Expression, Purification, and Crystallization of Bacteriorhodopsin and Its Derivatives.

From the Faculty of Georesources and Materials Engineering of the
RWTH Aachen University

Submitted by

Dmitry Bratanov, M. Sc

from Taganrog, Russian Federation

in respect of the academic degree of

Doctor of Natural Sciences

approved thesis

Advisors: Prof. Dr. Valentin Gordeliy
Prof. Dr. Georg Büldt

Date of the oral examination: 24.09.2014

Table of Contents

Abbreviations

bR	bacteriorhodopsin from *Halobium salinarum*
SRII	sensory rhodopsin II from *Natronomonas pharaonis*
hR	halorhodopsin from *Natronomonas pharaonis*
ESR	rhodopsin from *Exiguobacterium sibiricum*
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
DMPC	L-α-1,2-dimyristoylphosphatidylcholine
DDM	*n*-dodecyl-β-D-maltopyranoside
Sarkosyl	sodium *N*-lauroylsarcosinate
SDS	sodium dodecyl sulfate
MP	membrane protein
GPCR	G-protein coupled receptor
MISTIC	Membrane-Integrating Sequence for Translation Integral membrane protein Constructs
Lys	Thermostabilized T4 lysozyme
MO	monoglycerol monooleoyl, monoolein
NTR	non translating region
PCR	polymerase chain reaction

1 Introduction

Membrane proteins (MPs) are major functional components of cell membrane. They implement function of signal transduction, carry out passive and active transport across membrane, and mediate intercellular communications. Membrane proteins constitute about one third of human genome and are targets for more than 50-60% of modern drugs [1]. Availability of the structures of MPs with atomic resolution is crucial not only for drug design, but also for basic science. However, despite great interest and importance of MPs, their structures make up just about 1% of all high resolution structures deposited in Protein Data Bank [2-4].

Throughout the years X-ray crystallography became a method of choice to determine a high resolution 3D structure of MPs, since it allows determination of atomic coordinates for small and large protein molecules. However, its application has two general bottlenecks dramatically slowing down the MP studies: protein production and crystallization.

Generally, homologous expression of MPs is at very low level, making the purification of material from natural source virtually impossible. Unfortunately, the yield of heterologous overexpression is quite often poor. Purification of the functional MP is not a straightforward process requiring multiple steps. Despite the loss of material on each step can be acceptable, the total protein losses throughout the whole solubilization and purification process make a considerable part of total expressed protein. The final yield of pure functional MP is low in many cases, thus limiting further studies, especially structural studies.

However, even if the protein material is available in large quantities, the determination of high resolution MP structure using X-ray crystallography is not facile, while this method demands crystals. Unfortunately, MPs are often difficult to crystallize due to the inherent instability of solubilized MPs in detergent and steric hindrances caused by detergent micelles. The available methods of MP crystallization are far from what is required. Despite recently new promising approaches to crystallization of MP were developed [5-7], the determination of the MP structure at atomic resolution is still a great challenge.

In this work we are mostly focused on the issues of MP production and crystallization. The major object under study was bacteriorhodopsin from *Halobium salinarum* (bR), a light-driven pump that provides proton vectorial transport across the cell membrane of archaea [8, 9]. This protein is a striking example of difficulties that scientists encounter handling MPs. The case of bR

illustrates to the full extent the fails of functional heterologous expression of MPs and problems of MP crystallization.

Bacteriorhodopsin consists of 7 transmembrane α-helices with retinal molecule bound covalently to the Lys 216 residue [10]. As one of the simplest proteins involved in a key and universal step of energy production in living cells, bR is of great interest for bioenergetics. Due to the high level of homologous expression, ease of purification from the natural source [11], and high thermal and chemical stability [12] this protein has become the most studied MP being a model for expression, folding, and crystallization experiments.

Despite availibility from natural source and appearance of first crystals in 1980 [13], the high resolution 3D structure of bR was first obtained in 1996 [6], when a new *in meso* crystallization concept was introduced. *In meso* approach is based on the use of three-dimensional amphiphilic bilayer arrays forming interconnected bilayers pervaded by aqueous channels as crystallization medium. In more native lipid-based environment MP remains stable for a longer time, and its crystallization is not restrained by the bulky detergent micelle. This new approach allowed scientists to obtain atomic resolution structures of bR [6, 14] that significantly improved understanding of the vectorial proton transfer mechanism [15]. However, there are still a controversies regarding the structures of bR photocycle intermediates obtained by different groups [16] requiring new structural studies of the protein.

For the investigation of bR photocycle and determination of the atomic structures of photocycle intermediates an availability of different bR mutants is essential. The bR and its mutants are also highly demanded in the applied studies regarding possible bR applications in bioelectronics, optics, biophotonics, etc [17, 18]. But homologous production of bR and its mutants in halobacteria is laborious, time-, and resource-consuming. Therefore, the more efficient system for functional expression of bR would be of wide interest.

The functional expression of bR in *E. coli* would be the most simple, robust, and inexpensive system [19], and the efforts to reach this goal last already almost 30 years [20]. First studies have shown a low level of wild type bR expression in *E. coli* because of severe degradation of newly synthesized protein [20]. Application of exogenous *N*-terminal tags allowed to stabilize the protein, prevent its degradation, and increase the yield of modified bacterioopsin up to 17 mg of protein per liter of culture [20, 21]. Further progress was related to the use of the fusion proteins that boosted the expression level up to 100-200 mg/l [22-24]. However, the mentioned systems did not provide

functional bR expression: the protein had to be first undergone unfolding followed by refolding. Such behavior of bR is all the more surprising, as the expression of a number of closely related retinal proteins, namely sensory rhodopsin I and II from *H. salinarum* [25, 26], halorhodopsin and sensory rhodopsin II from *N. pharaonis* [27], bacteriorhodopsin from *Exiguobacterium sibiricum* [28], deltarhodopsin from *Haloterrigena* turkmenica [29], bacteriorhodopsins and sensory rhodopsins from *Haloarcula marismortui* [30], and others, was shown to be functional in *E. coli*. Solving this puzzle would shed some light on the obstacles found in heterologous MP expression in *E. coli*.

Why the other retinal proteins do express functionally in *E. coli* and bR does not? What is the difference between bR and, for instance, SRII that allows functional expression of SRII in *E. coli*? In the present study we have addressed this problem applying the protein complementarity approach. We constructed chimeric proteins between bR and complementary protein SRII, the only retinal protein expressed functionally in *E. coli* [27, 31] for which the high resolution 3D crystal structure was available to the moment we have started this study [32, 33]. This approach allowed us to localize quickly the reason for the lack of bR expression in *E. coli*. We suggest that this approach may have a general application.

The major problem of bR low expression could be caused by the unfavorable mRNA structure downstream of ribosome binding site or by presence of positively charged amino acids on the *N*-terminus of bR. The removal of the positive charge on the *N*-terminus of bR improved the yield of the protein up to 7 times comparing to native gene. Two silent mutations that destabilize the stem impeding the transcription of the bR gene gave 50-fold expression gain comparing to native gene. In addition, improving of expression conditions led to better incorporation of the newly synthesized protein into *E. coli* membrane resulting in considerable amount of the properly folded bR. We have developed the protocol for purification of the functional bR in non-denaturing conditions with the yield of 2.4±1.3 mg/L providing the protein most closely resembling bR from native purple membranes. Using developed protocol the V49A, D85N, and D96N mutants of bR were readily produced.

One of the major tasks of the present work was to obtain the bR-Lys fusion protein that can allow to resolve the controversies on bR structural studies. In 2007 the first high resolution structure of GPCR was published [34]. The one of the main features that made crystallization of the protein possible was the use of GPCR-Lys FP. Lysozyme reduced conformational heterogeneity stabilizing the protein and provided crystal contacts facilitating nucleation. In next few years several other

GPCR structures were published where Lys was used as a crystallization tag [35-38]. We suppose that bR would be a good starting model for investigation if the utilization of Lys as crystallization tags is general. As GPCRs have the same structural topology as bR, we suppose that crystallization of bR-Lys fusion protein have high chances for success. As *in meso* grown crystals of bR are prone to severe twinning the crystallization of bR-Lys fusion may produce twinning-free crystals of bR in another space group. Such crystals would facilitate the determination of the precise atomic structures of bR and its intermediates without bias caused by twinning. The other very important point is to check what is the influence of Lys as a crystallization tag on the structure of crystallized protein. The structure of bR ground state is well-described by several groups thus providing a good model for comparison of structure with tag and without it.

Realization of such goal requires a crystallization of the bR-Lys fusion protein demanding high yield expression system and effective crystallization approaches. Here, bR can serve as a guiding reference, as the approaches for the expression and crystallization of this protein have been developing throughout several decades.

In membranes of halobacteria bR forms the 2D crystalline lattice called purple membranes. Purple membranes consist of 25% of lipids and 75% of bR. The reported X-ray structures of bR were solved using protein preparations from purple membranes [5-7]. High resolution structures show that bR trimers are surrounded by the native lipid belt shielding the hydrophobic area of the protein [14]. Despite multiple protocols of *E. coli* expression there have not been reports about successful 3D crystallization of the heterologously expressed bR. The 2D crystals of bR expressed in *E. coli* were obtained from protein reconstituted into native *H. salinarum* lipids [39]. Taking into account the participation of the native lipids in the formation of the lattice of 2D and 3D bR crystals there were speculations about the influence of *H. salinarum* native lipids on the crystallization of bR. Are native lipids required for bR crystallization or they can be avoided?

Another important point is the use of detergents. Since the *in meso* grown crystals of bR were obtained from purple membranes solubilized in OG, the use of other detergents might impair the ability of bR to form crystals. In addition, there is a concept that for MP solubilization the long chain detergents are good as they stabilize the protein, while for crystallization the short chain detergents are better as they do not provide a steric conflicts and increase the curvature of lipid bilayers. The detergent exchange is very long and can be harmful for protein protein. Unfortunately, the role of detergent in the process of *in meso* crystallization is not well studied. Does OG form specific contacts to stabilize *in meso* grown bR crystals? Can we use the mixtures of detergents to

eliminate the step of detergent exchange? This and relevant questions are still open.

Thus, the detailed investigation of the influence of lipid/detergent environment on the *in meso* crystallization of bR is required. In the present study to establish the role of detergent in *in meso* crystallization we carried out the crystallization of the bR from *H. salinarum* purple membranes in three mixtures of detergents. The crystals appeared in all mixtures. Using the crystals grown in OG-MEGA10 mixtures the structure of bR ground state was solved at 1.45 Å resolution. This high resolution structure (along with the others solved by my colleagues) have proved that detergent molecules are not necessary important for the formation of contacts between protein molecules in the crystals, and that the mixtures of detergents can be also used in *in meso* crystallization. Thus, we have shown that detergent influences on the *in meso* crystallization not via providing a specific crystal contacts but by altering the physical properties of the lipid matrix. Moreover, the wild type bR and its mutants expressed in *E. coli* were also crystallized *in meso.* The crystals had the same hexagonal shape as crystals grown from purple membranes and diffracted up to 2.4 Å resolution when tested using synchrotron radiation, thus giving the first diffracting 3D crystals of bR expressed in *E. coli*. This way we have proved the quality of the obtained protein material. On the other hand, the well diffracting crystals obtained from the probes where the lipids from *H. salinarum* were not present confirm that native lipids are not required for the *in meso* crystallization of bR.

Thus, the goals of the present study were:

1. Development of *E. coli* based expression system that will enable functional expression of bacteriorhodopsin and its mutants with significantly high yield for the subsequent crystallographic studies of bacteriorhodopsin and its photocycle intermediates as well as industrial use. For the obtained protein non-denaturing purification protocol should be elaborated allowing the production of bacteriorhodopsin and its mutants with the purity sufficient for crystallization and other downstream applications.

2. Expression in and purification of the functional fusion protein between bacteriorhodopsin and lysozyme for crystallographic studies.

3. Investigation of the influence of the detergents and their mixtures, as well as native lipids on *in meso* crystallization of bR. The main practical objective was to clarify if it is possible a) to avoid the use of *H. salinarum* native lipids; and b) to employ the mixtures of detergents to

avoid the complete detergent exchange, which is usually recommended by standard protocols of crystallization.

2 Literature review

2.1 Membrane proteins.

Every cell is surrounded by biological membrane that acts as barrier between the inner space of the cell and external environment. Biomembrane essentially consists of lipid bilayer with embedded membrane proteins (MPs) realizing the functions attributed to biomembranes. Membrane proteins fulfill the interactions of cell with the outside medium and, thus, implement function of signal sensing, carry out passive and active transport across membrane, are responsible for cell-cell recognition and intercellular junctions as well as for some very important energetic processes in cell. The importance of MP studies can be illustrated by the fact that MPs are targets for more than 50% of modern medical drugs [1].

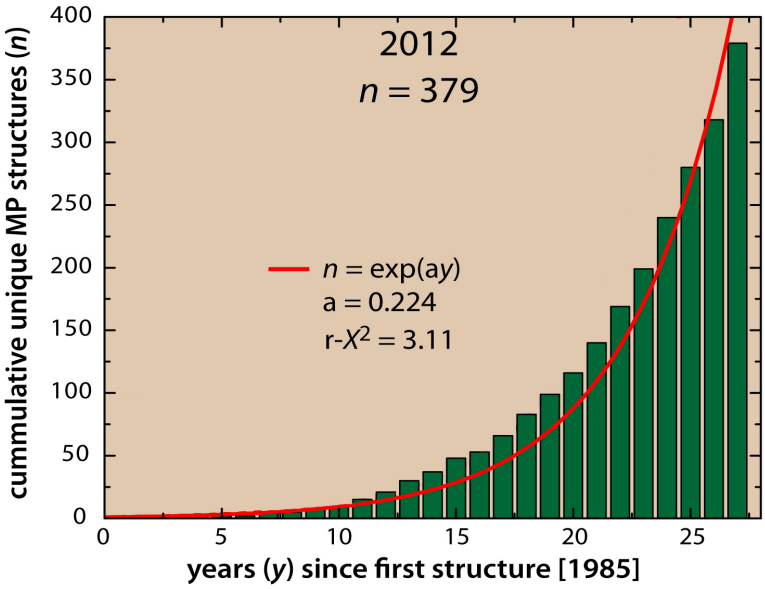

Figure 1: The amount of unique MP structures in Protein Data Bank. From http://blanco.biomol.uci.edu/mpstruc/.

For clear understanding of MP functions the structure of the MP at atomic resolution is required. High resolution structures of MPs are prerequisite for computer based drug design as well. However, in practice, solving the MP structure at high resolution was proved to be difficult being a great challenge in modern structural biology. Despite MPs constitute about one third of human genome, their structures make up just about 1% of all high resolution structures deposited in Protein Data Bank [2, 4]: from more than 92 000 of the total structures the MP structures comprise only ~1200 (statistic is available at http://blanco.biomol.uci.edu/mpstruc/). The count of the unique MPs is only ~400 (**Fig. 1**). Such lag between soluble and membrane protein structures is explained by the

amphiphilic nature of the MPs that allow them to settle in membrane but give a rise to problems when handle them outside of native lipid environment.

In spite of great variety of methods to obtain information about protein structure, throughout the years X-ray crystallography became a method of choice to determine a high resolution 3D structure of MPs, since it allows determination of atomic coordinates for comparatively large protein molecules. But this method has two general bottlenecks dramatically slowing down progress in the field: MP production and MP crystallization.

The first thing one will encounter trying to work with MPs is the availability of the purified protein material. Generally, in normal conditions MPs present in cell at very low level, so the purification of the material from natural source in most cases is virtually impossible. In such case heterologous overexpression is required, however, its yield is often poor. Moreover, the requirements [40] of MPs to the specific lipids from their native organisms can limit the folding of MPs during heterologous overexpression. Then, due to their amphiphilic nature MPs have to be transferred to solution (solubilized) before purification. Outside of their native lipid environment MPs are significantly less stable than in membrane. Organic solvents convert MPs to nonfunctional state and, thus, should be avoided. The selection of proper surfactant (detergent) and optimization of solubilization conditions are still a matter of trials and errors in looking for a compromise between the amount of total protein recovered and the fraction of functional protein in preparations. Additionally, the use of detergents considerably decrease the efficiency of protein purification methods mostly designed for water soluble proteins. Thus, purification of functional MP is not a straightforward process and requires multiple steps leading to the protein losses. Taking into account that typical concentrations of MP required for crystallization is over 10 mg/ml, the final yield of pure functional MP limits its further studies in many cases. Even utilization of the nanovolume *in meso* crystallization [41] requires at least 1 mg of the purified protein for extensive screening of crystallization conditions. Such amount of purified material is not readily available for a lot of the MPs, particularly for the pharmaceutically important human MPs.

Yet still if the protein is expressed and functionally purified with high yield, the solving of MP structure is not a low-duty affair. Because X-ray crystallography demands crystals. However, if for soluble proteins crystallization techniques are well developed and successfully applied in practice, crystallization of MPs is still a great challenge. The problems in crystallization originate from amphiphilic nature of MPs. Detergent micelle is a poor substitution of the native membrane, for this reason solubilized MP is inherently unstable. Even extremely stable in purple membranes bR

solubilized in OG quickly denatures under light illumination [42]. Growth of MP crystal can take up to 2 months and during all this time the protein should remain properly folded and functional. Crystallization of MPs by detergent can be also hampered by steric hindrances caused by detergent micelles. When MP is solubilized, detergent molecules form a belt covering the hydrophobic areas of MP but at the same time decreasing the possible area of protein-protein interactions. Recently several new promising approaches to crystallization of MPs were developed [5-7]. They employ as a crystallization media a special lipid environment that is closer to the native membrane than detergent micelle. Nevertheless, available methods of MP crystallization are far from what is required and solving of MP structure at atomic resolution is still a challenge. Bacteriorhodopsin from *H. salinarum* (bR) being the most studied MP to date is an illustrative example of state of art in MP research.

2.2 Bacteriorhodopsin.

2.2.1. General overview.

Bacteriorhodopsin from *H. salinarum* (bR) is a light-driven pump that provides proton vectorial transport across the cell membrane of the archaea [8, 43]. It consists of 7 transmembrane α-hilices

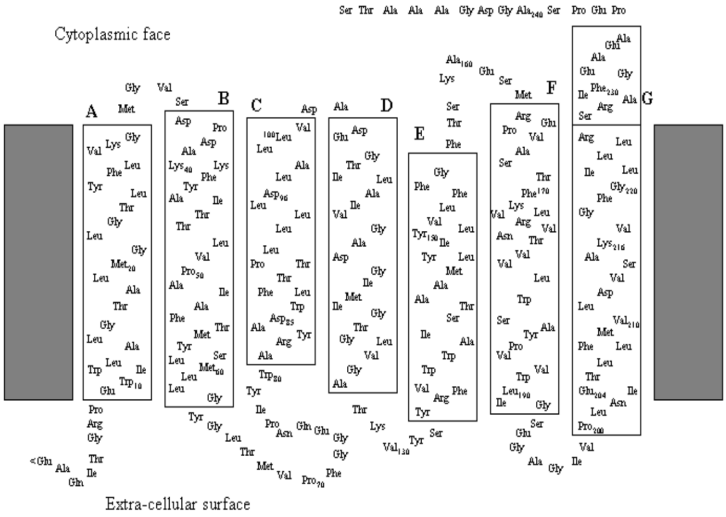

Figure 2: Schematic representation of bR primary and secondary structure. Modified from [44].

with retinal molecule (vitamin A derivative) bound covalently to the Lys 216 residue via protonated Schiff base [10]. Amino acid sequence of bR as well as nucleotide sequence of bR native gene are known (**Fig. 2**) [45, 46]. In nature bR is expressed by *H. salinarum*, archea that can be found in highly saline lakes or salt preserved food. The halobacterium cells can grow aerobically or anaerobically, but under low oxygen concentration and intense light they overexpress bR that forms

purple membranes. Being the natural 2D crystals of bR with P_3 symmetry [47] purple membranes can occupy up to 50% of the surface area of halobacteria. Bacteriorhodopsin comprises 75% of mass of purple membrane being the only protein in it [48]. Purple membranes can be easily isolated by several centrifugations [11] and the yield of homogeneous bR can reach 30 mg per liter of *H. salinarum* culture. In purple membrane bR is stable in wide range of pH (2-10) and temperatures (up to 100°C) [49]. Such eminent features of bR allowed it to become not only the model of protein transfer across membrane, but the general model of MP. Many new approaches to MPs (concerning expression, solubilization, crystallization) were first tried on bR and then applied to other proteins. As a result bR to date is the most studied MP.

Besides the great significance for science, bR have the important practical applications. There are dozens of patents (quickly reviewed in [50]) and proposed applications of bR. This protein was used to construct prototypical devices of fourier-transform holographic associative processors and 3D volume data storage [17, 18] and photosensors [51]. Application of bR in constructing of artificial retinas [52, 53] seems to be natural. Also, bR was proposed to use as a molecular switch [54] in nanobiotechnology. This protein also was used in construction of solar cells [55]. Thus, the studies of bR will find a use not only in pure, but also in applied science.

2.2.2. Photocycle of bR.

After absorption of photon initial storage of light energy in bR occurs by the isomerisation of retinal molecule. From this excited state the sequence of reversible conformational changes (photocycle) occurs in bR and results in the effective transfer of proton across membrane. In such a way the light energy is converted into the electrochemical gradient across the membrane of archea. Afterwards, this energy can be used by ATP synthase to produce ATP. Thus, archea is one of the simplest systems performing according to the chemiosmotic theory of energy coupling proposed by P. Mitchell in 1961 [56]. Moreover, the other proton pumps, for example cytochrome *c* oxidase or ATP syntase, contain proton translocation mechanism that can have a universal features in nature. Simplicity of bR that realize a direct transfer of protons across membrane attracted attention of scientific community to the details of its photocycle that was in focus of MP research over a period of two decades.

The details of photocycle taking place in bR after absorption of quantum of light by retinal are illustrated on **Fig. 3**. To date thanks to the huge efforts of different groups all over the world the huge volume of information about bR photocycle was obtained using different techniques:

mutational studies, UV-Vis spectroscopy, FTIR, NMR, electron, neutron, and X-ray diffraction. All this data can not be cited here, but they were carefully reviewed in numerous articles.

As of today the general steps of proton transfer by bR are considered to be as follows. The photocycle starts from isomerisation of retinal after photon absorption from *all-trans* to *13-cis* state to form a K intermediate state within several picoseconds (13-cis 15-anti in some studies). The resulting steric clash induces the cascade of thermal structural rearrangements K↔L↔M_1↔M_2↔N↔O↔bR resulting in the transfer of bR back to the ground state coupled with effective proton transfer across membrane.

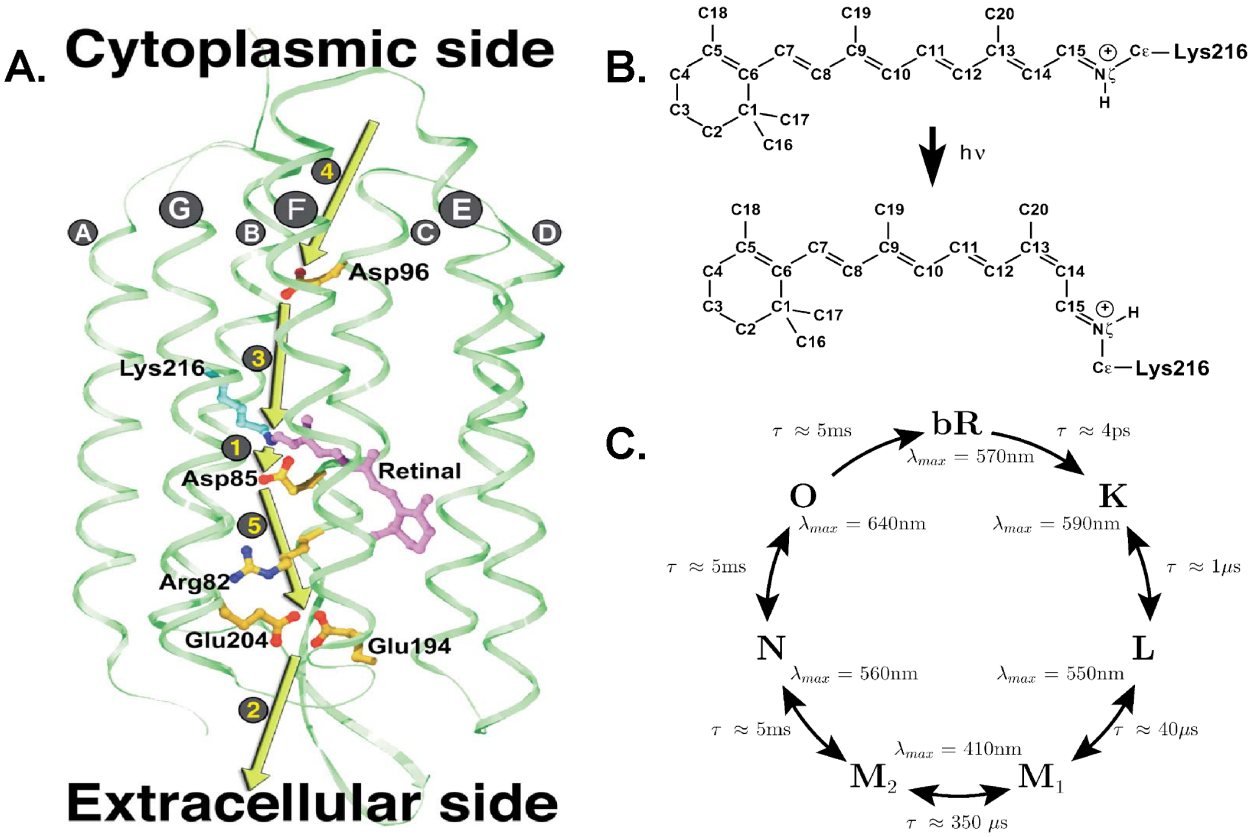

Figure 3: Photocycle of bR. **A.** Schematic representation of the proton transfer steps in the bR photocycle, overlaid on the ground state model (ribbon representation in green, showing helices A to G). Strategic residues that participate in the proton transfer (Asp96, Asp85, Arg82, Glu194, Glu204), and retinal bound to Lys216, are highlighted. The primary proton transfer (1) is from the Schiff base to Asp85. A proton is released to the extracellular medium (2) by the proton release group, thought to be Glu194 and Glu204. The Schiff base is subsequently reprotonated from Asp96 (3) which is then reprotonated from the cytoplasmic medium (4). The final proton transfer step (5) from Asp85 via Arg82 to the proton release group restores the ground state. **B.** The retinal chromophore is covalently bound to Lys216 via a protonated Schiff base. Following absorption of a photon the all-trans retinal is isomerized to the 13-cis configuration. **C.** The spectral intermediates, their absorption maxima, and their lifetimes at room temperature are shown. From [57].

The formation of blue-shifted L-state occurs in the time scale of μs and governs subsequent

protonation of Asp85. Main structural changes characterizing the L-state consist in the approaching of Asp85 to the protonated Schiff base combined with movement of C-helix, reorganization of retinal N-H bond of retinal, and changes in retinal binding pocket. During the transition from K-state to L the pK_a of Asp85 grows significantly allowing it to accept the proton from Schiff base. The primiry transfer of proton from Schiff base to Asp85 occurs during transition to M_1-state and takes around 40 μs. In the M-state itself two different spectrally silent states are distinguished: early M-state (M_1-state) and late M-state (M_2-state). The main changes in M_1-state occur in the extracellular part of bR, where water molecule W402 is disordered thus breaking the bond between Schiff base and Asp85. Along with that retinal adopts a relaxed 13-*cis* conformation. By this means the pK_a of Asp85 increases considerably and its protonation from Shiff base is not longer possible. Asp82 in this state is oriented to the extracellular side of bR that promote a proton release from, probably, Glu194 and Glu204. The transition from M1-state to M2-state takes around 350 μs. The main structural changes affect the cytoplasmic part of bR, where occur considerable shifts of F and G helices formed a hydrophobic plug. This plug prevented the back diffusion of protons through the membrane. The shifts of F and G helices form an opening towards towards cytoplasm, which is filled with several water molecules. Establishing of such formation decrease the pK_a of Asp96 and facilitate the subsequent reprotonation of the Schiff base. Thus, simultaneous increase of Asp85 pKa and decrease of Asp96 pKa form so called protonation switch. Some authors also claim that there is additional M_2'-state - late intermediate when the proton release occurs, but the general changes in the M-state are considered to be similar by all groups.

During the transition from M_2-state to N-state the reprotonation of Shiff base from Asp96 occurs. In O-state the Asp96 was reprotonated from the from cytoplasmic surface and the retinal molecule spontaneously reisomerised from 13-*cis* to all-*trans*. Then, water molecule W402 reorders in its position having the hydrogen bond with the protonated Schiff base that facilitate the reordering of Asp82 and Asp85. Finally, the proton release group is protonated from Asp85 to complete the photocycle.

Despite the general mechanism of the proton transfer by bR is known, there are dramatic controversies in fine details and some steps of the photocycle are still not clarified. For example, there is no structure of O-state of bR. The O-like structure of D85S mutant can not be accounted as a reliable source of information, because of considerable disturbance of the tertiary structure of bR which are not reported by other methods. Another important controversy is the scale of conformational changes in bR in M-state. While electron and neutron diffraction report the

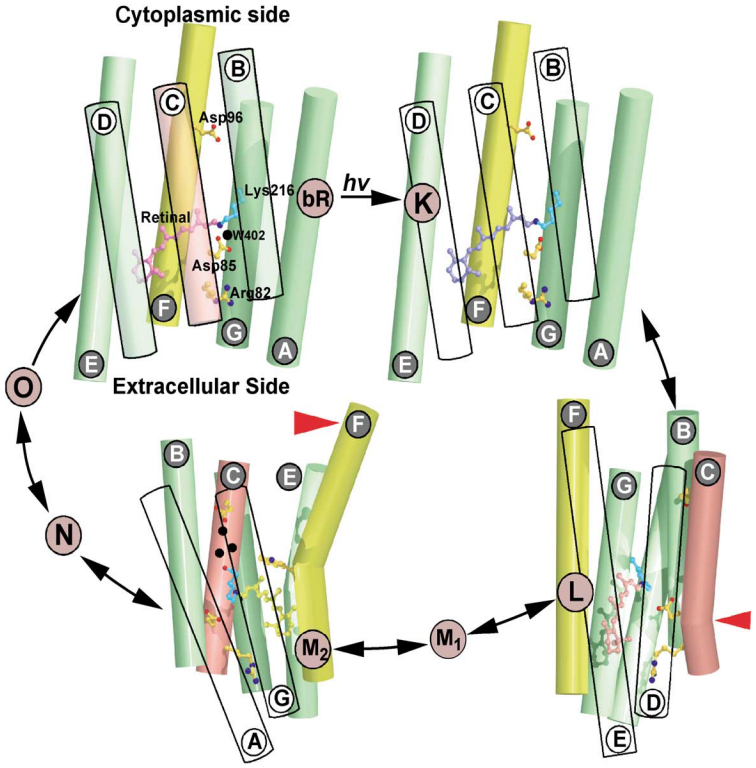

Figure 4: Schematic overview of the large conformational changes during the bR photocycle according to [57]. In the ground state of bR the protonated Schiff base nitrogen (blue) is H-bonded to Wat402 (black), which has a hydrogen bond to Asp85. A local flex of helix C in L-state (indicated by red arrow) allows Asp85 to approach the Schiff base enabling a proton transfer to Aps85. The reorientation of Arg82 towards the extracellular medium facilitates this bend of helix C, increases the pKa of Asp85, and governs the proton release to the extracellular medium. In M_2 an outward shift of the cytoplasmic part of helix F (indicated by red arrow) exposes key groups to the cytoplasmic medium allowing ordering of the water molecules (black) along the cytoplasmic half of the proton translocation channel. These structural changes allow the Schiff base to be reprotonated from Asp96, and then Asp96 to be reprotonated from the cytoplasm. Structural relaxation recovers the original ground state. For emphasis, helix C is colored red and helix F is colored yellow.

conformational changes up to 3.5 Å for F helix of bR, X-ray crystallography observed only 0.7 Å shift of the F helix. It is supposed that the crystal contacts in 3D crystal of bR restrain the conformational changes and, thus, do not reveal the real rearrangements of bR.

However, the most contradictory situation appeared to emerge around L-state structure. This is one of the most important states as it precedes the primary proton transfer from Schiff base to Asp85. However, there are several structures of L-state of bR that conflict with each other [16]. One

group was able to achieve the resolution of 2.1 Å [58, 59] using *in meso* grown crystals. The main structural changes in L-state are: 1) the shift of N-H bond of Schiff base to the cytoplasmic side of the protein; 2) disordering of the water molecule W402; 3) Asp85 approaches the Schiff base; 4) C helix is distorted to facilitate the approach of Asp85 to Schiff base. The proton transfer occurs directly from Schiff base to Asp85. Another group was able to produce the structure at 1.53 Å resolution using *in cubo* grown crystals [60, 61]. The main structural changes described are: 1) there is no considerable conformational changes in bR helices; 2) retinal is in exited distorted state; 3) the water molecule W402 is ordered and H-bonded to the Schiff base. In this model the proton transfer occures from Schiff base to Asp85 via water W402. The third model of L-state of bR [62] was solved at 2.4 Å resolution using the P622-symmetry crystals obtained by vesicles fusion. The features of this model are: 1) the shift of N-H bond of Schiff base to the cytoplasmic side of the protein; 2) the side chain of Leu93 is distorted providing a space to the water molecule W402 that is ordered and has a hydrogen bond to Schiff base. In this model the proton is transferred from Schiff base occurs via OH-group of Trp89. Thus, there are three different models of the L-state - the key intermediate to understanding the mechanism of proton transfer. Thus, further crystallographic studies of bR are required to understand the vital details of protein translocation by bR. For this purpose we should improve the established protocols for crystallization. Also, for the obtaining the structures of bR intermediate states we will require the fast and efficient system to produce different mutants of bR. In next chapters the current approaches to expression and crystallization of bR will be reviewed.

2.3 Crystallization of membrane proteins.

The first structure of MP was solved by H. Michel in 1984 [63]. Since that time several new approaches were introduced and to date there are four main approaches to crystallization of MPs. They are 1) *in surfo* crystallization; 2) in cubo crystallization; 3) crystallization by vesicle fusion; 4) bicelle crystallization. The first, *in surfo* method was introduce for soluble proteins and then adapted to MPs. The other methods were introduced especially for MPs and use lipids as a matrix for crystallization.

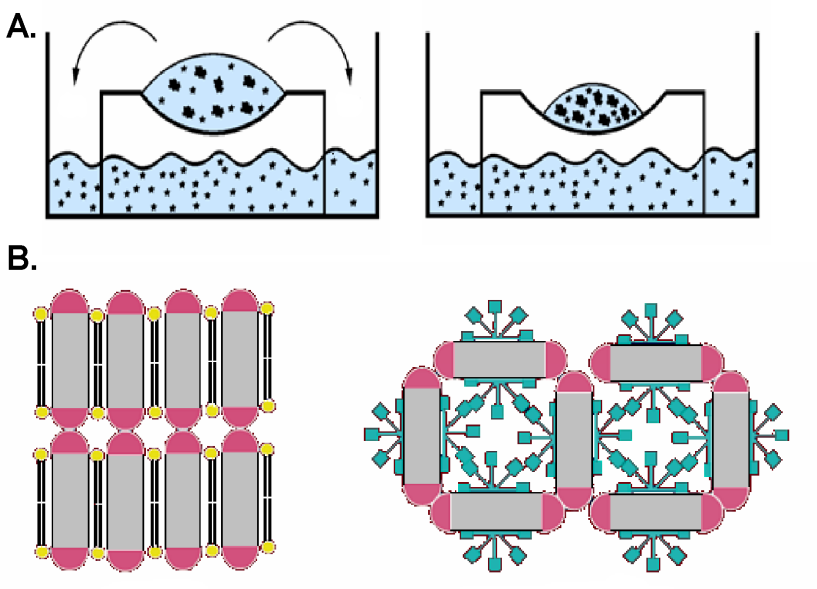

Figure 5: A. Illustration of the in surfo sitting drop method. The drop of the protein concentrated protein solution is placed into the reservoir with highly concentrated saline buffer. Then the system is sealed and during equilibration in the system crystallization of the membrane protein is possible. The picture is modified from [64]. **B.** The membrane protein crystal types according to Michel [65]. In type I crystals the protein molecules are organized in parallel layers and contact between proteins is established by in-plane hydrophobic surfaces, while in type II crystals the protein molecules interact by polar domains. The membrane proteins are presented as a cylinders with hydrophobic (grey) and polar (red) regions. The molecules of lipids are shown in yellow, detergent molecules - in green.

The first MPs were crystallized using *in surfo* approach. In essence, it is the conventional method to crystallize water soluble proteins (either sitting drop or hanging drop) adapted for crystallization of MPs directly from detergent (surfactant) solutions. Crystallization trials are organized as shown in **Fig. 5**. The concentrated solution of the protein solubilized in required detergent is placed on crystallization bridge. The bridge itself is located in reservoir filled with highly concentrated saline buffer. The system is sealed. During the equilibration of water activities in the protein drop and in the matrix solution water evaporates from the drop and the sample is concentrated. By that means the conditions in the sample can favor crystal nucleation and in such a case crystals can grow in the protein probe. This approach is a direct extension of the same methodology to the MPs and, thus, can utilize the vast range of technical advances in this area: there are automated crystallization dispensers, precipitant solutions, and crystallization additives. The purified protein samples also do not require any additional manipulations. These factors explain the popularity of this approach: to date most of the MP structures were solved using *in surfo* grown crystals.

However, despite the apparent facility and considerable advantages of this method, there are

considerable disadvantages of the *in surfo* approach. One of the main drawbacks is that crystallization occurs directly from protein solubilized in detergents. Detergent micelle is not a good substitute for the native lipid membrane. Solubilized MPs outside of native membrane tend to denature or aggregate. Another problem is associated with the presence of detergent molecules in the crystal. The *in surfo* grown crystals have type II crystal packing (**Fig. 5**) [65], where detergent molecule form a belt around hydrophobic region of MP as it was shown by neutron diffraction [66]. If the detergent micelle is big enough while the polar region is small enough, this micelle can cause a steric conflict leading to the inability to obtain well ordered crystals by *in surfo* method. It was exactly the case with bR. The first *in surfo* grown crystals were obtained in 1980 by H. Michel [13], however the diffraction was limited to 8 Å resolution. The polar domain of bR is extremely small, and it was not possible to obtain highly ordered bR crystals before Landau and Rosenbuch have not introduced a new method of MP crystallization 16 years later [6]. Moreover, the *in surfo* grown crystals tend to be of larger size than crystals grown by other methods, however, they are much sparser, so their diffraction power is lower. Thus, the crystallization of MPs using *in surfo* approach seems to be not optimal for crystallization of MPs.

