



**HAL**  
open science

# Interactions of light-driven membrane proteins with their environment

Maksim Rulev

► **To cite this version:**

Maksim Rulev. Interactions of light-driven membrane proteins with their environment. Structural Biology [q-bio.BM]. Université Grenoble Alpes [2020-..], 2021. English. NNT : 2021GRALV045 . tel-04055297

**HAL Id: tel-04055297**

**<https://theses.hal.science/tel-04055297>**

Submitted on 2 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 25 mai 2016

Présentée par

### Maksim RULEV

Thèse dirigée par **Gordon LEONARD**, SB group Leader, HDR, ESRF, Grenoble,

et codirigée par **Alexander POPOV**, ESRF, Grenoble, et par Valentin GORDELIY, HDR, ICS-6, Jülich, IBS, CNRS, Grenoble

préparée au sein du **Structural Biology Group (SB), European Synchrotron Radiation Facility (ESRF)**  
dans l'**École Doctorale Chimie et Sciences du Vivant**

## Interactions des protéines membranaires induites par la lumière avec leur environnement

## Interactions of light-driven membrane proteins with their environment

Thèse soutenue publiquement le **1 octobre 2021**,  
devant le jury composé de :

**Monsieur Gordon LEONARD**

CHERCHEUR HDR, ESRF - EUROPEAN SYNCHROTRON, Directeur de thèse

**Monsieur Johann P. KLARE**

CHERCHEUR, Universität Osnabrück, Rapporteur

**Madame Svetlana ANTONYUK**

PROFESSEUR ASSISTANT, The University of Liverpool, Rapporteur

**Monsieur Victor SOURJIK**

PROFESSEUR, Max Planck Institute for Terrestrial Microbiology, Examineur

**Monsieur Oded LIVNAH**

PROFESSEUR, The Hebrew University of Jerusalem, Examineur

**Monsieur Marc JAMIN**

PROFESSEUR, Université Grenoble Alpes, Président





## Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisors Dr. Gordon Leonard, Prof. Valentin Gordeliy, Dr. Alexander Popov for their continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge. This guidance helped me in all the time of research and during the writing of this thesis. I could not have imagined having better advisors and mentors for my Ph.D study.

Besides my advisors, I would like to thank the members of my thesis advisory committee: Prof. Georg Bueldt and Dr. Gleb Bourenkov for their insightful comments and encouragement, but also for the hard questions which motivated me to widen my research from various perspectives.

My sincere thanks also go to Dr. Montserrat Soler-Lopez (ESRF), Dr. Igor Melnikov (ESRF), and Dr. Philippe Carpentier (CEA, Grenoble), who provided me with the opportunity to join their teams and who gave me access to their laboratories and research facilities. Without their precious support it would not have been possible to conduct this research.

I am also sincerely grateful to Dr. Yury Rizhikau (MIPT, Russia), Prof. Alexander Kuklin (JINR, MIPT, Russia), Dr. Andrey Rogachev (JINR, MIPT, Russia), Dr. Philip Orekhov (MSU, Russia) and Dr. Ivan Gushchin (MIPT, Russia), Dr. Kirill Kovalev (EMBL, Hamburg) and PhD Roman Astashkin (IBS, Grenoble) with whom we conducted joint research, regularly discussed results and could exchange experiences. Without their valuable help, this research would have been impossible.

I would also like to express my gratitude to the group of Dr. Albert Guskov (University of Groningen, Groningen) for their invaluable help in conducting experiments on electron microscopy. I also thank Dr. Taras Balandin (ICS-6, Julich) for his valuable advice in the field of biochemistry.

I thank my fellow lab mates for stimulating discussions, for the sleepless nights when working together before deadlines, and for all the fun we have had in the last years. Also I thank my friends in the following institutions: ESRF, IBS, ICS-6, MIPT.

I am also grateful to my thesis defence jury: Prof. Marc Jamin, Dr. Johann P Klare, Dr. Svetlana Antonyuk, Prof. Victor Surjik and Prof. Oded Livnah for their interest in the work described here and for the time spent in judging and commenting on it.

Last, but not the least, I would like to thank my family for supporting me spiritually throughout writing of this thesis and my life in general.

## List of Abbreviations

2D	Two dimensional
CD	Circular dichroism
DDM	n-Dodecyl-B-D-maltoside
DLPC	1,2-dilauroyl-sn-glycero-3-phosphocholine
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DNA	Deoxyribonucleic acid
DPPC	1,2-dipalmitoyl-sn-glycero-3-phosphocholine
<i>E. Coli (Ec)</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ESRF	European Synchrotron Radiation Facility
Gly	Glycine
GPCR	G protein-coupled receptor
<i>H. Salinarum</i>	<i>Halobacteria Salinarum</i>
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
LB	Lycogeny broth
LCP	Lipidic cubic phase
LN	Liquid nitrogen
MD	Molecular dynamics
MO	Monoolein
MP	Membrane protein
MR	Microbial rhodopsin
MSP	Membrane scaffold protein
MX	Macromolecular crystallography
ND	Nanodisc

Ni-NTA	Nickel-Nitrilotriacetic Acid
NMR	Nuclear magnetic resonance
<i>Np</i>	<i>Natronomonas pharaonis</i>
OD	Optical Density
PDB	Protein Data Bank
PCR	Polymerase Chain Reaction
PMSF	Phenylmethylsulfonyl fluoride
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
PRG	Proton release group
TB	Terrific broth
TCS	Two-component signaling system
TM	Transmembrane
RPM	Rotations per minute
SANS	Small-angle neutron scattering
SAS	Small-angle scattering
SAXS	Small-angle X-Ray scattering
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TRIS	2-amino-2-(hydroxymethyl)propane-1,3-diol
SOC	Super Optimal broth with Catabolite repression
OPO	Main movable reflector
DPO	auxiliary movable reflector
NPT	Isothermal-isobaric
TES	2- {[1,3-Dihydroxy-2-(hydroxymethyl)propan-2-yl]amino} ethane-1-sulfonic acid

## Table of contents

Acknowledgements.....	2
List of Abbreviations.....	4
Table of contents.....	6
List of Figures.....	9
List of Tables.....	14
Abstract.....	15
Résumé.....	17
<b>1 Introduction.....</b>	<b>19</b>
<b>1.1 Microbial Rhodopsins.....</b>	<b>22</b>
<b>1.2 Two-component protein signaling systems.....</b>	<b>24</b>
<b>1.3 The complex between sensory rhodopsin II and its related transducer molecule (<i>NpSRII/NpHtrII</i>).....</b>	<b>36</b>
<b>1.4 Application of small-angle scattering methods for structural studies of membrane proteins.....</b>	<b>41</b>
<b>1.5 Noble gases and membrane proteins.....</b>	<b>43</b>
<b>1.6 Thermodynamic principles of noble gas-protein interactions.....</b>	<b>46</b>
<b>1.7 Anomalous Scattering of Noble Gases.....</b>	<b>50</b>
<b>2 Materials and Methods.....</b>	<b>54</b>
<b>2.1 Expression and purification of the <i>NpSrII/NpHtrII</i> complex.....</b>	<b>54</b>
<b>2.1.1 Expression and purification of DNA.....</b>	<b>55</b>
<b>2.1.2 Restrictase ligase cloning.....</b>	<b>55</b>
<b>2.1.3 Preparation of competent cells.....</b>	<b>56</b>
<b>2.1.4 Transformation of plasmid DNA into <i>E. coli</i> cells.....</b>	<b>56</b>
<b>2.1.5 Expression of proteins in <i>E. coli</i>.....</b>	<b>57</b>

2.1.6 Purification of <i>NpSRII/NpHtrII</i> <sub>137</sub> and <i>NpSRII/NpHtrII</i> .....	58
2.1.7 Nanodisc Assembly and Reconstitution of Membrane Proteins .....	58
2.2 Crystallization by the <i>in-meso</i> method.....	60
2.3 Negative stain electron microscopy .....	62
2.4 Cryo-electron microscopy .....	63
2.5 Small-angle scattering measurements.....	63
2.5.1 SAXS measurements at BM29, ESRF, Grenoble, France .....	63
2.5.2 SANS measurements at YuMO, IBR-2, JINR, Dubna, Russian Federation .....	64
2.5.3 Small-angle scattering data processing.....	66
2.6 Microbial rhodopsins derivatized with Noble Gases.....	66
2.6.1 Protein expression and purification .....	66
2.6.2 Crystallisation .....	66
2.6.3 High-pressure derivatization .....	67
2.6.4 Crystallographic data collection.....	68
2.6.5 Structure solution and refinement .....	69
2.5.6 Molecular dynamics simulations .....	70
3.Results.....	72
3.1 Investigation of the <i>NpSRII/NpHtrII</i> complex solubilized in DDM.....	72
3.1.1 Gene Optimization .....	72
3.1.2 Optimization of expression and purification protocols.....	72
3.1.3 Single particle cryo-electron microscopy analysis of <i>NpSrII/NpHtrII</i> .....	77
3.1.4 Crystallization of <i>NpSrII/NpHtrII</i> solubilized in DDM .....	80
3.1.5 Proteases activities .....	83

<b>3.1.6 Small-angle Scattering investigations of <i>Np</i>SrII/<i>Np</i>HtrII<sub>137</sub> and of full-length <i>Np</i>SrII/<i>Np</i>HtrII .....</b>	<b>83</b>
<b>3.2 <i>Np</i>SRII/<i>Np</i>HtrII reconstituted in nanodiscs.....</b>	<b>87</b>
<b>3.2.1 Crystallization and X-ray diffraction analysis.....</b>	<b>87</b>
<b>3.2.2 Small-angle scattering measurements from solutions of <i>Np</i>SRII/<i>Np</i>HtrII reconstituted in nanodiscs .....</b>	<b>90</b>
<b>3.2.3 Single particle cryo-electron microscopy analysis of <i>Np</i>SRII/<i>Np</i>HtrII reconstituted in nanodiscs .....</b>	<b>93</b>
<b>3.3 Membrane Proteins Interactions Probed by Noble Gases.....</b>	<b>97</b>
<b>3.3.1 Effect on MPs caused by noble gas binding .....</b>	<b>106</b>
<b>3.3.2 MD experiments strongly support and extend the results obtained using X-ray crystallography .....</b>	<b>107</b>
<b>3.3.3 Internal binding sites and their relevance to rhodopsin function...</b>	<b>108</b>
<b>4. Conclusions.....</b>	<b>111</b>
<b>References .....</b>	<b>115</b>
<b>Publications.....</b>	<b>129</b>

## List of Figures

<b>Figure 1.1.1</b>	<b>The structure of the archetypal microbial rhodopsin bR.....</b>	<b>22</b>
<b>Figure 1.2.1</b>	<b>Domain architectures of representative microbial.....</b>	<b>24</b>
<b>Figure 1.2.2</b>	<b>Schematic of the signaling cascade in the two-component negative phototaxis system <i>NpSRII/NpHtrII</i> in <i>N. Pharaonis</i>..</b>	<b>25</b>
<b>Figure 1.2.3</b>	<b>Different oligomeric states of chemoreceptor TCS. ....</b>	<b>26</b>
<b>Figure 1.2.4</b>	<b>The crystal structure of the sensory domain dimer of the aspartate receptor from <i>S. typhimurium</i> combined with a simulated transmembrane region (left) and a diagram of hydrogen bonds formed upon the binding of aspartate (right).....</b>	<b>27</b>
<b>Figure 1.2.5</b>	<b>Schematic representations of atomic models obtained by X-ray crystallography of the dimer and trimer of dimers of the <i>EcTsrQ</i> cytoplasmic domain. ....</b>	<b>28</b>
<b>Figure 1.2.6</b>	<b>The crystal structure of the 3-element poly-HAMP region of the <i>Aer2</i> receptor.....</b>	<b>29</b>
<b>Figure 1.2.7</b>	<b>The crystal structures of a <i>NarQ</i> fragment in apo- and holo- states. ....</b>	<b>30</b>
<b>Figure 1.2.8</b>	<b>Electron tomography studies of <i>Tar</i> dimers reconstituted in nanodiscs. ....</b>	<b>31</b>
<b>Figure 1.2.9</b>	<b>Schematic representations of signaling arrays formed by twelve chemoreceptor trimers of dimers. ....</b>	<b>32</b>
<b>Figure 1.2.10</b>	<b>A schematic showing two different signaling states of the cytoplasmic regions of <i>Tsr</i> trimers of dimers.....</b>	<b>33</b>
<b>Figure 1.2.11</b>	<b>Structural studies of chemoreceptors exploiting cryoelectron tomography and averaging of subtomograms.....</b>	<b>34</b>
<b>Figure 1.3.1</b>	<b>The crystal structure of <i>NpSRII/NpHtrII</i><sub>114</sub>.....</b>	<b>36</b>
<b>Figure 1.3.2</b>	<b>The different conformations adopted by <i>NpSRII/NpHtrII</i><sub>114</sub>. ....</b>	<b>37</b>
<b>Figure 1.3.3</b>	<b>Structural studies of <i>NpHtrII</i><sub>229-504</sub> in solution.....</b>	<b>38</b>



<b>Figure 1.3.4 The "two-level" equilibrium between dynamic (dHAMP) and compact (cHAMP) conformations.</b> .....	39
<b>Figure 1.3.5 Cryo-EM studies of a negative phototaxis system TCS.</b> .....	40
<b>Figure 1.4.1 Solution scattering studies of membrane proteins.</b> .....	41
<b>Figure 1.4.2 Schematic of an automatic software pipeline for structural studies of membrane proteins based on small-angle X-ray scattering data.</b> .....	42
<b>Figure 1.5.1 The GLIC pentameric Ligand-Gated Ion Channel in its open form complexed to Xe (cyan spheres).</b> .....	44
<b>Figure 1.5.2 Xe-derivatization of OprN.</b> .....	44
<b>Figure 1.5.3 The Ammonium Transporter Amt-1 from <i>A. fulgidus</i> (grey surface) in complex with Xe (cyan spheres).</b> .....	45
<b>Figure 1.6.1 The electronic structure of noble gases.</b> .....	46
<b>Figure 1.6.2 Schematic diagrams of three types of weak electromagnetic interactions.</b> .....	47
<b>Figure 1.6.3 Pressure-temperature phase diagrams of argon and krypton.</b> ...	50
<b>Figure 1.7.1 A vector diagram illustrating equation (2) and showing that in the presence of anomalous scattering Freidel's law is no longer obeyed and that <math> F_{hkl}  \neq  F_{-h-k-l} </math>.</b> .....	51
<b>Figure 1.7.1 Theoretical values of the anomalous scattering properties of Ar (top), Kr (middle) and Xe (bottom).</b> .....	52
<b>Figure 2.1.1 General scheme of the pET27bmod co-expression plasmid with inserted modified NpSR<sub>II</sub> and NpHtr<sub>II</sub> genes.</b> .....	54
<b>Figure 2.1.7.1 Schematic step-by-step diagram of the assembly of empty nanodiscs.</b> .....	59
<b>Figure 2.1.7.2 Schematic step-by-step diagram of the assembly of nanodiscs containing a membrane protein.</b> .....	60
<b>Figure 2.2.1 Schematic step-by-step diagram of the in meso crystallization of membrane proteins.</b> .....	61
<b>Figure 2.2.2 The procedure for preparing proteins for in meso crystallization.</b> .....	62

<b>Figure 2.2.3</b>	<b>The procedure of dispensing protein mixtures onto LCP plates</b>	<b>62</b>
<b>Figure 2.4.1</b>	<b>Talos Arctica electron microscope (Groningen University, Groningen, The Netherlands).</b>	<b>63</b>
<b>Figure 2.5.1.1</b>	<b>Scheme of BM29 BioSAXS, ESRF.</b>	<b>64</b>
<b>Figure 2.5.2.1</b>	<b>Schematic of the YuMO, IBR-2, JINR facility.</b>	<b>65</b>
<b>Figure 2.6.2.1</b>	<b>Photographs of crystals of microbial rhodopsins obtained during this work.</b>	<b>67</b>
<b>Figure 2.6.3.1</b>	<b>The ESRF device for high pressure noble gas derivatization.</b>	<b>68</b>
<b>Figure 3.1.2.1</b>	<b>SDS-PAGE analysis of protein expression.</b>	<b>73</b>
<b>Figure 3.1.2.2</b>	<b>SDS-PAGE analysis of further protein expression tests.</b>	<b>74</b>
<b>Figure 3.1.2.3</b>	<b>SDS-PAGE analysis of protein expression as a function of post-induction time.</b>	<b>75</b>
<b>Figure 3.1.2.4</b>	<b>Purification of <i>NpSrIIStrep/NpHtrIIHis<sub>6</sub></i> via nickel affinity chromatography.</b>	<b>76</b>
<b>Figure 3.1.2.5</b>	<b>Purification of <i>NpSrIIStrep/NpHtrIIHis<sub>6</sub></i> using gel filtration.</b>	<b>77</b>
<b>Figure 3.1.3.1.1</b>	<b>Cryo-EM micrographs collected for <i>NpSRII/NpHtrII</i>.</b>	<b>79</b>
<b>Figure 3.1.3.1.2</b>	<b>2D classifications obtained from the cryo-EM analysis of <i>NpSrII/NpHtrII</i> solubilized in DDM.</b>	<b>79</b>
<b>Figure 3.1.3.1.3</b>	<b>Cryo-EM analysis of <i>NpSrII/NpHtrII</i> solubilized in DDM.</b>	<b>80</b>
<b>Figure 3.1.4.1</b>	<b>Diamond and hexagonal shaped crystals obtained during the crystallisation trials for <i>NpSrII/NpHtrII</i> solubilized in DDM.</b>	<b>81</b>
<b>Figure 3.1.4.2</b>	<b>The crystal structure of the diamond-shaped crystals obtained.</b>	<b>81</b>
<b>Figure 3.1.4.3</b>	<b>The crystal structure analysis of the hexagonal plates shown in Figure 3.1.4.1.</b>	<b>82</b>
<b>Figure 3.1.6.1</b>	<b>SAS analysis of <i>NpSRII/NpHtrII<sub>137</sub></i> and of full-length <i>NpSRII/NpHtrII</i> at low salt concentration.</b>	<b>85</b>
<b>Figure 3.1.6.2</b>	<b>SANS scattering curve for <i>NpSRII/NpHtrII</i> at 4.0 M NaCl and corresponding distance distribution function.</b>	<b>86</b>

<b>Figure 3.1.6.3 The two different forms of full-length <i>NpSRII/NpHtrII</i> trimers of dimers constructed for use in fits of experimental and theoretical SANS scattering curves.....</b>	<b>86</b>
<b>Figure 3.2.1.1 Crystallization of <i>NpSRII/NpHtrII</i> complex in MSP1E3D1/DPPC nanodiscs using the in meso method.....</b>	<b>88</b>
<b>Figure 3.2.1.2 Diffraction pattern from a crystal of <i>NpSRII/NpHtrII</i> taken from crystallization well H6 (1.6M ammonium phosphate pH 8.8). ....</b>	<b>89</b>
<b>Figure 3.2.1.3 Micrographs of crystals obtained in wells A2 and H7. ....</b>	<b>89</b>
<b>Figure 3.2.2.1 SAXS analysis from solutions of <i>NpSRII/NpHtrII</i> embedded in nanodiscs. ....</b>	<b>90</b>
<b>Figure 3.2.2.2 The SAXS-obtained molecular envelope of <i>NpSRII/NpHtrII</i> embedded in nanodiscs. ....</b>	<b>91</b>
<b>Figure 3.2.2.3 SANS scattering curves obtained for <i>NpSRII/NpHtrII</i> in nanodiscs at three different heavy water concentrations.....</b>	<b>91</b>
<b>Figure 3.2.2.4 SAXS analysis, at increasing sucrose concentrations, of <i>NpSRII/NpHtrII</i> embedded in nanodiscs. ....</b>	<b>92</b>
<b>Figure 3.2.3.1 Cryo-electron micrograph for <i>NpSRII/NpHtrII</i> reconstituted in nanodiscs. ....</b>	<b>93</b>
<b>Figure 3.2.3.2 Classes obtained using 2D classification procedures for particles isolated in manual processing mode. ....</b>	<b>94</b>
<b>Figure 3.2.3.3 2D class averages obtained following automatic particle picking based on the template highlighted in Figure 3.1.3.2. ....</b>	<b>95</b>
<b>Figure 3.2.3.4 Three images of a cryo-EM derived three-dimensional de novo model of the <i>NpSRII/NpHtrII</i> complex reconstructed in nanodiscs. ....</b>	<b>95</b>
<b>Figure 3.2.3.5 Gel filtration analysis of three <i>NpSRII/NpHtrII</i> samples.....</b>	<b>96</b>
<b>Figure 3.2.3.6 Micrographs obtained using the negative contrast method for the full-size <i>NpSRII/NpHtrII</i> protein complex embedded in MSP1E1 lipid nanodiscs. ....</b>	<b>97</b>
<b>Figure 3.3.1 Overall cartoon views of the structures of derivatised proteins: tmBR-argon, tmBR-krypton, KR2-krypton, MAR-krypton.....</b>	<b>101</b>

<b>Figure 3.3.2 Illustrations of how noble gas atoms are arranged on the surfaces of the derivatized structures of tmBR-argon, tmBR-krypton, KR2-krypton, MAR-krypton. ....</b>	<b>103</b>
<b>Figure 3.3.3 Cartoon representation of the crystal structure of KR2 derivatised with krypton.....</b>	<b>104</b>
<b>Figure 3.3.2 Arrangement of peaks of the MD maps (sampled at <math>3.5 \times</math> r.m.s. level, navy blue chicken wire) on the surface of the proteins of tmBR-argon, tmBR-krypton, KR2-krypton, MAR-krypton.....</b>	<b>104</b>
<b>Figure 3.3.3.1 Schematic illustration of the positions of the internal noble gas atoms in the structures of tmBR (a, Ar; b, Kr) and MAR (c, Kr) and their environments.....</b>	<b>108</b>

## List of Tables

<b>Table 1.6.1 The van der Waals radii of some elements .....</b>	<b>47</b>
<b>Table 2.1.3.1 Procedure for the "Calcium" protocol for preparing competent cells.....</b>	<b>56</b>
<b>Table 2.1.7.1 Nanodisc – lipid ratios required to fill the full surface of the nanodisc.....</b>	<b>59</b>
<b>Table 3.1.2.1 OD<sub>600</sub> measurement as function of post-induction time .....</b>	<b>74</b>
<b>Table 3.1.3.1.1 A summary of the conditions used in grid preparation for cryo-EM studies of wild-type <i>NpSRII/NpHtrII</i>.....</b>	<b>78</b>
<b>Table 3.1.4.1 Data collection and refinement statistics for the crystal structure analysis of the Diamond- and hexagonal-shaped crystals shown in Figure 3.1.4.1. ....</b>	<b>82</b>
<b>Table 3.2.2.1 The theoretical values of the volume and number of electrons for the <i>NpSRII/NpHtrII</i> complex, MSP1E3D1, DMPC, and a <i>NpSRII/NpHtrII</i> trimer of dimers reconstructed in nanodiscs.....</b>	<b>92</b>
<b>Table 3.3.1 Data collection statistics for noble-gas derivatized membrane protein crystals. ....</b>	<b>99</b>
<b>Table 3.3.2 Refinement statistics for noble-gas derivatized membrane protein crystal structures described in this thesis. ....</b>	<b>100</b>
<b>Table 3.3.3 Statistics of the MD experiment and comparison with noble gas atom positions observed in the MX models described above.....</b>	<b>105</b>

## Abstract

In the study of protein structure, function and mechanism, transmembrane and membrane proteins (MPs) are of particular interest. MPs constitute 25% of all known proteins and are involved in a host of fundamental biological processes including ion transport and signal transduction across the membrane. Moreover, MPs are targeted by more than 60% of existing drugs. However, in contrast to water-soluble proteins, they are much less well understood. The primary reason for this is their amphiphilic character which leads to complications with their isolation from the membrane and subsequent purification. Indeed, for the study of MPs, one needs to extract them from their native lipid environments into more tractable membrane-mimicking systems. Such difficulties, together with the importance of MPs, make the corresponding research an exciting but challenging problem.

Two-component signaling systems (TCS) enable microorganisms to communicate with the environment, are present in all domains of life and are the most abundant signaling systems in Nature. TCS consist of a signal receptor/transducer and a response regulator. The former are usually transmembrane receptors (histidine kinases, chemoreceptors, photoreceptors) which have a similar, modular structure. Most TCS receptors/transducers function as higher order oligomers (trimers of dimers in the case of chemoreceptors). Moreover, the oligomeric state adopted is often dependent on external conditions, which makes structural study of such receptors a difficult task, particularly for full-length systems. The first part of the work described here focuses on structural studies of the full-length *NpSRII/NpHtrII* receptor/transducer TCS. Crystallization of the full-length system resulted in either very small, poorly diffracting crystals or samples in which the transducer region has undergone proteolysis. Solution scattering (SAXS/SANS) studies described here confirm that the oligomerization and folding of full-length *NpSRII/NpHtrII* strongly depends on ionic strength of its surrounding solution: at low salt concentration (150 mM NaCl) the complex forms dimers when solubilized in detergent; in high salt buffers (4.0 M NaCl), corresponding to native conditions,

these dimers associate to form trimers of dimers. The SAXS/SANS experiments carried out also confirm that full-length *NpSRII/NpHtrII* is highly conformationally dynamic, a fact which might explain the failure, despite promising initial conditions, for the electron microscopy analysis reported here to produce anything but a very low resolution reconstruction of *NpSRII/NpHtrII*.

While functional dependence of MPs on their surrounding lipids has been well established experimentally, the mechanisms by which lipids modulate MP structure and function are still poorly understood. To shed light on these, the second part of this thesis work examines, using inert gases as models, how small hydrophobic moieties interact with MPs at the molecular level. High pressure techniques were used to produce argon and krypton derivatives of crystals of three well studied MPs (two different proton pumps and a light-driven sodium ion pump). The crystal structures obtained show that a vast majority of argon and krypton binding sites were located on the outer hydrophobic surface of the MPs – a surface buried in the membrane and which usually accommodates the hydrophobic chains of annular lipids. Supplementary analysis by *in silico* molecular dynamics (MD) carried out here shows an even greater number of potential argon and krypton interaction sites on MP surface within the lipid bilayer. These results suggest that MPs are stabilised in an optimal functional conformation by the specific binding of lipids that energetically best fit the grooves on their hydrophobic surfaces. A concept of a general mechanism of allosteric regulation of MP function by lipids and its alteration by other hydrophobic molecules is thus proposed.

## Résumé

Dans l'étude de la structure, la fonction et le mécanisme des protéines, les protéines transmembranaires et membranaires (MP) présentent un intérêt particulier. Les MPs constituent 25% de toutes les protéines connues, sont impliquées dans une série de processus biologiques fondamentaux comme le transport d'ions et la transduction de signaux à travers la membrane. De plus, les MPs sont ciblés par plus de 60 % des médicaments existants. Cependant, contrairement aux protéines hydrosolubles, elles sont beaucoup moins bien comprises. La raison principale en est que leur caractère amphiphile entraîne des complications lors de leur isolement de la membrane et de leur purification ultérieure. En effet, pour l'étude des MPs, il faut les extraire de leur environnement lipidique natif pour les transformer en un système plus tractable imitant la membrane. De telles difficultés, ainsi que leur importance rendent l'étude des MPs passionnante mais compliquée.

Les systèmes de signalisation à deux composantes (TCS) permettent aux microorganismes de communiquer avec l'environnement, sont présents dans tous les domaines de la vie et sont les systèmes de signalisation les plus abondants dans la nature. Les TCS se composent d'un récepteur/transducteur de signal et d'un régulateur de réponse. Les premiers sont généralement des récepteurs transmembranaires (histidine kinases, chimiorécepteurs, photorécepteurs) qui ont une structure modulaire similaire. La plupart des récepteurs/transducteurs du TCS fonctionnent comme des oligomères d'ordre supérieur (trimères de dimères dans le cas des chimiorécepteurs). En outre, l'état oligomérique adopté dépend souvent de conditions externes, ce qui rend l'étude structurale de ces récepteurs difficile, en particulier pour les systèmes de grande longueur.

La première partie des travaux démontrés ici est focalisée sur les études structurales du TCS *NpSRII/NpHtrII*. Dans un premier temps, la cristallisation n'a produit que des cristaux de mauvaise qualité ou des échantillons dans lesquels le transducteur de signal a été protéolysé. Les expériences exploitant la diffusion des solutions (SAXS/SANS) ont montré que l'oligomérisation et le pliage de l'ensemble



*NpSRII/NpHtrII* complet dépendent fortement de la force ionique de la solution. À faible concentration en sel (150 mM NaCl), lorsqu'il est solubilisé dans un détergent, l'ensemble forme des dimères. Dans les tampons à forte teneur en sel (4.0 M NaCl), correspondant aux conditions natives, ces dimères s'associent pour former des trimères de dimères, comme c'est le cas pour les chimiorécepteurs. Les expériences SAXS/SANS ont confirmé que le complexe *NpSRII/NpHtrII* est hautement dynamique sur le plan de la conformation. Ce qui peut expliquer que, malgré les bonnes conditions initiales pour mener les expériences, les études cryo-EM ont produit une image qu'à très basse résolution.

Alors que la dépendance fonctionnelle des MPs par rapport aux lipides environnants a été bien établie expérimentalement, les mécanismes par lesquels les lipides modulent la structure et la fonction des MPs sont toujours mal compris. Pour les éclairer, ici il a été examiné comment de petites molécules hydrophobes comme les gaz inertes interagissent avec les MPs au niveau moléculaire. L'argon et du krypton ont été utilisés pour dériver à haute pression des cristaux de trois MPs bien étudiés (deux pompes à protons différentes et une pompe à sodium commandée par la lumière). Les structures obtenues par cristallographie aux rayons X ont montré qu'une grande majorité des sites de liaison de l'argon et du krypton sont situés sur la surface hydrophobe externe des MPs - une surface qui accueille généralement dans ses sillons des chaînes hydrophobes de lipides annulaires. Une analyse supplémentaire par la dynamique moléculaire *in silico* (MD) a montré un nombre encore plus important de positions de l'argon et du krypton sur la surface des MP à l'intérieur de la bicouche, en accord avec les données cristallographiques. Ces résultats suggèrent qu'un MP est stabilisé dans une conformation fonctionnelle optimale par la liaison spécifique des lipides qui s'adaptent le mieux énergétiquement aux rainures de sa surface hydrophobe. Un concept de mécanisme général de régulation allostérique de la fonction de la MP par les lipides et son altération par d'autres molécules hydrophobes est proposé.

## 1 Introduction

In the study protein structure, function and mechanism, transmembrane and membrane proteins (MPs) are of particular interest. MPs constitute 25% of all known proteins, are involved in a host of fundamental biological processes as ion transport and signal transduction across the membrane, enzymatic activity, etc., and are targeted by more than 60% of existing drugs. However, in contrast to water-soluble proteins, they are much less well understood. The primary reason for this is that their amphiphilic character leads to complications with their isolation from the membrane and their subsequent purification. Indeed, for the study of MPs, one needs to extract them from their native lipid environments into more tractable membrane-mimicking systems. Such difficulties, together with the importance of MPs, make the corresponding research an exciting but challenging problem.

Two-component systems (TCS) are the most common protein signaling systems in Nature, providing a microorganism's communication with its environment [1]. The first components of TCS – receptors /transducers - are generally membrane proteins. The second component is a response regulator that transmits information to gene regulation machineries. Three main types of membrane receptors (sensors), all having a similar modular structure are found in TCS: histidine kinases, chemoreceptors, and photoreceptors. Given their absence in mammals, TCS are potential targets for antimicrobial drugs. To provide a rational platform for the development of such drugs, and to fully understand the molecular mechanisms of signal transduction by TCS receptors through the membrane, it is necessary to have information on these proteins' atomic structures.

Despite the importance of TCSs and the scientific community's broad interest in their study, only high-resolution structure of fragments of these receptors/transducers have been obtained and described in the literature. This is because full-size TCS sensors are dynamic, which complicates their crystallization and subsequent study using X-ray crystallography [1]. There is thus a very high

interest in studying how the dynamics (that is, ordering at the secondary structure and polydispersity in the mutual arrangement of domains) of TCS, (and, indeed, many other biological systems) depends on external conditions. Part of the thesis work described here was therefore targeted at structural studies of a full-length TCS with the aim of elucidating further the dynamic processes involved in transmembrane signal transduction.

Another important problem relating to MPs is how their hydrophobic environment, and changes in this due to external conditions, influence their structure and function. Indeed, a very well-known field of MP research is the study of lipid-membrane protein interactions. It is also known that small hydrophobic molecules may influence the properties of MPs. Such molecules can be dissolved in hydrophobic region of the bilayer and, diffusing through this, may interact with a MP by binding to the pockets on its hydrophobic surface, competing with lipids. Such interaction may have a direct impact on protein stability and function. However, the details of such interactions are currently poorly studied.

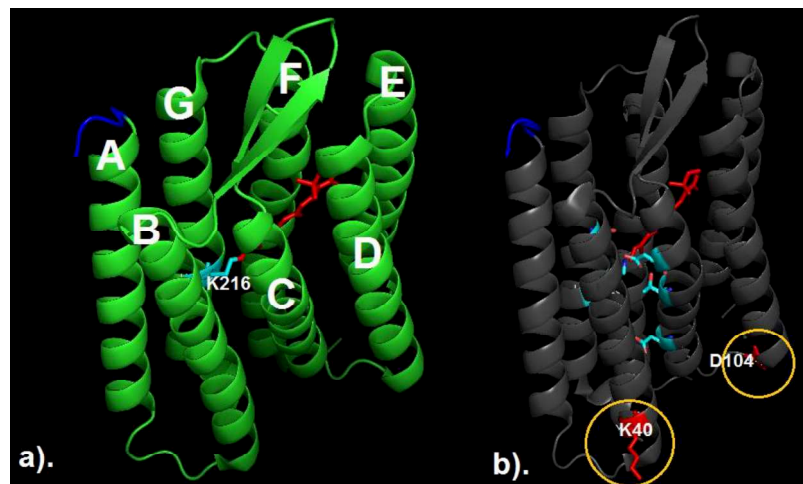
Under a pressurized atmosphere, noble gas atoms can quickly diffuse to potential interaction sites in proteins through channels within protein molecules. Here, the number and occupancy of binding sites depends on the applied pressure and the interaction, entirely reversible, of noble gas atoms with proteins results from non-covalent van der Waals forces. This led to an opinion that noble gas binding causes only minor disturbances to the surrounding molecular structures. Consequently, xenon and krypton derivatives of protein crystals were considered isostructural their native counterparts and, in X-ray crystallography, have mostly been exploited in experiments aimed at the experimental determination of phase information.

Nevertheless, the experiments alluded to above showed that Xe and Kr bind to a wide variety of sites in proteins, including closed intramolecular hydrophobic cavities, accessible hydrophobic binding sites, intermolecular cavities, channel pores, and accessible cavities on the surface of the molecule. Additionally, noble

gases - particularly Xe - have anesthetic properties, suggesting their use in this regard modifies the functionality of membrane proteins involved in the sensing of pain. Taken together, the above observations suggest that an excellent model for the influence of small molecules on the structure and function of MPs is their interactions with noble gases. With the aim of shedding light on how small hydrophobic molecules dissolved in lipid bilayers might interact with, and modify, the structures of MPs and to investigate how noble gases act as anesthetics, in this work derivatives of MPs were prepared by exposing crystals of native MPs to pressurized atmospheres of three noble gases. The derivative crystals thus obtained were analyzed using X-ray crystallography and the results are described here.

## 1.1 Microbial Rhodopsins

Microbial Rhodopsins (MRs) are light-sensitive membrane proteins that perform diverse functions as the primary solar energy harvesters on Earth [2]. The three-dimensional structure of MRs (Figure 1.1.1) comprises seven TM  $\alpha$ -helices (named A to G) linked by three extracellular and three intracellular loops. The light-sensitivity of the proteins is provided by a retinal cofactor, covalently bound to a conserved lysine residue in helix G via a Schiff base.



**Figure 1.1.1** The structure of the archetypal microbial rhodopsin bR. *a)* the 7TM  $\alpha$ -helices (in green) labelled A to G starting from extracellular N-terminus (depicted in blue), the conserved Lys216 is depicted in light blue, and the all-trans-retinal is shown in red. *b)* Lys40 and Asp104 (red) emphasized with yellow circles

The first known MR, the light-driven proton pump bacteriorhodopsin (bR, Figure 1.1.1), was discovered in the archaea *Halobacterium Salinarum* [3] and has since become the classical object for the studies of all other MRs and, indeed, membrane proteins in general. Structural studies of bR showed that water molecules and cavities occupied by them, play a crucial role in the function of MRs [4, 5]. This observation is supported by recent advances in the structural characterization of novel families of MRs, such as bacterial proton-pumping proteorhodopsins [6], cation and anion channelrhodopsins [7-11], bacterial ion pumps [12-14], viral rhodopsins [15-17] and heliorhodopsins [18-20].

Photochemical reactions, excited by the photoisomerization of the retinal chromophore, lead to different processes in different MRs. However, the biological functions of MRs generally fall into two categories:

1). Photo energy converters which convert light into an electrochemical potential to drive cells (i.e. light-driven ion pumps).

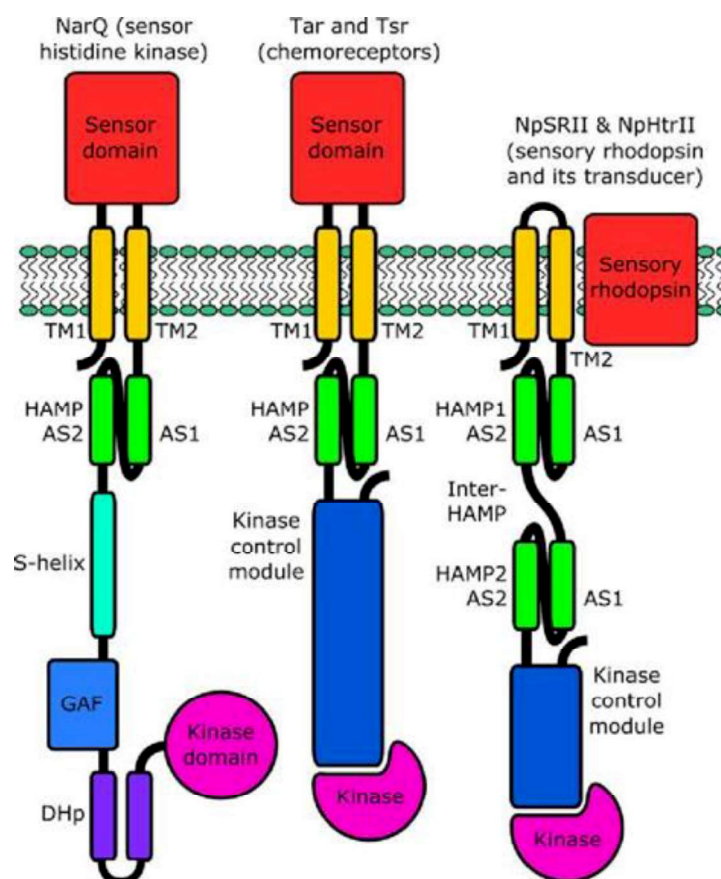
2). Photosensory receptors, which use light to sense the environment and trigger the regulation of cellular processes. Known methods for transmitting signals by sensory MRs are protein-protein interactions with membrane-embedded transducers, interaction of their cytoplasmic domains with soluble transducer proteins, enzymatic activity encoded in their cytoplasmic domain, and signal transmission by conduction through passive ion channels controlled by light [21-25].

Examples of MRs which operate as ion pumps include the KR2 sodium pump [26]. In KR the chromophore is linked to the residue K255, while the bR proton acceptor and donor aspartate residues (D96 and D85, respectively) are replaced by polar uncharged amino acids (N112 and Q123 in the sequence of KR2). In addition to this, T89, which is not strongly involved in proton transport in bR, is replaced by D116.

Photosensory systems include sensory rhodopsin II (SRII) [27] and sensory rhodopsin I (SRI) [28], which allow archaeobacteria to seek optimal lighting conditions for their energy needs while avoiding photooxidative damage. Both photoreceptors trigger a signaling chain that is closely related to that of the two-component eubacterial system. One significant, but not yet fully understood, problem concerns the molecular mechanism of signal transmission from receptors to cytoplasmic signaling components.

## 1.2 Two-component protein signaling systems

Two-component signaling systems (TCS) are represented across all domains representing the diversity of living beings. TCS are the most common signal transduction machineries in Nature. However, although recent work [29] has shown that TCS expressed in mammalian cells can mediate orthogonal signaling pathways, TCS are naturally absent in mammals, making them potential targets for new antimicrobial drugs [30]. However, both the design of potential new antibiotics and a full understanding of TCS signal transduction mechanisms will require detailed knowledge of the atomic-level structures of full-length TCS in order to understand the relationship between their structure and function.

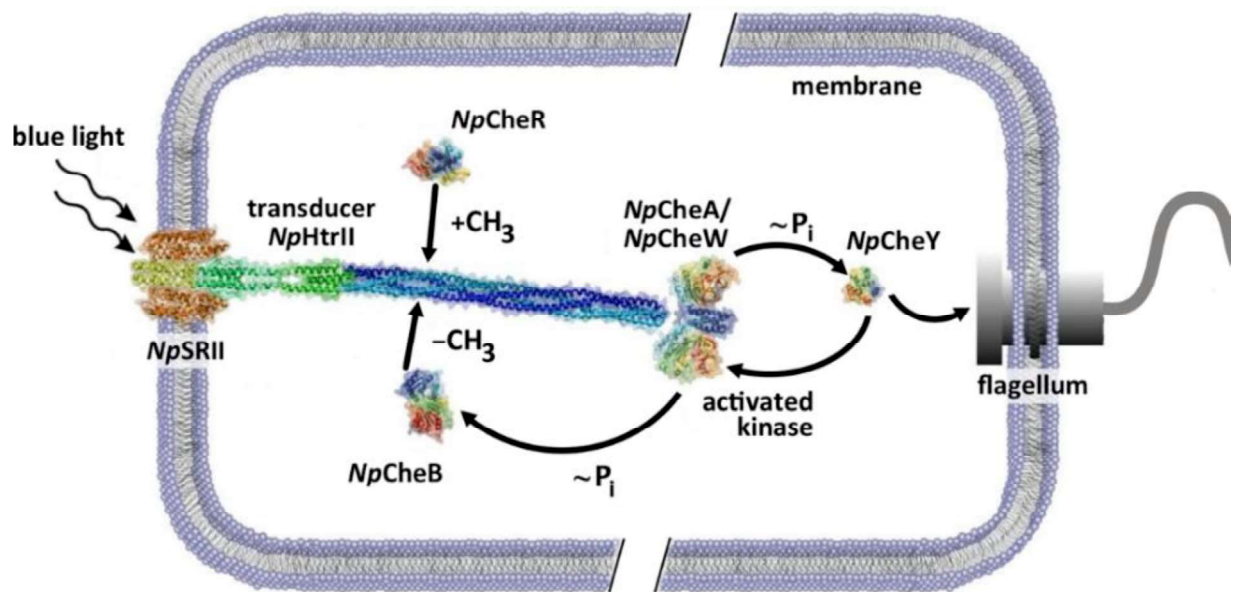


**Figure 1.2.1 Domain architectures of representative microbial TCS.** *TM1 and TM2 are transmembrane helices. AS1 and AS2 are HAMP domain helices. NarQ is E. coli nitrate and nitrite sensor, Tsr is E. coli serine chemoreceptor, NpSRII and NpHtrII are Natronomonas pharaonis sensory rhodopsin II and sensory rhodopsin transducer II. In chemo- and photoreceptor systems, the kinase is a different protein. While only monomers are shown, all of the proteins usually are dimeric with chemo- and photoreceptor dimers forming higher-order oligomeric assemblies. Figure adapted from [1].*

To date, among the most studied TCS have been the chemotaxis and photoreceptors of the TCS of motile bacteria and archaea, respectively (Figure 1.2.1)

[22, 31]. Chemotaxis receptors form homodimers consisting of an extracellular sensory domain, a transmembrane domain, and a long rod-like cytoplasmic domain. For photoreceptors, the sensory domain is a separate protein, a microbial rhodopsin, which forms a 2:2 complex with its related transducer.

The separate sensory domain excepted, photoreceptor transducers have a high homology level with chemoreceptors and a similar modular domain architecture [22, 28] (Figure 1.2.1). This domain architecture comprises a transmembrane (TM) region, two HAMP domains (connected by a helical inter- HAMP [32] region), and an adaptation module containing reversible methylation sites and glycine (Gly) flexible hinges [33]. The adjacent signaling region contains binding sites for the CheW adapter protein and CheA histidine kinase. Interestingly, bacterial chemoreceptors have only one HAMP domain [34], while archaeal transducers can include two HAMP domains.



