

RESEARCH ARTICLE

# Crystal Structure of *Escherichia coli*-Expressed *Haloarcula marismortui* Bacteriorhodopsin I in the Trimeric Form

Vitaly Shevchenko<sup>1,2</sup>, Ivan Gushchin<sup>2,3,4,5</sup>, Vitaly Polovinkin<sup>2,3,4,5</sup>, Ekaterina Round<sup>1</sup>, Valentin Borshchevskiy<sup>1,2</sup>, Petr Utrobin<sup>2,3,4</sup>, Alexander Popov<sup>6</sup>, Taras Balandin<sup>1</sup>, Georg Büldt<sup>2</sup>, Valentin Gordeliy<sup>1,2,3,4,5\*</sup>

1. Institute of Complex Systems (ICS-6) Structural Biochemistry, Research Centre Jülich GmbH, Jülich, Germany, 2. Laboratory for advanced studies of membrane proteins, Moscow institute of physics and technology, Dolgoprudnyi, Russia, 3. Univ. Grenoble Alpes, IBS, Grenoble, France, 4. CNRS, IBS, Grenoble, France, 5. CEA, IBS, Grenoble, France, 6. European Synchrotron Radiation Facility, Grenoble, France

\*[valentin.gordeliy@ibs.fr](mailto:valentin.gordeliy@ibs.fr)

These authors contributed equally to this work.



 OPEN ACCESS

**Citation:** Shevchenko V, Gushchin I, Polovinkin V, Round E, Borshchevskiy V, et al. (2014) Crystal Structure of *Escherichia coli*-Expressed *Haloarcula marismortui* Bacteriorhodopsin I in the Trimeric Form. PLoS ONE 9(12): e112873. doi:10.1371/journal.pone.0112873

**Editor:** Oleg Y. Dmitriev, University of Saskatchewan, Canada

**Received:** June 25, 2014

**Accepted:** October 17, 2014

**Published:** December 5, 2014

**Copyright:** © 2014 Shevchenko et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. Coordinates and structure factors have been deposited in the Protein Data Bank under the accession code 4PXK.

**Funding:** The work was supported by the program "Chaires d'excellence" edition 2008 of ANR France and CEA(IFS) - HGF(FZJ) STC 5.1 specific agreement. Part of this work was supported by BMBF (PhoNa - Photonic Nanomaterials). The work was supported by RFBR (research projects 13-04-91320 and 12-04-31290) and the Russian state Program for enhancing the competitiveness of MIPT among the world's leading research and education centers of the Ministry of education and science. The work used the platforms of the Grenoble Instruct centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). VP is deeply thankful to Fondation Nanosciences for financial support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

Bacteriorhodopsins are a large family of seven-helical transmembrane proteins that function as light-driven proton pumps. Here, we present the crystal structure of a new member of the family, *Haloarcula marismortui* bacteriorhodopsin I (*HmBRI*) D94N mutant, at the resolution of 2.5 Å. While the *HmBRI* retinal-binding pocket and proton donor site are similar to those of other archaeal proton pumps, its proton release region is extended and contains additional water molecules. The protein's fold is reinforced by three novel inter-helical hydrogen bonds, two of which result from double substitutions relative to *Halobacterium salinarum* bacteriorhodopsin and other similar proteins. Despite the expression in *Escherichia coli* and consequent absence of native lipids, the protein assembles as a trimer in crystals. The unique extended loop between the helices D and E of *HmBRI* makes contacts with the adjacent protomer and appears to stabilize the interface. Many lipidic hydrophobic tail groups are discernible in the membrane region, and their positions are similar to those of archaeal isoprenoid lipids in the crystals of other proton pumps, isolated from native or native-like sources. All these features might explain the *HmBRI* properties and establish the protein as a novel model for the microbial rhodopsin proton pumping studies.

## Introduction

Microbial rhodopsins are a large family of seven-helical transmembrane proteins that contain the covalently attached cofactor retinal [1]. Upon absorption of a photon, the retinal isomerizes and starts a series of structural transformations, correlated with spectral changes and called photocycle [1, 2]. Among the members of the family are light-driven proton, anion or cation pumps, light-gated ion channels and photoreceptors [1].

The most studied class of microbial rhodopsins are the proton pumps. It has been established that there are three distinct regions involved in the proton translocation: the proton donor site, the retinal binding pocket, and finally, the proton release region. In *Halobacterium salinarum* bacteriorhodopsin (*HsBR*), identified in 1971 by Oesterhelt and Stoekenius [3], the proton translocation cycle starts with isomerization of the retinal and transfer of the proton from the retinal Schiff base to the side-chain of the proton acceptor D85 [2]. In the next stage, the proton is transferred to the proton release group, consisting of two closely situated glutamates E194 and E204 and several water molecules. Then, the Schiff base is reprotonated from the proton donor D96, which itself is reprotonated from the cytoplasm. Although the general details of this proton pumping mechanism are well known, there are some discrepancies between the published structures of the intermediate states [4, 5], some of which might be explained by high radiation susceptibility of bacteriorhodopsin [6] or twinning of crystals [7]. Consequently, different details of the bacteriorhodopsin proton pumping still continue to be investigated, related to both the proton release [8] and the proton uptake [9].

In the recent years, structures of many microbial rhodopsins have been determined. Structures of four different archaeal proton pumps are known, *HsBR* [10, 11], archaerhodopsin-1 and -2 (*ar-1* and *ar-2*) [12], and deltarhodopsin-3 (*dr-3*) [13], as well as the structures of bacterial pumps xanthorhodopsin [14], various proteorhodopsins [15, 16] and eukaryotic *Acetabularia* rhodopsin [17] and *Chlamydomonas reinhardtii* channelrhodopsin [18]. Here, we present the structure of bacteriorhodopsin I from *Haloarcula marismortui* (henceforth *HmBRI*). *HmBRI* is one of the six *Haloarcula marismortui* retinylidene proteins and one of its two bacteriorhodopsins, whose sequences are 50% identical [19]. The protein can be expressed in *Escherichia coli* in large quantities and in fact can be used as a tag for production of other membrane proteins [20]. The structure reveals the conserved proton donor and retinal-binding pocket sites, and expanded proton release region. There are three additional inter-helical hydrogen bonds in the *HmBRI* transmembrane region. Despite heterologous expression, the protein assembles as a trimer in crystals. The loop between the helices D and E of *HmBRI* is extended, makes contacts with the adjacent protomer and appears to stabilize the interface. These features might result in a higher stability of *HmBRI* and explain its high expression level in *E. coli* cells. *E. coli* production allows for easy genetic manipulation and rapid production of mutants, which will be helpful

for studies of the proton transport mechanism as well as for protein engineering for optogenetics needs [1, 21].

## Results and Discussion

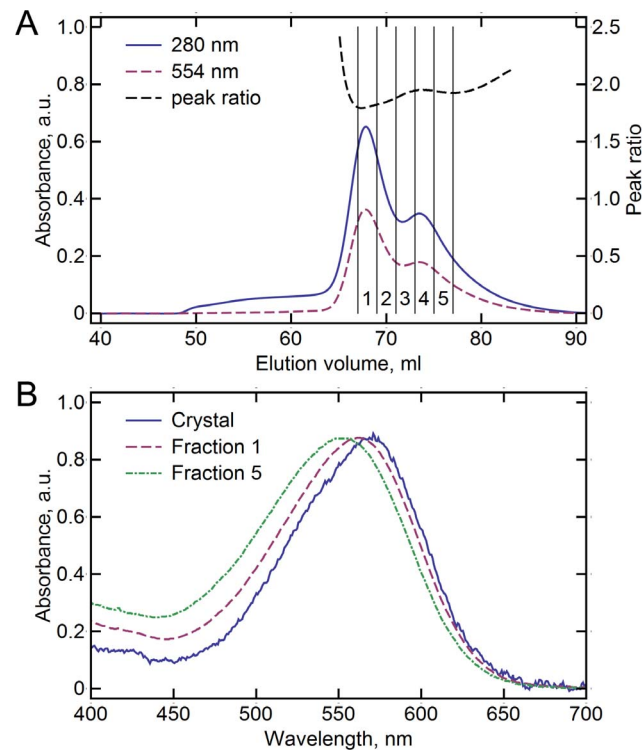
### HmBRI expression and spectroscopic characterization

*HmBRI* was heterologously expressed in *E. coli* and purified by nickel-affinity and size-exclusion (SEC) chromatography. The SEC elution profile revealed that the protein exists in two forms in the detergent solution (Figure 1A), presumably monomeric and oligomeric (most probably trimeric). Both forms behave identically on the sodium dodecyl sulfate polyacrylamide electrophoresis. Although the approximate molecular weights of the two species can be calculated as 130 kDa (the lighter form) and 210 kDa (the heavier form) on the basis of their elution volumes, this information is not enough to determine the oligomeric states, as the relation between the SEC mobility of a protein and its molecular weight is not accurate. Also, the size of the detergent micelle around the protein and amount of the lipids of the expression host carried with the protein are not known.

