ayeni, C., Breuer, S. & Torbett, B. E. Virtual Screening for HIV Protease Inhibitors: A Comparison of AutoDock 4 and Vina. *PLoS ONE* **5,** e11955 (2010).

218. Luecke, H., Schobert, B., Richter, H.-T., Cartailler, J.-P. & Lanyi, J. K. Structural Changes in Bacteriorhodopsin During Ion Transport at 2 Angstrom Resolution. *Science* **286,** 255–260 (1999).

219. Freier, E., Wolf, S. & Gerwert, K. Proton transfer via a transient linear water-molecule chain in a membrane protein. *Proc. Natl. Acad. Sci.* **108,** 11435–11439 (2011).

220. Gerwert, K., Freier, E. & Wolf, S. The role of protein-bound water molecules in microbial rhodopsins. *Biochim. Biophys. Acta BBA - Bioenerg.* doi:10.1016/j.bbabio.2013.09.006

221. Faham, S. *et al.* Side-chain contributions to membrane protein structure and stability. *J. Mol. Biol.* **335,** 297–305 (2004).

222. Joh, N. H. *et al.* Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* **453,** 1266–1270 (2008).

223. Essen, L.-O., Siegert, R., Lehmann, W. D. & Oesterhelt, D. Lipid patches in membrane protein oligomers: Crystal structure of the bacteriorhodopsin-lipid complex. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 11673–11678 (1998).

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

225. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinforma. Oxf. Engl.* **23,** 2947–2948 (2007).

226. Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7,** 439 (2006).

227. Grishaev, A. & Bax, A. An Empirical Backbone−Backbone Hydrogen-Bonding Potential in Proteins and Its Applications to NMR Structure Refinement and Validation. *J. Am. Chem. Soc.* **126,** 7281–7292 (2004).

228. Cooley, R. B., Arp, D. J. & Karplus, P. A. Evolutionary Origin of a Secondary Structure: π-Helices as Cryptic but Widespread Insertional Variations of α-Helices That Enhance Protein Functionality. *J. Mol. Biol.* **404,** 232–246 (2010).

229. Cartailler, J.-P. & Luecke, H. Structural and Functional Characterization of π Bulges and Other Short Intrahelical Deformations. *Structure* **12,** 133–144 (2004).

230. Vieira-Pires, R. S. & Morais-Cabral, J. H. 310 helices in channels and other membrane proteins. *J. Gen. Physiol.* **136,** 585–592 (2010).

231. Balashov, S. P. *et al.* Xanthorhodopsin: A Proton Pump with a Light-Harvesting Carotenoid Antenna. *Science* **309,** 2061–2064 (2005).

232. Cleland, W. W. & Kreevoy, M. M. Low-barrier hydrogen bonds and enzymic catalysis. *Science* **264,** 1887–1890 (1994).

233. Cleland, W. W., Frey, P. A. & Gerlt, J. A. The Low Barrier Hydrogen Bond in Enzymatic Catalysis. *J. Biol. Chem.* **273,** 25529–25532 (1998).

234. Frey, P. A., Whitt, S. A. & Tobin, J. B. A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* **264,** 1927–1930 (1994).

235. Ho, B. & Gruswitz, F. HOLLOW: Generating Accurate Representations of Channel and Interior Surfaces in Molecular Structures. *BMC Struct. Biol.* **8,** 49 (2008).

236. Lörinczi, É. *et al.* Voltage- and pH-Dependent Changes in Vectoriality of Photocurrents Mediated by Wild-type and Mutant Proteorhodopsins upon Expression in Xenopus Oocytes. *J. Mol. Biol.* **393,** 320–341 (2009).

237. Isom, D. G., Castañeda, C. A., Cannon, B. R. & E, B. G.-M. Large shifts in pKa values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci.* **108,** 5260–5265 (2011).

238. Balashov, S. P. *et al.* Reconstitution of Gloeobacter Rhodopsin with Echinenone: Role of the 4-Keto Group. *Biochemistry (Mosc.)* **49,** 9792–9799 (2010).

239. Imasheva, E. S., Balashov, S. P., Choi, A. R., Jung, K.-H. & Lanyi, J. K. Reconstitution of Gloeobacter violaceus Rhodopsin with a Light-Harvesting Carotenoid Antenna. *Biochemistry (Mosc.)* **48,** 10948–10955 (2009).

240. Khan, S. T., Nakagawa, Y. & Harayama, S. Krokinobacter gen. nov., with three novel species, in the family Flavobacteriaceae. *Int. J. Syst. Evol. Microbiol.* **56,** 323–328 (2006).

241. Chizhov, I. *et al.* The photophobic receptor from Natronobacterium pharaonis: temperature and pH dependencies of the photocycle of sensory rhodopsin II. *Biophys. J.* **75,** 999–1009 (1998).

242. Taniguchi, Y., Ikehara, T., Kamo, N., Yamasaki, H. & Toyoshima, Y. Dynamics of Light-Induced Conformational Changes of the Phoborhodopsin/Transducer Complex Formed in the n-Dodecyl β-d-Maltoside Micelle. *Biochemistry (Mosc.)* **46,** 5349–5357 (2007).