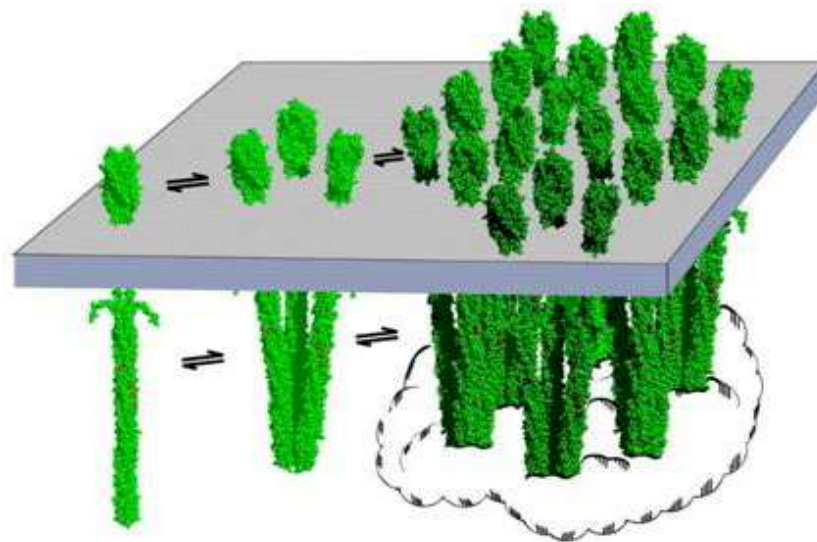
**Figure 1.2.2** Schematic of the signaling cascade in the two-component negative phototaxis system *NpSRII/NpHtrII* in *N. pharaonis*. Figure taken from [35].

In the *NpSRII/NpHtrII* photoreceptor of the haloalkaliphilic archaeon *Natronomonas pharaonis* [35], the subject of majority of the work described in this thesis manuscript, blue light stimulation of sensory rhodopsin II (*NpSRII*) induces structural and/or dynamic changes in the transducer (*NpHtrII*). The signal is then converted by two HAMP domains and transmitted via a 200 Å long cytoplasmic



kinase module to the outermost cytoplasmic region of *Np*HtrII. The transducer-activated histidine kinase CheA (linked to the adapter protein CheW) undergoes autophosphorylation and transfers the phosphate group to the CheY or CheB response regulators. CheY affects the displacement of the flagellum's rotation, while the metilesteraza CheB and the methyltransferase CheR control adaptation mechanisms (Figure 1.2.2).

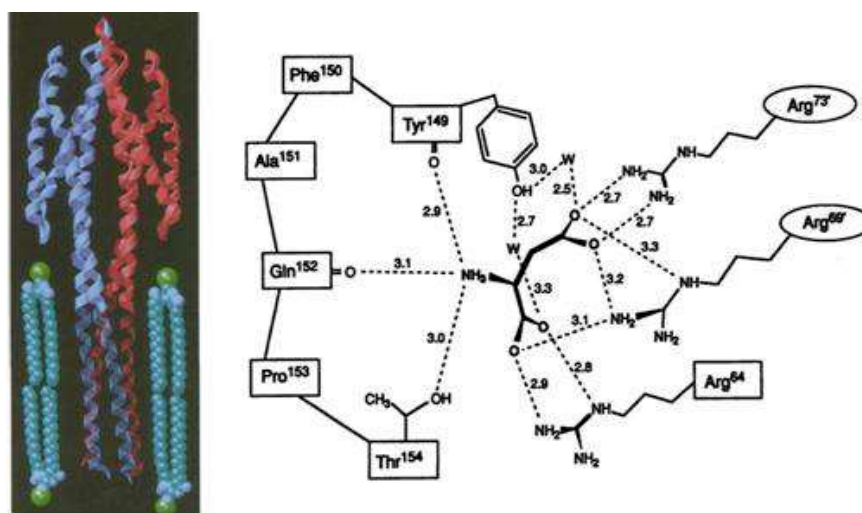
Despite an abundance of biochemical data, the molecular mechanism of a TCS such as *Np*SRII/*Np*HtrII has yet to be fully disclosed, although a general scheme assumes sequential dynamic changes in the cytoplasmic domains. Indeed, both the chemoreceptors and the transducers of sensory rhodopsins show different dynamics in adjacent modules, which correlates with a signal transmission along the cytoplasmic "core" [35, 36]. Homodimers of chemoreceptors (or rhodopsin-transducer complexes) in the cell membrane form trimers, functional units able to activate CheA [37]. Together with CheA and CheW, these dimers of trimers coalesce to form extended two-dimensional signaling arrays (Figure 1.2.3) [38], compact membrane super-complexes responsible for amplifying the incoming signal [39-41].



**Figure 1.2.3 Different oligomeric states of chemoreceptor TCS.** A schematic showing a chemoreceptor homodimer (left), a receptor trimer of dimers (center), and a two-dimensional membrane-associated signaling array consisting of dimers of trimers associated with CheA, and CheW (right). The cloud represents CheA and CheW. Figure taken from [38].

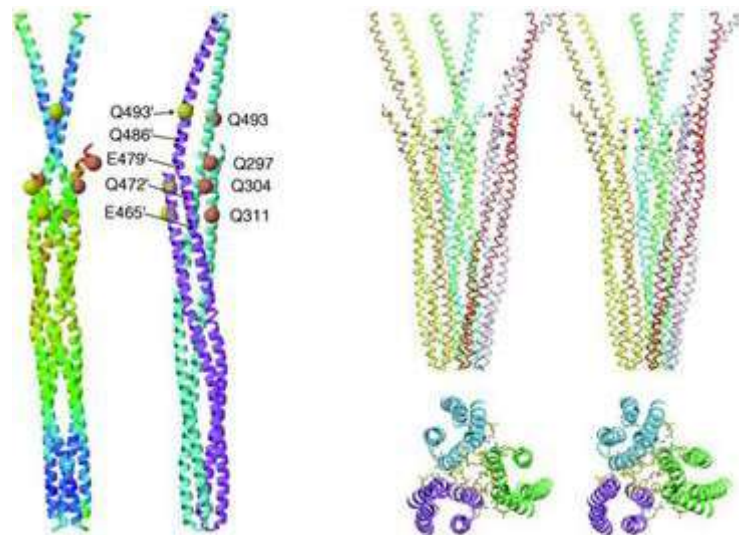
An essential condition for a complete understanding of the signal transduction mechanism of *NpSRII/NpHtrII* is high-resolution structural information for this full-length TCS receptor in its functional state (i.e. trimers of dimers). However, three-dimensional X-ray structures at close to atomic resolution have been reported only for fragments of chemoreceptors [42-44] while, for *NpSRII/NpHtrII* itself, structural data are available only for a truncated *NpSRII/NpHtrII*<sub>114</sub> complex [21] (see §1.3).

In the work reported in [42], the crystal structures of the ligand-binding domain of the aspartate receptor from *Salmonella typhimurium* in both its *apo*- form and aspartate-bound forms were obtained at resolutions of 2.4 Å and 2.0 Å, respectively. In the case of the *apo*-receptor (Figure 1.2.4), the two subunits in the dimer are related by a two-fold crystallographic axis. In the substrate-bound complex, aspartate binds to two helices from one subunit and one coil from the other in a positively charged pocket formed by three arginine residues. The substrate-bound dimer is thus not entirely symmetrical. As also reported, the ligand binding site is located more than 60 Å from the putative membrane surface and is situated at the interface between the two subunits making up the dimer.



**Figure 1.2.4** The crystal structure of the sensory domain dimer of the aspartate receptor from *S. typhimurium* combined with a simulated transmembrane region (left) and a diagram of hydrogen bonds formed upon the binding of aspartate (right). Lipid molecules are located according to the putative hydrophobic boundaries of the protein. The dimer has a length of more than 120 Å and a diameter of ~ 40 Å. In the panel on the right, amino acids from one subunit are shown in rectangles, those from the other subunit in ellipses. Water molecules are marked "W"; putative hydrogen bonds are shown as dotted lines with corresponding distances. Figure taken from [42].

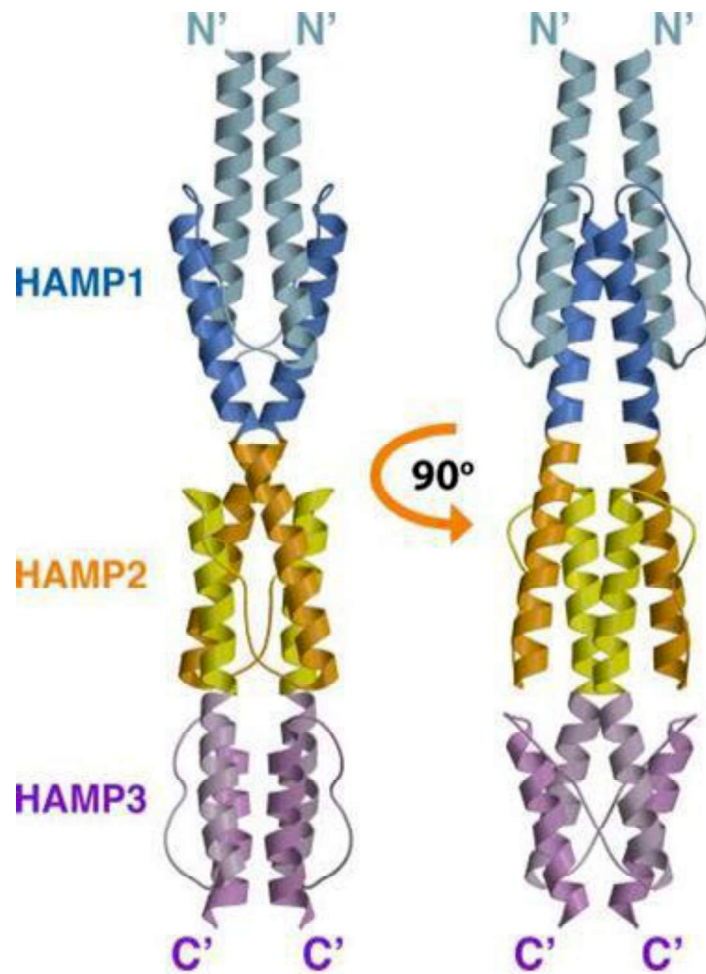
The crystal structure of the cytoplasmic domain of the serine chemoreceptor from *Escherichia coli* (EcTsrQ) is reported in [43] (Figure 1.2.5). Here, the two subunits making up the cytoplasmic domain dimer associate to form a long, supercoiled bundle about 200 Å in length. As also reported for the aspartate receptor from *S. typhimurium* [17], in the crystal the dimers associate to form trimers of dimers.



**Figure 1.2.5** Schematic representations of atomic models obtained by X-ray crystallography of the dimer and trimer of dimers of the EcTsrQ cytoplasmic domain. Left - two images of the cytoplasmic domain dimer with the image on the right showing the two subunits in purple and blue, respectively. Right - The EcTsrQ trimer of dimers formed by crystal packing. In both panels methylation sites are shown as solid spheres. Figure adapted from [43].

A structural study of the HAMP domains of the soluble N-terminal fragment of the Aer2 receptor from *Pseudomonas aeruginosa* has also been reported [44]. HAMP (found in Histidine kinases, Adenylyl cyclases, Methyl-accepting chemoreceptors, and Phosphatases) domains are widespread prokaryotic signaling modules and are found as either separate domains or chains (poly-HAMPs) in transmembrane and soluble proteins (see Figure 1.2.1). The three HAMP domains (3-element poly-HAMP) seen in the crystal structure reported in [44] illustrate the universality of their structure (Figure 1.2.6): they are formed from four  $\alpha$ -helices, two (AS1 and AS2) from each receptor subunit. In the crystal structure the three HAMP domains exhibit two different conformations, which differ in helix displacements, intersection, and rotation angles. These conformations are stabilized

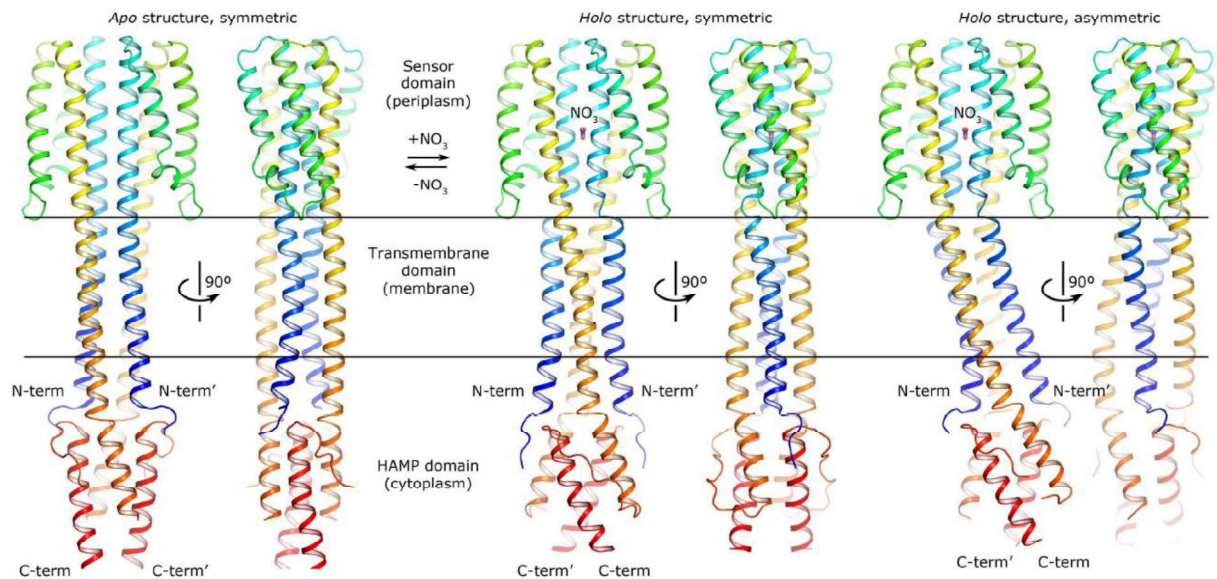
by various subsets of conserved amino acid residues. The authors assumed that the two conformations of the poly-HAMP chain described in [44] are representative of different states adapted during signal transduction.



**Figure 1.2.6** The crystal structure of the 3-element poly-HAMP region of the Aer2 receptor. The fragment contains three sequential and intertwined HAMP domains. The AS2 helix of the HAMP1 domain (blue) is adjacent to the AS1 helix of the HAMP2 domain (orange), and the helix AS2 of the HAMP2 domain (yellow) is adjacent to the AS1 helix of the HAMP3 domain (light purple). Figure taken from [44].

A significant advance in structural studies of high-resolution TCS membrane sensors was achieved in 2017 with a report describing the crystal structures of the histidine kinase NarQ (nitrite/nitrate sensor from *E. coli*) in both *apo*- and *holo*- (i.e. substrate-bound) forms [45]. The NarQ fragment used in this work comprised three domains: the sensory domain located on the outside of the cell membrane, the transmembrane domain, and the HAMP domain that follows it. The *holo*-NarQ homodimer exhibited two different conformations - symmetric and asymmetric (Figure 1.2.7).



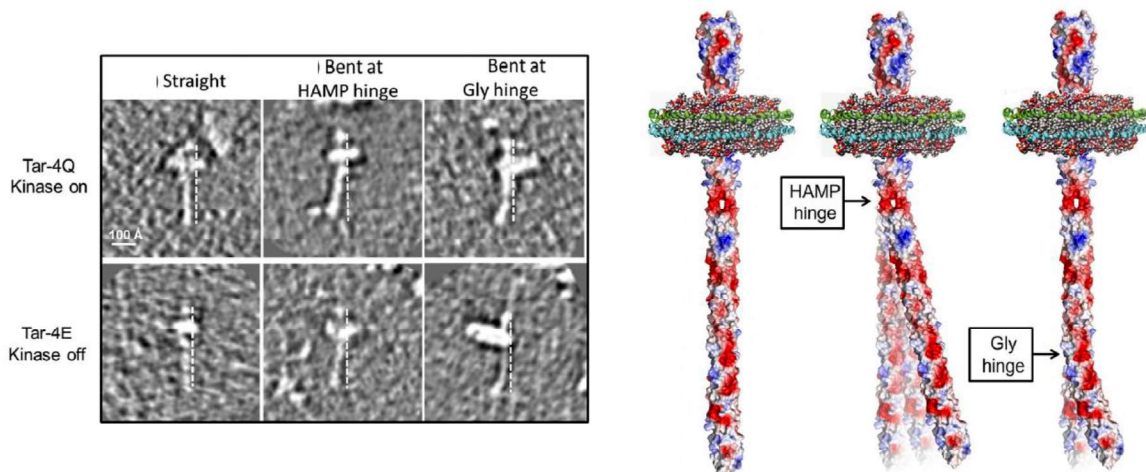


**Figure 1.2.7** The crystal structures of a NarQ fragment in *apo*- and *holo*- states. The protein is observed as an asymmetric dimer in its *apo*- (left) and *holo*- (middle) states in space groups  $I2_12_12_1$  and  $F222$ , respectively and as an asymmetric dimer in the *holo*- state in the space group  $P2$  (right). Figure from [45].

Although the authors of [45] achieved significant progress in structural studies of TCS sensors, the results reported are still far from providing high-resolution structural details of a full-length TCS sensor protein. Indeed, in addition to the three regions included in the crystal structures the authors described, the cytoplasmic region of NarQ also includes the following domains: S-helix, GAF-domain (the name is derived from various proteins in which it is found: GMP-specific phosphodiesterases, adenylate cyclases (Adenylyl cyclases) and FHLA); N-terminal dimerization and phosphotransfer domain (N-terminal dimerization and histidine phosphotransfer (DHp) domain); and a kinase domain (see Figure 1.2.1).

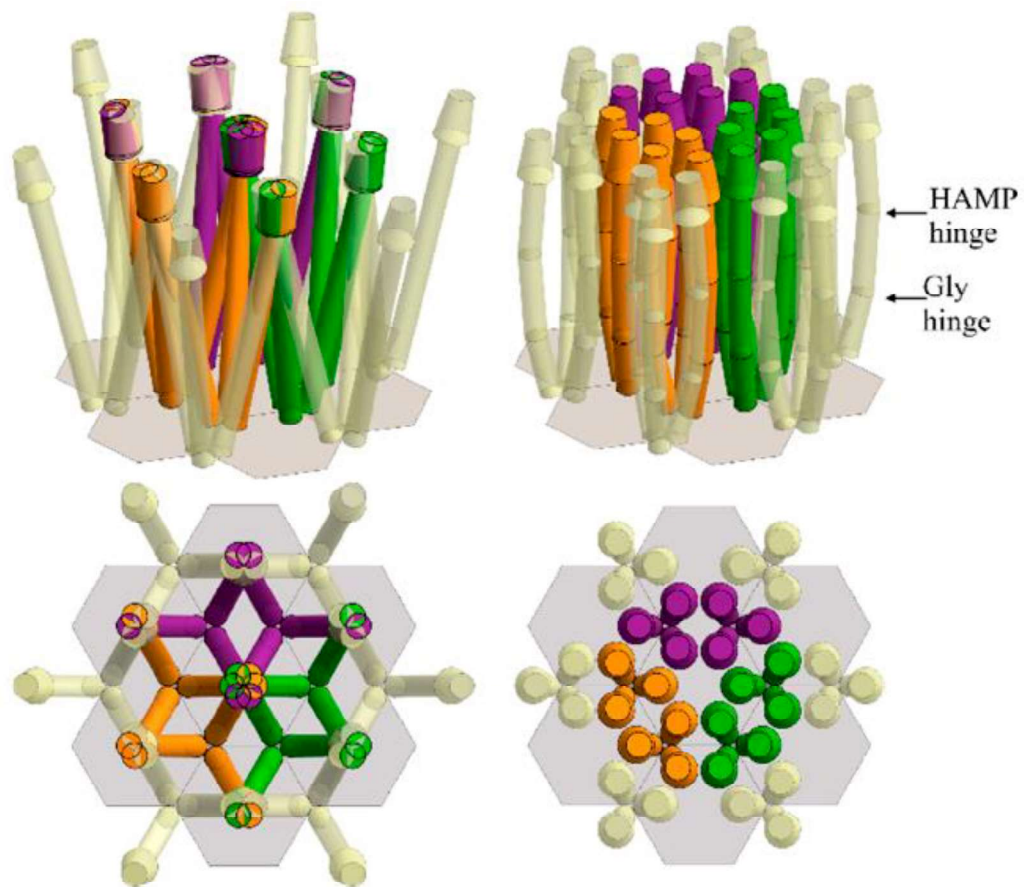
In addition to high-resolution structures of fragments of two-component systems obtained by X-ray diffraction, the literature describes several atomistic models of chemoreceptors from *E. coli*, which were constructed based on low-resolution data obtained using electron microscopy. These include a homodimer of the aspartate receptor Tar [33], trimers of dimers of the cytoplasmic domain of serine chemoreceptors in different signaling states [46], and the central signaling block of the chemoreceptor 2D array formed by a mixture of receptors with various adaptive modifications [47].

The study of the aspartate receptor Tar reported in [33] used contrast methods, transmission electron tomography, and image analysis to characterize its homodimeric state. The chemoreceptor's functionality was by ensured reconstituting the Tar dimers in nanodiscs prepared using lipids isolated from *E. coli* membranes. Single-particle analysis and tomography of particles in a three-dimensional matrix helped to identify two bending regions in the cytoplasmic domain of the chemoreceptor (Figure 1.2.8): a short two-strand break of  $\alpha$ -helices between the area of the HAMP-domain closest to the membrane and transmembrane helices (HAMP-hinge); aligned glycines in an extended four- $\alpha$ -helical super-helix (Gly-hinge). Electron microscopy data show angles of the HAMP-hinge bending ranging from  $0^\circ$  to  $13^\circ$ , and bending of the Gly-hinge varying from  $0^\circ$  to  $20^\circ$ .



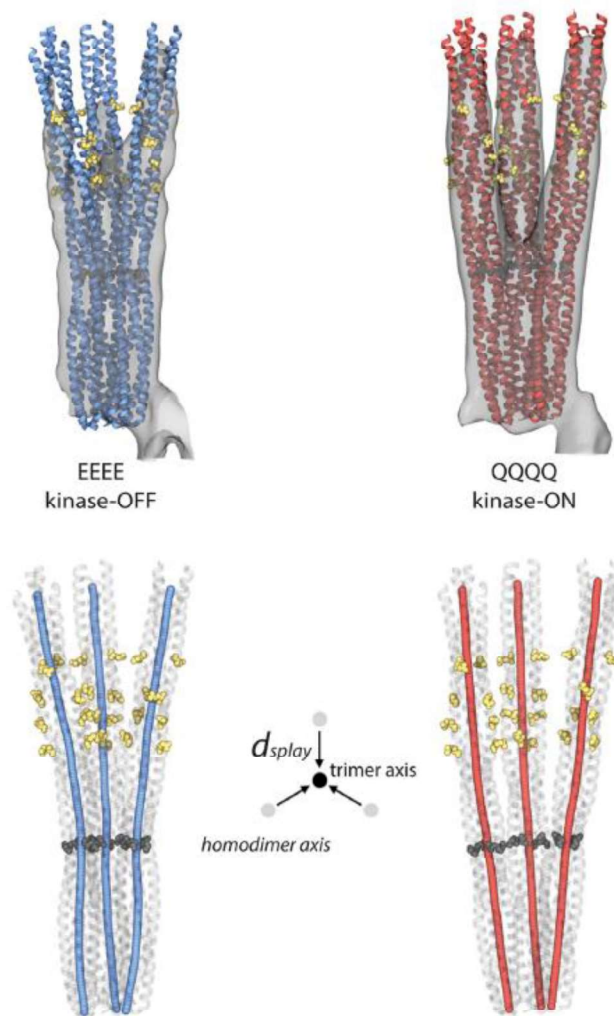
**Figure 1.2.8 Electron tomography studies of Tar dimers reconstituted in nanodiscs.** Right: Images obtained from an electron tomography analysis of Tar dimers reconstituted in nanodiscs suspended in a three-dimensional trehalose layer and stained with ammonium molybdate. Left: Models of the reconstructions of the straight and bent Tar chemoreceptor dimers obtained. Figure adapted from [33].

The authors of the work described in [33] assumed that bends in the flexible hinge region do not correlate with chemoreceptor signaling states. However, they did hypothesize that flexibility afforded by these hinge regions might be important in preventing steric hindrance between adjacent receptors in two-dimensional signaling arrays (Figure 1.2.9). It should also be noted that the variable hinge bending described in [33] resulted in fuzzy regions in the reconstruction obtained for the chemoreceptor, particularly in the middle region between the contacts with CheA/CheW and the membrane.



**Figure 1.2.9 Schematic representations of signaling arrays formed by twelve chemoreceptor trimers of dimers.** On the left, the forbidden orientation of the receptors is shown at zero values of the HAMP- and Gly-hinges (the so-called "straight" receptors). Shown on the right is the receptors' permitted orientation with a  $10^\circ$  deflection of the Gly-hinge and  $8^\circ$  of the HAMP-hinge. The top images correspond to a view approximately parallel to the membrane, the bottom images show a view perpendicular to the membrane as viewed from the periplasm. Figure taken from [33].

Cryo-electron tomography combined with molecular dynamics modeling was used in to study the serine chemoreceptor Tsr from *Escherichia coli* in the context of the native signaling array [46]. Here, signaling-related conformational changes were observed both at the individual dimer level and at the level of the trimer of dimers within the hexagonal signaling array. In particular, it was shown that individual Tsr dimers making up trimers of dimers exhibit asymmetrical flexibility, which is a signaling state function. The study also revealed that the compactness of Tsr trimers of dimers is changed when switching between different signal transmission states. This is achieved mainly due to the flexing in the Gly-hinge region (see above). A more compact conformation corresponds to an inactivated state of the kinase (kinase -OFF) while a more extended conformation related to an active state (kinase- ON) (Figure 1.2.10).

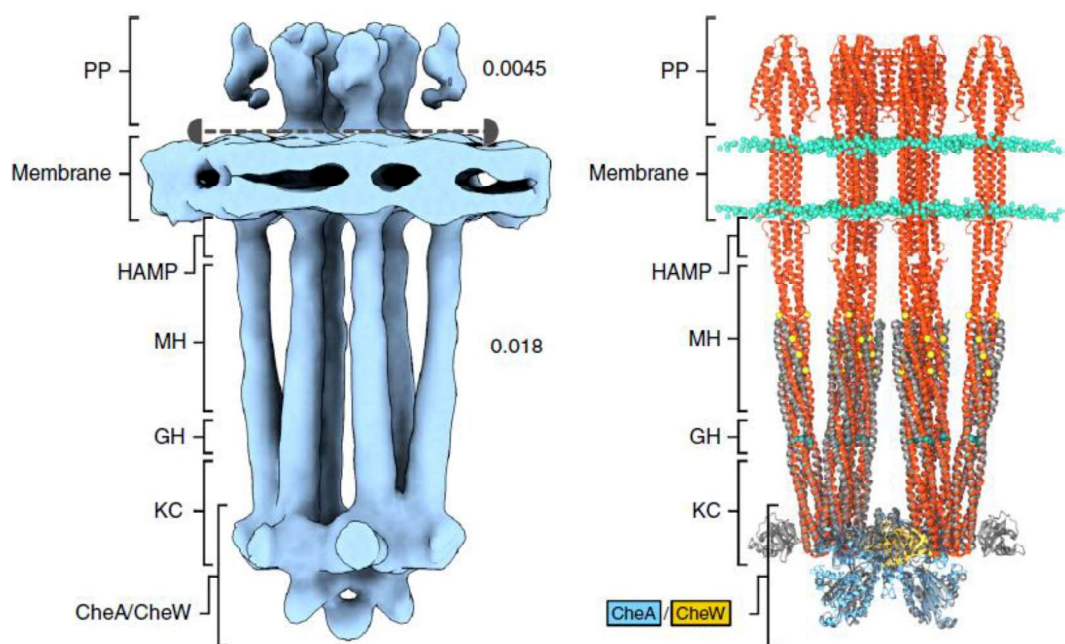


**Figure 1.2.10** A schematic showing two different signaling states of the cytoplasmic regions of Tsr trimers of dimers. Top - density maps (grey surfaces) obtained from cryo-electron tomography data for 4E (inactive, kinase-OFF) and 4Q (active, kinase-ON) states superimposed with secondary structure elements corresponding to atomic models of this region. Bottom - the central axes of homodimers in each of the two states of the trimers of dimers. Yellow beads indicate methylation sites, and dark gray beads indicate amino acids corresponding to the glycine flex hinges. Figure adapted from [46].

The results of the work described in [46] clearly suggest a decisive role of the Gly-hinge region in ensuring the receptor's flexibility to achieve various states during signal transduction while, at the same time, maintaining the structural constraints imposed by the membrane and by the signal array architecture. This contradicts the conclusions of [33], in which it was suggested that the conformational state did not correlate with protein signaling state. However, the inferences of the study reported in [46] seem to be more valid since the results were obtained from Tsr under native conditions, while in [33], only individual dimers of Tar were studied.



Very recently, work exploiting cryoelectron tomography and averaging of subtomograms in structural studies of chemoreceptors was reported [47]. Here, a specially optimized *E. coli* strain with smaller than usual cells, but with a higher relative proportion of the membrane area corresponding to chemoreceptor arrays was used. In [47], a three-dimensional reconstruction for a trimer of dimers, coupled with high-resolution structural information for individual domains and molecular dynamics simulations, allowed the modelling of the structure of a full-length chemoreceptor and its interactions with the kinases CheA/CheW (Figure 1.2.11).



**Figure 1.2.11 Structural studies of chemoreceptors exploiting cryoelectron tomography and averaging of subtomograms.** The isosurface obtained from cryo-electron tomography (left) and the molecular model, obtained by combining this with molecular dynamics simulations and existing high-resolution structural information, of a full-length chemoreceptor trimer of dimers in a complex with CheA and CheW (right) are shown. Figure from [47].

Most of the studies of TCS described above relate to either chemoreceptors or histidine kinases. One of the difficulties in photoreceptor research is that their host organisms are halophilic archaea. Their habitat is associated with high salt concentrations, the maintenance of which is necessary for proper protein folding and oligomerization. While this does not pose too many problems for structural studies of the transmembrane domains of photoreceptors [21, 23], this is not the case when

working with full-length proteins or their fragments containing cytoplasmic or periplasmic domains.

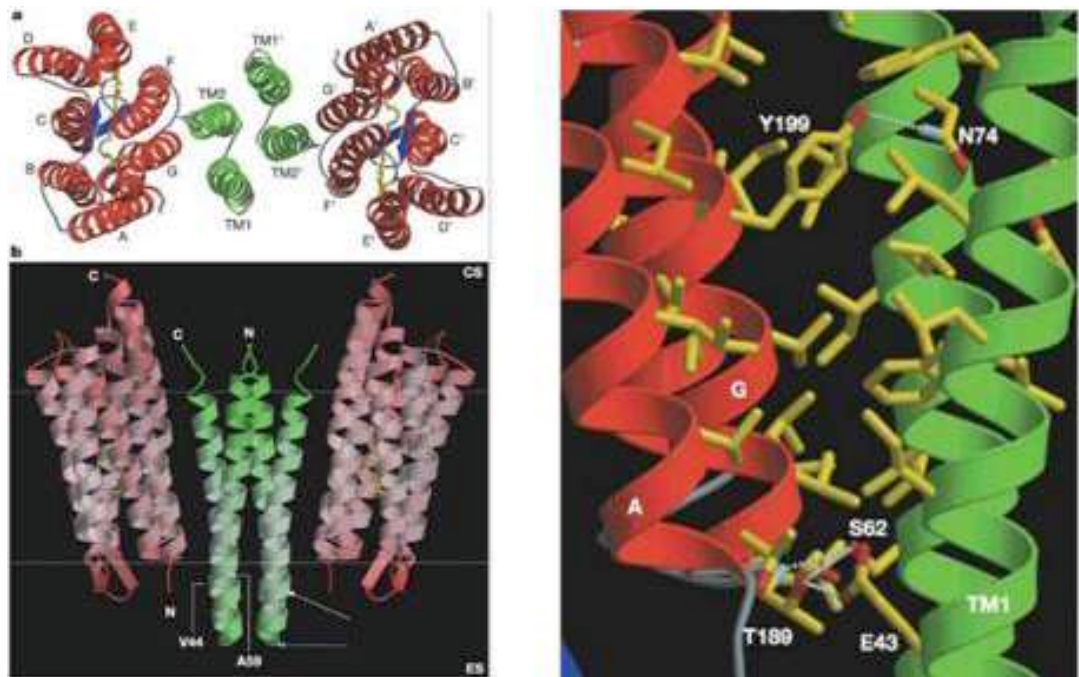
Indeed, the need to maintain high concentrations of salts and/or other substances is a challenge for X-ray crystallography and cryo-electron microscopy (cryo-EM) studies targeting full-length photoreceptors. However, for cryo-EM this situation is not unpromising due to developments both in the field of the study of halophilic archaeal cells (for example, *Halobacterium salinarum*) at high salt concentrations [48] and in the field of cryo-electron microscopy with negative contrast (so-called cryo-negative staining), which is achieved through the use of ammonium molybdate [49-51].

### 1.3 The complex between sensory rhodopsin II and its related transducer molecule (*NpSRII/NpHtrII*)

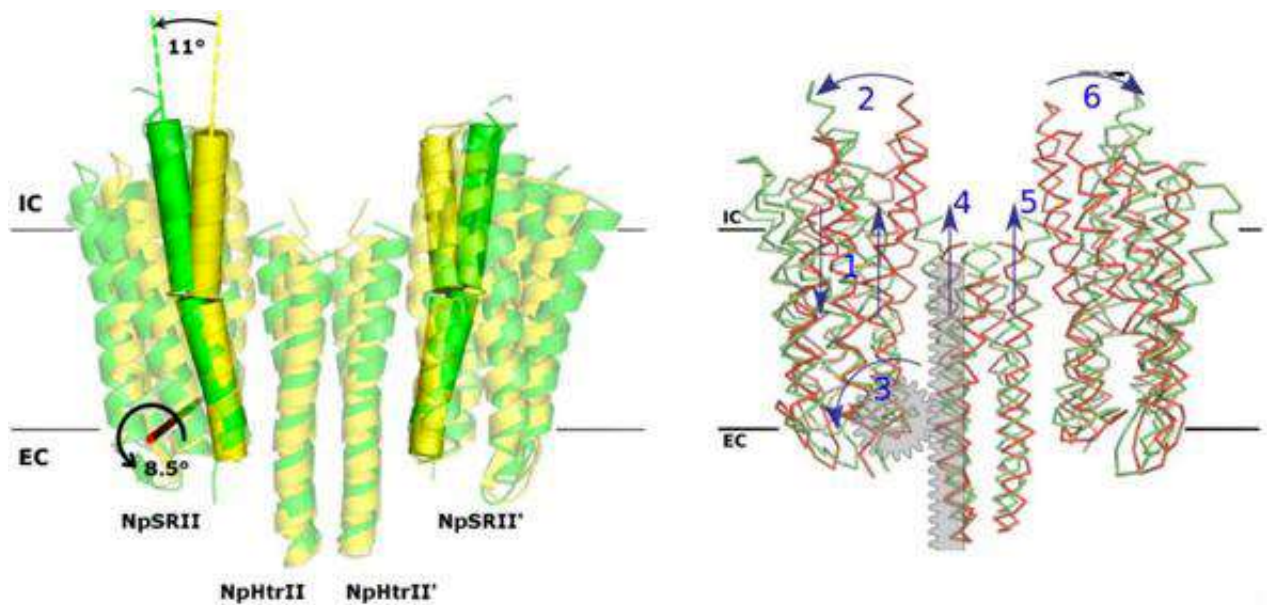
*NpSRII/NpHtrII* mediates negative phototaxis in the halobacterium *Natronomonas pharaonis* and is the sensor (first component) of a TCS. The HtrII transducer is similar to chemoreceptors in domain architecture [52] (see Figure 1.2.1, Figure 1.2.2) and, similarly to chemoreceptors, the functional state of full-length *NpSRII/NpHtrII* is a trimer of homodimers [53].

Sensory rhodopsin II from *N. pharaonis* (*NpSRII*) is a photoactive transmembrane seven-helix protein with an absorption maximum in the optical range at a wavelength of 498 nm [54, 55]. Upon exposure to light *NpSRII* combined with its HtrII transducer induces a cascade of cytoplasmic signals relayed via the TM2 helix of the transducer.

In 2002, the crystal structure, at 1.94 Å resolution, of a complex between SRII in its ground state and a truncated transducer (*NpSRII/NpHtrII*<sub>114</sub>) was reported [21]. This made it possible to explain the complex's first signal transmission stage with the functional complex adopting a dimeric form.

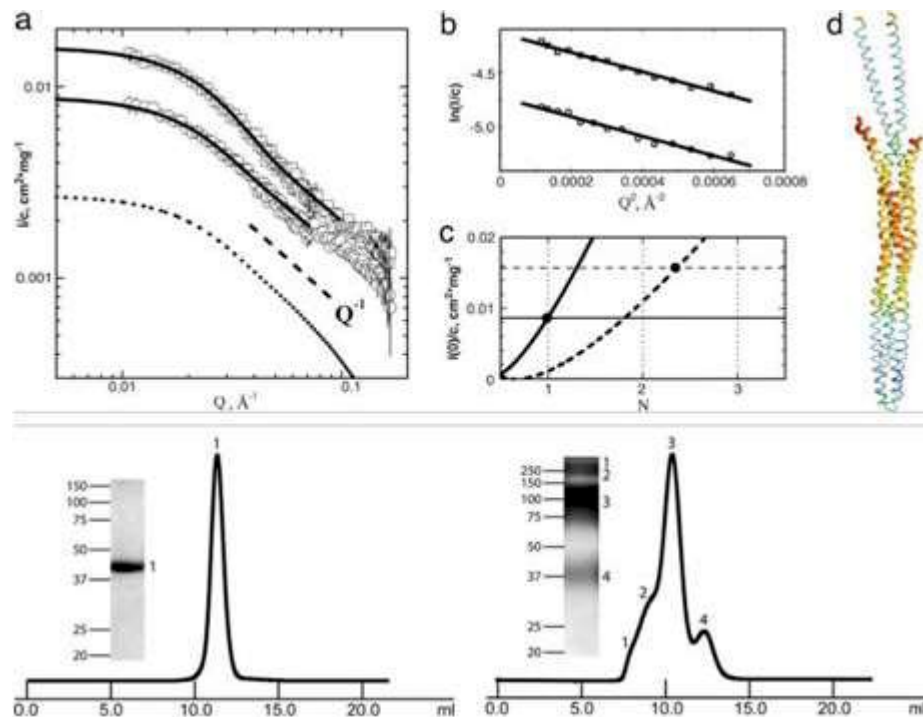


**Figure 1.3.1** The crystal structure of *NpSRII/NpHtrII*<sub>114</sub>. The crystal structure of *NpSRII/NpHtrII*<sub>114</sub> dimer (left) and a diagram of hydrogen bonds and van der Waals contacts between *NpSRII* and *NpHtrII*<sub>114</sub>. *NpSRII* is shown in red, *NpHtrII*<sub>114</sub> in green. Figure taken from [21].



**Figure 1.3.2** The different conformations adopted by *NpSRII/NpHtrII*<sub>114</sub>. Left: The V- (green) and U-shaped (yellow) forms of the *NpSRII/NpHtrII*<sub>114</sub> homodimer (parallel to the membrane). The G-helix of *NpSRII* is shown as a cylinder. The axis of rotation of *NpSRII* as a rigid body during the transition from "U" to "V"-shape is shown in red, and the corresponding angle shown. The right panel shows the signal transmission rack model corresponding to the normal mode mobility analysis. Figure taken from [25].

This work was complemented with high-resolution structures of the complex with intermediate states of the SRII photocycle [23]. In 2017 the crystal structures of the ground and active M-states of the *NpSRII/NpHtrII*<sub>114</sub> complex in the space group  $I2_12_12_1$  were reported [25]. Here, the relative orientation of the symmetrical parts of the dimer is parallel (that is, it is "U" -shaped, see Figure 1.3.2), which distinguishes it from the "V" -shaped structure of the *NpSRII/NpHtrII*<sub>114</sub> complex described earlier in the space group  $P2_12_12$  [25]. However, the structures of the monomers individually are very similar. The authors observed a piston shift of both transmembrane helices of the transducer, which coincides with the receptors' rotation. However, this cannot be caused by the transmembrane portion of proteins since the interactions are almost the same in both states. Since the HAMP domain is sterically instead limited, inhibiting conformational changes in the U-shape, the transition from U-shape to V-shape may be involved in signal propagation.



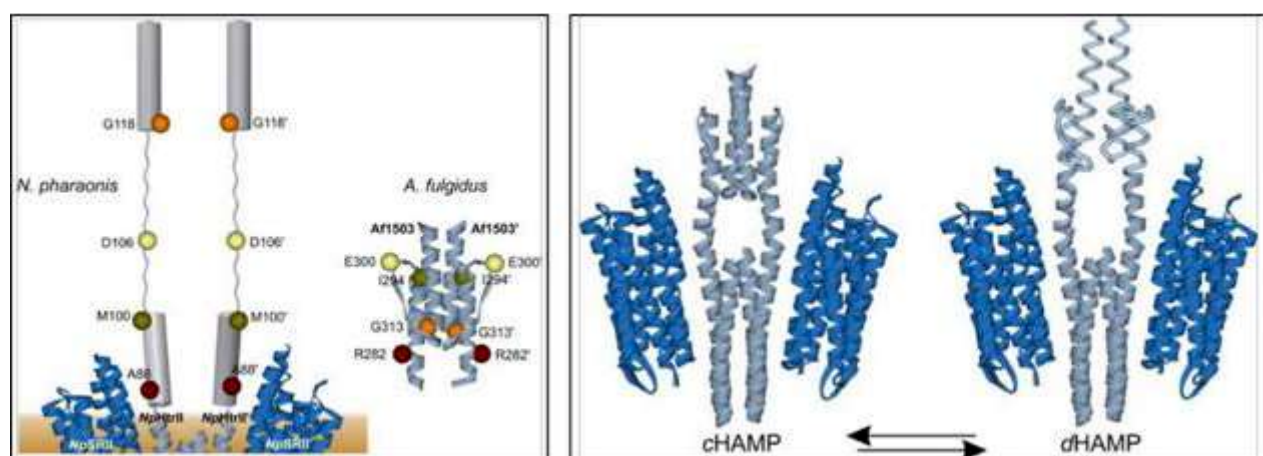
**Figure 1.3.3 Structural studies of *NpHtrII*<sub>229-504</sub> in solution.** Top - (a) concentration-normalized solution scattering curves for *NpHtrII*<sub>229-504</sub> in low and high salt conditions. (b, c) - Guinier plots and the dependence of the concentration-normalized scattering intensity at the zero angle  $I(0)$  on the aggregation number. (d) Image of a *Tsr-cyt* model obtained by X-ray diffraction analysis for a fragment of the serine receptor's cytoplasmic domain (PDB code 1QU7 [43]), corresponding to a fragment homologous to *NpHtrII*<sub>229-504</sub>. Bottom - the results of the dependence of optical density on volume and images of polyacrylamide gel obtained for *NpHtrII*<sub>229-504</sub> samples after cross-linking in low (left) and high (right) salt conditions. Figure adapted from [56].

As noted earlier, *N. pharaonis* is an extremophile that lives in a saturated saline solution at alkaline pH values. Indeed, the optimal growth conditions for the organism correspond to a salt concentration of 3.5 M NaCl and pH 8.5 [56] and it has been shown that the structure and oligomeric state of the cytoplasmic domain of *NpHtrII*<sub>229-504</sub> strongly depends on salt concentration [57]. Specifically, at low salt concentration (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.76 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4) the fragment was in a monomeric state, while increasing the salt concentration to 4.0 M NaCl saw the formation both of dimers (Figure 1.3.3) and of higher order oligomers (probably, trimers of trimers). Moreover, circular dichroism (CD) measurements also reported in [56] showed that at low salt conditions *NpHtrII*<sub>229-504</sub> is disordered while at 4.0 M NaCl (as well as 4.0 M KCl and 40% ammonium sulfate) CD spectra show a typical  $\alpha$ -helical profile. This latter result



confirms the critical importance of maintaining physiological high salt conditions for proper protein folding.

The results reported in [56, 57] were supplemented by further CD measurements in various different buffer solutions [58]. These showed that an increase in glycerol concentration leads to a transition in structure from disordered to  $\alpha$ -helical. Glycerol has little effect on the ionic strength of a solution, but it can significantly reduce water activity [56, 57]. As *NpHtrII*<sub>229-504</sub> contains many charged amino acids it might be expected that ionic strength is a critical parameter affecting protein folding and oligomerization. However, the results reported in [58] suggest that water activity is also of importance for folding. Nevertheless, as the question of the effect of water activity on the oligomerization *NpHtrII*<sub>229-504</sub> has not been considered further, it is impossible to attribute crucial importance to this observation.