To determine the potential differences between the two forms, the protein was characterized spectroscopically. The lighter form has the absorption maxima at 554 nm in the light-adapted state, and the heavier form at 562 nm (Table 1). The absorption maxima of both forms are blue-shifted by ~4 nm in the dark-adapted state. Thus, the protein's absorption maximum depends on its oligomeric state. The small variation of the absorbance maximum position with dark adaptation as compared to other bacteriorhodopsin-like proton pumps might be explained by the nature of the retinal  $\beta$ -ionone ring-proximal side-chain, corresponding to M145 in *HsBR*. It has been shown, that the M145F *HsBR* mutation, reducing the steric conflict, results in much lower occupancy of the 13-*cis* retinal conformer in the dark-adapted state and a smaller absorption maximum shift [22]. In *HmBRI*, the residue is L149, and its position is closer to that of F145 in the ar-2 structure, than to that of M145 in *HsBR* or ar-1 structures [12, 23].

*HmBRI*, reconstituted into *H. salinarum* polar lipids, revealed the spectroscopic characteristics very similar to those of the heavier detergent form. Such red shift of the absorption maximum of the protein in the membrane relatively to the absorption maximum in the detergent solution was also observed in the case of *HsBR* [24, 25]. Finally, *HmBRI* in crystals, where it is expected to be in the light-adapted state, absorbed at even higher wavelengths (Figure 1B, Table 1).

Since *HmBRI*, similarly to others light-driven archaeal proton pumps, is expected to form trimers in the native membrane (and forms trimers in crystal), and the heavier detergent form absorbs light similarly to the protein reconstituted into the native-like membranes, we suggest that the two observed forms of the protein in detergent are monomers and trimers.



**Figure 1. *HmBRI* size-exclusion chromatography (SEC) elution profile and visual light absorbance spectra in different environments.** (A) SEC elution profile of *HmBRI*. Two distinct peaks are observed, that correspond supposedly to monomeric *HmBRI* (at 73.4 ml, peak ratio is ~1.95) and oligomeric *HmBRI* (at 67.8 ml, peak ratio is ~1.85). Five 2 ml fractions (1–5) were collected for subsequent crystallization. (B) Spectra of the light-adapted *HmBRI* in the hexagonal crystals, the first and the fifth SEC fractions.

doi:10.1371/journal.pone.0112873.g001

### Crystallization and crystal packing of *HmBRI*

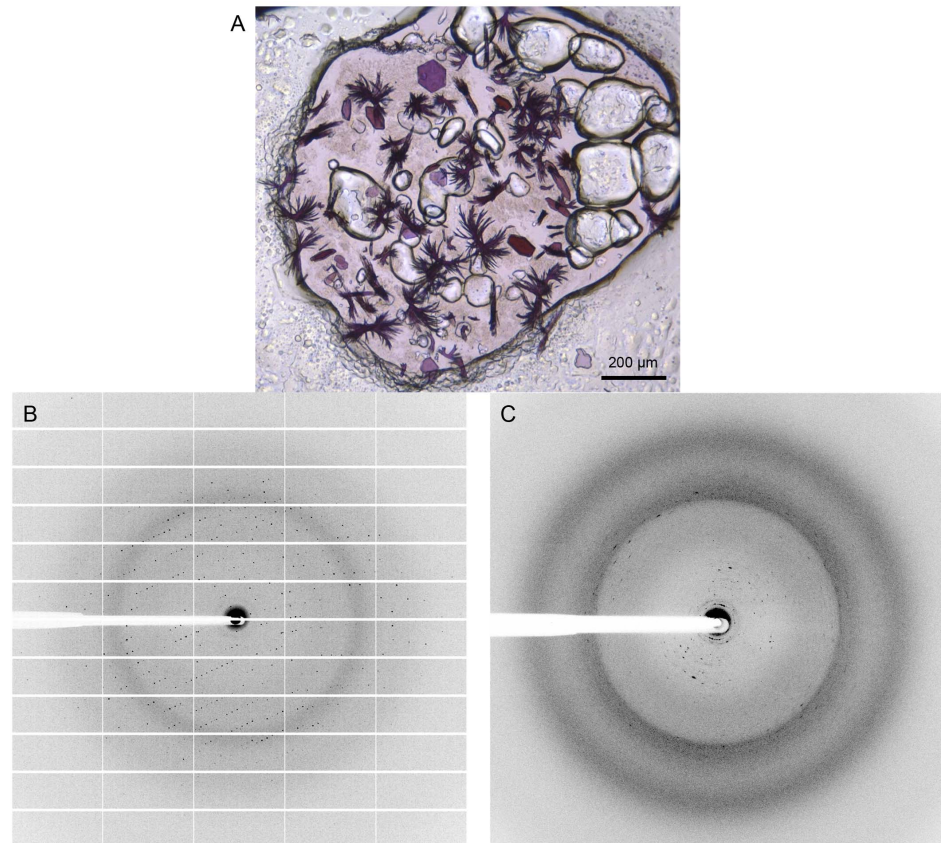
The *HmBRI* protein was crystallized using the *in meso* approach [26, 27] similarly to our previous work [15, 28]. Two types of crystals appeared in the same crystallization wells: regularly shaped hexagonal crystals and bunches of needles (Figure 2A), both approaching 100 μm in the largest dimension. Absorption maximum of *HmBRI* in the hexagonal crystals is at ~567 nm (Figure 1B). While the hexagonally shaped crystals produced good diffraction patterns up to the

**Table 1. *HmBRI* absorption maxima in different environments (nm).**

Environment	Dark-adapted state	Light-adapted state
Crystal	n/d	~566
<i>H. salinarum</i> polar lipids	559	563
DDM Fraction 1* (SEC)	559	562
DDM Fraction 3* (SEC)	552	555
DDM Fraction 5* (SEC)	549	554

\* See Figure 1A for definition of the SEC fractions.

doi:10.1371/journal.pone.0112873.t001



**Figure 2. *HmbRI* crystals and diffraction.** (A) *HmbRI* crystals in the crystallization well. Two types of crystals are observed: regular shaped hexagonal crystals and bunches of needles. (B) Example of the diffraction from the hexagonal crystals. Resolution at the detector edge is 2.4 Å. (C) Example of the diffraction from needle-shaped crystals. Resolution at the detector edge is 2.5 Å. The diffraction patterns are smeared and observed up to the resolution of 3.5 Å.

doi:10.1371/journal.pone.0112873.g002

resolution of 2.3 Å (Figure 2B, Table 2), the needles produced overlapping diffraction patterns up to the resolution of 3.5 Å, which could not be indexed (Figure 2C).

The spacegroup of the hexagonally shaped crystals was determined to be P321, cell parameters  $a, b=102.9$  Å,  $c=60.0$  Å. The protein is in the same arrangement as archaerhodopsin-2 in the trimeric form (PDB ID 2EI4 [29], space group P321, cell parameters  $a, b=98.2$  Å,  $c=56.2$  Å). Typically for *in meso* crystallization, *HmbRI* crystals are organized as stacks of the membrane-like layers (Figure 3A). The unit cell spans one such layer, and contains two *HmbRI* trimers, oriented in opposite directions (Figures 3A and 3B).