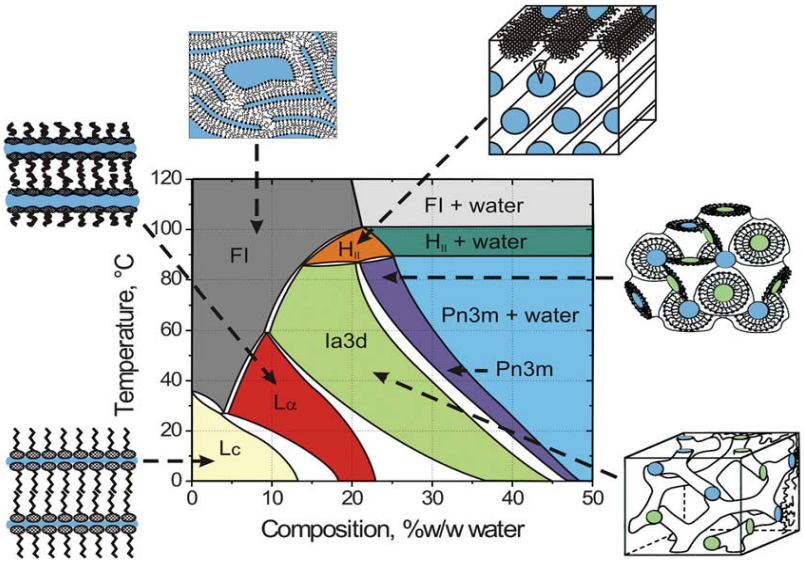

Figure 6: Temperature-composition phase diagram of the monoolein/water system determined under "conditions of use" in the heating and cooling directions from 20 °C. A cartoon representation of the various phase states is included in which colored zones represent water. Taken from [67].

To overcome the limitations of *in surfo* crystallization method in 1996 Landau and Rosenbuch introduced a novel *in meso* approach for crystallization of MPs [6]. Appealing feature of this approach is the use of amphiphilic lipid bilayer as a crystallization media in contrast to traditional *in surfo* method. The *in meso* approach is based on the use of lipid mesophases. Typically, monoglycerol monooleoyl (C18:1c9; monoolein, MO) is used as a host lipid for *in meso*

crystallization, the temperature-composition phase diagram of MO/water system is presented in **Fig. 6**. The mechanism of *in meso* crystallization is still not clearly understood, but proposed to be as follows. First, the solution of solubilized protein is added to the molten MO and the cubic phase Pn3m forms spontaneously within several days [6]. Initiation of crystallization by addition of dry salt dramatically decreases the lattice constant of the phase, i.e. shrinks the phase [68]. This increases the energy of the single protein in curved lipid bilayer, thus favoring crystal nucleation [69]. The addition of liquid precipitant, in contrast, leads to the swelling of the phase, thus destabilizing it [70] and promoting nucleation. The growth of crystal nucleus induce the transition of Pn3m cubic phase into Lα lamellar phase [71] that forms lipid conduit feeding the growing crystal from the bulk of cubic phase [72].

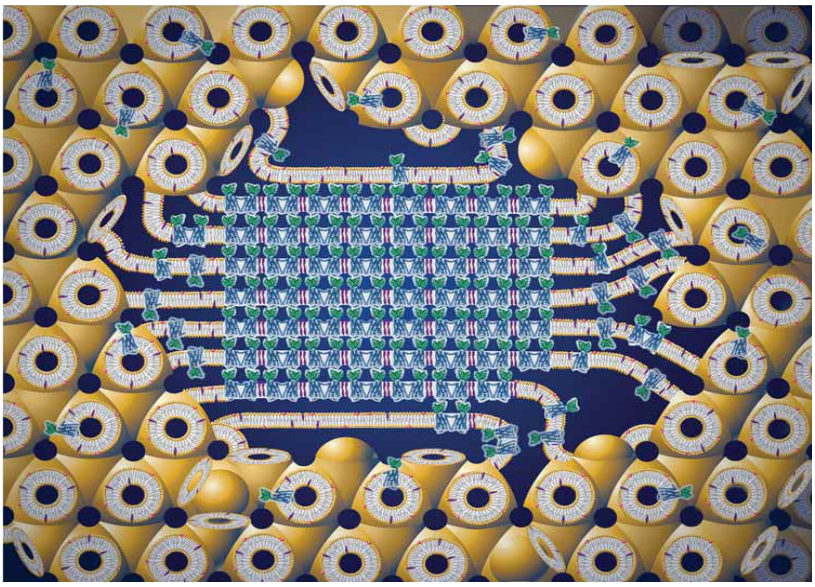

Figure 7: Cartoon representation of the events proposed to take place during the crystallization of an integral membrane protein from the lipidic cubic mesophase. The process begins with the protein reconstituted into the curved bilayers of the 'bicontinuous' cubic phase (tan). Added 'precipitants' shift the equilibrium away from stability in the cubic membrane. This leads to phase separation wherein protein molecules diffuse from the continuous bilayered reservoir of the cubic phase by way of a sheet-like or lamellar portal to lock into the lattice of the advancing crystal face (midsection of figure). Protein (blue-green), bilayer and aqueous channels (dark blue) have been drawn to scale. The lipid bilayer is 40 Å thick. From [70].

The *in meso* approach was introduced more than a decade ago and is credited with ~10% of all integral MP structures [72]. It was considerably developed over this time: the introduction of the crystallization screen solutions [73], nanovolume dispenser system [41, 74], and MP stability assay [75] considerably expand the experimental potential of this method, while controlling of *in meso* crystallization [76] may provide the ample grounds for rationalization of crystallization experiments. These step by step improvements finally led to a great success in structural biology: G-protein coupled receptor was crystallized for the first time using *in meso* approach in 2007 [34].

At the same time, the alternative lipid-based crystallization methods were developing. Except *in meso* approach there are two other methods that involve more native lipid environment for crystallization of MPs, namely, vesicles fusion and crystallization from bicelles. Both methods as well as *in meso* crystallization of bR give 3D crystals of type I (**Fig. 5**), where two neighboring MPs interact in the plane of the layer by hydrophobic contacts.

In case of vesicles fusion crystallization occurs from natural protein enriched membranes, for example, purple membranes of bR. The vesicles of 50 nm in size are formed during incubation of the isolated membranes with small amount of neutral detergent in presence of precipitant at elevated temperatures. When the vesicle solution is cooled down and concentrated, crystals can be obtained by sitting drop method [7]. The vesicles fusion method was introduced for of bR [7], but then the structures of halorhodopsin from *Natronomonas pharaonis*, archaerhodopsins from *Halorubrum sp. aus*, and deltarhodopsin-3 from *Haloterrigena thermotolerans* were obtained using this method. However, the 3D crystals were obtained in all these cases by utilizing naturally occurring protein enriched membranes, which are very rare case, therefore, a general applicability of this method is open to question.

The second method, crystallization from bicelles, employs the artificial disc-shaped bilayer micelles that are formed in a special mixtures of a lipid and a detergent and called bicelles [5]. Membrane protein is first incorporated into bicelles at low temperatures (+4°C), and then usual sitting drop crystallization trials are set with bicelles solution. The crystallization occurs in gel-like lipid phase, that forms at the elevated temperatures (room temperature or higher) in the probes. The advantage of crystallization from bicelles is that at low temperatures the crystallization media is liquid thus facilitating the handling of protein samples and crystals. The method seems to be general as the list of proteins crystallized from bicelles include different types of MPs: bR [5], mouse voltage dependent anion channel [77], xanthorhodopsin from *Salinibacter ruber* [78], rhomboid protease from *E. coli* [79], and even GPCRs (β_1-adrenergic receptor from turkey [80] and human β_2-adrenergic receptor [81]).

As a result of the enormous efforts in the field of MP crystallization the number of MP structures steadily grows (**Fig. 1**). However, still the count of the unique MPs is ~400, while only human GPCRs amount to 800 [82, 83]. Still there are no structures of many important therapeutically relevant human proteins, as well as proteins important for our general understanding of biological processes in living cells. Thus, the available methods of MP crystallization are far from what is required and recently appeared new promising approaches to crystallization of MPs

should be developed. The idea to use crystallization tags seems attractive.

It worth noting that the first structure of GPCR [34] was obtained using Lys as crystallization tag. Later several more GPCR structures were obtained using Lys by the same group [35, 37, 38, 84, 85] as well as by another groups [86-89]. In these cases Lys stabilized inherently flexible GPCRs and provided protein-protein crystal contacts to produce crystals. Crystallization tags and, in general, expansion of a polar surface of MPs was first employed to crystallize cytochrome *c* oxidase [90] in complex with antibody fragment and was addressed in further studies [91, 92]. For soluble proteins the use of fusion proteins was first apprehended with concern on the subject of their structural heterogeneity. However, the first high resolution crystal structures allayed these anxieties [93], and new crystallization tags were proposed (for example, [94, 95]). In the field of MPs the use of crystallization tags was first proposed in 1994 [96], then cytochrome(b562), flavodoxin, Lys, beta-lactamase, and heat shock ATPase domain were examined as crystallization tags [97], but without considerable success. Then aspartyl transcarbamylase was fused to bR [98] to try this water soluble protein as a crystallization tag, but the fusion protein appeared to be prone to aggregation and resisted crystallization trials. Thus, the idea of crystallization tags for MPs was neglected for a while and drawn considerable attention only after crystallization of human β_2-adrenergic receptor using Lys [34]. Though, the versatility of Lys for crystallization of MPs in general, despite massive crystallization of GPCRs, is still questionable because to date only GPCRs were crystallized using this crystallization tag. Accordingly, we are addressing this problem fusing Lys to bR. The choice of bR as a model protein has a considerable advantage due to the previous studies of this protein: as bR one of the most studied MPs, the enormous volume of information concerning expression, purification, crystallization, and functioning is available. Particularly important is that the structure of bR ground state is known, and thus comparison of the structures of bR crystallized with and without crystallization tag can help to establish the influence of crystallization tag on the structure of MP: is the structure native or disturbed by crystallization tag? On the other hand, there is a lot of uncertainty in the details of bR photocycle. There is a certain probability that bR-Lys fusion protein will give a crystals in a space group where the twinning is not possible. Thus, the crystallization of this fusion protein can be useful not only from the perspective of examination of the crystallization tag idea, but also in a context of investigation of bR itself.

Decided to crystallize the bR-Lys we should first establish the expression system to obtain sufficient amount of purified functional homogeneous bR-Lys fusion protein. Here, the information considering expression of wild type bR can come in handy. The most popular system for functional

expression of bR is by far the homologous expression in halobacteria. For example, in [99] the authors achieved in the transformants the yield of bR equal to yield from wild type strains, and the expression of bR mutants was established in the system. Employing an analogous homologous *H. salinarum* system along with the expression of wild type gene established the expression of fusion proteins [100]. Particularly, the bR-aspartyl transcarbamylase fusion protein was expressed and purified in this system [98]. The yield of functional fusion protein in non-denaturing contitions was 7 mg per liter of culture, and the protein formed purple membranes. However, our experiments (chapter 3.1) have shown that in case of Lys fusion the protein do not form purple membranes. In addition, the cells were not colored indicating that the high expression yield of functional fusion protein was not achieved. Therefore, we had to reconsider the choice of expression system for bR-Lys fusion protein. The functional expression in *E. coli* would be the most simple, robust, and inexpensive system [19]. The efforts to reach the expression of bR in *E. coli* last already almost 30 years.

2.4 The expression of bR in E. coli.

2.4.1. The use of exogenous *N*-terminal sequences to stabilize bR.

The first paper addressed the bR expression in *E. coli* appeared in 1987 [101]. As bR gene contains 13 amino acid long signal peptide that is very different from *E. coli* signal sequences and is unlikely removed by *E. coli* signal peptidase, the *H. salinarum* leader peptide was substituted for the signal peptide of *E. coli* outer membrane lipoprotein with additional amino acid sequences of three different lengths coming from lipoprotein. Resulting fusion protein amino acid sequences were inserted into pIN plasmid under control of *lac* repressor. The yield of expression of bR fusion proteins in *E. coli* was estimated as 0.05% of total protein (or, according to our estimations, 0.4 mg per liter of culture). The expression of bR was deleterious for cell growth depending on the *N*-terminal amino acid sequence. The level of bR mRNA was comparable with the level of bR mRNA found in *H. salinarum*. Taking into account the *N*-terminal degradation of the expressed protein and its low half life times *in vivo* (15-40 minutes) the authors explained the low yield of protein expression by extensive degradation of newly synthesized bR in cells. Nevertheless, the protein was solubilized by Triton X-100 from membranes and purified using immunoaffinity column of bR monoclonal antibody and preparative SDS-PAGE. Denaturated protein was refolded using standard DMPC/CHAPS protocol to the extent of 55%. The proton pumping activity of bR was shown.

In the subsequent paper [20] the bR gene (with one additional *N*-terminal methionine) was

introduced into pJP plasmid containing T5 promoter under control or without control of *lac* repressor. Despite the high level of transcription of the gene, expression of the protein was not detected. The variation of promoter, ribosome binding site as well as the spacing between Shine-Dalgarno sequence and the initiation codon had no effect on expression level of bR. Then, the codons of first six amino acids in bR gene were altered to introduce the codons preferred by *E. coli*. In case of controlled expression these changes had no effect on the bR yield, while uncontrolled expression led to the significant yield of bR. It is worth noting that the authors did not relate the changes of the first bR codons with the changes of mRNA secondary structure, but rather attended to the primary nucleotide sequence. The difference between controlled and uncontrolled expression was attributed to the palindromic sequence of *lac* operator in mRNA structure that could form a hairpin loop structure and have a negative influence on the ribosome-binding site immediately downstream. However, the uncontrolled expression of bR gene was extremely deleterious to the cells and led to the severe plasmid instability. The authors had to use expression system, P_L promoter controlled by temperature sensitive repressor. When the expression of bR gene was induced by temperature shift to 42°C, bR made 8-10% of newly synthesized protein. However, due to intensive degradation of the synthesized bR (half-life time was estimated as 8-10 minutes) only 30-40% of bR was stable and recoverable. Thus, the expression yield of bR was estimated as 0.5% of total protein (or, according to our estimations, 4 mg per liter of culture). Importantly, there appeared to be two phases in the turn over of bR, an early rapid degradation and then a long refractory phase where bR is rather stable. The protein was stabilized by the entire signal peptides of *E. coli* outer membrane protein A and β-lactamase, but the signal sequences were not cleaved providing 21- and 23 amino acids long *N*-terminally extended bR. The synthetic bR gene facilitating DNA manipulations gave the same bR yield as the gene with altered codons [102]. The protein was purified in several steps [103]. First, bR was extracted from *E. coli* membranes into chloroform/methanol-based solvent system. Then the protein was either purified using preparative SDS-PAGE or delipidated. Delipidation was achieved either by extraction of lipids using chloroform/methanol mixture or by gel filtration in chloroform/methanol-based solvent. The delipidated samples were purified using either hydroxylapatite or ion exchange chromatography in organic solvent. The protein transfer from organic solvents to SDS by phase separation followed by renaturation of the samples using DMPC/CHAPS standard approach. The efficiency of purification was 70-80%. The purity of the nonfunctional preparations was estimated as 90-99%, while the regeneration efficiency was 55-80%. Interestingly, the renaturation efficiency of bR from *H. salinarum* decreased from 95% to 85-90% when the protein was subjected to identical procedures (i.e. exposed to organic solvents and etc.). The proton pumping activity and light adaptation of bR

were shown. Thus, using the described expression and purification system for the first time it became possible to obtain the milligram quantities of sufficiently pure functional bR for structural studies.

In order to establish the functional expression of bR in *E. coli* the same authors investigated the insertion of expressed in *E. coli* bR into membrane *in vivo* modifying the *N*-terminal amino acid sequence [104]. The T7 RNA polymerase system under control of *lac* operator was used. The bR was expressed with one of the following modifications of *N*-terminus: an additional methionine, additional signal sequence of *E. coli* lipoprotein, additional signal sequence of *E. coli* outer membrane protein, or native bR gene with additional 13 amino acid long *H. salinarum* signal sequence. The degradation of the protein was estimated to be 80% for native signal sequence, 70% for protein with additional methionine, 20% in case of *E. coli* signal peptides. The expressed protein was located in both outer and cytoplasmic membranes. The *H. salinarum* signal sequence was not processed, and only 1-2% of both bR with and without *H. salinarum* signal sequence bounded retinal. The cleavage of lipoprotein and outer membrane protein was partial, 15% and 50% of total protein synthesized, respectively. The protein in the outer membrane was not processed at all, while in inner membrane the protein was processed to higher extent. Unprocessed protein did not bind retinal, but the fractions of the processed protein that bound retinal was 20% for both proteins. The authors suggested the presence of two differently folded states of bR, one of them binds retinal, while the other do not.

A new expression system [21] employed the stabilization of bR protein with *N*-terminal amino acid sequence. Nucleotide sequence encoding 13 heterologous residues derived partially from β-galactosidase were added to the *N*-terminus of bR under control of wild-type *lac* promoter, and the resulting fusion protein was expressed in *E. coli*. The protein expression was deleterious for cell growth. The half-life time of expressed bR was 26 minutes. The protein was found in cell membrane, inner membrane contained 82% of the synthesized protein. The protein was heterogeneous due to the modifications of the *N*-terminus. The expression level of bR was estimated as 5.8% of total cell protein or 17.6 mg per liter of cells.

The expressed bR was first solubilized in 2% SDS and then purified by multiple passes runs of high performance size exclusion chromatography [105]. Then the protein was renaturated using standard DMPC/CHAPS(O) approach followed by another size exclusion chromatography to remove residual SDS, lipids, and retinal. In order to remove unfolded bR and change detergent to nonylglucoside the dye-ligand affinity chromatography was performed. The efficiency of

purification was ~50%, the functionality of purified protein was supposed to be 100%. The proton pumping activity was shown. The described system was used to produce several mutants of bR. Finally, despite *N*-terminal heterogeneity of produced bR, using the purified protein reconstituted into *H. salinarum* native lipids the 2D crystals of bR were obtained in P3 symmetry. At room temperature the electron diffraction to a resolution exceeding 3.0 Å was observed and the projection difference Fourier maps were obtained for wild type protein and D96N mutant [39]. The same system was used to express bR-aspartyl transcarbamylase fusion protein [98], however the better yield and formation of purple membranes *in vivo* in *H. salinurum* prompted the authors to use halobacterial expression system.

Thus, the first progress in the expression of bR in *E. coli* was attributed to the stabilization of bR *N*-terminus with exogenous peptides to prevent degradation of the protein. This approach allowed to obtain the bR expression yields up to 18 mg per liter of culture. However, the functional expression of bR was not achieved, and the protein purification was carried out under denaturing conditions. As the result, the purified bR was heterogeneous and contained additional amino acids on the *N*-terminus.

In 1999 one more important work concerning expression of bR appeared [27]. The *C*-terminally *His*-tagged gene of mature bR was inserted into pET vector employing T7 RNA polymerase system under control of *lac* operator. The *His*-tagged bR was expressed in *E. coli* at 37°C. The protein was solubilized in DM from membranes isolated by ultracentrifugation and purified using metal affinity chromatography. Investigation of photocycle kinetics of the purified protein reconstituted into native *H. salinarum* lipids showed that the last stages of photocycle of bR expressed in *E. coli* are retarded comparing to homologously expressed bR. The total yield of the protein was estimated as 0.084 mg per liter of culture. The purity of the preparations was maximum 30% judging from UV-Vis spectrum. The important point is that the expression was carried out with retinal added and purification was performed under nondenaturing conditions. In next study [29] bR was used as a reference for expression of *Haloterrigena turkmenica* deltarhodopsin in the similar system as in [27]. It was mentioned that when cells are cultivated with addition of retinal they are not colored, however when the isolated membrane were solubilized in DDM, the suspension became purple. But this phenomenon was neither described in details nor examined in followed publications.

In addition, besides the modifications of bR gene by the short peptides used for expression of bR gene in *E. coli*, the utilization of fusion proteins to boost the expression yield of bR was examined as well.

2.4.2. *The use of fusion proteins for expression of* bR.

Besides the use of short *N*-terminal peptide derived from β-galactosidase to prevent the degradation of bR, the 1-47 and 601-1008 amino acid fragment of β-galactosidase was used as a fusion protein for expression of bR [24]. The synthetic bR gene was appended to this fragment via short linker containing IgA protease cleavage site. The gene of the fusion protein was clone into pLZ plasmid downstream the wild type *lac* promoter/operator region and expressed in *E. coli*. Induction of expression did not have any adverse effect on the cell growth, as the fusion protein was expressed as inclusion bodies. Simple washing of the cell lysate allowed to obtain 150-200 mg of insoluble fusion protein per liter of cell culture. Then the protein was cleaved by IgA protease in heterogeneous phase, and the bR was selectively extracted to the organic solvent. Next, the protein transfer to SDS aqueous solution followed by the renaturation of bR in phosphatidylcholine/CHAPS micelles. The proton pumping activity was demonstrated. Finally, the anion exchange resin was used to change the detergent to OG. The total yield of purified bR was 30-50 mg per liter of culture, however, the efficiency of retinalization was only 25%. The protein was not purified thereafter leading to the low quality preparations.

In the next study [22], the bR was fused to *E. coli* maltose binding protein without periplasmatic localization signal to target the resulting fusion protein to inclusion bodies. The synthetic bR gene with appended nucleotide sequence encoding 1D4 purification tag was inserted downstream the *MalE* gene of *E. coli*. The resulting maltose binding protein-bR fusion protein was expressed in *E. coli* under control of *tac* promoter. The expressed fusion protein was stable and accumulated in cells as inclusion bodies yielding 170 mg of protein per liter of culture. The inclusion bodies were isolated from cell lizate, washed and solubilized in 8M urea which was removed by dialysis. At this stage the purified fusion protein was in aggregated state, however it did not precipitate. The fusion protein contained Fxa cleavage site between maltose binding protein and bR and was cleaved by trypsin in presence of SDS. The proteolytic products were separated using gel filtration. Then, bR was refolded using standard DMPC/CHAPS approach followed by further purification on either ion exchanger or 1D4 immunoaffinity column. Finally, the dye-ligand affinity chromatography was performed to remove unfolded protein. Dark adaptation and photocycle of the protein were demonstrated. The final yield of the functional bR was assessed as 6-10 mg per liter of culture (the purification efficiency ~10-15%) with 100% of protein being functional. Using isoelectric focusing the purified bR was shown to contain two different protein species. The purified protein contained additional six *N*-terminal amino acids as well as optional nine amino acid long 1D4 purification tag

on the *C*-terminus. Keeping in mind the idea of crystallization tags the authors purified functional maltose binding protein-bR fusion protein. However, no further information concerning this matter was published.

Recently, a new fusion protein, MISTIC (Membrane-Integrating Sequence for Translation of IM protein Constructs), was used for expression of bR in *E. coli* [23]. The *His*-tagged MISTIC-bR fusion protein was inserted into pET plasmid with T7 RNA polymerase system under control of *lac* operator. The fusion protein was expressed in *E. coli* at 18°C and was found in membranes. It was solubilized in Sarkosyl from membranes isolated by ultracentrifugation. MISTIC-bR fusion protein was purified from crude membrane extract using metal affinity chromatography. Then the fusion protein was cleaved on-column at thrombin cleavage site between MISTIC and bR. Detergent exchange from Sarkosyl to SDS was achieved on column by subsequent washes containing required detergent. Eluted from column bR was refolded using standard DMPC/CHAPS approach. To remove the residual lipids and retinal and exchange detergent to nonylglucoside the second run of metal affinity chromatography was used. Proton pumping activity was demonstrated for the purified protein. The protein contains additional *N*-terminal amino acids (at least, Gly-Ser derived from thrombin cleavage site) and *C*-terminal *His*-tag. The authors claim the yield of functional purified protein to be 120 mg of purified functional bR per liter of culture with 100% of protein being functional.

2.4.3. Expression of another retinal proteins in *E. coli*.

Thus, despite that the different approaches were proposed for expression of bR in *E. coli*, the high yield functional expression of bR in *E. coli* was not achieved. The problem in expression of bR is even more surprising taking into account that there are many reports describing the functional expression of another retinal proteins in *E. coli*. Namely, sensory rhodopsin I and II from *H. salinarum* [25, 26], halorhodopsin and sensory rhodopsin II from *N. pharaonis* [27], bacteriorhodopsin from *Exiguobacterium sibiricum* [28], deltarhodopsin from *Haloterrigena* turkmenica [29], bacteriorhodopsins and sensory rhodopsins from *Haloarcula marismortui* [30], as well as two microbial rhodopsins from *Haloquadratum walsbyi* [106] and two rhodopsins from *Krokinobacter eikastus* (the first, KR1, being a prototypical proton pump, while the second, KR2, pumps sodium ions outward, [107]) were expressed in *E. coli* with the yield more than 1 mg per liter of culture. The proteins were folded *in vivo*, when cell cultures were grown in presence of retinal. In addition, in these cases the reasonable extent of purification was acheived under non-denaturing conditions by single step of metal affinity chromatography. So, why the other retinal

proteins do express functionally in *E. coli* and bR does not? Thus, one of the main objectives of the present study is to give the answer for this question.

2.4.4. The role of lipid/detergent environment in crystallization *in meso*.

The change of expression system from homologous *H. salinarum* to heterologous *E. coli* as well as the respective change of purification protocol can affect not only expression, but also crystallization. At first, the problem of detergent influence on the process of *in meso* crystallization appears. A lot of residual detergent derived from solubilization appears in crystallization probes. Large quantities of detergent were shown to destabilize lipid cubic phase [108]. The detergent was not observed in *in meso* grown crystals of bR by MALDI mass spectrometry [109]. However, the *in meso* grown crystals of bR were obtained using the protein solubilized from purple membranes in OG, but not in other detergents. Is OG use is crucial for successful crystallization of bR or one can use other detergents in *in meso* crystallization. Or mixture of detergents? Is the type of the detergent important or, maybe, its concentration? Does detergent have any influence on crystallization process at all?

Another even more important question is the influence of native lipids in crystallization of bR. Understanding that native lipids are important for expression, crystallization, and function of MPs steadily grows [40, 110]. The native lipids from *H. salinarum* were shown to play an important role in formation of purple membranes and *in meso* grown crystals of bR [109, 111, 112]. In addition, the only available crystals of bR expressed in *E. coli* (2D crystallization) were obtained from protein reconstituted into native *H. salinarum* lipids [39]. As the lipid composition of *H. salinarum* membranes is very different from lipid composition of *E. coli* membranes, bR or bR-Lys fusion protein expressed in *E. coli* and devoid of native *H. salinarum* native lipids can appear to be not crystallizable. Therefore, the question concerning the influence of native lipids on the *in meso* crystallization should be examined.

2.5 Conclusions.

The main purpose of the project is the X-ray crystallography studies of bR to clarify the appeared controversies in the mechanism of proton pumping. Such study requires sufficient amount of pure functional protein and new approaches to crystallization of the protein. We suggested that the functional expression of bR in *E. coli* would be the most simple, robust, and inexpensive approach satisfying these conditions. However, despite that the different protocols were proposed

for expression of bR in *E. coli*, the high yield functional expression of bR in *E. coli* was not achieved. The existing protocols for crystallization of bR are well developed, but twinning of the bR crystals obtained by *in meso* approach and insufficient ordering of crystals obtained by vesicle fusion do not allow to elucidate the mechanism of proton pumping by bR. Utilization of crystallization tags seems to be a promising approach to produce *in meso* highly ordered nontwinned bR crystals with Lys being one of the emerging candidates on the role of fusion protein. We proposed to use the *E. coli* system developed for functional expression of bR to produce functional bR-Lys fusion protein. Crystallizing bR and bR-Lys fusion protein one should also pay attention to the influence of lipid/detergent environment on the process of *in meso* crystallization. In spite of importance of this matter for success of crystallization, the role of lipid and detergent in crystallization *in meso* was not carefully explored. Thus, the main objectives of this study, namely, establishing of *E. coli*-based system for functional expression of bR and bR-Lys fusion protein and subsequent crystallization of these proteins as well as investigation of the influence of lipid/detergent environment on *in meso* crystallization, are relevant for structural studies of bR.

Moreover, bR is a model MP. Therefore, findings concerning development of the *E. coli* system for functional expression of bR can be useful for expression of other MP targets in *E. coli*. In addition, currently the versatility of crystallization tag approach was not examined as well as the influence of crystallization tags on structure and function of a crystallized protein. Since the structure of bR ground state is well established, expression and crystallization of bR-Lys fusion protein will allow to compare the structures of bR crystallized with and without crystallization tag and thus estimate a structural changes imposed by crystallization tag. Crystallization of bR without native *H. salinarum* lipids will allow to investigate the influence of native lipids on structure and function of MP. Thus, solving of the main problems formulated above will provide important information concerning expression and crystallization of MPs in general and thus facilitate the progress in the field of structural studies of MPs.

3 Results and discussion

3.1 H. salinarum expression

To explore the generality of crystallization tag idea using bR as a model MP and obtain in another space group crystals of bR that may be twinning free and, therefore, may help to obtain precise atomic structure of bR one should first obtain the functional bR-Lys fusion protein in amounts sufficient for crystallographic studies. The *H. salinarum* based homologous expression system is preferable because, in spite of the difficulties of work with archaea, this system [113] can give the yields of functional bR mutants up to 30 mg per liter of culture (wild type yield, [99]) and of fusion proteins [98] up to 7 mg per liter of culture. We used the system of pEF191 *Huloferux volcunii* - *Escherichia coli* shuttle vector that bear bR coding sequence with short flanking untranslated regions of native bR gene and bR-deficient *H. salinarum* L33 strain bearing impaired bR gene [99]. Transformation of L33 cells with pEF191 plasmid results in incorporation of recombinant bR gene into homologous bR locus and leads to the production of native bR. We have introduced Lys coding sequence to the pEF191 shuttle vector at two different positions: between Met163 and Arg164 of EF-loop and to the *C*-terminus of bR that gave the bRL and bRLE fusion proteins, respectively.

The plasmids pEF191bRL and pEF191bRLE containing coding sequences of fusion proteins between bR and Lys were constructed as described in Material and methods. First colonies appeared 15-20 days after transformation of L33 strain of *H. salinarum* with the plasmids. Colonies were separated and analyzed for the presence of Lys gene. The cells transformed with pEF191bRLE plasmid did not show Lys coding sequence in genome. We suppose that this is spontaneous mutants with mevinoline resistance as it was reported before [99] and that genetic recombination did not occur properly in this case. The reason might be the disturbance of structure of bR gene by Lys insertion on *C*-terminus of bR. In previous study [114] severe deterioration of expression yield of wild type bR with *C*-terminal *His*-tag was shown in archaea.

Both wild type bR and bR-Lys fusion protein genes were present in the genomic DNA of the cells transformed with pEF191bRL plasmid. Clones were separated by several passages of cells on agar plates with antibiotic. On the second passage cells showed different morphology: some colonies had typical for *H. salinarum* L33 yellow color while the others appeared to be white. We suppose that in white colonies the genetic recombination have successfully occurred and bRL gene incorporated into genome L33 cells.

The cells transformed with pEF191bRL plasmid were cultivated in liquid media. The culture did not gain purple color characteristic for the cells that express functional bR on the high level and form purple membranes. Therefore, we can conclude that insertion of Lys into E-F loop of bR prevent formation of purple membranes in archaea, and functional expression of the fusion protein, if present, is on the low level. In previous study [98] aspartyl transcarbamylase fusion was shown to impair threefold the expression of bR in *H. salinarum*, however the synthesized protein amounted to 7 mg per liter of culture and purple membranes were formed. Our findings suggest that ability of bR for functional accommodation of massive insertions is limited. As the high level of functional bR expression along with formation of purple membranes facilitating the purification of the target protein was supposed to be the main advantage of the *H. salinarum* expression system [11], the absence of purple membrane formation make archaea not promising for expression of bR-Lys fusion proteins.

3.2 MISTIC constructs.

3.2.1. Expression and purification of MISTIC fusion proteins.

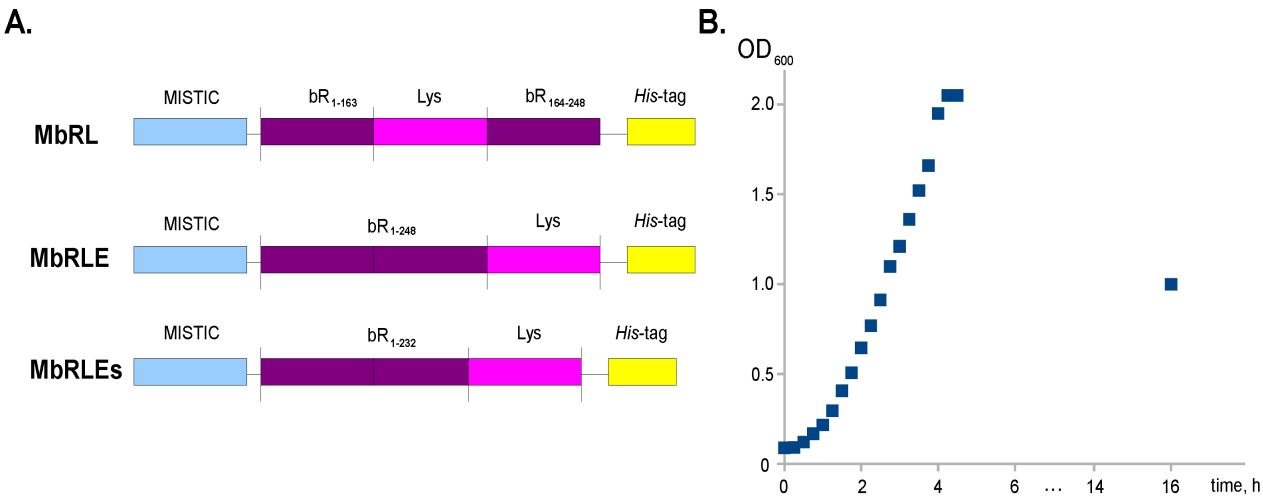

Figure 8: A. The schematic presentation of the MbRL, MbRLE, and MbRLEs fusion proteins. MISTIC is shown in pale blue, bR in violet, Lys in pink, and *His*-tag in yellow. The truncated bR in construct MbRLEs contains amino acids from 1 to 232. In MbRL construct Lys is inserted between Met163 and Arg164. **B.** The plot illustrating the optical density OD_{600} of *E. coli* SE1 cells bearing the gene of fusion protein with functional Lys throughout the culture growth. After overnight incubation of the culture most of the cells were dead and the cell pellets were soft and mucoid.

When *H. salinarum* system have not shown an appropriate expression of bR-Lys fusion protein, we have decided to change expression system. One of the most well-developed, facile, and inexpensive expression systems to date are *E. coli* based systems [19]. Expression of bR in *E. coli* was tried with a limited success many times (see chapter2.4for review). MISTIC (Membrane-

Integrating Sequence for Translation of IM protein Constructs) have revealed itself as a promising fusion protein, as it boosts expression of bR and other MPs to high levels and targets them to the *E. coli* membrane [23, 115].

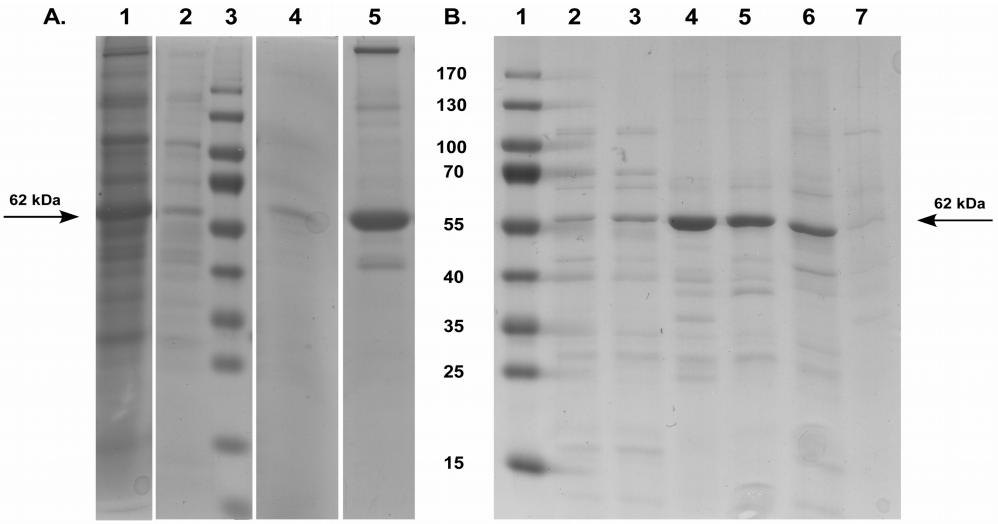

Figure 9: **A.** SDS-PAGE analysis illustrating the progress of the purification under denaturing conditions of MISTIC-bR-Lys fusion proteins. The MbRL purified in Sarkosyl is presented as example. 1 - cell suspension before solubilization; 2 - fraction solubilized in Sarkosyl; 3 - Protein Ladder; 4 - Ni-NTA column wash; 5 - Ni-NTA column eluate. **B.** The screen of detergents for solubilization of MISTIC-bR-Lys fusion proteins with MbRLEs as example. The lanes 2-7 on SDS-PAGE are eluates from Ni-NTA column in different detergents. 1 - Protein Ladder; 2 - DM; 3 - DDM; 4 - Sarkosyl; 5 - Fos12; 6 - LDAO; 7 - SDS. Band corresponding to the target protein is indicated.

The MISTIC gene was amplified by PCR from PSCodon-MhCx26 obtained previously in our laboratory and added to the *N*-terminus of the bR-Lys fusion proteins. We have constructed three variants of MISTIC-bR-Lys fusion proteins: MbRL, MbRLE and MbRLEs (**Fig. 8**). Fusion protein MbRL contains Lys inserted into E-F loop of bR between residues Met 163 and Arg 164. Fusion proteins MbRLE and MbRLEs contain Lys appended to the *C*-terminus of full-length bR and bR truncated at Glu 232, respectively. The MISTIC-bR-Lys fusion proteins were inserted into pSCodon vector, which uses the T7 RNA-polymerase expression system [116] with a separate-component-stabilization system [117] and contains genes of tRNAs for rare codons of *E.* coli.

First experiments showed that, when MISTIC-bR-Lys fusion proteins are expressed in *E. coli*, Lys is properly folded and displays its hydrolytic activity. The overnight grown induced cells achieved the OD = 1.0÷1.5 comparing to the OD = 3.0÷5.0 for the cells that harbored control plasmids without functional Lys gene and were grown under same conditions (**Fig. 8**). The cell pellets were soft and mucoid indicating the destruction of the cellular membranes during cultivation.