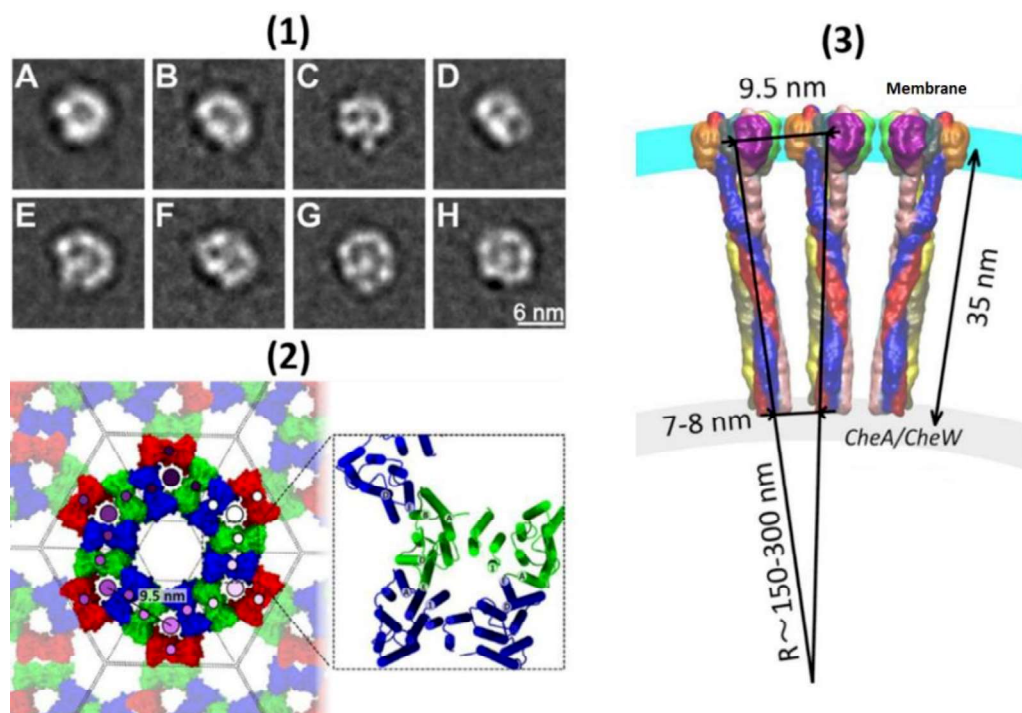


**Figure 1.3.4** The "two-level" equilibrium between dynamic (dHAMP) and compact (cHAMP) conformations. *Left:* HAMP domain models for *HtrII* from *N. Pharaonis* and *Af1503* from *A. fulgidus*. *Ca* atoms of amino acids that have undergone mutation are presented as coloured spheres. *Right:* the 'two-level' equilibrium of *NpHtrII*<sub>157</sub> showing the corresponding compact (cHAMP) and dynamic (dHAMP) conformations of HAMP1. Figure taken from [59].

The effect of salt concentration on the dynamic properties of *NpSRII/NpHtrII* was considered in [59]. Here, the authors focused their attention on the ordering of the HAMP1 domain of the *NpHtrII* dimer as studied using paramagnetic resonance methods. Using a series of spin-labeled mutants of the transducer fragment, *NpHtrII*<sub>157</sub>, which contains only the first HAMP domain following the TM2 helix, they showed the HAMP domain to be in a "two-level" equilibrium between highly

dynamic (dHAMP) and more compact (cHAMP) conformations (see Figure 1.3.4). The structural properties of cHAMP are consistent with the four-helix model obtained by NMR for the HAMP domain from *Archaeoglobus fulgidus* [60].

It should be noted that, because of the above-described features, the angles of bending of the HAMP1-hinge (and, probably, the HAMP2- and Gly-hinges) in the case of photoreceptors from halophilic harps can adopt a broader range of values that seen for the Tar chemoreceptor from *E. coli* [33]. In particular, this remark applies to the case of relatively low salt conditions.

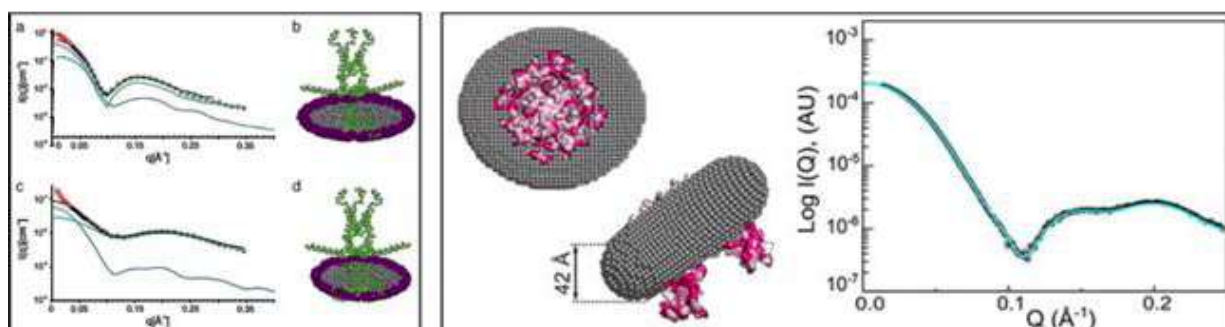


**Figure 1.3.5** Cryo-EM studies of a negative phototaxis system TCS. 2D class averages (1) for full-length *NpSRII/NpHtrII* and models of two-dimensional photoreceptor packing dimer trimers into hexagonal arrays (2, 3). Figure adapted from [53].

For a TCS negative phototaxis system to function, signal amplification is required. This is achieved by forming two-dimensional membrane arrays. As noted above, the key "building block" of such an array is a trimer of photo- (chemo-) receptor homodimers. Work reported in [53], carried out using electron microscopy coupled with calculations using molecular dynamics, showed that for sensory rhodopsins I and II in complex with their related transducers elements of the hexagonal arrays could be formed even in the presence of detergent (Figure 1.3.5).

## 1.4 Application of small-angle scattering methods for structural studies of membrane proteins

The exploitation of small-angle scattering (SAS) methodologies such as Small-Angle X-ray Scattering (SAXS) or Small-Angle Neutron Scattering (SANS) for the structural studies of MPs requires the maintenance of the MP under study in a close-to-native state. Here, lipid vesicles, phospholipid nanodiscs, bicelles, and detergent micelles [39–41] are examples of settings used to mimic the membrane environment. If existing models are available to help in the interpretation of solution scattering data, the detergent belt must be modelled either explicitly or with pseudo-atoms [61–68] (Figure 1.4.1). This can be done with the program MEMPROT [69], which accepts the atomic model of the protein and SAXS data as input data, and at the output gives the optimal parameters of the detergent belt, using the CRY SOL program [70] of the ATSAS software package [71] to calculate the theoretical SAXS curves.



**Figure 1.4.1** Solution scattering studies of membrane proteins. Examples of experimental SAXS curves from solutions of membrane proteins, their model approximations, and images of hybrid models, of which the approximations were obtained. The left and right panels show results presented in [68] and [61], respectively.

Structural modeling of MPs from SAXS data is nontrivial, and the necessary computational procedures require further formalization and simplification. In [72], an automated software pipeline (see Figure 1.4.2), integrated with the ISPyB information management system [73], is proposed. The pipeline is aimed at the preliminary analysis of SAXS data and the first stage of MP reconstruction in detergent solutions to simplify high-throughput studies, especially on synchrotron beamlines. The program queries the available *a priori*



information in the ISPyB database via dedicated services, estimates parameters from the SAXS data without using models (scattering invariants), and generates preliminary models using either *ab initio*, methods based on a known high-resolution structure, or mixed/hybrid methods. Automated analysis results can be verified online using the industry-standard ISPyB interface, and simulation design parameters can be used to further in-depth simulations outside the pipeline.

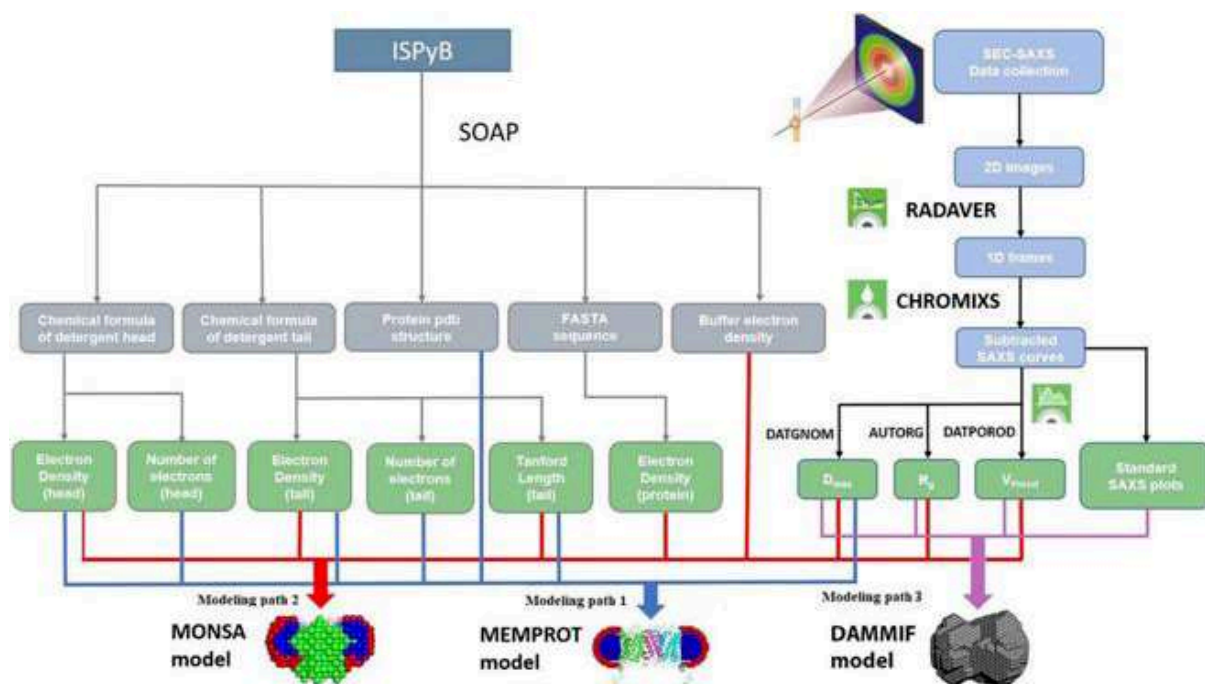
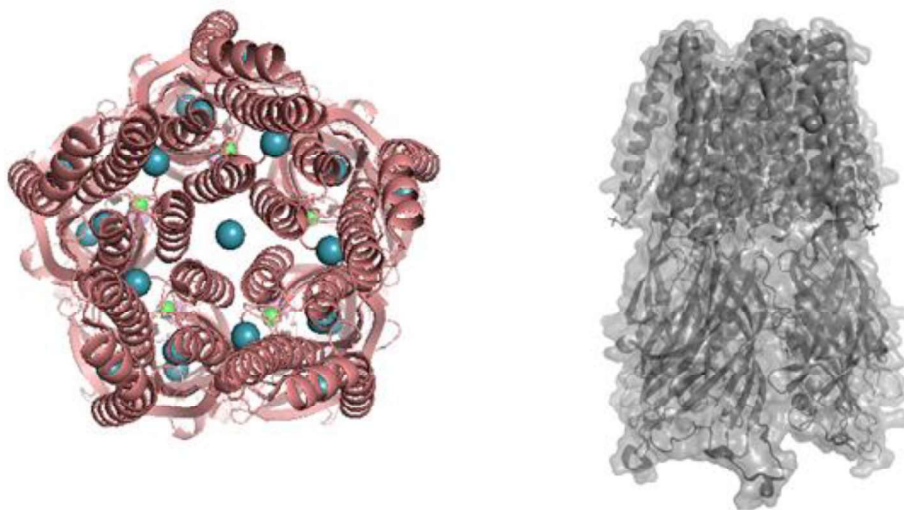


Figure 1.4.2 Schematic of an automatic software pipeline for structural studies of membrane proteins based on small-angle X-ray scattering data. Figure from [72].

## 1.5 Noble gases and membrane proteins

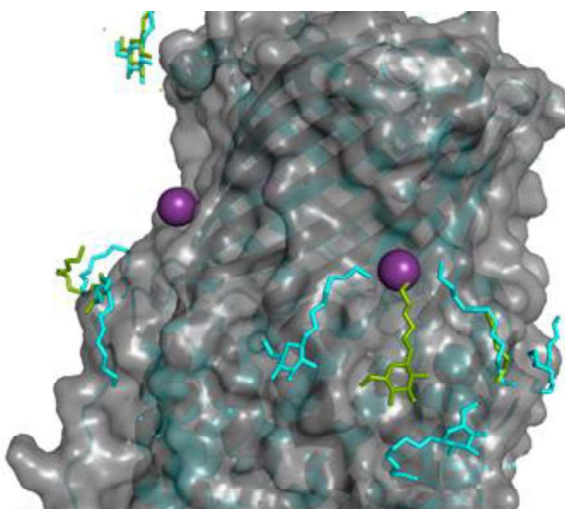
The study of membrane proteins with noble gases is of particular interest to researchers but, until now, has usually focused on the use of xenon (Xe). This is probably because the binding of Xe to a protein molecule can be induced at much lower pressure ( $\sim 20$  bar) than is required in the case of other noble gases such as krypton (Kr,  $\sim 180$  bar) or argon (Ar,  $\sim 2000$  bar) [74]. However, where noble gases are used as probes to map intramolecular channels [75-77] in proteins Xe is not the most suitable choice, due to its larger Van der Waals radius (216 pm). Nevertheless, Xe-derivatisation of MPs has led to interesting results:

1. Sauguet and colleagues [75] studied Xe binding to GLIC, a pentameric ligand-controlled ion channel whose action is inhibited by xenon. The work identified Xe atoms binding in three major regions: 1) in hydrophobic cavities, 2) at the interface between adjacent subunits, 3) in hydrophobic pockets. However, the most interesting Xe position identified is in the center of the ion channel in its open form (Figure 1.5.1). The Xe atom does not replace the lipid located in the center but binds precisely above it. As explained by the authors, this blocks the ion channel. Indeed, ordered water molecules (including a pentagon) have been observed at this location in the GLIC open form and have been shown to be essential for ion translocation by allowing the permeant ion to cross hydrophobic barriers. This result can be interesting in the study of viral rhodopsins where a lipid (retinal, see §1.1) is also found at the molecule center.



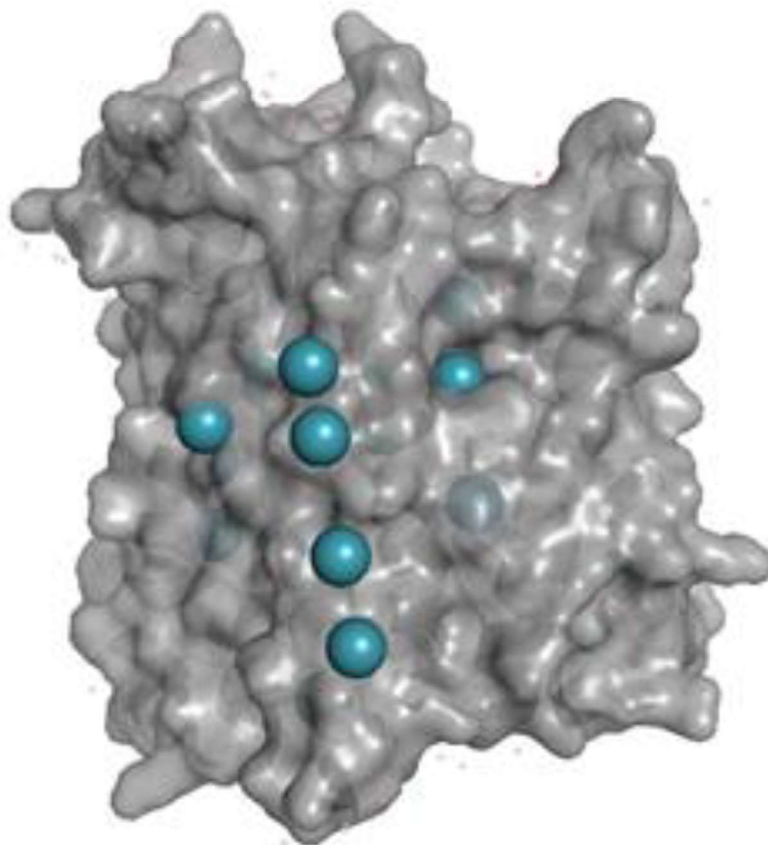
**Figure 1.5.1** The GLIC pentameric Ligand-Gated Ion Channel in its open form complexed to Xe (cyan spheres). Of note is the Xe atom found in the center of the pentamer and which blocks the ion channel. Figure adapted from [75].

2. Xe derivatization of the efflux pump OprN [76]. Here, the authors tried to identify channels which would allow small ligands to diffuse through the membrane (Figure 1.5.2). They identified 14 binding sites in hydrophobic regions, both within the protein molecule and in hydrophobic pores on the surface. In particular, Xenon binding to OprN highlights a strong discontinuity in the hydrophobic gradient along the channel at this level which might be an additional driving force in the passive mechanism for drug transport out of the cell.



**Figure 1.5.2** Xe-derivatization of OprN. OprN is shown as a grey surface, Xe atoms as magenta spheres. Detergent molecules are shown as green (native structure) or cyan (Xe-derivatized structure) sticks. Figure adapted from [76].

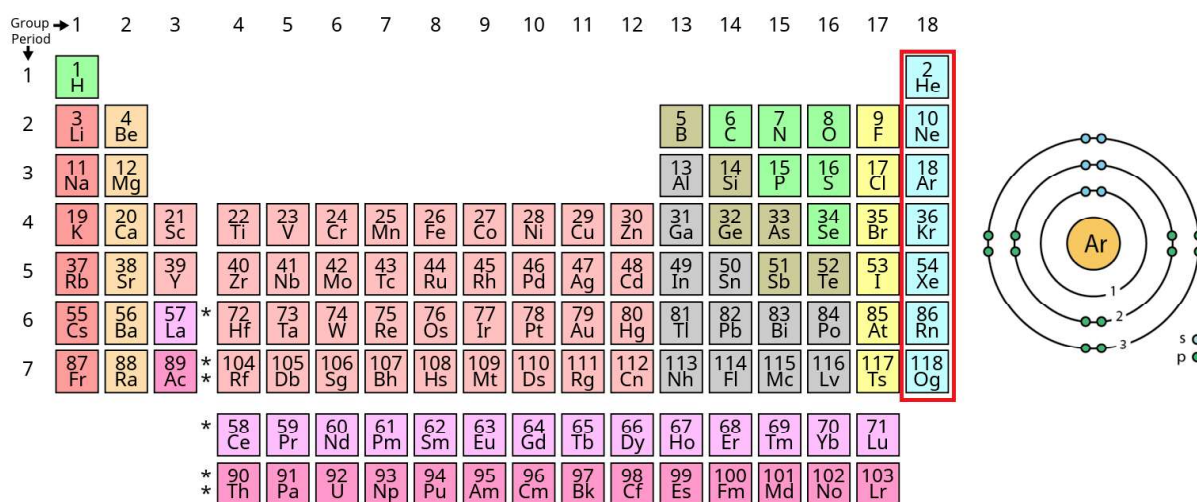
3. The archaeal ammonium transporter Amt-1 was also investigated using Xe-derivatisation [77]. This allowed the authors to explain pumping through the molecule of the  $\text{NH}_4^+$  ions. In this case, and due to its chemical inertia and its similar size to the ammonium ion, Xe is an excellent tool for channel exploration. As in previous studies, Xe bound either in hydrophobic cavities or in pores on the surface of a molecule.



**Figure 1.5.3** The Ammonium Transporter Amt-1 from *A. fulgidus* (grey surface) in complex with Xe (cyan spheres). *Figure adapted from [77].*

## 1.6 Thermodynamic principles of noble gas-protein interactions

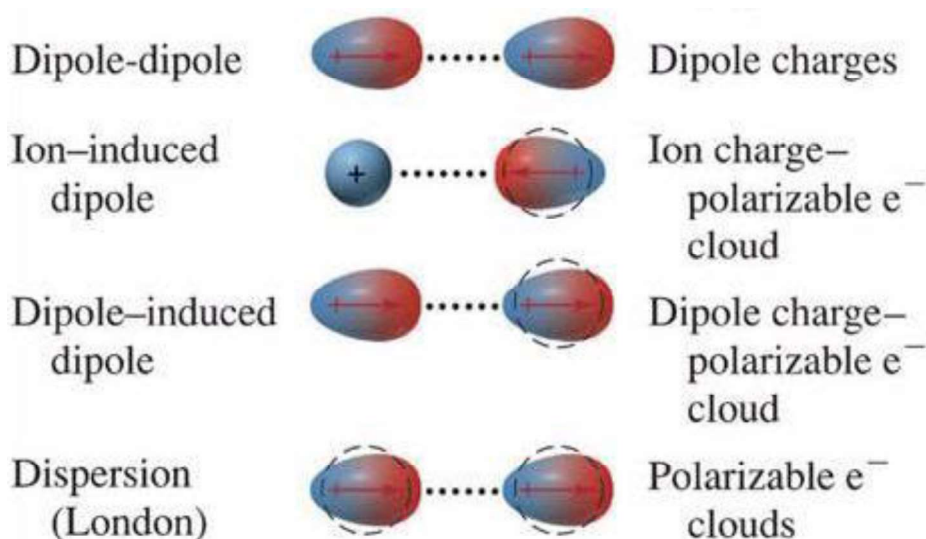
Noble gases do not have free electronic orbitals (Figure 1.6.1). Therefore, they cannot easily interact with other atoms and, hence, are often called the ‘inert gases’. The only possible interaction mechanism for a noble gas atom is the Van der Waals interaction. In modern science, this term is usually applied to forces arising from a molecule’s potential to be polarized and to form dipoles [78].



**Figure 1.6.1** The electronic structure of noble gases. *Left: The periodic table with the noble gases highlighted in red. Right: A schematic diagram of the electronic structure of argon showing that its s and p orbitals are completely occupied by electrons. Figure adapted from [79].*

In fact, van der Waals interactions consist of three types of weak electromagnetic interactions (Figure 1.6.2):

- Orientational forces, dipole-dipole attraction between molecules that are constant dipoles. An example is HCl in liquid and solid-state. The interaction energy is inversely proportional to the cube of the distance between the dipoles.
- Induction attraction (polarization attraction). The interaction between a constant dipole and induced dipole. This interaction's energy is inversely proportional to the sixth power of the distance between the dipoles.
- Dispersion attraction (London forces, dispersion forces). These are due to interactions between instantaneous and induced dipoles. This interaction's energy is inversely proportional to the sixth power of the distance between the dipoles.



**Figure 1.6.2 Schematic diagrams of three types of weak electromagnetic interactions.** *Ion-induced and dipole-induced attractions are two instances of induction attraction (see main text above). Figure adapted from [80].*

As noble gas atoms have a zero net charge and are spherically symmetric, coulomb interactions, hydrogen-bonding, and permanent dipole-dipole interactions cannot be involved their binding to proteins. Thus, the only possible attractive interactions between noble gas atoms and proteins are due to charge (i.e. ion)-induced, dipole-induced, and London (dispersion) forces. The critical physical parameter in these interactions is the noble gas atoms' electronic polarizability. The usual repulsive forces between atoms and molecules in close contact with each other also play an important role because they determine the minimum allowed size of a cavity in order for a particular noble gas atom to occupy it [81]. The characteristic bond length is formed from the van der Waals radii of the elements forming the bond. Table 1.6.1 shows these radii for the elements making up protein molecules and the same inert gases.

**Table 1.6.1 The van der Waals radii of some elements**

Ar	188 pm	N	155 pm
Kr	202 pm	C	170 pm
Xe	216 pm	S	180 pm
O	152 pm	H	120 pm



Under the influence of an external electric field  $E$ , an electric dipole  $P$  is induced in a noble gas atom. The induced dipole can then establish attractive electrostatic interactions with surrounding electric charges or dipoles. This is the physical basis of the forces involved in noble gas binding to proteins. Assuming a linear response, the induced dipole is proportional to the external field strength ( $P \propto E$ ), where the proportionality coefficient is named the electronic polarizability. This is a characteristic parameter for a given atom type and, qualitatively, it expresses the ease with which the electron cloud can be displaced. Most atoms and small molecules have relatively low polarizabilities (1.63 for Ar, 1.60 for O<sub>2</sub>, 1.48 for H<sub>2</sub>O). However, for larger molecules and for diatomic molecules which have many electrons, the values are higher (4.6 for Cl<sub>2</sub>, 4.5 for C<sub>2</sub>H<sub>6</sub>, 8.2 for CHCl<sub>3</sub>). It is noticeable that even though it is a single atom, the polarizability of Xe is quite large (4.00). For Kr, the polarizability is 2.46, which is significantly lower and explains why Kr usually binds with a weaker occupancy to known Xe-binding sites [81]. In this regard, under normal conditions, we do not observe the formation of protein-noble gas complexes and pressurized atmospheres of noble gases are required to achieve this.

The binding energy is characterized by a change in Gibbs free energy, consisting of enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ ) factors (equation 1, below). The enthalpy factor is not a variable quantity, and is determined empirically. The entropy factor depends on the thermodynamic parameters: temperature, volume, and pressure. Therefore, when changing one of these parameters, we can change the entropy factor and, consequently, the change in Gibbs energy.

$$\Delta G = \Delta H - T\Delta S; (1)$$

$$\Delta S = S_2 - S_1 = k \ln \frac{w_2}{w_1},$$

where  $w$  is thermodynamic probability, with  $w = V^{N_A}$  ( $V$ - volume,  $N_A$  - Avogadro number). Thus,

$$\Delta S = k \ln \left| \frac{V_2}{V_1} \right|^{N_A} = k N_A \ln \frac{V_2}{V_1} = R \ln \frac{V_2}{V_1},$$

where the Universal Gas Constant  $R=8.31 \text{ J/mole}\cdot\text{K}$ . From the equation of state of an ideal gas, it follows that for 1 mole of gas:

$$V = \frac{RT}{P}$$

Where  $T$  – temperature,  $P$  – pressure. For isothermal process:

$$\Delta S = R \ln \frac{RT/p_2}{RT/p_1} = R \ln \frac{p_1}{p_2}.$$

Substituting Gibbs free energy into the equation, we get:

$$\Delta G = \Delta H - T\Delta S = \Delta H - TR \ln \frac{p_1}{p_2}.$$

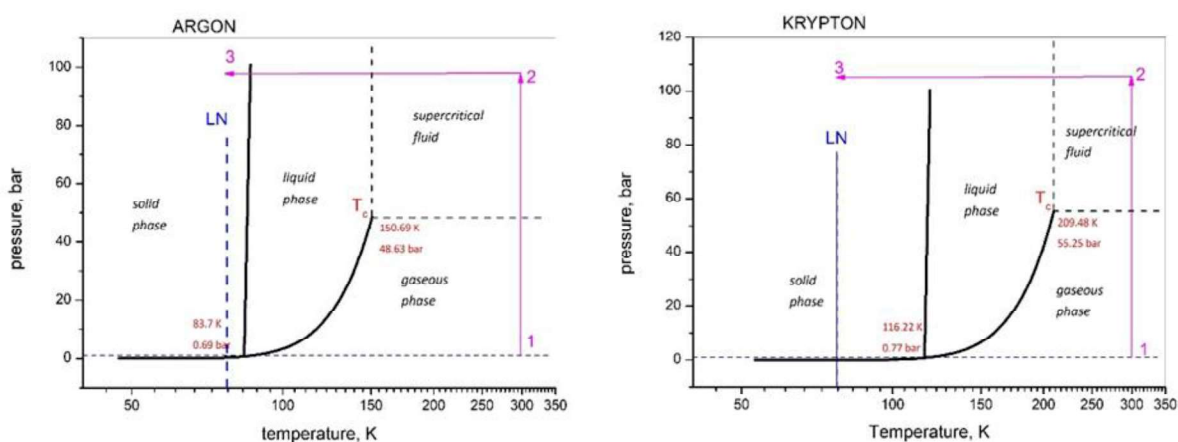
As a result, with increasing pressure, the entropy factor becomes more negative. The Gibbs free energy increases, which allows the noble gas to form a complex with a protein molecule. The binding of Xe to a protein molecule requires a much lower pressure ( $\sim 20 \text{ bar}$ ) than for Kr ( $\sim 180 \text{ bar}$ ) or Ar ( $\sim 2000 \text{ bar}$ ) due to differences in atomic polarizability. However, an advantage of using Kr and Ar is that the atoms of these gases are much smaller than Xe and so they have access to smaller cavities which may be important for functionality.

Pressure-temperature phase diagrams for Ar and Kr are shown in Figure 1.6.3. Based on these – and similar diagrams for Xe – a protocol for the preparation of noble-gas derivatives of crystals of biological macromolecules, including MPs, is as published in paper [82].

In this preparation process, a crystal of a protein is harvested and placed in a pressurizing device at room temperature. The pressure of noble gas atmosphere is increased without changing the temperature. At this time, the entropy factor of Gibbs energy decreases, increasing the possibility of the binding of noble gas atoms to



molecules in the crystal (Step 1, in Figure 1.6.3). The temperature is then lowered without changing the pressure (Step 2). At this stage, an important point is the crystal's presence in the noble gas's liquid phase so that during the liquid-solid transition, the noble gas does not destroy the protein crystal. At this stage, the entropy factor increases, preventing new gas atoms from binding. Then the pressure is released while maintaining the temperature (Step 3). Such a crystal preparation procedure allows controlling the crystal quality and thermodynamically explains the formation of a protein-noble gas complexes in crystals.



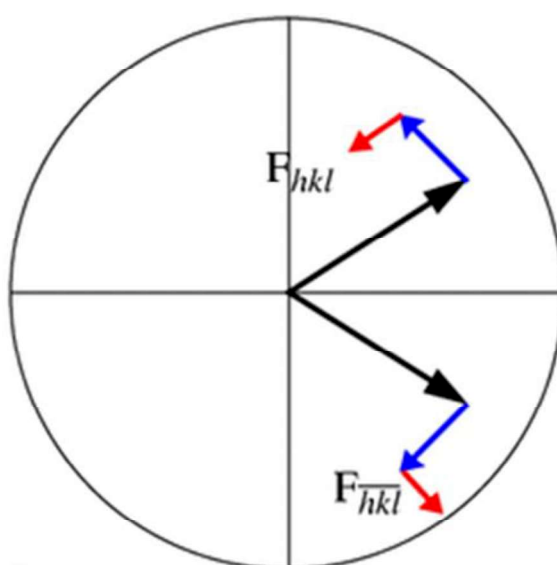
**Figure 1.6.3 Pressure-temperature phase diagrams of argon and krypton.** *The temperature of liquid nitrogen (LN) shown as a blue dashed line. The three main steps in the preparation of noble gas derivatives (see main text) are highlighted in magenta. Data taken from web database [83] (<https://webbook.nist.gov/chemistry/>).*

## 1.7 Anomalous Scattering of Noble Gases

When collecting data from a crystal by macromolecular crystallography (MX), the intensity of the diffracted waves is measured. From these intensities we get the amplitudes of the scattered waves, but we lose information about the phase. This is commonly known as the "phase problem". The most common method for finding phases is molecular replacement, but the method of finding the position of heavy atoms (either native to the protein or added before the experiment) is also used [84].

$$f = f_0 + f' + if'' \quad (2)$$

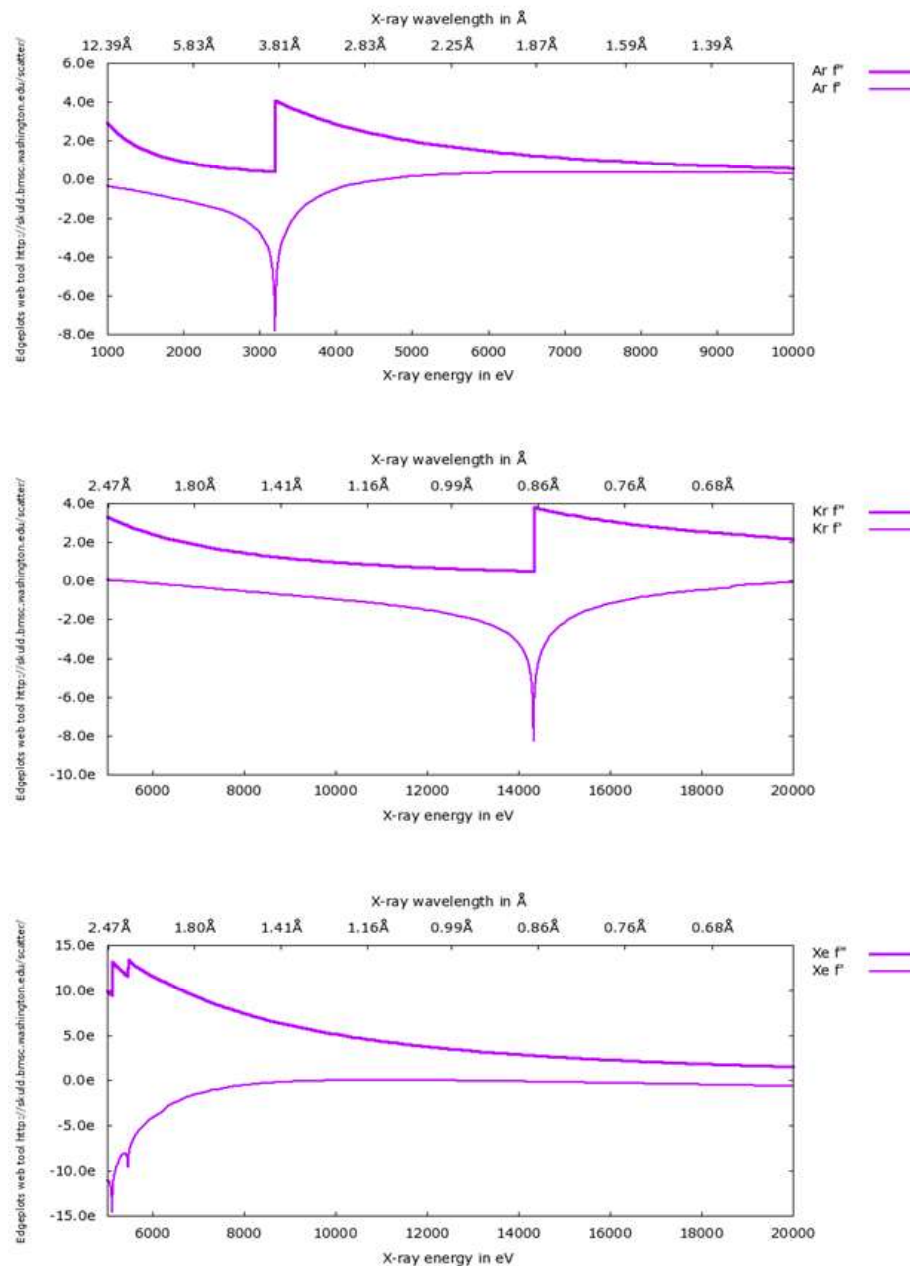
The atomic scattering factor (equation 2) contains three components: the normal scattering term  $f_o$ , which depends on the scattering angle, and two terms  $f'$  and  $f''$ , which do not depend on the scattering angle, but on the wavelength. These represent anomalous scattering that most strongly occurs at atom absorption edges, when the energy of an absorbed X-ray photon is sufficient to propel an electron out of the inner shell. Here, the dispersion term  $f'$  changes the normal scattering coefficient, while the absorption term  $f''$  is ahead of the phase by  $90^\circ$  (Figure 1.7.2). According to Friedel's law  $|F_{hkl}| = |F_{-h-k-l}|$ . However, in the presence of anomalous atoms, this is violated, causing anomalous differences ( $\Delta F_{\text{ano}} = F_{hkl} - F_{-h-k-l}$ ) (Figure 1.7.1).



**Figure 1.7.1** A vector diagram illustrating equation (2) and showing that in the presence of anomalous scattering Friedel's law is no longer obeyed and that  $|F_{hkl}| \neq |F_{-h-k-l}|$ . Figure from [http://skuld.bmsc.washington.edu/scatter/AS\\_Bijvoet.html](http://skuld.bmsc.washington.edu/scatter/AS_Bijvoet.html).

An increasing number of protein structures are now being phased using only one diffraction dataset using the single wavelength anomalous dispersion/diffraction (SAD) method [85]. A SAD experiment allows calculation of anomalous differences ( $\Delta F_{\text{ano}} = |F_{hkl}| - |F_{-h-k-l}|$ ) which are then used as estimates of the heavy atom contribution to scattering and allow direct or Patterson methods to be used to determine the position of the heavy atom substructure which is then used in crystal structure phasing protocols.

In MX, the use of noble gases is often associated with their anomalous dispersion properties (Figure 1.7.1).



**Figure 1.7.1** Theoretical values of the anomalous scattering properties of Ar (top), Kr (middle) and Xe (bottom). Generated at [86].

Of the noble gases exploited in this work, Xe shows the highest anomalous scattering with  $f'' \sim 12 - 13e^-$  at its L- absorption edges. Unfortunately, at an energy of around 5 keV ( $\sim 2.5 \text{ \AA}$ ) these are outside the routinely accessible energy ranges of most tunable synchrotron-based MX beamlines [87, 88]. However, at more

accessible energies (*i.e.* 6 keV,  $\lambda = 2.06 \text{ \AA}$ ) this signal is still more than  $10e^-$ , while at  $E = 12 \text{ keV}$  ( $\lambda = 1.0 \text{ \AA}$ ) it is still considerable ( $\sim 4e^-$ ). The K- absorption edge of Kr ( $f'' \sim 4 e^-$ ) occurs at  $E \sim 14.3 \text{ keV}$  ( $\lambda = 0.87 \text{ \AA}$ ). While this is smaller than is available for Xe, the energy at which it occurs does allow for the possibility of the use of Kr derivatives of protein crystals in multi-wavelength phasing protocols [89]. The K-absorption edge of Ar occurs at  $E \sim 3 \text{ keV}$  ( $\lambda = \sim 4 \text{ \AA}$ ), well outside the routinely accessible range of most tunable synchrotron-based MX beamlines. However, the  $f''$  ( $\sim 2e^-$ ) available at  $E = 6 \text{ keV}$  is small, but still exploitable.

In the early years of MX, the positions of one [90, 91] or two [92] Xe atoms bound to myoglobins were revealed using structure factor differences between Xe-derivatized and native crystals. Increasing the pressure of the Xe atmosphere from 1-2.5 to 7-10 bars in the derivatization process allowed Xe binding sites to achieve sustainable higher occupancy (in the range from 0.4 to 1.0) and 3 Xe positions were observed derivatized crystals of metmyoglobin [93].

First proposed in 1967 [94], the first phasing of a crystal structure using the anomalous signal from Xe was finally published three decades later and required improvements in data collection and treatment procedures as well as in pressurizing techniques [95]. While there are now several reports of the *de novo* structure solution of different proteins using Xe and Kr derivatives [81, 96, 97], (Ar has never been used to phase crystal structures), there is only one report description of the successful use of Xe to phase the crystal structure of a MP.[98].

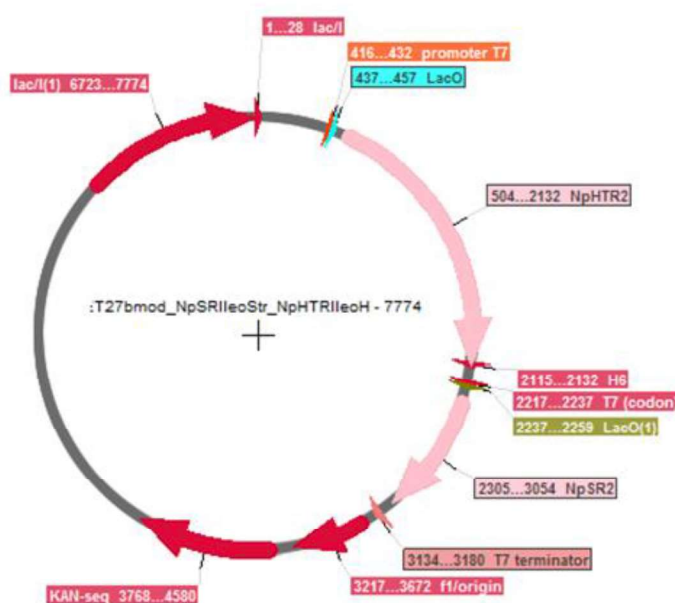
However, the use of noble gases in MX is not confined to the phasing of crystal structures. Using anomalous difference maps, the positions of noble gas atoms, particularly Xe and Kr, can easily be located in the structures of macromolecules even at low resolution [99]. Moreover, due to their high solubility in lipids, noble gas atoms can be used to probe the hydrophobic environment in membrane protein crystals [100, 101]. This aspect of the noble-gas derivatization of MPs will be discussed in more detail later in this thesis (§ 4).

## 2 Materials and Methods

### 2.1 Expression and purification of the *Np*SrII/*Np*HtrII complex

In the studies of *Np*SrII/*Np*HtrII described in this thesis work, proteins were either expressed separately in plasmids and then combined to create a complex or they were co-expressed in a single plasmid. Several different expression systems were used.

When expressed separately, the sequences of the *Np*SrII (UniProt ID P42196) and *Np*HtrII (UniProt ID P42259) genes were optimized for codon composition for expression in *E. coli* using the GeneArt services online program (<https://www.thermofisher.com/fr/fr/home/global/forms/geneart-genes-synthesis-requests.html>) [101]. A strep-tag II (WSHPQFEK with an ENS linker) and a poly-histidine (His<sub>6</sub>) tag (also with an ENS linker) were added to the C-termini of *Np*SrII and *Np*HtrII, respectively. The codon-optimised *Np*SrII and *Np*HtrII genes were cloned into the pSCodon1.2 vector separately using the NheI-AccIII and NdeI-NheI restriction sites for circular PCR due to its small size. The gene for the truncated *Np*HtrII<sub>137</sub> transducer was obtained using PCR. Positive clones were identified using sequencing techniques.



**Figure 2.1.1** General scheme of the pET27bmod co-expression plasmid with inserted modified *Np*SrII and *Np*HtrII genes.

For co-expression, the genes *NpSRII* and *NpHtrII* (or *NpHtrII*<sub>137</sub>) were inserted into a pET27bmod vector at the restriction sites BglII-NotI and NotI-BlpI (Figure 2.1.1).

### **2.1.1 Amplification and purification of DNA**

A standard setup (Plasmid Miniprep, Evrogen) was used for isolation and purification of plasmid DNA from *E. coli* cells with the commercially available Cleanup Standard kit (Evrogen) used to isolate and purify DNA fragments from agarose gels or reaction mixtures. Isolation and purification were performed according to the manufacturer's standard protocols. For cell growth, LB liquid nutrient medium (Lysogeny broth, 0.5% yeast extract, 1% tryptone, 1% NaCl) and solid LB agar medium (LB + 1.5% bactoagar) were used.

### **2.1.2 Restrictase ligase cloning**

Cloning using the restriction enzyme ligase method was carried out in two stages. In the first stage, the DNA fragment containing the target site for cloning was restricted using two restriction enzymes. Typically, restriction protocols involved using a temperature of 37 °C and incubation for 1-2 hours. Simultaneously with the original DNA, the vector into which the DNA fragment is planned to be inserted is subjected to restriction. If restriction enzymes require different buffer compositions for functioning, restriction reactions must be carried out sequentially, replacing the buffer between the first and second restriction reactions. When using restriction check enzymes from the Fast Digest (FD) series (Thermo Fisher), there are no problems with buffer incompatibility since their restriction enzymes are optimized to function in the same buffer (FD or FD Green). Restriction enzymes from New England Biolabs were also exploited, and here the Cutsmart buffer was used. The restriction product was isolated and purified from the gel using a Cleanup Standard kit (Evrogen). The compositions of all buffers are unavailable for commercial reasons.

The second stage of restriction enzyme ligase cloning involves a ligation procedure. This was carried out at 22 °C for 10-60 minutes using the DNA ligase of phage T4 (Thermo Fisher). Recommended vector/insert molar ratio varies in the range 1:1 - 1:5. The amount of the enzyme used was based on the manufacturer's recommendations. The ligation mixture was used to transform *E. coli* strain DH5α cells.

### 2.1.3 Preparation of competent cells

In this work, the "calcium" method - named after the main active ingredient (CaCl<sub>2</sub>) - for preparing competent cells was used. This procedure is detailed in Table 2.1.3.1.

**Table 2.1.3.1 Procedure for the "Calcium" protocol for preparing competent cells.**

Day 1	Collect part of the frozen glycerin drain (you need to act quickly, avoiding defrosting) with a previously fired loop.
	Streak the LB agar plate.
	Set the drain immediately to -80 °C.
	Leave plates for 5 minutes, then place upside down in a 37°C incubator for 16-20 hours.
Day 2	Select one colony and subculture in 5 ml LB.
	Grow the culture overnight at 37 °C in a shaker at 250 rpm.
Day 3	Inoculate 1 ml of saturated overnight culture in 100 ml of LB medium.
	All necessary equipment and components are pre-cooled: solutions, centrifuge, pipette tips, Falcons and Eppendorf tubes. If possible, work in a cold room.
	Grow in a shaker at 37 °C until OD <sub>600</sub> = 0.4 (usually 2-3 hours). Place in an ice bath for 10 minutes. [Cells should never touch anything warm after that]
	Transfer the culture to two pre-cooled 50 ml tubes. Centrifuge at 2700 g for 10 minutes at 4 °C.
	Remove the medium, resuspend the cell pellet in 1.6 ml of cooled 100 mM CaCl <sub>2</sub> solution, stirring.
	Incubate on ice for 30 minutes.
	Centrifuge at 2700 g for 10 minutes at 4 °C.
	Remove the medium, resuspend the cell pellet in 1.6 ml of the ice-cold solution of 80 mM MgCl <sub>2</sub> + 20 mM CaCl <sub>2</sub> , and gently shake it on ice.
	Incubate on ice for 20 minutes.
	Combine cells into one tube, add 0.5 ml of chilled 80% glycerin, and shake to mix.
Freeze 100 µl aliquots in liquid nitrogen. Store at -80 °C.	

### 2.1.4 Transformation of plasmid DNA into *E. coli* cells

An aliquot of frozen competent cells was removed from -80 °C storage and transferred to ice. Plasmid DNA or ligase mixture was sterilely added to the cells.