The inter-layer crystal contacts rely on the ionic interactions between the sulfate ions and the side-chains of the helix F lysine 176 and arginine 172 of the opposing *HmbRI* molecules (Figure 3C). The second partially occupied sulfate ion binding site is observed close to the first one. However, the second sulfate does not participate in inter-molecular contacts and is bound to the residues K176 and

**Table 2.** Crystallographic data collection and refinement statistics.

Data collection	
Space group	P321
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	102.85, 102.85, 60.00
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 120
Resolution (Å)	60–2.5 (2.63–2.5)*
<i>R</i> <sub>merge</sub> (%)	15.4 (94.9)*
<i>I</i> / $\sigma$ <i>I</i>	9.1 (3.0)*
<i>CC</i> <sub>1/2</sub>	0.99 (0.64)*
Completeness (%)	98.7 (97.5)*
Multiplicity	5.4 (5.4)*
Refinement	
Resolution	60–2.5 Å
No. reflections	12817 (623**)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	19.16%/23.72%
Number of atoms	
Protein and retinal	1827
Water	19
Lipid fragments	188
Sulfate ions	10
Average B-factor (Å <sup>2</sup> )	
Protein and retinal	38.8
Water	40.8
Lipid fragments	53.5
Sulfate ions	63.9
R.m.s. deviations	
Bond lengths	0.010 Å
Bond angles	1.7°

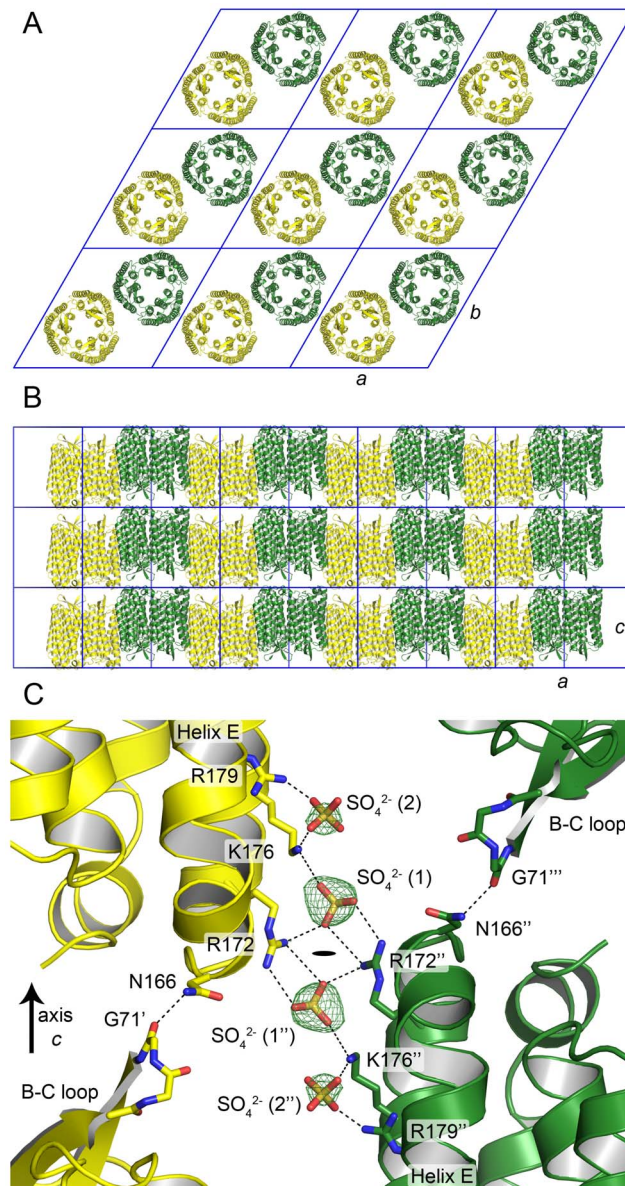
\* Values in parentheses are for the highest-resolution shell.

\*\* Number of reflections that are not used for refinement (free reflections).

doi:10.1371/journal.pone.0112873.t002

R179 from the same *HmBRI* molecule. A sulfate ion at the similar position is observed in the structure of deltarhodopsin-3 [13], where it is bound by R160 of the helix E and R174 of the helix F. The other inter-layer contact in *HmBRI* crystals is formed by the hydrogen bond between the G71 oxygen of one *HmBRI* molecule and N166 amine nitrogen of another (Figure 3C).

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

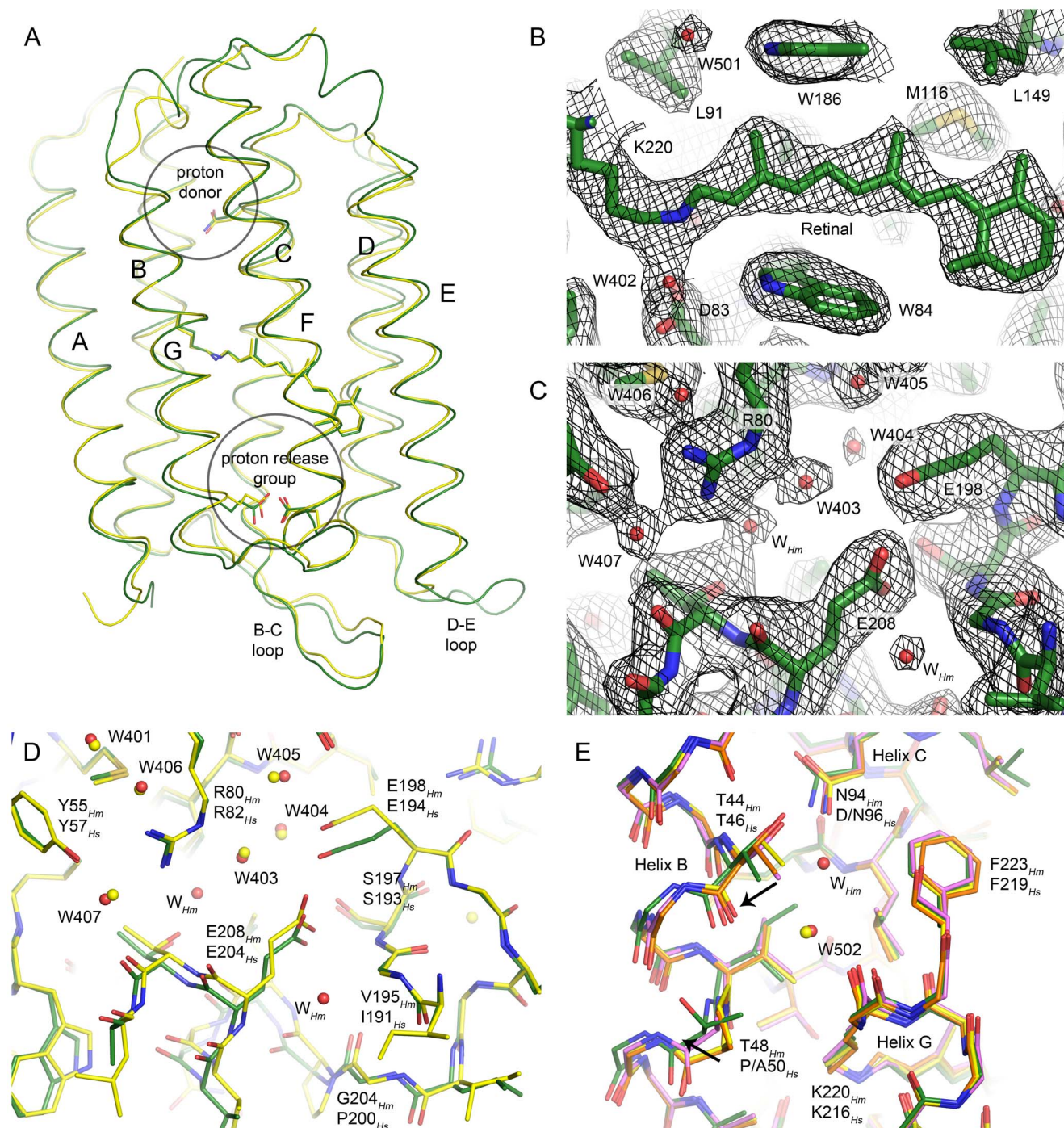


**Figure 3. HmBRI packing in hexagonal crystals.** (A) Packing in the membrane plane. There are two trimers in the unit cell (shown in yellow and green) that are oriented in opposite directions. (B) Packing of the membrane-like layers. (C) Crystal contacts between the layers. The contacts rely on the ionic interactions between the  $\text{SO}_4^{2-}$  ions and the R172 and K176 side-chains, and on the hydrogen bond between the N166 side-chain of one protomer and the G71 backbone oxygen of another. A second putative  $\text{SO}_4^{2-}$  ion is observed close to the first one that interacts with K176 and R179 side-chains. Potential interactions are marked by dashed lines.  $F_o - F_c$  difference electron densities before the insertion of the ions into the model are shown at the level of  $3\sigma$ . The symbols ', ' and ''' denote different crystallographic symmetry-related HmBRI molecules. The black oval denotes the crystallographic symmetry rotation axis ( $C_2$ ).

doi:10.1371/journal.pone.0112873.g003

### Structure of HmBRI

Similarly to other microbial rhodopsins, *HmBRI* is organized as a bundle of seven transmembrane helices (A-G) connected by relatively short loops, with the