Overnight culture was harvested by centrifugation, resuspended and lysed using microfluidizer. The membranes were isolated by ultracentrifugation. The solubilization test included 6 different

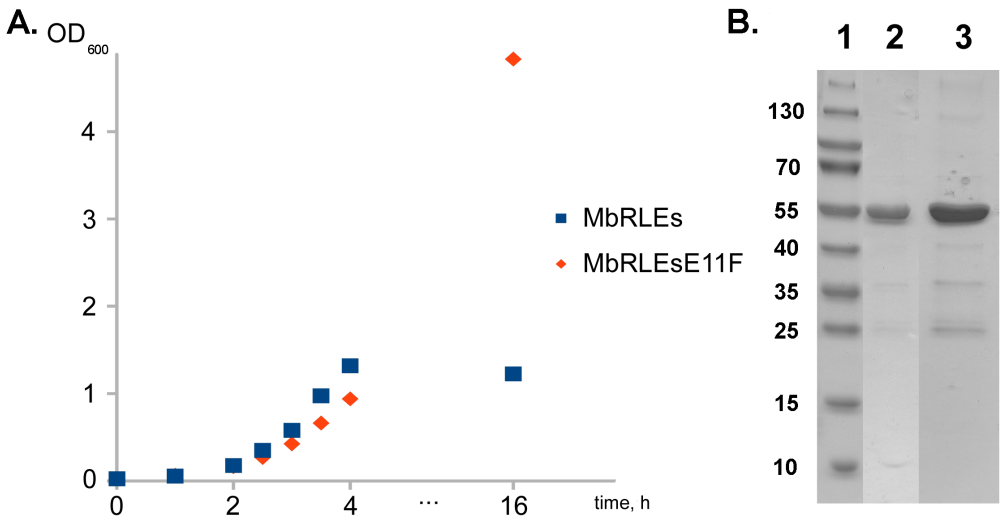

Figure 10: Expression of the MISTIC-bR-LysE11F fusion protein with abolished enzymatic activity of Lys. **A.** The plot illustrating the kinetics of culture growth of *E. coli* SE1 cells bearing fusion protein gene containing Lys with abolished enzymatic activity comparing to the cells bearing fusion protein gene with functional Lys. After overnight incubation of the culture the cells reached 5-fold higher OD_{600} and the cell pellets were hard in contrast to the cells with functional Lys gene. **B.** SDS-PAGE illustrating the difference in protein yields of MISTIC-bR-Lys fusion proteins with functional and with nonfunctional Lys. 1 - Protein Ladder; 2 - MbRLEs Ni-NTA column eluate; 3 - MbRLEsE11F Ni-NTA column eluate.

detergents: DM, DDM, Sarkosyl, Fos12, LDAO and SDS. After overnight solubilization the suspensions were clarified by centrifugation, diluted, and loaded on Ni-NTA resin. The columns were washed with buffer containing 20 mM imidazole, and the target proteins were eluted in buffer containing 200 mM imidazole. The progress of the protein purification is presented on the **Fig. 9**. The SDS-PAGE analysis have shown that MISTIC-bR-Lys fusion proteins are not soluble in DM and DDM, while LDAO, FOS12, and Sarkosyl are the most suitable detergents, as they allow to solubilize majority of the target protein (**Fig. 9**). SDS gave the lowest purification yield because the solubilized protein did not bind to the column. The FOS12, DDM and Sarkosyl were considered as the best detergents: FOS12 and Sarkosyl provided the best overall yield of purification, while DDM is compatible with Fxa protease and, thus, do not require detergent exchange. The SDS-PAGE analysis did not reveal presence of the target protein in column flowthrough and wash for the selected detergents. The yields for these detergents were about 3 mg (for FOS12 and Sarkosyl) and 0.3 mg (for DDM) of purified protein per liter of culture as quantified by coomassie-stained SDS-PAGE.

To increase the total yield of the protein we abolished the activity of Lys. The E11F mutation of lysozyme was previously reported to eliminate its enzymatic activity [118]. We have introduced

E11F mutation into Lys coding sequence of MbRLEs with wild type Lys. The E11F mutant was expressed and purified by the same protocol as MbRLEs fusion protein. The final OD of the overnight grown cell culture was 4.8 in comparison to the OD 1.2 of the cells contained plasmid with functional Lys coding sequence (**Fig. 10**). The cell pellets were solid indicating the inactivation of the Lys. The introduction of the E11F mutation increased the expression yield threefold from 3 to 10 mg per liter of culture as seen from the **Fig. 10**.

3.2.2. Cleavage and renaturation of MISTIC fusion proteins.

Cleavage.

The purified MbRL, MbRLE and MbRLEs fusion proteins were concentrated to 0.25 mg/ml and dialyzed against Factor Xa cleavage buffer with 0.2% DDM, as these detergent and protease are compatible. After dialysis, 5 µg of each protein was incubated 24 h at 20°C with 2.0, 0.4, 0.08, 0.016 and 0.0032 units of protease in 20 µl aliquotes. Results of the cleavage screening are

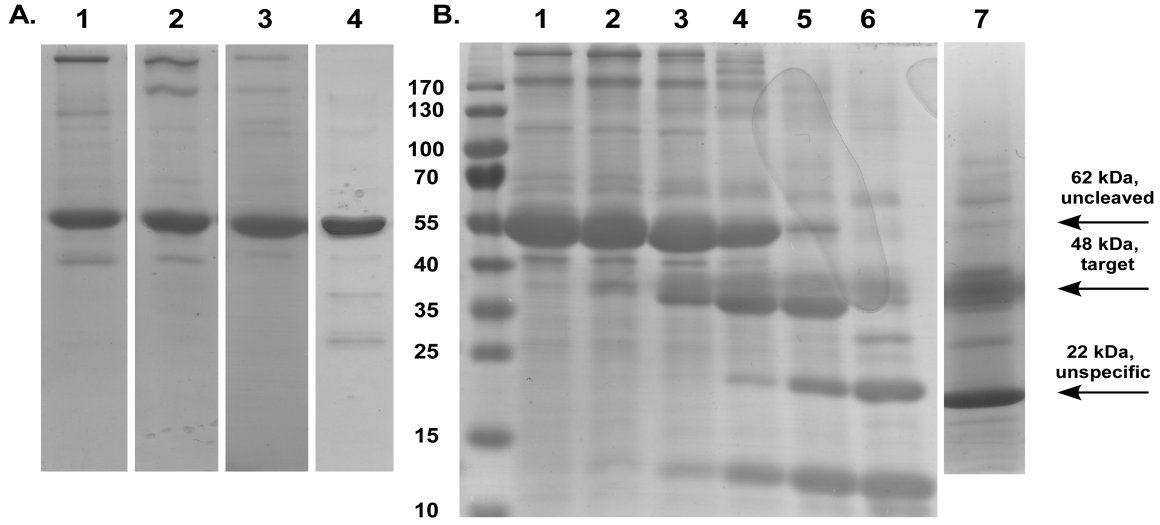

Figure 11: SDS-PAGE illustrating the purity of the preparations before cleavage and screening of conditions for the protease Factor Xa cleavage. **A.** The eluates after Ni-NTA column were used to access the purity of the preparations: 1 - MbRL; 2 - MbRLE; 3 - MbRLEs; 4 - MbRLEsE11F. **B.** The protease digestion of MbRLEs fusion protein under different conditions: lanes 1 - 6 illustrate different protease concentrations (units of protease per 5 µg of protein), line 7 - prolonged time of incubation with optimal concentration (0.4 units per 5 µg of protein) of protease. 1 - Protein Ladder; 2 - 0.0032; 3 - 0.016; 4 - 0.08;5 - 0.4;6 - 2.0; 7 - 40 h incubation of the reaction (instead of 16 h). The bands corresponding to the target cleaved, not cleaved, and nonspecific products are indicated.

presented on **Fig. 11**. Surprisingly, Factor Xa did not show considerable activity at recommended pH 6.5, but at pH 8.0 the activity of the enzyme was acceptable. For constructs MbRL, MbRLE and MbRLEs the optimal cleavage was observed at a ratio of 0.4 units per 5 µg of protein. SDS-PAGE analysis have clearly shown the presence of the cleaved target bR-Lys fusion protein, not cleaved

MISTIC-bR-Lys fusion protein and unknown product of unspecific cleavage, corresponding to the band at 22 kDa on acrylamide gel in 7:1:2 ratio, respectively (as quantified by SDS-PAGE densitometry, **Fig. 11**). Prolongation of incubation time with the optimal protease concentration to the 48, 72 or 96 h led to the complete degradation of the protein (**Fig. 11**).

Purification of cleavage mixture.

To separate the target cleaved product from the impurities we used size-exclusion chromatography on Superose 6PG column with 180 ml bed volume. The elution profile had a clear peak at 124.4 ml corresponded to the uncleaved protein, target and unspecific cleavage products as followed from SDS-PAGE analysis (**Fig. 12**). Thus, the gel-filtration was proved to be inefficient for separation of the target fusion protein from impurities.

Ion-exchange chromatography could be another useful method to separate fragments after

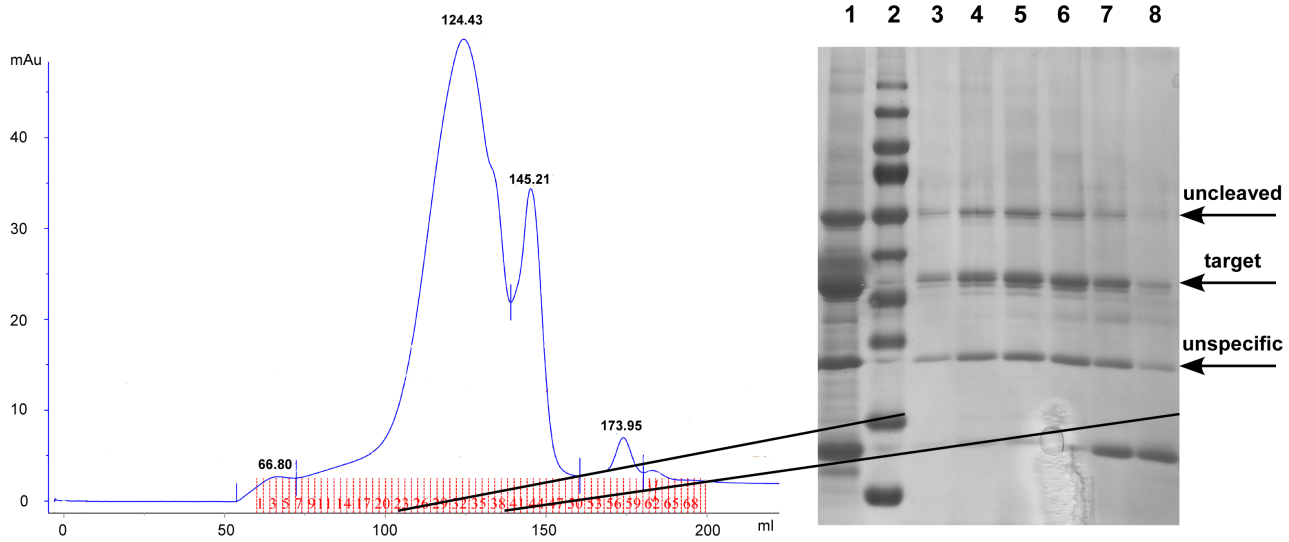

Figure 12: Separation of the Factor Xa cleavage products by gel-filtration. Elution profile shows two distinct peaks. The corresponding fractions were analyzed on SDS-PAGE: 1 - the sample after proteolysis before gel-filtration; 2 - Protein Ladder; 3 - fraction 23; 4 - fraction 28; 5 - fraction 33; 6 - fraction 35; 7 - fraction 37; 8 - fraction 40. Fractions are indicated in red on elution profile. Fractions from 23 to 38 correspond to the first peak, fractions from 39 to 46 correspond to the second peak. The uncleaved, unspecific and target products of proteolysis are indicated. SDS-PAGE clearly shows that the target protein can not be separated by gel-filtration from the other products of reaction.

proteolysis. Despite relatively low overall charge of the total MISTIC-bR-Lys fusion protein at neutral pH, the protein is strongly polarized. The fusion protein has an isoelectric point (pI) of 6.68, while pI are 4.57, 5.55, and 9.76 for MISTIC, bR, and Lys, respectively. We used Sepharose High Performance strong cation exchange column that was supposed to bind Lys positively charged at pH 7.0. The elution profile had two distinct peaks at 23.76 and 28.90 ml as seen on **Fig. 13**. According to SDS-PAGE analysis, the three proteins (uncleaved protein, target and unspecific cleavage

products) bound to the column and were eluted as a single peak what did not allow to separate the target protein from contaminants.

Renaturation of uncleaved protein.

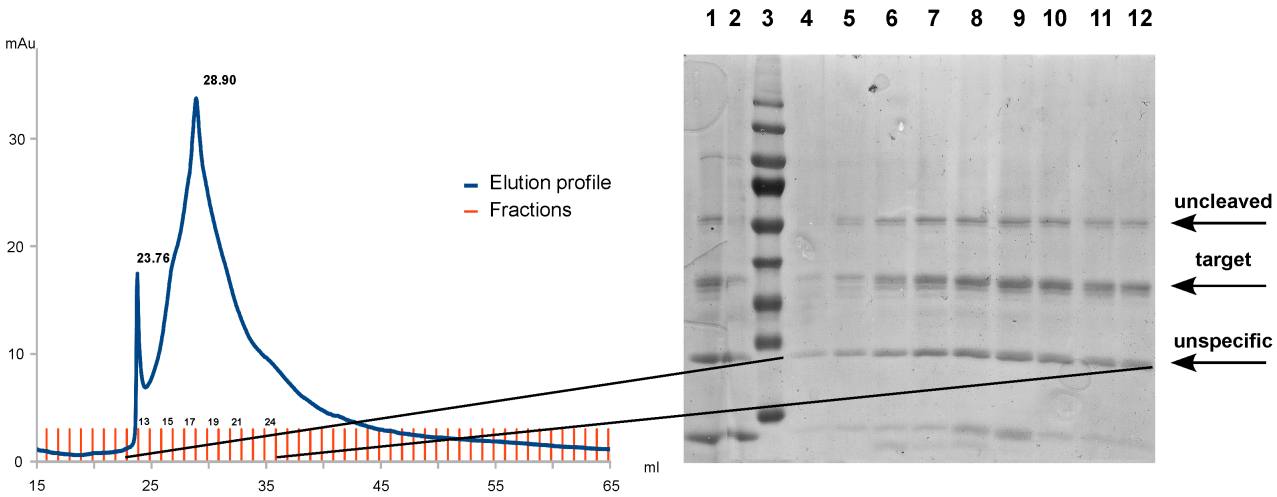

Figure 13: Separation of the Factor Xa cleavage products on ion-exchange column. Elution profile shows two distinct peaks. The corresponding fractions were analyzed on SDS-PAGE: 1 - the sample after proteolysis before column; 2 - column flowthrough; 3 - Protein Ladder; 4 - fraction 12; 5 - fraction 13; 6 - fraction 15; 7 - fraction 16; 8 - fraction 17; 9 - fraction 18; 10 - fraction 19; 11 - fraction 21; 12 - fraction 22. Fractions are indicated in red on elution profile. Fractions from 12 and 13 correspond to the first peak, fractions from 14 to 25 correspond to the second peak. The uncleaved, unspecific and target products of proteolysis are indicated. SDS-PAGE clearly shows that the target protein can not be separated from the other products of reaction by the used ion-exchanger at chosen conditions.

The reason for the presence of the different proteolytic products could be the inherent heterogeneity of the fusion protein denatured in strong detergents that give a rise to different accessibility of the cleavage site of protease in different molecules. If we managed to refold the protein to the functional state first, then it might become homogeneous and might not give the unspecific cleavage sites. The most simple and effective system to refold bR is DMPC/CHAPS bicelles [119].

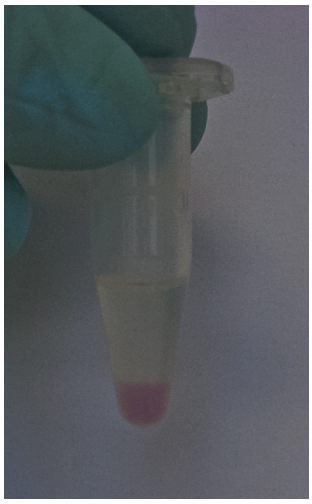

Figure 14: The pink pellet in the renaturation sample of MbRLEs denatured in SDS. The renaturation was carried out according to standard procedure of bR renaturation in DMPC/CHAPS vesicules. After overnight incubation the protein was found in the pink pellet.

The MbRLEs fusion protein solubilized in DDM, FOS12, or Sarkosyl was purified as described. Then samples were concentrated to 0.25 mg/ml and dialyzed overnight with 3 changes of buffer against 50 mM Na/Na-Pi pH 6.0 supplied one of the following detergents: 0.2% DDM, 0.2% FOS12, 0.2% SDS (the protein for this sample was solubilized in Sarkosyl), or 0.5% Sarkosyl. Then, the equal volume of DMPC/CHAPS bicelles contained 40 μM of *all-trans* retinal was added and samples were left at room temperature with stirring for retinalization. After overnight incubation the samples with DDM, FOS12, and Sarkosyl did not change color indicating that retinal did not incorporate into bR in those samples. However, SDS sample turned pink demonstrating that renaturation of bR-Lys fusion protein was achieved. To remove redundant detergent and, thus, improve the incorporation of the protein into lipids we added to the samples Bio-Beads adsorbent assuming its capacity as 70 mg of detergent per 1 g of beads. But no visible changes occurred in the samples. The samples in SDS after retinalization contained pink pellet (**Fig. 14**), probably being the liposomes with incorporated properly folded bR fusion protein. We have tried to resolubilize the pellet in either DDM or FOS12, but 1.5% DDM did not solubilize the protein, while 1% FOS12 denatured the protein leading to the loss of color by MbRLEs fusion protein.

Renaturation of cleaved protein directly from proteolytic mixture.

Failed to renature the MbRLEs fusion protein tried to alter purification strategy. First, we cleaved this fusion protein and then tried to renature the bRLEs fusion directly in the proteolytic mixture without separation of the products. However, when the proteolytic mixture was dialyzed against renaturation buffer, we observed the heavy precipitation of the bR-Lys fusion protein. We suggest that MISTIC artificially maintained the solubility of the bR-Lys fusion membrane protein, but when MISTIC was cleaved the fusion protein lost its solubility and precipitated.

Conclusions.

Thus, the resulting fusion protein turned out to be extremely polarized being difficult to handle. Despite the high yield achieved by use of MISTIC this fusion tag prevents bR-Lys fusion protein from proper folding *in vivo* and *in vitro*. After cleavage of MISTIC unfolded protein seems to loose immediately its solubility at pH it should be functionally refolded. Moreover, the employment of the strong detergents to solubilize MISTIC probably led to the inherent heterogeneity of the protein and unspecific cleavage by Factor Xa protease. The proteolytic products were difficult to separate aggravating the situation. Therefore, the strategy employed the simultaneous use of two different fusion tags, MISTIC as expression tag and Lys as crystallization tag, did not pay for itself. Having all these problems concerning the use of two fusion proteins for expression of bR-Lys fusion protein we decided to avoid such strategy in the further development of the project.

As native bR gene is notoriously known for its low expression yield in *E. coli* [20], we intended first to find out the reason of this phenomenon and develop the expression system that will allow the functional expression of bR in *E. coli* with a high yield. Thereafter, utilizing the bR expression system we might be able to produce functional bR-Lys fusion protein without need for cleavage of the fusion tags or renaturation of the target protein.

3.3 Complementary protein approach.

Throughout the years the functional bR expression was tried many times. Application of exogenous *N*-terminal tags allowed to stabilize the newly synthesized protein preventing its degradation and increasing bR yield [20, 21]. Further progress was related to the use of the fusion proteins that boosted the expression level [22-24]. However, despite the high yield of the synthesized bR mentioned systems did not provide functional bR expression: the protein should have been first undergone unfolding followed by refolding.

Such behavior of bR is all the more surprising, as the expression of a number of closely related retinal proteins, namely sensory rhodopsin I and II from *H. salinarum* [25, 26], halorhodopsin and sensory rhodopsin II from *N. pharaonis* [27], bacteriorhodopsin from *Exiguobacterium sibiricum* [28], deltarhodopsin from *Haloterrigena turkmenica* [29], bacteriorhodopsins and sensory rhodopsins from *Haloarcula marismortui* [30], and others, was shown to be functional in *E. coli*. Why the bR does not express functionally in *E. coli*? What is the difference between bR and another retinal proteins that allows their functional expression in *E. coli*?

In present study we have introduced the complementary protein approach that allowed us to localize and solve the problem of bR low expression yield in a few steps. Schematically this approach is presented on **Fig. 15**. Given two homologous proteins, the first one, the complimentary protein, gives the high expression yield in the chosen expression system in contrast to the other, target protein, one can construct chimeric proteins containing different parts of the proteins. If one of the constructed proteins has high expression level comparing to the protein under study, it can be proceeded further. In the second iteration of the problem finding the insert from the complimentary protein is divided into parts and each part is substituted to the counterpart from the original protein resulting in the new pair of chimeric proteins. Analyzing the expression yields of these chimeric proteins one can identify the part impairing expression. In the same manner we can finally reveal the underlying problem that do not allow the expression of the target protein in the selected organism.

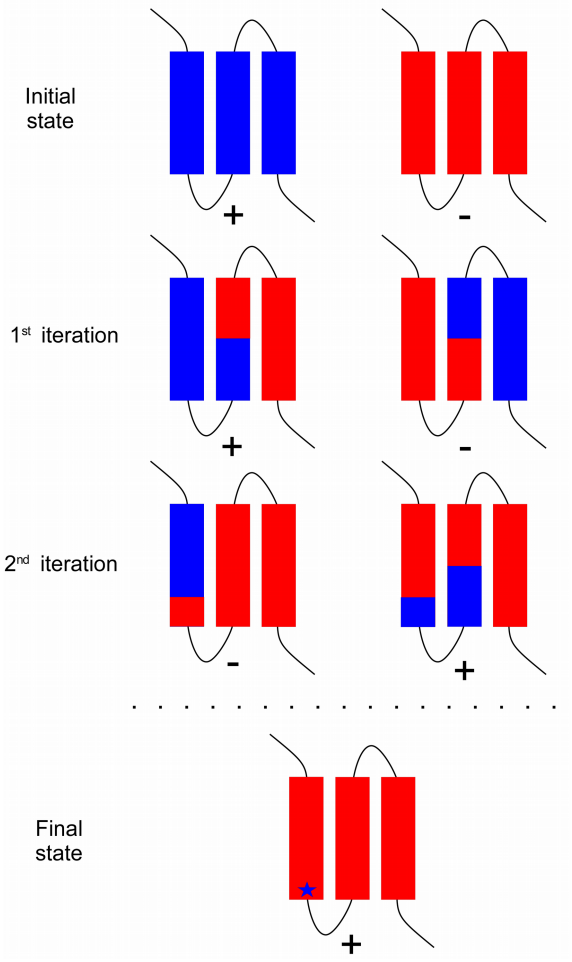

Figure 15: A general representation of the complementary protein approach. Given two homologous MPs with high (blue) and low (red) level expression in a selected organism, one can construct chimeric proteins (1st iteration) comprising the part of target protein (the red one) and complementary protein (the blue one). Analyzing expression yields of the chimeric proteins one can conclude which part of the protein is responsible for a heterologous expression failure. In the next step, one can divide the new insert of the complementary protein to parts and construct next pair chimeric proteins (2nd iteration) to localize more precisely the problem place. Acting in the same manner one can finally localize the problem of lack of high yield expression of target protein in a limited number of steps. Such approach allow one to reduce the required number of genetic constructs from 2^N point mutations to $\sim 2 \cdot \log_2 N$ chimeric proteins, where N is the number of amino acids in the target protein.

This approach has a limited application requiring the existence of the localizable "point" reason for a protein expression fail. If the lack of expression is caused by extensive factor influencing the vast area of the protein (for example, positive charge of the extracellular face of the protein or commonly encountering rare codons), the complimentary protein approach will not be able to cope

with the problem. However, the closer is the resemblance of the complimentary protein to the protein under study, the higher should be the probability of success. Additional rational considerations related to the target protein and chosen expression system can increase the chances of progress.

In a favorable case the described approach might lead to the issue finding in 10 steps (~$\log_2 N$, where N is the number of amino acids in the target protein) for the protein of a typical size less than 1000 amino acids accounting the finite set of amino acids in the protein. Therefore, the localization of the expression problem can be accomplished by constructing only $2 \times 10 = 20$ chimeric proteins instead of 2^N mutant proteins (the number of possible single mutations if this mutations substitute amino acids from target protein for amino acids from complementary protein). Moreover, additional information and rational considerations about the protein structure or the expression system used can decrease the number of required chimeric proteins. Even if the described approach failed to perform at certain step, the information obtained on previous iterations can help to narrow the area of problem search. Thus, we suppose that the complementary protein approach can have a general application.

In the present study we addressed the problem of low yield expression of bR in *E. coli* using complementary protein approach. Previous publications, and particularly [104], concerning this problem gave us an idea that the low yield of bR expression and the lack of functional expression are aroused by the ineffective incorporation of the newly synthesized protein into *E. coli* membrane. We have applied the complementary protein approach to the bR and used the SRII as a complementary protein, since this is the only retinal protein expressed functionally with the high yield in *E. coli* [27, 31] for which the high resolution 3D crystal structure was available to the moment we have started this study [32, 33].

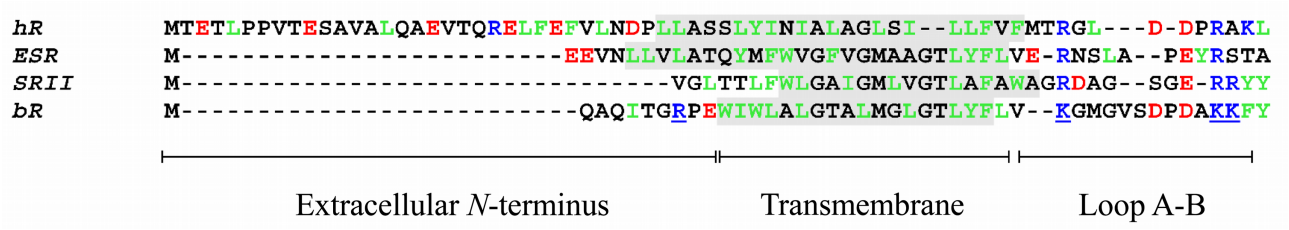

Figure 16: Sequence alignment of HR, ESR, SRII and bR amino acids of helix A with adjacent regions. This proteins were chosen because the structural data allowing the reliable assigning of transmembrane regions were available only for these proteins. Positively charged, negatively charged and hydrophobic amino acids are marked with blue, red, and green colors, respectively. Transmembrane domains of the helix A are marked with light grey according to the structural data. The features, which may effect expression of the protein, were revealed in bR and are underlined.

3.4 Functional expression of wild type bR and its mutants in E. coli.

3.4.1. Amino acid alignment.

In spite of high level of amino acid similarity bR did not show functional expression in *E. coli* as opposed to SRII, hR, and ESR. Trying to solve this riddle we made amino acid sequence alignment of bR to the aforementioned homologous retinal proteins shown in **Fig. 16**. Previous results indicated that the difficulties of bR expression in *E. coli* could be attributed to the impediment of insertion of newly synthesized bR into *E. coli* membrane [104].The "positive inside" rule states that MP topogenesis is controlled by positively charged amino acids [120] that keep the cytoplasmic parts of the MP in cytoplasm by interactions with negatively charged headgroups of anionic lipids and withstand both Sec-dependent and independent translocation against membrane potential [121, 122]. As observed *in vitro* using lysates and inner membrane vesicles prepared from *E. coli* as a model bR inserts co-translationally into *E. coli* membrane in Sec-dependent manner [123] demonstrating that amino acid sequence of the first transmembrane helix and adjacent regions is particularly important for the insertion of integral MPs into membrane. Therefore, analyzing the amino acid sequence alignment of bR to the other homologous retinal proteins we attended only to the region of first transmembrane helix and adjacent amino acids.

In this region we found (**Fig. 16**) two distinct features of bR that could be a factor limiting the ability of the helix to insert into *E. coli* membrane according to "positive inside" rule. First, positively charged Arg7 on the extracellular *N*-terminus of bacteriorhodopsin deviates from the "positive inside" rule. Second, bR has different positively charged amino acid pattern in the cytoplasmic loop A-B, where lysines substitute for arginines in the vicinity of the membrane. Such a charge distribution on the membrane flanking regions of the first helix could compromise folding and expression rate of bR.

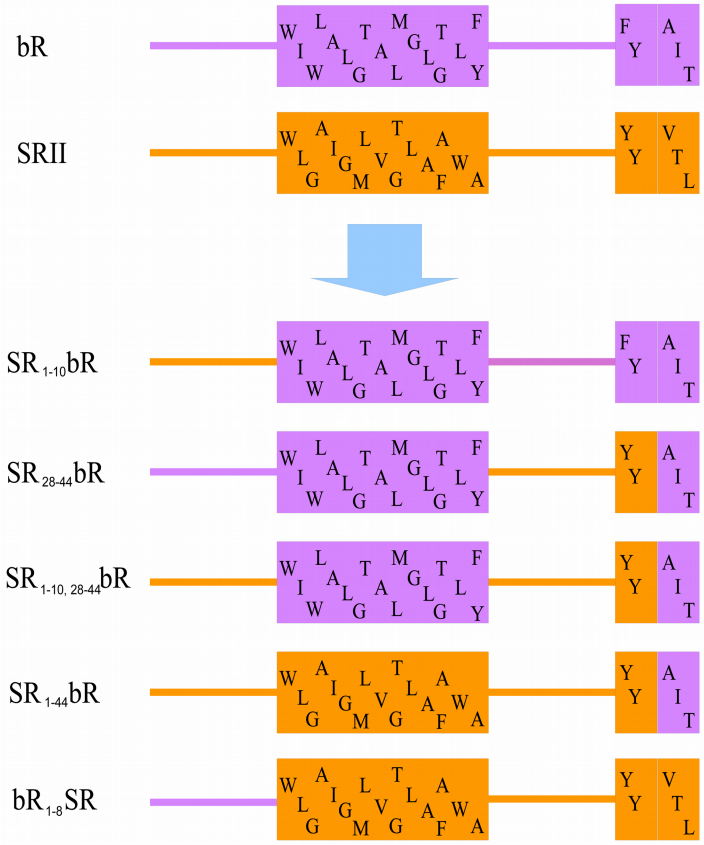

Figure 17: Design of chimeric proteins between bR and SRII. Parts of bR with corresponding amino acid sequence are shown in purple, for SRII the color is orange. Subscript in the name of the construct indicates amino acids of bR that were replaced by the counterparts from SRII and *vice versa*. *N*-terminus, first α-helix and the beginning of the second α-helix are shown.

3.4.2. Expression and purification of chimeric proteins between bR and SRII.

To investigate whether the charge distribution in the helix A region influences bR expression level in *E. coli* and the insertion of the protein into bacterial membrane we applied the complimentary protein approach to bR. We chose SRII as a complimentary protein since it is expressed in *E. coli* functionally with a high yield [27] and satisfies the "positive inside" (**Fig. 16**). In addition, SRII was the only retinal protein expressed in *E. coli* for which the high resolution structure was available. We constructed several chimeric proteins combining counterparts of bR and SRII as illustrated on **Fig. 17**. Following the protein complementarity approach, at first we replaced initial 43 amino acids of bR for the corresponding 36 amino acids of SRII composing chimeric

protein SR_{1-44}bR (as described in Materials and methods). Second, to elucidate the influence of extracellular, transmembrane and cytoplasmic regions independently we constructed in a similar way the chimeric proteins SR_{1-10}bR, SR_{28-44}bR, and $SR_{1-10,28-44}$bR, respectively.

The chimeric protein genes were inserted to pSCodon plasmid under control of inducible T7/*lac* promoter/operator. The *E. coli* SE1 cells were transformed with pSCodon plasmid carrying genes of target chimeric proteins to produce expression strains.

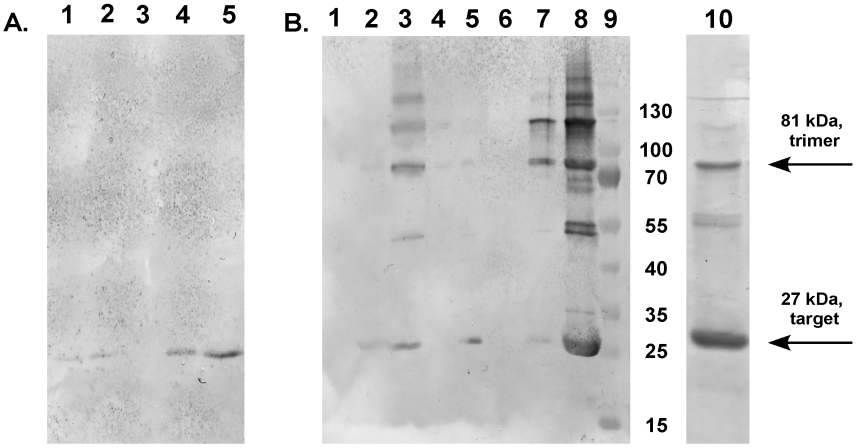

Figure 18: The purification of the chimeric proteins analyzed with SDS-PAGE and Western blotting. **A.** Western blotting with anti-*His*-tag antibodies showing the efficiency of purification of SR_{1-44}bR chimeric protein in different detergents. 1 - DM; 2 - DDM; 3 - OG; 4 - Fos12; 5 - Sarkosyl. Relying on these data we chose Sarkosyl to solubilize the chimeric proteins as the most effective detergent. **B.** Western blotting with anti-*His*-tag antibodies (and SDS-PAGE, lane 10) illustrates the process of the purification of chimeric protein in Sarkosyl (SR_{1-10}bR was taken as an example). 1 - cytoplasm; 2 - isolated membranes; 3 - solubilization suspension after ultracentrifugation; 4 - Ni-NTA column flowthrough; 5 - wash A; 6 - wash B; 7 - wash C; 8 - Ni-NTA column eluate; 9 - Protein Ladder; 10 - Ni-NTA column eluate (SDS-PAGE). The protein shows the distinct tendency to form multimers in Sarkosyl, monomer and trimer are indicated.

First, SR_{1-44}bR chimeric protein was expressed and purified. The progress of the purification process is illustrated in **Fig. 18**. The expression level of the protein was assessed to be 7.0±1.0 mg of protein per liter of culture from SDS-PAGE and BCA assay using Ni-NTA purified product. The screening of the detergents for solubilization of the chimeric protein showed (**Fig. 18**) that the strong detergents (FOS-12, Sarkosyl) solubilize the protein completely, but mild non-ionic detergents were not so efficient. DM, DDM can solubilize only up to 40-50% of the protein (DM, DDM). The protein resisted solubilization in OG. Thus, in the further experiments we purified the protein under semi-denaturing conditions employing Sarkosyl for solubilization.

To investigate the influence of the *N*-terminus, A-B loop, and first transmembrane region we expressed simultaneously SR_{1-44}bR, SR_{1-10}bR, SR_{28-44}bR, and $SR_{1-10,28-44}$bR chimeric proteins. The proteins were purified using Ni-NTA resin. Solubilization pellet, Ni-NTA flowthough and wash did

not contain considerable amount of the target proteins (**Fig. 18**, panel B, lanes 4-7). The yield and purity of the proteins were assessed from SDS-PAGE (**Fig. 19**, panel A, lanes 1-4). The purity of the Ni-NTA purified chimeric proteins was minimum 90%. The SR_{1-10}bR and $SR_{1-10,28-44}$bR constructs showed the same yield as SR_{1-44}bR chimeric protein, while SR_{28-44}bR had expression level nearly as low as native bR gene (**Fig. 19**). This data allowed us to localize the problem of bR expression in *E. coli* at the *N*-terminus of the protein.

To evaluate the efficiency of the expression of bR in *E. coli* comparing to the other retinal proteins, which have been shown to express in *E. coli* with the high yield, we have expressed and purified chimeric protein SR_{1-10}bR, native bR gene, SRII and ESR. The *E. coli* BL21 (DE3) cells were transformed with pET27bmod plasmid bearing SRII gene while *E. coli* SE1 cells were transformed with pSCodon plasmid bearing ESR, SR_{1-10}bR, or native bR genes. The proteins were expressed at the same conditions. The ESR and SRII were purified under non-denaturing conditions, while SR_{1-10}bR and bR were solubilized in NLS and detergent was then exchanged to SDS according to the protocol described in Materials and methods. Acrylamide gel with the Ni-NTA eluates loaded is presented in the **Fig 19**. According to the densitometry of the SDS-PAGE the yields of the proteins were 21 mg, 14 mg, 8 mg, and 0.15 mg per liter of culture for ESR, SRII,

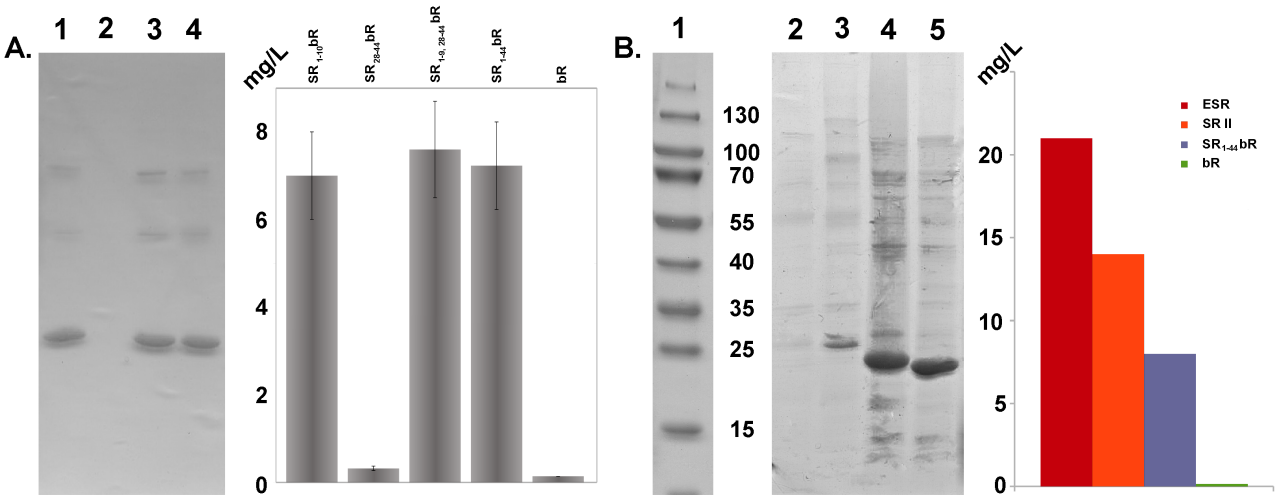

Figure 19: The SDS-PAGE analysis of expression of the chimeric proteins and its comparison with the other retinal proteins. **A.** Analysis of the Ni-NTA eluates of the chimeric proteins purified in Sarkosyl. 1 - SR_{1-10}bR; 2 - SR_{28-44}bR; 3 - $SR_{1-10,28-44}$bR; 4 - SR_{1-44}bR. The yields of the chimeric proteins were quantified using BCA protein assay and SDS-PAGE densitometry, the results are presented as histogram. **B.** Comparison of the yield of chimeric protein with the yields of the other retinal proteins. After purification on the Ni-NTA column eluates were loaded on SDS-PAGE and yields were quantified using BCA protein assay and SDS-PAGE densitometry, the results are presented as histogram. 1 - Protein Ladder; 2 - bR; 3 - SR1-44bR; 4 - SRII; 5 - ESR.

SR_{1-10}bR, and native bR respectively. Thus, the exchange of the first 10 amino acids of bR to their counterparts from SRII allowed us to increase ~50-fold the yield of the bR and made its yield

comparable to the yields of the other retinal proteins.