The added DNA solution volume should not exceed 10% of the aliquot's cell suspension volume. The mixture was incubated on ice for 30 minutes, after which the cells were subjected to heat shock (42 °C, ~ 1 minute), and the mixture was incubated on ice for a further 5-10 minutes. 0.4-0.9 ml of LB medium was then added to a test tube with cells. The cells were placed in a shaker for 30-60 minutes, 37 °C, 200-250 rpm. Super Optimal broth with Catabolite repression (SOC) was used instead of LB at this stage increases the number of colonies. The bacteria were then ground on a Petri dish with a solid nutrient medium LB-agar with the required antibiotic's addition (or a mixture of antibiotics). The Petri dish was incubated at 37 °C overnight.

## **2.1.5 Expression of Proteins in *E. coli***

### *2.1.5.1 Co-expression of NpSRII and NpHtrII<sub>137</sub>*

This non-chimeric complex was expressed in *E. coli* strain BL21-AI (DE3) (Stratagene, La Jolla, Calif.). The cell culture was grown at 37 °C in Terrific Broth (TB) medium supplemented with 100mM Na / K-phosphate buffer (pH = 6.7), 25mM ammonium sulfate, and 100mg/L kanamycin. Induction with Isopropyl-β-D-galactopyranoside (IPTG) (1 mM) was carried out at OD<sub>600</sub>=1.6-2.0. Simultaneously with IPTG, purified trans-retinal in ethanol was added to a final concentration of 5-10 μM. Subsequently, the cells were cultured for 3-4 hours at 37 °C.

### *2.5.1.2 Co-expression of NpSRII and NpHtrII by autoinduction*

*NpSRII* and full-length *NpHtrII* were co-expressed in *E. coli* cells, strain BL21 (DE3) (Stratagene, La Jolla, CA). The culture was grown at 37 °C in TB-5052 medium [85] supplemented with 100 mM Na / K-Pi (pH = 6.7), 25 mM ammonium sulfate, and 100 mg/L kanamycin. Upon reaching OD<sub>600</sub>=1.0-1.2, purified trans-retinal in ethanol was added to a final concentration of 5-10 μM. Cells were grown at 20 °C for 16-20 hours (a similar procedure was described in [25] as applied to the expression of a single *NpSRII* alone.



### 2.1.5.3 Co-expression of *NpSRII* and *NpHtrII* in *E. coli* strain BL21-AI

*NpSRII* and full-length *NpHtrII* were also co-expressed in *E. coli* cells, strain BL21-AI (Invitrogen, Carlsbad, Calif.). Here, the culture was grown at 37 °C in TB medium supplemented with 70 mM Na/K-Pi (pH 6.7) and 50mg/L kanamycin. Upon reaching  $OD_{600}=1.6 - 20$ , purified trans-retinal in ethanol was added to a final concentration of 5-10  $\mu$ M as were arabinose (to a concentration of 0.1%) and IPTG (to a concentration of 2 mM). Then the cells were grown for another 4 hours.

In all three scenarios above, the maximum protein yield was approximately 1-2mg per liter of cell culture, with the induction protocol with IPTG and arabinose (see above) showing better reproducibility.

### 2.1.6 Purification of *NpSRII/NpHtrII*<sub>137</sub> and *NpSRII/NpHtrII*

After expression, the cells were pelleted by centrifugation and then resuspended in a buffer containing 25 mM Na/Na-Pi (pH 8.0), 150 mM NaCl, 1 mM PMSF, and cOmplete protease inhibitors (Roche, Switzerland). The cells were lysed using a Microfluidizer M-110P (Microfluidics, Massachusetts, USA). After centrifugation of the lysate, the membrane pellet was collected, then solubilized in 1% n-Dodecyl  $\beta$ -D-maltoside (DDM). The solubilized solution was separated from the membrane residues by centrifugation. The supernatant was used for the subsequent purification of the *NpSRII/NpHtrII*<sub>137</sub> or *NpSRII/NpHtrII* using metal chelate chromatography (HisTrap™ HP 5 ml, GE Healthcare, Illinois, USA) and gel filtration (Superose6 10/300 GL 24 ml, GE Healthcare, USA). As already noted above, using this protocol the final yield is 5 mg and 1-2 mg per liter of medium for *NpSRII/NpHtrII*<sub>137</sub> and *NpSRII/NpHtrII*, respectively.

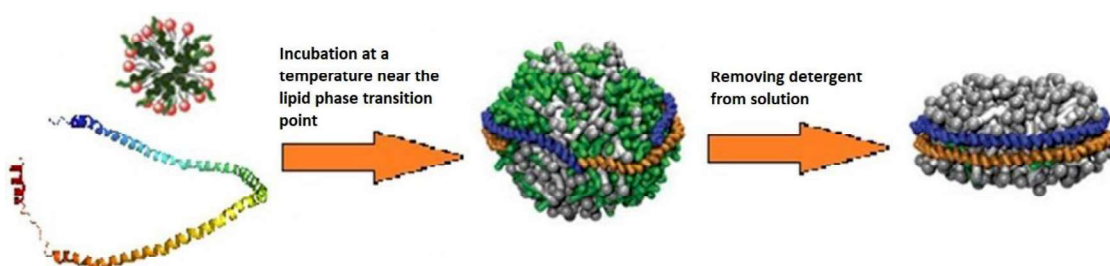
### 2.1.7 Nanodisc Assembly and Reconstitution of Membrane Proteins

The preparation of nanodiscs (NDs) containing embedded proteins requires both the expression and purification of membrane scaffold protein (MSP) and the preparation of lipid micelles using a detergent.

To prepare "empty" nanodiscs (Figure 2.1.7.1), lipid/detergent micelles (usually lipid:detergent ratio is 1:2) are mixed with MSP [102]. To reconstitute membrane protein in NDs a MP solubilized in detergent is added to the mixture of MSP and lipid-detergent micelles. The correct proportions of the above ingredients must be observed. The ratios of lipid molecules to one MSP molecule for the assembly of empty nanodiscs are given in Table 2.1.7.1.

**Table 2.1.7.1 Nanodisc – lipid ratios required to fill the full surface of the nanodisc.**

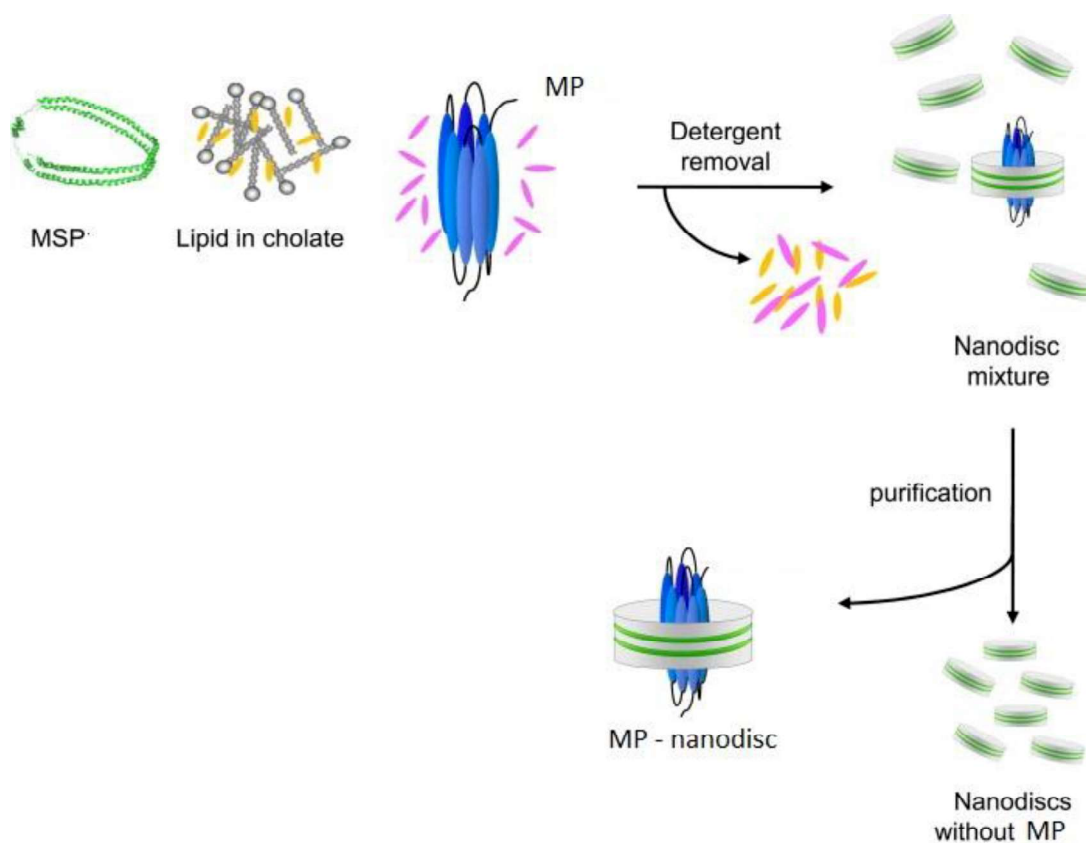
	Lipid molecule		
	POPC	DPPC	DMPC
MSP1D1	65	90	80
MSP1E1D1	85	115	100
MSP1E2D1	105	145	130
MSP1E3D1	130	180	160



**Figure 2.1.7.1 Schematic step-by-step diagram of the assembly of empty nanodiscs. Figure taken from [102].**

The mixture is incubated for 1-2 hours at a temperature close to, but not lower than, the phase transition point of the lipid used and the nanodiscs are formed after the detergent is removed from the solution. This can be done in two ways: dialysis, in which the concentration of the detergent in the solution is reduced by 100 times during a buffer exchange procedure (preferably repeated 3-4 times), or the use of Bio-Beads (Bio-Rad), which adsorb detergents on their surface. A combination of the two methods can be used to better remove detergent. The last stage in nanodisc preparation is gel filtration, which separates the suspension components by size. This stage is necessary to remove aggregates.

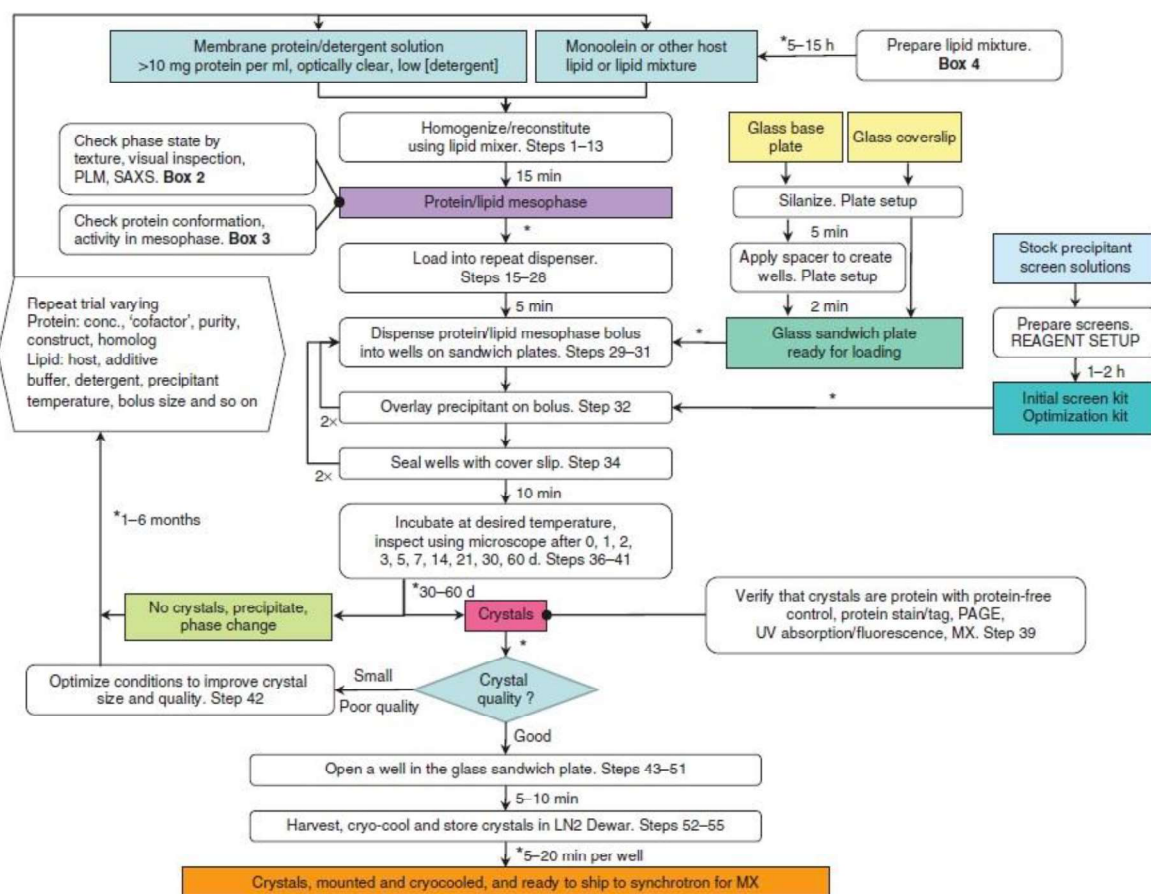
For embedding MPs in nanodiscs (Figure 2.1.7.1), a similar protocol is used as for assembling empty nanodiscs, but with the addition of a MP. Separation of protein-containing and empty nanodiscs is also performed using gel filtration. The calculation of the amount of lipid molecules to use should take into account the displacement of lipid molecules by the transmembrane region of the protein [103].



**Figure 2.1.7.2** Schematic step-by-step diagram of the assembly of nanodiscs containing a membrane protein. *Figure adapted from [103].*

## 2.2 Crystallization by the *in-meso* method

The crystals obtained in this study were obtained by crystallization using the *in meso* method. The general scheme of *in meso* crystallization is described in [104] and is shown in Figure 2.2.1.



**Figure 2.2.1** Schematic step-by-step diagram of the *in meso* crystallization of membrane proteins. Figure taken from [104].

The more practical aspects of *in meso* crystallization procedure are shown in Figure 2.2.2 and Figure 2.2.3. Two Hamilton syringes (Hamilton Research): one containing the MP (in this work a concentration of 20 mg/ml was used), the other containing monoolein (MO) are prepared. The contents of the two syringes are then mixed to external homogeneity (color and composition).

Once the contents of the two syringes are satisfactorily mixed, droplets are dispensed into a 96-well LCP plate (Qiagen, Germany), at 100nL of mesophase per well, using an NT8 crystallization robot (Formulatrix, USA) at the Partnership for Structural Biology (PSB; Grenoble, France). The resulting drops are then covered with 800nL of precipitant solution. In this work, standard kits of precipitant solutions, QCP1 and QCP2 (Qiagen, Germany), were used.

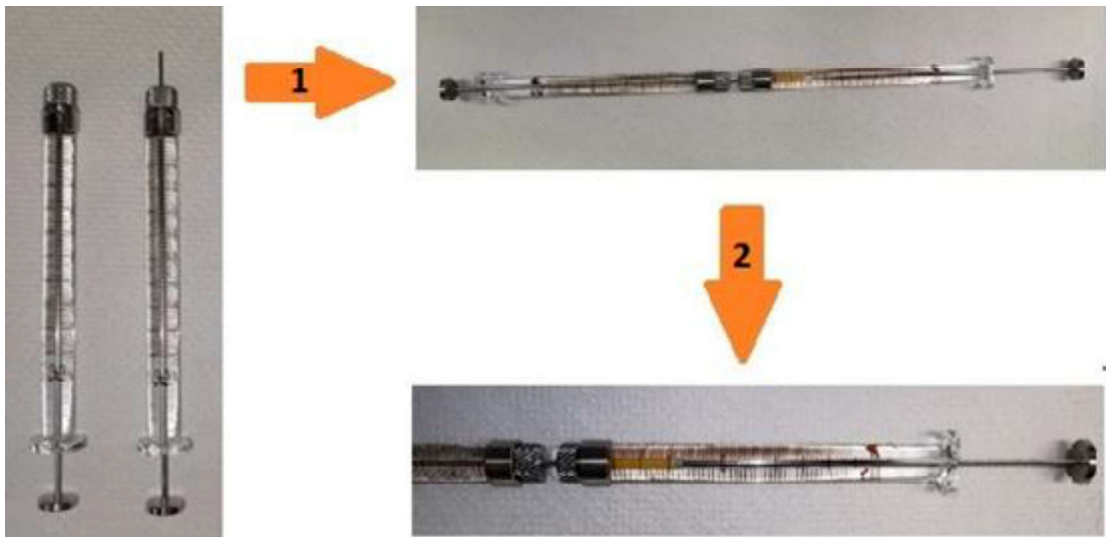


Figure 2.2.2 The procedure for preparing proteins for in meso crystallization.

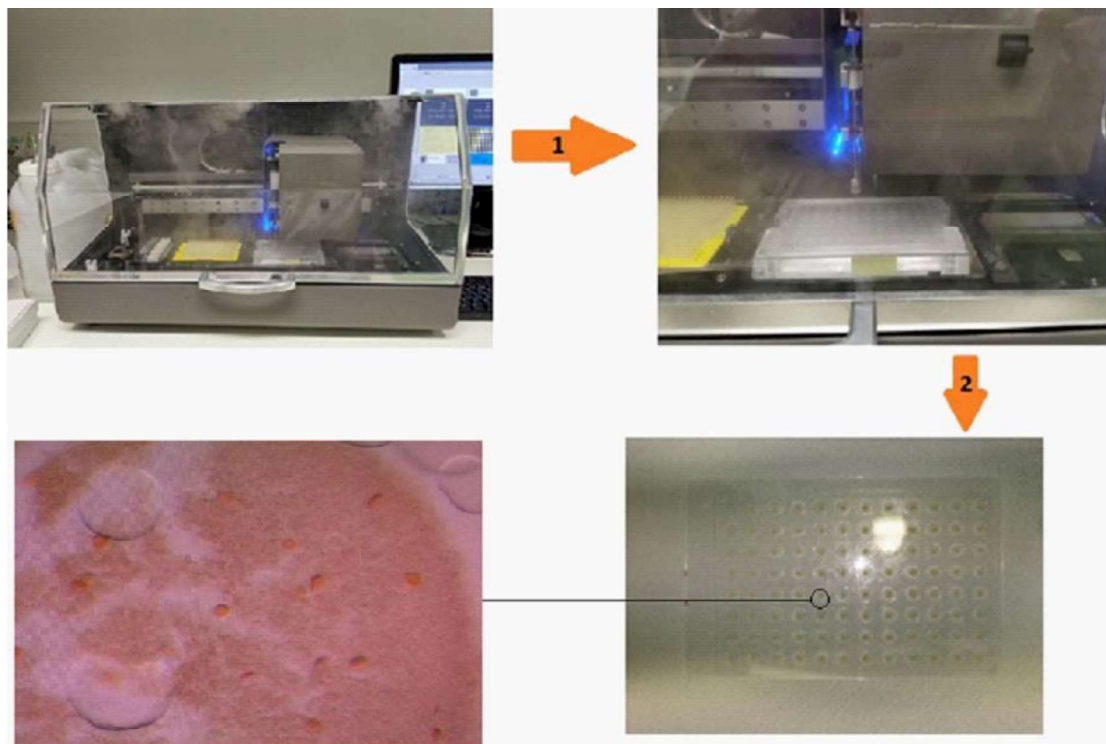


Figure 2.2.3 The procedure of dispensing protein mixtures onto LCP plates

### 2.3 Negative stain electron microscopy

The *NpSRII/NpHtrII* protein solution in a 150 mM NaCl buffer, 25 mM Na / K-Pi, and 0.01% DDM (pH 8.0) was inserted between mica and a carbon film and then stained with a solution of 2% sodium silicotungstate. The carbon film was then applied to a copper mesh and dried at room temperature. Micrographs of negatively stained particles were obtained at low dose at the Partnership for Biology (PSB)

Electron Microscopy Platform (<https://www.psb-grenoble.eu/spip.php?rubrique34>) using a Tecnai F20 (FEI) electron microscope with a FEG (Field Emission Electron Gun) at 200 kV, equipped with a Ceta (FEI, The Netherlands) 4kx4k camera. For visualization, the defocus value was set to  $\sim 4 \mu\text{m}$ .

## 2.4 Cryo-electron microscopy

Cryo-electron microscopy (Cryo-EM) measurements were carried out at Maastricht University using a Tecnai Arctica microscope (FEI, The Netherlands) (Figure 2.4.1), at Groningen University with a Talos Arctica cryo-electron microscope (FEI, The Netherlands) and at The Netherlands Center for Electron Nanoscopy (NeCEN, The Netherlands) with a Titan Krios K2 cryo-electron microscope. For blotting, carbon-copper meshes were used, the sample was applied using a Vitrobot™ device (ThermoFisher).

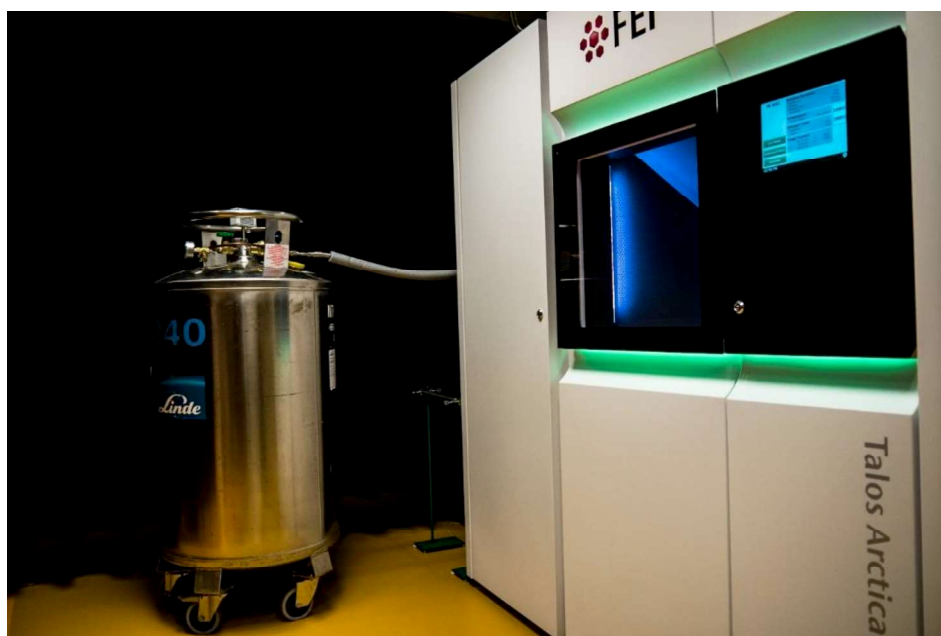


Figure 2.4.1 Talos Arctica electron microscope (Groningen University, Groningen, The Netherlands).

## 2.5 Small-angle scattering measurements

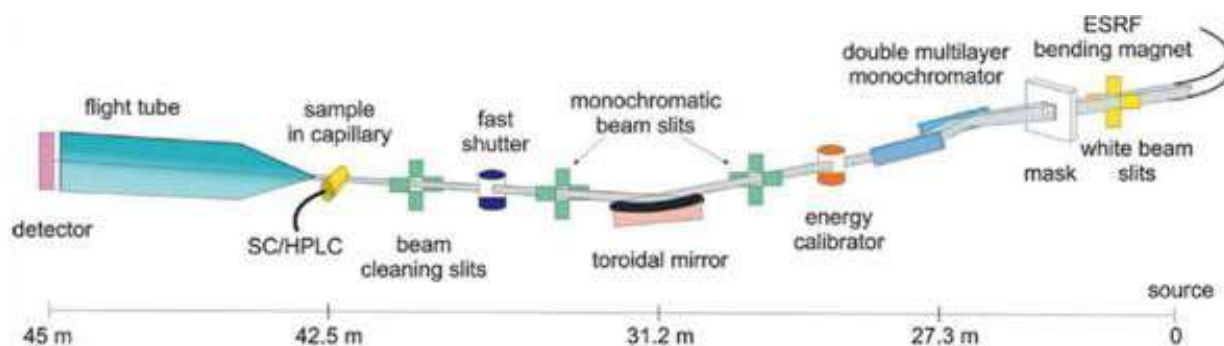
### 2.5.1 SAXS measurements at BM29, ESRF, Grenoble, France

Beamline BM29 (Figure 2.5.1) of the European Synchrotron Radiation Facility (ESRF) is a tool designed for Small-angle X-ray Scattering (SAXS) measurements from solutions of biological macromolecules [105]. The X-ray



wavelength range is  $\lambda=0.82 - 1.77 \text{ \AA}$ . The flux density in the collimated beam is  $1.3 \times 10^{13}$  photons/sec at  $\lambda=1.13 \text{ \AA}$ . The attainable range in modulus of the scattering vector ( $q$ ) is  $0.001 - 0.5 \text{ \AA}^{-1}$ . To set the wavelength, a double multilayer monochromator is used (energy bandpass  $(\Delta E/E) \sim 10^{-2}$ ).

The beamline's sample changer, Pilatus 1M 2D detector and the connecting vacuum tube are located on the marble table. BM29 is also equipped with a chromatograph, which allows SAXS measurements of various fractions to be carried out directly following gel filtration. The diameter of the capillary with the irradiated sample is 1.8 mm. The dimensions of the collimated beam are  $500 \times 500 \mu\text{m}^2$ .



**Figure 2.5.1.1** Scheme of BM29 BioSAXS, ESRF. *Figure taken from [105].*

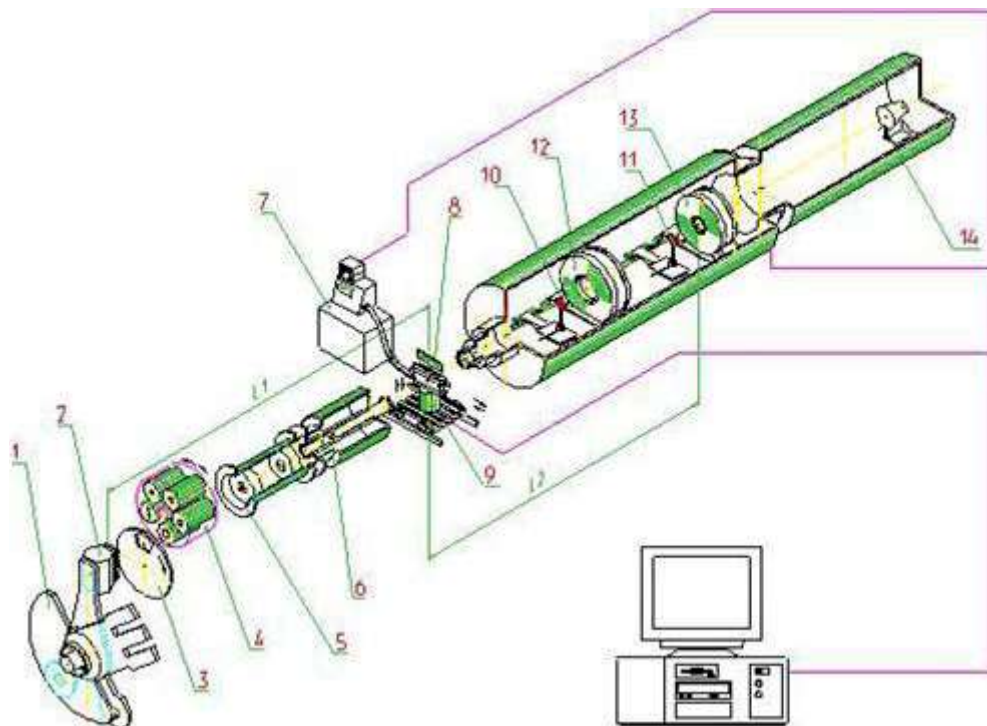
## 2.5.2 SANS measurements at YuMO, IBR-2, JINR, Dubna, Russian Federation

The fast pulsed reactor (IBR-2) operates as part of the Frank Laboratory of Neutron Physics (FLNP) of the Joint Institute for Nuclear Research (JINR, Dubna, Russia) [106]. It is the only nuclear reactor in the world with a movable reflector.

The movable reflector is a complex mechanical system. It consists of a main and additional rotating movable reflectors (OPO and DPO). The rotors of OPO and DPO rotate in opposite directions at different speeds, the rotation frequencies being 5 and 10 Hz, respectively. At the moment when both reflectors are aligned, a power pulse is generated near the reactor zone. Next, the neutrons hit the water moderator, the flux of which is  $\sim 10^{16}$  n/(cm<sup>2</sup> x sec). The peak power is 1850 MW. These indicators make the IBR-2 one of the most powerful research reactors in the world.

At the exit of the moderator the neutrons have a Maxwellian momentum distribution, the peak position and width of which correspond to the neutron wavelength range of 0.1 - 10 Å. The non-monochromaticity of the beam is turned into an advantage due to the use of the time-of-flight technique - scanning along the wavelength is carried out by measuring the neutron flight time from the moment of the flash.

YuMO is a small-angle neutron scattering (SANS) installation operating at IBR-2 (Figure 2.5.2.1). A vanadium standard is used for intensity calibration, and a two-detector system is used to expand the range of the values of the scattering vector modulus which can be measured [107]. YuMO has the following characteristics [108]: the thermal neutron flux on the sample  $4 \times 10^7$  n/cm<sup>2</sup>/sec, the range of used wavelengths is  $\lambda=0.5 - 8$  Å, an accessible  $q$  range of 0.001 - 0.5 Å<sup>-1</sup>, the beam diameter is on the sample can be selected in the range of 8 - 22 mm.



**Figure 2.5.2.1 Schematic of the YuMO, IBR-2, JINR facility.** 1 - double reflector; 2 - reactor zone with moderator; 3 - breaker; 4 - the first collimator; 5 - vacuum tube; 6 - the second collimator; 7 - thermostat; 8 - sample rack; 9 - goniometer; 10-11 - vanadium (Vd) standard; 12 - ring-wire detector; 13 - position-sensitive detector "Volga"; 14 - direct beam detector.



### **2.5.3 Small-angle scattering data processing**

The SAXS and SANS data collected at BM29 (ESRF) and YuMO (FLNP) and were processed using the ATSAS [109] and BioXTAS RAW [110] software packages. The pair correlation function  $p(r)$  and the normalized  $I(q)$  were obtained using the GNOM program, which implements the inverse Fourier transform method [111]. The CRY SOL and CRYSON [70, 112] programs were used to calculate theoretical solution scattering curves from atomic models of macromolecules and fit these to experimental data. MEMPROT [69] (Figure 1.4.2) was used to construct a pseudo-atomic model of the detergent belt surrounding transmembrane region of the proteins studied and to approximate the experimental SAXS data using hybrid models combining pseudo-atomic models of the detergent belt and atomic models of proteins. Before starting the MEMPROT program, the center of the transmembrane part of the protein was placed at the origin, and the direction of the normal vector to the membrane plane was selected along the z axis.

## **2.6 Microbial rhodopsins with Noble Gases**

### **2.6.1 Protein expression and purification**

KR2, tmbR and MAR were obtained from the Gordel'iy group, ICS-6, FCJ, Julich, Germany. Protein expression and purification protocols are standard for rhodopsins and are described in the following publications: tmbR [113], KR2 [26], MAR [114].

### **2.6.2 Crystallisation**

All crystals were grown using the *in meso* approach (see above; [115]) used in previous work [21]. MAR, tmbR, and KR2 in their respective crystallisation buffers were added to the monoolein-formed lipidic phase (Nu-Chek Prep, USA). The protein-LCP (lipidic cubic phase) mixtures (100-nl aliquots for all proteins) were spotted on 96-well LCP glass sandwich plates (Marienfeld, Germany) and covered with 800 nl of precipitant solution using the NT8-LCP crystallization robot

(Formulatrix, USA) from PSB, Grenoble, France. Crystals were grown at 20 °C and reached their final sizes within 1 to 8 weeks.

The best crystals of tmbR were obtained using 3.0 M Ammonium Sulfate and 0.1 M Sodium Acetate (pH 4.6) (Qiagen, Germany) as the precipitant solution and grew to 50-100  $\mu\text{m}$  in size. For KR2, the best crystals were obtained using 2.0 M sodium malonate (pH 4.3) (Hampton Research, USA) as the precipitant solution and grew up to 70-100  $\mu\text{m}$  in size. The best crystals of MAR were obtained using 2.6 M  $(\text{NH}_4)_2\text{SO}_4$  and 0.1 M sodium acetate pH 5.2 (Qiagen, Germany) as the precipitant and were 100 to 150  $\mu\text{m}$  in size.

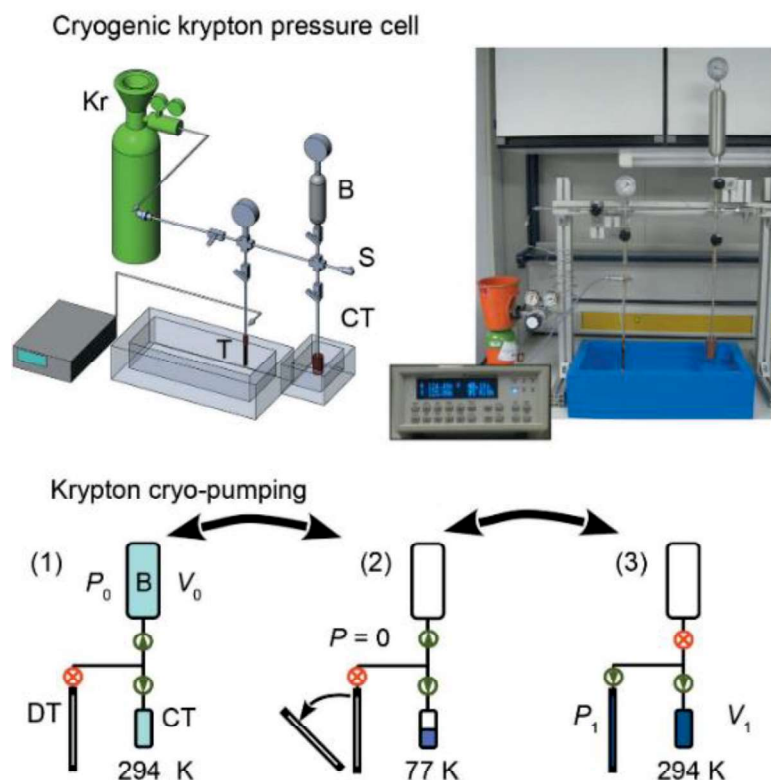


**Figure 2.6.2.1** Photographs of crystals of microbial rhodopsins obtained during this work. *A*, tmbR (purple, plate form). *B*, KR2 (violet, plate form). *C*, MAR (red, rod form).

Once crystals reached their final size, crystallization wells were opened as described elsewhere [13]. For cryo-protection for native data collection, drops containing the protein-mesophase mixture were covered with 50  $\mu\text{l}$  of the respective precipitant solution before they were mounted in MicroMounts (MiTeGen, USA), flash-cooled, and stored in liquid nitrogen.

### 2.6.3 High-pressure derivatization

High pressure derivatization with noble gases was carried out the ESRF's High Pressure Macromolecular Crystallography (HPMX) facility (Figure 2.5.3.1) [82].



**Figure 2.6.3.1** The ESRF device for high pressure noble gas derivatization. *Figure taken from [82].*

Cryo-protected crystals were mounted into the pressurizing device which was then used derivatize the crystals using a pressurized atmosphere of the noble gas desired. Ar derivatization was performed using an atmosphere pressurized at ~2000 bar, for Kr derivatization the pressure used was ~200 bar.

## 2.6.4 Crystallographic data collection

X-ray diffraction data collection was carried out at ESRF beamlines ID23-1 [87] and ID29 [116]. Protein crystals were mounted on SPINE standard sample holders [117] (for data collection at 100 K) or in capillaries (room-temperature data collection). Crystals were located and their diffraction properties characterized using X-ray mesh scans analyzed using Dozor-MeshBest [118, 119]. The X-ray beam size at the sample position was chosen to be in the range from 10 to 50  $\mu\text{m}$  FWHM in accordance with Dozor-MeshBest analysis. Experimental parameters for optimal data collection were designed using the program BEST [120]. Diffraction data were recorded at  $\lambda=1.85 \text{ \AA}$  and  $\lambda=0.972 \text{ \AA}$  for Ar-derivatized crystals and at  $\lambda=0.861\text{-}0.863 \text{ \AA}$  for crystals derivatized with Kr. Diffraction data from native (i.e. non-

derivatized) crystals were recorded at  $\lambda=0.972$  Å. In all cases diffraction images were recorded on a Pilatus 6M detector (Dectris, Switzerland) and processed with XDS [121] and XSCALE [121], with the latter also being used to merge and scale different data sets. Structure factor amplitudes, anomalous differences ( $|F_{hkl} - F_{-h-k-l}|$ ,  $\Delta F_{\text{ano}}$  (DANO)) and  $R_{\text{Free}}$  labels were generated using the CCP4 programs POINTLESS and AIMLESS [122].

### 2.6.5 Structure solution and refinement

The native crystal structures of the same crystal forms as featured in this work had been solved previously in other works, and the protein chains from those structures were used as initial models for crystal structure refinement (MAR: PDB code 5JSI [114]; KR2: PDB code 6RF7 [123]; bR: PDB code 3MBV [124]). Anomalous difference Fourier maps were calculated with FFT [125] from the CCP4 package using DANO and  $(\alpha_{\text{calc}} + 90^\circ)$  as coefficients ( $\alpha_{\text{calc}}$  = calculated phases from the final refined models). In cases where multiple data sets were available from a given crystal form, anomalous difference maps were generated for each data set and those which showed consistency in anomalous difference map peaks were merged and refinement carried out with the merged data. In the case of tmbR-Ar crystals the anomalous dataset recorded at  $\lambda=1.85$  Å was used to calculate anomalous difference maps, with that recorded at  $\lambda=0.972$  Å (and to higher resolution) used to carry out refinement. Final structural models were obtained by alternating cycles of manual building in Coot [126] followed by refinement in REFMAC5 [127]. Omit ( $|mF_{\text{obs}} - DF_{\text{calc}}|$ ;  $\alpha_{\text{calc}}$ ) maps were used to verify the presence of lipid fragments, ions, and water molecules. Individual noble gas atoms were located by choosing peaks in the anomalous difference maps (in cases with multiple anomalous maps from different data sets each map was taken into account) greater than the  $3 \times$  r.m.s. level and positioned in accordance with peaks in electron density ( $|2mF_{\text{obs}} - DF_{\text{calc}}|$ ;  $\alpha_{\text{calc}}$ ) maps. Their occupancies were then refined in REFMAC5 maintaining their B-factors similar to those of the surrounding protein structure.

## 2.6.6 Molecular dynamics simulations

The systems used for molecular dynamics simulations consisted of protein chain monomer models of tmBR, KR2, and MAR. In all simulations, the proteins were embedded in a model detergent bilayer consisting of DPPC/DLPC lipids and solvated with water molecules with counter ions using TIP3P [128] through the CHARMM-GUI web-service [129]. Following this process, a small fraction (2%) of water molecules were replaced by atoms of specific noble gases.

The recommended CHARMM-GUI protocols were followed for the initial energy minimization and equilibration of the systems. The atoms of proteins and lipids in the systems were subject to harmonic positional restraints. The steepest descent minimization (5000 steps) was followed by a series of short equilibration simulations (25-50 ps) in the NPT ensemble [130] using the Berendsen thermostat and barostat [131] with the restraints on lipids gradually released. For the production simulations, the Nose–Hoover thermostat [132] and Parrinello–Rahman barostat [133] were used. Here, the C $\alpha$ -atoms of the protein backbone were constrained using the harmonic potential with the force constant of 100 kJ/mol/nm<sup>2</sup> to prevent large fluctuations of the protein structure, which could obscure further analysis. The temperature and pressure were set to 323.15 K and 1 bar with temperature and pressure coupling time constants of 1.0 ps<sup>-1</sup> and 0.5 ps<sup>-1</sup>, respectively. All MD simulations were performed with GROMACS version 2020.2 [134]. A time step of 2 fs was used for all production simulations. The CHARMM36 force field [135] was used for the proteins, lipids, and ions, while the parameters for the noble gases were obtained from [136]. All production simulations were run for 100 ns. The density maps for the noble gases were calculated using the volmap utility in VMD [137] after roto-translational alignment of a protein and the treatment of periodic boundary conditions. Only the last 90 ns of the production simulations were used for the analysis.

Density peaks were identified on the density maps sampled on a 0.5 Å grid, above (mean + 5 x r.m.s.d.) level as local maxima in a moving window 2 Å in size.

Peaks within 4 Å of crystallographically observed noble gas sites were counted as “atom matches”. Density peaks clusters were identified by hierarchical cluster analysis using a cut-off distance of 5 Å between cluster centres. Clusters containing no “atom matches” were counted as “not present”.

## 3. Results

### 3.1 Investigation of *NpSR*II/*NpHtr*II protein complex solubilized in DDM

#### 3.1.1 Gene Optimization

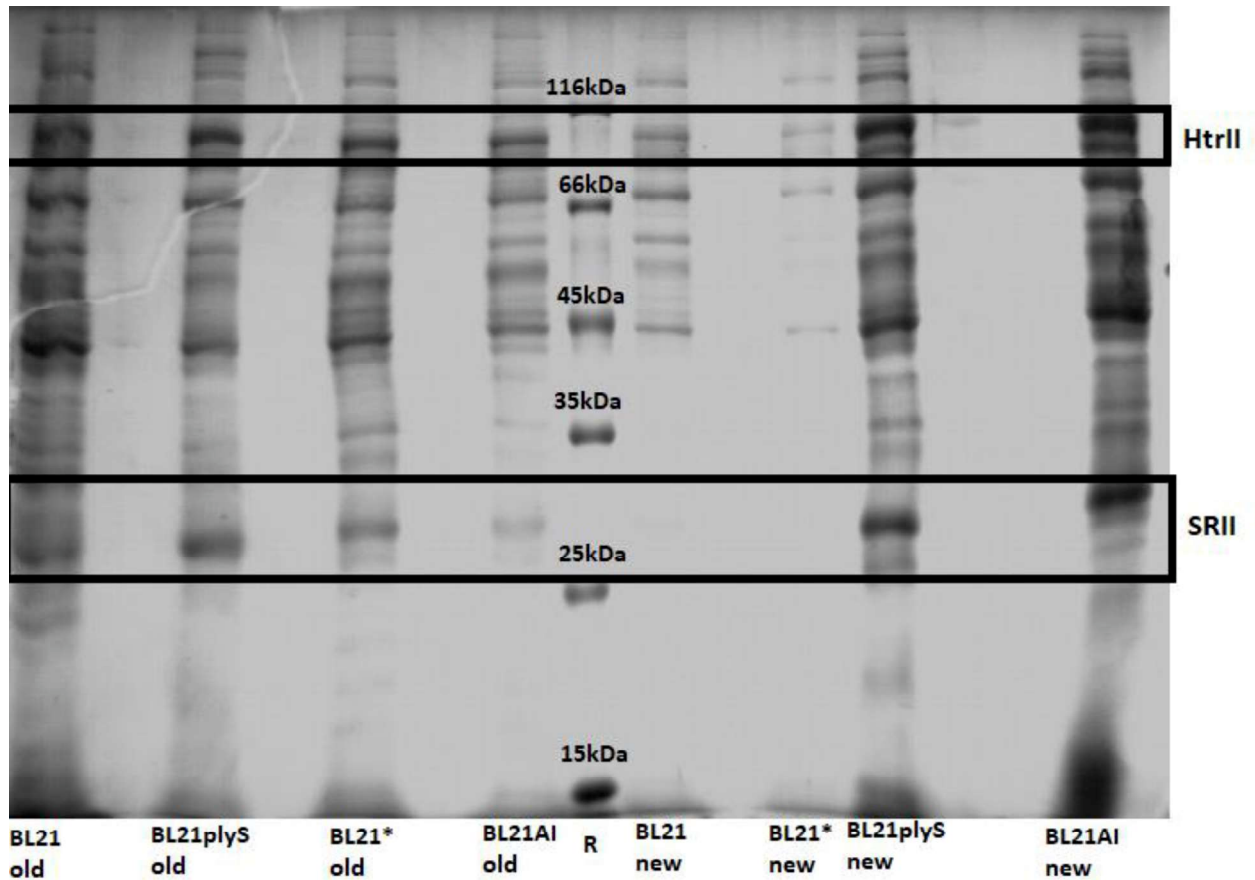
The original genes used in this work were obtained from our collaborators in the Borshchevskiy Team from the Moscow Institute of Physics and Technology (MIPT, Dolgoprudny, Russia). For co-expression, they were encoded in a pET27bmod vector containing XhoI and NdeI restriction sites. Protein expression was carried out in *E. coli* cells, Rosetta2 strain. The protein yield was 5-6 mg from 6 liters of medium, but this amount of protein is not enough to carry out experiments planned in this thesis work. For this reason, it was decided to optimize the genes of both *NpSR*II and *NpHtr*II for expression in *E. coli*, without modifying their amino acid sequences. The optimized genes were inserted into separate (pSC\_SRIIStrep; pSC\_HtrIIHis<sub>6</sub>) small plasmids and then inserted into a co-expression vector (pET27bmod\_SRIIStrep\_HtrIIHis<sub>6</sub>). A hexa-histidine (His<sub>6</sub>) purification tag was added to the C-terminus of *NpHtr*II, while a streptavidin (Strep) tag was encoded to the C-terminus of *NpSR*II. This was done to get a cleaner sample before the SEC.