**Figure 4. Crystallographic structure of the *HmBRI* D94N mutant.** (A) Comparison of the *HmBRI* backbone structure (green) with that of *HsBR* [23] (yellow). (B)  $2F_o - F_c$  electron density maps in the vicinity of the retinal. The maps are contoured at the level of  $1.5 \sigma$ . (C)  $2F_o - F_c$  electron density maps in the proton release region. The maps are contoured at the level of  $1.2 \sigma$ . (D) Comparison of the *HmBRI* proton release group (green) with that of *HsBR* [23] (yellow). Overall, the conformations of the side-chains and positions of water molecules are very similar. However, the water accessible space is larger in *HmBRI*, and additional water molecules are observed ( $W_{Hm}$ ). One of the reasons for this difference might be the substitution of *HsBR*' proline 200 with the glycine 204 in *HmBRI*, that allows unlatching of the extracellular part of the helix G. (E) Comparison of the *HmBRI* proton donor region with that of wild-type *HsBR* [23] (green) and its D96N [30] (orange) and P50A [32] (magenta) mutants. It appears that in *HmBRI* the effects of the D94N mutation and the P50<sub>Hs</sub> → T48<sub>Hm</sub> substitution combine and result in a larger displacement of the helix B relative to the helices C and G (black arrows), as similar displacements are present in the P50A and D96N mutants of *HsBR*, albeit with a smaller amplitude.

doi:10.1371/journal.pone.0112873.g004

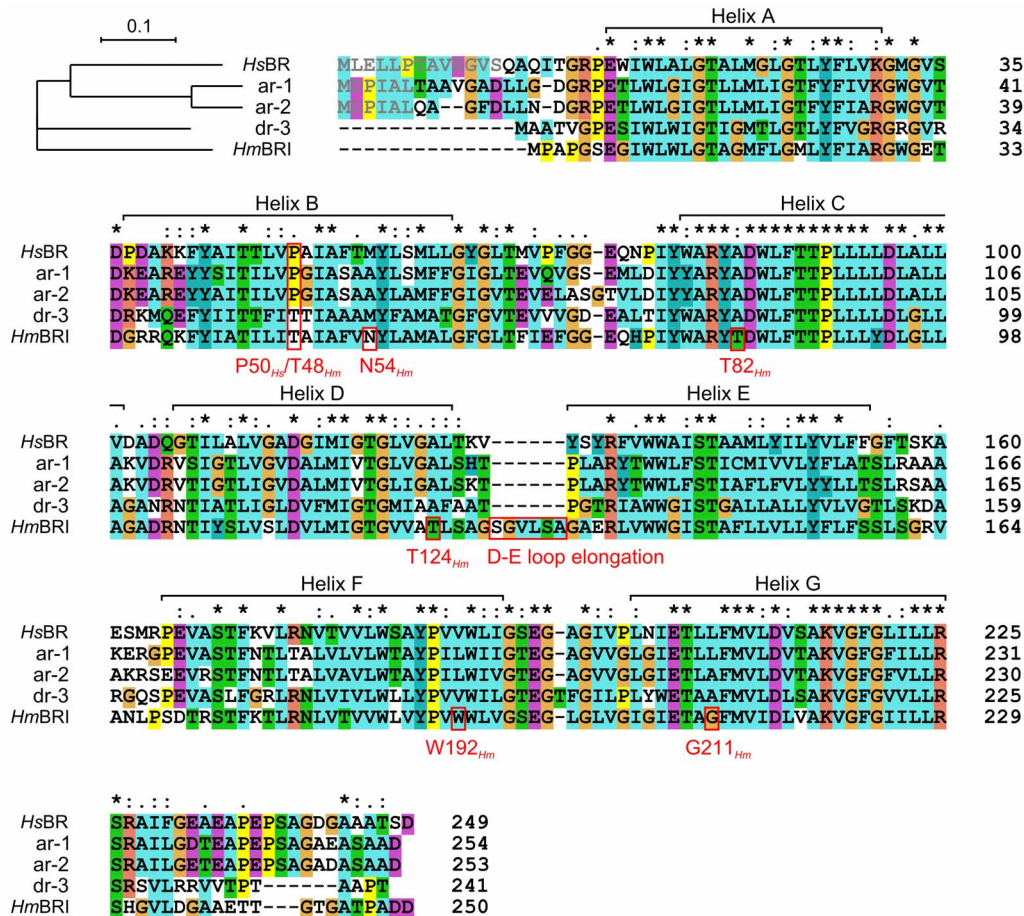


exception of the elongated beta hairpin-forming B-C loop, which caps the extracellular side, and the extended D-E loop that is unique to *HmBRI* (Figures 4A and 5). The root mean square deviation (RMSD) of the backbone heavy atoms positions between the *HmBRI* and *HsBR* (PDB ID 1C3W) structures is  $\sim 0.6$  Å. In the native *HmBRI*, all the important proton translocation pathway residues, corresponding to D96, D85 and D212, R82, E194 and E204 in *HsBR*, are conserved. For the crystallization trials, we have used the D94N mutant of *HmBRI* (which has higher expression level), corresponding to the D96N mutant of *HsBR*. This mutation is not expected to introduce the structural changes in the protein ground state other than in the vicinity of D94 [2, 30]. Consequently, we will start with the discussion of the retinal binding pocket and the proton release group of *HmBRI*.

At the core of the *HmBRI* retinal binding pocket is the cofactor retinal in the all-*trans* conformation (Figure 4B). Bound to the Schiff base is the water molecule W402, coordinated by the aspartates 83 and 216. This arrangement, observed in bacteriorhodopsin-like proton pumps [2, 23] as well as in proteorhodopsins [15, 16] appears to be critical for the proton pumping.

The proton release group configuration in *HmBRI* is similar to that in the highest-resolution ground state *HsBR* structure (PDB ID 1C3W), as the R80<sub>Hm</sub> side-chain (corresponds to R82<sub>Hs</sub>) is turned towards the Schiff base, and the E198<sub>Hm</sub> (E194<sub>Hs</sub>) and E208<sub>Hm</sub> (E204<sub>Hs</sub>) side-chains are close to each other (Figures 4C and 4D). Positions of the water molecules, such as 401, 403, 404, 405 and 407, are similar to those in *HsBR* (Figure 4D). However, the extracellular part of the *HmBRI*'s helix G appears to be slightly unlatched, which results in its outward displacement. As a consequence, there is more water-accessible space in the *HmBRI* proton release region, and two more water molecules are observed (Figures 4C and 4D). It is not clear how this alteration could affect the proton release mechanism [2], which possibly involves formation of Zundel cation [8, 31].

Finally, we analyze the *HmBRI* proton donor region and the effects of the introduced *HmBRI* mutation D94N and the P50<sub>Hs</sub> → T48<sub>Hm</sub> substitution (Figure 4E). Structures of the D96N and P50A mutants of *HsBR* have been determined before [30, 32] and allow us to compare the effect of double "mutation" in *HmBRI*. Structure of the *HsBR* D96N mutant reveals the outward motion of T46 side-chain, which is hydrogen-bonded to this residue 96. At the same time, positions of the residues close to T46 in sequence, such as P50, remain unchanged (Figure 4E). In the P50A *HsBR* mutant, on the contrary, the backbone around T46 is relatively unchanged, meanwhile the mutated A50 is moved away from the helix C. The *HmBRI* D94N structure reveals the most pronounced motion of the helix B relative to the helix C, which seems to be explained by the combination of the effects of the mutation D94N and substitution P50<sub>Hs</sub> → T48<sub>Hm</sub>. It must be noted that, while the structures of ar-1 and ar-2 are closer to *HsBR* in this region, dr-3 also has P50<sub>Hs</sub> → T substitution (Figure 5). Its structure is intermediate between *HmBRI* and *HsBR* P50A mutant [13, 32].



**Figure 5. Phylogenetic tree and sequence alignment of *HsBR*, *ar-1*, *ar-2*, *dr-3* and *HmBRI*.** The unique *HmBRI* D-E loop elongation, the residues that result in novel inter-helical hydrogen bonds, and the P50<sub>Hs</sub>/T48<sub>Hm</sub> position are highlighted. The propeptide residues that are cleaved in the mature protein *in vivo* are shown in grey. The asterisk indicates the fully conserved residues, the colon and period indicate conservation between groups of strongly similar properties scoring > 0.5 and ≤ 0.5 correspondingly in the Gonnet PAM 250 matrix [49].