Interestingly, the construct $bR_{1-8}SR$ containing first 10 amino acids of bR instead of their counterparts on the *N*-terminus of SRII yielded 6.0±1.8 mg/l comparing to 18.1±2.8 mg/l for SRII. In the study [124] the authors accomplished the spectral tuning of SRII by constructing chimeric proteins between SRII and bR and were able to produce in small amount the functional chimera comprising A-C helices of bR and D-G helices of SRII expressed in *E. coli*. In case of $bR_{1-8}SR$

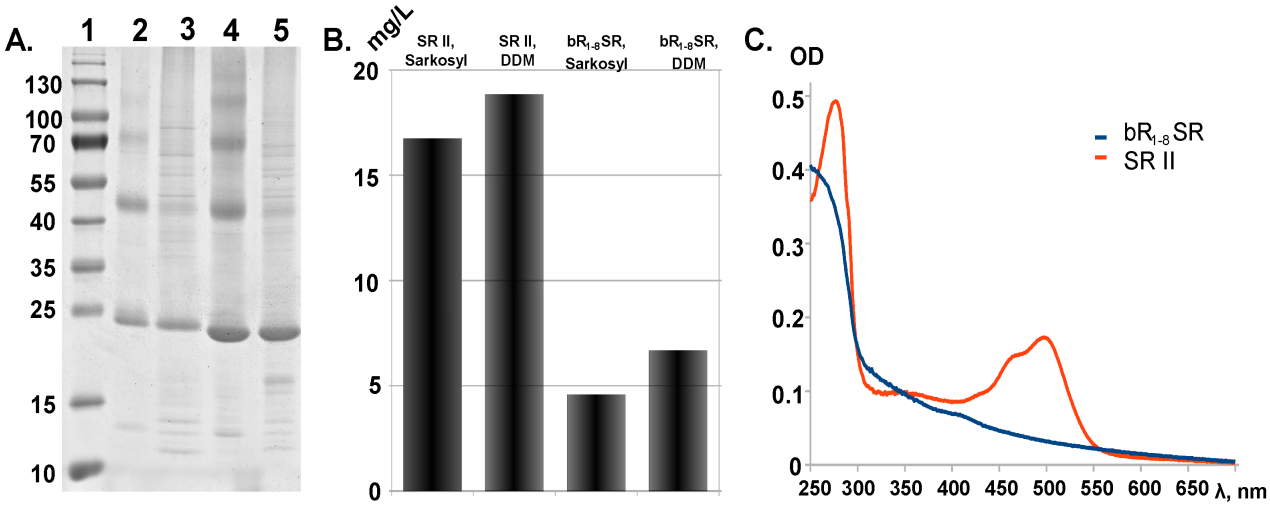

Figure 20: Expression of $bR_{1-8}SR$ chimeric protein and its comparison to SRII. Both proteins were purified in two different detergents (DDM and Sarkosyl) on Ni-NTA column. The eluates were loaded on SDS-PAGE: 1 - Protein Ladder; 2 - $bR_{1-8}SR$ in Sarkosyl; 3 - $bR_{1-8}SR$ in DDM; 4 - SRII in Sarkosyl; 5 - SRII in DDM. The yields of the protein were quantified using SDS-PAGE densitometry and BCA protein assay kit, the results are presented as histogram. The UV-Vis spectra of the proteins solubilized in DDM were measured. In contrast to SRII, $bR_{1-8}SR$ did not show retinal peak.

construct we have not obtained functional protein under non-denaturing conditions. We suppose that the introduction of charged Arg7 from bR into the hydrophobic interface of helix A brakes its hydrophobic interactions with the other helices leading to the destabilization of the SRII structure and misfolding of the $bR_{1-8}SR$ chimera. Thus, the suppressing influence of the *N*-terminus of bR have been supported by deterioration of functional SRII expression in *E. coli* when *N*-terminus of SRII was replaced with its counterpart from bR.

3.4.3. Functionality of the chimeric proteins and their characterization.

When the chimeric protein was expressed with retinal added to the cell culture and purified under non-denaturing conditions using DDM, it did not appear to be colored and, thus, functional. To acquire functionality the protein must be subjected renaturation. The prevailing protocol for bacterioopsin retinalization employs the DMPC/CHAPS vesicles and protein solubilized in SDS [119]. Attempts to renaturate bR directly from Sarkosyl were not successful as well as attempts to

renaturate bR in SDS obtained by the exchange of Sarkosyl to SDS using dialysis. However, when the detergent exchange was carried out directly on the Ni-NTA column, the chimeric protein was able to refold (**Fig. 21**). The amount of functional protein was calculated from absorbance maximum of retinal at 555 nm (**Fig. 21**) and taking into account the extinction coefficient of bR in DMPC/CHAPS vesicles to be 55 300 $cm^{-1}M^{-1}$ [125]. The total protein was estimated from BCA protein assay. The amount of the functional protein was assessed as 91%. Taking into account that the purity of the Ni-NTA purified samples was previously determined to be ~90% we concluded that the chimeric protein was retinalized completely.

After renaturation the protein samples contained excessive amount lipid/detergent vesicles and free retinal. These components can impede the spectroscopic and crystallography studies, thus we should have removed them. For this purpose we used size-exclusion chromatography (SEC). The renatureted chimeric protein in DMPC/CHAPS mixture was loaded on the Superose6HR column with bed volume 24 ml and eluted in 50 mM Na/Na-Pi pH 6.0, 0.1% DDM. Protein elution profile (**Fig. 21**) had only one distinct peak at 15.1 ml corresponding to the radius of gyration of 4.9 nm (or 170 kDa molecular weight) according to column calibration. The peak was very narrow (0.5 ml,

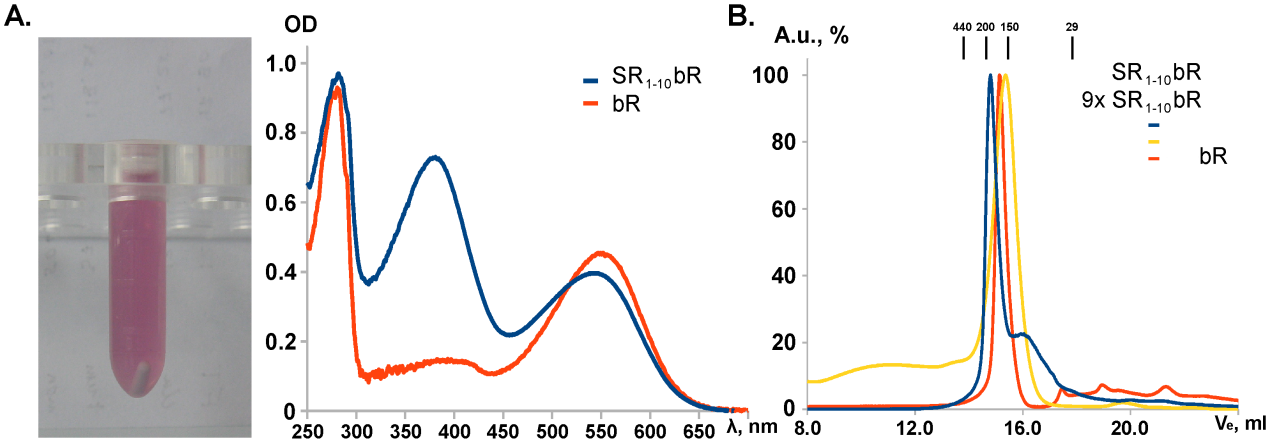

Figure 21: Renaturation of SR_{1-10}bR chimeric protein and its characterization. **A.** The renaturated chimeric protein had a distinct purple color. UV-Vis spectrum showed that sample has strong retinal absorption peak characteristic to functional bR shifted due to the absorption of the residual free retinal. **B.** Analysis of the renaturated protein using gel-filtration. The protein eluted as a monomer in DMPC/CHAPS bicell. As a reference we used native bR from purple membranes in Amphipol A8-35. The maxima of elution peaks corresponding to the water-soluble proteins (with their molecular weights) used for calibration of the column are indicated by arrows.

while for MPs characteristic values are 0.8 ml or wider) but had a clear wing in the area of the smaller molecular weight.

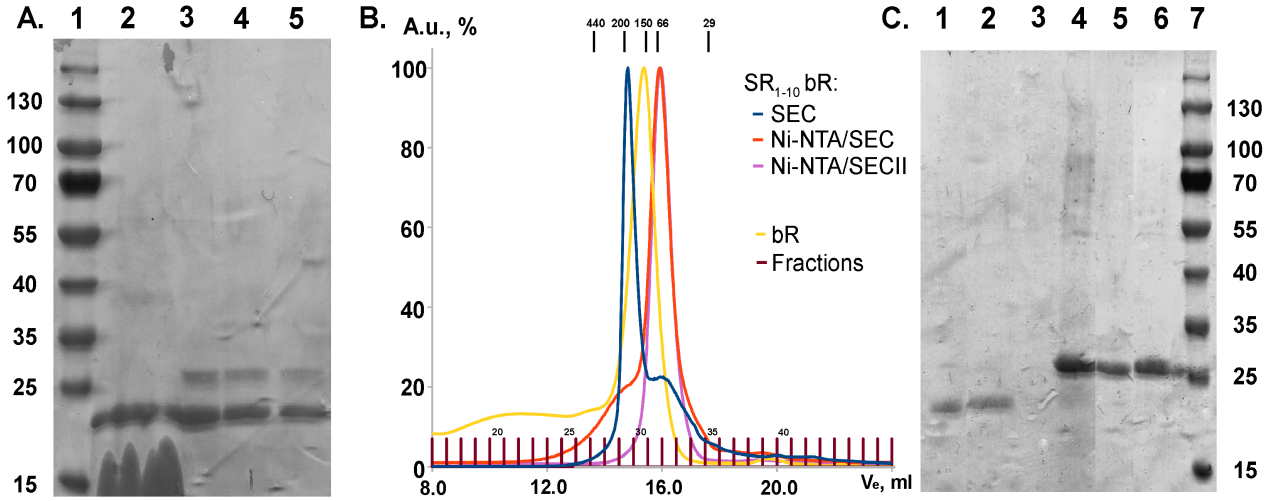

Figure 22: Delipidation of the SR_{1-10}bR reconstitution mixture. **A.** SDS-PAGE illustrating the delipidation of the reconstitution mixture using gel-filtration. 1 - Protein Ladder; 2 - concentrated sample before loading on the column; 3 - gel-filtration, fraction 29; 4 - gel-filtration, fraction 30; 5 - gel-filtration, fraction 32. **B.** Normalized SEC elution profiles of SR_{1-10}bR reconstitution mixture. The samples were: concentrated renaturation mixture directly (blue), renaturation mixture after delipidation on Ni-NTA column (red), and the second run on gel-filtration column of the sample delipidated using Ni-NTA (violet). As a reference we used bR from purple membranes in Amphipol A8-35 (yellow). The maxima of elution peaks corresponding to the water-soluble proteins (with their molecular weights) used for calibration of the column are indicated by arrows. Fractions collected are indicated in dark red. **C.** SDS-PAGE analysis of the delipidation using Ni-NTA and SEC. 1 - gel-filtration, fractions 26-32; 2 - Ni-NTA column flowthrough; 3 - Ni-NTA column wash; 4 - Ni-NTA column eluate; 5 - gel-filtration after Ni-NTA, fractions 31-32; 6 - gel-filtration II after Ni-NTA, fractions 30-33; 7 - Protein Ladder. Protein in lipids corresponds to the band ~23 kDa, whereas delipidated protein corresponds to the band 26kDa.

Relying on this information we concluded that on SEC-column the partial delipidation of the protein did occur. SDS-PAGE confirmed this finding (**Fig. 22**). The sample before loading on the chromatographic column showed the single band and excessive amount of lipid (seen on the gel as a smeared patch in the range of low molecular weights), whereas the samples collected after the column showed two different bands. While the low molecular weight band (~22 kDa) corresponds to the protein in lipid before gel-filtration, the second band (~26 kDa) matches the protein purified on the Ni-NTA resin. Unfortunately, delipidation of the samples was not complete leading to the non-homogenous preparations. The fractions pooled we not able to be concentrated to the volume requiring for the successive loading on the column turning into gel-like phase. Moreover, the excessive lipids resulted in the clogging of the column. Thus, this protocol did not allow the delipidation of the samples with high efficiency and on the large scale forcing us to try another approaches.

To achieve complete delipidation we used the Ni-NTA resin second time: the protein after renaturation was loaded in a batch mode to the Ni-NTA column equilibrated with 50 mM Na/Na-Pi

pH 6.0, 0.1% DDM buffer and incubated overnight for the improved binding of the protein to the column. As NG is known as the detergent that provides the bR photocycle most close to one of bR in purple membranes, we used this detergent in the elution buffer. Attempts to elute the protein using low pH (acetic buffer with pH 4.5) were not successful. Further attempts to elute the protein with imidazole and/or DDM resulted in complete denaturation of the protein. However, combining DDM with imidazole in elution buffer at pH 6.0 resulted in the successful elution of the protein in the next preparation.

After subsequent Ni-NTA column the protein eluates were loaded into SEC-column and eluted in 50 mM Na/Na-Pi pH 6.0, 0.1% DDM buffer. The elution profile showed only one distinct peak at 15.9 ml corresponding to the radius of gyration of 3.7 nm (or 65 kDa molecular weight) according to column calibration corresponding to the mostly delipidaded protein (**Fig. 22**). The fractions corresponding to this peak were collected and successive loading of the protein on SEC-column indicated delipidation of the sample. SDS-PAGE proved this observation showing the most of the protein in the 27 kDa band corresponding to the delipidated protein (**Fig. 22**). However, analysis of the probes taken from the second Ni-NTA column flowthrough and wash as well as lipid pellet before first gel-filtration indicated that significant part of the protein (at least 30%) was lost during removal of excessive lipids.

Expression of the chimeric proteins between bR and SRII showed that the complementary approach is a valuable tool to localize the reason of the expression failure. The exchange of first 10 amino acids of bR to their counterparts from SRII increased the expression level 50-fold. The expressed protein can be purified under denaturing conditions using Ni-affinity chromatography to the purity over 90%. Despite the fact that chimeric protein was not expressed functionally in *E. coli* it could be easily renaturated with high efficiency using standard approach employing DMPC/CHAPS vesicle. The high protein purity as well as delipidation of the samples could be achieved by means of the successive Ni-affinity and size exclusion chromatographies. However, such approach to the functional purification of the bR is not straightforward and do not give any significant advantage over the other published methods (see chapter2.4for review).

3.5 Identification of the reason of the low level expression of bR in *E. coli*.

3.5.1. Expression of the point mutants R7E and R7Q of bR.

One of the goals of the study was the expression of the native bR in *E. coli*. Consequently, we should clarify the reason of the increase in the yield of the chimeric protein $SR_{1\text{-}10}$bR to be able to produce the protein with amino acid sequence most close to the sequence of native bR. Within the short fragment identified by protein complementarity approach the most probable amino acids impeding expression of bR could be suggested using rational considerations. We supposed that the reasons of low expression level of bR in *E. coli* could be the presence of positively charged Arg7 on the *N*-terminus side of the helix A of bR or the unfavorable mRNA structure of bR gene near the ribosome binding site. We investigated this factors separately either by introducing point mutations that eliminate Arg7 or by optimizing mRNA structure with silent mutations in bR gene.

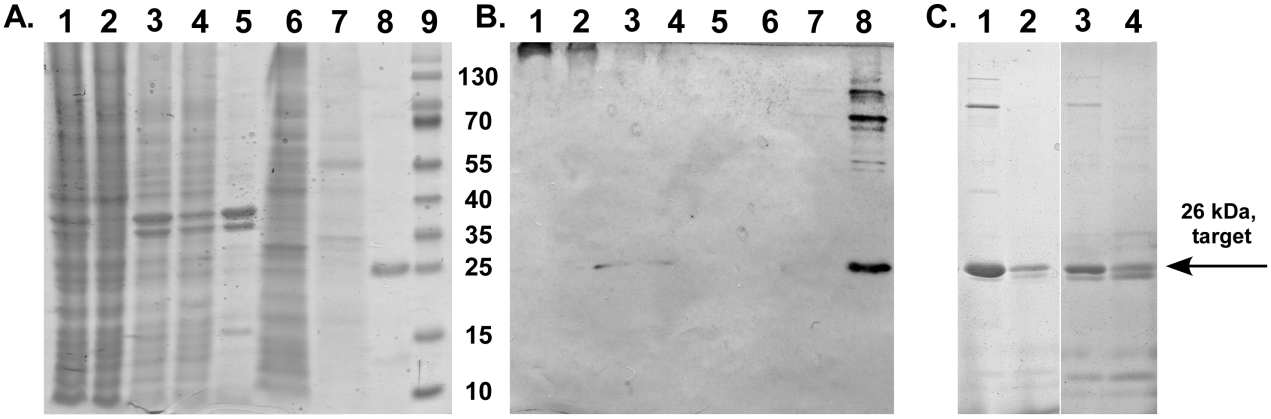

Figure 23: Expression of the point mutants R7Q and R7E, their purification and comparison to native bR and $SR_{1\text{-}10}$bR chimeric protein. **A.** SDS-PAGE illustrating the progress of purification of R7Q mutant protein. 1 - cell lysate; 2 - cytoplasm; 3 - membrane suspension; 4 - solubilized fraction; 5 - nonsoluble material; 6 - Ni-NTA column flowthrough; 7 - Ni-NTA column wash; 8 - Ni-NTA column eluate; 9 - Protein Ladder. **B.** Corresponding anti-*His*-tag immunoblot. The lanes 1-8 are the same as on SDS-PAGE. **C.** Comparison of protein yields of point mutants of bR and reference proteins $SR_{1\text{-}10}$bR and bR. 1 - $SR_{1\text{-}10}$bR; 2 - bR; 3 - R7Q; 4 - R7E.

First, we have replaced the Arg7 amino acid with neutral or negatively charged residues constructing the point mutants R7Q and R7E, respectively. According to "positive inside" rule [120] this residue could impede the insertion of the bR into *E. coli* membrane, thus diminishing the protein yield. To explore this hypothesis we expressed and purified the point mutants according to the same protocol as chimeric proteins. SDS-PAGE analysis have showed that there was no considerable losses of protein during purification (**Fig. 23**). Using BCA Protein Assay and SDS-PAGE analysis we found the yields of the point mutants to be 0.96±0.20 mg and 0.37±0.10 mg of protein per liter of culture for the R7Q and R7E mutants, respectively.

For comparison, the yield of the SR_{1-10}bR chimeric protein and native bR were 7.0±1.0 mg and 0.14±0.04 mg per liter of culture, respectively. The yield of the point mutants is considerably higher than that of native bR, on the other hand, it is still significantly lower than for the SR_{1-10}bR. We can suggest the simultaneous influence of several factors on the expression of bR in *E. coli*. It is first influenced by the "positive inside" rule, since the elimination of Arg7 increases the expression yield of point mutants 3- and 7-fold. However, the expression gain of SR_{1-10}bR is over the 50 times. Thus, despite its strong influence it is not only "positive inside" rule that caused the increase of expression yield of the chimeric protein comparing to wild type bR. We suggested that there is an additional effect, and the reason of the observed increase in protein yield could be the influence of the mRNA structure of bR gene downstream the ribosome binding site.

3.5.2. Optimization of bR mRNA.

The stability of mRNA folding downstream the ribosome binding is known to have a strong

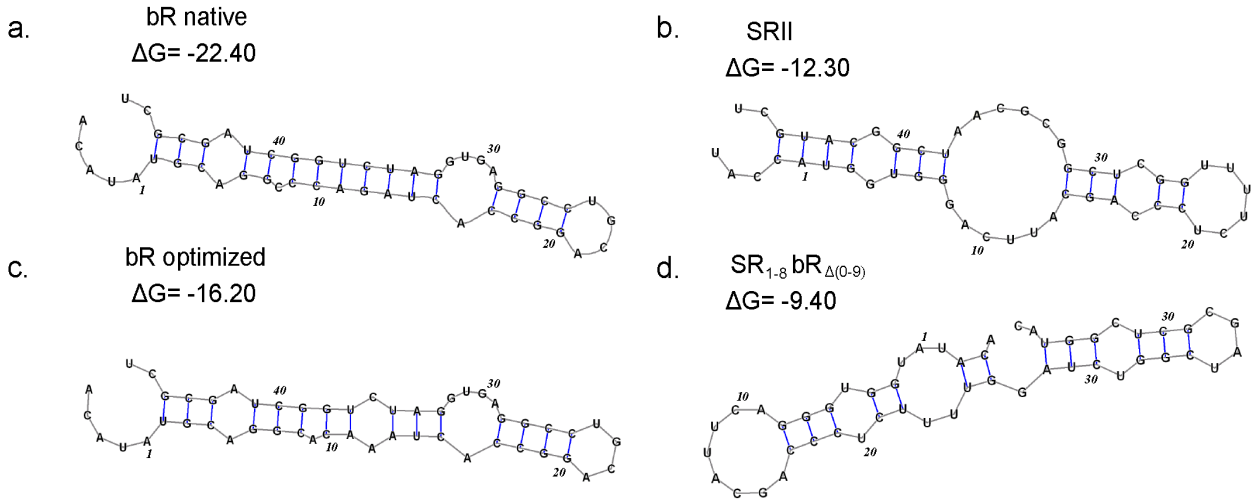

Figure 24: The mRNA shapes of the native bR, SRII, chimeric protein SR_{1-10}bR, and optimized bR gene for the region (-4, +47) with numbering starting from ATG start codon. This region corresponds to amino acids from Met0 to Leu15. Free energies presented are expressed in kcal/mol. The values of mRNA free energy for SRII and SR_{1-10}bR are considerably smaller than those of native bR gene. mRNA structure optimization of bR by silent mutations have increased the free energy of the region (-4, +47) by 6.2 kcal/mol.

impact on the protein expression level [126]. Particularly illustrative is the study with green fluorescent protein, when the yield of the protein in *E. coli* varied 250-fold across the synthetic gene library encoding the same amino acid sequence [127]. The authors have shown that the stability of mRNA folding near ribosome binding site and associated rate of translation initiation play the predominant role in the defining the expression level of the protein. In the other work [20] it has been shown that the variation of nucleotide sequence near ribosome binding site significantly

influences the expression level of bR in *E. coli*. Our purpose was to examine whether the increase of the protein yield of chimeric proteins can be explained by the impaired translation initiation due to possible unfavorable mRNA structure of bR near ribosome binding site.

Using the mRNA structure prediction software *mRNAshapes* [128] we have found a putative stem structure at the start of the bR gene (**Fig. 24**). As the stem stretches beyond nucleotide 37, we had to expand the region under consideration to (-4, 47) comparing to previous study [127]. This region corresponds to amino acids from Met0 to Leu15. The free folding energy of mRNA for the (-4, 47) region of bR is predicted to be -22.4 *kcal/mol*, while the corresponding value for SRII, ESR, and hR genes are -12.3, -5.2, -13.1 *kcal/mol* indicating the decreased stability of the mRNA structures at the 5'-terminus near ribosome binding site. The change of first 9 amino acids of bR to the counterparts of SRII have considerably weakened the interactions between the 6-15 and 33-41 nucleotides and increased the free energy by 13.0 *kcal/mol* (**Fig. 24**). And *vice versa,* the folding energy of $bR_{1-8}SR$ is -15.5 *kcal/mol* explaining the lower yield of this chimera. Since the expression level of the native bR, SRII, $SR_{1-10}bR$, and $bR_{1-8}SR$ have correlated with the stability of the (-4, 47) region of mRNA structure, we introduced into wild type bR gene two silent mutations (C9A, G12A corresponding to amino acids Ala2 and Gln3) that increase the free energy by 6.2 *kcal/mol* and significantly reduce the stem stability (**Fig. 24**).

3.5.3. Expression in *E. coli* of the mRNA-optimized bR and its purification under denaturing conditions.

The mRNA optimized bR gene was inserted into pSCodon plasmid, expressed in *E. coli* SE1 cells, and purified under denaturing conditions as described in Materials and methods similar to the chimeric proteins. SDS-PAGE analysis have showed that there was no protein in the probes from solubilization pellet, column flowthrough and washes, while cytoplasm contained considerable amount of target protein (**Fig. 25**). The purity of the preparations was accessed by SDS-PAGE to be higher than 90%. The protein could be renaturated to functional state in the same fashion as $SR_{1-10}bR$ with the high efficiency. The expression yield was estimated by SDS-PAGE and BCA protein assay kit as 7.6±2.9 mg per liter of culture. This yield is essentially the same as the yield of the chimeric protein $SR_{1-10}bR$ ($p = 0.74$). Thus, we have proved that the expression gain of the chimeric protein was due to the alteration of the mRNA structure of bR near ribosome binding site. With two point mutations in the bR gene, which do not change the amino acid sequence, we have achieved the expression level 50 times higher than for the native gene.

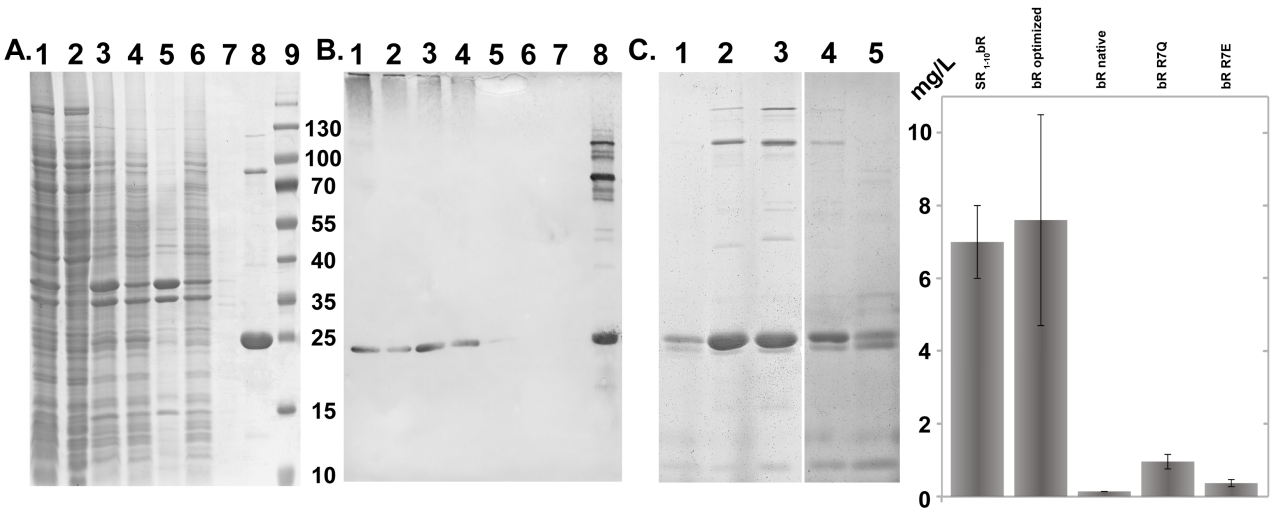

Figure 25: Expression of the mRNA optimized bR and its purification under denaturing conditions. **A.** SDS-PAGE illustrating the progress of protein purification. 1 - cell lysate; 2 - cytoplasm; 3 - membrane suspension; 4 - solubilized fraction; 5 - nonsoluble material; 6 - Ni-NTA column flowthrough; 7 - Ni-NTA column wash; 8 - Ni-NTA column eluate; 9 - Protein Ladder. **B.** Corresponding anti-*His*-tag immunoblot. The lanes 1-8 are the same as on SDS-PAGE. **C.** Comparison of protein yields of the optimized bR, point mutants, and reference proteins SR_{1-10}bR and bR. 1 - SR_{1-10}bR; 2 - bR; 3 - optimized bR; 4 - R7Q; 5 - R7E. The protein yields were quantified using SDS-PAGE densitometry and BCA protein assay kit, the results are presented as histogram.

The optimization of mRNA has also significantly increased the bR expression rate in previous study with the use of pJP plasmid with T5 promoter under control or without control of *lac* operator [20]. However, the authors found in the mRNA 5' untranslated region the hairpin structure that encloses *lac* operator and hence impairs mRNA translation initiation, while uncontrolled T5 promoter led to the deleterious growth of *E. coli* cells. In case of pSCodon plasmid we have not observed any adverse influence of *lac* operator on the protein yield. This discrepancy could be explained by the insufficient spacing between *lac* operator and ribosome binding site in the pJP plasmid. To prevent translation suppression *lac* operator the authors of those study used P_L promoter repressed by temperature sensitive repressor. Transcription was induced by temperature shift from 30 to 42°C resulting in the rapid degradation of the newly synthesized protein under the chosen expression conditions. Therefore, the total yield of unfolded bR was approximately 2 times lower than achieved in the present study.

In general, a combination of the complementary protein approach with rational considerations allowed us to determine the reason of the low level of bR expression in *E. coli*. We have attributed the problem to the unfavorable mRNA structure of bR downstream the ribosome binding site. The improved expression system and more optimal conditions of protein expression allowed us to achieve additional gain in protein yield comparing to the previous studies. The yield achieved was 50 times higher than for the native gene being comparable with the yield of native *H. salinarum* system [11].

3.6 Expression in *E. coli* and functional purification under non-denaturing conditions of bR and its mutants.

3.6.1. Purification of bR in non-denaturing conditions.

When we expressed mRNA optimized wild-type bR in *E. coli* with retinal added into cell culture and used mild detergent DDM for the solubilization of the protein after membrane isolation, the protein retained its functionality during the purification process under non-denaturing conditions. The progress of purification is illustrated by SDS-PAGE in **Fig. 26**. The first step of affinity chromatography led to the mixed preparations of functional bR as well as not properly folded bacterioopsin and *E. coli* MPs impurities. The protein yield was estimated to be 7.0±2.8 mg per litre of culture, as measured by BCA protein assay. Sample homogeneity was confirmed by coomassie-stained SDS-PAGE (**Fig. 26**, A).

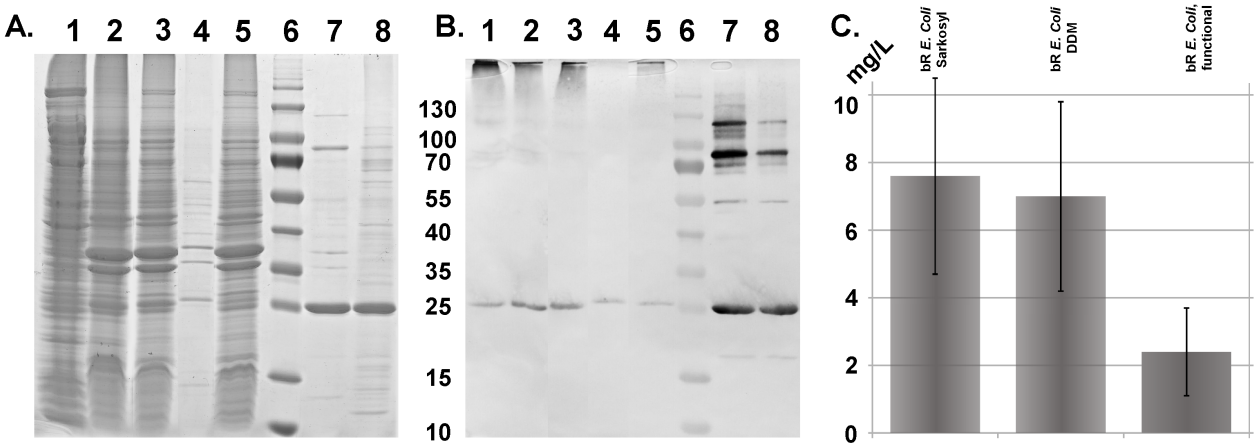

Figure 26: Expression of the mRNA optimized bR and its purification under non-denaturing conditions. **A.** SDS-PAGE illustrating the progress of protein purification. 1 - cytoplasm; 2 - membrane suspension; 3 - solubilized fraction; 4 - nonsoluble material; 5 - Ni-NTA column flowthrough; 6 - Protein Ladder; 7 - Ni-NTA column eluate (protein solubilized in Sarkozyl); 8 - Ni-NTA column eluate (protein solubilized in DDM). **B.** Corresponding anti-*His*-tag immunoblot. The lanes 1-8 are the same as on SDS-PAGE. **C.** Comparison of protein yields of the bR using different purification protocols. 1 - under denaturing conditions; 2 - under non-denaturing conditions, total yield; 3 - under non-denaturing conditions, functional protein. The protein yields were quantified using SDS-PAGE densitometry and BCA protein assay kit, the results are presented as histogram.

As solubilized bR is unstable at alkaline pH and imidazole is harmful for the protein, we removed imidazole and adjusted pH to 6.0 by dialysis. During the pH exchange the protein heavily precipitated. The pellet was not colored and constituted of bR according to SDS-PAGE analysis suggesting aggregation of the misfolded protein. After discarding the pellet the most of the functional bR still remained in the supernatant, but UV-Vis spectroscopy have shown that the sample was not yet free of aggregates and protein contaminations.

For further purification of bR we have used size-exclusion chromatography (**Fig. 27**). Protein elution profile had two distinct peaks at 69.7 ml and 86.5 ml corresponding to bR aggregates and functional bR. The colored fractions from the latter peak were pooled, mixed and concentrated. The UV-Vis absorbance spectrum exhibited the retinal absorption peak at 555.5±1.0 nm (**Fig. 27**, C) matching absorption maximum of the dark-adapted bR from purple membranes of *H. salinarum* solubilized in DDM [129]. The peak ratio $A_{280}/A_{\lambda max}$ of 1.5 was achieved demonstrating that the purity of the protein is consistent with that of bR solubilized from native purple membranes of *H. salinarum* and, thus, satisfies the requirements for the use of this protein in different applications in science and industry. Storage stability of the protein was analyzed 5 days after purification by gel-filtration on the same column. Bacteriorhodopsin eluted as single and symmetric (asymmetry index 1.05) peak indicating size homogeneity of the final product (**Fig. 27**). Thereby we have shown that the aggregates were removed completely from the samples and the purified protein had no tendency to denature or form aggregates upon storage. The final yield of the purified functional wild type bR expressed in *E. coli* was 2.4±1.3 mg of the protein per liter of culture corresponding to 15-35% of total synthesized bR (**Fig. 26**).

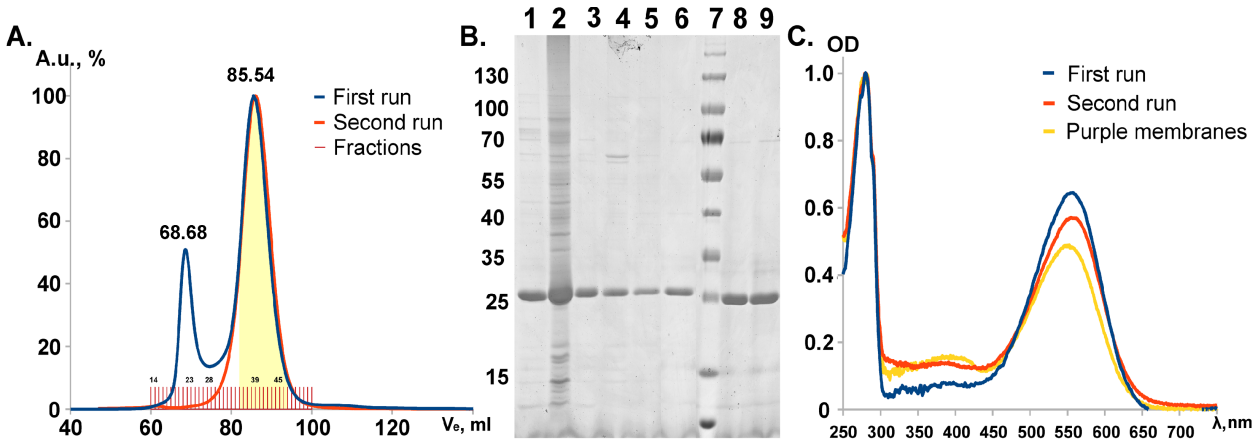

Figure 27: Characteization of bR expressed in *E. coli*. **A.** Elution profile of wild type bR on Sephacryl S200HR. The first run is shown in blue and the second in red. Peak at 68.68 ml in the first run corresponds to aggregated colorless protein. Second peak at 85.54 ml is the target protein. The fractions are indicated and those that were pooled are shown by pale yellow color. After 5 days the purified protein was subjected to the second round of gel-filtration, but it did not show any significant amount of aggregates. **B.** SDS-PAGE analysis of fractions obtained from Sephacryl S200HR column. 1 - eluate from Ni-NTA column; 2 - pellet after dialysis; 3 - sample before gel-filtration; 4 - fraction 23; 5 - fraction 28; 6 - fraction 34; 7 - Protein Ladder; 8 - fraction 39; 9 - fraction - 45. **C.** UV-Vis absorption spectra of the samples of wild type bR expressed in *E. coli* (after first and second run of gel-filtration) and of protein from native purple membranes solubilized in OG. The spectra were normalized by absorbance at 280 nm. Proteins were solubilized in DDM and purified using non-denaturing protocol. The spectrum of wild type bR has exhibited the retinal absorption peak at 555.5±1.0 nm matching absorption maximum of the delipidated dark-adapted bR from purple membranes. The peak ratio $A_{280}/A_{\lambda max}$ of 1.5 was achieved demonstrating the high purity of the protein.

3.6.2. Factors influencing the expression of wild type bR in *E. coli*.

Karnik et al. [104] have shown that only 1-2% of bR synthesized in *E. coli* bound retinal, when cells were incubated at 37°C after induction. We suppose that the cultivation temperature shows significant influence on the yield of the functional bR in case of *E. coli* expression. Indeed, it is recommended [130] to express MPs at lowered temperatures (20-30°C) to reduce the rate of the protein synthesis, facilitate its membrane insertion and proper folding. As shown *in vitro* bR insertion into membrane occurs co-translationally in Sec-dependent manner [123]. Its overexpression can overload the cell translocation system resulting in misfolded protein and increased rate of protein degradation. Using similar mutations to optimize mRNA structure (-16.2 vs. -12.3 kcal/mol) we have obtained the doubled translational yield of bR comparing to previous study [20]. However, the authors of this work employed denaturing protocol involving organic solvents, therefore, they had to renaturate the protein using DMPC/CHAPS vesicles, whereas our protocol takes advantage of non-denaturing conditions, thus allowing to produce functional protein straightforward.

It is also worth noting that high rate of bR misfolding can be attributed to unfavorable lipid composition of *E. coli* membrane. The PE lipids, a major component of *E. coli* membrane, were shown to decrease bR regeneration yield *in vitro* [131]. At the same time, the folding of bR was shown to follow a two-stage model with at least one transition state [132]. Therefore, *E. coli* lipids might influence the transition state and favor bR misfolding.

The described protocol allows one to avoid in expression and purification strategy steps of protein extraction with organic solvents, use of fusion proteins as expression drivers, solubilization of protein in denaturing detergents and protein renaturation from SDS, which inherent to the existing protocols [20-24]. We have achieved expression in *E. coli* and purification of functional wild type bR under non-denaturing conditions in quantities sufficient for structural biology studies and other purposes. The protein is stable, homogeneous and resembles the native bR from purple membranes.

3.6.3. Purification of V49A, D85N, and D96N mutants of bR in non-denaturing conditions.

One of the main advantages of *E. coli* expression system over *H. salinarum* is the considerably reduced time required to produce the mutants of interest facilitating the intense study of the target

protein. Using mRNA-optimized bR gene we have introduced mutations V49A, D85N, and D96N and utilizing the non-denaturing purification protocol we readily obtained functional mutant proteins with yields of 0.3, 3.8, and 8.8 mg per liter of culture, respectively. The mutant proteins had essentially the same gel-filtration elution profile as wild-type bR (**Fig. 28**). V49A and D96N UV-Vis spectra were simular to the wild-type bR, while D85N mutant exhibited characteristic maximum retinal absorbance at 598.0 nm and showed higher peak ratio $A_{280}/A_{\lambda max}$ comparing to the wild type protein and D96N mutant (**Fig. 28**). Thus, we have proven that the suggested approach is also efficient for fast production of bR mutants.