#### 3.1.2 Optimization of expression and purification protocols

Test expression was performed under conditions that were selected for the old plasmid, but the yield of the protein was significantly lower than previously. In this regard, it was decided to search for optimized expression conditions (selection of cell strain, search for the best OD<sub>600</sub> to induce cell culture and IPTG concentrations). Here two strategies were adopted:

1. A comparative analysis of *E. coli* strains for the new and old expression plasmids was carried out. Cells were grown in TB-medium (80 ml) at T = 37 °C to OD<sub>600</sub> ~ 1.5, then IPTG was added up to 2mM as was antibiotic 100µM if necessary (ampicillin plasmid) and growth was continued at the same temperature for 4 hours. At the end of the process, cells were selected and further solubilized in TES-buffer (100 mM TRIS, 30 mM EDTA, 1% SDS) for 12% sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis (Figure

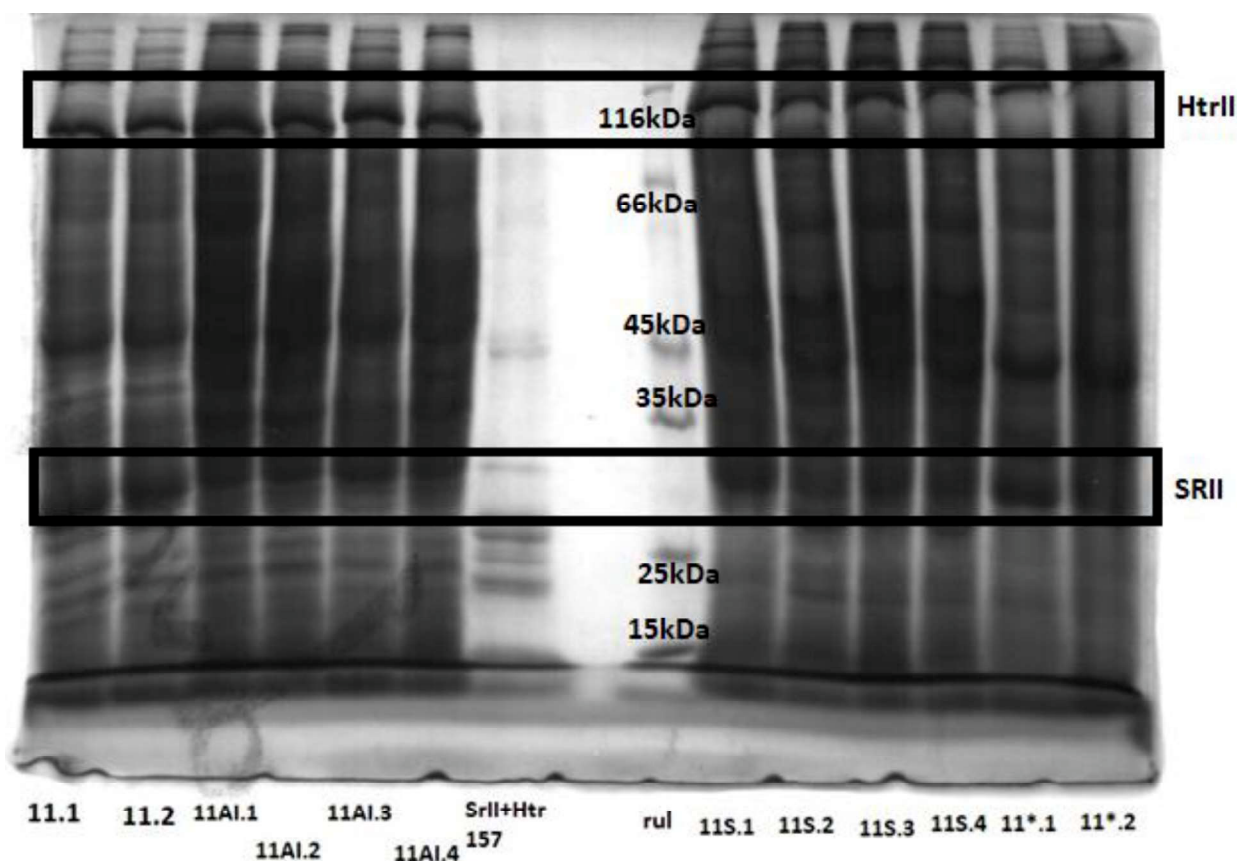
3.2.2.1). According to the results of the gel analysis, the most promising kanamycin plasmids were selected and another test expression was performed.



**Figure 3.1.2.1 SDS-PAGE analysis of protein expression.** 12% SDS-gel analysis of old and new (i.e. *E. Coli optimized*) plasmids in four *E. Coli* strains (*BL21(DE3)*, *BL21(DE3)plyS*, *BL21\*(DE3)*, *BL21(DE3)AI*). “R” is molecular weight marker.

2. All previously tested *E. coli* strains were grown in TB-medium (70 ml) at of 37 °C to  $OD_{600} \sim 2$  (approximately 4 hours). For experiments in the cell strains BL21 and BL21\* two protocols were then used: add IPTG up to 2mM and leave to grow at 37 °C; add IPTG to 2 mM and continue cell growth at 25 °C. For the cell strains BL21PlyS and BL21AI, 4 paths were chosen: IPTG 2mM, 1% arabinose, cell growth at 37 °C; IPTG 2mM, cell growth at 37 °C; IPTG 2mM, 1% arabinose, cell growth at 25 °C; IPTG 2mM, cell growth at 25 °C. In all cases the culture was grown for 4 hours. The whole culture was then centrifuged and the same mass of cells (40 mg) was taken to accurately determine the amount of protein produced. Next, solubilization was performed in TES-buffer following which the samples were applied on a 12% SDS gel (Figure 3.2.2.2).



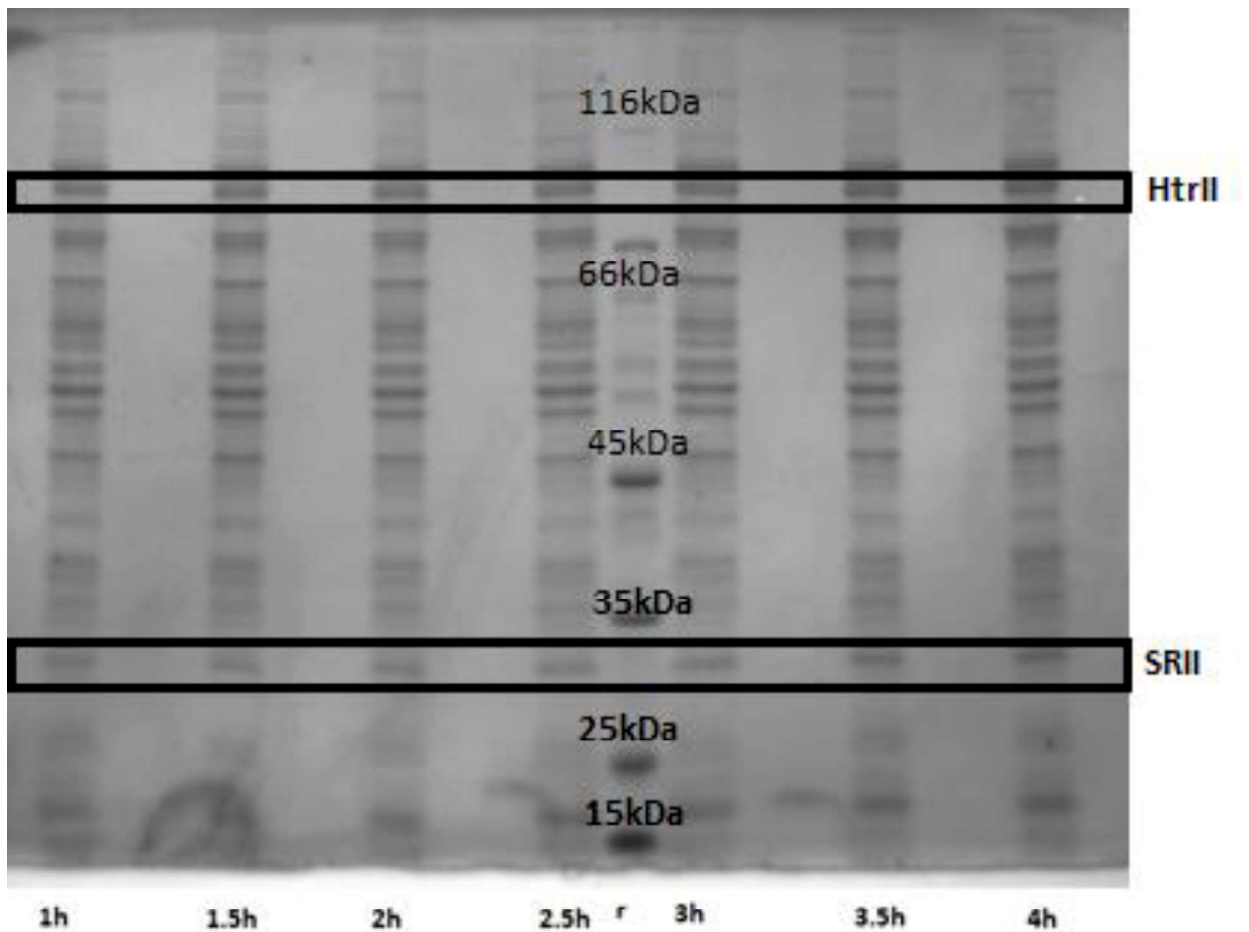


**Figure 3.1.2.2 SDS-PAGE analysis of further protein expression tests.** 12% SDS-gel analysis for *E. Coli* strains (*BL21(DE3)* (11.1, 11.2), *BL21(DE3)plyS* (11S.1, 11.2), *BL21\*(DE3)* (11\*.1, 11\*.2, 11\*.3, 11\*.4) and *BL21(DE3)AI* (11AI.1, 11AI.2, 11AI.3, 11AI.4). See main text for conditions used. “rul” is molecular marker.

According to the results, the most suitable cell strain was BL21AI coupled with cell growth at 37 °C and the addition of 2 mM IPTG supplemented with 1% arabinose to induce expression. For larger scale cell-growth/expression, cells were grown in TB-medium (1L) at 37 °C, 250 RPM. After 3 hours cell growth, IPTG and arabinose were added at  $OD_{600} = 2.2$ . An hour after induction  $OD_{600}$  was measured. This measurement was repeated every 30 minutes until 4 hours had elapsed post-induction. These measurements (Table 3.2.2.1) coupled with 12% SDS-PAGE analysis (Figure 3.2.2.3) confirmed that cell growth stopped 3 hours after induction.

**Table 3.1.2.1  $OD_{600}$  measurement as function of post-induction time**

Time after induction	1h	1.5h	2h	2.5h	3h	3.5h	4h
$OD_{600}$	3.44	3.76	4.00	4.04	4.28	4.28	4.20



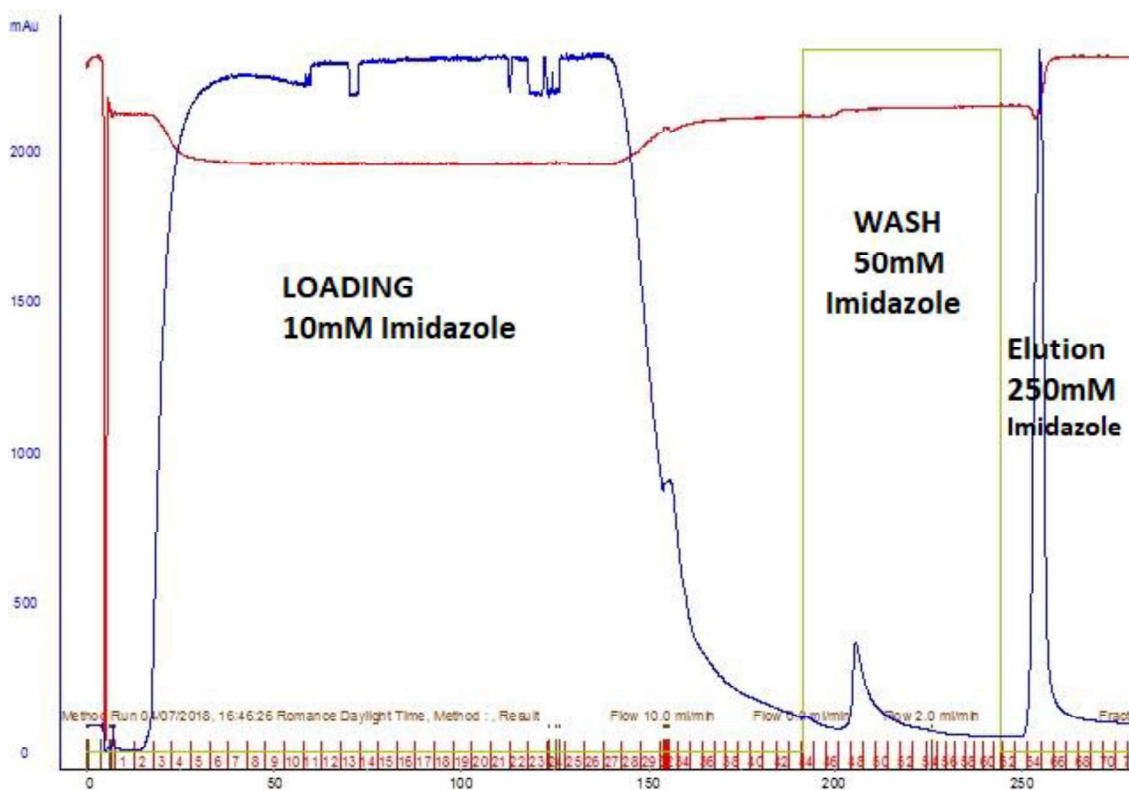
**Figure 3.1.2.3 SDS-PAGE analysis of protein expression as a function of post-induction time.** *Column 'r' is a molecular weight marker. SrII<sub>Strep</sub> and HtrII<sub>His6</sub> are indicated with rectangles.*

After deposition, the cells were opened using a French Press. Here, cells were resuspended in a lysis buffer (150 mM NaCl, 25 mM NaPi, 1mM EDTA, 0.5 mM PMSF, pH 7.5) at a ratio of 2-3 ml of buffer per 1 g of cells and the suspension was pumped through a pressure cell at 2200 Mpa (2 repetitions).

Membranes containing the target protein(s) were precipitated by centrifuging at 60,000g for 3 hours at 10 °C. The pellet was resuspended in a solubilization buffer comprising 150 mM NaCl, 25 mM NaPi, pH 7.5 supplemented with 1-2% DDM (n-Dodecyl β-D-maltoside). Solubilization was carried out for 10-20 hours at 4 °C, with constant vigorous stirring. The membranes that were not solubilized were precipitated by centrifugation, also at 60,000g, for 2 hours at 10 °C.

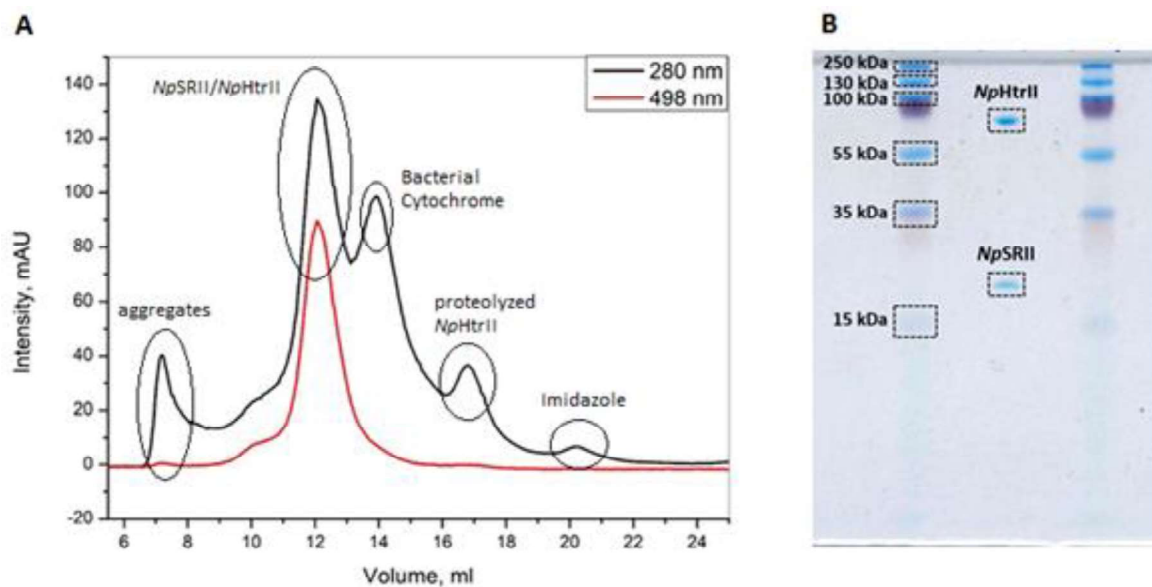
For purification, the supernatant was applied to a Ni-NTA carrier column (His-Trap) at a rate of 0.5 to 1.0 ml / min, with a preliminary addition of 10 mM

Imidazole. The binding of *NpSrIIStrep/NpHtrIIHis<sub>6</sub>* with divalent nickel ions is due to the presence of the C-terminal histidine tag added to *NpHtrII*. After applying the supernatant, the column was washed with 50 mM imidazole to remove proteins non-specifically associated with Ni-NTA. Then, the target protein was washed from the column with buffer containing 25 mM NaPi (pH 7.5), 150 mM NaCl, 250 mM imidazole and 0.05% DDM. (Figure 3.1.2.4)



**Figure 3.1.2.4 Purification of *NpSrIIStrep/NpHtrIIHis<sub>6</sub>* via nickel affinity chromatography.** *Ni-NTA* curve (blue) obtained by monitoring the  $\lambda=280$  nm absorption profile. The procedure consists of three stages: Loading, Wash and Elution.

After the Ni-NTA step, *NpSrIIStrep/NpHtrIIHis<sub>6</sub>* was further purified using size exclusion chromatography (SEC) on an Akta Purifier (GE, USA) equipped with a Superose6, 24 ml column (GE, USA). The equilibrating buffer comprised 150 mM NaCl, 15 mM NaPi (pH 7.5), 0,05% DDM. Proteins were separated by absorption profile at  $\lambda=498$  nm (photoactive wavelength of SRII) (Figure 3.1.2.5.A). Fractions containing *NpSrIIStrep/NpHtrIIHis<sub>6</sub>* were collected and concentrated to that required for crystallization (~30 mg/ml). That the purified protein was truly a *NpSrIIStrep/NpHtrIIHis<sub>6</sub>* complex was confirmed by SDS-PAGE analysis (Figure 3.1.2.5 B).



**Figure 3.1.2.5 Purification of *NpSrIIStrep/NpHtrIIHis<sub>6</sub>* using gel filtration.** “A” shows the gel filtration profiles for *NpSrII/NpHtrII* after Ni-NTA. The black curve is absorption at  $\lambda = 280$  nm absorption profile, the red curve is absorption at  $\lambda = 498$  nm (photoactive wavelength of SrII) absorption profile. “B” is 12% SDS-gel of *NpSrII/NpHtrII* after SEC.

### 3.1.3 Single particle cryo-electron microscopy analysis of *NpSrII/NpHtrII*

In parallel with crystallographic studies (see §3.1.4 below) electron microscopy experiments were carried out.

#### 3.1.3.1 Grid preparation and derivatized

The first stage of this approach was a search for optimal parameters for preparing grids. In this regard, a work plan (Table 3.1.3.1.1) was drawn up:

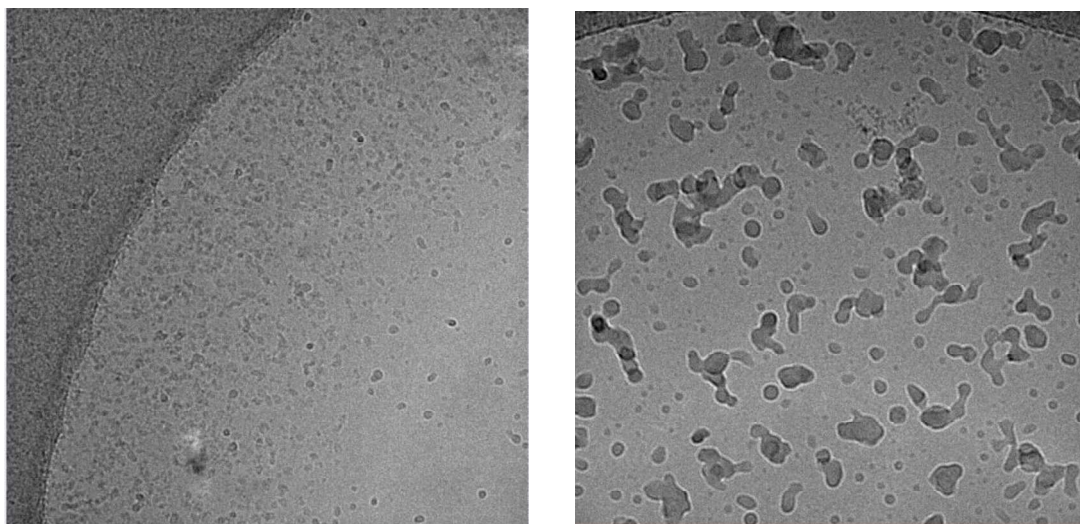
- Low salt conditions (150 mM NaCl). Preparation of grids under initial conditions (detergent, Cu-C grid) at different concentrations (1.0 and 0.5 mg/ml)
- Preparation of high salt (4M NaCl) grids for the available protein (detergent, Cu-C grid) at different concentrations (2.1 mg/ml)
- Testing of a selection of conditions for glow discharging and applying protein to the grid (blotting angle, humidity, temperature; number of depositions between blotting)

**Table 3.1.3.1.1 A summary of the conditions used in grid preparation for cryo-EM studies of wild-type *NpSRII/NpHtrII*.**

	Detergent	Grid type	Buffer	Concentration	Blot conditions (angle, time, humidity, temperature)	Glow
1	DDM	Cu-C	150mM NaCl 25mM NaP 0.05% DDM pH 8.0	1.0 mg/ml	0° 1s 100% 25 °C	2 min, 20mA, negative
2	DDM	Cu-C	150mM NaCl 25mM NaPi 0.05% DDM pH 8.0	0.5 mg/ml	0° 1s 100% 25 °C	2 min, 20mA, negative
3	DDM	Cu-C	4M NaCl 25mM NaPi 0.05% DDM pH 8.0	2.0 mg/ml	0° 1s 100% 25 °C	2 min, 20mA, negative
4	DDM	Cu-C	4M NaCl 25mM NaPi 0.05% DDM pH 8.0	1.0 mg/ml	0° 1s 100% 25 °C	2 min, 20mA, negative
5	DDM	Cu-C R2/2 300	150mM NaCl 25mM NaPi 0.05% DDM pH 8.0	1.0 mg/ml	0° 2s 100% 22 °C	30 s, 20mA, negative
6	DDM	Cu-C R2/2 300	150mM NaCl 25mM NaPi 0.05% DDM pH 8.0	2.0 mg/ml	0° 2s 100% 22 °C	30 s, 20mA, negative
7	DDM	Cu-C R2/2 300	150mM NaCl 25mM NaPi 0.05% DDM pH 8.0	4.0 mg/ml	0° 2s 100% 22 °C	30 s, 20mA, negative
8	DDM	Cu-C R2/2 300	150mM NaCl 25mM NaPi 0.05% DDM pH 8.0	7.0 mg/ml	0° 2s 100% 22 °C	30 s, 20mA, negative

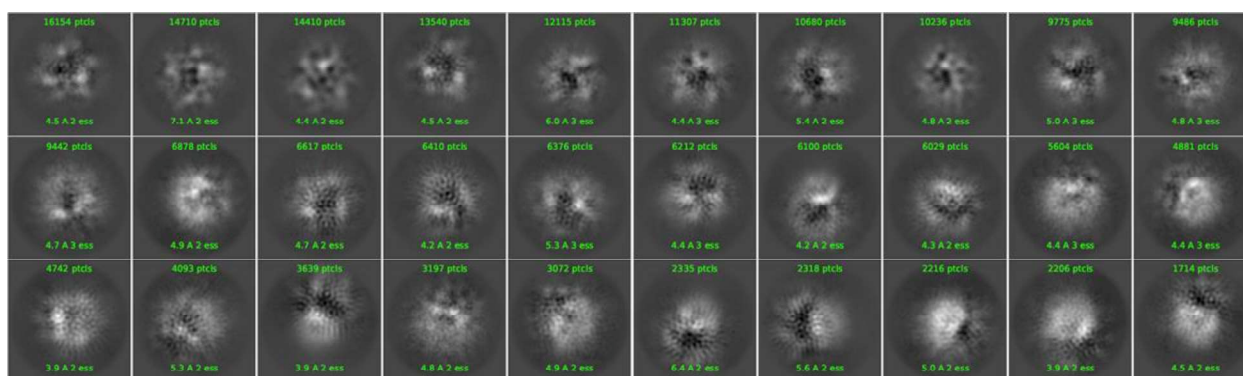
Example micrographs are shown in Figure 3.1.3.1.1. Large numbers of detergent bubbles could be seen in the images and the protein also tends to gather the edge of the metal mesh. The likely reason for this distribution is thin ice in the middle of the grid. The problem of the formation of a large number of detergent bubbles could be solved by reducing the concentration of the detergent or upon applying a larger sample volume during application. To avoid uneven distribution, reducing the blotting time, changing the blot angle, using Au-C grids, use grids with smaller hole sizes were utilized. Unfortunately, time did not allow for further testing and cryo-EM measurements went ahead with the samples to hand.





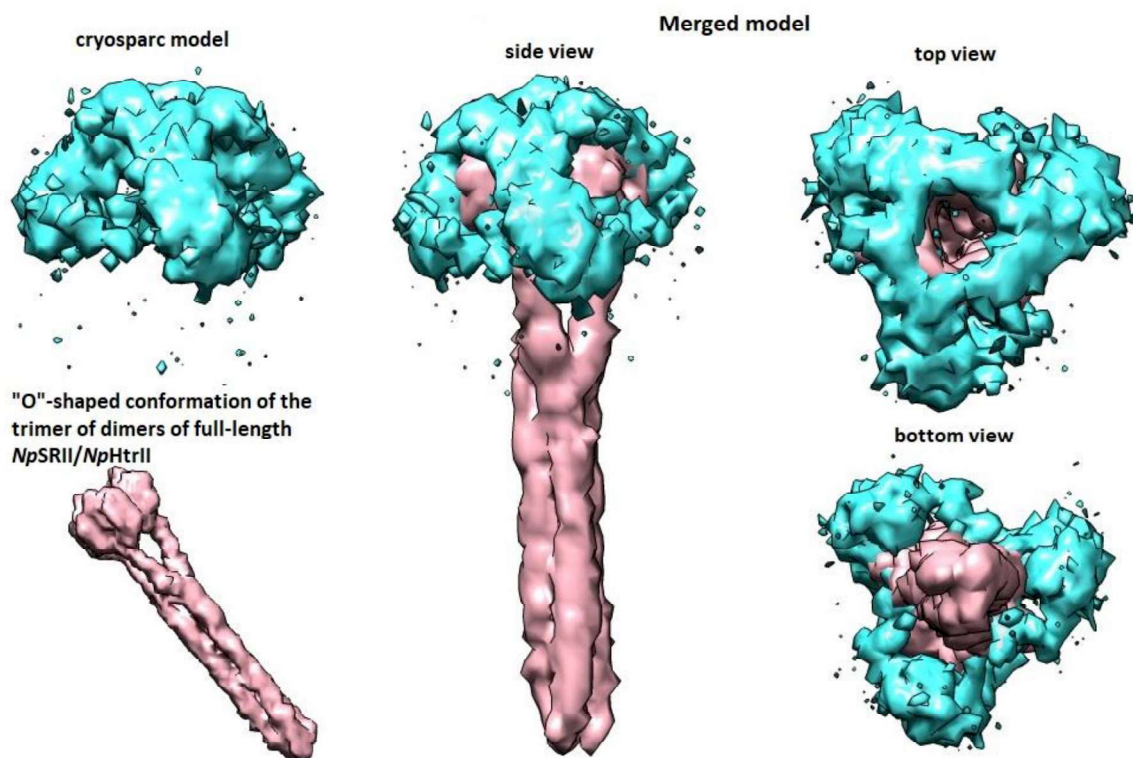
**Figure 3.1.3.1.1 Cryo-EM micrographs collected for *NpSrII/NpHtrII*.** Left: Grid 6; Right Grid 8. See Table 3.1.3.1.1 for full details of grid preparation conditions.

Cryo-EM measurements for the *NpSrII/NpHtrII* complex solubilized in DDM were carried out at Netherlands Centre for Electron Nanoscopy (NeCEN, Leiden, The Netherlands) using a Titan Krios K2 cryo-electron microscope. 2000 images were collected. The results of 2D classification was shown in Figure 3.1.3.1.2. As can be seen, the particles on the grid exhibited a marked preferred orientation and only a “top” view was visible.



**Figure 3.1.3.1.2 2D classifications obtained from the cryo-EM analysis of *NpSrII/NpHtrII* solubilized in DDM.**

Data processing was carried out using the cryoSPARC software package [138]. After 3D reconstruction an “O-shaped” model was obtained. At first, it was thought that it was a model of the transmembrane part of the complex, but after processing we understood that it is detergent around the transmembrane region (Figure 3.1.3.1.3).



**Figure 3.1.3.1.3 Cryo-EM analysis of *NpSrII/NpHtrII* solubilized in DDM.** 3D model (cyan) obtained from the cryo-EM analysis and a theoretical surface model of the *NpSrII/NpHtrII* complex (pink). Only the detergent (cyan) could be modelled using cryo-EM.

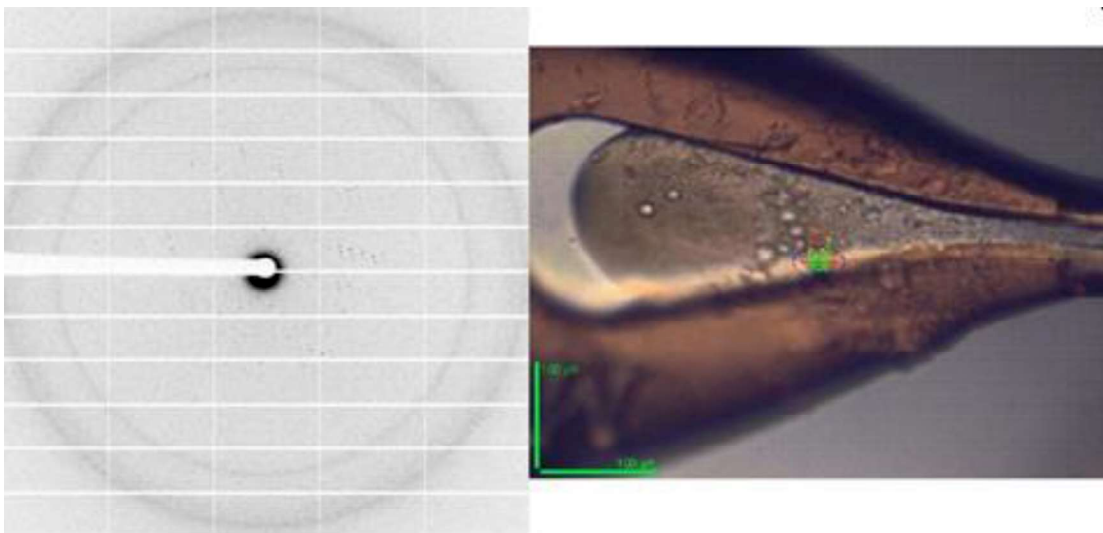
### 3.1.4 Crystallization of *NpSrII/NpHtrII* solubilized in DDM

*In meso* crystallization of *NpSrII/NpHtrII* solubilized in DDM was carried out according to standard protocols (see §3.1.1). QCP1 and QCP2 precipitation buffers (Qiagen, Germany) were used for the test crystallization screen. After 1 month, crystals of two different forms were obtained: Diamond-shaped plates (Figure 3.2.4.1, left; precipitant solution: 1.8 M – 2.4 M  $(\text{NH}_4)_2\text{SO}_4$  pH 8.5 – 9.3); Hexagonal-shaped plates (Figure 3.2.4.1, right; precipitant solution 1.8 M – 2.4 M  $(\text{NH}_4)_2\text{SO}_4$  pH 5.6 – 7.5)

X-ray diffraction analysis of both crystal forms was then carried out at beamline ID23-1 [87] of the ESRF (Table 3.1.4.1). The best results were shown in Figures 3.1.4.2 and 3.1.4.3.

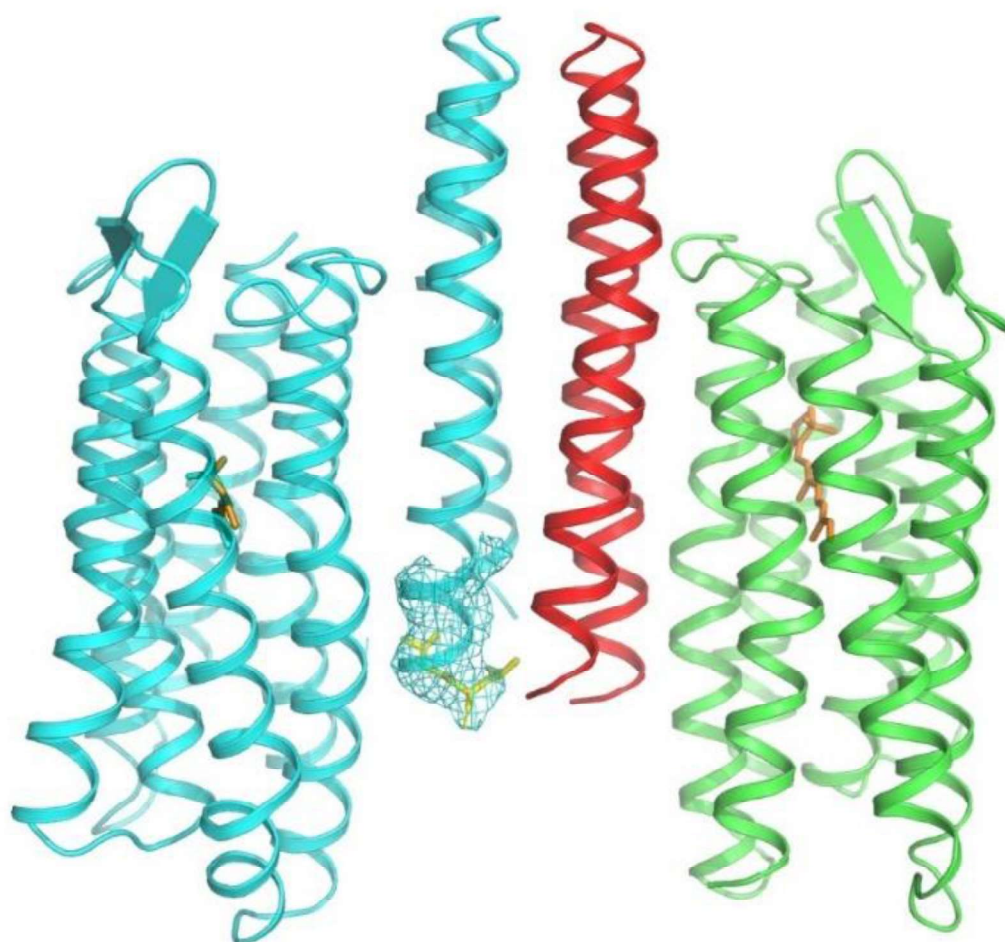


**Figure 3.1.4.1** Diamond- (left) and hexagonal- (right) shaped crystals obtained during the crystallisation trials for *NpSRII/NpHtrII* solubilized in DDM.



**Figure 3.1.4.2** The crystal structure of the diamond-shaped crystals obtained. *Top: Diffraction image and photograph of crystal mounted in a loop. Bottom: The asymmetric unit of the crystal structure of diamond-shaped crystal shown in Figure 3.1.4.1. As can be seen, these proved to be crystals of only the NpSRII sensor domain of NpSRII/NpHtrII. NpSRII molecules are shown as a green C<sub>α</sub> trace, retinal molecules in stick representation (orange).*





**Figure 3.1.4.3** The crystal structure analysis of the hexagonal plates shown in Figure 3.1.4.1. These proved to be crystals of a NpRSII/NpHtrII<sub>84</sub> dimer, which had been previously reported [24]. Part of the electron density for the truncated transducer domain is shown as blue chicken wire.

**Table 3.1.4.1** Data collection and refinement statistics for the crystal structure analysis of the Diamond- and hexagonal-shaped crystals shown in Figure 3.1.4.1. Values for the highest resolution shells are given in parentheses.

	<i>Diamonds</i>	<i>Hexagonal plates</i>
Beamline	ID 23-1, ESRF	ID 23-1, ESRF
Wavelength (Å)	0.976	0.976
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2	P6 <sub>4</sub>
Cell dimensions	$a = 61.4 \text{ \AA}$ $b = 89.1 \text{ \AA}$ $c = 151.3 \text{ \AA}$	$a = b = 66.3 \text{ \AA}$ $c = 171.25 \text{ \AA}$ $\gamma = 120^\circ$
Resolution range (Å)	48.00 – 2.81	47.74 – 3.8
R <sub>merge</sub> (%)	34.8 (141.2)	43.5 (156.1)
<I/σ(I)>	4.3 (1.13)	4.2 (1.14)
CC <sub>1/2</sub> (%)	99.7 (45.0)	99.5 (45.5)
Completeness (%)	99.21 (97.8)	99.9 (95.6)
<b>Refinement</b>		
Resolution(Å)	2.83	3.9
R <sub>work</sub> /R <sub>free</sub> (%)	23.7/25.8	26.7/28.0

### 3.1.5 Proteases activities

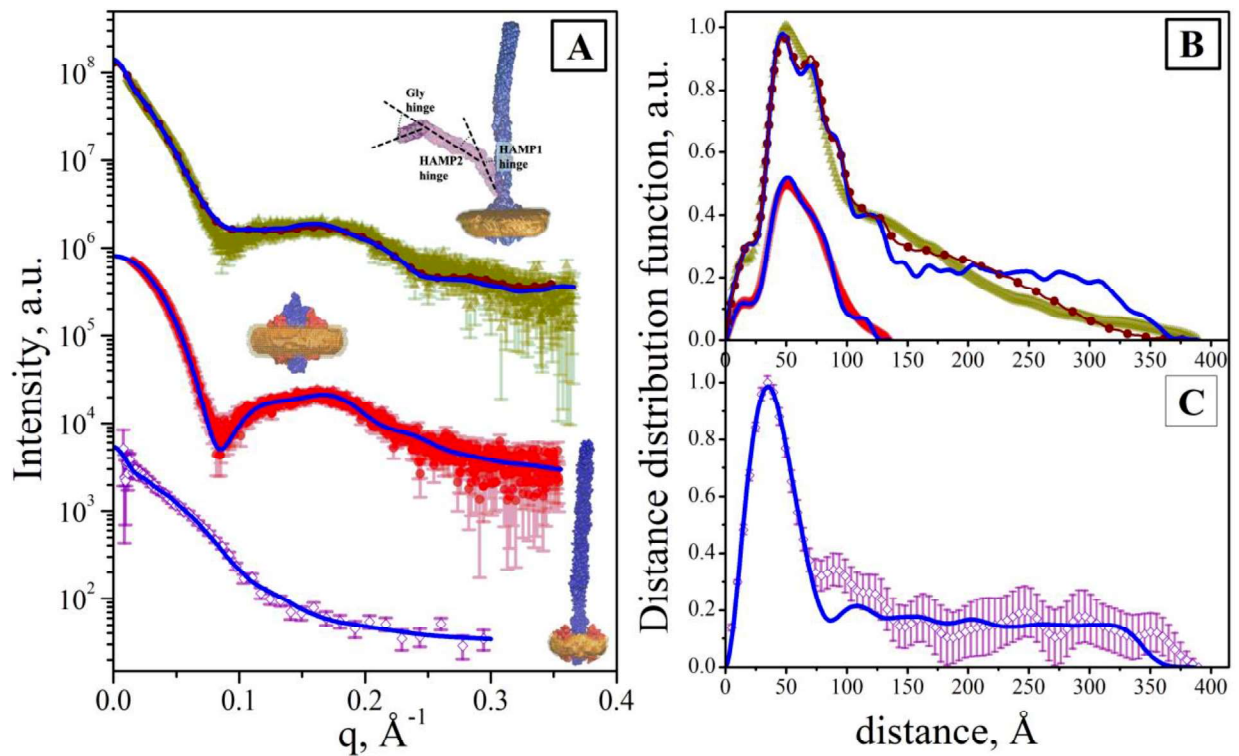
As can be seen in §3.1.4, upon crystallization of the complex of sensory rhodopsin *NpSRII* with its related transducer *NpHtrII*, two crystal forms were obtained: crystals of *NpSRII* alone (diamond-shaped crystals; Figure 3.1.4.1 and Figure 3.2.4.2) and crystals of *SRII* in complex with only the TM domain of *NpHtrII* (hexagonal plates; Figure 3.1.4.1 and Figure 3.1.4.3). This latter crystal form was obtained both during crystallization of the full-length complex (see above), and for truncated constructs containing one or two HAMP domains (data not shown here). Obviously, the transducer was cleaved. Knowledge of exactly at which site and which protease is responsible for this might help avoid proteolysis either by using inhibitors during protein preparation or by designing a construct more resistant to proteolysis.

From the crystal structure shown in Figure 3.1.4.2 it was clear that the *HtrII* proteolysis site was located after the amino acid 84. However, that more residues were not seen in the electron density may be due to disorder. Further experiments, not detailed here, suggested that the *HtrII* proteolysis site was between residues 84 and 98. This region corresponded to the putative AS1 helix of the HAMP domain of *HtrII*, thus suggesting that this region needed to be proteolysis protected if crystals of full-length *NpSRII/NpHtrII* were to be obtained.

### 3.1.6 Small-angle Scattering investigations of *NpSrII/NpHtrII*<sub>137</sub> and of full-length *NpSrII/NpHtrII*

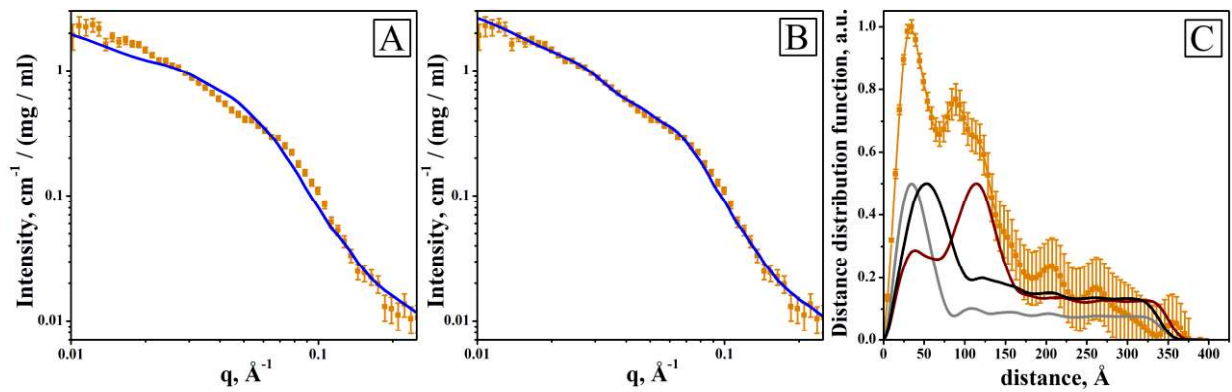
Small-angle scattering was used to investigate the oligomeric state in solution of *NpSRII/NpHtrII*<sub>137</sub> and of full-length *NpSRII/NpHtrII*. At low salt conditions SAXS scattering curves for *NpSRII/NpHtrII*<sub>137</sub>, measured at BM29 of the ESRF, were fitted well ( $\chi^2 = 1.2$ ) using a model of a dimer of the truncated complex surrounded by a detergent belt (Figure 3.6.1A, middle), strongly suggesting that, under the conditions employed, interactions between the membrane and/or HAMP1 domains of *NpHtrII*<sub>137</sub> were sufficient to induce dimerization.

Low salt condition SANS scattering curves (Figure 1.3.6.1A, bottom) for full-length *NpSRII/NpHtrII* were fitted well ( $\chi^2 = 0.7$ ) using an optimized dimer model containing a ‘straight’ adaptational domain. However, subsequent SAXS experiments, collected on ESRF BM29, allowed access to a wider range of scattering vectors compared to those accessible in the SANS experiments and suggested that this model does not fully describe the overall structure of a full-length *NpSRII/NpHtrII* dimer (Figure 1.3.6.1A, top). Here, a fit between the experimental SAXS scattering curve and that calculated from the ‘straight transducer’ model by yielded with  $\chi^2 = 5.0$ . This can be improved ( $\chi^2 = 3.3$ ) using a calculated scattering curve obtained by combining a series of models in which bends of between  $-90^\circ$  to  $90^\circ$  (with the step of  $30^\circ$ ) were induced in the HAMP1-, HAMP2- and Gly- hinges. This procedure also produced a noticeably improved fit of the distance distribution functions ( $P(r)$ ) calculated from the experimental and theoretical scattering curves, particularly in the  $160\text{--}\text{\AA}$  -  $370\text{ \AA}$  region (Figure 1.3.6.1B). This clearly suggests a dynamic structural ‘polydispersity’ for full-length *NpSRII/NpHtrII* dimers and is in line with EM studies showing that bending of Tar chemoreceptor dimers occurs at flexible HAMP and/or Gly hinges [33] (see Figure 1.2.8). While it had been suggested that such bending might play a crucial role in the formation of the hexagonal arrays of membrane receptors in complex with kinases CheA and CheW [37] it could not be ruled out that the structural ‘polydispersity’ observed for full-length *NpSRII/NpHtrII* in SAXS experiments at low salt conditions might be a result of low ionic strength in which the transducer domain is partially disordered.