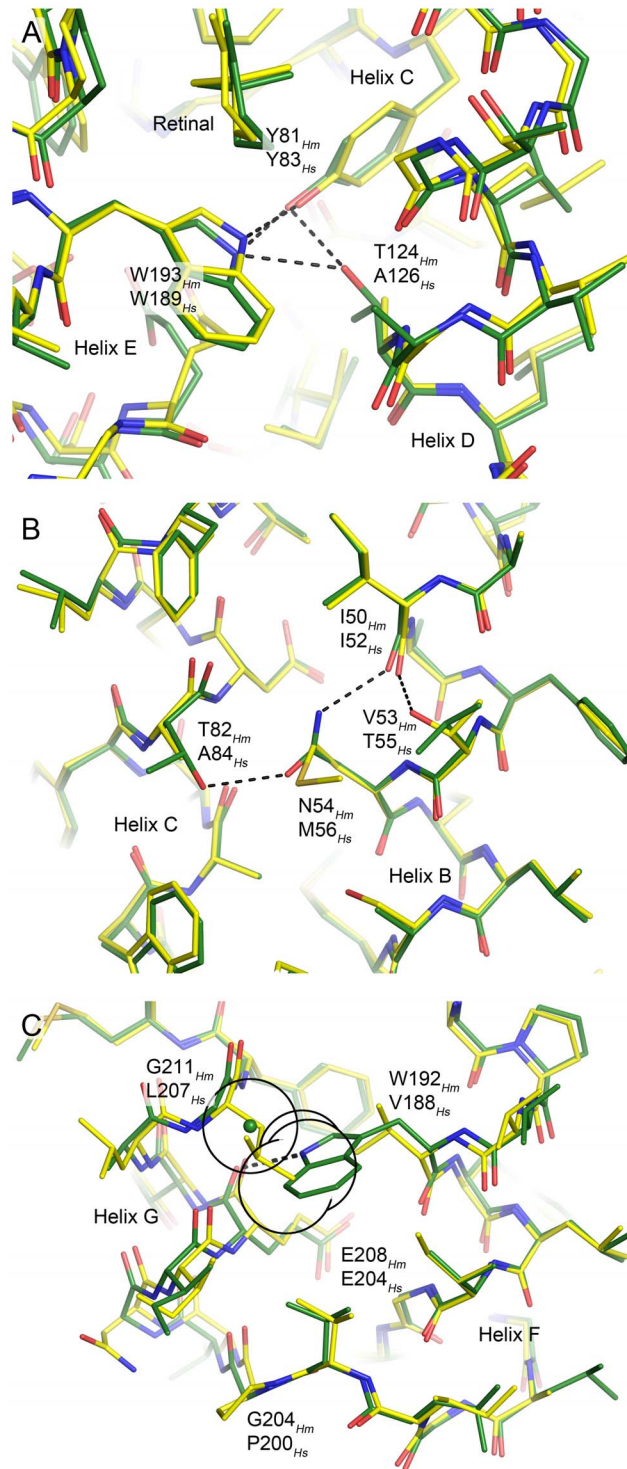
doi:10.1371/journal.pone.0112873.g005

### Novel inter-helical hydrogen bonds in *HmBRI*

There are several local alterations present in the *HmBRI* structure relative to the structure of *HsBR* and other bacteriorhodopsin-like proton pumps resulting from amino acid substitutions. First, T124 of *HmBRI* helix D is hydrogen-bonded to the tyrosyl oxygen of the helix C tyrosine 81 and to the indole nitrogen of the helix F tryptophan W193 (Figure 6A). While Y81 and W193 that are situated in the vicinity of the retinal β-ionone ring are extremely conserved among archaeal rhodopsins, the threonine amino acid is observed only in *HmBRI* (Figure 5).

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

The first coupled substitution, M56<sub>Hs</sub> → N54<sub>Hm</sub> and A84<sub>Hs</sub> → T82<sub>Hm</sub> results in introduction of the hydrogen bond between the helices B and C (Figure 6B). Interestingly, the intra-helical hydrogen bond between the I52 backbone oxygen and T55 side-chain that existed in *HsBR* is replaced with the hydrogen bond



**Figure 6. The novel inter-helical hydrogen bonds in HmBRI (green) relative to HsBR (yellow).** In each case, the structure alignment was done locally to emphasize the local effects. (A) The substitution  $A126_{Hs} \rightarrow T124_{Hm}$  results in two novel hydrogen bonds connecting the helix D to helices C and E. (B) The coupled substitution  $M56_{Hs} \rightarrow N54_{Hm}$ ,  $A84_{Hs} \rightarrow T82_{Hm}$  results in introduction of the hydrogen bond between the helices B and C. Interestingly, the intra-helical hydrogen bond between the I52 backbone oxygen and T55 is replaced with the hydrogen bond between the homologous I50 backbone oxygen and N54 side-chain amine.

(C) Coupled substitution  $V188_{Hs} \rightarrow W192_{Hm}$ ,  $L207_{Hs} \rightarrow G211_{Hm}$  results in introduction of the hydrogen bond between the helices F and G (E208 backbone oxygen and W192 indole nitrogen). Interestingly, the glycine is the only possible amino acid at the position 211, as the  $C_{\beta}$  atom of any other amino acid would create a steric conflict with W192 side-chain. The hypothetical position of the residue 211  $C_{\beta}$  atom is shown by the green sphere, and its van der Waals radius, as well as those of proximal W192 heavy atoms, is shown as a black circle. Introduction of the bulky tryptophan at the position 192 might be another reason for the divergence of the helices F and G in *HmBRI*, and, as a consequence, the bigger water-accessible volume of the proton release region.

doi:10.1371/journal.pone.0112873.g006

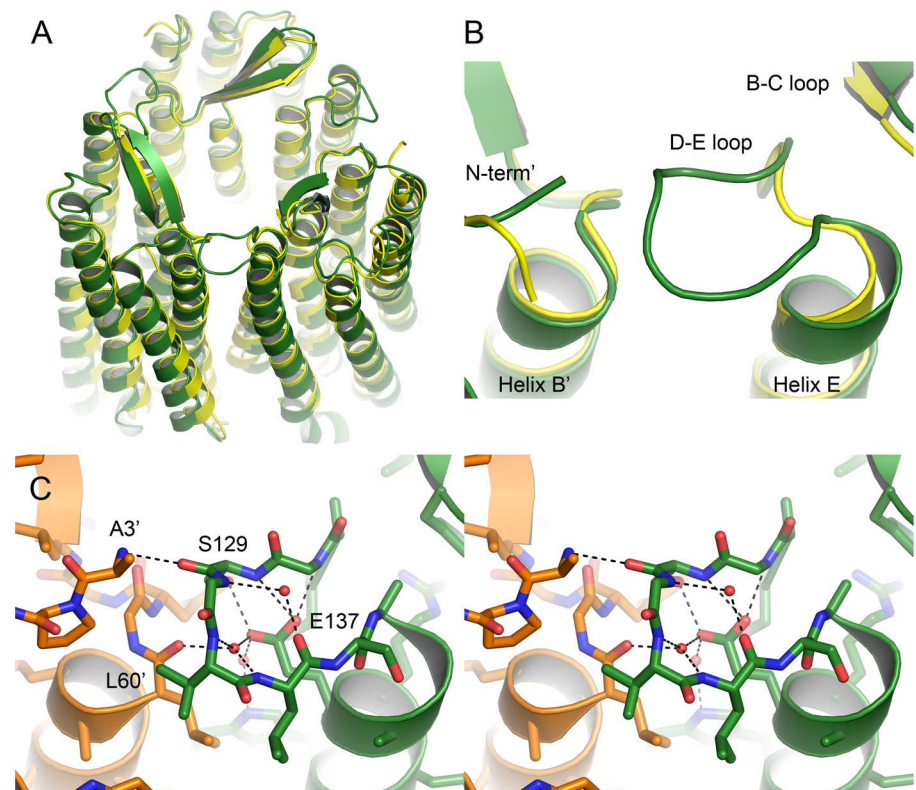
between the homologous I50 backbone oxygen and N54 side-chain amine in *HmBRI*. None of these bonds is present in ar-1, ar-2 or dr-3.