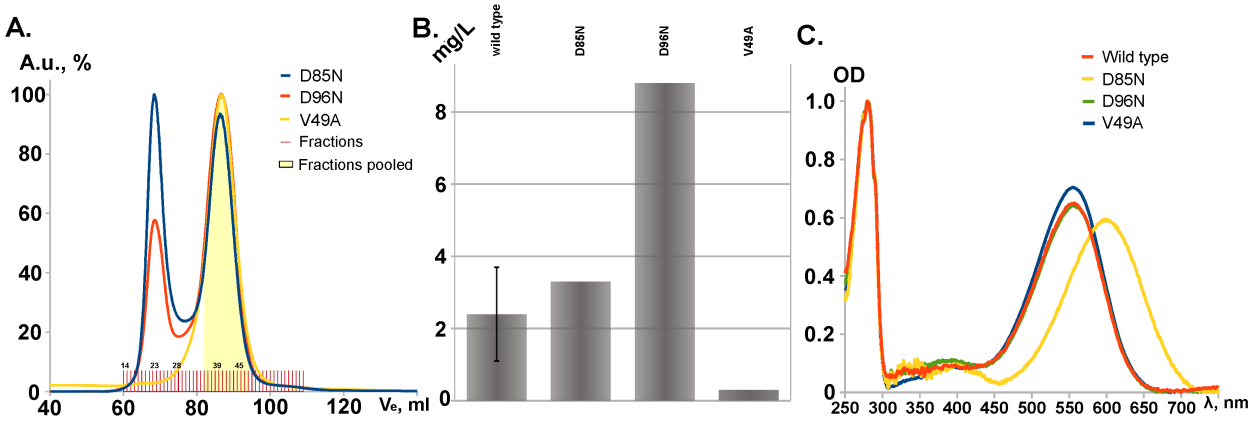

Figure 28: Characterization of bR mutants expressed in *E. coli*. **A.** Elution profile of bR mutants on Sephacryl S200HR. The elution profile for D85N mutant is shown in blue, for D96N in red, for V49A (second run) in yellow. Peak at 68 ml corresponds to aggregated colorless protein. Second peak at 85 ml is the target protein. The fractions are indicated and those that were pooled are shown by yellow color. **B.** Comparison of yields of the bR mutants. The protein yields were quantified using BCA protein assay kit and UV-Vis spectroscopy, the results are presented as histogram. **C.** UV-Vis spectra of wild type bR and its mutants. The spectrum for wild type protein is shown in red, for D85N mutant in yellow, for D96N in green, for V49A in blue.

It is worth emphasizing the higher expression level of D85N and D96N mutants of bR. It is not surprising. The mutation D94N in the bacteriorhodopsin from *Haloarcula marismortui* (corresponding to D96N mutation in bR) have increased 10-fold the yield of the functional protein [133]. It was also shown [134] that insertion of bR helix C into membrane *in vitro* is impeded by two aspartic acid residues within transmembrane region. In the present work D85N and D96N mutations led to the 1.5 to 4-fold increase in the yield of functional protein relative to wild type bR, the fraction of properly folded protein increased as well from 25±10% to 35% and 60% of the total yield of D85N and D96N mutants, respectively. This increase may be explained by improved incorporation of the newly synthesized protein into *E. coli* membrane that reduces protein degradation and facilitates the correct folding of the bR mutants.

3.6.4. Expression and functional purification of bR-Lys fusion protein under non-denaturing conditions.

The further development of the project in line with the objectives of this study supposed the adaptation of the developed expression system for production of the functional bR-Lys fusion proteins. To achieve the goal we have constructed again three variants of bR-Lys fusion protein gene in pSCodon plasmid: pSCodonbRL, pSCodonbRLE, and pSCodonbRLEs as it is described in Materials and methods. The proteins were expressed in *E. coli* and purified under non-denaturing conditions similar to the bR and its mutants. The probes of the cytoplasm fractions, solubilization pellets, column flowthroughs and washes were analyzed on SDS-PAGE and Western blotting showed that only solubilization pellets contained considerable part of the target protein. However, the part of the protein in the pellet was only 20-30% of total expression yield, and, as pellet was not colored, consisted mostly from not properly folded protein. The total yields (translational yield) of the bR-Lys fusion proteins were estimated by BCA assay to be 5.6 mg, 4.0 mg, and 4.1 mg per liter of culture for bRL, bRLE, and bRLEs, respectively. The purity of the preparations was estimated to be more then 90% using SDS-PAGE.

The bRL fusion protein was not colored and, thus, functional, whereas the eluates corresponding to the bRLE and bRLEs exhibited a faint purple color. This proteins (bRLE and bRLEs) initially purified with Ni-NTA affinity chromatography were further purified using gel-filtration according to the procedure described for the wild type bR. The proteins were loaded on SephacrylS200HR column with bed volume of ~160 ml and eluted in 50 mM Na/Na-Pi pH 6.0, 0.1% DDM. Protein elution profiles exhibited two distinct peaks at 60.5 ml and 84.9 ml. The colored fractions corresponded to the peak at 84.9 ml, were pooled, concentrated and analyzed using UV-Vis spectroscopy. The spectra exhibited two characteristic peaks at 280 and 560 nm corresponding to the absorption of protein amino acids and retinal. Using Protein Calculator (available at http://www.scripps.edu/~cdputnam/protcalc.html) the reliable peak ratio that similar to 1.5 for bR for the bR-Lys chimeric proteins was calculated to be 2.13. After purification on SephacrylS200HR we were able to obtain the proteins with the peak ratios 2.34 and 2.63 for bRLE and bRLEs, respectively, that corresponds to 91% and 81% of functional protein in the preparations. The yields of the functional proteins were 0.83 mg and 0.41 mg what means that only 20% of bRLE and 10% of bRLEs total protein is properly folded. The lower fractions of folded protein comparing to the wild type bR may be explained by complexity of the 3D structure of the fusion proteins that decreases the probability of the proper folding. However, we were able to purify the functional protein to the purity sufficient for crystallization and started crystallization. Before starting

crystallization trials we launched the experiments to investigate the influence of detergent and native lipids on crystallization of bR.

3.7 The influence of detergent on in meso crystallization of bR.

3.7.1. Crystallization of bR from purple membranes in the mixtures of detergents.

As it was stated before, to develop an effective approach for the crystallization of bR-Lys fusion

	OG	MEGA-10					
Cdet	10.0	1.0	2.1	4.2	6.5		
Cdet	8.5	1.0	2.1	4.2	6.5	10.5	
Cdet	7.0		2.1	4.2	6.5	10.5	12.0
Cdet	3.5			4.2	6.5	10.5	12.0
Cdet	1.2			4.2	6.5	10.5	12.0
Cdet	0.6			4.2	6.5	10.5	12.0

	OG	DDM					
Cdet	13.0	1.0	2.4	3.8	5.1		
Cdet	11.5	1.0	2.4	3.8	5.1		
Cdet	10.0	1.0	2.4	3.8	5.1		
Cdet	8.5	1.0	2.4	3.8	5.1	10.2	
Cdet	7.0		2.4	3.8	5.1	10.2	20.4
Cdet	3.5			3.8	5.1	10.2	20.4

	OG	CYMAL-5					
Cdet	10.0	1.7	3.4	6.7	10.0		
Cdet	8.5	1.7	3.4	6.7	10.0	13.2	
Cdet	7.0		3.4	6.7	10.0	13.2	18.8
Cdet	3.5			6.7	10.0	13.2	18.8

Table 1: Crystallization conditions of homologously expressed wild type bR (from purple membranes) in mixtures of detergents. The concentrations of the detergents are indicated in arbitrary units. The three mixtures of detergents were used: OG-MEGA10, OG-CYMAL5, and OG-DDM.

protein and, thus, to have more chances to crystallize the fusion protein spending less protein material we need first to investigate the influence of lipid/detergent environment on the process of *in meso* crystallization. My task was to explore the *in meso* crystallization of homologously expressed bR from purple membranes in mixtures of different detergents to establish the influence of detergent on crystallization of MPs. This study was carried out as a part of the work under the project devoted to the investigation of the role of lipid/detergent environment in *in meso* crystallization in our group in collaboration with my colleagues.

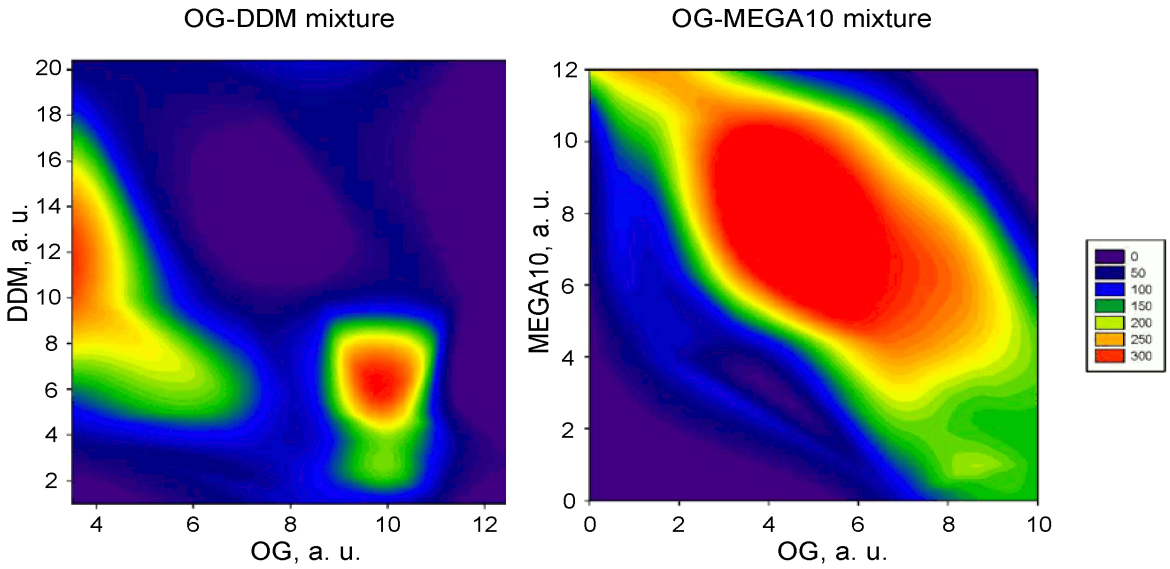

Figure 29: The crystallization diagrams showing the maximal size of the crystals obtained depending on the concentrations of the detergents in arbitrary units. The diagrams for the OG-DDM (left) and OG-MEGA10 (right) are presented. The size of the crystals indicated by color: the biggest crystals correspond to the red color, the blue color corresponds to the absence of crystals. The biggest crystals reached 300-350 μm.

The crystallization was done according the modified protocol of Landau and Rosenbuch [6], as described in Materials and Methods. We chose mixtures OG-MEGA10, OG-DDM, and OG-CYMAL5 as this detergents have shown the best results for bR crystallization in single detergent (the data were obtained in our laboratory). The the concentrations of the detergents that were used

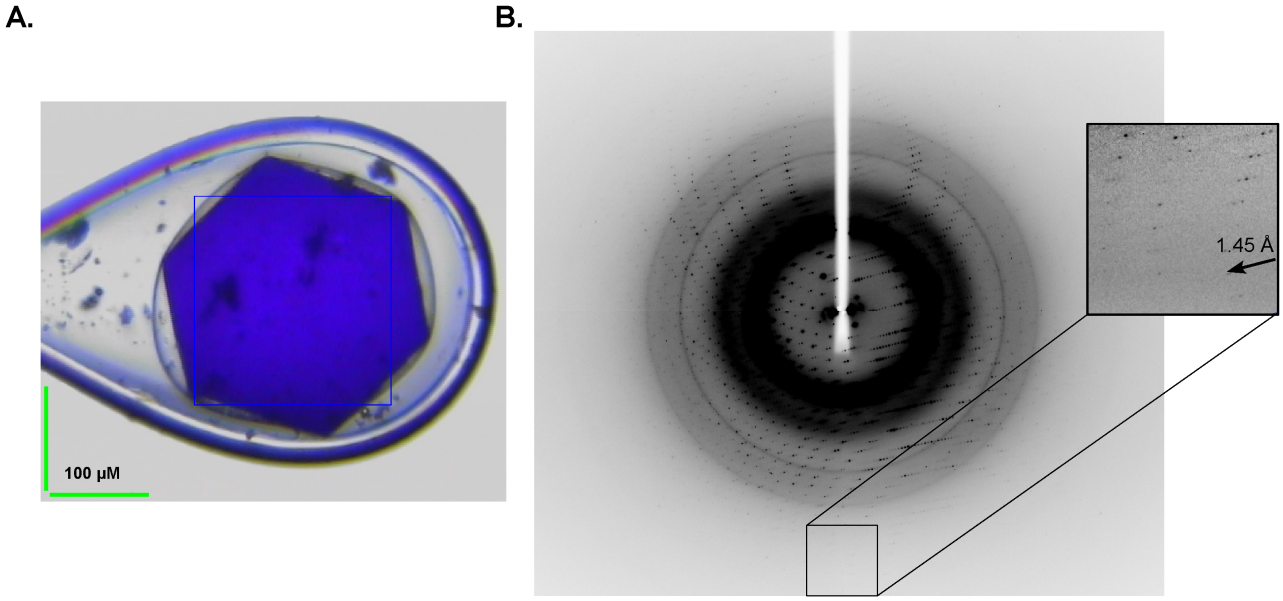

Figure 30: Crystals of wild type bR from *H. salinarum* purple membranes. A. Crystals of wild type bR were obtained using in cubo crystallization approach in the mixtures of detergents. The crystals we up to 350 μm. B. The datasets from single crystals were obtained at synchrotron radiation. The highest resolution seen is 1.45 Å (inset).

for screening are summarized in the Tab. 1. One of the most important data obtained from this experiments is the maximal size of crystals that could be grown under certain conditions. For the mixtures OG-MEGA10 and OG-DDM the crystallization diagrams are presented in **Fig. 29**. The mixture OG-CYMAL5 did not give such a big crystals, but, however, there was a plenty of crystals with the size up to 100 μm. The OG-MEGA10 mixture gave the highly ordered crystals with size up to 250-300 μm. In OG-DDM mixture the crystals grew up to 300-350 μm in size, however most of the big crystals had a shape of several thin plates stuck together.

3.7.2. Data collection and model building.

The biggest highly ordered crystals were tested on the in-house X-ray radiation source. The X-ray diffraction data sets at synchrotron radiation (ID 14-1, Grenoble, France) were taken from the 5 best crystals according to the results of the tests. One crystal from OG-MEGA10 mixture (345OME) diffracted up to 1.45 Å resolution. It was grown in the mixture containing 3.5 a.u. OG and 10.5 a.u. MEGA10. Two crystals grown in OG-DDM mixture (365ODM and 385ODM) gave the diffraction up to 1.7 Å resolution. Both crystals were obtained at 7 a.u. OG and 2.4 a.u. DDM.

The diffraction patterns were processed with MOSFILM [135] and SCALA [136]. For crystals

Dataset (model)	345OME	365ODM	385ODM
Data collection			
Space group	$P6_3$		
Cell dimensions, Å×Å×Å	60.78×60.78×110.42	61.00×61.00×110.13	60.86×60.86×110.40
Resolution, Å	19,90-1,45	55,07-1,70	110,36-1,70
Number of reflections processed	276274	185578	185967
Number of unique reflections	40710	23324	23278
Twinning fraction, %	49.4	38.6	32.6
I/σ(I)	18.4 (4.1)	14.1 (3.3)	13.4 (4.0)
R_{sym}, %	6.0 (36.9)	8.1 (37.2)	9.5 (33.0)
Completeness, %	99.7 (98.9)	91.4 (98.8)	91.6 (97.2)
Model building			
R_{work}, %	10.44	13.6	15.56
R_{free}, %	12.76	14.48	18.71
RMSD bond length, Å	0.014	0.013	0.017
RMSD bond angle, °	1.656	1.502	1.813
Overall B-factor, Å^2	15.70	22.00	17.50
Protein B-factor, Å^2	13.03	19.15	14.12
Water B-factor, Å^2	24.90	31.67	25.42
Lipid B-factor, Å^2	46.66	56.23	54.46
Monooleoyl B-factor, Å^2	37.70	51.08	43.22

Table 2: Data collection and refinement statistics. The data are presented for the datasets (models) 345OME, 365ODM, and 385ODM. The data for the highest resolution shell are shown in parentheses.

with relatively low (less than 35%) twinning ratio the detwinning procedure was used, while for the cases of perfect twinning the build-in algorithms of CCP4 package [137] were used. The initial phasing was done using molecular replacement method realized in program MOLREP [138] with poly-Ala model of high resolution bR structure (derived from 1C3W structure [14] in PDB Data Bank [2]). Protein auto-building was accomplished by ARP/wARP [139]. The initial model was refined using Refmac [140].

All three crystals of bR belonged to the $P6_3$ space group with the cell dimensions 61×61×110 $Å^3$. The crystals exhibited twinning characteristic for the most bR crystals grown *in meso*. The 365ODM and 385ODM crystals had a relatively low twinning ratio 38.6% and 32.6%, respectively, while 345OME crystal had a perfect twinning of 49.4%. The diffraction patterns of crystals grown in OG-DDM mixture had intensive rings from powder diffraction on ice and remains of the cubic phase that deteriorated the quality of diffraction data. As the 365ODM crystal had higher twinning ratio the corresponding structure had the noisier electron densities and higher temperature factors of the atoms comparing to the 385ODM. The models built using 365ODM and 385ODM had the R_{work}/R_{free} factors 13.60%/14.48% and 15.56%/18.71%, respectively. The statistics of the data collection and model building is presented in the Tab. 2.

Since the main objective of the study was the lipid/detergent environment of the protein, we will mostly concentrate on the aspects related to this matter. In general, the models built do not have significant differences from the other high resolution models, so the aspects considering the functionally important amino acids as well as water molecules network inside bR will not be described.

3.7.3. The general overview of the model of ground state of bR.

As there was no considerable differences between the models corresponding to the crystals of bR grown in the OG-MEGA10 and OG-DDM mixtures, we will describe the model based on the diffraction data with highest resolution, namely 345OME. In case there is significant difference between the structures, it will be specially specified.

Amino acid sequence of bR consists of 248 amino acids, 227 of them were observed in crystallographic structures. In accordance to previous data, bR contains 7 transmembrane helices. The loop B-C include 20 amino acids, 14 of them form β-sheet, while the rest of the loops are short consisting of 7-10 amino acids. *N*- and *C*-termini are disordered, thus, amino acids from 1 to 4 and

from 235 to 248 are not included to the model. The E-F loop is flexible as well and was not built due to the insufficient quality of electron densities in the corresponding region. Besides the

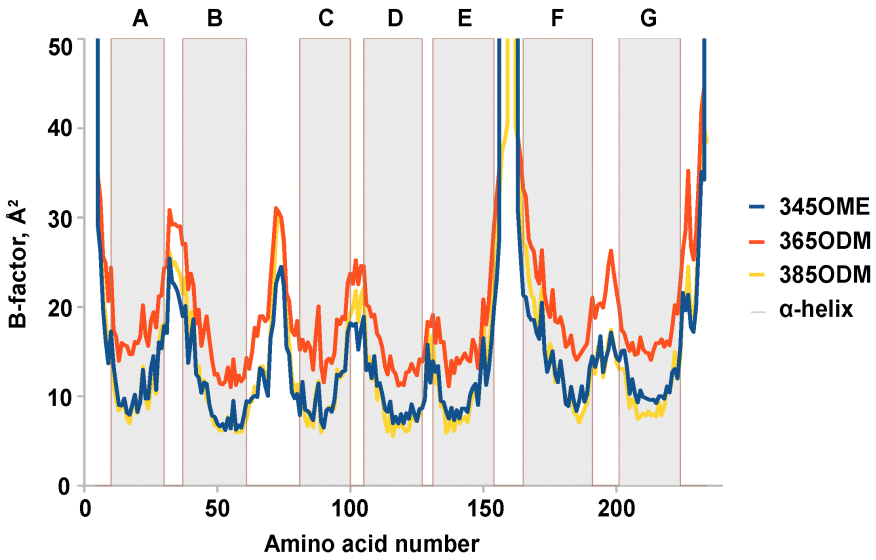

Figure 31: The plot illustrating the distribution of average amino acid B-factor for the models built. The dependence of average B-factor on the amino acid number for model 345OME is shown in blue, for model 365OME in red, and for 385OME in yellow. The α-helices of the protein are indicated by gray.

molecules of bR the model have also the water (54 for 345OME, 62 for 365ODM, 44 for 385ODM) and lipid molecules. The location of the secondary structure elements agrees with the previously published. The dependance of the B-factors of the amino acid backbone is presented on the **Fig. 31**, the mean B-factors of protein, waters and lipids are summarized in Table. 2.

Molecules of the protein form trimers packed into the layers parallel to the crystallographic plane a/b. The arrangement of the bR molecules in such layers is similar to that in native purple membranes. The organization of bR trimers is shown in **Fig. 32**. The trimers are stabilized by hydrophobic contacts between α-helices, intermolecular salt bridge between Lys40 and Asp104, and specific contacts with MO, which will be described later.

3.7.4. Lipid/detergent environment of bR molecules in the crystal.

One of the main objective of this study and, particularly, the model building was to investigate whether detergents and lipids participate in the formation of the bR crystals grown *in meso*. We have solved three structures of bR from crystals grown in different mixtures of detergents as well as in different single detergents. The lipid/detergent environment was carefully modeled and all the structures as well as electron densities were compared.

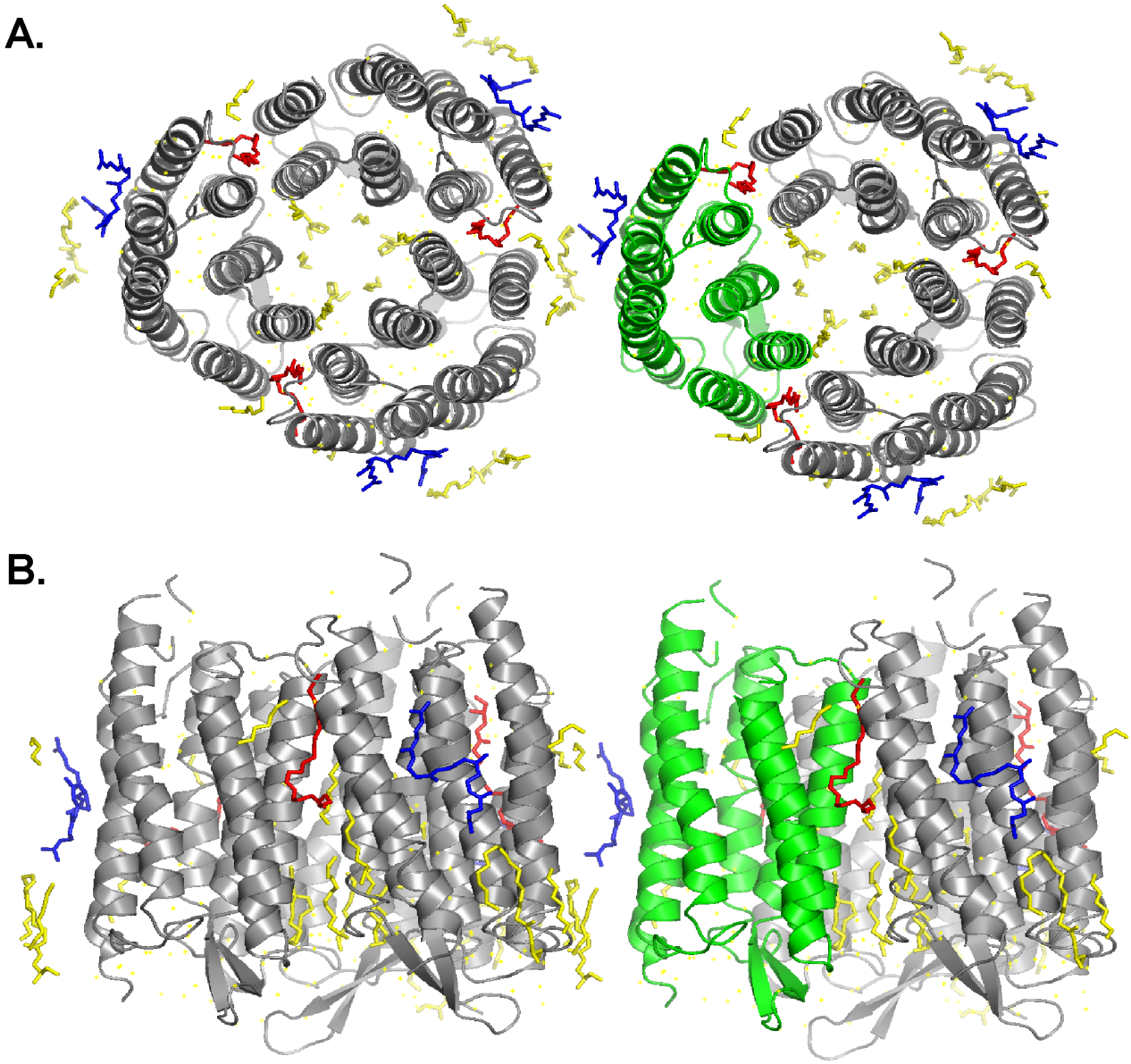

Figure 32: Organization of bR trimers in crystals and lipid environment of bR molecules. **A.** Cytoplasmic view. **B.** Membrane plane view. The bR molecule is shown in green, the symmetry related bR molecules within the trimer as well as molecules of the neighbouring trimer are shown in gray. The MO molecules are red, squalene molecules are blue, and other aliphatic chains are yellow.

On the $2F_o$-F_c electron density maps we have found a lot of elongated blobs around hydrophobic core of the protein along crystallographic axis *c*, which were interpreted as the aliphatic chains of the lipid molecules. Unfortunately, the lipid headgroups were disordered that did not allow us to identify the lipids precisely. In structure 345OME we have identified 7 aliphatic chains (total 59 atoms, average B-factor is 42.10 Å^2), in 365ODM - 7 chains (total 59 atoms, average B-factor is 53.11 Å^2), and in 385ODM - 5 chains (total 40 atoms, average B-factor is 50.17 Å^2). In the last structure it was possible to model one molecule of archaeol (B-factor is 55.44 Å^2) as well. In **Fig. 32** the molecules of lipids are shown in yellow, blue, and red. All the parts of the molecules

match well the molecules of the archaeol modeled in the published structure. Unfortunately, the headgroups were not resolved in our structures. Possible explanations are either that the lipid headgroups are not ordered properly or that the symmetry related positions are occupied by different types of the lipid molecules. In addition, in all the structures the molecule of squalene (shown in blue in **Fig. 32**) was resolved in the crevasse formed by bR amino acids on the surface of the protein near π-bulge. Its location coincides with the location of squalene in the published structures, but the B-factors are considerably higher than for the other lipids being ~65 Å^2.

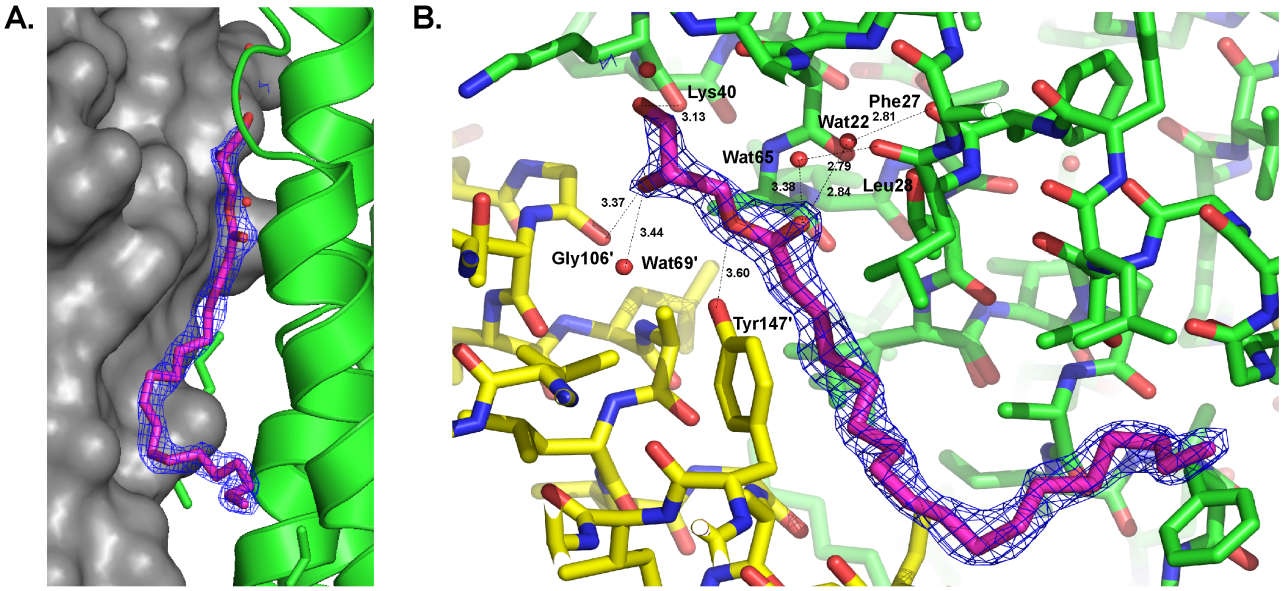

Figure 33: The molecule of MO stabilizes the trimer of bR in crystals. **A.** The position of the MO molecule between two molecules of bR in trimer. One molecule of bR is green and represented by cartoon, while the symmetry related molecule is gray and presented as a surface. **B.** The specific interactions with bR molecules that allow to the MO molecule to stabilize the trimer. The distances between atoms are indicated in angstroms, prime indicates the symmetry related bR molecule. The MO is colored magenta.

Besides native lipids we have observed the molecule of MO (**Fig. 33**). It stabilizes the trimer formed by bR molecules by hydrogen bonds with Tyr147 of one bR molecule and Phe27 of the other. Therefore, the successful use of the MO as a crystallization matrix for *in meso* crystallization may be not coincidence, as this lipid specifically bound to the bR molecules.

From the other hand, in the structures of bR obtained using the crystals grown from the different detergent mixtures (as well as single detergents) the molecules of the detergent were not found. What is more, the electron densities corresponding to the different models (obtained using different detergents) looks similar indicating that detergent molecules are not included into the crystal. This finding confirm the previous results: the clear detergent molecules were not found in the other published models of bR up to date. Mass-spectrometry also did not observed detergent in the

crystals of bR [109].

Thus, we have confirmed that detergents do not participate in the formation of the bR crystal *in cubo*. The presence of the optimal concentrations of detergent for bR crystallization can be explained by possible influence of detergent on the physical properties of the crystallization matrix. The native lipids from *H. salinarum* were shown to play important role in the formation of the crystals [109, 111], however, the fact that the headgroups are not ordered may indicate that it is non-specific hydrophobic interactions that are important, but not the type of the lipid. The crystallization of the bR expressed in *E. coli* could help to investigate this problem. The ability of this protein to produce highly ordered 3D crystals will show that only hydrophobic interactions are important in this case.

3.8 Crystallization of bR expressed in E. coli.

Having obtained the data that detergent do not influence significantly the crystallization of bR from *H. salinarum* we started the crystallization trials of bR expressed in *E. coli* and solubilized in DDM. This samples do not have native lipids from *H. salinarum* bound to bR and crystallization experiments should help to reveal the role that native lipids play in the crystallization process.

We set up crystallization trials using nanovolume *in meso* high throughput membrane

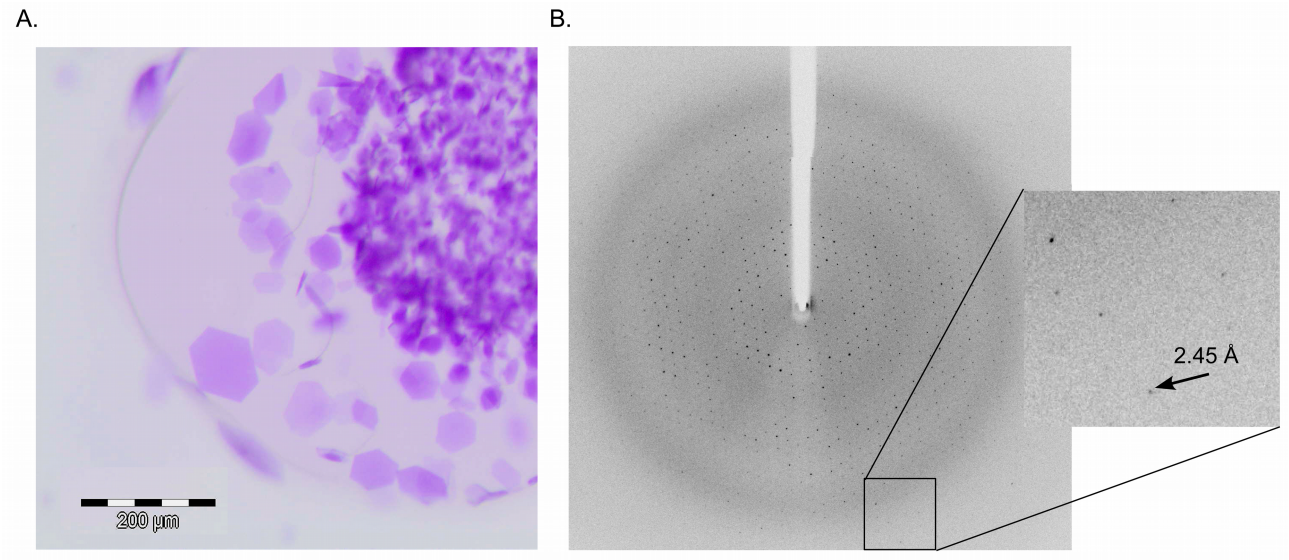

Figure 34: Crystals of bR D96N mutant expressed in *E. coli*. **A.** Crystals of bR D96N mutant expressed in *E. coli* were obtained using *in meso* nanovolume crystallization approach. The crystals we up to 120 µm. **B.** The single crystal (without further optimization of crystallization conditions) was tested at synchrotron radiation. The highest resolution seen is 2.5 Å (inset).

crystallization platform in IBS, Grenoble. The conditions used included different detergent and protein concentrations for wild type bR and D85N and D96N mutants, we have used Qigen Cubic

Phase I and Cubic Phase II precipitation solutions kits. The crystals were obtained for all three proteins. The crystals of wild type bR and D96N mutants had a shape of thin hexagonal plates up to 120 μm in longest direction. We have observed a thin needles up to 250 μm as well as star-like clusters of short thin needles. The best crystals (**Fig. 34**) were obtained using 0.1M Tris pH 8.8, 2.6M ammonium sulfate precipitation solution with protein concentration being 39 mg/ml. The D85N mutant gave only needles up to 200 μm in size. The obtained crystals were tested on synchrotron radiation (ID 14-1, ESRF, Grenoble) without further optimization of crystallization parameters. The crystals of D96N mutant diffracted up to 2.4 Å resolution, while D85N crystals up to 3.5 Å. The photos of the crystals and corresponding diffraction patterns are presented in **Fig. 34**.

The grow of the well ordered crystals of bR expressed in *E. coli* proved the highest quality of the produced protein, which is comparable to the quality of the protein solubilized from native purple membranes. Successful crystallization of the protein shows the aptness of the presented approach for expression of bR and its mutants suitable for all scientific and industrial applications. We suppose that this result may be explained by avoiding of denaturation of the protein throughout purification that is inherent to the other existing protocols [20-24]. Such approach allowed us to obtain the first reported 3D crystals of bR expressed in *E. coli*.

However, the most valuable finding is about the role of the native lipids from *H. salinarum*. Previously the archea lipids were found to be very specific stabilizing bR molecules in trimers inside the 3D crystal [109, 111]. Bacteriorhodopsin expressed in *E. coli* does not contain the specific lipids from *H. salinarum*. Moreover, the lipid compositions of *E. coli* and *H. salinarum* differ dramatically. The ability to grow the crystals of bR without involvement of *H. salinarum* native lipids indicate that stabilization of the trimers in the bR crystals is not based on specific interactions, but on nonspecific interactions of amphifilic nature. Thus, we showed that molecules of native lipids and detergents do not participate in specific interactions required for the formation of 3D crystals of bR. Apparently, native lipids are bound to bR by nonspecific amphifilic interactions forming a belt that masks the hydrophobic areas of the protein. Also, it seems that both detergents and native lipids influence the physical properties of *in meso* crystallization matrix in such a way affecting the process of crystallization.

Having obtained this information we started the crystallization trials with bRLE and bRLEs proteins using nanovolume *in meso* high throughout crystallization system. The crystallization trials are ongoing.

4 Materials and methods

4.1 Materials.

All the salts and media components were either from AppliChem (Darmstadt, Germany) or Sigma-Aldrich (St. Louis, USA) of analysis quality or higher. All enzymes were from Thermo Fisher Scientific (Rockford, USA). DM, DDM, MEGA10, CYMAL5 was from Affymetrix (Santa Clara, USA), Sarcosyl from AppliChem, retinal, DMPC, and CHAPS from Sigma-Aldrich.

4.2 DNA manipulation

Polymerase chain reactions were carried out using Phusion Hot Start II DNA Polymerase from Fermentas, Thermo Scientific (Rockford, USA). All the primers were synthesized by MWG Operon, Eurofins (Ebersberg, Germany). Commonly, to 37.5 µl of sterile water we added 1 µl of matrix, 5x Phusion High Fidelity Reaction buffer, 20 pM of each primer, 5 nM of each deoxyribonucleotide, and 1 unit of polymerase. Total volume was 50 µl. Reactions were carried out using Arktik Thermal Cycler from Thermo Scientific (Rockford, USA). Program used was usually as follows: 120 s at 98°C to activate enzyme, then 40 cycles of reaction and final elongation at 72°C for 300 seconds. The body of the cycle consisted of denaturation step at 98°C for 20 seconds, annealing step at 66-70°C for 20 seconds and elongation step at 72°C for 40-80 seconds depending on the length of the amplified region.

Restriction enzymes used were of mostly Fast Digest line from Fermentas, Thermo Scientific (Rockford, USA). Typically, reaction mixture was composed from DNA solution, sterilized mQ water, 10x Fast Digest buffer and 1 µl of enzyme. Total volume was generally 20 µl for plasmid and 40 µl for PCR product. Reaction mixture was incubated at 37°C from 0.5 to 2 h depending on the amount of DNA to be cut.

Ligation of the DNA fragments was achieved using T7 DNA Ligase from Fermentas, Thermo Scientific (Rockford, USA). Typically, DNA fragments in water or low salt buffer were mixed in appropriate ratios, T7 Ligase buffer was added according to final volume of the sample and then 1 µl of ligase was added. Reaction mixture was incubated at 20-22°C at least 2 hours. Generally, DNA insert was in 3-5 excess in respect to the vector DNA.

A horizontal agarose gels were used for visual analysis of DNA. Agarose was from Bio-Rad

(Hercules, USA). Gels were prepared from TAE buffer (40mM Tris, 20mM acetic acid and 1mM EDTA, pH 8.5) with 1% agarose and 1:20000 of GelRed solution from Biotium (Hayward, USA) for in-gel staining. The DNA fragments were separated applying 10 V/cm voltage to the gel. When needed DNA fragments were cut and eluted from gel using NucleoSpin Gel and PCR clean-up Kit from Macherey-Nagel (Dueren, Germany). The signal was visualized using manual gel documentation system InGenius from Syngene (Cambridge, UK).