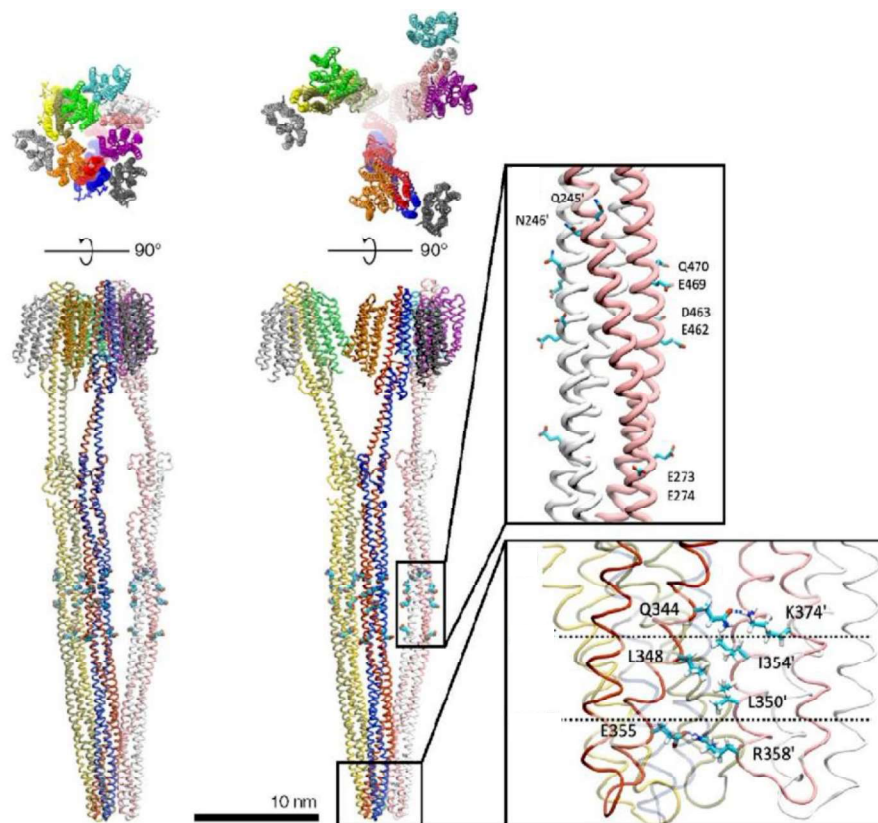


**Figure 3.1.6.1 SAS analysis of *NpSRII/NpHtrII*<sub>137</sub> and of full-length *NpSRII/NpHtrII* at low salt concentration.** *A*: bottom – experimental SANS scattering curve for *NpSRII/NpHtrII* and CRYSON fit ( $\chi^2 = 0.7$ , corresponding blue curve) based on a model (shown on the right of the curves) of a *NpSRII/NpHtrII* dimer with a detergent corona. Middle – experimental SAXS scattering curve for the *NpSRII/NpHtrII*<sub>137</sub> dimer and MEMPROT fit ( $\chi^2 = 1.2$ , corresponding blue line) based on a model (shown above the corresponding curve) of the *NpSRII/NpHtrII*<sub>137</sub> dimer with a detergent corona. Top – experimental SAXS scattering curve for full-length *NpSRII/NpHtrII* and two theoretical approximations. The first (blue solid curve) is a CRYSON3 fit ( $\chi^2 = 5.0$ ) based on “a strai”ht” model of the *NpSRII/NpHtrII* full-length dimer. The second (brown pointed curve) is a fit ( $\chi^2 = 3.3$ ) based on a combination of modified models of the *NpSRII/NpHtrII* dimer. While the difference between the two fits is not immediately apparent the improved fit of the second approximation is evident when considering of the relative residuals of the fit for the scattering vector values  $q < 0.04 \text{ \AA}^{-1}$ . In the representations of the atomic models, transducers are shown in blue or violet, the detergent belt in orange. *B*: distance distribution functions calculated from the SAXS scattering curves shown in *A* (with the same designations as in *A*). *C*: distance distribution functions calculated from small angle neutron scattering data shown in *A* (with the same designations as in *A*).

While both SAXS and SANS experiments clearly suggested that *NpSRII/NpHtrII* mainly forms dimers at low ionic strength, this was not the case at increased salt concentrations. Figure 3.1.6.2 shows the SANS scattering curve for the full length *NpSRII/NpHtrII* measured in a buffer containing 4.0M NaCl. This could not be satisfactorily matched by a theoretical curve calculated from the polydisperse dimers described above and to obtain a satisfactory fit it was necessary to calculate a theoretical scattering curve based upon a mixture *NpSRII/NpHtrII* dimers and trimers of dimers.



**Figure 3.1.6.2 SANS scattering curve for NpSRII/NpHtrII at 4.0 M NaCl and corresponding distance distribution function.** A: experimental scattering curve for NpSRII/NpHtrII at 4.0 M NaCl (orange squares) fitted with  $\chi^2 = 5.5$  to a theoretical curve (blue line) calculated for a mixture of NpSRII/NpHtrII dimers and trimers of dimers which inter-dimer contacts are induced both between the transmembrane regions of dimers and their cytoplasmic tips. B: experimental scattering curve for NpSRII/NpHtrII at 4.0 M NaCl (orange squares) fitted with  $\chi^2 = 1.3$  to a theoretical curve (blue line) calculated for a mixture of NpSRII/NpHtrII dimers a “d ”tri”od” shaped trimers of dimers. C: distance distribution functions calculated from the experimental curve shown in A and B (orange squares), and theoretical distance distribution functions of the dimers (grey line), t “e ”tri”od”-shaped trimers of dimers (brown line) and other conformation of trimer of dimers of the NpSRII/NpHtrII (black line). Distance distribution functions were normalized to obtain maximum values of 1.0 for the experimental and 0.5 for the theoretical curves.



**Figure 3.1.6.3 The two different forms of full-length NpSRII/NpHtrII trimers of dimers constructed for use in fits of experimental and theoretical SANS scattering curves.** Left: A trimer of dimers in which inter-dimer contacts are induced both between the transmembrane regions of dimers and their cytoplasmic tips. Right: A tripod-shaped trimers of dimers in which inter-dimer contacts are formed only between the tips of the cytoplasmic domains of each dimer. Figure adapted from [139].



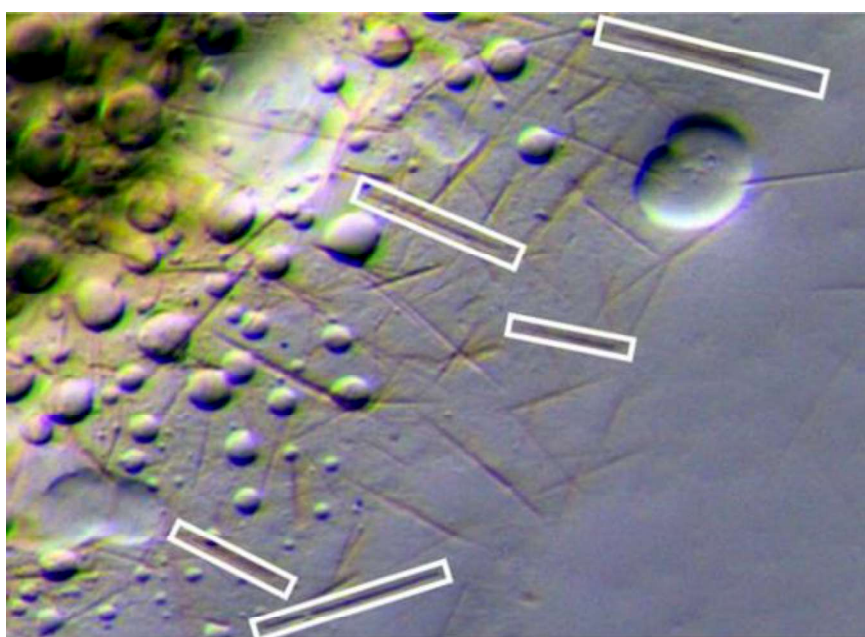
Two different forms of trimers of dimers were constructed for use in our calculations (Figure 3.1.6.3). In the first conformation (parallel), similar to a model of the trimer of dimers previously proposed, the direction of bending angles at the HAMP and Gly hinges was such that inter-dimer contacts are induced both between the transmembrane regions of dimers and their cytoplasmic tips. For the second model, the direction of bending angles at the HAMP and Gly hinges was introduced such that the transmembrane regions at the base of the trimer of dimers adopt a tripod-type disposition and inter-dimer contacts are formed only between the tips of the cytoplasmic domains of each dimer. In both cases, a detergent belt was applied as described for the SANS analysis at low salt. Fitting the experimental SANS scattering curve to a theoretical curve based on a mixture of *NpSRII/NpHtrII* dimers and tripod-like trimers of dimers resulted in  $\chi^2 = 1.3$  (Figure 3.1.6.2B). The same exercise using a theoretical curve based on a mixture of *NpSRII/NpHtrII* dimers and parallel conformation trimers of dimers resulted in  $\chi^2 = 5.5$  (Figure 3.1.6.2A). SANS experiments at NaCl concentrations 1.4 M and 2.8 M produced similar results but indicated an increasing proportion of tripod-style trimers of dimers with increasing salt concentration. Resulting weight fractions of tripod-shaped trimers of dimers are  $21 \pm 3\%$ ,  $29 \pm 3\%$  and  $36 \pm 2\%$  for 1.4 M, 2.8 M and 4.0 M, respectively. These results strongly support that the formation of *NpSRII/NpHtrII* trimers of dimers at high salt concentration is likely to occur through the contacts of the cytoplasmic tips of dimers with their transmembrane parts remaining unconnected.

## **3.2 *NpSRII/NpHtrII* reconstituted in nanodiscs**

### **3.2.1 Crystallization and X-ray diffraction analysis**

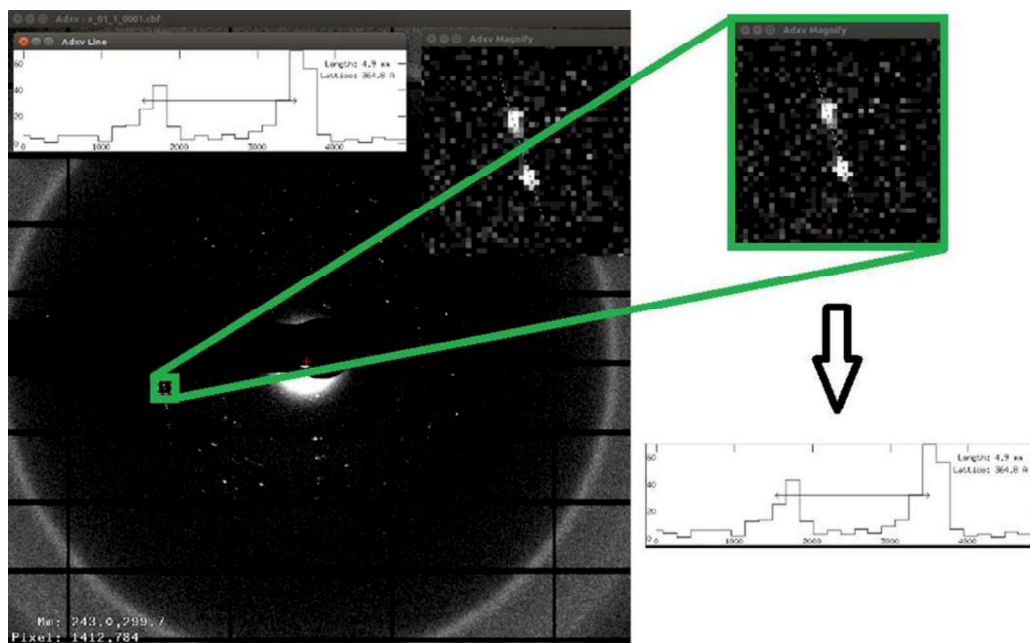
The primary goal of crystallization is to obtain high-resolution atomic structural details using X-ray crystallography. As crystallization of full-length *NpSRII/NpHtrII* solubilized in DDM was unsuccessful (see §3.1.4) and since there are precedents for the successful crystallization of membrane proteins, in particular bacteriorhodopsin bR from *H. Salinarum* [140], reconstructed in nanodiscs, it was decided to attempt the crystallization of *NpSRII/NpHtrII* complex in

MSP1E3D1/DPPC nanodiscs using the *in meso* method. Here, a protein produced using a gene not optimized for expression in *E. coli* K12 cells was used. As a result of crystallization experiments, crystals were formed in the LCP plate wells, where the QCP1 'it's precipitates were used. Crystals formed only in wells corresponding to alkaline pH values and high concentrations of ammonium phosphate (namely, in wells F6, G6, H6 (1.6M ammonium phosphate pH 7–0 - 8.2); F9, G9, H9 (2.2M ammonium phosphate pH 7–0 - 8.2); G12 (2.8M ammonium phosphate pH 7.6)). A micrograph of the crystals obtained well F6 is shown in Figure 3.2.1.1.



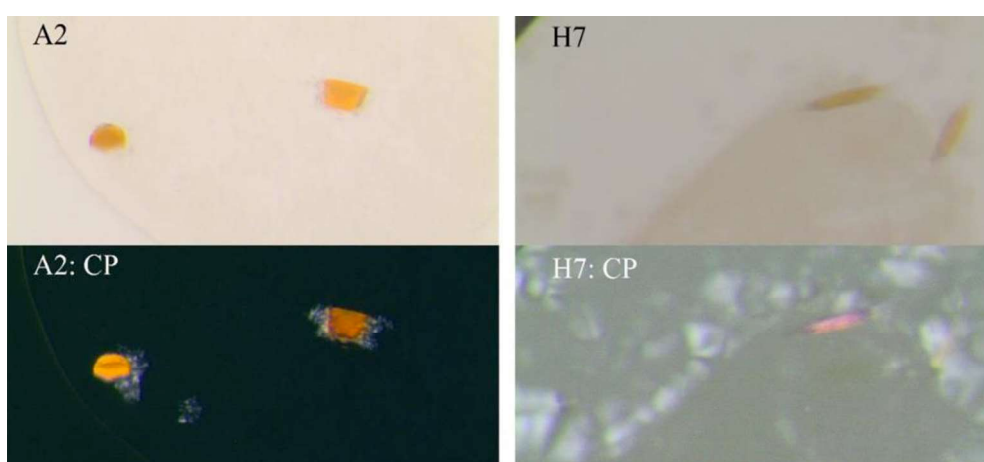
**Figure 3.2.1.1** Crystallization of *NpSR*II/*NpHtr*II complex in MSP1E3D1/DPPC nanodiscs using the *in meso* method. Micrograph of crystals in well F6 (1.6M Ammonium phosphate, pH 7.6) of the LCP plate. Several crystals are marked with rectangles.

Several long needles-shaped crystals between 5  $\mu\text{m}$  and 15  $\mu\text{m}$  thick were taken from well H6 of the LCP plate for X-ray studies. Unfortunately, the quality of the obtained diffraction patterns (Figure 3.2.1.2) did not allow the elucidation the atomic structure. However, the lattice parameters derived indicated that the crystals appeared to correspond to those of full-size *NpSR*II/*NpHtr*II complex, with a characteristic length in the range of 350-400  $\text{\AA}$  (Figure 3.2.1.2.)



**Figure 3.2.1.2** Diffraction pattern from a crystal of *NpSRII/NpHtrII* taken from crystallization well H6 (1.6M ammonium phosphate pH 8.8). On the right is an enlarged region of the image showing reflections located at a distance corresponding to a lattice parameter in direct space of  $\sim 365 \text{ \AA}$ . This suggests that the crystal corresponds to that of a full-length *NpSRII/NpHtrII* complex.

X-ray diffraction measurements were also carried out with crystals from wells A2 (0.1M Sodium acetate pH 4.6) and H7 (1.6M Ammonium phosphate pH 8.8) (Figure 3.2.1.3). Here, the lattice parameters and space group coincided with those previously obtained in the literature for the transmembrane part of the complex with a truncated transducer [21]. Since the full-size complex was used for the experiments, that these crystals contained only a truncated transducer region indicates auto-proteolysis in the sample (see §3.1.5).



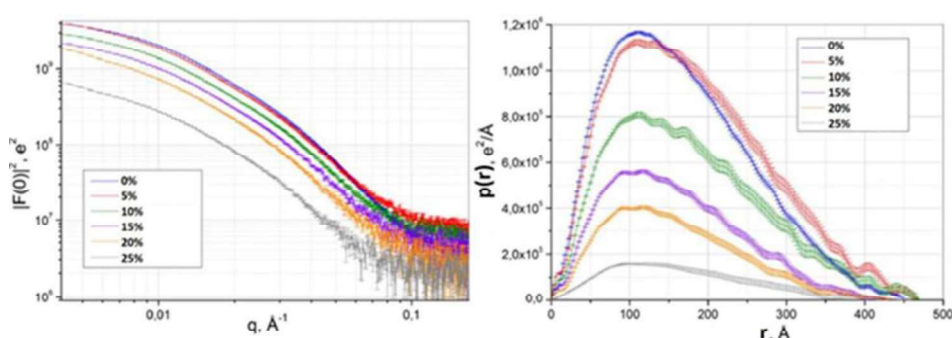
**Figure 3.2.1.3** Micrographs of crystals obtained in wells A2 and H7. The size of crystals in well A2 is  $100 \times 80 \mu\text{m}^2$ , the size of those in H7  $\sim 80 \times 15 \mu\text{m}^2$ .



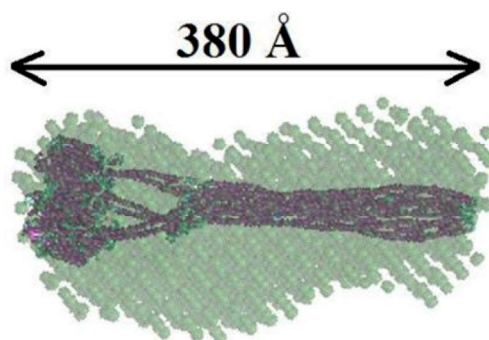
### 3.2.2 Small-angle scattering measurements from solutions of *NpSR*II/*NpHtr*II reconstructed in nanodiscs

For small-angle scattering (SAS) measurements in nanodiscs, the *NpSR*II/*NpHtr*II complex was expressed in the *E. coli* Rosetta2 (DE3) strain and solubilized in DDM. The complex was reconstructed into nanodiscs prepared from the DMPC lipid and the MSP1E3D1 membrane framework protein. The ratio *NpSR*II/*NpHtr*II: MSP1E3D1: DMPC was adjusted 3/3: 1: 120 to obtain three dimers of the complex on one nanodisk. The calculated total molecular weight of a particle of a *NpSR*II/*NpHtr*II trimer of dimers embedded in nanodiscs is 766 kDa.

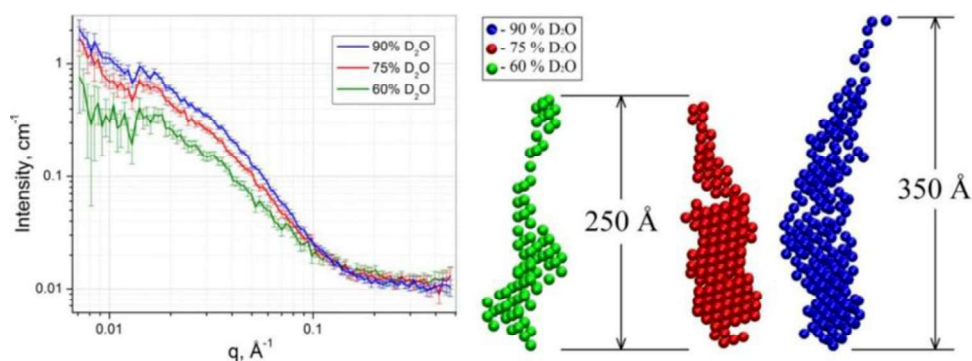
Small-angle X-ray (SAXS) (Figure 3.2.2.1 and 3.2.2.2) and neutron (SANS) scattering (Figure 3.2.2.3) measurements were carried out on the sample prepared using the protocol described above. *Ab-initio* molecular envelopes obtained from both SAXS (Figure 3.2.2.2) and SANS experiments (Figure 3.2.2.3, left panel) indicate that the *NpSR*II/*NpHtr*II complex in nanodiscs is full length with a longest dimension of between 360 Å and 380 Å as determined by SANS and SAXS experiments, respectively. However, this distance alone does not confirm the correct oligomeric state (i.e. a trimer of dimers) of the complex. For this contrast variation measurements were carried out. In SAXS experiments this was achieved by varying sucrose concentration [141], while in SANS the concentration of D<sub>2</sub>O [142] was modified.



**Figure 3.2.2.1** SAXS analysis from solutions of *NpSR*II/*NpHtr*II embedded in nanodiscs. *Left:* SAXS scattering curves normalized to the volume concentration of trimers of protein dimers, reduced to units of the square of the form factor, expressed in electronic units, obtained for *NpSR*II/*NpHtr*II in nanodiscs and at sucrose concentrations from 0% to 25% (step 5%). *Paired-distance distribution functions* are shown to the right of the scattering curves.

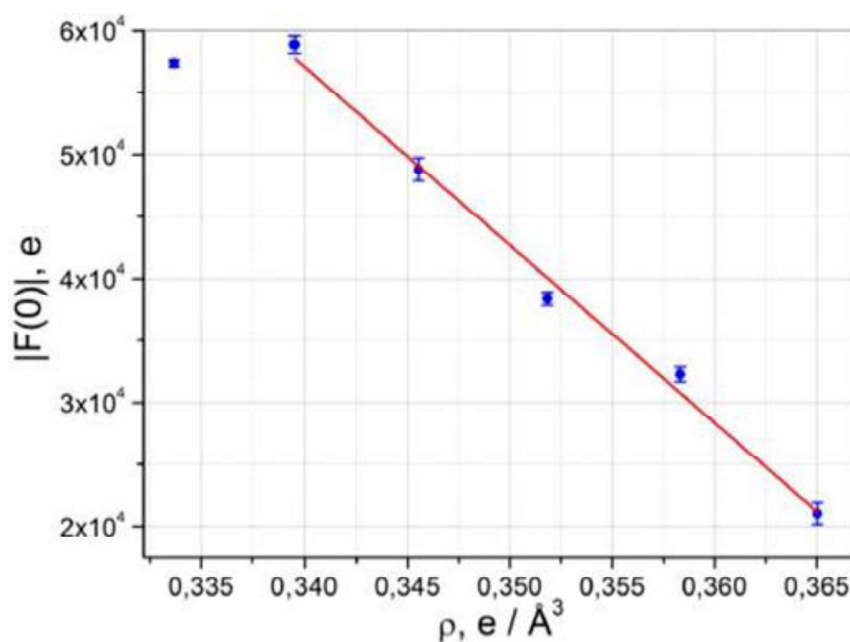


**Figure 3.2.2.2** The SAXS-obtained molecular envelope of *NpSRII/NpHtrII* embedded in nanodiscs. An image of a single-phase *ab initio* molecular envelope (green spheres) for *NpSRII/NpHtrII* structure inserted in nanodiscs obtained by processing small-angle X-ray scattering data (blue curve on the left side of Figure 3.1.2.3) using the DAMMIF program of the ATSAS software package, in comparison with atom “c ‘Y’-shaped model of a trimer of dimers of the *NpSRII/NpHtrII* complex.



**Figure 3.2.2.3** SANS scattering curves obtained for *NpSRII/NpHtrII* in nanodiscs at three different heavy water concentrations. Small-angle neutron scattering curves were obtained for *NpSRII/NpHtrII* in nanodiscs at three different heavy water concentrations: 60%, 75%, and 90%. On the right is a set of *ab initio* molecular envelopes obtained using the ATSAS software package.

The results of SAXS experiments (normalized scattering curves, distance distribution functions) in which the concentration of sucrose in the buffer was varied is shown in Figure 3.2.2.1. As can be seen from the plots of the pair-distribution function,  $D_{\max}$  varies from 380 Å (0 % sucrose) to 460 Å (25 % sucrose), consistent with the full-length protein in all cases. From the values of the intensity at the zero angle ( $|F(0)|$ ) at different buffer densities (i.e. sucrose concentrations), it is possible to recover information about the volume of particles, the number of electrons (total scattering length), and average electron density (see Figure 3.2.2.4 and Table 3.2.2.1).



**Figure 3.2.2.4 SAXS analysis, at increasing sucrose concentrations, of *NpSRII/NpHtrII* embedded in nanodiscs.** The dependence of the value of the form factor's modulus at zero scattering angle  $|F(0)|$  vs the solvent electron density (increasing sucrose concentration). Blue dots correspond to experimental data and the red line to the linear fit. It should be noted that the point corresponding to 0% sucrose, does not fit the linear relationship of the rest of the points.

**Table 3.2.2.1 The theoretical values of the volume and number of electrons for the *NpSRII/NpHtrII* complex, MSP1E3D1, DMPC, and a *NpSRII/NpHtrII* trimer of dimers reconstructed in nanodiscs.** Also included are the experimentally obtained values (the volume and number of electrons correspond to the slope and free term of the linear approximation in Figure 3.2.2.4).

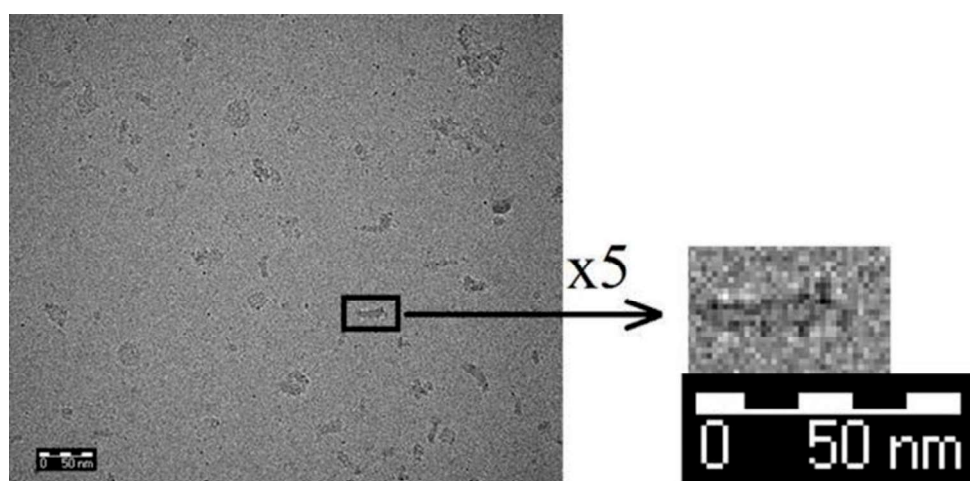
Parameter	<i>NpSRII/NpHtrII</i>	MSP	DMPC	<i>NpSRII/NpHtrII</i> in ND (trimer of dimers, theory)	<i>NpSRII/NpHtrII</i> in ND (experimentally determined)
Mass (Da)	84291	32618	678	733681	-
Volume ( $\text{\AA}^3$ )	105084	40861	1333	1032247	$(1400 \pm 100) \times 10^3$
No of electrons	45014	17426	374	394696	$(540 \pm 30) \times 10^3$
Electron Density ( $e^-/\text{\AA}^3$ )	0.4284	0.4265	0.2805	0.3824	$(0.3933 \pm 0.0018)$

As can be seen from Table 3.2.2.1, the experimentally determined partial volume and number of electrons in a particle are approximately 40% higher than the theoretical values for a *NpSRII/NpHtrII* trimer of dimers reconstituted in nanodiscs. There are two possible non-mutually exclusive explanations for this discrepancy. First, sucrose molecules can bind to a protein, changing its structural details and leading to overestimated parameters derived from SAXS data. This assumption may

be supported by the fact that in Figure 3.2.2.4, the point corresponding to the zero concentration of sucrose, does not fall on the linear relationship along with the rest of the points. Secondly, there may be partial aggregation and sample inhomogeneity. Partial aggregation may be supported because the particle shape obtained by processing the SAXS data, even for the case with zero sucrose concentration, is at least twice as thick as the atomic model (see Figure 3.2.2.2). This may be because the protocol for protein reconstitution in nanodiscs assumes a molar ratio of complex/MSP = 3/1, leading to the replacement of too many lipid molecules by the protein, leading to destabilization nanodiscs and partial aggregation of membrane proteins embedded in NDs.

### 3.2.3 Single particle cryo-electron microscopy analysis of NpSRII/NpHtrII reconstituted in nanodiscs

Cryo-EM measurements for *NpSRII/NpHtrII* reconstituted in nanodiscs were carried out in Maastricht, Germany using an ARCTICA cryo-electron microscope. A total of 1285 movies were taken, with 96 frames taken for each (16 fractions, 6 frames per fraction). The total dose used was  $78 \text{ e}^-/\text{\AA}^2$  ( $4.8 \text{ e}^-/\text{\AA}^2$  per fraction). An example of a micrograph obtained is shown in Figure 3.2.3.1. The pixel size in the image is  $1.27 \text{ \AA}$ .

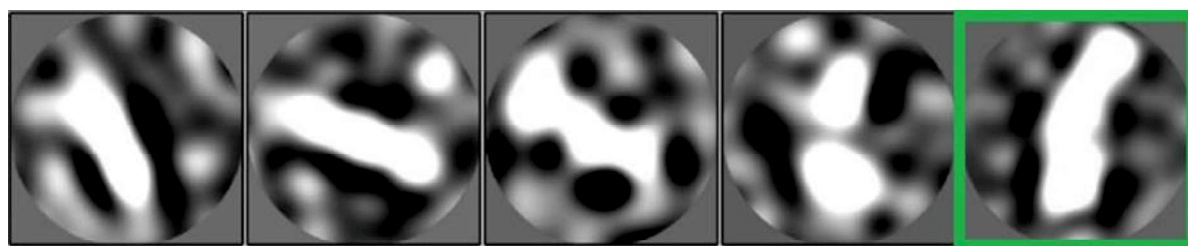


**Figure 3.2.3.1** Cryo-electron micrograph for *NpSRII/NpHtrII* reconstituted in nanodiscs. *On the right is a comparison of one of the particles with a scale of 50 nm.*

Comparison of the particle length with the scale shown in Figure 3.2.3.1 shows a length of  $\sim 300 \text{ \AA}$ , consistent with that expected for a *NpSRII/NpHtrII* dimer,

especially when considering that the particle may be oriented not exactly parallel to the plane of the grid cryo-EM grid. However, only a small fraction of the particles in the micrograph shown in Figure 3.2.3.1 can be compared to any possible projection of a *NpSRII/NpHtrII* dimer. This was true for all micrographs obtained. Moreover, there are a large number of aggregates of different sizes in the micrographs, making it difficult to automatically search for particles.

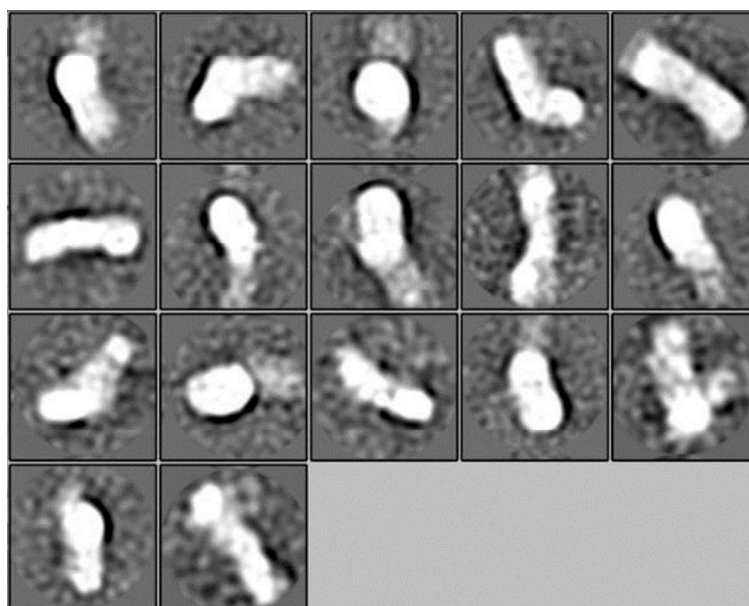
Using a manual isolation procedure, 27 particles were identified (using a particle diameter of 355.6 Å) for input into a 2D-classification procedure from which 5 classes were obtained (Figure 3.2.3.2). Data processing was carried out using a software package RELION [143]. Of course, 27 particles is too small for a fully-fledged 2D classification. Indeed, it is recommended to use a minimum of 100-150 particles in the initial stages. However, even to choose these 27 particles, it was necessary to manually view about 200 micrographs.



**Figure 3.2.3.2** Classes obtained using 2D classification procedures for particles isolated in manual processing mode. The frame size corresponds to 355.6 Å. The class that was used as a template for the automatic particle search is marked with a green border.

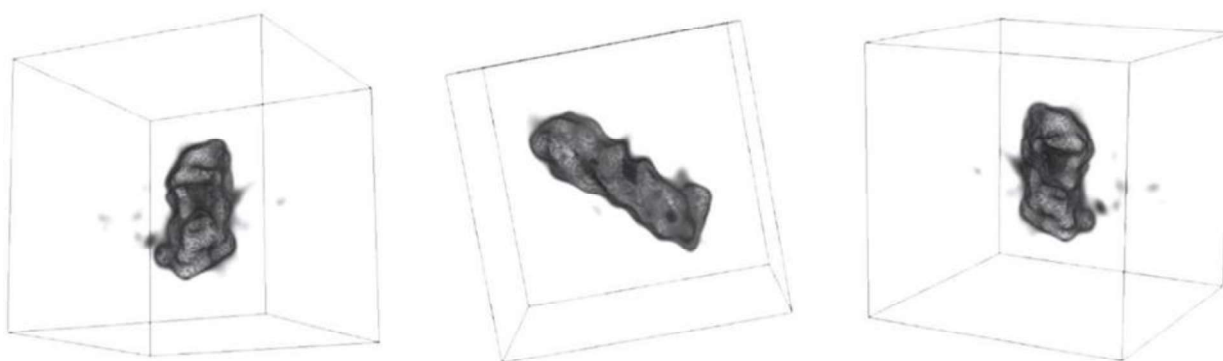
One of the 2D classes obtained (highlighted in green in Figure 3.2.3.2) was used as a template to automatically find other particles. Here, the following conditions were used: peak threshold = 1.0; minimum separation between particles = 355.6 Å; minimum signal-to-noise ratio = 1.0. As a result, 4510 particles were obtained. For these particles, a 2D classification with 100 grades was applied. After that, using the subset selection option, 17 classes were selected (see Figure 3.2.3.3). The number of particles corresponding to these 17 classes is about 1250.





**Figure 3.2.3.3** 2D class averages obtained following automatic particle picking based on the template highlighted in Figure 3.1.3.2.

Upon closer examination, it transpired that of the ~1250 particles corresponding to 17 selected classes, about 90% visually differ from these classes. Using subset sampling, only about 150 particles (which visually correspond to the 17 selected classes) were selected from these 1250 particles. These selected ~150 particles were used to generate the *de novo* 3D model shown in Figure 3.2.3.4. While this shows some features that might be expected for a *NpSR II/NpHtr II* complex reconstructed in nanodiscs there is not enough particle statistics to obtain structural information with a higher resolution.

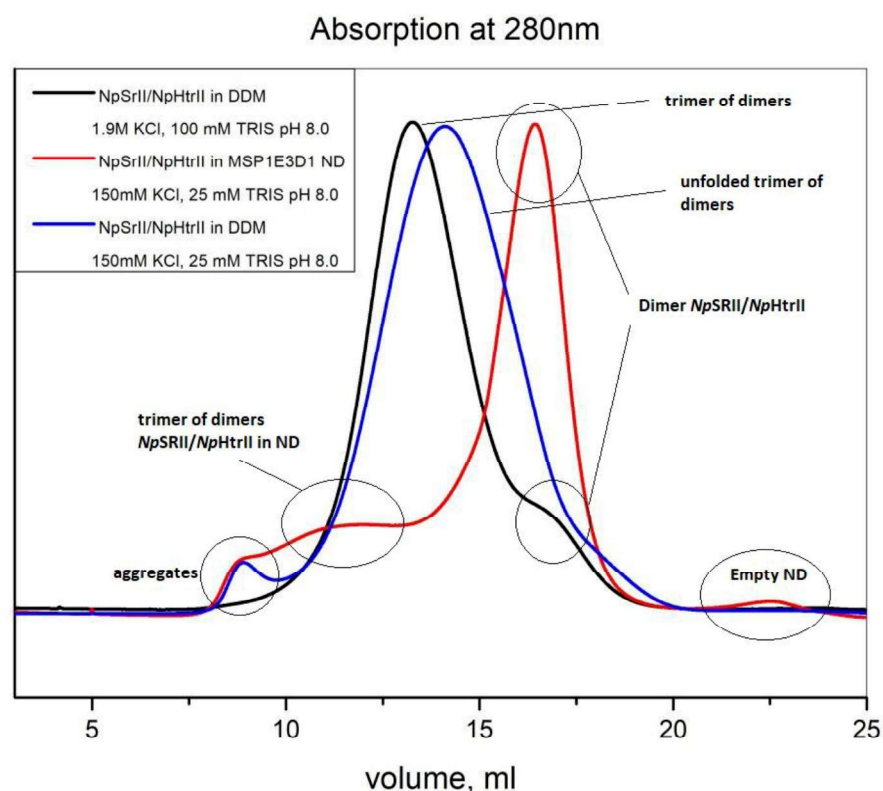


**Figure 3.2.3.4** Three images of a cryo-EM derived three-dimensional *de novo* model of the *NpSR II/NpHtr II* complex reconstructed in nanodiscs. The cube edge is 355.6 Å.

It is known that the full-length *NpSR II/NpHtr II* complex assembles into a functional state (a trimer of dimers) in the presence of a high-molar salt buffer (4 M

NaCl). The use of a buffer of such molarity in cryo-electron microscopy experiments is difficult, because there is no contrast between the protein and the solution in the micrograph; this consideration is confirmed experimentally (one grid was tested) so it was decided to try to see whether cryo-EM negative contrast methods [50, 51] might improve the contrast available in such high salt conditions.

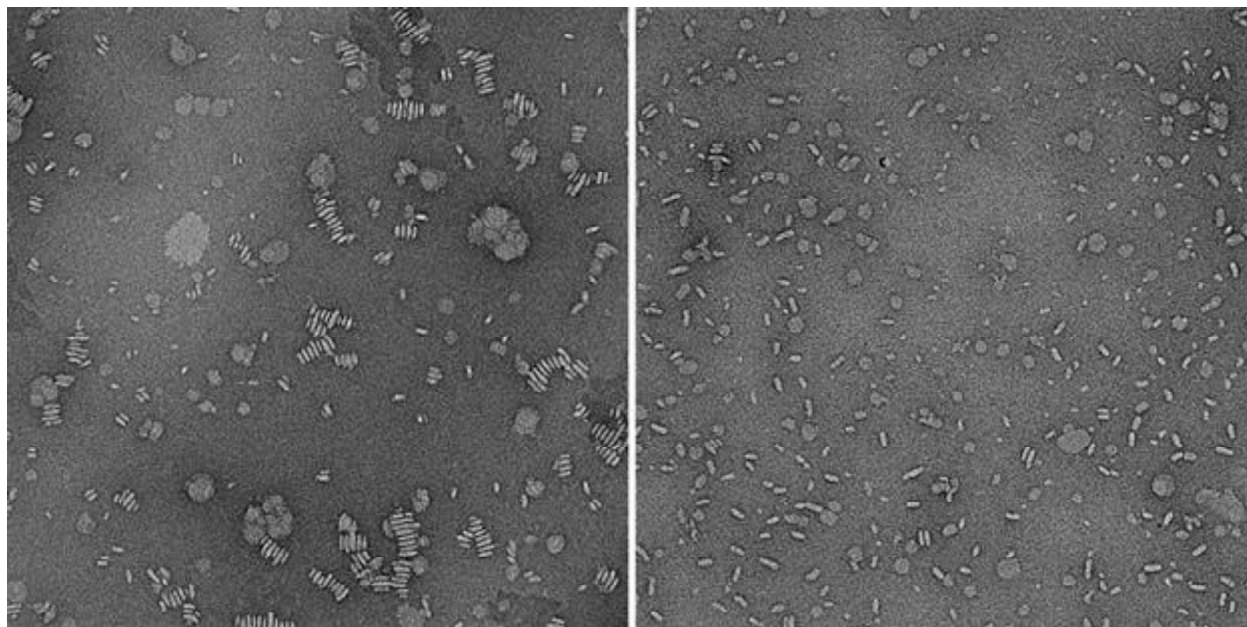
Here, the complex was embedded into nanodiscs formed with MSP1E1 lipoprotein and POPC lipid (SigmaAldrich) in a buffer with a high salt and concentration according to known protocols. Next, gel filtration purification of the protein complex embedded into the nanodiscs was carried out in a buffer comprising 25 mM Tris-HCl, 150 mM NaCl, pH 8.0 (Figure 3.2.3.5).



**Figure 3.2.3.5 Gel filtration analysis of three *NpSrII/NpHtrII* samples.** The black curve is for *NpSrII/NpHtrII* in high salt concentration solubilized in DDM, the blue curve is for *NpSrII/NpHtrII* in low salt concentration solubilized in DDM, the red curve is *NpSrII/NpHtrII* reconstructed in MSP nanodiscs at high salt concentration and equilibrated in low salt concentration.

The obtained sample was investigated by the negative contrast method at various concentrations (2 mg/ml, 1 mg/ml, 0.5 mg/ml). It was noted on the

micrographs that the sample aggregates very strongly and the nanodiscs are collected in “stacks” (Figure 3.2.3.6). The acidity of the buffer was changed to pH 7.5 and the step of concentrating the built-in complex was excluded from the protocol. Due to this, it was possible to significantly improve the homogeneity of the sample but it remained impossible to take this analysis further.



**Figure 3.2.3.6** Micrographs obtained using the negative contrast method for the full-size *NpRSII/NpHtrII* protein complex embedded in MSP1E1 lipid nanodiscs. *Left: An image captured in a standard buffer. Right: image captured in the optimized buffer. Clearly, the assembly of nanodiscs into “stacks” is observed in a standard buffer.*

### 3.3 Membrane Proteins Interactions Probed by Noble Gases

The initial aim of this thesis work was to elucidate, at high resolution, the three-dimensional structure of the full-length *NpRSII/NpHtrII* photo axis sensor in its physiologically relevant state. Unfortunately, and as detailed above, X-ray crystallography and cryo-EM experiments in which the sensor was either solubilized in detergent or reconstituted in nanodiscs were unsuccessful in achieving this aim. Moreover, while SAS experiments did allow the construction of a model of the *NpRSII/NpHtrII* trimer of dimers, SAXS and SANS are low-resolution techniques and the resulting model relies heavily on existing high-resolution structural information gathered for various fragments of *NpRSII/NpHtrII* in previous work. As



it was clear that optimizing sample preparation to allow successful high-resolution studies of full-length *NpRSII/NpHtrII* trimers or dimers would require a much longer time-frame than available during this thesis, it was therefore decided to shift the focus of this work towards an investigation of noble gas-membrane protein interactions, with the idea that noble gas atoms could serve as models for interactions of membrane proteins with small molecules which can diffuse through the lipid bilayer environment of membrane proteins. Such interactions are known to modify the functionality of membrane proteins [144].