243. Yan, B., Takahashi, T., Johnson, R. & Spudich, J. L. Identification of signaling states of a sensory receptor by modulation of lifetimes of stimulus-induced conformations: the case of sensory rhodopsin II. *Biochemistry (Mosc.)* **30,** 10686–10692 (1991).

244. Hein, M., Wegener, A. A., Engelhard, M. & Siebert, F. Time-Resolved FTIR Studies of Sensory Rhodopsin II (NpSRII) from Natronobacterium pharaonis: Implications for Proton Transport and Receptor Activation. *Biophys. J.* **84,** 1208–1217 (2003).

245. Bergo, V., Spudich, E. N., Spudich, J. L. & Rothschild, K. J. Conformational changes detected in a sensory rhodopsin II-transducer complex. *J. Biol. Chem.* **278,** 36556–36562 (2003).

246. Furutani, Y. *et al.* FTIR Spectroscopy of the O Photointermediate in pharaonis Phoborhodopsin. *Biochemistry (Mosc.)* **43,** 5204–5212 (2004).

247. Inoue, K., Sasaki, J., Spudich, J. L. & Terazima, M. Laser-Induced Transient Grating Analysis of Dynamics of Interaction between Sensory Rhodopsin II D75N and the HtrII Transducer. *Biophys. J.* **92,** 2028–2040 (2007).

248. Subramaniam, S. B. ., Tittor, J. O. . & Lanyi, J. H. . Protein conformational changes in the bacteriorhodopsin photocycle. *J. Mol. Biol.* **287,** (1999).

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

250. Bergo, V. B., Spudich, E. N., Rothschild, K. J. & Spudich, J. L. Photoactivation Perturbs the Membrane-embedded Contacts between Sensory Rhodopsin II and Its Transducer. *J. Biol. Chem.* **280,** 28365 –28369 (2005).

251. Yoshida, H., Sudo, Y., Shimono, K., Iwamoto, M. & Kamo, N. Transient movement of helix F revealed by photo-induced inactivation by reaction of a bulky SH-reagent to cysteine-introduced pharaonis phoborhodopsin (sensory rhodopsin II). *Photochem. Photobiol. Sci. Off. J. Eur. Photochem. Assoc. Eur. Soc. Photobiol.* **3,** 537–542 (2004).

252. Yang, C.-S., Sineshchekov, O., Spudich, E. N. & Spudich, J. L. The Cytoplasmic Membrane-proximal Domain of the HtrII Transducer Interacts with the E-F Loop of Photoactivated Natronomonas pharaonis Sensory Rhodopsin II. *J Biol Chem* **279,** 42970–42976 (2004).

253. Sasaki, J., Nara, T., Spudich, E. N. & Spudich, J. L. Constitutive activity in chimeras and deletions localize sensory rhodopsin II/HtrII signal relay to the membrane-inserted domain. *Mol. Microbiol.* **66,** 1321–1330 (2007).

254. Sudo, Y., Furutani, Y., Kandori, H. & Spudich, J. L. Functional importance of the interhelical hydrogen bond between Thr204 and Tyr174 of sensory rhodopsin II and its alteration during the signaling process. *J. Biol. Chem.* **281,** 34239 (2006).

255. Ito, M. *et al.* Steric Constraint in the Primary Photoproduct of Sensory Rhodopsin II Is a Prerequisite for Light-Signal Transfer to HtrII. *Biochemistry (Mosc.)* **47,** 6208–6215 (2008).

256. Furutani, Y., Sudo, Y., Kamo, N. & Kandori, H. FTIR Spectroscopy of the Complex between pharaonis Phoborhodopsin and Its Transducer Protein. *Biochemistry (Mosc.)* **42,** 4837–4842 (2003).

257. Sudo, Y., Furutani, Y., Shimono, K., Kamo, N. & Kandori, H. Hydrogen Bonding Alteration of Thr-204 in the Complex between pharaonis Phoborhodopsin and Its Transducer Protein. *Biochemistry (Mosc.)* **42,** 14166–14172 (2003).

258. Lanyi, J. K. Proton transfers in the bacteriorhodopsin photocycle. *Biochim. Biophys. Acta BBA - Bioenerg.* **1757,** 1012–1018 (2006).

259. Kim, K. K., Yokota, H. & Kim, S.-H. Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. *Nature* **400,** 787–792 (1999).

260. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16,** 404–405 (2000).

261. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252,** 1162–1164 (1991).

262. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **16,** 1339–1350 (2004).

263. Anantharaman, V., Balaji, S. & Aravind, L. The signaling helix: a common functional theme in diverse signaling proteins. *Biol. Direct* **1,** 25 (2006).

264. Lindorff-Larsen, K. *et al.* Systematic Validation of Protein Force Fields against Experimental Data. *PLoS ONE* **7,** e32131 (2012).

265. Beauchamp, K. A., Lin, Y.-S., Das, R. & Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **8,** 1409–1414 (2012).

266. Ostermeier, C. & Michel, H. Crystallization of membrane proteins. *Curr. Opin. Struct. Biol.* **7,** 697–701 (1997).

267. Rao, S. T. & Rossmann, M. G. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **76,** 241–256 (1973).