Top10 *E. coli* strain (Invitrogen, Carlsbad, USA) was used as a cloning strain. Chemically competent cells we prepared in-house according to a standard protocol [141]. For transformation 1 µl of plasmid solution or 2-5 µl of ligation mixture were added to frozen cells. They were incubated on ice for 30-60 minutes and after a heat-shock at 42°C for 90 seconds the cells were incubated for an additional 5 minutes. Then 800 µl of SOC media was added and cells were let to recover at 37°C for 30 minutes with gentle shaking. Then cells were pelleted and resuspended in 50 µl SOC media and applied on LB-agar plates with suitable antibiotic.

SE1 *E. coli* strain (Delphi Genetics, Charleroi, Belgium) was used as expression strain in combination with pSCodon plasmid (Delphi Genetics, Charleroi, Belgium). Competent cells were prepared by supplier and used according to supplier recommendations. Shortly, transformation protocol was essentially the same as with Top 10 strain except 350 ml of the supplied recovery media was added to the cells for recovery and cells were not pelleted before applying on agar plate.

BL21 *E. coli* strain were used for expression of SRII protein and its derivative. The competent cells were prepared in-house by TSS-procedure [142] and transformation protocol was the same as for Top 10 strain.

4.3 Cloning of genetic constructs

Plasmid pEF191 containing wild type bR nucleotide sequence was kindly provided by D. Oesterhelt [99]. Plasmid pET27bmod_SOPII containing wild type SRII with *His*-tag appended to its C-terminus was kindly provided by M. Engelhard [27]. Stabilized lysozyme gene [143] optimized for expression in *Halobium Salinarum* and MISTIC (M110 from [115]) were synthesized by MWG Operon, Eurofins (Ebersberg, Germany). All the DNA manipulations were accomplished as it was described above and the accuracy of all constructs was proved by sequencing of the modified parts of the plasmids by MWG company. The nucleotide sequences of the major constructs (only open reading frames) are presented in Appendix.

4.3.1. pEF191 constructs for homologous expression of bR in *H. salinarum*.

To obtain the pEF191-bRL plasmid the synthesized Lys was inserted into pEF191 plasmid at *NsiI* and *ClaI* restriction sites. We introduced these sites into the plasmid by PCR using primers containing these sites. The gene that was synthesized by MWG company already contained these sites and was cut out directly by restriction enzymes. Three insertions containing bRI and bRII parts and Lys gene were ligated at *BamHI* and *HindIII* sites.

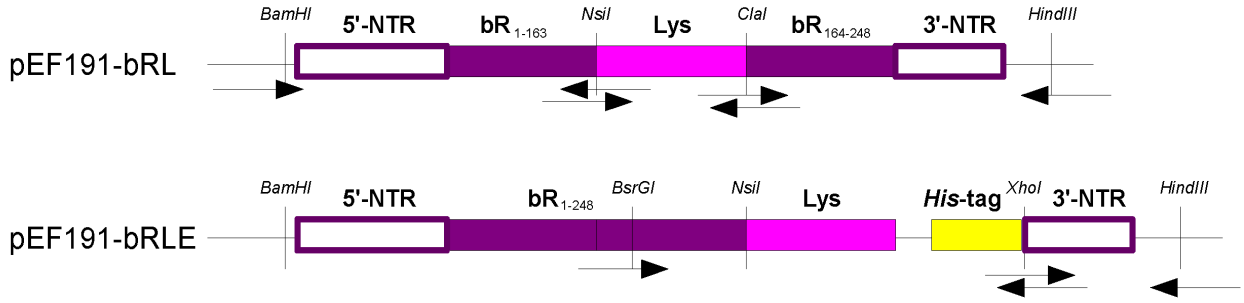

Figure 35: Cloning of pEF191 based plasmids bearing the gene of bR-Lys fusion proteins. In bRL fusion protein Lys is put into bR between Met163 and Arg164, while in bRLE construct it is appended to the bR gene. The 5'-NTR and 3'-NTR of the bR gene from *H. salinarum* are shown in purple hollow rectangles, Lys is pink, bR is purple, *His*-tag is yellow. The restriction enzymes are indicated by vertical lines, while PCR primers indicated by arrows.

The pEF191-bRLE plasmid was obtained using *His*-taged Lys (not described here). The PCR was carried out using primers complementary to the bR gene before *BsrGI* restriction site and *His*-tag with *XhoI* site. The *XhoI* restriction site was added to the 3'-NTR using PCR with primer containing this site. Two insertions containing Lys gene and 3'-NTR were ligated into pEF191-bRLE plasmid at *BsrGI* and *HindIII* sites.

4.3.2. Constructs containing MISTIC-bR-Lys fusion protein.

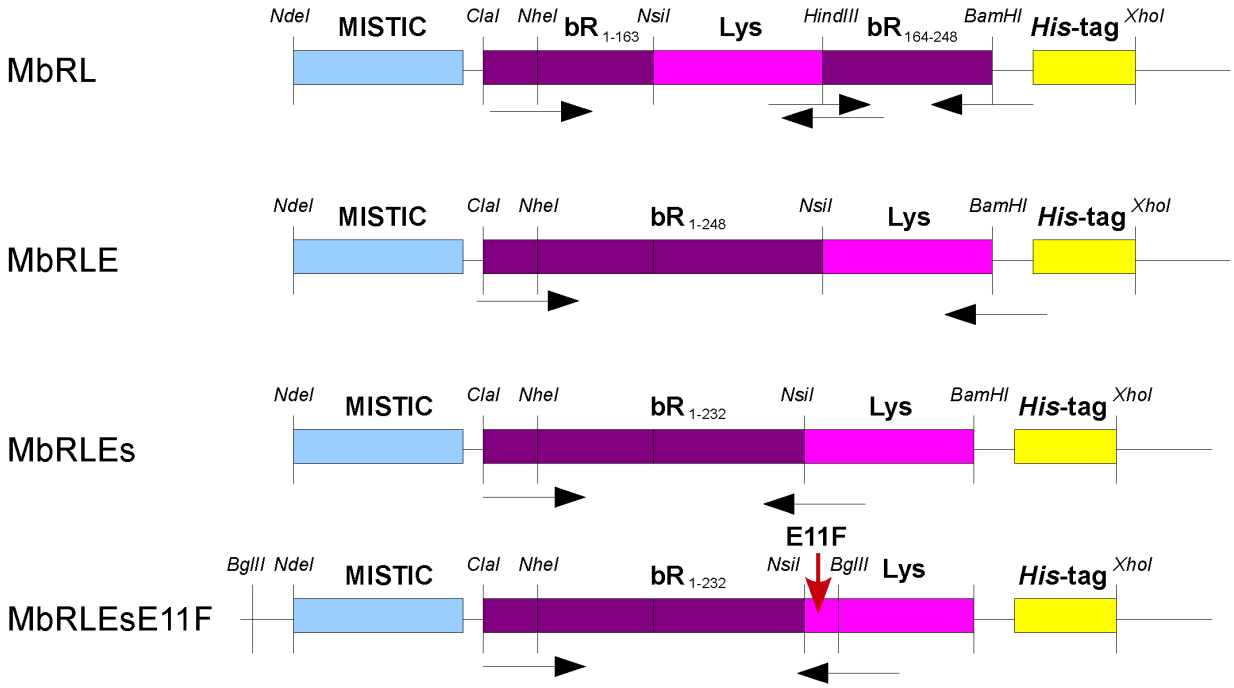

Figure 36: Cloning of pSCodon based plasmids bearing the gene of MISTIC-bR-Lys fusion proteins. In MbRL fusion protein Lys is put into bR between Met163 and Arg164, while in MbRLE and MbRLEs construct it is appended to the native and truncated at Glu232 bR gene, respectively. MbRLEsE11F construct contains E11F mutation in Lys gene that abolishes activity of Lys. The MISTIC gene is light blue, Lys is pink, bR is purple, *His*-tag is yellow. The restriction enzymes are indicated by vertical lines, while PCR primers are indicated by arrows. The E11F mutation in Lys is shown by red arrow.

The MISTIC gene was synthesized by MWG company, cloned into pSCodon vector containing *His*-tag and was routinely used in our laboratory for expression of different fusion proteins. This construct contains Factor Xa protease sites (IEGR) separated from MISTIC by 9 amino acid linker GPGGSSGAS and from the target protein and *His*-tag by GSS and SGAP linkers, respectively. Thus, after proteolys with Factor Xa the *N*-terminus of target protein would not contain any additional amino acids, while *C*-terminus would contain GSSIEGR amino acids. The MbR fusion protein gene inserted into pSCodon vector. Using this plasmid we carried out cloning to obtain MISTIC-bR-Lys fusion protein genes according to the strategy presented in **Fig. 36**.

The MbRL construct was obtained from pEF191-bRL construct using two PCR. First reaction amplified bRI part of bR including *NheI* restriction site and Lys introducing *HindIII* restriction site appended to the Lys gene, while second reaction introduced *HindIII* and *BamHI* restriction sites flanking bRII part of bR. The pSCodon-MbR plasmid was cut at *NheI* and *BamHI* restriction sites and two insertions containing bRI and Lys and bRII genes were ligated into pSCodon vector giving

pSCodon-MbRL plasmid.

The MbRLE construct was obtained using pEF191-bRLE construct. The PCR amplified the bR gene including *NheI* restriction site and Lys gene introducing *BamHI* restriction site (**Fig. 36**). The *NheI*/*BamHI*-cut PCR product was ligated into pSCodon vector using these sites. The MbRLEs construct was obtained from MbRLE by PCR with the same forward but different reverse primer that appended *NsiI* restriction site to the Glu232 of bR. The *NheI*/*NsiI* insertion was ligated into pSCodon vector obtained from pSCodon-MbRLE cut at *NheI* and *NsiI* restriction sites.

To introduce E11F mutation we used PCR. The closest suitable restriction site was *BglII* located just 13 bp downstream. We amplified bR gene including *NheI* restriction site and the beginning of Lys gene up to *BglII* restriction site using reverse primer containing E11F mutation. The *NheI*/*BglII*-cut PCR product was ligated to the same sites of pSCodon-bRLEs plasmid. The *BglII* restriction site is present in pSCodon plasmid as well. The *NheI*/*BglII*-cut plasmid had three fragments: 6162, 710, and 512 bp. We mixed 6162 and 512 bp fragments and PCR products for ligation.

4.3.3. Construction of chimeric proteins.

The realization of the complementary protein approach required the construction of the chimeric proteins between SRII and bR. We have costructed chimeric proteins SR_{1-44}bR, SR_{1-10}bR, $SR_{1-10,\ 28-44}$bR, SR_{28-44}bR, and bR_{1-8}SR according to the strategies presented in **Fig. 37**.

The wild type bR gene was used as a negative control of bR expression. The construct was produced from MbR construct by PCR using positive primer introducing *NdeI* restriction site to the 5'-terminus of bR gene and reverse primer complementary to the DNA downstream of the *BsrGI* restriction site. The *NdeI*/*BsrGI*-cut PCR product was ligated to the pSCodon-MbR plasmid cut by the same enzymes.

The chimeric protein SR_{1-44}bR was constructed by PCR. By silent mutations G144A and C147A of bR we introduced *SpeI* restriction site into bR gene: the first PCR added the *NdeI* and *SpeI* restriction site flanking SRII gene region coding amino acids from 1 to 44, while the second PCR introduce *SpeI* site into bR gene. Two inserts were ligated into pSCodon-bR plasmid at *NdeI* and *BsrGI* sites.

First ten amino acids of SRII were introduced into bR gene by PCR using 50 bp long forward primer containing *NdeI* restriction site and first 30 nucleotides of SRII. The pSCodon-bR plasmid

was cut by *NdeI* and *BsrGI* restriction enzymes and PCR product cut by the same enzymes was ligated into it.

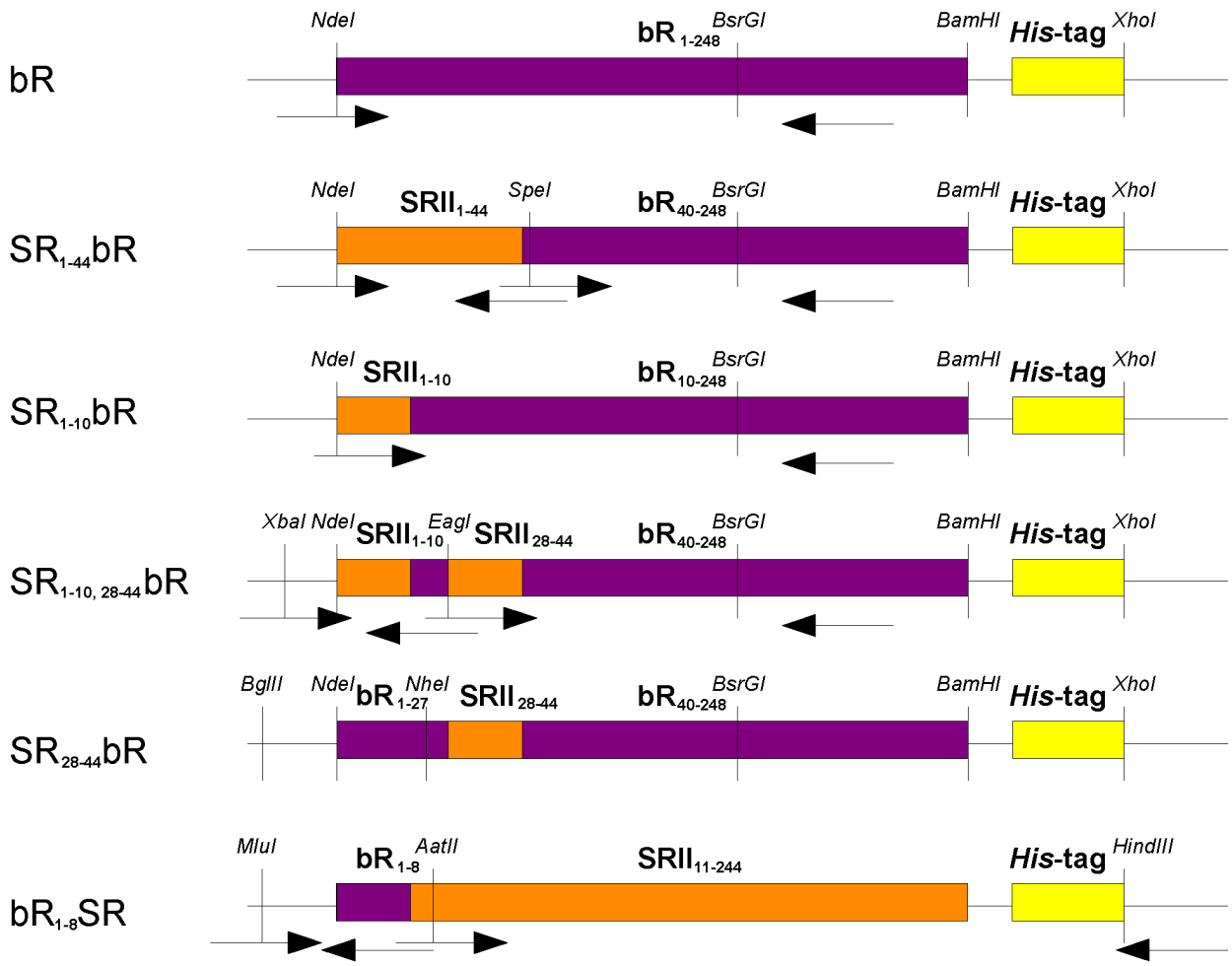

Figure 37: Cloning of the chimeric proteins between SRII and bR. The subscripts indicate which amino acids of SRII were introduced to bR and *vice versa*. The **Fig. 17** shows the amino acid sequence of the *N*-terminus of the constructed chimeric proteins. The SRII gene is orange, bR is purple, *His*-tag is yellow. The restriction enzymes are indicated by vertical lines, while PCR primers indicated by arrows.

The cloning of the $SR_{1\text{-}10,\ 28\text{-}44}$bR chimeric protein was accomplished by two PCRs. Fortunately, the junction of bR and SRII genes contained *EagII* restriction site. We carried out first PCR using pSCodon- $SR_{1\text{-}10}$bR plasmid as a matrix, thus amplifying the amino acids from 1 to 10 of SRII inside bR gene. The second PCR was carried out using pSCodon- $SR_{1\text{-}44}$bR plasmid as a matrix, thus amplifying the amino acids from 28 to 44 of SRII inside bR gene. These two inserts were ligated at *XbaI* and *BsrGI* sites into pSCodon-bR plasmid. The $SR_{28\text{-}44}$bR chimeric protein gene was easily obtained from $SR_{1\text{-}10,\ 28\text{-}44}$bR gene replacing its *BglII*/*NheI* fragment by counterpart from pSCodon-bR plasmid that do not contain amino acids from SRII on the *N*-terminus of the protein.

To investigate the influence of bR *N*-terminus on expression of SRII in *E. coli* we have constructed also the chimeric protein $bR_{1\text{-}8}SR$. The amino acids from 1 to 8 were introduced to SRII by PCR using 60 bp long reverse primer containing required nucleotide sequence from bR and *AatII* restriction site. Two PCR fragments were ligated into pET27bmod plasmid at *MluI* and *HindIII* restriction sites.

4.3.4. Cloning of R8E and R8Q mutants, optimization of bR mRNA, and mutation of mRNA optimized bR.

The cloning of the R8Q and R8E point mutants of bR was accomplished similar to the construction of pSCodon-bR plasmid (**Fig. 37**). We used two different forward primers containing CAA and GAA nucleotides (instead of CGT in bR) for R8Q and R8E mutants, respectively. The inserts were ligated into pSCodon-bR plasmid at *NdeI* and *BsrGI* restriction sites.

The optimization of mRNA was carried out using *mRNAshapes* [128] software. Previously [127], it was shown that stabilization of the (-4, 37) region of mRNA can significantly reduce expression of the protein. Using *mRNAshapes* we have found that 5'-terminus of bR contains stem structure that extended further than (-4, 37) region. Thus, we had to extend the region under investigation from (-4, 37) to (-4, 47) to include the observed stem. To compare the stabilization of the *N*-terminus of the different proteins we calculated the mRNA folding free energies of the (-4, 47) region. Silent mutations C9A and G12A were shown to disrupt the stem structure destabilizing the 5'-terminus of bR mRNA.

To introduce C9A and G12A mutations to the bR gene the two oligonucleotides (for sense and antisense strands) were synthesized to fill the gap between *NdeI* and *NheI* restriction sites. The oligos were mixed and ligated into pSCodon-bR plasmid cut by these enzymes.

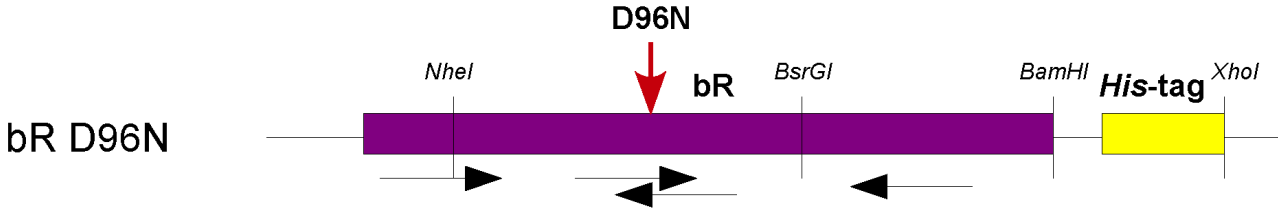

Figure 38: Cloning of the bR mutants using D96N mutant as example. The gene bR is purple, *His*-tag is yellow. The restriction enzymes are indicated by vertical lines, while PCR primers indicated by arrows.

Having obtained functional expression of bR in *E. coli* we started to produce the mutants required for the project devoted to the intermediate states of bR. We obtained V49A, D85N, and

D96N mutants.

The D96N mutant was constructed using PCR in two steps as there was no suitable restriction site nearby. At first, two PCR were carried out as illustrated on **Fig. 38**. First reaction introduced the D96N mutation with reverse primer amplifying nucleotide sequence from *NheI* site to D96N mutation site. Second reaction introduced the D96N mutation with forward primer amplifying nucleotide sequence from D96N mutation site to *BsrGI* restriction site. The reverse primer of first reaction and forward primer of second reaction were designed to contain 18 nucleotides overlapping. After in-gel purification of the PCR products we mixed them to use these products as a matrix for PCR, added primers (forward of first reaction and reverse of second reaction), and carried out new PCR. The resulting elongated PCR product was cut at *NheI* and *BsrGI* sites and ligated into pSCodon-bR plasmid using these sites. The V49A mutation was introduced similarly. The D85N mutant of bR was already present in our laboratory so we have just cut it at *NheI* and *BsrGI* sites and ligated this insertion to pSCodon-bR plasmid at the same sites.

4.3.5. Cloning of the bR-Lys fusion proteins.

As we already had all the required parts to accomplish the construction of the bR-Lys fusion proteins genes we utilized only restriction reaction to produce these genes. Also we wanted to introduce the E11F mutation into Lys to abolish the enzymatic activity of this protein. The strategy of the cloning is presented on **Fig. 39**.

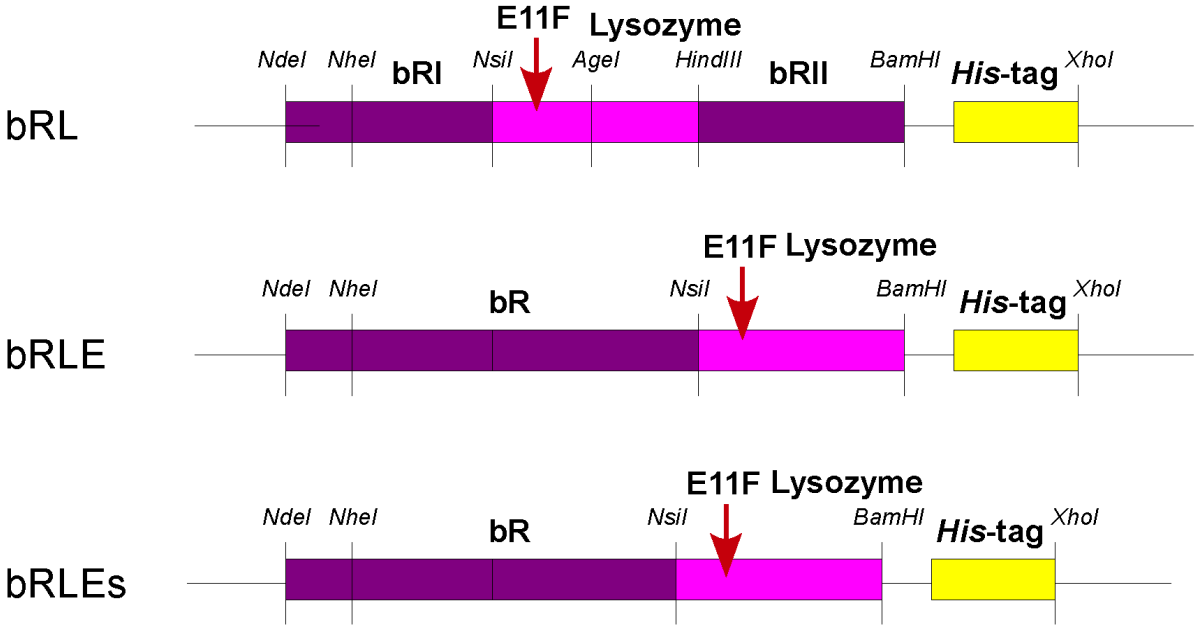

Figure 39: Cloning of the bR-Lys fusion proteins with RNA optimized bR. Mutation E11F was introduced into all three variations of fusion proteins. The Lys gene is pink, bR is purple, *His*-tag is yellow. The restriction enzymes are indicated by vertical lines, while PCR primers indicated by arrows. Location of E11F mutation in Lys gene is indicated by red arrow.

The bRL, bRLE, and bRLEs fusion protein genes were constructed using constructs MbRL, MbRLE, MbRLEs, and MbRLEsE11F constructs presented in **Fig. 36**. All three new fusion proteins contained E11F mutation in Lys gene and RNA optimized bR gene. We constructed first the bRLEs construct just exchanging the *BsrGI*/*XhoI* fragments of pSCodon-bR^{RNA} plasmid containing RNA optimized bR gene and pSCodon-MbRLEsE11F plasmid.

To construct the bRLE construct we took the *NheI*/*NsiI* fragment from MbRLE construct and *NsiI*/*XhoI* fragment from bRLEs construct and ligated them into pSCodon-bR_RNA vector at the *NheI* and *XhoI* sites. The bRL construct was assembled from three insertions and pSCodon-bR_RNA vector. The inserts were: 1) *NheI*/*NsiI* fragment of MbRL construct; 2) *AgeI*/*XhoI* fragment of MbRL construct; 3) *NsiI*/*AgeI* fragment of bRLE construct. The first two fragments introduced *NsiI* and *HindIII* restriction sites into bR gene as well as the 3'-terminus of Lys gene, while third fragment contained E11F mutation that was absent in MbRL construct. Three inserts were mixed and ligated into pSCodon-bR_RNA vector at the *NheI* and *XhoI* sites.

4.4 Transformation of H. salinarum and screening of clones.

The transformation of *H. salinarum* L33 strain with pEF191 plasmid bearing genes of bR-Lys fusion proteins was carried out according to [144] with slight modifications. In brief, the cells were

grown in 1% pepton L37 from Oxoid, Thermo Fisher Scientific (Rockford, USA), 4.3 M NaCl, 80 mM $MgSO_4$, 27 mM KCl, 10 mM $Na_3C_6H_5O_7$ pH 6.5 until the OD_{560} of 0.8. The 2 ml of cells were harvested by centrifugation and resuspended in 150 µl of 2M NaCl, 27mM KCl, 15% (w/v) sucrose buffer. Then 15 µl of EDTA was added, then after 10 minutes 5 µg of plasmid DNA was added in 10 µl of 1M NaCl, 13.5 mM KCl, 7.5% (w/v) sucrose solution. After 5 minutes 175 µl of 60% PEG_{600} was added, cells were shaked vigorously and incubated for 30 minutes. The cells were recovered overnight at 37°C in 10 ml of growth media with 15% sucrose. Then the cells were collected by centrifugation and plated on the agar plates containing 3 µg/ml mevinolin from Sigma-Aldrich (St. Louis, USA). The first colonies were seen after 14 days at 37°C.

The screening of clones was accomplished using PCR amplification of genomic DNA. First, the genomic DNA was isolated from cells lysated by osmotic shock and purified by sodium acetate precipitation. Then the genomic DNA was cut to smaller pieces by *HindIII* restriction enzyme which is not present in bR and Lys gene. The restriction mixture was taken as a matrix for PCR with (5'-GGTTTGGAATCCGCGTGTCGGCTC-3') forward primer and either (5'-GCTGTCCTGCGTGTGCGATCAGTC-3') or (5'-GTGTCGTTGAAAAGCCGCGCCGGTT-3') reverse primers for bRL and bRLE constructs, respectively. To differentiate correct clones the PCR product was cut by *BglII* restriction enzyme. According to the amount and size of the bands from cut PCR product on agarose gel we could distinguish if the genetic recombination took place. Finally, the PCR product was sequenced by MWG company using the same primers as were used for PCR to prove accuracy of genetic recombination.

4.5 Growth media and cultivation of E. coli cells.

During DNA manipulations we mostly used LB and SOC media. The LB media contains 1% w/v bactotryptone, 1% w/v yeast extract, 1% w/v NaCl. The pH of the LB media is adjusted to 7.0 with NaOH. Generally, to suppress the expression of the target protein in culture we added to the media also 2% w/v glucose. The SOC media contains 2% w/v bactotryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$. The pH of the SOC media is adjusted to 7.0 with NaOH.

The cells were plated on LB agar plates made from the described LB media with addition of 1.5% w/v agar and proper antibiotic. The pSCodon vector contains gene providing the resistivity to ampicillin, we used ampicillin concentrations between 100 and 250 µg/ml. The pET vector contains gene providing the resistivity to kanamycin, we used kanamycin concentration of 50 µg/ml.

For protein expression we used the ZYP-5052 autoinduction media according to [145]. The media was composed from ZY media (1% w/v bactotryptone, 1% w/v yeast extract) with stock solutions 20×P (1 M Na_2HPO_4, 1 M KH_2PO_4, 0.5 M $(NH_4)_2SO_4$) and 50×5052 (25% glycerol, 2.5% glucose, 10% α-lactose monohydrate) and supplied with suitable antibiotics. The expression of the membrane proteins in preparative scales was carried out in 2L baffled flasks with 400 ml ZY media using Infors HT Minitron shaker (Bottmingen/Basel, Switzerland) at 120 rpm.

Several colonies of cell culture from a fresh plate (either fresh transformation of the cells with a plasmid or cells freshly grown from stock at -80°C) were used to inoculate 20 ml of the fresh media for overnight preculture that was grown ~16 h at 37°C and vigorous shaking in baffled flasks. Next morning the cell culture was diluted 20-40 fold with a fresh media and grown at 37°C and vigorous shaking until reached the OD_{600} of 2.0. Then the culture was diluted again 20 fold with a fresh media without glucose (maximal residual concentration of glucose is 0.05% w/v) and grown at 37°C and vigorous shaking until reached the OD_{600} of 1.5. At this moment we decreased the temperature to 20°C, added retinal (50 mM of retinal in ethanol solution, 400 µL per flask) if needed, and let the cells to grow overnight. The cells were harvested by centrifugation at 5 000 rpm for 30 min in JLA-8000 rotor for Avanti J-26 XP preparative centrifuge (Beckman, Brea, USA).

4.6 Protein purification.

4.6.1. Protein handling.

The protein probes were taken at different stages of protein purification. During protein solubilization and affinity chromatography we took the amount of the sample normalized to the amount of total cell lysate. Generally, we take 40 µl of the total lysate (cell pellets from liter of culture were resuspended in 50 ml of buffer) and prepare from it SDS-PAGE probe with the total volume of 200 µl. All the following samples are made to contain the same amount of total lysate. Probes of eluates from Ni-NTA column (Qigen, Hilden, Germany) are generally highly concentrated to estimate the amount of impurities and, thus, not normalized to the lysate probes.

The probes of total lysate, cytoplasm, solubilized and nonsolubilized fractions were precipitated using trichloroacetic acid. For this we added 2.25×V of acetone and 0.35×V of trichloroacetic acid to a protein probe. The samples were incubated at -80°C for at least 2 h, then protein was precipitated by centrifugation at 14 000 rpm in a bench centrifuge. The pellets were washed with ice-cold acetone twice and then resuspended in 200 µl 1×SDS-PAGE loading buffer.

For SDS-PAGE analysis we used in-house made 8-16% discontinuous gradient SDS polyacrylamide gels. From 5 to 20 µl of protein probe was loaded per lane and the ~10 V/cm voltage was applied. The gels were stained with coomassie brilliand blue from Sigma-Aldrich (St. Louis, USA) and destained with 3% acetic acid.

For Western blotting we used a liquid transfer from gel to nitrocellulose membrane. The voltage applied was ~1.8 V/cm^2. The transfer solution contained 47.9 mM Tris-base, 38.6 mM glycine, 0.0385% SDS, 20% methanol. After transfer nitrocellulose membrane was incubated with the 5% solution of non-fat dry milk (Bio-Rad Laboratories, Hercules, USA) in TBST buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.05% Tween 20, 5% glycerol). The primary antibodies were monoclonal *His*-probe antibodies produced in mouse from Santa Cruz Biotechnology (Dallas, USA), while the secondary antibodies were anti-mouse IgG- horseradish peroxidase conjugated antibodies produced in goat from Sigma-Aldrich (St. Louis, USA). The membranes were stained with NBT/BCIP Substrate Solution from Thermo Scientific (Rockford, USA).

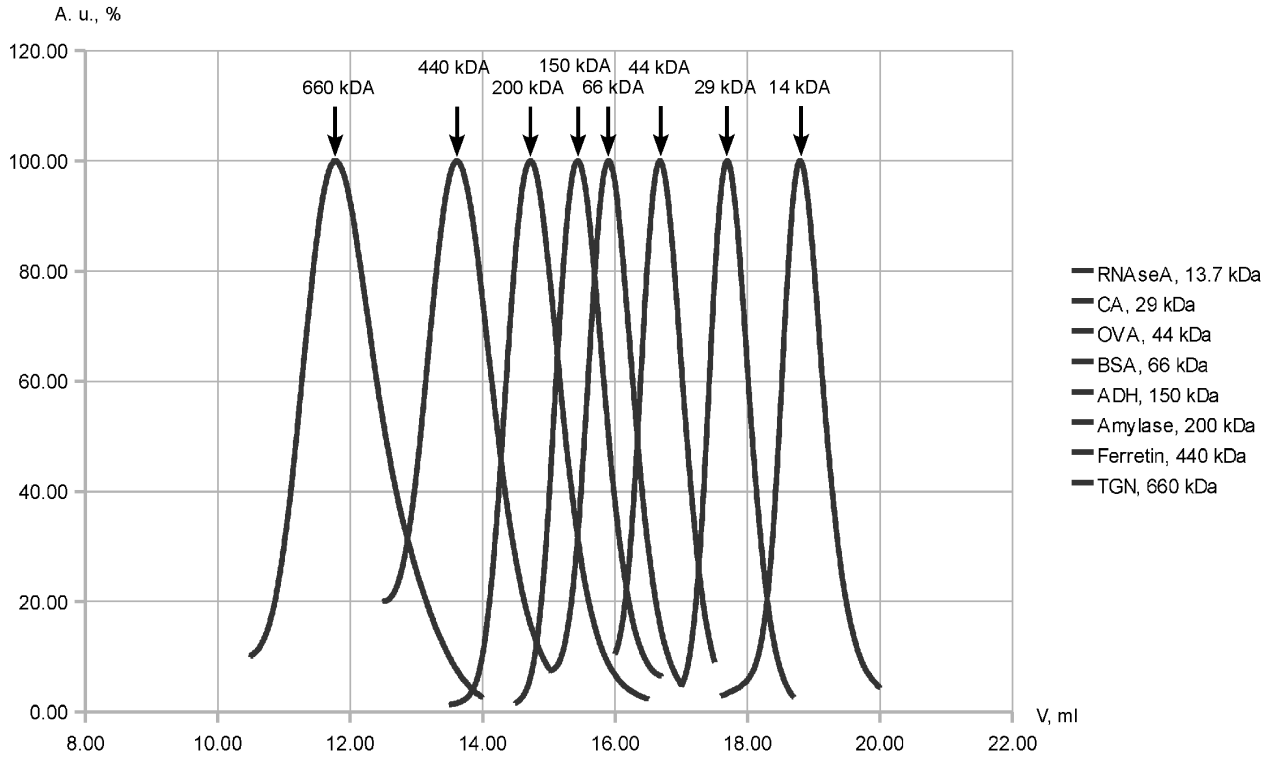

Figure 40: Calibration of Superose6HR column. The proteins used: RNAseA (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (44 kDa), albumin (66 kDa), alcohol dehydrogenase (150 kDa), amylase (200 kDa), ferretin (440 kDa), and thyroglobulin (660 kDa).

Size exclusion chromatography was carried out using Superose 6pg column with 180 ml bed volume (preparative scale), Superose6HR column with 24 ml bed volume (analytical scale), Sephacryl S200HR column with 165 ml bed volume (preparative scale), all from GE Healthcare

(Little Chalfont, UK). The column run and fraction collection was automated using ÄKTAprime plus system from GE Healthcare (Little Chalfont, UK). For ion-exchange chromatography we used prepacked HiTrap Sp Hp column with 1 ml bed volume from GE Healthcare (Little Chalfont, UK). The analytical column Superose6HR was calibrated with Gel Filtration Markers Kit for Protein Molecular Weights 12-200 kDa as well as RNAseA (~14 kDa), ovalbumin (44 kDa), ferretin (440 kDa), and thyroglobulin (660 kDa) all from Sigma-Aldrich (St. Louis, USA) as presented on **Fig. 40**. The preparative scale columns were not calibrated.

To estimate the protein concentration we used SDS-PAGE densitometry, UV-Vis spectroscopy, and BCA Protein Assay Kit. To estimate the purity and concentration of the Ni-NTA purified samples we scanned the gel using Epson Perfection V750 Pro scanner (Suwa, Japan). The images were transferred to grayscale and inverted, then the protein band densities were calculated using ImageJ software. Total protein concentration was measured by BCA Protein Assay Kit from Thermo Fisher Scientific(Rockford, USA) following supplier protocol. UV-Vis absorbance spectrum was measured on Shimadzu UV-2450 spectrophotometer (Kyoto, Japan). The fraction of the functional protein was accessed as absorbance ratio $A_{280}/A_{\lambda max}$, where $A_{\lambda max}$ is maximum absorbance of retinal [146]. Extinction coefficient for retinal in *H.salinarum* purple membranes was assumed to be 42 000 $cm^{-1}M^{-1}$ [11], for solubilized bR 63 000 $cm^{-1}M^{-1}$ [147], while for bR in DMPC/CHAPS bicelles 55 300 $cm^{-1}M^{-1}$ [125].

4.6.2. Screening of detergents for solubilization of membrane proteins.

To establish the purification protocol for particular MP we should first select the proper detergent for its solubilization. As the yield of MPs is generally rather low in order to visualize the efficiency of solubilization by different detergents we employed the additional purification step of Ni-NTA affinity chromatography before loading the samples on SDS-PAGE.

The cell pellets from 100 ml of cell culture were resuspended in 20 mM Tris-HCl pH 8.0, 5% glycerol buffer adjusting the volume of cell suspension to 50 ml per each liter of cell culture. The cells were homogenized by vigorous stirring, then 1 mg of DNAse I and 10 mg of lysozyme, both from Sigma-Aldrich (St. Louis, USA), were added for each liter of cell culture. The cell suspension was incubated for 2 h at 4°C with stirring and then lysated by three passes through either French Press from SLM-Aminco (Irvine, USA) or micro-fluidizer M-110P from Microfluidics (Newton, USA). The cell suspension was divided to equal parts and the detergents were added directly to the suspension. Generally we used 1% DM, 1% DDM, 1% FOS-10, 1% FOS-12, 1% LDAO, 1% OG,

2% Sarkosyl, and 1% SDS. After detergents dissolved in suspension we incubated the samples overnight at 4°C (except SDS sample which was kept at room temperature) for complete solubilization.

Insoluble material was removed by ultracentrifugation in Ti-70 (Beckman, Brea, USA) rotor at 35 000 rpm for 1h. Supernatant was 5 times diluted with 20 mM Tris-HCl pH 8.0, 100 mM NaCl buffer. Suspension was loaded in batch mode on the 0.5 ml of Ni-NTA resin (Qiagen, Hilden, Germany) equilibrated with the same buffer. The column was first washed with 3 CV of 20 mM Tris-HCl pH 8.0, 100 mM NaCl buffer supplied with the required detergent. We used concentrations 0.1% DM, 0.1% DDM, 0.1% FOS-10, 0.1% FOS-12, 0.1% LDAO, 0.8% OG, 0.25% Sarkosyl, and 0.1% SDS. The protein was eluted with 3 CV of 20 mM Tris-HCl pH 8.0, 300 mM imidazole buffer supplied with the required detergent. The eluate samples were loaded on SDS-PAGE. The gels were stained with coomassie, the Western blotting with anti-*His* antibodies was carried out as well.