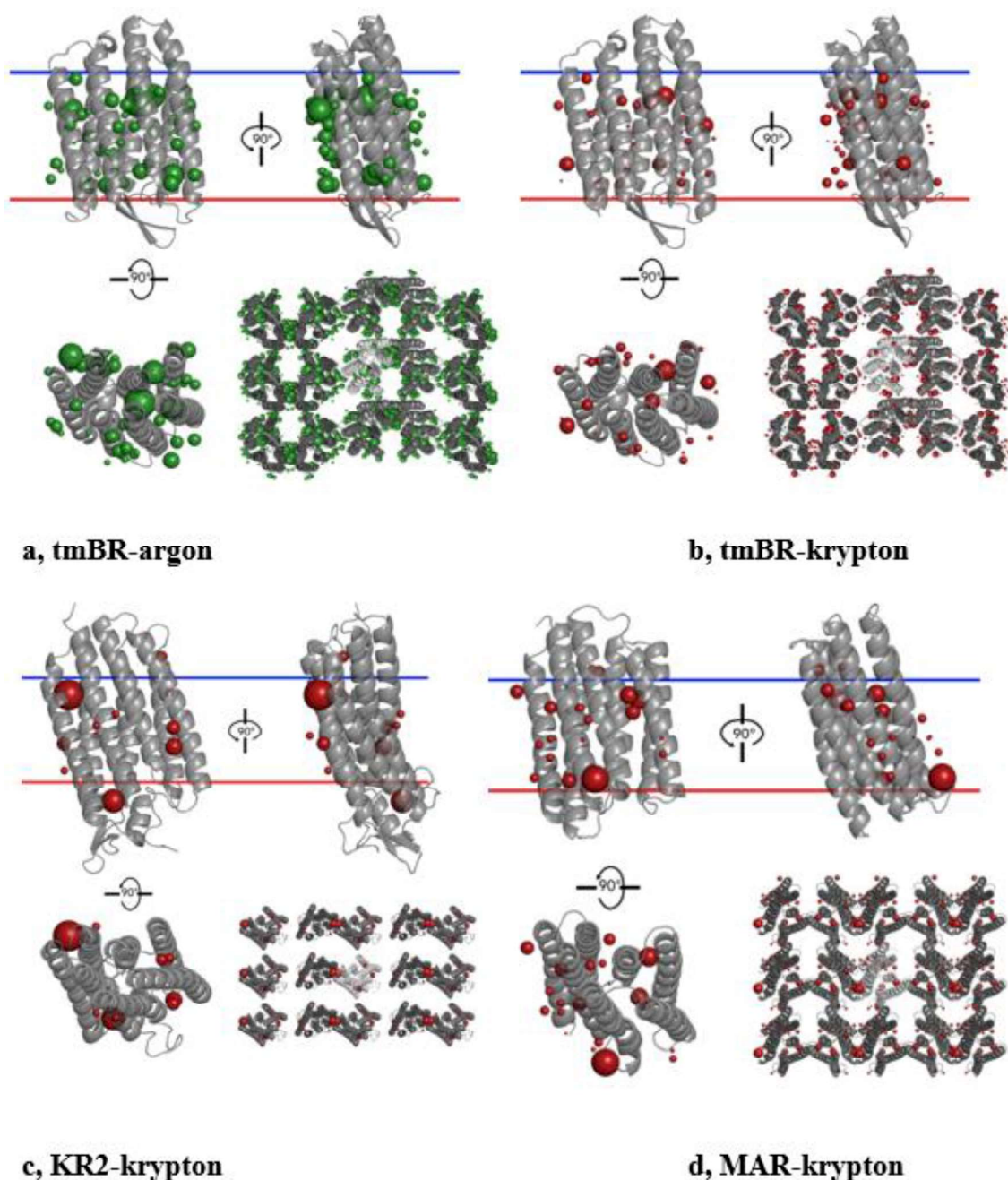
With this in mind, crystals of three membrane protein systems were prepared: a triple mutant of bacteriorhodopsin (tmBR) [124], the proton pumping rhodopsin MAR [114] and the monomeric form of the sodium pump KR2 [123]. The crystals obtained were derivatized with krypton under pressure of  $130\pm 5$  bar using the ESRF high pressure facility [82] for macromolecular crystallography (MX). Crystals of tmBR were also derivatized with argon under  $2000\pm 10$  bar. For each target, multiple data sets, aiming either at maximizing anomalous signal or data set resolution, were collected (Table 3.3.1). Native data sets were also collected from crystals of the same batch for tmBR and MAR (Table 3.3.1). Kr/Ar atoms (Figure 3.3.1) were located using anomalous difference maps. Datasets of highest resolution were chosen for crystal structure refinement (Table 3.3.2).

**Table 3.3.1 Data collection statistics for noble-gas derivatized membrane protein crystals. Highest resolution shells are shown in parentheses.**

Protein acronym	KR2	MAR			tmBR				
	krypton	krypton		native (2 datasets merged)	krypton		argon		native
Dataset No	1	merged	5		3	merged	anomalous	high- resolution	
Wavelength, Å	0.864	-	0.861		0.863	-	1.850	0.972	
Space group	I222	P2 <sub>1</sub>			C222				
a, Å	40.5	50.9	50.9	50.3	115.3	115.4	115.5	115.7	115.9
b, Å	82.2	40.31	40.38	40.0	120.0	119.8	119.2	119.2	119.4
c, Å	233.4	60.81	60.62	60.2	36.5	36.4	36.3	36.4	36.2
β, °	-	101.30	101.36	102.08	-				
Resolution, Å	100-2.6 (2.8-2.6)	100-2.25 (2.35-2.25)	100-2.25 (2.35-2.25)	100-1.85 (1.90-1.85)	100-2.0 (2.1-2.0)	100-2.0 (2.1-2.0)	100-2.2 (2.3-2.2)	100-1.65 (1.70-1.65)	100-1.70 (1.74-1.70)
Multiplicity	6.9 (7.0)	9.3 (5.1)	3.5 (3.5)	3.2 (1.9)	6.7 (6.7)	15.2 (10.2)	6.3 (5.3)	3.1 (3.1)	3.5 (3.5)
Completeness (%)	99.9 (100.0)	99.5 (96.8)	99.4 (97.9)	98.7 (95.6)	99.7 (99.8)	100 (100)	99.2 (96.7)	99.0 (99.4)	99.4 (99.6)
R <sub>meas</sub> (%)	19.2 (211.7)	21.7 (139.2)	14.0 (129.9)	11.7 (101.3)	14.9 (186.0)	21.2 (185.0)	3.3 (48.1)	5.0 (176.9)	8.8 (95.7)
<I/σ(I)>	9.22 (1.07)	9.42 (1.19)	7.74 (1.14)	8.37 (1.13)	9.09 (1.49)	11.09 (1.63)	37.72 (4.52)	16.29 (0.92)	10.52 (1.33)
CC <sub>1/2</sub> (%)	99.7 (40.0)	99.6 (45.0)	99.6 (49.1)	99.6 (45.5)	99.8 (51.7)	99.8 (62.8)	100 (95.9)	99.9 (34.0)	99.8 (59.1)
XDS SigAno (100Å- 10Å shell)	1.51	1.45	1.35	-	2.21	2.88	4.45	1.18	-
CC <sub>ano</sub> , %	61	62	54	-	83	85	93	45	-

**Table 3.3.2 Refinement statistics for noble-gas derivatized membrane protein crystal structures described in this thesis.**

<b>Protein acronym</b>	<b>KR2</b>	<b>MAR</b>		<b>tmBR</b>		
<i>Derivative</i>	Kr	Kr	Native	Kr	Ar	Native
<i>Resolution limit, Å</i>	2.6	2.25	1.85	2.0	1.65	1.70
<i>R<sub>work</sub>/R<sub>free</sub> (%)</i>	21.6/26.9	17.8/24.5	16.6/21.2	17.0/23.2	16.3/20.5	15.4/19.5
<i>Number in the ASU:</i>						
<i>Protein residues</i>	272	217	217	225	228	225
<i>Atoms</i>	2444	1983	2024	2095	2240	2131
<i>Water molecules</i>	81	99	137	106	119	110
<i>Atoms of lipids</i>	194	162	177	206	219	243
<i>Other ions</i>	Na <sup>+</sup>	-	-	SO <sub>4</sub> <sup>-</sup>	2 SO <sub>4</sub> <sup>-</sup>	2 SO <sub>4</sub> <sup>-</sup>
<i>Noble gas atoms:</i>	11	19	-	35	47	-
<i>Internal</i>	0	6	-	2	5	-
<i>Surface</i>	7	8	-	11	15	-
<i>In crystal contacts</i>	4	5	-	22	27	-
<i>R.m.s.d. with native C<sub>α</sub> positions, Å</i>	0.31	0.26	-	0.11	0.11	-
<i>Average B-factors, Å<sup>2</sup></i>						
<i>Overall</i>	46.8	34.6	34.3	29.4	27.5	22.5
<i>Protein atoms</i>	51.8	35.3	33.6	28.1	25.5	19.2
<i>Water molecules</i>	61.5	51.5	54.6	52.2	44.4	38.8
<i>Atoms of lipids</i>	51.2	34.6	34.8	31.8	30.2	22.5
<i>Noble gas atoms (internal)</i>	-	32.9	-	23.6	22.3	-
<i>Noble gas atoms (surface)</i>	48.5	33.8	-	25.0	21.7	-
<i>Noble gas atoms (crystal contacts)</i>	44.4	40.5	-	28.4	22.6	-
<i>Average occupancy</i>						
<i>Lipid atoms</i>	0.59	0.46	0.39	0.45	0.40	0.35
<i>Noble gas atoms (internal)</i>	-	0.30	-	0.25	0.50	-
<i>Noble gas atoms (surface)</i>	0.32	0.22	-	0.22	0.29	-
<i>Noble gas atoms (crystal contacts)</i>	0.38	0.26	-	0.16	0.30	-
<i>Ramachandran plot, %</i>						
<i>Preferred</i>	96.7	99.1	99.5	99.1	98.5	99.1
<i>Allowed</i>	3.3	0.9	0.5	0.9	1.5	0.9
<i>Outliers</i>	0	0	0	0	0	0

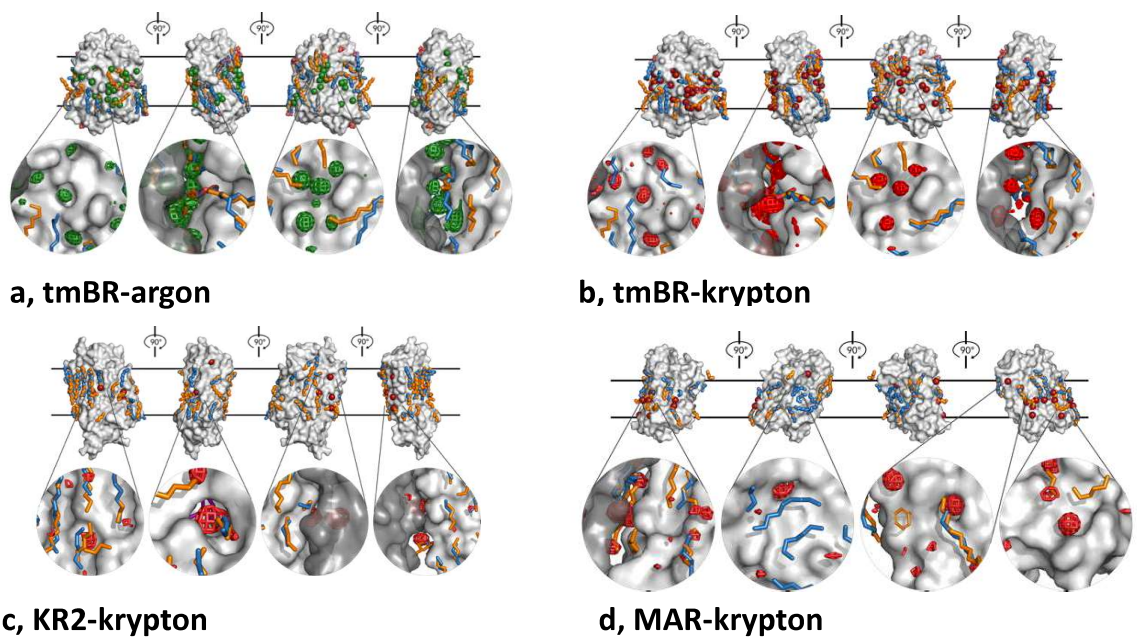


**Figure 3.3.1 Overall cartoon views of the structures of derivatised proteins: tmBR-argon (a), tmBR-krypton (b), KR2-krypton (c), MAR-krypton (d).** *The represented views are side views in membrane (top left and top right figures), top view perpendicular to membrane (bottom left figure), and crystal packing in XY-plane (bottom right figure). Noble gas atoms are shown as red (krypton) or green (argon) spheres. Sphere size is proportional to atom occupancy.*

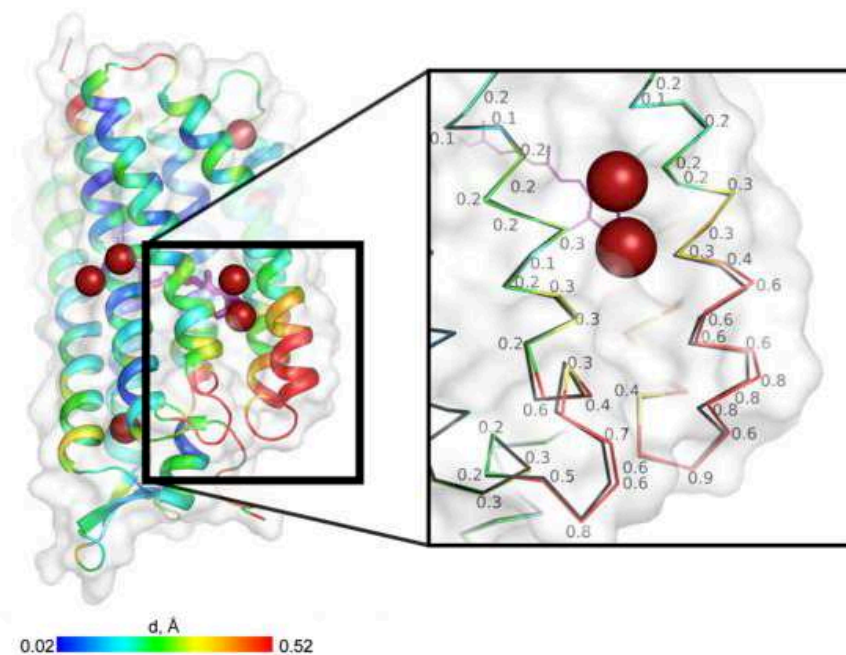
In contrast to previously reported studies [145], most of the noble gas atoms were observed on the outer hydrophobic surface (buried in the membrane) of the MPs and only a small portion of the atoms bound in internal cavities. An explicit correlation is observed between the number of found positions and the resolution of the diffraction data set.

In total, 47, 35, 11 and 19 noble gas atom positions were identified – many with partial occupancy - in the derivatized structures of tmBR-argon, tmBR-krypton, KR2 and MAR, respectively. These fall into two main groups (Table 3.3.2): atoms in internal cavities, defined as being confined inside the volume of the 7-helix bundles characteristic of all three membrane proteins (*1<sup>st</sup> group*); atoms on the hydrophobic surface the proteins. This latter category can be sub-divided into two further groups: those exposed to a bulk lipid bilayer region (*2<sup>nd</sup> group*); those involved in crystal contacts – in the plane of the membrane - between MP molecules (*3<sup>rd</sup> group*). The majority of binding sites belong to the 3<sup>rd</sup> group. However, the average occupancy of atoms in crystal contacts are not higher than of those on the surface (*2<sup>nd</sup> group*) (Table 3.3.2); internal sites (*1<sup>st</sup> group*) have higher occupancy on average.

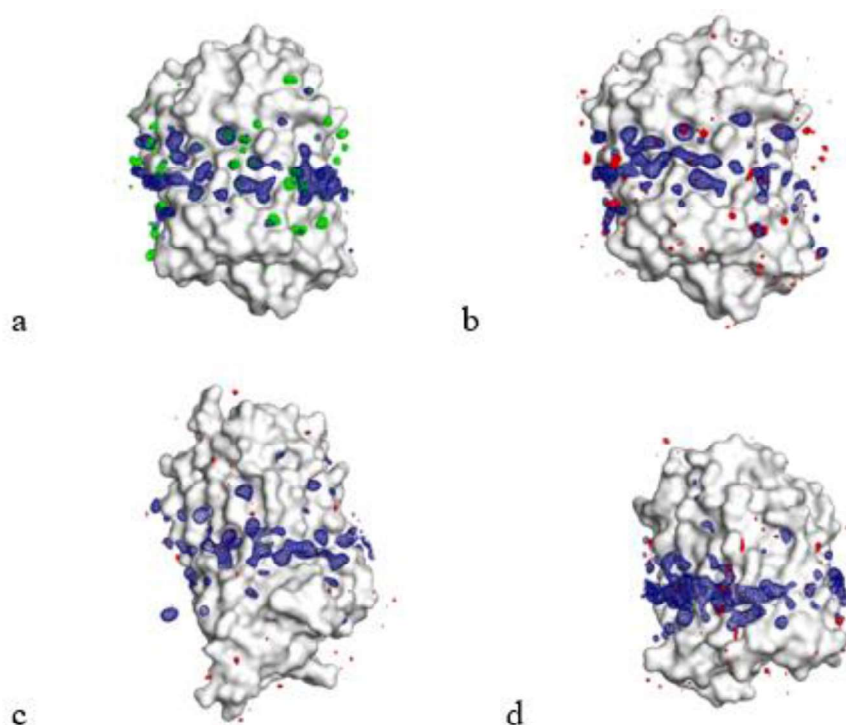
Most of the noble gas atom positions seen the bulk lipid bilayer region were close to the positions of lipid acyl chain fragments observed in the comparable native crystal structures (Figure 3.3.2). At many such sites (11, 13, 8 and 6, for tmBR-argon, tmBR-krypton, MAR-krypton and KR2-krypton, respectively), noble gas atoms replaced, at least partially, lipid atoms, with an altered conformation of lipid chains sometimes being clearly visible (Figure 3.3.2a). However, in most of the cases any alternate conformations were not clearly discernible. The other visible lipid fragments were found at the same positions they occupy in the native structures. Overall, the derivatized structures do not differ significantly from their native counterparts with  $C_{\alpha}$  r.m.s.d. values not exceeding 0.3 Å (Table 3.3.2). However, the case of KR2 is particular because the backbone of helix E is displaced about 1 Å near binding of a krypton atom in the pocket proximal to  $\beta$ -ionone ring of the retinal (Figure 3.3.3). It should also be noted that KR2 case is unique due to the absence of noble gas atoms in internal cavities. There is a good consistency between Ar and Kr binding positions in tmBR structures: of 35 determined Kr positions, only two with low occupancy (0.1) do not match any Ar sites identified.



**Figure 3.3.2** Illustrations of how noble gas atoms are arranged on the surfaces of the derivatized structures (tmBR (a, Ar; b, Kr), KR2- $\text{C}$ (c), MAR-Kr (d)). Each panel shows four side views in the membrane that differ by  $90^\circ$  rotation. The membrane region is bordered with two black parallel lines. Kr atoms are shown as red spheres, Ar as green spheres. Lipid fragments present in the derivatized structures are shown in sky-blue; lipid fragments from native structures are shown in orange for comparison. At the bottom of each panel four representative noble gas binding sites in each view are enlarged and the anomalous density maps from either krypton (red chicken wire) or argon (green chicken wire) are drawn at the level of  $3.0 \times r.m.s.$  Symmetry-related molecules are shown in transparent dark to better illustrate the contacts.



**Figure 3.3.3** Cartoon representation of the crystal structure of KR2 derivatised with krypton. *Cartoon colour illustrates  $C\alpha$ -position discrepancy (in  $\text{\AA}$ ) between the derivatised and the native structure (with scale in the bottom). In the zoomed window those discrepancies are shown for each atom. Krypton atoms are shown as red spheres.*

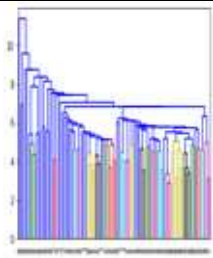
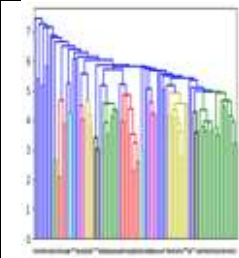
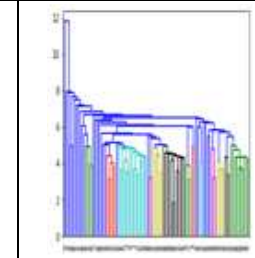
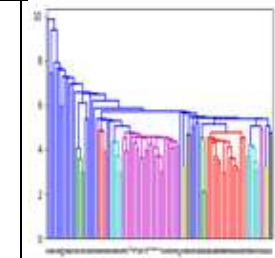


**Figure 3.3.1** Arrangement of peaks of the MD maps (sampled at  $3.5 \times r.m.s.$  level, navy blue chicken wire) on the surface of the proteins (a, tmBR-argon, b, tmBR-krypton, c, KR2-krypton, d, MAR-krypton). *Anomalous maps are shown (sampled at  $3.0 \times r.m.s.$  level) in red (krypton) and green (argon).*



We also analyzed the probability density maps obtained by molecular dynamics (MD) simulations (MD maps, see Materials and Methods § 2.5.6) of Kr and Ar interacting with the three proteins. Peaks in the MD maps (indicating increased presence of Ar/Kr, see Methods) massively populated the hydrophobic surface of the proteins (Figure 3.3.1). These peaks were systematically clustered (Figure 3.3.1).

**Table 3.3.3 Statistics of the MD experiment and comparison with noble gas atom positions observed in the MX models described above.**

	<b>KR2 dimer (Kr)</b>	<b>MAR single molecule (Kr)</b>	<b>tmBR single molecule (Kr)</b>	<b>tmBR single molecule (Ar)</b>
<i>Noble gas atoms (MX)</i>	11	19	35	47
<i>Matches</i>	11	16	30	38
<i>“Surface” atom matches</i>	7/7	8/8	11/11	15/15
<i>“Crystal” atom matches</i>	4/4	5/5	18/22	23/27
<i>“Internal” atom matches</i>	-	3/6	1/2	0/5
<i>Number of MD density peaks</i>	98	96	96	102
<i>Number of MD density peak clusters</i>	59	37	39	38
<i>Clusters not present in the MX model</i>	53	31	26	23
<i>Dendrogram for density peak clusters</i>				

The MD simulations were carried out using monomers of tmBR and MAR, and on a dimer of KR2. As can be seen (Table 3.3.3, Figure 3.3.4), there is a good correspondence between positions of noble gas atoms as determined by MX and in MD simulations. It is not surprising however, that internal noble gas positions and the positions in crystal contacts were not entirely reproduced by the MD (Table 3.3.3) due to model differences (absence of crystal contacts and MD experiment time). On the other hand, the more crucial noble gas positions “on the surface”

coincided well with the MD map peaks (Table 3.3.3). In total, the MD experiment revealed a greater number of noble gas positions than in the crystallographic models (Table 3.3.3).

### **3.3.1 Effect on MPs caused by noble gas binding**

In the experiments described above, it is clearly shown that noble gas atoms readily interact with the outer hydrophobic surface of MPs by non-specific binding and that, in many cases, this binding competes with that of lipids.

A closer look at the underlying nature of the interactions helps in understanding the above observations. Noble gas atoms interact with the hydrophobic surface of membrane protein in a similar fashion as do lipids (i.e. via weak Van der Waals forces). Most of binding sites usually are of hydrophobic nature (apolar binding pockets) exploiting London (induced dipole-induced dipole) interactions. In terms of binding energy, this largely depends on the binding site itself but will typically be in the region of a few kcal/mol (~several  $kT$  at 300K). Energy dissipation that significant upon binding can trigger substantial changes in MP stability and structure.

Matthews and co-workers [146] demonstrated that polarisability is the main factor for a noble gas atom to bind into a protein cavity. For this reason, Xe has the highest binding affinity (which is also in line with anaesthetic potencies [147]) among the non-radioactive noble gases. Using noble gas atmosphere of higher pressure can increase binding energy by the law of mass and therefore increases the occupancies of noble gas atoms at their binding sites [147] making them visible in crystallographic experiments. Another advantage of the method, as has been discussed earlier, is in flash cooling of the pressurized sample that slows down dissolution of noble gas atoms from their binding sites in a crystal.

Nevertheless, it is questionable how these energetic effects could occur within a MP. Binding to an internal cavity could evidently affect protein function and influence the dynamics. However, in the case of KR2 the absence of noble gas atoms in the central cavities indicates that a potential modulation, if it occurs, is certainly

of allosteric nature. It is evident that krypton binding has a significant impact on the structure of KR2, considering the 1 Å-shift of the end of  $\alpha$ -helix E (Figure 3.3.3). Further studies should reveal the importance of this and the functional outcome. We suggest that this outcome should be significant, similarly to another work on wild-type bacteriorhodopsin (wtBR) [145]. Here, Hayakawa and co-workers demonstrated that a xenon atom binds to the pre-existing hydrophobic cavity hidden between the C and D helices (located at the same depth in the membrane as Asp96, a key residue of proton uptake pathway). In the non-photolysed state of wtBR, Xe binding did not cause a large deformation of the cavity. However, the photocycle of purple membranes was remarkably modified in the presence of xenon: the decay of the M state was significantly accelerated while the decay of the equilibrium states N and O was slowed down.

### **3.3.2 MD experiments strongly supports and extends the results obtained using X-ray crystallography**

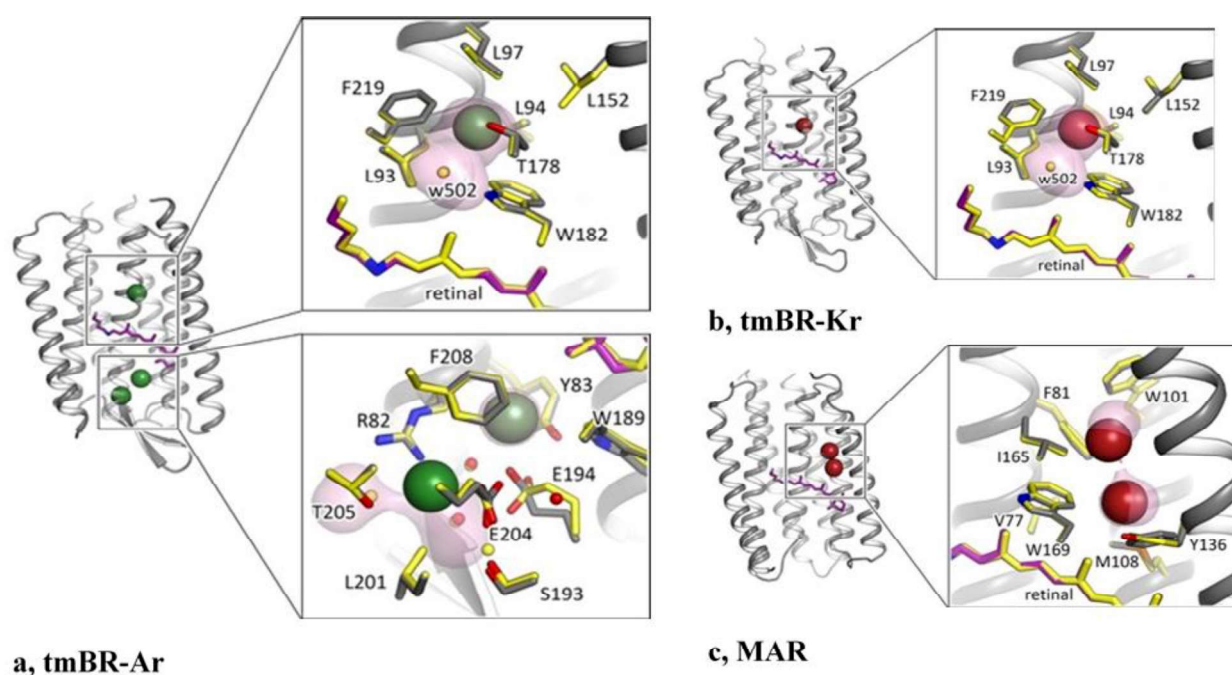
The MD simulations described above were performed partly to validate whether such an *in silico* method can generate a reliable information about binding positions of noble gas atoms at the surface of MPs. X-ray crystallography, despite being a real experimental approach has limitations such as protein crystal contacts which block access to a part of MP surface. That is why those noble gas positions in crystal contacts (3<sup>rd</sup> group) were poorly matched in the MD experiment, where a single molecule of MP has been studied. This is preferable because the part of MP surface buried between crystal contacts may be important in terms of noble gas binding. Although we still observe a partial match in the 3<sup>rd</sup> group between crystallographic model and MD simulation results which might indicate that some sites between crystal contacts are still accessible.

Strikingly, all the positions belonging to the 2<sup>nd</sup> group (“surface” positions) were entirely represented in the MD maps and additional binding sites not present in the crystallographic models were observed in the MD simulations. Given the full match in the 2<sup>nd</sup> group these predicted positions have high chances to be real and

should be taken into account when discussing noble gas interactions with MPs. That not all of the positions observed in the MD simulations were observed in the X-ray models could perhaps be explained by relatively low resolution of some X-ray structures and low occupancies of noble gas atoms observed in these. This hypothesis is supported by the case of KR2 where at the resolution of 2.5 Å lesser number of the noble gas atoms were identified despite larger hydrophobic surface. Overall, the MD method has been shown to provide a reliable result and could be used to predict noble gas positions on MP surface.

### 3.3.3 Internal binding sites and their relevance to rhodopsin function

The noble gas derivatization experiments with tmBR helped to visualize small hydrophobic cavities often found in the cytoplasmic inner parts of rhodopsins. The role of such cavities, however, remains poorly understood. One possible hypothesis is that these cavities can accommodate diffuse water molecules, which are relocated during photocycles and thus may play the key role in the function of rhodopsins.



**Figure 3.3.3.1** Schematic illustration of the positions of the internal noble gas atoms in the structures of tmBR (a, Ar; b, Kr) and MAR (c, Kr) and their environments. The protein structures are shown in grey cartoon representation, retinal molecules are colored in purple, krypton atoms in red and argon atoms in green. Panels on the right of the cartoon models illustrate noble gas atom environments and compare side chain arrangements with the native crystal structures in those regions (shown in yellow).

As a demonstration of this concept, in the crystal structure of the early L intermediate state of bR a continuous chain of five water molecules was predicted to be formed between the retinal Schiff base and the proton donor, D96. However, in the ground state crystal structure only two water molecules were identified in the cytoplasmic inner part of the protein [113]. In the tmBR-argon and tmBR-krypton crystal structures determined here, a small cavity near T178 residue and the water molecule HOH502 is occupied either by an argon atom or a krypton atom, respectively (Figure 3.3.3.1). Under physiological conditions, and taking into consideration the distances to nearby residues and the cavity volume, this cleft is suitable for accommodation of an additional water molecule. This cavity is also found near helix E of bR, thus another possibility is that it may be a part of ion pathway. Indeed, in the light-driven chloride pump halorhodopsin from *N. Pharaonis* (pHR) a channel containing a chain of water molecules is transiently formed in this region in the N-intermediate state. This channel was suggested to be a chloride-release pathway [148].

The absence of krypton atoms observed in the interior of the light-driven sodium pump KR2 protomer supports both the absence of hydrophobic cavities inside the protein and the lack of a need for such cavities for sodium pumping. Indeed, while the extracellular inner region of KR2 is polar, as in many other MRs, in its ground state structure, its cytoplasmic region also contains a large polar ion-uptake cavity, filled with water molecules [18]. The recently reported structure of the O-state of KR2 has shown that the cytoplasmic region is completely restored in the intermediate [26]. Together with the absence of the pronounced M-state, presumably corresponding to the sodium translocation through the cytoplasmic side of the protein, it also suggests the absence of large conformational changes in the cytoplasmic inner part of KR2 during photocycle. All this makes the presence of additional hydrophobic cavities or/and water molecules in the region unnecessary.

Finally, yet importantly, in the tmBR-argon crystal structure two argon atoms were observed in the extracellular part of the protein (Figure 3.3.3.1). These atoms

interact with R82, a key determinant of the vectorial proton translocation, and are located close to E194 and E204, the so-called 'proton release group' (PRG) of bR. The binding of argon causes notable rearrangements of the PRG. This observation supports the high flexibility of the region, was hypothesized and demonstrated recently by the identification of two conformations of the PRG in the ground state of bR [149]. Thus, argon may stabilize one of the conformations, presumably the minor one, which was not identified in the native structure determined here.

## 4. Conclusions

The present thesis work comprised two parts:

1. Structural studies of the full-length *NpSRII/NpHtrII* receptor/transducer complex, with the aim of determining the signal transmission mechanism through the membrane.

2. Binding of Noble gases to membrane proteins.

During the first part of the research plan, an important result was that a genetically optimized construct was created and optimized for the heterologous co-expression of *NpSRII* and full-length (or truncated) *NpHtrII* in *Escherichia coli*. The integrity of the complexes obtained was not compromised during subsequent solubilization in detergent and purification. The complexes thus obtained were then subject to various analyses (electron microscopy, X-ray crystallography, X-ray/neutron solution scattering) in order to try to obtain structural details which might shed light on the signal transduction mechanism of *NpSRII/NpHtrII*.

1. Negative contrast cryo-EM data obtained for full length *NpSRII/NpHtrII* solubilized on detergent mainly contains images of particles with sizes in the region of 50-100 Å, the shape of which corresponds to the TM domains of the *NpSRII/NpHtrII* dimer. The particles form oligomers of various shapes. One can distinguish patterns similar to the "O"-shaped conformation of a trimer of dimers, in which dimer-dimer contacts are also present in the membrane region. However, considering the inconsistency of the "O"-shaped conformation with the results of small-angle scattering experiments, it was concluded that this result is likely to be an artefact associated with the detergent surface around the transmembrane part of the protein.

2. Crystals obtained using the *in meso* crystallization method for the *NpSRII/NpHtrII* complex reconstructed in detergent were characterized by X-ray diffraction analysis. Here, the crystals obtained were either those of *NpSRII* alone or of dimers of *NpSRII/NpHtrII*<sub>184</sub> previously demonstrated in the literature.



Refinement of the structural models obtained was not taken to completion as neither crystal form obtained contained full length *NpSRII/NpHtrII*. The crystal form of *NpSRII/NpHtrII*<sub>184</sub> obtained suggests that *NpHtrII* underwent proteolysis during the crystallization procedure. In this work, potential sites of transducer proteolysis have been identified and it seems necessary to create mutations in the transducer region in the amino acid residue range 84 to 96 to avoid transducer proteolysis in future crystallization experiments.

3. Structural studies of *NpSRII/NpHtrII* reconstituted in nanodiscs was then carried out. Here, crystallization experiments resulted in crystals with a unit cell dimension consistent with the expected longest dimension of full length *NpSRII/NpHtrII*. However, the crystals obtained diffracted only to very low resolution and the analysis was not carried further. Electron microscopy studies of *NpSRII/NpHtrII* reconstituted in nanodiscs showed the samples to exhibit a high degree of inhomogeneity and resulted in a very low resolution reconstruction. This may be due to the destabilization of nanodiscs and/or partial aggregation of membrane proteins embedded in the nanodiscs, perhaps caused by the fact that the complex/MSP = 3/1 ratio is not optimal. This conclusion is confirmed by the SANS experiments (see below), which are in the best agreement with the "tripod" model.

4. Although not high-resolution structural techniques, X-ray/neutron solution scattering studies (SAXS/SANS) were also carried out with the aim of providing a structural model of *NpSRII/NpHtrII* at close to physiological conditions. Under conditions of low ionic strength (150 mM NaCl), solution scattering experiments showed that *NpSRII/NpHtrII* forms dimers which are partially disordered and highly dynamic. This result showed that *NpSRII/NpHtrII* is incorrectly oligomerized in low-molecular-weight salt buffers and that high salt buffers are necessary for this.

5. Under conditions of high ionic strength (4 M NaCl, close to physiological conditions), solution scattering experiments showed that *NpSRII/NpHtrII* forms physiologically relevant 'tripod-shaped' trimers of dimers in which dimer-dimer contacts are formed only in the cytoplasmic part of the protein. Under these

conditions the proportion of amino acid residues involved in forming  $\alpha$ -helices is 63%. These experiments thus showed that high salt conditions are crucial for structural studies of the physiologically relevant oligomeric form of *NpSRII/NpHtrII*. Future structural studies should thus make use of this knowledge.

Structural studies of *NpSRII/NpHtrII* proved even more complex problem than expected. To obtain high-resolution structural data using X-ray crystallography will require the elimination of the cause of proteolysis of the transducer regions during crystallization trials by introducing mutations and characterizing the effects of these before carrying out crystallization trials. Electron microscopy studies will also require a different approach. The high salt conditions needed to maintain *NpSRII/NpHtrII* in a physiologically relevant state will require the use of negative contrasting (in cryo mode) with ammonium molybdate. This will increase the ionic strength of the protein solution, leading to proper folding. Also, due to the understanding of oligomerization's dependence on salt concentration, it is possible to perform functional tests depending on the salt concentration in the buffer. This will provide a complete understanding of the signaling process depending on the complex's oligomeric state.

A striking result of the second part of this work is that atoms of noble gases readily interact with the external hydrophobic surface of MPs by nonspecific binding and can compete with the binding of lipids in this environment. However, several factors (for example, structure resolution, a crystal system based on a detergent, gas pressure) likely limited the number of noble gas atoms observed in these studies, which thus probably did not reveal fully the possible extensive interaction of these atoms with the hydrophobic surface of MPs. Moreover, although an accurate experimental approach, MX has limitations. In particular crystal contacts can block access to areas of the MP surface.

MD experiments were thus performed to check whether *in silico* methods can generate reliable information on the binding positions of noble gas atoms on the MP

surface. Overall, the MD method has shown a reliable result and can thus be used to predict a noble gas position on the MP surface.

Noble gases interact with proteins through weak van der Waals forces, mostly based on the London interaction (spontaneously induced dipoles). Therefore, the binding sites are of a hydrophobic nature (apolar pockets). Thus, the noble gas binding process on the hydrophobic surface of MPs is a good imitation (or competitor) of protein-lipid interaction.

Although the work described here shows unequivocally that atoms of noble gases can replace lipids at their binding sites on MPs, it remains unclear how this might affect MP function. In order to better understand this, it would be essential to carry out structural studies of viral rhodopsins [15, 16], which present transmembrane channels closed by lipid molecules. Another class of MPs to study would be GPCR proteins [150], which are more suitable objects for research than rhodopsins due to their more flexible surfaces. It would also be interesting to continue the study of GLIC protein with other inert gases [75].

The work was done in collaboration with various groups. I directly performed the expression and purification of *Np*SrII/*Np*HtrII, crystallization and X-ray studies, and sample preparation for EM experiments. I also participated in the electron microscopy, X-ray diffraction and SAXS experiments described in this work and was involved in the subsequent processing of data and interpretation of results.

Concerning the section of the work describing noble gas binding to membrane proteins, the purified proteins studied were gifts from collaborators in the V. Gordeliy group. I participated in the crystallization of the proteins, preparation of crystals for noble gas derivatization and X-ray diffraction measurements, the collection of diffraction data and their subsequent processing (after publishing PDB IDs will be 7Q35, 7Q36, 7Q37 and 7Q38).

## References

1. Gushchin, I. and V. Gordeliy, *Transmembrane Signal Transduction in Two-Component Systems: Piston, Scissoring, or Helical Rotation?* *Bioessays*, 2018. **40**(2): p. 1700197.
2. Gómez-Consarnau, L., et al., *Microbial rhodopsins are major contributors to the solar energy captured in the sea.* *Science advances*, 2019. **5**(8): p. eaaw8855.
3. Oesterhelt, D. and W. Stoeckenius, *Rhodopsin-like protein from the purple membrane of Halobacterium halobium.* *Nature new biology*, 1971. **233**(39): p. 149-152.
4. Luecke, H., et al., *Structure of bacteriorhodopsin at 1.55 Å resolution.* *Journal of molecular biology*, 1999. **291**(4): p. 899-911.
5. Gerwert, K., E. Freier, and S. Wolf, *The role of protein-bound water molecules in microbial rhodopsins.* *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 2014. **1837**(5): p. 606-613.
6. Gushchin, I., et al., *Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria.* *Proc Natl Acad Sci U S A*, 2013. **110**(31): p. 12631-6.
7. Volkov, O., et al., *Structural insights into ion conduction by channelrhodopsin 2.* *Science*, 2017. **358**(6366).
8. Kim, Y.S., et al., *Crystal structure of the natural anion-conducting channelrhodopsin Gt ACRI.* *Nature*, 2018. **561**(7723): p. 343-348.
9. Kato, H.E., et al., *Structural mechanisms of selectivity and gating in anion channelrhodopsins.* *Nature*, 2018. **561**(7723): p. 349-354.
10. Oda, K., et al., *Crystal structure of the red light-activated channelrhodopsin Chrimson.* *Nature communications*, 2018. **9**(1): p. 1-11.

11. Li, P., et al., *Allele-specific CRISPR-Cas9 genome editing of the single-base P23H mutation for rhodopsin-associated dominant retinitis pigmentosa*. The CRISPR journal, 2018. **1**(1): p. 55-64.
12. Kato, H.E., et al., *Structural basis for Na<sup>+</sup> transport mechanism by a light-driven Na<sup>+</sup> pump*. Nature, 2015. **521**(7550): p. 48-53.
13. Gushchin, I., et al., *Crystal structure of a light-driven sodium pump*. Nature structural & molecular biology, 2015. **22**(5): p. 390.
14. Kim, K., et al., *Crystal structure and functional characterization of a light-driven chloride pump having an NTQ motif*. Nature communications, 2016. **7**(1): p. 1-10.
15. Bratanov, D., et al., *Unique structure and function of viral rhodopsins*. Nature communications, 2019. **10**(1): p. 1-13.
16. Zabelskii, D., et al., *Viral rhodopsins 1 are an unique family of light-gated cation channels*. Nature communications, 2020. **11**(1): p. 1-16.
17. Needham, D.M., et al., *A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators*. Proceedings of the National Academy of Sciences, 2019. **116**(41): p. 20574-20583.
18. Kovalev, K., et al., *High-resolution structural insights into the heliorhodopsin family*. Proceedings of the National Academy of Sciences, 2020. **117**(8): p. 4131-4141.
19. Lu, Y., et al., *Crystal structure of heliorhodopsin 48C12*. Cell research, 2020. **30**(1): p. 88-90.
20. Shihoya, W., et al., *Crystal structure of heliorhodopsin*. Nature, 2019. **574**(7776): p. 132-136.
21. Gordeliy, V.I., et al., *Molecular basis of transmembrane signalling by sensory rhodopsin II–transducer complex*. Nature, 2002. **419**(6906): p. 484.
22. Klare, J.P., I. Chizhov, and M. Engelhard, *Microbial rhodopsins: scaffolds for ion pumps, channels, and sensors*, in *Bioenergetics*. 2007, Springer. p. 73-122.

23. Moukhametzianov, R., et al., *Development of the signal in sensory rhodopsin and its transfer to the cognate transducer*. Nature, 2006. **440**(7080): p. 115.
24. Ishchenko, A., et al., *Ground state structure of D75N mutant of sensory rhodopsin II in complex with its cognate transducer*. J Photochem Photobiol B, 2013. **123**: p. 55-8.
25. Ishchenko, A., et al., *New insights on signal propagation by sensory rhodopsin II/transducer complex*. Scientific reports, 2017. **7**: p. 41811.
26. Kovalev, K., et al., *Molecular mechanism of light-driven sodium pumping*. Nature communications, 2020. **11**(1): p. 1-11.
27. Gushchin, I., et al., *Active state of sensory rhodopsin II: structural determinants for signal transfer and proton pumping*. Journal of molecular biology, 2011. **412**(4): p. 591-600.
28. Hoff, W.D., K.-H. Jung, and J.L. Spudich, *Molecular mechanism of photosignaling by archaeal sensory rhodopsins*. Annual review of biophysics and biomolecular structure, 1997. **26**(1): p. 223-258.
29. Mazé, A. and Y. Benenson, *Artificial signaling in mammalian cells enabled by prokaryotic two-component system*. Nature Chemical Biology, 2020. **16**(2): p. 179-187.
30. Bem, A.E., et al., *Bacterial histidine kinases as novel antibacterial drug targets*. ACS chemical biology, 2015. **10**(1): p. 213-224.
31. Hazelbauer, G.L., J.J. Falke, and J.S. Parkinson, *Bacterial chemoreceptors: high-performance signaling in networked arrays*. Trends in biochemical sciences, 2008. **33**(1): p. 9-19.
32. Koch, M.K., et al., *Physiological sites of deamidation and methyl esterification in sensory transducers of Halobacterium salinarum*. Journal of molecular biology, 2008. **380**(2): p. 285-302.
33. Akkaladevi, N., et al., *Flexible hinges in bacterial chemoreceptors*. Journal of bacteriology, 2018. **200**(5).
34. Aravind, L. and C.P. Ponting, *The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many*

- prokaryotic signalling proteins*. FEMS Microbiol Lett, 1999. **176**(1): p. 111-6.
35. Orekhov, P.S., et al., *Signaling and Adaptation Modulate the Dynamics of the Photosensory Complex of Natronomonas pharaonis*. PLoS Comput Biol, 2015. **11**(10): p. e1004561.
  36. Bartelli, N.L. and G.L. Hazelbauer, *Differential backbone dynamics of companion helices in the extended helical coiled-coil domain of a bacterial chemoreceptor*. Protein Science, 2015. **24**(11): p. 1764-1776.
  37. Li, M., et al., *Chemotaxis kinase CheA is activated by three neighbouring chemoreceptor dimers as effectively as by receptor clusters*. Molecular microbiology, 2011. **79**(3): p. 677-685.
  38. Li, M. and G.L. Hazelbauer, *Core unit of chemotaxis signaling complexes*. Proceedings of the National Academy of Sciences, 2011. **108**(23): p. 9390-9395.
  39. Parkinson, J.S., G.L. Hazelbauer, and J.J. Falke, *Signaling and sensory adaptation in Escherichia coli chemoreceptors: 2015 update*. Trends in microbiology, 2015. **23**(5): p. 257-266.
  40. Sourjik, V. and J.P. Armitage, *Spatial organization in bacterial chemotaxis*. The EMBO journal, 2010. **29**(16): p. 2724-2733.
  41. Bray, D., M.D. Levin, and C.J. Morton-Firth, *Receptor clustering as a cellular mechanism to control sensitivity*. Nature, 1998. **393**(6680): p. 85-88.
  42. Milburn, M.V., et al., *Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand*. Science, 1991. **254**(5036): p. 1342-1347.
  43. Kim, K.K., H. Yokota, and S.-H. Kim, *Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor*. Nature, 1999. **400**(6746): p. 787-792.
  44. Airola, M.V., et al., *Structure of concatenated HAMP domains provides a mechanism for signal transduction*. Structure, 2010. **18**(4): p. 436-448.