The second coupled substitution  $V188_{Hs} \rightarrow W192_{Hm}$  and  $L207_{Hs} \rightarrow G211_{Hm}$  results in introduction of the hydrogen bond between the helices F and G (namely, between the E208 backbone oxygen and W192 indole nitrogen; [Figure 6C](#)). Similar tryptophan nitrogen – backbone oxygen hydrogen bonds have already been observed in retinylidene proteins, however, always in special cases. The W138-P186 bond in *HsBR* (W142-P190 in *HmBRI*) links the helices E and F, moved away from each other by the intercalating retinal. The similar bond in proteorhodopsins (residues W154-I187 in ESR) brings the cytoplasmic end of the helix F closer to the helix E, but causes a transformation of the  $\alpha$ -helix to a complex conformation including a  $\pi$ -helix like element [15]. In *HmBRI*, glycine is the only possible amino acid at the position 211, as the  $C_{\beta}$  atom of any other amino acid would create a steric conflict with the W192 side-chain ([Figure 6C](#)). Thus, *HmBRI* structure presents an example of the inter-helical hydrogen bond involving tryptophan that does not result in disruption of normal  $\alpha$ -helical structure. Interestingly, the distance between the helices F and G is also elevated in ar-1 and ar-2, the reason for which is most probably the bigger-volume leucine amino acid at the place of *HsBR*'s valine. In dr-3, despite an additional amino acid in the F-G loop, the distance between the helices F and G is the same as in *HsBR*.

Although the stabilization of membrane proteins due to the hydrogen bonds in the transmembrane region might be moderate [33], presence of three additional bonds at the same time should undoubtedly stabilize *HmBRI* and might be a reason for its high *E. coli* expression level [20].

### HmBRI trimers and its D-E loop

Similarly to other archaeal proton pumps [10–13] and despite being heterologously expressed in a bacterium *Escherichia coli*, *HmBRI* assembles as a trimer in crystals ([Figure 7A](#)). The RMSD between the backbone heavy atoms positions in the *HmBRI* and *HsBR* trimers is  $\sim 1.0$  Å. This value reflexes the slightly different orientation of the *HmBRI* protomers in crystals, where the extracellular sides of the trimers align well, but the cytoplasmic sides are slightly displaced ([Figure 7A](#)). Such trimeric organization of *HmBRI* is similar to that of deltarhodopsin-3 [13], as the RMSD of atomic positions is  $\sim 0.45$  Å when either monomers or trimers are compared.



**Figure 7. Structure of the *HmBRI* trimer and its D-E loop.** (A) Comparison of the *HmBRI* trimer structure (green) with that of *HsBR* [23] (yellow). *HmBRI* trimer aligns well in the extracellular region, but the protomers are slightly rotated at the cytoplasmic side. (B) Magnification of the D-E loop. Unlike in other trimerizing retinylidene proteins, in *HmBRI* the loop is extended and makes contact to the adjacent protomer. (C) Wall-eyed stereogram of the *HmBRI* D-E loop. The adjacent protomer is shown in orange and its residues are marked by a prime. Three structural water molecules are observed that stabilize the loop structure.

doi:10.1371/journal.pone.0112873.g007

However, there is an important detail that separates *HmBRI* from the others bacteriorhodopsin-like proton pumps (Figure 7B). The *HmBRI*'s D-E loop is 6 residues longer than the loops of other trimeric proton pumps and makes contacts with the adjacent protomer (Figures 5 and 7B-C). At the core of the loop are the buried E137 side-chain and three structural water molecules, participating in numerous hydrogen bonds (Figure 7C). The loop expands the trimerization interface and might stabilize the trimeric assembly.

### Lipid molecules around *HmBRI*

Interestingly, all the retinylidene proteins, for which the trimeric assembly was observed in crystal, were purified from natural or natural-like sources [12, 13, 23, 34–38]. The *HmBRI* crystals, presented here, were grown with the protein expressed in *Escherichia coli*, a bacterium, although the protein itself originates from the archaeon *Haloarcula marismortui*. The ordered lipid tails, belonging either to the *E. coli* lipids or the *in meso* crystallization matrix lipid

monooleoyl, are observed at the same positions as the native lipids in other structures (Figure 8). First, there are three paired hydrophobic tails at the extracellular side of the intra-trimer cavity of *HmBRI* (Figure 8A), where binding of highly specific sulfated triglycoside lipid S-TGA-1 is observed in *HsBR* crystals [34–36]. Second, there are ordered hydrophobic tails on the outer surface of the trimer (Figures 8B and 8C), whose position is also very similar to that in the crystals of *HsBR* and other proteins. It is remarkable that the lipid binding mode is conserved not only across the different proteins of the family but even across the different kinds of the hydrophobic tail moieties of the lipid molecules (branched isoprenoid chains in archaea and straight fatty acid chains in bacteria). Thus, stronger ability to assemble into a trimer and bind the host lipids might be another reason for the efficient expression of *HmBRI* in *E. coli* cells.

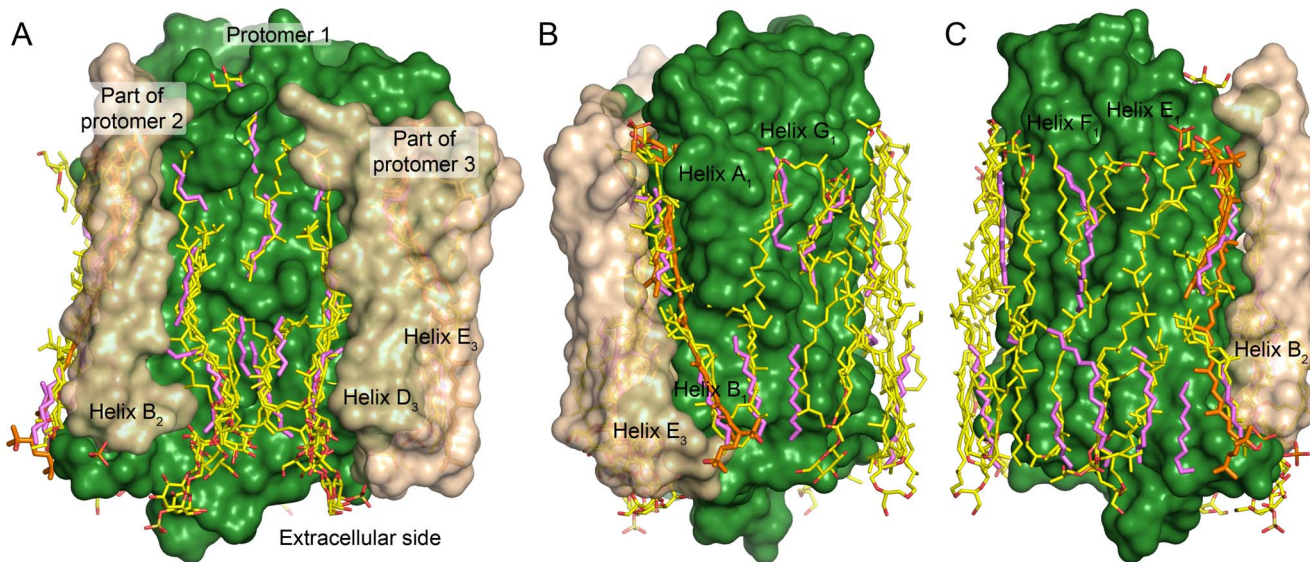
## Conclusions

Here, we have presented the structure of the D94N mutant of *Haloarcula marismortui* bacteriorhodopsin I at the resolution of 2.5 Å. The protein is trimeric in the crystal. Its retinal-binding pocket and proton acceptor site are similar to those of other archaeal proton pumps, but its proton release region is extended and contains additional water molecules. There are three novel inter-helical hydrogen bonds and a unique extended loop between the helices D and E that makes contacts with the adjacent protomer and might stabilize the trimer. Many lipidic hydrophobic tail groups are discernible in the membrane region, and their position is similar to that of archaeal lipids in the crystals of other proton pumps, isolated from native or native-like sources. Being easily expressed in *E. coli* and crystallized using the *in meso* approach, *HmBRI* can become a complementary model for studies of the light-driven proton pumping and more general topics such as the link between the protein structure, stability, lipid binding and expression level.

## Materials and Methods

### Construction of Expression Plasmid

*Haloarcula marismortui* bacterio-opsin gene (*bop*, UniProt ID Q5UXY6) with D94N mutation was synthesized *de novo*. The nucleotide sequence was optimized for *E. coli* expression using the GeneOptimizer software (Life Technologies, USA). The gene was introduced into the pSCodon1.2 expression vector (Staby Codon T7, Eurogentec, Belgium) via *NdeI* and *XhoI* restriction sites, resulting in pSCo-*HmBRI*-His<sub>6</sub> expression plasmid. Consequently, the expressed construct harbored an additional C-terminal tag with a sequence LEHHHHHH.