4.6.3. Protein purification under denaturing conditions.

When it was established that Sarkosyl is the best detergent for solubilization of different bR constructs we optimized purification protocol to obtain better purity of the protein on preparative scale. For this purpose we isolated membranes from total lysate, washed the Ni-NTA column with imidazole containing buffer and used larger volumes of cell culture. When the in-column detergent exchange from Sarkosyl to SDS was required we used another sequence of washing buffers.

The cell pellets from 1 to 5 liters of cell culture were resuspended in 20 mM Tris-HCl pH 8.0, 5% glycerol buffer adjusting the volume of cell suspension to 50 ml per each liter of cell culture. The cells were homogenized by vigorous stirring, then 1 mg of DNAse I and 10 mg of lysozyme were added for each liter of cell culture. The cell suspension was incubated for 2 h at 4°C with stirring and then lysated by three passes through either french press or micro-fluidizer M-110P (Microfluidics, Newton, USA). Then 5M NaCl was added to a final concentration of 200 mM and suspension was layered over a glycerol cushion (1 ml - 90%, 1 ml - 80%, 1ml - 60%) in two 32 ml tubes. The total membranes were isolated by ultracentrifugation in SW-32Ti rotor (Beckman, Brea, USA) at 28 000 rpm for 1h. The supernatant was discarded and glycerol cushion containing membranes was resuspended in 50 ml of 20 mM Tris-HCl pH 8.0, 100 mM NaCl. Sarkosyl was added to the final concentration of 2% and membranes were solubilized overnight with stirring at 4˚C.

Insoluble material was removed by ultracentrifugation in Ti-70 (Beckman, Brea, USA) rotor at 35 000 rpm for 1h. Supernatant was 5 times diluted with 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM imidazole buffer. Suspension was loaded in batch mode on the 5 ml of Ni-NTA resin (Qiagen, Hilden, Germany) equilibrated with the same buffer. The column was first washed with 3 CV of 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM imidazole, 0.25% Sarkosyl. The protein was eluted with 3 CV of 100 mM Na_2HPO_4 pH 8.0, 100 mM NaCl, 300 mM imidazole, 0.25% Sarkosyl.

When the detergent exchange was required, the sequence of washing buffers was as follows: 3 CV of 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.25% Sarkosyl, 20 mM imidazole buffer, then 5 CV of 100 mM Na_2HPO_4 pH 8.0, 100 mM NaCl, 0.2% SDS, and 3 CV of 100 mM Na_2HPO_4 pH 8.0, 0.2% SDS buffer. The protein was eluted with 3 CV of 100 mM Na_2HPO_4 pH 8.0, 0.2% SDS, 300 mM imidazole. To remove imidazole the samples were dialysed against 0.8 L of 50 mM NaH_2PO_4 pH 6.0, 0.2% SDS.

4.6.4. The cleavage of MISTIC fusion tag, renaturation of the protein in DMPC/CHAPS bicelles, delipidation of the protein.

The MISTIC fusion protein was cleaved with Factor Xa protease from Qigen (Hilden, Germany). The analytical scale preparations were concentrated to 0.25 mg/ml and dialyzed against 20 mM Tris-HCl pH 8.0, 50 mM NaCl buffer supplied with 0.2% DDM. Afterwards, 1 mM of $CaCl_2$ was added to the reaction mixture. After dialysis 5 µg of protein was incubated 24 h at 20°C with 1.0, 0.2, 0.04, 0.008 and 0.0016 µl of protease with concentration 2 units/µl. During prolonged incubation with protease were used 0.4 units/µl protease concentration in the sample and took probes every 24 hours during 4 days.

When needed the proteins were renaturated in DMPC/CHAPS bicelles according to modified protocol of [119]. Mixed DMPC/CHAPS bicelles were prepared by stirring DMPC from Sigma-Aldrich (St. Louis, USA) in 50 mM sodium phosphate buffer pH 6.0 for 1 h followed by sonication. Then CHAPS was added and the resulting clear micellar solution was stored at room temperature and used within 24 h. Briefly, 0.5 mg/ml of the protein in 0.2% SDS (w/v) was mixed with an equal volume of mixed DMPC/CHAPS micelles, containing all-trans-retinal. Final concentrations were 0.5 mg/ml protein, 0.1% SDS, 1% DMPC and 0.75% CHAPS, 50 mM phosphate, and 25 µM retinal. The samples were protected from light. Absorption and fluorescence spectra of regenerated bR were measured after overnight incubation.

For delipidation of the protein we used Superose6HR and Ni-NTA column. For gel-filtration the sample was 9-fold concentrated, loaded on the column, and then eluted with 50 mM phosphate buffer with 0.2% DDM. Before loading to Ni-NTA column the sample was 5-fold diluted. The loading of the sample on the column was accomplished in a batch-mode overnight at room temperature. Then the column was intensively washed with 50 mM phosphate buffer pH 6.0, 0.2% DDM buffer. Then the protein was eluted with 50 mM phosphate buffer pH 6.0, 100 mM NaCl, 0.2% DDM, 300 mM imidazole buffer. After elution the samples were dialyzed against 50 mM phosphate buffer pH 6.0, 100 mM NaCl, 0.2% DDM buffer to remove imidazole.

4.6.5. Protein purification under non-denaturing conditions and characterization of the protein.

DDM was added to the final concentration of 1% to the membranes isolated as described in chapter 4.6.3 and then membranes were solubilized overnight with stirring at 4˚C. Insoluble material was removed by ultracentrifugation in Ti-70 rotor at 35 000 rpm for 1h. Supernatant was 5 times diluted with 20 mM Tris-HCl pH 8.0, 100 mM NaCl buffer and 10 mM imidazole was added. Suspension was loaded on the 5 ml of Ni-NTA resin equilibrated with the same buffer. The column was washed with 10 CV of 50 mM NaH_2PO_4 pH 6.0, 100 mM NaCl, 0.2% DDM, 30 mM imidazole. The protein was eluted with 3 CV of 50 mM NaH_2PO_4 pH 7.4, 100mM NaCl, 0.2% DDM, 300 mM imidazole. Only coloured fractions were pooled. To remove imidazole the samples were immediately dialysed against 0.6 L of 50 mM NaH_2PO_4 pH 6.0, 100 mM NaCl for 2 hours and removed from dialysis buffer, then after 10 hours dialysis was continued for additional 2 hours against fresh buffer.

After the dialysis the protein heavily precipitated. The white pellet was separated by centrifugation at 5000 rpm from colored solution and discarded. Protein was concentrated to the volume of 2 ml by ultrafiltration and applied to 165 ml Sephacryl S200HR (GE Healthcare, UK) column equilibrated with 50 mM NaH_2PO_4 pH 6.0, 100 mM NaCl, 0.1% DDM. Peak of colored functional protein could be easily separated from the peak of the aggregated protein.

The samples were protected from the light. Total protein concentration was measured by BCA Protein Assay Kit following supplier protocol. To access the protein purity the samples were analyzed on 8-16% gradient SDS-PAGE. The fraction of the functional protein was accessed from UV-Vis spectra as absorbance ratio $A_{280}/A_{\lambda max}$, where $A_{\lambda max}$ is maximum absorbance of retinal. The concentration of functional protein was accessed by retinal absorbance using extinction coefficient

63 000 $cm^{-1}M^{-1}$ [147].

4.7 Crystallization of wild type bR from purple membranes, data collection and model building.

Purple membranes were produced from *H. salinarum* S9 bR overexpressing strain according to [11]. The solubilization and crystallization were carried out according to the protocols routinely used in our laboratory and described in [147].

The concentration of the detergent in solubilized protein was determined by weighting. First, thin glass plate was weighted and then 10 μL of the solubilized protein was dropped on the plate. The liquid was evaporated during 2 h in desiccator, then the glass plate was weighted again. Subtracting from the final mass of the plate its initial mass as well as calculated mass of the protein and salt we can estimate the amount of detergent in the sample. Comparison with the Fourier transformed infrared spectroscopy showed that this method gives acceptable results.

The crystallization probes were inspected in light microscope. The crystals reached the size up to 350 μm. The selected probes transferred to the 3M sodium phosphate buffer pH 5.6 supplied with 0.1% OG where the cubic phase was dissolved. The crystals were fished out from mother liquor with cryoloop (Hampton Research, Aliso Viejo, USA) and flash cooled in liquid nitrogen. The frozen crystals were tested using in-house rotating anode X-ray generator (Nonius FR 591, Rotterdam, Netherlands), then X-ray diffraction data (wavelengths 0.934 Å and 0.976 Å) were collected at the beamlines ID14-1 and ID23-1 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using a PILATUS 6M detector.

The diffraction patterns were processed with MOSFILM [135] and SCALA [136]. For crystals with relatively low (less than 35%) twinning ratio the detwinning procedure was used, while for the cases of perfect twinning the build-in algorithms of CCP4 package [137] were used. The initial phasing was done using molecular replacement method realized in program MOLREP [138] with poly-Ala model of high resolution bR structure (derived from 1C3W structure [14] in PDB Data Bank [2]). Protein auto-building was accomplished by ARP/wARP [139]. The initial model was refined using Refmac [140]. The model was visualized using Coot [148]. The pictures were created using Pymol [149].

4.8 Crystallization of wild type bR and its mutants expressed in E. coli, data collection.

The crystals of wild type bR and its D85N and D96N mutants expressed in *E. coli* were grown using *in meso* approach using nanovolume robotic system Formulatrix NT8 (Waltham, USA). The purified protein in crystallization buffer was added to the MO-based lipid mesophase. The best crystals were obtained using the protein concentration of 39 mg/ml and 0.1M Tris pH 8.8, 2.6M ammonium sulfate precipitation solution. The crystals were grown at 22°C.

X-ray diffraction data (wavelengths 0.934 Å and 0.976 Å) were collected at the beamlines ID14-1 and ID23-1 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using a PILATUS 6M detector. Diffraction patterns were investigated using the MOSFLM software [135] from the CCP4 program suite [137].

5 Summary.

Energy production in living cells is among the most important questions in biology and for the modern technology. Bacteriorhodopsin (bR) being a simplest tool to produce electrochemical gradient of protons across membrane, a key and universal step of energy production in living cells, have received a lot of attention. It became one of the model membrane proteins (MPs), also because of its relative abundance in nature and relative ease of purification from natural source, *H. salinarum*. But homologous production of bR and its mutants in halobacteria is laborious, time-, and resource-consuming, thus restraining the research studies. Unique properties of bR photocycle make it useful and promising in a wide variety of technical and medical applications, thus giving rise to growing need of this protein in basic and applied science. Despite the availability of the atomic structures there are still controversies on the structures of bR photocycle intermediates obtained by different groups and the mechanism of the vectorial proton transfer by bR is not yet clear and requires new structural studies. For the investigation of the bR photocycle and obtaining of the atomic structures of photocycle intermediates availability of different bR mutants is essential. Archaea is not very suitable for fast production of the required mutants or large-scale protein production for industry, therefore the heterogeneous expression would be a desirable option as a source of wild type and mutant bR. An *Escherichia coli* based expression system is the most preferable because of its facility and robustness. The functional expression of bR in *E. coli* was intensively tried over the period of last 30 years but unsuccessfully. This is quite surprising considering the fact that there are several examples (SRII, hR and bR homologs from other archaea) of functional expression of retinal proteins in *E. coli*. The work presented addressed the problem of bR functional expression in *E. coli*.

At first, the hypothesis that the low yield of bR expression can be attributed to the low rate of protein insertion into bacterial membrane was verified. We introduced complementary protein approach that allowed us to localize the problem in MP expression using finite number of steps. It is based on constructing chimeric proteins between a protein of interest and complementary homologous protein expressed in chosen system with high yield. To investigate the role of the first transmembrane helix in insertion of bR into membrane the chimeric proteins, where different regions on the *N*-terminus of bR were replaced by the corresponding parts of SRII, have been expressed in *E. coli*. The substitution of first ten amino acids of bR for the corresponding eight amino acids from SRII was shown to increase the protein expression yield more than 50-fold making it comparable with the yields of reference retinal proteins known to express at high level in

E. coli. We suppose also that complementary protein approach may have a general application and can be used for MPs that are difficult to express.

Initially, we examined one of the possible reasons of such a dramatic effect – the presence of positively charged Arg7 on the *N*-terminus of bR that deviates from "positive inside" rule. This rule is based on idea that it is energetically unfavorable process to translocate the positive charge across lipid membrane. The exchange of *N*-terminus removes the positive charge on the *N*-terminus of the protein. According to the "positive inside" rule this could result in better incorporation of the protein to the membrane and thus increase the protein yield. To verify this idea we expressed bR point mutants R7Q and R7E where the only positive charge in the extracellular *N*-terminus was changed to neutral and negative charges, respectively. Although the yield of the point mutants were visibly higher than of wild type gene, it was considerably lower than that of chimeric protein indicating that presence of positive charge on the *N*-terminus of bR is not the main reason of its poor expression in *E. coli*.

Using RNA modeling software a putative stem structure in the *N*-terminus of bR mRNA was found revealing another possible reason of low bR expression in *E. coli*. The expression yield of optimized wild type bR gene was on the same level as yield of the chimera between bR and SRII, in contrast to the expression of native gene. Thus, the low yield of bacteriorhodopsin native gene in *E. coli* is mainly attributed to the unfavorable mRNA structure of native gene close to the ribosome binding site.

When we used non-denaturing conditions, the protein have retained its functionality during purification process. Using affinity and size-exclusion chromatography we were able to purify the functional wild type bR expressed in *E. coli* to homogeneity. The purity and functionality of the protein obtained were confirmed spectroscopically. The yield of functional protein was 2.4±1.3 mg of protein per liter of culture what is sufficient for a large-scale crystallization and industrial use.

One of the main advantages of *E. coli* expression system is the considerably shorter time required to produce the mutants of interest facilitating the intense study of the target protein. Using mRNA-optimized bR gene we introduced mutations V49A, D85N, and D96N of the key amino acids for proton pumping in bR and utilizing the non-denaturing purification protocol we have readily obtained functional mutant proteins in short time with yields of 0.3, 3.8, and 8.8 mg per liter of culture, respectively. We suppose that increased yield of D85N and D96N mutants can be explained by better incorporation of positively charged C-helix of bR into *E. coli* membrane. As

translocation of positive charge across membrane is energetically unfavorable process, the substitution of either Asp85 or Asp 96 with neutral Asn may favor insertion of bR into membrane.

One more task of the present work was to obtain the bR-Lys fusion protein that might help resolving the controversies in bR structural studies. Throughout the years over almost a hundred of bR structures were deposited in Protein Data Bank solved by both electron and X-ray crystallography. But still there are inconsistencies in details of proton translocation by bR. In 2007 the first structure of GPCR was obtained using GPCR-Lys fusion protein. We suppose that bR would be a good starting model for investigation of versatility of the utilization of Lys as crystallization tag. As most of the crystal contacts were formed by Lys and GPCRs have topology similar to bR, we expect that bR-Lys can readily form crystals. As *in meso* grown crystals of wild type bR are prone to twinning we hope that crystallization of bR-Lys fusion protein will provide the twinning-free crystals in another space group allowing to determine precise structures of bR ground state as well as its intermediates during photocycle.

Realization of such project requires a crystallization of the fusion protein between membrane bR and water soluble Lys, thus demanding high yield expression system and effective crystallization approaches. Here, bR can serve as a guiding reference. The previously reported X-ray structures of bR were solved using protein preparations from purple membranes. High resolution structures show that bR trimers are surrounded by the native lipid belt shielding the hydrophobic area of the protein. Despite multiple protocols of *E. coli* expression there have not been reports about successful 3D crystallization of the heterologously expressed bR. The 2D crystals of bR expressed in *E. coli* were obtained from protein reconstituted into native *H. salinarum* lipids. Since bR-Lys fusion protein expressed in *E. coli* would not have *H. salinarum* lipids bound and purification strategy of bR expressed in *E. coli* is based on the application of DDM instead of usual OG, the careful investigation of the influence of lipid/detergent environment on *in meso* crystallization would significantly benefit to the crystallization trials with *E. coli* expressed bR, its mutants and fusion proteins.

To study influence of detergent on the *in meso* crystallization we set large-scale crystallization trials with homologously expressed bR from purple membranes using the following mixtures of detergents: OG-DDM, OG-MEGA10, OG-CYMAL5. The crystals of different size (up to 300 μm in OG-DDM mixture) and quality were obtained in all three mixtures. Crystallization diagrams presenting the dependence of the size of the grown crystals on the crystallization conditions showed that not the type of the detergent influences the *in meso* crystallization, but the normalized amount

of detergent in the sample indicating that detergent alters the physical properties of the crystallization matrix.

The X-ray diffraction data sets at synchrotron radiation (ID 14-1, Grenoble, France) were taken from the best crystals according to the results of the in-house tests. Three crystals (one from OG-MEGA10 and two from OG-DDM mixtures) gave a diffraction 1.45 and 1.70 Å, respectively. The full datasets were collected and three structures of bR were solved. The structures obtained did not show significant differences in electron densities and no electron densities were modeled as detergent molecules. These structures (along with the structures solved by my colleagues using the crystals grown from different detergents) proved that detergent molecules do not participate in the formation of the crystal lattice of bR crystals. Moreover, there is no need to exchange the detergent from DDM, which is used for purification, to OG, which is generally used for *in meso* crystallization of bR from purple membranes.

Then, using the protein obtained by expression of RNA optimized gene in *E. coli* the crystals of wild type bR and D85N and D96N bR mutants were grown. Crystals were tested under synchrotron radiation and gave a diffraction up to 2.5Å resolution for D96N crystals and 3.7Å for D85N without optimization of crystal growth. These are the first 3D crystals of bR which expressed in *E. coli* that demonstrate the aptness of the presented approach for expression of bR and its mutants suitable for many scientific and industrial applications. In addition, the successful crystallization of protein isolated from *E. coli* demonstrated that *H. salinarum* native lipids are not required for the formation of well ordered bR crystals.

Using the mRNA optimized wild type bR gene we have expressed two bR-Lys fusion proteins in *E. coli* with yield up to 0.9 mg of functional protein per liter of culture. Using metal affinity and size-exclusion chromatography under non-denaturing conditions the functional protein was purified to homogeneity and *in meso* crystallization trials are ongoing.

To summarize in brief:

- We introduced and verified complementary protein approach that is based on constructing chimeric proteins between a target protein and complementary homologous protein expressed functionally in chosen system with high yield. We suppose that such approach can be used for MPs in general allowing to localize the underlying issues in the expression of target protein using a limited number of steps.

• Using complementary protein modular approach we found that replacement of first 10 amino acids of bR dramatically improves the yield of bR. The yield of chimeric protein is 50 times higher and was attributed mainly to the optimization of mRNA structure.

• The point mutations R8Q and R8E also lead to the increase in bR expression yield in 6.8 and 2.7 times, respectively. Thus, the "positive inside" rule considerably influences the expression level of bR in *E. coli*.

• Wild-type bR expressed in *E. coli* using RNA optimized gene retained its functional activity during purification under non-denaturing conditions. We purified the protein to homogeneity using metal affinity and size exclusion chromatography. Thus, our study established for the first time the protocol of wild type bR expression in *E. coli* and its functional purification in non-denaturing conditions to homogeneity in amounts sufficient for a large-scale crystallization.

• The potency of the presented approach was proved by expression in *E. coli* and functional purification under non-denaturing conditions of V49A, D85N, D96N mutants of bR.

• Bacteriorhodopsin from purple membranes was crystallized in OG-MEGA10, OG-DDM, and OG-CYMAL5 detergent mixtures giving the crystals diffracted up to 1.45 Å resolution. Three solved structures of bR in the ground state have not reveal the detergent in crystal lattice. Crystallization diagrams, however, indicate that detergent do influence the crystallization process altering the physical properties of amphiphilic crystallization matrix.

• The first 3D crystals of wild-type bR and its D85N and D96N mutants expressed in *E. coli* were grown from the obtained protein material. First crystals were tested on synchrotron beam line and gave a diffraction up to 2.5Å resolution for D96N and 3.7Å for D85N mutants even without optimization of crystallization parameters. We have demonstrated the aptness of the presented approach for expression of bR and its mutants suitable for many scientific and industrial applications. Successful crystallization of protein material expressed in *E. coli* demonstrated that *H. salinarum* native lipids are not required for the formation of well ordered bR crystals.

• Employing the described system for bR expression in *E. coli* we expressed and functionally purified under non-denaturing conditions the bR-Lys fusion proteins with a yield enough for crystallographic studies. The crystallization trials are ongoing.

The development of the *E. coli* based system for functional expression of bR and its mutants

opens new perspectives for the studies of this MP. Particularly, a readily available mutants of bR will facilitate structural studies helping to understand fundamentally the mechanism of proton transfer in the cells. Bacteriorhodopsin mutants may find their applications in multiple studies arising in bioelectronics, biophotonics, etc. The *E. coli* expression system provides many opportunities to investigate the influence of the lipid/detergent environment on the folding and crystallization of bR *in vivo* and *in vitro*. Also, we suggest that introduced here the complimentary protein approach may find a general application for MP expression. Very valuable seems to be the production of the functional bR-Lys fusion protein. The crystallization of this fusion protein can give an important information concerning the generality of the crystallization tag strategy and can allow to determine the atomic structures of bR and its intermediate states using twinning-free data. Thus, the results presented in this thesis are of great relevancy and, we hope, will facilitate the progress in the field.

Publication list

Bratanov D., Balandin T., Round E., Schevchenko V., Gordeliy V. An Approach to Functional Heterologous Expression of Membrane Proteins in *E. coli*. The Case of Bacteriorhodopsin. [in preparation]

Borshchevskaya Y., Borshchevskiy V., Round E., Bratanov D., Büldt G., Gordeliy V. Crystallographic study of detergent role in stabilization of membrane protein crystals grown *in meso*. [in preparation]

Acknowledgements

At first, I'm deeply thankful to my supervisor Prof. Dr. Valentin Gordeliy, who has been mentoring me over my undergraduate and PhD work, for his support and valuable discussions. Without his care and help my scientific career would not be possible. Also I would like to thank Taras Balandin for his huge contribution to the work of our laboratory in general and this work in particular. I appreciate the help of my friends and colleagues in Research Centre Jülich: Alex Volkov, Vitaliy Shevchenko, Andrii Ishchenko, and Maria Silacheva. I would like to express my gratitude to Christian Baeken for his technical assistance and help in a lot practical aspects of my life. I'm very grateful to Prof. Dr. Georg Büldt for the opportunity to work at ICS-5 in Research Centre Jülich.

A considerable part of this work was done in cooperation with the lab of membrane transporters in IBS, Grenoble. Most nanovolume crystallization trials were set up there and it would not be possible without the help from my colleagues: Ekaterina Round, Vitaliy Polovinkin, Ivan Gushchin, Pavel Chervakov, and Petr Utrobin. I would like to thank Prof. Dr. Eva Pebay-Peyroula for the support of this collaboration.

The data collection was performed at ESRF in Grenoble and I'm thankful to all my colleagues from Jülich and Grenoble as well as ESRF staff (and, in particular, to Dr. Alexander Popov) for their help in these experiments and useful recommendations.

In the end, I would like to thank my parents and grandparents for supporting me throughout all my studies. And I am really thankful to my wife, who makes my life wonderful.

Appendix

Nucleotide sequences of important genetic constructs (only open reading frames).

MbRL

atgttttgtacatttttgaaaaacatcaccggaagtgggacatactgttagaaaaaagcacgggtgtgatggaagctatgaaagtgacgagtgaggaaaaggaacagctgagc
acagcaatcgaccgaatgaatgaaggactggacgcgtttatccagctgtataatgaatcggaaattgatgaaccgcttattcagcttgatgatgatacagccgagttaatgaagc
aggcccgagatatgtacggccaggaaaagctaaatgagaaattaaatacaattattaaacagattttatccatctcagtatctgaagaaggagaaaaagaaggcccgggtggc
agcagcggtgcatcgattgaaggtcgtcaggcccagatcaccggacgtccggagtggatctggctagcgctcggtacggcgctaatgggactcgggacgctctatttcctcg
tgaaagggatgggcgtctcggacccagatgcaaagaaattctacgccatcacgacgctcgtcccagccatcgcgttcacgatgtacctctcgatgctgctggggtatggcctc
acaatggtaccgttcggtggggagcagaaccccatctactgggcgcggtacgctgactggctgttcaccacgccgctgttgttgttagacctcgcgttgctcgttgacgcggat
cagggaacgatccttgcgctcgtcggtgccgacggcatcatgatcgggaccggcctggtcggcgcactgacgaaggtctactcgtaccgcttcgtgtggtgggcgatcagc
accgcagcgatgctgtacatcctgtacgtgctgttcttcgggttcacctcgaaggccgaaagcatgatgcataacatcttcgagatgctccgcatcgacgaggggctccgcctg
aagatctacaaggacacggaggggtactacaccatcggcatcgggcacctcctgaccaagagcccgtccctcaacgccgcgaaaagcgagctggacaaggcgattggcc
gcaacaccaacggcgtcatcaccaaggacgaggccgagaagctcttcaaccaggacgtggatgcggcggtgcgcggcatcctccgcaacgccaagctgaaaccggtcta
tgactcgctcgacgccgtccgacgtgccgccctcatcaacatggtgttccagatgggcgaaacgggtgtcgcgggcttcacgaactcgctgcggatgctccagcagaaacg
ctgggacgaagcggccgtcaacctcgcgaaatcgcggtggtacaaccaaacgcccaaccgcgccaaacgggtcatcaccaccttccgaaccggcacgtgggacgcctac
aagaacctgaagcttcgccccgaggtcgcatccacgttcaaagtactgcgtaacgttaccgttgtgttgtggtccgcgtatcccgtcgtgtggctgatcggcagcgaaggtgcg
ggaatcgtgccgctgaacatcgagacgctgctgttcatggtgcttgacgtgagcgcgaaggtcggcttcgggctcatcctcctgcgcagtcgtgcgatcttcggcgaagccga
agcgccggagccgtccgccggcgacggcgcggccgcgaccagcgacggatccggcattgaaggtcgtagcggtgctccgcatcaccatcaccatcaccatcactaa

MbRLE

atgttttgtacatttttgaaaaacatcaccggaagtgggacatactgttagaaaaaagcacgggtgtgatggaagctatgaaagtgacgagtgaggaaaaggaacagctgagc
acagcaatcgaccgaatgaatgaaggactggacgcgtttatccagctgtataatgaatcggaaattgatgaaccgcttattcagcttgatgatgatacagccgagttaatgaagc
aggcccgagatatgtacggccaggaaaagctaaatgagaaattaaatacaattattaaacagattttatccatctcagtatctgaagaaggagaaaaagaaggcccgggtggc
agcagcggtgcatcgattgaaggtcgtcaggcccagatcaccggacgtccggagtggatctggctagcgctcggtacggcgctaatgggactcgggacgctctatttcctcg
tgaaagggatgggcgtctcggacccagatgcaaagaaattctacgccatcacgacgctcgtcccagccatcgcgttcacgatgtacctctcgatgctgctggggtatggcctc
acaatggtaccgttcggtggggagcagaaccccatctactgggcgcggtacgctgactggctgttcaccacgccgctgttgttgttagacctcgcgttgctcgttgacgcggat
cagggaacgatccttgcgctcgtcggtgccgacggcatcatgatcgggaccggcctggtcggcgcactgacgaaggtctactcgtaccgcttcgtgtggtgggcgatcagc
accgcagcgatgctgtacatcctgtacgtgctgttcttcgggttcacctcgaaggccgaaagcatgcgccccgaggtcgcatccacgttcaaagtactgcgtaacgttaccgtt
gtgttgtggtccgcgtatcccgtcgtgtggctgatcggcagcgaaggtgcgggaatcgtgccgctgaacatcgagacgctgctgttcatggtgcttgacgtgagcgcgaaggt
cggcttcgggctcatcctcctgcgcagtcgtgcgatcttcggcgaagccgaagcgccggagccgtccgccggcgacggcgcggccgcgaccacccggatgcataacatc
ttcgagatgctccgcatcgacgaggggctccgcctgaagatctacaaggacacggaggggtactacaccatcggcatcgggcacctcctgaccaagagcccgtccctcaac
gccgcgaaaagcgagctggacaaggcgattggccgcaacaccaacggcgtcatcaccaaggacgaggccgagaagctcttcaaccaggacgtggatgcggcggtgcg
cggcatcctccgcaacgccaagctgaaaccggtctatgactcgctcgacgccgtccgacgtgccgccctcatcaacatggtgttccagatgggcgaaacgggtgtcgcggg
cttcacgaactcgctgcggatgctccagcagaaacgctgggacgaagcggccgtcaacctcgcgaaatcgcggtggtacaaccaaacgcccaaccgcgccaaacgggtc
atcaccaccttccgaaccggcacgtgggacgcctacaagaacctgggatccggcattgaaggtcgtagcggtgctccgcatcaccatcaccatcaccatcactaa

MbRLEs

atgttttgtacatttttgaaaaacatcaccggaagtgggacatactgttagaaaaaagcacgggtgtgatggaagctatgaaagtgacgagtgaggaaaaggaacagctgagc
acagcaatcgaccgaatgaatgaaggactggacgcgtttatccagctgtataatgaatcggaaattgatgaaccgcttattcagcttgatgatgatacagccgagttaatgaagc
aggcccgagatatgtacggccaggaaaagctaaatgagaaattaaatacaattattaaacagattttatccatctcagtatctgaagaaggagaaaaagaaggcccgggtggc
agcagcggtgcatcgattgaaggtcgtcaggcccagatcaccggacgtccggagtggatctggctagcgctcggtacggcgctaatgggactcgggacgctctatttcctcg
tgaaagggatgggcgtctcggacccagatgcaaagaaattctacgccatcacgacgctcgtcccagccatcgcgttcacgatgtacctctcgatgctgctggggtatggcctc
acaatggtaccgttcggtggggagcagaaccccatctactgggcgcggtacgctgactggctgttcaccacgccgctgttgttgttagacctcgcgttgctcgttgacgcggat
cagggaacgatccttgcgctcgtcggtgccgacggcatcatgatcgggaccggcctggtcggcgcactgacgaaggtctactcgtaccgcttcgtgtggtgggcgatcagc
accgcagcgatgctgtacatcctgtacgtgctgttcttcgggttcacctcgaaggccgaaagcatgcgccccgaggtcgcatccacgttcaaagtactgcgtaacgttaccgtt
gtgttgtggtccgcgtatcccgtcgtgtggctgatcggcagcgaaggtgcgggaatcgtgccgctgaacatcgagacgctgctgttcatggtgcttgacgtgagcgcgaaggt
cggcttcgggctcatcctcctgcgcagtcgtgcgatcttcggcgaaatgcataacatcttcgagatgctccgcatcgacgaggggctccgcctgaagatctacaaggacacgg
aggggtactacaccatcggcatcgggcacctcctgaccaagagcccgtccctcaacgccgcgaaaagcgagctggacaaggcgattggccgcaacaccaacggcgtcat
caccaaggacgaggccgagaagctcttcaaccaggacgtggatgcggcggtgcgcggcatcctccgcaacgccaagctgaaaccggtctatgactcgctcgacgccgtc
cgacgtgccgccctcatcaacatggtgttccagatgggcgaaacgggtgtcgcgggcttcacgaactcgctgcggatgctccagcagaaacgctgggacgaagcggccgt
caacctcgcgaaatcgcggtggtacaaccaaacgcccaaccgcgccaaacgggtcatcaccaccttccgaaccggcacgtgggacgcctacaagaacctgggatccggc
attgaaggtcgtagcggtgctccgcatcaccatcaccatcaccatcactaa

bR

atgcaggcccagatcaccggacgtccggagtggatctggctagcgctcggtacggcgctaatgggactcgggacgctctatttcctcgtgaaagggatgggcgtctcggac

ccagatgcaaagaaattctacgccatcacgacgctcgtcccagccatcgcgttcacgatgtacctctcgatgctgctggggtatggcctcacaatggtaccgttcggtgggga
gcagaaccccatctactgggcgcggtacgctgactggctgttcaccacgccgctgttgttgttagacctcgcgttgctcgttgacgcggatcagggaacgatccttgcgctcgt
cggtgccgacggcatcatgatcgggaccggcctggtcggcgcactgacgaaggtctactcgtaccgcttcgtgtggtgggcgatcagcaccgcagcgatgctgtacatcct
gtacgtgctgttcttcgggttcacctcgaaggccgaaagcatgcgccccgaggtcgcatccacgttcaaagtactgcgtaacgttaccgttgtgttgtggtccgcgtatcccgtc
gtgtggctgatcggcagcgaaggtgcgggaatcgtgccgctgaacatcgagacgctgctgttcatggtgcttgacgtgagcgcgaaggtcggcttcgggctcatcctcctgc
gcagtcgtgcgatcttcggcgaagccgaagcgccggagccgtccgccggcgacggcgcggccgcgaccagcgacggatccggcattgaaggtcgtagcggtgctccgc
atcaccatcaccatcaccatcactaa

$SR_{1\text{-}10,\ 28\text{-}44}bR$

atggtgggacttacgaccctcttttggatctggctagcgctcggtacggcgctaatgggactcgggacgctctatttcggccgtgacgccggaagcggcgagcgacggtact
acgccatcacgacactagtcccagccatcgcgttcacgatgtacctctcgatgctgctggggtatggcctcacaatggtaccgttcggtggggagcagaaccccatctactgg
gcgcggtacgctgactggctgttcaccacgccgctgttgttgttagacctcgcgttgctcgttgacgcggatcagggaacgatccttgcgctcgtcggtgccgacggcatcatg
atcgggaccggcctggtcggcgcactgacgaaggtctactcgtaccgcttcgtgtggtgggcgatcagcaccgcagcgatgctgtacatcctgtacgtgctgttcttcgggttc
acctcgaaggccgaaagcatgcgccccgaggtcgcatccacgttcaaagtactgcgtaacgttaccgttgtgttgtggtccgcgtatcccgtcgtgtggctgatcggcagcga
aggtgcgggaatcgtgccgctgaacatcgagacgctgctgttcatggtgcttgacgtgagcgcgaaggtcggcttcgggctcatcctcctgcgcagtcgtgcgatcttcggcg
aagccgaagcgccggagccgtccgccggcgacggcgcggccgcgaccagcgacggatccggcattgaaggtcgtagcggtgctccgcatcaccatcaccatcaccatc
actaa

$bR_{1\text{-}8}SR$

atgcaggcccagatcaccggacgtccggagtggctcggcgcaatcggcatgctcgtcggcacgctcgcgttcgcgtgggccggccgtgacgccggaagcggcgagcga
cggtactacgtgacgcttgtcggcatcagtggtatcgcagcagtcgcctacgtcgtcatggcgctgggcgtcggctgggttcccgtggccgaacggactgtttttgccccccg
gtacattgactggattctcacaaccccgctcatcgtctacttcctcgggctgcttgcggggcttgatagtcgggagttcggcatcgtcatcacgctcaacaccgtggtcatgctcg
ccggcttcgccggggcgatggtgcccggtatcgagcgctacgcgctgttcggcatgggggcggtcgcattcctcggactggtctactacctcgtcgggccgatgaccgaaa
gtgccagccagcggtcctccggaatcaagtcgctgtacgtccgcctccgaaacctgacggtcatcctctgggcgatttatccgttcatctggctgcttggaccgccgggcgtg
gcgctgctgacaccgactgtcgacgtggcgcttatcgtctaccttgacctcgtcacgaaggtcggattcggcttcatcgcactcgatgctgcggcgacacttcgggccgaaca
cggcgaatcgctcgctggcgtcgatactgacgcgcctgcggtcgccgacgagaattcgcaccaccaccaccaccaccactaa

References

[1] **Overington JP, Al-Lazikani B, Hopkins AL**. (**2006**) How many drug targets are there? *Nat Rev Drug Discov.* 5: 993-6.
[2] **Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE**. (**2000**) The Protein Data Bank. *Nucleic Acids Res.* 28: 235-42.
[3] **Tusnády GE, Dosztányi Z, Simon I**. (**2004**) Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics* 20: 2964-72.
[4] **White SH,**. (**2009**) Biophysical dissection of membrane proteins. *Nature* 459: 344-6.
[5] **Faham S, Bowie JU**. (**2002**) Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J Mol Biol.* 316(1): 1-6.
[6] **Landau EM, Rosenbusch JP**. (**1996**) Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc Natl Acad Sci U S A.* 93(25): 14532-5.
[7] **Takeda K, Sato H, Hino T, Kono M, Fukuda K, Sakurai I, Okada T, Kouyama T**. (**1998**) A novel three-dimensional crystal of bacteriorhodopsin obtained by successive fusion of the vesicular assemblies. *J Mol Biol.* 283(2): 463-74.
[8] **Kayushin LP, Skulachev VP**. (**1974**) Bacteriorhodopsin as an electrogenic proton pump: reconstitution of bacteriorhodopsin proteoliposomes generating delta psi and delta pH. *FEBS Lett.* 39: 39-42.
[9] **Oesterhelt D, Stoeckenius W**. (**1973**) Functions of a new photoreceptor membrane. *Proc Natl Acad Sci U S A.* 70: 2853-7.
[10] **Huang KS, Liao MJ, Gupta CM, Royal N, Biemann K, Khorana HG**. (**1982**) The site of attachment of retinal in bacteriorhodopsin. The epsilon-amino group in Lys-41 is not required for proton translocation. *J Biol Chem.* 257: 8596-9.
[11] **Lorber B, DeLucas LJ**. (**1990**) Large scale preparation of homogeneous bacteriorhodopsin. *FEBS Lett.* 261: 14-8.
[12] **Brouillette CG, McMichens RB, Stern LJ, Khorana HG**. (**1989**) Structure and thermal stability of monomeric bacteriorhodopsin in mixed phospholipid/detergent micelles. *Proteins* 5: 38-46.
[13] **Michel H, Oesterhelt D**. (**1980**) Three-dimensional crystals of membrane proteins: bacteriorhodopsin. *Proc Natl Acad Sci U S A.* 77: 1283-5.
[14] **Luecke H, Schobert B, Richter HT, Cartailler JP, Lanyi JK**. (**1999**) Structure of bacteriorhodopsin at 1.55 A resolution. *J Mol Biol* 291(4): 899-911.
[15] **Lanyi JK,**. (**2004**) Bacteriorhodopsin. *Annu Rev Physiol.* 66: 665-88.
[16] **Lanyi JK,**. (**2004**) What is the real crystallographic structure of the L photointermediate of bacteriorhodopsin? *Biochim Biophys Acta.* 1658: 14-22.
[17] **Birge RR, Gillespie NB, Izaguirre EW, Kusnetzow A, Lawrence AF, Singh D, Song QW, Schmidt E, Stuart JA, Seetharaman S, Wise KJ**. (**1999**) Biomolecular Electronics: Protein-Based Associative Processors and Volumetric Memories. *J Phys Chem B.* 103(49): 10746-66.
[18] **Hampp N,**. (**2000**) Bacteriorhodopsin as a Photochromic Retinal Protein for Optical Memories. *Chem Rev.* 100: 1755-76.
[19] **Francis DM, Page R**. (**2010**) Strategies to optimize protein expression in E. coli. *Curr Protoc Protein Sci.* Chapter 5: Unit 5.24.1-29.
[20] **Karnik SS, Nassal M, Doi T, Jay E, Sgaramella V, Khorana HG**. (**1987**) Structure-function studies on bacteriorhodopsin. II. Improved expression of the bacterio-opsin gene in Escherichia coli. *J Biol Chem.* 262: 9255-63.
[21] **Shand RF, Miercke LJ, Mitra AK, Fong SK, Stroud RM, Betlach MC**. (**1991**) Wild-type and mutant bacterioopsins D85N, D96N, and R82Q: high-level expression in Escherichia coli. *Biochemistry* 30: 3082-8.
[22] **Chen GQ, Gouaux JE**. (**1996**) Overexpression of bacterio-opsin in Escherichia coli as a

water-soluble fusion to maltose binding protein: efficient regeneration of the fusion protein and selective cleavage with trypsin. *Protein Sci.* 5: 456-67.
[23] **Nekrasova OV, Wulfson AN, Tikhonov RV, Yakimov SA, Simonova TN, Tagvey AI, Dolgikh DA, Ostrovsky MA, Kirpichnikov MP**. (**2010**) A new hybrid protein for production of recombinant bacteriorhodopsin in Escherichia coli. *J Biotechnol.* 147: 145-50.
[24] **Pompejus M, Friedrich K, Teufel M, Fritz HJ**. (**1993**) High-yield production of bacteriorhodopsin via expression of a synthetic gene in Escherichia coli. *Eur J Biochem.* 211: 27-35.
[25] **Mironova OS, Efremov RG, Person B, Heberle J, Budyak IL, Büldt G, Schlesinger R**. (**2005**) Functional characterization of sensory rhodopsin II from Halobacterium salinarum expressed in Escherichia coli. *FEBS Lett.* 579: 3147-51.
[26] **Schmies G, Chizhov I, Engelhard M**. (**2000**) Functional expression of His-tagged sensory rhodopsin I in Escherichia coli. *FEBS Lett.* 466: 67-9.
[27] **Hohenfeld IP, Wegener AA, Engelhard M**. (**1999**) Purification of histidine tagged bacteriorhodopsin, pharaonis halorhodopsin and pharaonis sensory rhodopsin II functionally expressed in Escherichia coli. *FEBS Lett.* 442: 198-202.
[28] **Petrovskaya LE, Lukashev EP, Chupin VV, Sychev SV, Lyukmanova EN, Kryukova EA, Ziganshin RH, Spirina EV, Rivkina EM, Khatypov RA, Erokhina LG, Gilichinsky DA, Shuvalov VA, Kirpichnikov MP**. (**2010**) Predicted bacteriorhodopsin from Exiguobacterium sibiricum is a functional proton pump. *FEBS Lett.* 584: 4193-6.
[29] **Kamo N, Hashiba T, Kikukawa T, Araiso T, Ihara K, Nara T**. (**2006**) A light-driven proton pump from Haloterrigena turkmenica: functional expression in Escherichia coli membrane and coupling with a H+ co-transporter. *Biochem Biophys Res Commun.* 341: 285-90.
[30] **Fu HY, Lin YC, Chang YN, Tseng H, Huang CC, Liu KC, Huang CS, Su CW, Weng RR, Lee YY, Ng WV, Yang CS**. (**2010**) A novel six-rhodopsin system in a single archaeon. *J Bacteriol.* 192: 5866-73.
[31] **Shimono K, Iwamoto M, Sumi M, Kamo N**. (**1997**) Functional expression of pharaonis phoborhodopsin in Escherichia coli. *FEBS Lett.* 420: 54-6.
[32] **Gordeliy VI, Labahn J, Moukhametzianov R, Efremov R, Granzin J, Schlesinger R, Büldt G, Savopol T, Scheidig AJ, Klare JP, Engelhard M**. (**2002**) Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* 419(6906): 484-7.
[33] **Royant A, Nollert P, Edman K, Neutze R, Landau EM, Pebay-Peyroula E, Navarro J**. (**2001**) X-ray structure of sensory rhodopsin II at 2.1-A resolution. *Proc Natl Acad Sci U S A.* 98: 10131-6.
[34] **Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC**. (**2007**) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318(5854): 1258-65.
[35] **Chien EY, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, Stevens RC**. (**2010**) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330: 1091-5.
[36] **Chun E, Thompson AA, Liu W, Roth CB, Griffith MT, Katritch V, Kunken J, Xu F, Cherezov V, Hanson MA, Stevens RC**. (**2012**) Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* 20: 967-76.
[37] **Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP, Stevens RC**. (**2008**) The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322: 1211-7.
[38] **Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC**. (**2010**) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330: 1066-71.