**Figure 8. Comparison of the ordered lipidic tails observed in *HmBRI* structure with those observed in the structures of trimeric *HsBR* [23, 34–36], *ar-2* [29] and *dr-3* [13].** The *HmBRI* surface is shown in green, parts of the adjacent protomers are in beige, *HmBRI* lipids are in magenta, the other lipids are in yellow and bacterioruberin molecules observed in *aR-2* and *dR-3* structures are in orange. (A) Lipids inside the trimer. (B) Lipids close to the helices A, B and G. (C) Lipids close to the helices E and F. To obtain the positions of the lipidic moieties observed in the structures of the other proteins, their trimeric assemblies were aligned first to *HmBRI* trimer.

doi:10.1371/journal.pone.0112873.g008

### Protein Expression and Purification

An *E. coli* strain SE1 cells (Staby Codon T7, Eurogentec, Belgium) were transformed with the pSC-*oHmBRI*-His<sub>6</sub> plasmid. The cells were grown at 37°C in shaking baffled flasks in an auto-inducing medium ZYP-5052 [39] containing 100 mg/L ampicillin. After the glucose level in the growing bacterial culture dropped below 10 mg/L, the incubation temperature was reduced to 20°C and incubation continued overnight. Collected cells were disrupted using the M-110P Lab Homogenizer (Microfluidics) at 25000 psi in a buffer containing 20 mM Tris-HCl pH 8.0, 5% glycerol and 50 mg/L DNase (Sigma-Aldrich, USA). Membrane fraction of cell lysate was obtained by ultracentrifugation at 90000 g for 1 h at 4°C. The pellets were resuspended in a buffer containing 20 mM Tris-HCl pH 8.0, 0.1 M NaCl and 1% DDM (Anatrace, Affymetrix, USA). All-trans-retinal (Sigma-Aldrich, USA) was added to 10 μM and immediate red shift of the solution color was observed. The mixture was left overnight for solubilization. Insoluble fraction part was removed by ultracentrifugation at 90000 g for 1 h at 4°C. The supernatant was loaded on Ni-NTA column (Qiagen, Germany) and the His-tagged protein was eluted in a buffer containing 20 mM Tris-HCl pH 7.5, 0.1 M NaCl, 50 mM EDTA, 0.01% DDM. The eluate was subjected to size-exclusion chromatography (125 ml Superdex 200 PG, GE Healthcare Life Sciences, USA) in a buffer containing 50 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> pH 7.5, 0.1 M NaCl, 0.01% DDM. Five protein-containing colored fractions were collected, pooled and concentrated to 40 mg/ml for crystallization.

## Spectroscopic characterization

Absorption spectra of *HmBRI* in solution were collected using the UV-2401PC spectrometer (Shimadzu, Japan). To measure the *HmBRI* absorption spectra in *H. salinarum* polar lipids, the protein was reconstituted into the liposomes: the polar lipids were added to the solubilized protein in the ratio 3:1 by mass and the detergent was removed by overnight incubation with Bio-Beads SM-2 (Bio-Rad, USA) at 4°C. Absorption spectra of *HmBRI* in crystals were collected at the ID-29S cryobench laboratory at the European Synchrotron Radiation Facility (ESRF), Grenoble, France [40] and corrected for the background with a linear function to match the scattering profiles of the protein in solution.

## Crystallization details

The crystals were grown using the *in meso* approach [26, 27], similarly to our previous work [15, 28]. The solubilized protein in the crystallization buffer was added to the monooleoyl-formed lipidic phase (Nu-Chek Prep, USA). The best crystals were obtained using the protein concentration of 20 mg/mL and the 0.1M Tris pH 8.8 and 2.6M Ammonium Sulfate precipitate solution from the Qiagen Cubic Phase I screening kit (Qiagen, Germany). Crystallization trials were set up using the NT8 robotic system (Formulatrix, USA). The crystals were grown at 22°C. Regular shaped hexagonal crystals reaching 100 µm in size appeared in approximately 4 weeks.

## Acquisition and treatment of diffraction data

X-ray diffraction data (wavelength 0.976 Å) were collected at the beamline ID23-1 of the ESRF, using a PILATUS 6M detector. Diffraction patterns were integrated using MOSFLM [41]. The diffraction data were analyzed using the POINTLESS software [42], and the space group was determined to be P321. The reflexes' intensities were scaled using the SCALA software from the CCP4 program suite [43]. The crystals were twinned, with the twin fraction in the range 0–25%. The crystal with the best diffraction was not twinned and was used for structure determination. The data were collected in two wedges from different parts of the same crystal. The diffraction was slightly anisotropic ( $I/\sigma I$  of  $\sim 2.5$  at 2.5 Å along the directions *a* and *b*,  $I/\sigma I$  of  $\sim 2.0$  for all the reflections in the range 2.5 ÷ 2.65 Å along the direction *c*). The data statistics are presented in the Table 2.

## Structure determination and refinement

Initial phases were successfully obtained in the P321 space group by a molecular replacement (MR) method using MOLREP [44]. There is one monomer in the asymmetric unit in this space group. The *HmBRI* homology model for MR search was built using the automatic modeling server RaptorX [45]. The R-factors were  $\sim 45\%$  immediately after the MR step and  $\sim 40\%$  after 4 cycles of automatic refinement with REFMAC5 [46]. The initial MR model was then iteratively



refined using REFMAC5 [46] and Coot [47]. The final steps of the refinement were conducted using the PHENIX software [48].

### Accession codes

Coordinates and structure factors have been deposited in the Protein Data Bank under the accession code 4PXK.

### Acknowledgments

The diffraction experiments were performed on the beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to local contacts at ESRF for providing assistance in using these beamlines. We are also thankful to David von Stetten and Antoine Royant (ESRF Cryobench laboratory) for help with collection of *HmBRI* crystals visual absorption spectra, to Alexey Vlasov, Yury Ryzhykau and Egor Zinovev for help with protein expression and purification, and to Francisco Rodriguez-Valera for fruitful discussions.

### Author Contributions

Wrote the paper: IG VG. Expressed and purified the protein: VS. Supervised the expression and purification: TB. Crystallized the protein: VP. Helped with automated crystallization: ER PU. Measured the absorption spectra: VS IG VP. Collected the data and solved the structure: IG. Helped with data collection: AP. Helped with structure analysis: VB GB. Supervised the project: VG. Analyzed the results and prepared the manuscript with input from all the other authors: IG VG.

### References

1. Ernst OP, Lodowski DT, Elstner M, Hegemann P, Brown LS, et al. (2014) Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chem Rev* 114: 126–163. doi:10.1021/cr4003769.
2. Lanyi JK (2004) Bacteriorhodopsin. *Annual Review of Physiology* 66: 665–688. doi:10.1146/annurev.physiol.66.032102.150049.
3. Oesterhelt D, Stoerkenius W (1971) Rhodopsin-like Protein from the Purple Membrane of *Halobacterium halobium*. *Nature* 233: 149–152. doi:10.1038/10.1038/newbio233149a0.
4. Hirai T, Subramaniam S (2009) Protein Conformational Changes in the Bacteriorhodopsin Photocycle: Comparison of Findings from Electron and X-Ray Crystallographic Analyses. *PLoS ONE* 4: e5769. doi:10.1371/journal.pone.0005769.
5. Wickstrand C, Dods R, Royant A, Neutze R (n.d.) Bacteriorhodopsin: Would the real structural intermediates please stand up? *Biochimica et Biophysica Acta (BBA) - General Subjects*. Available: <http://www.sciencedirect.com/science/article/pii/S0304416514002190>. Accessed 2014 October 4.
6. Borshchevskiy VI, Round ES, Popov AN, Büldt G, Gordeliy VI (2011) X-ray-Radiation-Induced Changes in Bacteriorhodopsin Structure. *Journal of Molecular Biology* 409: 813–825. doi:10.1016/j.jmb.2011.04.038.