[39] **Mitra AK, Miercke LJ, Turner GJ, Shand RF, Betlach MC, Stroud RM**. (**1993**) Two-dimensional crystallization of Escherichia coli-expressed bacteriorhodopsin and its D96N variant: high resolution structural studies in projection. *Biophys J.* 65: 1295-306.
[40] **Opekarová M, Tanner W**. (**2003**) Specific lipid requirements of membrane proteins - a putative bottleneck in heterologous expression. *Biochim Biophys Acta.* 1610: 11-22.
[41] **Cherezov V, Caffrey M**. (**2006**) Picolitre-scale crystallization of membrane proteins. *J Appl Cryst.* 39: 604-6.
[42] **Mukai Y, Kamo N, Mitaku S**. (**1999**) Light-induced denaturation of bacteriorhodopsin solubilized by octyl-beta-glucoside. *Protein Eng.* 12: 755-9.
[43] **Oesterhelt D, Stoeckenius W**. (**1973**) Functions of a new photoreceptor membrane. *Proc Natl Acad Sci U S A* 70: 2853-7.
[44] **Saitô H, Yamaguchi S, Ogawa K, Tuzi S, Márquez M, Sanz C, Padrós E**. (**2004**) Glutamic acid residues of bacteriorhodopsin at the extracellular surface as determinants for conformation and dynamics as revealed by site-directed solid-state 13C NMR. *Biophys J.* 86: 1673-81.
[45] **Dunn R, McCoy J, Simsek M, Majumdar A, Chang SH, Rajbhandary UL, Khorana HG**. (**1981**) The bacteriorhodopsin gene. *Proc Natl Acad Sci U S A.* 78: 6744-8.
[46] **Khorana HG, Gerber GE, Herlihy WC, Gray CP, Anderegg RJ, Nihei K, Biemann K**. (**1979**) Amino acid sequence of bacteriorhodopsin. *Proc Natl Acad Sci U S A.* 76: 5046-50.
[47] **Henderson R,**. (**1975**) The structure of the purple membrane from Halobacterium hallobium: analysis of the X-ray diffraction pattern. *J Mol Biol.* 93: 123-38.
[48] **Oesterhelt D, Stoeckenius W**. (**1971**) Rhodopsin-like protein from the purple membrane of Halobacterium halobium. *Nat New Biol.* 233: 149-52.
[49] **Brouillette CG, Muccio DD, Finney TK**. (**1987**) pH dependence of bacteriorhodopsin thermal unfolding. *Biochemistry* 26: 7431-8.
[50] **Trivedi S, Choudhary OP, Gharu J**. (**2011**) Different proposed applications of bacteriorhodopsin. *Recent Pat DNA Gene Seq.* 5: 35-40.
[51] **Takamatsu S, Hoshino K, Matsumoto K, Miyasaka T, Shimoyama I**. (**2011**) The photo charge of a bacterioRhodopsin electrochemical cells measured by a charge amplifier. *IEICE Electronics Express* 8: 505-11.
[52] **Frydrych M, Silfsten P, Parkkinen S, Parkkinen J, Jaaskelainen T**. (**2000**) Color sensitive retina based on bacteriorhodopsin. *Biosystems* 54: 131-40.
[53] **Tukiainen T, Lensu L, Parkkinen J**. Temporal Characteristics of Artificial Retina Based on Bacteriorhodopsin and Its Variants. *Springer Berlin Heidelberg*, 2007.
[54] **Patil AV, Premaraban T, Berthoumieu O, Watts A, Davis JJ**. (**2012**) Engineered bacteriorhodopsin: a molecular scale potential switch. *Chemistry* 18(18): 5632-6.
[55] **Thavasi V, Lazarova T, Filipek S, Kolinski M, Querol E, Kumar A, Ramakrishna S, Padrós E, Renugopalakrishnan V**. (**2009**) Study on the feasibility of bacteriorhodopsin as bio-photosensitizer in excitonic solar cell: a first report. *J Nanosci Nanotechnol.* 9: 1679-87.
[56] **Mitchell P,**. (**1961**) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* 191: 144-8.
[57] **Neutze R, Pebay-Peyroula E, Edman K, Royant A, Navarro J, Landau EM**. (**2002**) Bacteriorhodopsin: a high-resolution structural view of vectorial proton transport. *Biochim Biophys Acta.* 1565: 144-67.
[58] **Edman K, Royant A, Larsson G, Jacobson F, Taylor T, van der Spoel D, Landau EM, Pebay-Peyroula E, Neutze R**. (**2004**) Deformation of helix C in the low temperature L-intermediate of bacteriorhodopsin. *J Biol Chem.* 279: 2147-58.
[59] **Royant A, Edman K, Ursby T, Pebay-Peyroula E, Landau EM, Neutze R**. (**2000**) Helix deformation is coupled to vectorial proton transport in the photocycle of bacteriorhodopsin. *Nature* 406: 645-8.
[60] **Lanyi JK, Schobert B**. (**2003**) Mechanism of proton transport in bacteriorhodopsin from

crystallographic structures of the K, L, M1, M2, and M2' intermediates of the photocycle. *J Mol Biol.* 328: 439-50.
[61] **Lanyi JK, Schobert B**. (**2007**) Structural changes in the L photointermediate of bacteriorhodopsin. *J Mol Biol.* 365: 1379-92.
[62] **Kouyama T, Nishikawa T, Tokuhisa T, Okumura H**. (**2004**) Crystal structure of the L intermediate of bacteriorhodopsin: evidence for vertical translocation of a water molecule during the proton pumping cycle. *J Mol Biol.* 335: 531-46.
[63] **Deisenhofer J, Epp O, Miki K, Huber R, Michel H**. (**1984**) X-ray structure analysis of a membrane protein complex. Electron density map at 3 A resolution and a model of the chromophores of the photosynthetic reaction center from Rhodopseudomonas viridis. *J Mol Biol.* 180(2): 385-98.
[64] **McPherson A,**. Crystallization of biological macromolecules. *Cold Spring Harbor Laboratory Press, New York*, 1999.
[65] **Ostermeier C, Michel H**. (**1997**) Crystallization of membrane proteins. *Curr Opin Struct Biol.* 7(5): 697-701.
[66] **Roth M, Lewit-Bentley A, Michel H, Deisenhofer J, Huber R, Oesterhelt D**. (**1989**) Detergent structure in crystals of a bacterial photosynthetic reaction centre *Nature* 340: 659-62.
[67] **Cherezov V, Clogston J, Papiz MZ, Caffrey M**. (**2006**) Room to move: crystallizing membrane proteins in swollen lipidic mesophases. *J Mol Biol.* 357(5): 1605-18.
[68] **Efremov R, Shiryaeva G, Bueldt G, Islamov A, Kuklin A, Yaguzhinsky L, Fragneto-Cusani G, Gordeliy V**. (**2005**) SANS investigations of the lipidic cubic phase behavior in course of bacteriorhodopsin crystallization *Journal of Crystal Growth* 275: e1453-e1459.
[69] **Grabe M, Neu J, Oster G, Nollert P**. (**2003**) Protein interactions and membrane geometry. *Biophys J.* 84(2 Pt 1): 854-68.
[70] **Caffrey M, Cherezov V**. (**2009**) Crystallizing membrane proteins using lipidic mesophases. *Nat Protoc.* 4: 706-31.
[71] **Nollert P, Qiu H, Caffrey M, Rosenbusch JP, Landau EM**. (**2001**) Molecular mechanism for the crystallization of bacteriorhodopsin in lipidic cubic phases. *FEBS Lett.* 504(3): 179-86.
[72] **Cherezov V, Caffrey M**. (**2007**) Membrane protein crystallization in lipidic mesophases. A mechanism study using X-ray microdiffraction. *Faraday Discuss.* 136: 195-212; discussion 213-29.
[73] **Cherezov V, Fersi H, Caffrey M**. (**2001**) Crystallization screens: compatibility with the lipidic cubic phase for in meso crystallization of membrane proteins. *Biophys J.* 81(1): 225-42.
[74] **Cherezov V, Peddi A, Muthusubramaniam L, Zheng YF, Caffrey M**. (**2004**) A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr D Biol Crystallogr.* 60(Pt 10): 1795-807.
[75] **Liu W, Hanson MA, Stevens RC, Cherezov V**. (**2010**) LCP-Tm: an assay to measure and understand stability of membrane proteins in a membrane environment. *Biophys J.* 98: 1539-48.
[76] **Kubicek J, Schlesinger R, Baeken C, Büldt G, Schäfer F, Labahn J**. (**2012**) Controlled in meso phase crystallization - a method for the structural investigation of membrane proteins. *PLoS One* 7: e35458.
[77] **Ujwal R, Cascio D, Colletier JP, Faham S, Zhang J, Toro L, Ping P, Abramson J**. (**2008**) The crystal structure of mouse VDAC1 at 2.3 A resolution reveals mechanistic insights into metabolite gating. *Proc Natl Acad Sci U S A.* 105(46): 17742-7.
[78] **Luecke H, Schobert B, Stagno J, Imasheva ES, Wang JM, Balashov SP, Lanyi JK**. (**2008**) Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proc Natl Acad Sci U S A.* 105: 16561-5.
[79] **Vinothkumar KR,**. (**2011**) Structure of rhomboid protease in a lipid environment. *J Mol Biol.* 407: 232-47.
[80] **Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AG, Tate CG, Schertler GF**. (**2008**) Structure of a beta1-adrenergic G-protein-coupled

receptor. *Nature* 454: 486-91.
[81] **Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, Schertler GF, Weis WI, Kobilka BK**. (**2007**) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 450: 383-7.
[82] **Bjarnadóttir TK, Gloriam DE, Hellstrand SH, Kristiansson H, Fredriksson R, Schiöth HB**. (**2006**) Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics* 88: 263-73.
[83] **Lehnert U, Xia Y, Royce TE, Goh CS, Liu Y, Senes A, Yu H, Zhang ZL, Engelman DM, Gerstein M**. (**2004**) Computational analysis of membrane proteins: genomic occurrence, structure prediction and helix interactions. *Q Rev Biophys.* 37: 121-46.
[84] **Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC, Iwata S**. (**2011**) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475: 65-70.
[85] **Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, Liu W, Thompson AA, Huang XP, Carroll FI, Mascarella SW, Westkaemper RB, Mosier PD, Roth BL, Cherezov V, Stevens RC**. (**2012**) Structure of the human κ-opioid receptor in complex with JDTic. *Nature* 485: 327-32.
[86] **Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, Kobilka BK**. (**2012**) Structure of the δ-opioid receptor bound to naltrindole. *Nature* 485: 400-4.
[87] **Hollenstein K, Kean J, Bortolato A, Cheng RK, Doré AS, Jazayeri A, Cooke RM, Weir M, Marshall FH**. (**2013**) Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature* 499: 438-43.
[88] **Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah ST, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK**. (**2011**) Crystal structure of the β2 adrenergic receptor-Gs protein complex. *Nature* 477: 549-55.
[89] **White JF, Noinaj N, Shibata Y, Love J, Kloss B, Xu F, Gvozdenovic-Jeremic J, Shah P, Shiloach J, Tate CG, Grisshammer R**. (**2012**) Structure of the agonist-bound neurotensin receptor. *Nature* 490: 508-13.
[90] **Iwata S, Ostermeier C, Ludwig B, Michel H**. (**1995**) Structure at 2.8 A resolution of cytochrome c oxidase from Paracoccus denitrificans. *Nature* 376: 660-9.
[91] **Lim HH, Fang Y, Williams C**. (**2011**) High-efficiency screening of monoclonal antibodies for membrane protein crystallography. *PLoS One* 6: e24653.
[92] **Röthlisberger D, Pos KM, Plückthun A**. (**2004**) An antibody library for stabilizing and crystallizing membrane proteins - selecting binders to the citrate carrier CitS. *FEBS Lett.* 564: 340-8.
[93] **Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B**. (**2003**) Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* 12: 1313-22.
[94] **Corsini L, Hothorn M, Scheffzek K, Sattler M, Stier G**. (**2008**) Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein Sci.* 17: 2070-9.
[95] **Niemann HH, Schmoldt HU, Wentzel A, Kolmar H, Heinz DW**. (**2006**) Barnase fusion as a tool to determine the crystal structure of the small disulfide-rich protein McoEeTI. *J Mol Biol* 356: 1-8.
[96] **Privé GG, Verner GE, Weitzman C, Zen KH, Eisenberg D, Kaback HR**. (**1994**) Fusion proteins as tools for crystallization: the lactose permease from Escherichia coli. *Acta Crystallogr D Biol Crystallogr.* 50: 375-9.
[97] **Engel CK, Chen L, Privé GG**. (**2002**) Insertion of carrier proteins into hydrophilic loops of the Escherichia coli lactose permease. *Biochim Biophys Acta.* 1564: 38-46.
[98] **Turner GJ, Miercke LJ, Mitra AK, Stroud RM, Betlach MC, Winter-Vann A**. (**1999**)

Expression, purification, and structural characterization of the bacteriorhodopsin-aspartyl transcarbamylase fusion protein. *Protein Expr Purif.* 17: 324-38.
[99] **Ferrando E, Schweiger U, Oesterhelt D**. (**1993**) Homologous bacterio-opsin-encoding gene expression via site-specific vector integration. *Gene* 125: 41-7.
[100] **Turner GJ, Reusch R, Winter-Vann AM, Martinez L, Betlach MC**. (**1999**) Heterologous gene expression in a membrane-protein-specific system. *Protein Expr Purif.* 17: 312-23.
[101] **Dunn RJ, Hackett NR, McCoy JM, Chao BH, Kimura K, Khorana HG**. (**1987**) Structure-function studies on bacteriorhodopsin. I. Expression of the bacterio-opsin gene in Escherichia coli. *J Biol Chem.* 262: 9246-54.
[102] **Nassal M, Mogi T, Karnik SS, Khorana HG**. (**1987**) Structure-function studies on bacteriorhodopsin. III. Total synthesis of a gene for bacterio-opsin and its expression in Escherichia coli. *J Biol Chem.* 262: 9264-70.
[103] **Braiman MS, Stern LJ, Chao BH, Khorana HG**. (**1987**) Structure-function studies on bacteriorhodopsin. IV. Purification and renaturation of bacterio-opsin polypeptide expressed in Escherichia coli. *J Biol Chem.* 262: 9271-6.
[104] **Karnik S, Doi T, Molday R, Khorana HG**. (**1990**) Expression of the archaebacterial bacterio-opsin gene with and without signal sequences in Escherichia coli: the expressed proteins are located in the membrane but bind retinal poorly. *Proc Natl Acad Sci U S A.* 87: 8955-9.
[105] **Miercke LJ, Betlach MC, Mitra AK, Shand RF, Fong SK, Stroud RM**. (**1991**) Wild-type and mutant bacteriorhodopsins D85N, D96N, and R82Q: purification to homogeneity, pH dependence of pumping, and electron diffraction. *Biochemistry* 30: 3088-98.
[106] **Sudo Y, Ihara K, Kobayashi S, Suzuki D, Irieda H, Kikukawa T, Kandori H, Homma M**. (**2011**) A microbial rhodopsin with a unique retinal composition shows both sensory rhodopsin II and bacteriorhodopsin-like properties. *J Biol Chem.* 286: 5967-76.
[107] **Inoue K, Ono H, Abe-Yoshizumi R, Yoshizawa S, Ito H, Kogure K, Kandori H**. (**2013**) A light-driven sodium ion pump in marine bacteria. *Nat Commun.* 4: 1678.
[108] **Misquitta Y, Caffrey M**. (**2003**) Detergents destabilize the cubic phase of monoolein: implications for membrane protein crystallization. *Biophys J.* 85(5): 3084-96.
[109] **Belrhali H, Nollert P, Royant A, Menzel C, Rosenbusch JP, Landau EM, Pebay-Peyroula E**. (**1999**) Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 A resolution. *Structure* 7(8): 909-17.
[110] **Jensen MØ, Mouritsen OG**. (**2004**) Lipids do influence protein function-the hydrophobic matching hypothesis revisited. *Biochim Biophys Acta.* 1666(1-2): 205-26.
[111] **Essen L, Siegert R, Lehmann WD, Oesterhelt D**. (**1998**) Lipid patches in membrane protein oligomers: crystal structure of the bacteriorhodopsin-lipid complex. *Proc Natl Acad Sci U S A.* 95: 11673-8.
[112] **Krebs MP, Isenbarger TA**. (**2000**) Structural determinants of purple membrane assembly. *Biochim Biophys Acta.* 1460: 15-26.
[113] **Ni BF, Chang M, Duschl A, Lanyi J, Needleman R**. (**1990**) An efficient system for the synthesis of bacteriorhodopsin in Halobacterium halobium. *Gene* 90: 169-72.
[114] **Bartus CL, Jaakola VP, Reusch R, Valentine HH, Heikinheimo P, Levay A, Potter LT, Heimo H, Goldman A, Turner GJ**. (**2003**) Downstream coding region determinants of bacterio-opsin, muscarinic acetylcholine receptor and adrenergic receptor expression in Halobacterium salinarum. *Biochim Biophys Acta.* 1610: 109-23.
[115] **Dvir H, Choe S**. (**2009**) Bacterial expression of a eukaryotic membrane protein in fusion to various Mistic orthologs. *Protein Expr Purif.* 68: 28-33.
[116] **Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW**. (**1990**) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* 185: 60-89.
[117] **Szpirer CY, Milinkovitch MC**. (**2005**) Separate-component-stabilization system for protein and DNA production without the use of antibiotics. *Biotechniques* 38: 775-81.

[118] **Shoichet BK, Baase WA, Kuroki R, Matthews BW**. (**1995**) A relationship between protein stability and protein function. *Proc Natl Acad Sci U S A.* 92: 452-6.
[119] **Booth PJ, Riley ML, Flitsch SL, Templer RH, Farooq A, Curran AR, Chadborn N, Wright P**. (**1997**) Evidence that bilayer bending rigidity affects membrane protein folding. *Biochemistry* 36: 197-203.
[120] **von Heijne G,**. (**1989**) Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* 341: 456-8.
[121] **Bogdanov M, Heacock P, Guan Z, Dowhan W**. (**2010**) Plasticity of lipid-protein interactions in the function and topogenesis of the membrane protein lactose permease from Escherichia coli. *Proc Natl Acad Sci U S A.* 107: 15057-62.
[122] **van Klompenburg W, Nilsson I, von Heijne G, de Kruijff B**. (**1997**) Anionic phospholipids are determinants of membrane protein topology. *EMBO J.* 16: 4261-6.
[123] **Raine A, Ullers R, Pavlov M, Luirink J, Wikberg JE, Ehrenberg M**. (**2003**) Targeting and insertion of heterologous membrane proteins in E. coli. *Biochimie* 85: 659-68.
[124] **Shimono K, Hayashi T, Ikeura Y, Sudo Y, Iwamoto M, Kamo N**. (**2003**) Importance of the broad regional interaction for spectral tuning in Natronobacterium pharaonis phoborhodopsin (sensory rhodopsin II). *J Biol Chem.* 278: 23882-9.
[125] **Booth PJ, Farooq A, Flitsch SL**. (**1996**) Retinal binding during folding and assembly of the membrane protein bacteriorhodopsin. *Biochemistry* 35: 5902-9.
[126] **Sprengart ML, Porter AG**. (**1997**) Functional importance of RNA interactions in selection of translation initiation codons. *Mol Microbiol.* 24: 19-28.
[127] **Kudla G, Murray AW, Tollervey D, Plotkin JB**. (**2009**) Coding-sequence determinants of gene expression in Escherichia coli. *Science* 324: 255-8.
[128] **Steffen P, Voss B, Rehmsmeier M, Reeder J, Giegerich R**. (**2006**) RNAshapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics* 22: 500-3.
[129] **Sasaki T, Demura M, Kato N, Mukai Y**. (**2011**) Sensitive detection of protein-lipid interaction change on bacteriorhodopsin using dodecyl β-D-maltoside. *Biochemistry* 50: 2283-90.
[130] **Samuelson J,**. Production of Membrane Proteins: Strategies for Expression and Isolation. *Wiley-VCH Verlag GmbH & Co.*, 2011.
[131] **Curran AR, Templer RH, Booth PJ**. (**1999**) Modulation of folding and assembly of the membrane protein bacteriorhodopsin by intermolecular forces within the lipid bilayer. *Biochemistry* 38: 9328-36.
[132] **Booth PJ, Clarke J**. (**2010**) Membrane protein folding makes the transition. *Proc Natl Acad Sci U S A.* 107: 3947-8.
[133] **Hsu MF, Yu TF, Chou CC, Fu HY, Yang CS, Wang AH**. (**2013**) Using Haloarcula marismortui Bacteriorhodopsin as a Fusion Tag for Enhancing and Visible Expression of Integral Membrane Proteins in Escherichia coli. *PLoS One* 8: e56363.
[134] **Hunt JF, Rath P, Rothschild KJ, Engelman DM**. (**1997**) Spontaneous, pH-dependent membrane insertion of a transbilayer alpha-helix. *Biochemistry* 36: 15177-92.
[135] **Leslie AGW, Powell HR**. (**2007**) Processing diffraction data with mosflm. *NATO Science Series* 245: 41-51.
[136] **Evans PR,**. (**2006**) Scaling and assessment of data quality. *Acta Cryst D.* 62: 72-82.
[137] **Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS**. (**2011**) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr.* 67: 235-42.
[138] **Vagin A, Teplyakov A**. (**1997**) MOLREP: an automated program for molecular replacement. *J Appl Cryst.* 30: 1022-5.
[139] **Langer G, Cohen SX, Lamzin VS, Perrakis A**. (**2008**) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc.* 3: 1171-9.

[140] **Murshudov G, Vagin A, Dodson E**. (**1996**) Application of Maximum Likelihood Refinement. *Proceedings of Daresbury Study Weekend* 4.
[141] **Inoue H, Nojima H, Okayama H**. (**1990**) High efficiency transformation of Escherichia coli with plasmids. *Gene* 96: 23-8.
[142] **Chung CT, Niemela SL, Miller RH**. (**1989**) One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A.* 86: 2172-5.
[143] **Matsumura M, Becktel WJ, Levitt M, Matthews BW**. (**1989**) Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proc Natl Acad Sci U S A.* 86: 6562-6.
[144] **Dyall-Smith M,**. The halohandbook *Protocols for halobacterial genetics*, 2009.
[145] **Studier WF,**. (**2005**) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif.* 41: 207-34.
[146] **Miercke LJ, Ross PE, Stroud RM, Dratz EA**. (**1989**) Purification of bacteriorhodopsin and characterization of mature and partially processed forms. *J Biol Chem.* 264: 7531-5.
[147] **Gordeliy VI, Schlesinger R, Efremov R, Büldt G, Heberle J**. Crystallization in lipidic cubic phases: a case study with bacteriorhodopsin. *Springer*, 2003.
[148] **Emsley P, Cowtan K**. (**2004**) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 60: 2126-32.
[149] **The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC**.

Abstract.

Energy production in a living cells is among the most important questions in biology and for the modern technology. Bacteriorhodopsin (bR) being a simplest tool to produce electric potential across membrane have received a lot of attention. It became one of the model membrane proteins, also because of its relative abundance in nature and relative ease of purification from natural source, *H. salinarum*. Unique properties of bR photocycle make it useful and promising in a wide variety of technical applications, thus giving rise to growing need of this protein. Despite the availability of the atomic structures there are still controversies in mechanism of proton pumping by bR. Homologous production of bR and its mutants in halobacteria is laborious, time-, and resource-consuming, therefore facile and robust *E. coli* expression system would be of wide interest. Recently several structures of GPCRs were obtained using GPCR-lysozyme fusion protein where lysozyme served as a crystallization tag. We suppose that bR would be a good model for investigation of versatility of lysozyme as crystallization tags. As *in meso* grown crystals of bR are prone to twinning crystallization of bR-lysozyme fusion protein could provide twinning-free crystals allowing to clarify the details of bR photocycle.

In this work, it was suggested that the low yield of bR expression in *E. coli* can be attributed to the poor insertion of the protein into membrane. We have introduced protein complementary approach that may allow to localize the problem in membrane protein expression using finite number of steps. It is based on constructing of chimeric proteins between a protein of interest and complementary homologous protein expressed with high yield. Applying this approach we showed that the substitution of first ten amino acids of bR for the corresponding eight amino acids from SRII increase the expression yield of bR more than 50-fold. The reason for high yield of the chimera could be the positively charged Arg7 on the *N*-terminus of bR that deviates from “positive inside” rule and absent in the chimera. We expressed bR mutants R7Q and R7E where this positive charge was substituted for neutral and negative charges, respectively. Although the yields of the mutants were higher than of wild type gene, they were still considerably lower than yield of chimera. Thus, the positive charge on the *N*-terminus of bR is not the reason of its poor expression in *E. coli*. A putative stem structure 5'-end of bR mRNA was proposed to be another reason of low bR expression in *E. coli*. The expression yield of optimized wild type bR gene was on the same level as yield of the chimera. Therefore, the low yield of bacteriorhodopsin native gene in *E. coli* was attributed to the unfavorable mRNA structure of native gene close to the ribosome binding site. When purified under non-denaturing conditions, the protein have retained its functionality. The yield of functional homogenious protein was 2.4±1.3 mg per liter of culture what is sufficient for a

large-scale crystallization and industrial use. Using this approach we produced as well functional V49A, D85N, and D96N mutants of bR in short time with yields of 0.3, 3.8, and 8.8 mg per liter of culture, respectively. We suppose that increased yield of D85N and D96N mutants can be explained by better incorporation of positively charged C-helix of bR into *E. coli* membrane.

The second goal, crystallization of the bR-lysozyme fusion protein, demands a high yield expression system and effective crystallization approaches. Here, bR can serve as a reference. High resolution structures show that bR trimers are surrounded by the belt of native lipid. Despite multiple protocols of *E. coli* expression, 3D crystallization of this protein was not reported. Since expressed in *E. coli* bR-lysozyme fusion protein would not have *H. salinarum* lipids bound and purification of *this protein* is based on the application of DDM instead of usual OG, careful investigation of the influence of lipid/detergent environment on *in meso* crystallization is important.

To study the influence of detergent on the *in meso* crystallization we set large-scale crystallization trials with homologously expressed bR in mixtures of detergents. The crystals of different size (up to 300 μm) were obtained. Three crystals grown in mixtures of detergents radiation gave at synchrotron a diffraction up to 1.45 Å. The full datasets were collected and three structures of bR were solved. We have not observed detergent molecules on the electron densities corresponding to the structures. These experiments showed that detergent molecules do not participate in the formation of the crystal lattice of bR. Moreover, there is no need to exchange the detergent from DDM used for purification to OG that is generally used for *in meso* crystallization of homologously expressed bR. Then, using the protein expressed *E. coli* the crystals of wild type bR and D85N and D96N bR mutants were grown. Crystals were tested under synchrotron radiation and gave a diffraction up to 2.5Å resolution. The first 3D crystals of bR expressed in *E. coli* demonstrate that expression of bR and its mutants in *E. coli* is suitable for scientific and industrial applications. In addition, the successful crystallization of protein isolated from *E. coli* demonstrated that *H. salinarum* lipids are not strictly required for grow of well ordered bR crystals.

Using the optimized bR gene we have expressed bR-lysozyme fusion proteins in *E. coli* with yield up to 0.9 mg of functional protein per liter of culture. This protein was purified under non-denaturing conditions to homogeneity and *in meso* crystallization trials are ongoing.

Zusammenfassung

Die Energieproduktion in lebenden Zellen ist eine der wichtigsten Fragen der modernen Biotechnologie. Bacteriorhodopsin (bR) als eines der einfachsten Mittel zur Erzeugung eines elektrischen Potenzials an einer Membran hat in diesem Zusammenhang große Beachtung gefunden. Auch aufgrund seines häufigen Vorkommens in der Natur in *H. salinarum* und der relativ unproblematischen Aufreinigung gilt es heute als Modellmembranprotein. Dank seines einzigartigen Photozyklus ist bR ein nützliches, vielversprechendes Protein für eine große Bandbreite technischer Anwendungen, was zu einem steigenden Bedarf geführt hat. Obwohl die Struktur des Proteins aufgeklärt ist, sind Details des Protonenpumpmechanismus von bR noch immer umstritten. Die homologe Produktion von bR und seiner Varianten in Halobakterien ist arbeitsaufwändig, langwierig und ressourcenintensiv. Ein einfaches und robustes *E.-coli*-Expressionssystem würde daher auf großes Interesse stoßen. In der letzten Zeit wurden verschiedene GPCR-Strukturen mit Hilfe von GPCR-Lysozym-Fusionsproteinen mit Lysozym als Kristallisations-Tag hergestellt. Es ist anzunehmen, dass bR ein geeignetes Modell für die Untersuchung der Verwendungsmöglichkeiten von Lysozym als Kristallisations-Tag sein kann. *In meso* gewachsene bR-Kristalle neigen zur Zwillingsbildung. Durch Kristallisation des bR-Lysozym-Fusionsproteins können zwillingsfreie Kristalle gewonnen werden, um die Details des Photozyklus von bR näher zu untersuchen.

In dieser Arbeit wurde die Vermutung aufgestellt, dass die geringe Ausbeute bei der bR-Expression in *E. coli* auf den unzureichenden Einbau des Proteins in die Membran zurückzuführen ist. Es wurde ein Ansatz mit komplementären Proteinen eingeführt, um das Problem bei der Membranproteinexpression mit einer endlichen Anzahl von Schritten zu lokalisieren. Dieser Ansatz basiert auf der Herstellung von Fusionsproteinen, die aus dem untersuchten Protein und einem geeigneten komplementären und homologen Protein bestehen, das mit hoher Ausbeute exprimiert wird. Mithilfe dieses Verfahrens konnte mittels Substitution der ersten zehn Aminosäuren von bR durch die ersten acht Aminosäuren von SRII die Expressionsausbeute von bR um das Fünfzigfache erhöht werden. Grund für die hohe Ausbeute an Fusionsprotein ist möglicherweise, dass das positiv geladene Arg7 am N-Terminus von bR (eine Abweichung von der ‚Innen-positiv-Regel') in diesem Protein nicht vorliegt. Es wurden die bR-Varianten R7Q und R7E exprimiert, bei denen die positive Ladung jeweils durch eine neutrale bzw. negative Ladung ersetzt wurde. Obwohl die Ausbeute beider Varianten die des Wildtyps überstieg, lag sie doch deutlich niedriger als bei dem Fusionsprotein. Die positive Ladung am N-Terminus von bR ist daher nicht der Grund für seine schwache Expression in *E. coli*. Eine vermeintliche Stielstruktur am 5'-Ende der bR-mRNA wurde als weiterer möglicher Grund für die niedrige bR-Expression in *E. coli* vermutet. Die Expressionsausbeute lag bei dem optimierten bR-Wildtypgen auf dem gleichen Niveau wie bei dem Fusionsprotein. Die geringe Ausbeute bei dem natürlich vorkommenden Bacteriorhodopsin-Gen in *E. coli* wurde auf seine ungünstige mRNA-Struktur im Bereich der Ribosom-Bindungsstelle zurückgeführt. Bei der Aufreinigung unter nicht denaturierenden Bedingungen bleibt die Funktionalität des Proteins erhalten. Die Ausbeute an funktional homogenem Protein lag bei 2,4 ± 1,3 mg pro Liter Kultur, was für eine Kristallisation im Großmaßstab und industrielle Nutzung

ausreichend ist. Mit diesem Ansatz konnten innerhalb kurzer Zeit funktionelle V49A-, D85N- und D96N-Varianten von bR mit einer Ausbeute von 0,3 mg, 3,8 mg bzw. 8,8 mg pro Liter Kultur hergestellt werden. Es ist anzunehmen, dass die erhöhte Ausbeute der D85N- und D96N-Varianten durch einen erfolgreicheren Einbau der positiv geladenen C-Helix von bR in die *E.-coli*-Membran zurückzuführen ist.

Das zweite Ziel, die Kristallisation des bR-Lysozym-Fusionsproteins, erfordert ein hochproduktives Expressionssystem und effektive Kristallisationsansätze. Hier kann bR als Referenz dienen. Hochauflösende Strukturanalysen zeigen, dass bR-Trimere von einem Gürtel aus nativen Lipiden umgeben sind. Obwohl eine Reihe von Expressionsprotokollen für *E. coli* vorliegen, wurde bisher nicht von einer 3-D-Kristallisation dieses Proteins berichtet. Da das bR-Lysozym-Fusionsprotein bei einer Expression in *E. coli* keine *H.-salinarum*-Lipide binden würde und bei der Aufreinigung dieses Proteins DDM statt des üblichen OGs verwendet würde, muss hier sorgfältig untersucht werden, wie sich eine eine solche Lipid/Detergens-Umgebung auf die *In-meso*-Kristallisation auswirkt.

Um den Einfluss des Detergens auf die *In-meso*-Kristallisation zu untersuchen, wurden großangelegte Kristallisationsversuche mit homolog exprimiertem bR in Detergenziengemischen durchgeführt. Ergebnis waren Kristalle verschiedener Größe (bis zu 300 µm). Bei drei in Gemischen von Detergenzien gewachsenen Kristallen wurde bei der Beugungsanalyse im Synchrotron eine Auflösung von bis zu 1.45 Å gemessen. Die vollständigen Datensätze wurden gesammelt und die drei bR-Strukturen aufgeklärt. Die Elektronendichten, die diesen Strukturen entsprechen, zeigen keine Detergenzienmoleküle. Diese Versuche belegen, dass die Detergenzienmoleküle bei der Entstehung des bR-Kristallgitters keine Rolle spielen. Darüber hinaus besteht keine Notwendigkeit, das für die Aufreinigung verwendete Detergens DDM durch OG zu ersetzen, das im Allgemeinen für die *In-meso*-Kristallisation von homolog exprimiertem bR verwendet wird. Im Anschluss wurden Kristalle des bR-Wildtyps und der Varianten D85N und D96N unter Verwendung des in *E. coli* exprimierten Proteins gezüchtet. Die Untersuchung der Beugung mit Synchrotronstrahlung zeigte eine Auflösung von bis zu 2.5Å. Die ersten 3-D-Kristalle des in *E. coli* exprimierten bR belegen, dass die Expression von bR und seinen Varianten in *E. coli* für wissenschaftliche und industrielle Anwendungen geeignet ist. Darüber hinaus zeigt die erfolgreiche Kristallisation von aus *E. coli* isoliertem Protein, dass *H.-salinarum*-Lipide für das Wachstum regelmäßiger Kristalle nicht unbedingt erforderlich sind.

Mit dem optimierten bR-Gen konnten die bR-Lysozym-Fusionsproteine in *E. coli* mit einer Ausbeute von bis zu 0,9 mg funktionellem Protein pro Liter Kultur exprimiert werden. Dieses Protein wurde unter nicht denaturierenden Bedingungen bis zur Homogenität aufgereinigt. Die *In-meso*-Kristallisationsversuche laufen derzeit noch.