7. **Borshchevskiy V, Efremov R, Moiseeva E, Büldt G, Gordeliy V** (2010) Overcoming merohedral twinning in crystals of bacteriorhodopsin grown in lipidic mesophase. *Acta Crystallographica Section D Biological Crystallography* 66: 26–32. doi:10.1107/S0907444909042838.
8. **Freier E, Wolf S, Gerwert K** (2011) Proton transfer via a transient linear water-molecule chain in a membrane protein. *PNAS* 108: 11435–11439. doi:10.1073/pnas.1104735108.
9. **Wang T, Sessions AO, Lunde CS, Rouhani S, Glaeser RM, et al.** (2013) Deprotonation of D96 in Bacteriorhodopsin Opens the Proton Uptake Pathway. *Structure* 21: 290–297. doi:10.1016/j.str.2012.12.018.
10. **Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R** (1996) Electron-crystallographic Refinement of the Structure of Bacteriorhodopsin. *Journal of Molecular Biology* 259: 393–421. doi:10.1006/jmbi.1996.0328.
11. **Pebay-Peyroula E, Rummel G, Rosenbusch JP, Landau EM** (1997) X-ray Structure of Bacteriorhodopsin at 2.5 Angstroms from Microcrystals Grown in Lipidic Cubic Phases. *Science* 277: 1676–1681. doi:10.1126/science.277.5332.1676.
12. **Enami N, Yoshimura K, Murakami M, Okumura H, Ihara K, et al.** (2006) Crystal Structures of Archaeorhodopsin-1 and -2: Common Structural Motif in Archaeal Light-driven Proton Pumps. *Journal of Molecular Biology* 358: 675–685. doi:10.1016/j.jmb.2006.02.032.
13. **Zhang J, Mizuno K, Murata Y, Koide H, Murakami M, et al.** (2013) Crystal structure of deltarhodopsin-3 from *Haloterrigena thermotolerans*. *Proteins: Structure, Function, and Bioinformatics* 81: 1585–1592. doi:10.1002/prot.24316.
14. **Luecke H, Schobert B, Stagno J, Imasheva ES, Wang JM, et al.** (2008) Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *PNAS* 105: 16561–16565. doi:10.1073/pnas.0807162105.
15. **Gushchin I, Chervakov P, Kuzmichev P, Popov AN, Round E, et al.** (2013) Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. *PNAS* 110: 12631–12636. doi:10.1073/pnas.1221629110.
16. **Ran T, Ozorowski G, Gao Y, Sineshchikov OA, Wang W, et al.** (2013) Cross-protomer interaction with the photoactive site in oligomeric proteorhodopsin complexes. *Acta Crystallogr D Biol Crystallogr* 69: 1965–1980. doi:10.1107/S0907444913017575.
17. **Wada T, Shimono K, Kikukawa T, Hato M, Shinya N, et al.** (2011) Crystal Structure of the Eukaryotic Light-Driven Proton-Pumping Rhodopsin, *Acetabularia* Rhodopsin II, from Marine Alga. *Journal of Molecular Biology* 411: 986–998. doi:10.1016/j.jmb.2011.06.028.
18. **Kato HE, Zhang F, Yizhar O, Ramakrishnan C, Nishizawa T, et al.** (2012) Crystal structure of the channelrhodopsin light-gated cation channel. *Nature* 482: 369–374. doi:10.1038/nature10870.
19. **Fu H-Y, Lin Y-C, Chang Y-N, Tseng H, Huang C-C, et al.** (2010) A Novel Six-Rhodopsin System in a Single Archaeon. *J Bacteriol* 192: 5866–5873. doi:10.1128/JB.00642-10.
20. **Hsu M-F, Yu T-F, Chou C-C, Fu H-Y, Yang C-S, et al.** (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. doi:10.1371/journal.pone.0056363.
21. **Deisseroth K** (2011) Optogenetics. *Nat Meth* 8: 26–29. doi:10.1038/nmeth.f.324.
22. **Ihara K, Amemiya T, Miyashita Y, Mukohata Y** (1994) Met-145 is a key residue in the dark adaptation of bacteriorhodopsin homologs. *Biophysical Journal* 67: 1187–1191. doi:10.1016/S0006-3495(94)80587-9.
23. **Luecke H, Schobert B, Richter H-T, Cartailier J-P, Lanyi JK** (1999) Structure of bacteriorhodopsin at 1.55 Å resolution. *Journal of Molecular Biology* 291: 899–911. doi:10.1006/jmbi.1999.3027.
24. **Reynolds JA, Stoeckenius W** (1977) Molecular weight of bacteriorhodopsin solubilized in Triton X-100. *PNAS* 74: 2803–2804.
25. **Dencher NA, Kohl KD, Heyn MP** (1983) Photochemical cycle and light-dark adaptation of monomeric and aggregated bacteriorhodopsin in various lipid environments. *Biochemistry* 22: 1323–1334.
26. **Landau EM, Rosenbusch JP** (1996) Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *PNAS* 93: 14532–14535.

27. **Caffrey M, Cherezov V** (2009) Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* 4: 706–731. doi:10.1038/nprot.2009.31.
28. **Gordeliy VI, Labahn J, Moukhametzianov R, Efremov R, Granzin J, et al.** (2002) Molecular basis of transmembrane signalling by sensory rhodopsin II–transducer complex. *Nature* 419: 484–487. doi:10.1038/nature01109.
29. **Yoshimura K, Kouyama T** (2008) Structural Role of Bacterioruberin in the Trimeric Structure of Archaeorhodopsin-2. *Journal of Molecular Biology* 375: 1267–1281. doi:10.1016/j.jmb.2007.11.039.
30. **Luecke H, Schobert B, Richter H-T, Cartailler J-P, Lanyi JK** (1999) Structural Changes in Bacteriorhodopsin During Ion Transport at 2 Angstrom Resolution. *Science* 286: 255–260. doi:10.1126/science.286.5438.255.
31. **Gerwert K, Freier E, Wolf S** (n.d.) The role of protein-bound water molecules in microbial rhodopsins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. Available: <http://www.sciencedirect.com/science/article/pii/S000527281300162X>. Accessed 18 December 2013.
32. **Faham S, Yang D, Bare E, Yohannan S, Whitelegge JP, et al.** (2004) Side-chain contributions to membrane protein structure and stability. *J Mol Biol* 335: 297–305.
33. **Joh NH, Min A, Faham S, Whitelegge JP, Yang D, et al.** (2008) Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* 453: 1266–1270. doi:10.1038/nature06977.
34. **Essen L-O, Siebert 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: 11673–11678.
35. **Belrhali H, Nollert P, Royant A, Menzel C, Rosenbusch JP, et al.** (1999) Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Structure* 7: 909–917. doi:10.1016/S0969-2126(99)80118-X.
36. **Takeda K, Sato H, Hino T, Kono M, Fukuda K, et al.** (1998) A novel three-dimensional crystal of bacteriorhodopsin obtained by successive fusion of the vesicular assemblies. *Journal of Molecular Biology* 283: 463–474. doi:10.1006/jmbi.1998.2103.
37. **Kolbe M, Besir H, Essen L-O, Oesterhelt D** (2000) Structure of the Light-Driven Chloride Pump Halorhodopsin at 1.8 Å Resolution. *Science* 288: 1390–1396. doi:10.1126/science.288.5470.1390.
38. **Kouyama T, Kanada S, Takeguchi Y, Narusawa A, Murakami M, et al.** (2010) Crystal Structure of the Light-Driven Chloride Pump Halorhodopsin from *Natronomonas pharaonis*. *Journal of Molecular Biology* 396: 564–579. doi:10.1016/j.jmb.2009.11.061.
39. **Studier FW** (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* 41: 207–234. doi:10.1016/j.pep.2005.01.016.
40. **Royant A, Carpentier P, Ohana J, McGeehan J, Paetzold B, et al.** (2007) Advances in spectroscopic methods for biological crystals. 1. Fluorescence lifetime measurements. *Journal of Applied Crystallography* 40: 1105–1112. doi:10.1107/S0021889807044196.
41. **Leslie AGW, Powell HR** (2007) Processing diffraction data with mosflm. In: Read RJ, Sussman JL, editors. *Evolving Methods for Macromolecular Crystallography*. NATO Science Series. Springer Netherlands. pp. 41–51. Available: [http://link.springer.com/chapter/10.1007/978-1-4020-6316-9\\_4](http://link.springer.com/chapter/10.1007/978-1-4020-6316-9_4). Accessed 21 November 2012.
42. **Evans P** (2005) Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography* 62: 72–82. doi:10.1107/S0907444905036693.
43. **Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, et al.** (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography* 67: 235–242. doi:10.1107/S0907444910045749.
44. **Vagin A, Teplyakov A** (2009) Molecular replacement with MOLREP. *Acta Crystallographica Section D Biological Crystallography* 66: 22–25. doi:10.1107/S0907444909042589.
45. **Källberg M, Wang H, Wang S, Peng J, Wang Z, et al.** (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7: 1511–1522. doi:10.1038/nprot.2012.085.
46. **Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, et al.** (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography* 67: 355–367. doi:10.1107/S0907444911001314.

47. **Emsley P, Cowtan K** (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* 60: 2126–2132. doi:10.1107/S0907444904019158.
48. **Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, et al.** (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* 66: 213–221. doi:10.1107/S0907444909052925.
49. **Gonnet GH, Cohen MA, Benner SA** (1992) Exhaustive matching of the entire protein sequence database. *Science* 256: 1443–1445.