45. Gushchin, I., et al., *Mechanism of transmembrane signaling by sensor histidine kinases*. Science, 2017. **356**(6342): p. eaah6345.
46. Yang, W., et al., *In situ conformational changes of the Escherichia coli serine chemoreceptor in different signaling states*. MBio, 2019. **10**(4): p. e00973-19.
47. Burt, A., et al., *Complete structure of the chemosensory array core signalling unit in an E. coli minicell strain*. Nature communications, 2020. **11**(1): p. 1-9.
48. Bollschweiler, D., et al., *Cryo-electron microscopy of an extremely halophilic microbe: technical aspects*. Extremophiles, 2017. **21**(2): p. 393-398.
49. Adrian, M., et al., *Cryo-negative staining*. Micron, 1998. **29**(2-3): p. 145-160.
50. De Carlo, S. and H. Stark, *Cryonegative staining of macromolecular assemblies*, in *Methods in enzymology*. 2010, Elsevier. p. 127-145.
51. De Carlo, S. and J.R. Harris, *Negative staining and cryo-negative staining of macromolecules and viruses for TEM*. Micron, 2011. **42**(2): p. 117-131.
52. Zhang, P., et al., *Direct visualization of Escherichia coli chemotaxis receptor arrays using cryo-electron microscopy*. Proceedings of the National Academy of Sciences, 2007. **104**(10): p. 3777-3781.
53. Orekhov, P., et al., *Sensory Rhodopsin I and Sensory Rhodopsin II Form Trimers of Dimers in Complex with their Cognate Transducers*. Photochem Photobiol, 2017. **93**(3): p. 796-804.
54. Hirayama, J., et al., *Photocycle of phoborhodopsin from haloalkaliphilic bacterium (Natronobacterium pharaonis) studied by low-temperature spectrophotometry*. Biochemistry, 1992. **31**(7): p. 2093-2098.
55. Chizhov, I., et al., *The photophobic receptor from Natronobacterium pharaonis: temperature and pH dependencies of the photocycle of sensory rhodopsin II*. Biophysical Journal, 1998. **75**(2): p. 999-1009.
56. Falb, M., et al., *Living with two extremes: conclusions from the genome sequence of Natronomonas pharaonis*. Genome research, 2005. **15**(10): p. 1336-1343.

57. Budyak, I.L., et al., *Shape and oligomerization state of the cytoplasmic domain of the phototaxis transducer II from Natronobacterium pharaonis*. Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15428-33.
58. Budyak, I.L., et al., *Flexibility of the Cytoplasmic Domain of the Phototaxis Transducer II from Natronomonas pharaonis*. Journal of Biophysics, 2008. **2008**.
59. Doebber, M., et al., *Salt-driven equilibrium between two conformations in the HAMP domain from Natronomonas pharaonis: the language of signal transfer?* J Biol Chem, 2008. **283**(42): p. 28691-701.
60. Hulko, M., et al., *The HAMP domain structure implies helix rotation in transmembrane signaling*. Cell, 2006. **126**(5): p. 929-40.
61. Berthaud, A., et al., *Modeling detergent organization around aquaporin-0 using small-angle X-ray scattering*. Journal of the American Chemical Society, 2012. **134**(24): p. 10080-10088.
62. Hernando, M., et al., *Solution structure and oligomeric state of the E. coliglycerol facilitator*. Biochimica et Biophysica Acta (BBA)-Biomembranes, 2020. **1862**(5): p. 183191.
63. Ginsburg, A., et al., *D+: software for high-resolution hierarchical modeling of solution X-ray scattering from complex structures*. Journal of applied crystallography, 2019. **52**(1): p. 219-242.
64. Cooley, R.B., J.P. O'Donnell, and H. Sondermann, *Coincidence detection and bi-directional transmembrane signaling control a bacterial second messenger receptor*. Elife, 2016. **5**: p. e21848.
65. Wojtowicz, H., et al., *Structural basis of the signalling through a bacterial membrane receptor HasR deciphered by an integrative approach*. Biochemical Journal, 2016. **473**(14): p. 2239-2248.
66. Gao, Y., et al., *Isolation and structure–function characterization of a signaling-active rhodopsin–G protein complex*. Journal of Biological Chemistry, 2017. **292**(34): p. 14280-14289.

67. Chen, P.-c. and J.S. Hub, *Structural properties of protein–detergent complexes from SAXS and MD simulations*. The journal of physical chemistry letters, 2015. **6**(24): p. 5116-5121.
68. Jessen, C.M., et al., *Mapping of unfolding states of integral helical membrane proteins by GPS-NMR and scattering techniques: TFE-induced unfolding of KcsA in DDM surfactant*. Biochimica et Biophysica Acta (BBA)-Biomembranes, 2012. **1818**(9): p. 2290-2301.
69. Pérez, J. and A. Koutsioubas, *Memprot: a program to model the detergent corona around a membrane protein based on SEC–SAXS data*. Acta Crystallographica Section D: Biological Crystallography, 2015. **71**(1): p. 86-93.
70. Svergun, D., C. Barberato, and M.H. Koch, *CRY SOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates*. Journal of applied crystallography, 1995. **28**(6): p. 768-773.
71. Petoukhov, M.V., et al., *New developments in the ATSAS program package for small-angle scattering data analysis*. Journal of applied crystallography, 2012. **45**(2): p. 342-350.
72. Molodenskiy, D.S., H.D. Mertens, and D.I. Svergun, *An automated data processing and analysis pipeline for transmembrane proteins in detergent solutions*. Scientific reports, 2020. **10**(1): p. 1-11.
73. De Maria Antolinos, A., et al., *ISPyB for BioSAXS, the gateway to user autonomy in solution scattering experiments*. Acta Crystallographica Section D: Biological Crystallography, 2015. **71**(1): p. 76-85.
74. Winkler, D., et al., *Massive in silico study of noble gas binding to the structural proteome*. Journal of chemical information and modeling, 2019. **59**(11): p. 4844-4854.
75. Sauguet, L., et al., *Structural basis for xenon inhibition in a cationic pentameric ligand-gated ion channel*. PLoS One, 2016. **11**(2): p. e0149795.
76. Ntsogo Enguéné, Y.V., et al., *Xenon for tunnelling analysis of the efflux pump component OprN*. PloS one, 2017. **12**(9): p. e0184045.

77. Andrade, S.L., et al., *Crystal structure of the archaeal ammonium transporter Amt-1 from Archaeoglobus fulgidus*. Proceedings of the National Academy of Sciences, 2005. **102**(42): p. 14994-14999.
78. Herczyński, A., *Bound charges and currents*. American Journal of Physics, 2013. **81**(3): p. 202-205.
79. <https://en.wikipedia.org/wiki/Nonmetal>.
80. <https://chem.fsu.edu/chemlab/chm1046course/interforces.html>.
81. Schiltz, M., R. Fourme, and T. Prangé, *Use of noble gases xenon and krypton as heavy atoms in protein structure determination*. Methods in enzymology, 2003. **374**: p. 83-119.
82. Lafumat, B., et al., *Gas-sensitive biological crystals processed in pressurized oxygen and krypton atmospheres: deciphering gas channels in proteins using a novelsoak-and-freeze'methodology*. Journal of Applied Crystallography, 2016. **49**(5): p. 1478-1487.
83. Linstrom, P.J. and W.G. Mallard, *The NIST Chemistry WebBook: A chemical data resource on the internet*. Journal of Chemical & Engineering Data, 2001. **46**(5): p. 1059-1063.
84. Taylor, G.L., *Introduction to phasing*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(4): p. 325-338.
85. Wang, B.-C., *Resolution of phase ambiguity in macromolecular crystallography*. Methods in enzymology, 1985. **115**: p. 90-112.
86. [http://skuld.bmsc.washington.edu/scatter/AS\\_periodic.html](http://skuld.bmsc.washington.edu/scatter/AS_periodic.html).
87. Nurizzo, D., et al., *The ID23-1 structural biology beamline at the ESRF*. Journal of synchrotron radiation, 2006. **13**(3): p. 227-238.
88. McCarthy, A.A., et al., *ID30B—a versatile beamline for macromolecular crystallography experiments at the ESRF*. Journal of synchrotron radiation, 2018. **25**(4): p. 1249-1260.
89. Finke, A.D., et al., *Advanced crystallographic data collection protocols for experimental phasing*, in *Nucleic Acid Crystallography*. 2016, Springer. p. 175-191.

90. Schoenborn, B.P., H.C. Watson, and J.C. Kendrew, *Binding of xenon to sperm whale myoglobin*. *Nature*, 1965. **207**(4992): p. 28-30.
91. Schoenborn, B. and C. Nobbs, *The binding of xenon to sperm whale deoxymyoglobin*. *Molecular pharmacology*, 1966. **2**(5): p. 495-498.
92. Schoenborn, B.P., *Structure of alkaline metmyoglobin-xenon complex*. *Journal of molecular biology*, 1969. **45**(2): p. 297-303.
93. Tilton Jr, R.F., I.D. Kuntz Jr, and G.A. Petsko, *Cavities in proteins: structure of a metmyoglobin xenon complex solved to 1.9 Å*. *Biochemistry*, 1984. **23**(13): p. 2849-2857.
94. Schoenborn, B.P., *Binding of Cyclo propane to Sperm Whale Myoglobin*. *Nature*, 1967. **214**(5093): p. 1120-1122.
95. Vitali, J., et al., *Using xenon as a heavy atom for determining phases in sperm whale metmyoglobin*. *Journal of applied crystallography*, 1991. **24**(5): p. 931-935.
96. Fourme, R., et al., *Better structures from better data through better methods: a review of developments in de novo macromolecular phasing techniques and associated instrumentation at LURE*. *Journal of synchrotron radiation*, 1999. **6**(4): p. 834-844.
97. Kim, C.U., Q. Hao, and S.M. Gruner, *Solution of protein crystallographic structures by high-pressure cryocooling and noble-gas phasing*. *Acta Crystallographica Section D: Biological Crystallography*, 2006. **62**(7): p. 687-694.
98. Hilf, R.J. and R. Dutzler, *X-ray structure of a prokaryotic pentameric ligand-gated ion channel*. *Nature*, 2008. **452**(7185): p. 375-379.
99. Murray, J., et al., *Analysis of xenon binding to photosystem II by X-ray crystallography*. *Photosynthesis research*, 2008. **98**(1): p. 523-527.
100. Sauer, O., et al., *Low-resolution detergent tracing in protein crystals using xenon or krypton to enhance X-ray contrast*. *Acta Crystallographica Section D: Biological Crystallography*, 2002. **58**(1): p. 60-69.

101. <https://www.thermofisher.com/fr/fr/home/global/forms/geneart-genes-synthesis-requests.html>. Available from: <https://www.thermofisher.com/fr/fr/home/global/forms/geneart-genes-synthesis-requests.html>.
102. Bayburt, T.H. and S.G. Sligar, *Membrane protein assembly into Nanodiscs*. FEBS letters, 2010. **584**(9): p. 1721-1727.
103. Inagaki, S., R. Ghirlando, and R. Grishammer, *Biophysical characterization of membrane proteins in nanodiscs*. Methods, 2013. **59**(3): p. 287-300.
104. Caffrey, M. and V. Cherezov, *Crystallizing membrane proteins using lipidic mesophases*. Nature protocols, 2009. **4**(5): p. 706-731.
105. Pernot, P., et al., *Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution*. Journal of synchrotron radiation, 2013. **20**(4): p. 660-664.
106. Dragunov, Y.G., et al., *Modernization of the IBR-2 pulsed research reactor*. Atomic Energy, 2012. **113**(1): p. 29-38.
107. Kuklin, A., A.K. Islamov, and V. Gordeliy, *Scientific reviews: Two-detector system for small-angle neutron scattering instrument*. Neutron News, 2005. **16**(3): p. 16-18.
108. Kuklin, A., et al. *Neutronographic investigations of supramolecular structures on upgraded small-angle spectrometer YuMO*. in *Journal of Physics: Conference Series*. 2017. IOP Publishing.
109. Manalastas-Cantos, K., et al., *ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis*. Journal of Applied Crystallography. **54**(1).
110. Nielsen, S., et al., *BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis*. Journal of applied crystallography, 2009. **42**(5): p. 959-964.
111. Semenyuk, A. and D. Svergun, *GNOM—a program package for small-angle scattering data processing*. Journal of applied crystallography, 1991. **24**(5): p. 537-540.

112. Svergun, D., et al., *Protein hydration in solution: experimental observation by x-ray and neutron scattering*. Proceedings of the National Academy of Sciences, 1998. **95**(5): p. 2267-2272.
113. Bratanov, D., et al., *An approach to heterologous expression of membrane proteins. The case of bacteriorhodopsin*. PLoS One, 2015. **10**(6): p. e0128390.
114. Melnikov, I., et al., *Fast iodide-SAD phasing for high-throughput membrane protein structure determination*. Sci Adv, 2017. **3**(5): p. e1602952.
115. Landau, E.M. and J.P. Rosenbusch, *Lipidic cubic phases: a novel concept for the crystallization of membrane proteins*. Proceedings of the National Academy of Sciences, 1996. **93**(25): p. 14532-14535.
116. de Sanctis, D., et al., *ID29: a high-intensity highly automated ESRF beamline for macromolecular crystallography experiments exploiting anomalous scattering*. Journal of synchrotron radiation, 2012. **19**(3): p. 455-461.
117. Cipriani, F., et al., *Automation of sample mounting for macromolecular crystallography*. Acta Crystallographica Section D: Biological Crystallography, 2006. **62**(10): p. 1251-1259.
118. Zander, U., et al., *MeshAndCollect: an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines*. Acta Crystallographica Section D: Biological Crystallography, 2015. **71**(11): p. 2328-2343.
119. Melnikov, I., et al., *The complex analysis of X-ray mesh scans for macromolecular crystallography*. Acta Crystallographica Section D: Structural Biology, 2018. **74**(4): p. 355-365.
120. Bourenkov, G.P. and A.N. Popov, *Optimization of data collection taking radiation damage into account*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(4): p. 409-419.
121. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(2): p. 133-144.



122. Evans, P., *Scaling and assessment of data quality*. Acta Crystallographica Section D: Biological Crystallography, 2006. **62**(1): p. 72-82.
123. Kovalev, K., et al., *Structure and mechanisms of sodium-pumping KR2 rhodopsin*. Science advances, 2019. **5**(4): p. eaav2671.
124. Borshchevskiy, V., et al., *Isoprenoid-chained lipid  $\beta$ -Xyloc16+ 4—A novel molecule for in meso membrane protein crystallization*. Journal of crystal growth, 2010. **312**(22): p. 3326-3330.
125. Murshudov, G.N., et al., *Efficient anisotropic refinement of macromolecular structures using FFT*. Acta Crystallographica Section D: Biological Crystallography, 1999. **55**(1): p. 247-255.
126. Emsley, P., et al., *Features and development of Coot*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(4): p. 486-501.
127. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures*. Acta Crystallographica Section D: Biological Crystallography, 2011. **67**(4): p. 355-367.
128. Mark, P. and L. Nilsson, *Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K*. The Journal of Physical Chemistry A, 2001. **105**(43): p. 9954-9960.
129. Jo, S., et al., *CHARMM-GUI: a web-based graphical user interface for CHARMM*. Journal of computational chemistry, 2008. **29**(11): p. 1859-1865.
130. McDonald, I., *NpT-ensemble Monte Carlo calculations for binary liquid mixtures*. Molecular Physics, 1972. **23**(1): p. 41-58.
131. Lemak, A. and N. Balabaev, *On the Berendsen thermostat*. Molecular Simulation, 1994. **13**(3): p. 177-187.
132. Evans, D.J. and B.L. Holian, *The nose–hoover thermostat*. The Journal of chemical physics, 1985. **83**(8): p. 4069-4074.
133. Saito, H., et al., *Molecular collective dynamics in solid para-hydrogen and ortho-deuterium: The Parrinello–Rahman-type path integral centroid molecular dynamics approach*. The Journal of chemical physics, 2003. **119**(2): p. 953-963.

134. Pronk, S., et al., *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit*. Bioinformatics, 2013. **29**(7): p. 845-854.
135. Huang, J., et al., *CHARMM36m: an improved force field for folded and intrinsically disordered proteins*. Nature methods, 2017. **14**(1): p. 71-73.
136. Vrabec, J., J. Stoll, and H. Hasse, *A set of molecular models for symmetric quadrupolar fluids*. The Journal of Physical Chemistry B, 2001. **105**(48): p. 12126-12133.
137. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*. Journal of molecular graphics, 1996. **14**(1): p. 33-38.
138. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination*. Nature methods, 2017. **14**(3): p. 290-296.
139. Ryzhykau, Y.L., et al., *Molecular model of a sensor of two-component signaling system*. Scientific reports, 2021. **11**(1): p. 1-15.
140. Bayburt, T.H., Y.V. Grinkova, and S.G. Sligar, *Assembly of single bacteriorhodopsin trimers in bilayer nanodiscs*. Archives of biochemistry and biophysics, 2006. **450**(2): p. 215-222.
141. Garcia-Diez, R., C. Gollwitzer, and M. Krumrey, *Nanoparticle characterization by continuous contrast variation in SAXS with a solvent density gradient*. arXiv preprint arXiv:1409.1776, 2014.
142. Li, M., et al., *Morphological characterization of DMPC/CHAPSO bicellar mixtures: a combined SANS and NMR study*. Langmuir, 2013. **29**(51): p. 15943-15957.
143. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3*. elife, 2018. **7**: p. e42166.
144. Joshi, M.K., et al., *Importance of specific native lipids in controlling the photocycle of bacteriorhodopsin*. Biochemistry, 1998. **37**(41): p. 14463-14470.
145. Hayakawa, N., et al., *Effect of xenon binding to a hydrophobic cavity on the proton pumping cycle in bacteriorhodopsin*. Journal of molecular biology, 2008. **384**(4): p. 812-823.









146. Quillin, M.L., et al., *Size versus polarizability in protein-ligand interactions: binding of noble gases within engineered cavities in phage T4 lysozyme*. *Journal of molecular biology*, 2000. **302**(4): p. 955-977.
147. Trudell, J.R. *A unitary theory of anesthesia based on lateral phase separations in nerve membranes*. in *The Journal of the American Society of Anesthesiologists*. 1977. The American Society of Anesthesiologists.
148. Kouyama, T., et al., *Crystal structures of the L1, L2, N, and O states of pharaonis halorhodopsin*. *Biophysical journal*, 2015. **108**(11): p. 2680-2690.
149. Hasegawa, N., et al., *X-ray structure analysis of bacteriorhodopsin at 1.3 Å resolution*. *Scientific reports*, 2018. **8**(1): p. 1-8.
150. Lerch, M.T., et al., *Viewing rare conformations of the  $\beta$ 2 adrenergic receptor with pressure-resolved DEER spectroscopy*. *Proceedings of the National Academy of Sciences*, 2020. **117**(50): p. 31824-31831.

## Publications

1. Kovalev, K., Astashkin, R., Gushchin, I., Orekhov, P., Volkov, D., Zinovev, E., ... & Gordeliy, V. *Molecular mechanism of light-driven sodium pumping*. Nature communications, 2020, **11**(1), 1-11.
2. Maliar, N., Kovalev, K., Baeken, C., Balandin, T., Astashkin, R., Rulev, M., ... & Gordeliy, V. *Crystal Structure of the N112A Mutant of the Light-Driven Sodium Pump KR2*. Crystals, 2020, **10**(6), 496.
3. Zabelskii, D., Alekseev, A., Kovalev, K., Rankovic, V., Balandin, T., Soloviov, D., ... & Gordeliy, V. *Viral rhodopsins I are an unique family of light-gated cation channels*. Nature communications, 2020, **11**(1), 1-16.
4. Ryzhykau, Y. L., Orekhov, P. S., Rulev, M. I., Vlasov, A. V., Melnikov, I. A., Volkov, D. A., ... & Gordeliy, V. I. *Molecular model of a sensor of two-component signaling system*. Scientific Reports, 2020, **11**(1), 1-15

For the first and second articles I performed the determination of the oligomeric state of the protein depending on pH of the solution. For the third article I performed with protein crystallization and partial collection of diffraction data. In fourth article I purified and expressed the *NpSRII/NpHtrII* samples used and prepared these for SAXS experiments. I participated in the SAXS experiments, including the processing of the data obtained. I also contributed to the writing of the manuscript, including discussion of the results obtained in SANS experiments.

# Molecular mechanism of light-driven sodium pumping

Kirill Kovalev<sup>1,2,3,4,5</sup>, Roman Astashkin<sup>1,4</sup>, Ivan Gushchin <sup>4</sup>, Philipp Orekhov <sup>4</sup>, Dmytro Volkov <sup>2,3</sup>, Egor Zinovev<sup>2,3,4</sup>, Egor Marin <sup>4</sup>, Maksim Rulev<sup>2,3,6</sup>, Alexey Alekseev<sup>2,3,4,5</sup>, Antoine Royant<sup>1,6</sup>, Philippe Carpentier<sup>6,7</sup>, Svetlana Vaganova<sup>2,3</sup>, Dmitrii Zabelskii <sup>2,3,4</sup>, Christian Baeken<sup>2,3</sup>, Ilya Sergeev<sup>4</sup>, Taras Balandin<sup>2,3</sup>, Gleb Bourenkov<sup>8</sup>, Xavier Carpena <sup>9</sup>, Roeland Boer <sup>9</sup>, Nina Maliar <sup>4</sup>, Valentin Borshchevskiy<sup>2,3,4</sup>, Georg Büldt<sup>4</sup>, Ernst Bamberg<sup>4,10</sup> & Valentin Gordeliy<sup>1,2,3,4</sup>✉

The light-driven sodium-pumping rhodopsin KR2 from *Krokinobacter eikastus* is the only non-proton cation active transporter with demonstrated potential for optogenetics. However, the existing structural data on KR2 correspond exclusively to its ground state, and show no sodium inside the protein, which hampers the understanding of sodium-pumping mechanism. Here we present crystal structure of the O-intermediate of the physiologically relevant pentameric form of KR2 at the resolution of 2.1 Å, revealing a sodium ion near the retinal Schiff base, coordinated by N112 and D116 of the characteristic NDQ triad. We also obtained crystal structures of D116N and H30A variants, conducted metadynamics simulations and measured pumping activities of putative pathway mutants to demonstrate that sodium release likely proceeds alongside Q78 towards the structural sodium ion bound between KR2 protomers. Our findings highlight the importance of pentameric assembly for sodium pump function, and may be used for rational engineering of enhanced optogenetic tools.

<sup>1</sup>Institut de Biologie Structurale (IBS), Université Grenoble Alpes, CEA, CNRS, Grenoble, France. <sup>2</sup>Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich GmbH, Jülich, Germany. <sup>3</sup>JuStruct: Jülich Center for Structural Biology, Forschungszentrum Jülich GmbH, Jülich, Germany. <sup>4</sup>Research Center for Molecular Mechanisms of Aging and Age-related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia. <sup>5</sup>Institute of Crystallography, RWTH Aachen University, Aachen, Germany. <sup>6</sup>European Synchrotron Radiation Facility Grenoble, Grenoble, France. <sup>7</sup>Institut de Recherche Interdisciplinaire de Grenoble (IRIG), Laboratoire Chimie et Biologie des Métaux (LCBM), Université Grenoble Alpes, CEA, CNRS, Grenoble, France. <sup>8</sup>European Molecular Biology Laboratory, Hamburg unit c/o DESY, Hamburg, Germany. <sup>9</sup>XALOC beamline, ALBA synchrotron (CELLS), Cerdanyola del Valles, Catalunya, Spain. <sup>10</sup>Max Planck Institute of Biophysics, Frankfurt am Main, Germany. ✉email: [valentin.gordeliy@ibs.fr](mailto:valentin.gordeliy@ibs.fr)

**M**icrobial rhodopsins (MRs) are transmembrane light-sensitive proteins, found in Archaea, Bacteria, Eukaryota, and also viruses<sup>1</sup>. They possess diverse biological functions and are the core of breakthrough biotechnological applications, such as optogenetics. MRs are composed of seven transmembrane  $\alpha$ -helices (A–G) with the cofactor retinal covalently bound to the lysine residue of helix G via the Schiff base (RSB). Due to a conflict of a presence of a cation close to the RSB proton, it was believed that Na<sup>+</sup>-pumping rhodopsins could not exist in nature. Despite this paradigm, the first light-driven Na<sup>+</sup> pump KR2 was identified in *Krokinobacter eikastus* in 2013<sup>2</sup>. Its functional and structural properties were extensively studied<sup>3–5</sup>. KR2 contains a characteristic for all known Na<sup>+</sup>-pumping rhodopsins (NaRs) set of N112, D116, Q123 residues in the helix C (NDQ motif). It was shown that the protein pumps Na<sup>+</sup> when its concentration is much higher than that of H<sup>+</sup>, which is characteristic for physiological conditions, otherwise it acts as a H<sup>+</sup> pump<sup>6</sup>. An extensive mutational analysis of KR2 indicated key functional residues, such as N112, D116, Q123, but also H30, S70, R109, R243, D251, S254, G263<sup>2,3,7,8</sup>. Moreover, potassium-pumping and potassium-channeling variants of KR2 were designed, making the protein a potential tool for optogenetics<sup>3–5,9</sup>.

A key question remains to be answered: what is the mechanism of pumping. Indeed, principles of Na<sup>+</sup> transport by KR2 and other NaRs remain unclear. After light excitation the photocycle starts with the retinal isomerization from all-*trans* to 13-*cis* configuration<sup>10</sup>. In the Na<sup>+</sup>-pumping mode the protein is characterized by the K, L/M, and O intermediates<sup>10</sup> (Fig. 1a). It is known that upon retinal isomerization the proton is translocated from the RSB to the D116 during M-state formation<sup>2,11</sup>. Uptake of sodium occurs in the M-to-O transition. It was hypothesized that Na<sup>+</sup> may pass the cytoplasmic gate comprised by Q123 and the neutralized RSB, and binds in the central region near D116, N112, and presumably D251 transiently in the O-state<sup>4,10,12</sup>. It was also suggested that with the decay of the O-state, Na<sup>+</sup> is released via R109 and the cluster of E11, E160, and R243 to the extracellular space<sup>3,5</sup>. Therefore, the O-state is considered to be the key for elucidation of the Na<sup>+</sup>-pumping mechanism<sup>10</sup>.

It is important that the protein always forms pentamers being reconstituted into lipid membrane<sup>13</sup>. KR2 pentamers appear also under physiological conditions in crystals<sup>3,4,12</sup> and detergent micelles<sup>4,14</sup>. Hence, KR2 is considered to be a pentamer in the native membrane (Fig. 1a). Pentamerization is important for Na<sup>+</sup> pumping by KR2<sup>4</sup>. Particularly, all previous functional investigations of KR2 were performed on the pentameric form of the protein. Moreover, it was recently reported that in the ground state under physiological conditions (PDB ID: 6REW<sup>4</sup>) KR2 has a large water-filled cavity near the RSB (Schiff base cavity). This conformation of the protein was called ‘expanded’<sup>4,12</sup>, and only occurs in the pentameric form of KR2. On the contrary, in the monomeric form the large cavity is absent, and the protein is in another conformation, called ‘compact’<sup>4,12</sup>. Thus, to elucidate the mechanism of light-driven sodium pumping one needs to study the biologically relevant pentameric form of KR2, where it forms the ‘expanded’ conformation in the ground state.

We should note that uncovering of the mechanism of Na<sup>+</sup> pumping is of great importance and simultaneously is a challenge. First of all, it should be remarkably different from that of H<sup>+</sup> pumping<sup>10</sup> and, therefore, huge amount of our knowledge on the H<sup>+</sup> pumping mechanism obtained with a classic proton pump bacteriorhodopsin (BR)<sup>15,16</sup> could not be applied straightforward to NaRs. Not only there is a conflict of simultaneous presence of two positive charges in close proximity, the protonated RSB (RSBH<sup>+</sup>) and Na<sup>+</sup> in the case of KR2, but also there are fundamental differences in the translocation of the

proton and other cations. Namely, proton transport in biological objects implies ion tunneling, which is dramatically hampered in case of larger cations. Furthermore, non-proton cation pumps cannot utilize the Grotthuss mechanism for ion translocation<sup>15</sup>. Hence, pathways of proton in light-driven pumps cannot be the same as those in the cation pumps like NaRs. All history of studies of BR, halorhodopsin (HR) and also sensory rhodopsin II (SRII) shows that high-resolution structures of the intermediate states of a rhodopsin are key for the understanding of the mechanisms. It is even more valid in the case of NaRs, as Na<sup>+</sup> inside the protein is absent in the ground state of KR2, which provides a wide room for speculations on the mechanism of Na<sup>+</sup> transport.

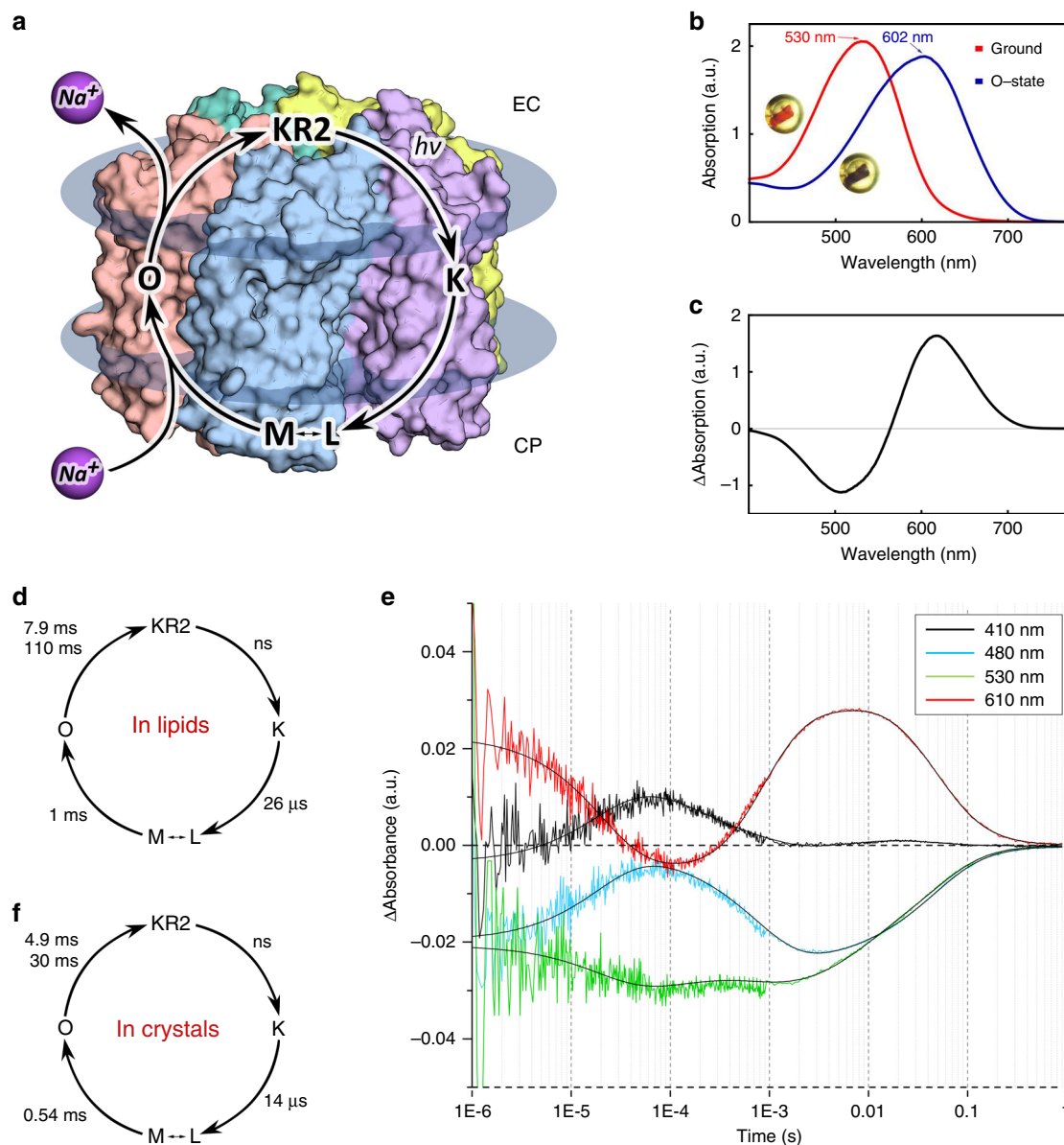
Here we present the structures of the O-state of physiologically relevant pentameric form of KR2—the key intermediate of Na<sup>+</sup> pumping—and functionally important D116N and H30A mutants of the protein. The structure of the O-state reveals Na<sup>+</sup>-binding site inside the rhodopsin and together with the structures of the mutants allows us to elucidate key determinants of cation pumping by KR2.

## Results

**The structure of the O-state of pentameric KR2.** We crystallized KR2 in the functional state (at pH 8.0) using in meso approach similarly to our previous works<sup>4,17,18</sup>. To verify that the protein in crystals undergoes the same photocycle as in lipids, we performed time-resolved visible absorption spectroscopy on KR2 microcrystal slurries. The experiments showed that similar to the protein in detergent micelles and lipids, crystallized KR2 also forms characteristic K-, L/M- and O-states (Fig. 1). Then the O-state was trapped using an approach described in ref. <sup>18</sup>. In brief, we illuminated KR2 crystals with 532-nm laser to freeze-trap the O-state. The cryostream was blocked for 1 s during laser exposure and released back before turning off the laser. This procedure allows accumulation and trapping in crystals of the dominant intermediate of the protein photocycle<sup>18,19</sup>. Single-crystal spectrophotometry showed that after the procedure nearly all proteins in the crystal were trapped in the red-shifted intermediate state (Fig. 1b, c). KR2 undergoes only two red-shifted states during photocycle: the K- and the O-states (Fig. 1f). As the difference electron density maps (described in details below) do not indicate even a low fraction of the ground state in the structure, show the Na<sup>+</sup> bound near the RSB region, which is characteristic for the O-state of KR2, and the O-state is a dominant intermediate of the KR2 photocycle, we consider the trapped intermediate as solely the O-state.

To verify that flash-cooling does not affect the conformation of the O-state, we solved X-ray structures of the illuminated KR2 at room temperature (RT) during continuous 532-nm laser illumination using both single-crystal and serial millisecond crystallography. These approaches allow to detect the dominant conformational changes of the photocycle<sup>19</sup>. We collected RT crystallographic data at 2.6 Å resolution by merging of three complete datasets obtained from three single crystals (see “Methods” section and Supplementary material). We have also collected more than 1.2M diffraction patterns using the stream of mesophase containing KR2 microcrystals injected to the X-ray beam, which were processed to 2.5 Å resolution (see “Methods” section and Supplementary material). Structure refinement identified nearly 1/1 ratio of the ground/O-states populations in crystals in both cases (see Supplementary material).

The structures of the O-state at RT are identical to each other and also to that at 100 K, which indicates that cryo-cooling does not affect the O-state of KR2 (Supplementary Fig. 9). Hence, we describe further only the structure of cryo-trapped intermediate,



**Fig. 1 Spectroscopy of KR2 in crystals.** **a** Scheme of the KR2 photocycle indicates that Na<sup>+</sup> binding occurs transiently in the red-shifted O-state. **b** UV-visible absorption spectra measured in crystallo at 100 K of the Ground state (red) and the O-state of KR2 (insets: photos of the same frozen KR2 crystal in the cryoloop before and after laser illumination, near the corresponding spectra). **c** Difference spectrum calculated between the blue and red spectra shown in **b**. **d** Photocycle of KR2 reconstituted in DOPC. **e** Time traces of absorption changes of KR2 microcrystals at 410 (black), 480 (light blue), 530 (green), and 610 nm (red) probe wavelengths. Black lines indicate fitting lines based on the sequential kinetic model shown in **f**. Photocycle of KR2 in microcrystals, determined in the present work.

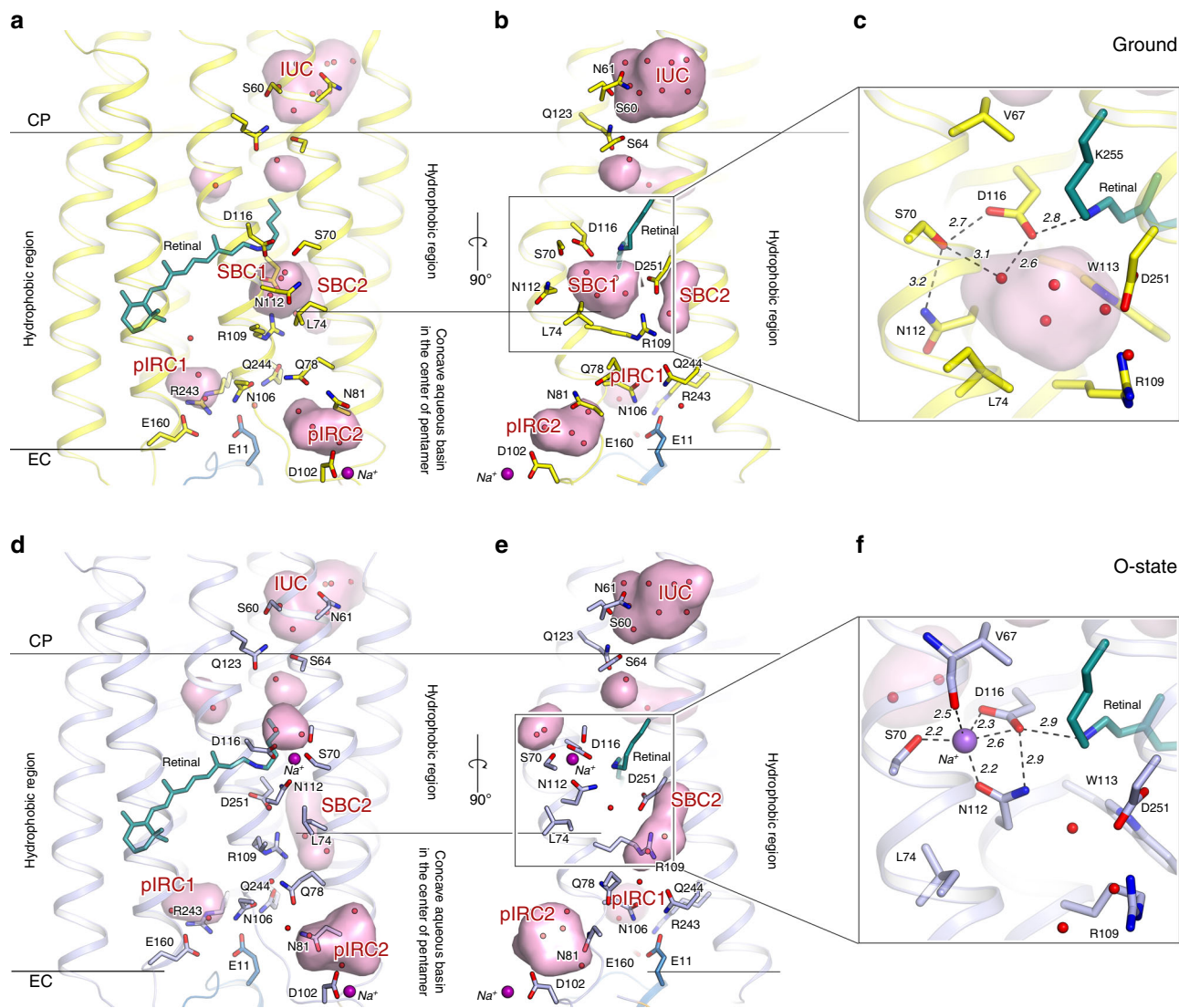
since it has higher resolution (2.1 Å) and occupancy of the O-state (100%).

Using the crystals with the trapped intermediate, we solved the structure of the O-state at 2.1 Å (Supplementary Table 1). The crystal symmetry and lattice parameters are the same as described previously for the ground state of the protein, with one KR2 pentamer in the asymmetric unit<sup>3,4</sup>. The structure demonstrates notable rearrangements compared to the ground state of KR2 (Fig. 2 and Supplementary Fig. 1). The root mean square deviation (RMSD) between the backbone atoms of the pentamers and protomers of the ground (PDB ID: 6REW<sup>4</sup>) and the O-states (present work) are 0.55 and 0.53 Å, respectively. The main changes occur in the extracellular parts of the helices B and C, which are shifted by 1.0 and 1.8 Å, respectively (Supplementary

Fig. 1). Helices A, D and G are also displaced by 0.7 Å in the extracellular regions (Supplementary Fig. 1).

**Retinal-binding pocket of KR2 in the O-state.** The polder<sup>20</sup> electron density maps built around the retinal cofactor strongly suggest its all-trans configuration in the O-state, distorted around C<sub>14</sub> atom (Supplementary Figs. 2 and 3). Indeed, the fitting of the electron density maps with either 13-cis or the mixture of all-trans/13-cis retinal results in the appearing of strong negative peaks of the F<sub>o</sub>-F<sub>c</sub> difference electron maps at the level higher than 3σ. Moreover, our data shows that 13-cis configuration would result in the steric conflict between C<sub>15</sub> and C<sub>20</sub> atoms of the retinal and W113 and W215 residues, respectively. Therefore, all-trans retinal was modeled into final structure of the O-state.





**Fig. 2 Overall comparison of the ground and O-states of KR2.** **a, d** Side view of the KR2 protomer in the ground (yellow, PDB ID: 6REW) and O- (blue, present work) states. **b, e** View from the side of the helices A and B. Membrane hydrophobic/hydrophilic boundaries were calculated using PPM server<sup>56</sup> and are shown with the black lines. The membrane boundary at the extracellular side is located at two levels for the inner and outer parts of the KR2 pentamer, respectively. Helices A and B face the concave aqueous basin, formed in the central pore of the pentamer and helices C–G face the lipid bilayer, surrounding the pentamer. Water molecules are shown as yellow and blue spheres for ground and O-state, respectively. Helices A and B are hidden for clarity. **c, f** Detailed view of the RSB region of the ground and the O-state of KR2. Cavities (ion-uptake cavity—IUC; the Schiff base cavities 1 and 2—SBC1 and SBC2, respectively; putative ion-release cavities 1 and 2—pIRC1 and pIRC2, respectively) inside the protein were calculated using HOLLOW<sup>57</sup> shown in pink and marked with red labels. Retinal cofactor is colored teal. Water molecules are shown with red spheres. Sodium ion is shown with a purple sphere. Hydrogen bonds involving S70, N112, D116, D251, and RSB are shown with black dashed lines. The lengths of the shown hydrogen bonds are shown with bold italic numbers and are in Å. Helix A and SBC2 are hidden for clarity.

The positions of the residues comprising the retinal pocket, particularly W113, D251, D116, I150, Y218, and W215 are correspondingly shifted relative to those in the ground state (Supplementary Fig. 3). Surprisingly, all-*trans* configuration of the retinal in the O-state is in contrast to the recently published time-resolved Fourier-transform infrared spectroscopy (FTIR) data, where authors suggested 13-*cis* configuration in the O-state of NaRs<sup>21</sup>, but is in line with the data on another light-driven Na<sup>+</sup> pump from *Gillisia limnaea* (GLR), published in 2014, where the authors report a distorted all-*trans* configuration of retinal in the O-state<sup>22</sup>. In ref. <sup>21</sup> authors found a broad peak at 940 cm<sup>-1</sup> in the FTIR spectrum of the O-state of KR2 containing 12, 14-D<sub>2</sub> retinal, which was interpreted as 13-*cis* configuration of the

retinal. Our data demonstrate that although retinal is all-*trans* in both the ground and the O-state, it is kinked notably around C<sub>14</sub> atom only in the intermediate, but not in the ground state (Supplementary Figs. 2 and 3). This distortion may result in the appearing of the peak at 940 cm<sup>-1</sup> described in ref. <sup>21</sup>. We also cannot exclude that the inconsistency of the results on the retinal configuration in the O-state may originate from the different conditions and protein environment during the experiments.

The all-*trans* retinal configuration in the O-state of KR2 obtained in the present work means that relative location of the RSBH<sup>+</sup> and D116 side chain is similar to that in the ground state (PDB ID: 6REW<sup>4</sup>). RSBH<sup>+</sup> is hydrogen bonded to D116 and the distance between them is 2.9 Å in the O-state (Fig. 2f). The

existence of this hydrogen bond is supported by time-resolved resonance Raman spectroscopy<sup>11</sup>. Indeed, in ref. 11 authors reported that C=N stretching frequencies of the RSB are very similar between the ground state (1640 cm<sup>-1</sup>) and the O-state (1642 cm<sup>-1</sup>). The C=N stretching frequency is a sensitive marker for the hydrogen bond strength of the protonated RSB. The similar frequencies of the C=N stretching mode support that relative locations of the RSBH<sup>+</sup> and D116 side chain are similar to those in the ground state.

**Sodium-binding site inside the protein.** The crystal structure of the O-state of KR2 clearly reveals the Na<sup>+</sup>-binding site near the RSB, comprised of S70, N112, and D116 side chains and main chain oxygen of V67 (Fig. 2f and Supplementary Figs. 2 and 4). Previous mutational analysis confirms the importance of these residues for KR2 pumping activity. Indeed, D116 is crucial for KR2 functioning<sup>2</sup>, and N112 determines ion selectivity<sup>7</sup>. Substitution of S70 with threonine or alanine dramatically decreases Na<sup>+</sup>-pumping activity of KR2<sup>5,9</sup>. The mean distance between Na<sup>+</sup> and the coordinating oxygen atoms is 2.3 Å (Fig. 2f and Supplementary Fig. 2).

While in the ground state KR2 is in the ‘expanded’ conformation, in the O-state N112 is flipped towards S70 and D116, therefore the overall configuration is similar to that of the ‘compact’ conformation of KR2<sup>4,12</sup> (Fig. 2, Supplementary Figs. 5 and 6). This is also evidenced by disappearance of the big polar cavity near the RSB (SBC1) and enlargement and elongation of the cavity near R109–D251 pair (SBC2) in the O-state (Fig. 2 and Supplementary Fig. 6). Four water molecules, filling the SBC1 in the ‘expanded’ ground state (Fig. 2c), are displaced as follows: two of them are found in the small cavity formed in the intermediate near S70 at the pentamerization interface, one remains at the same place and is coordinated by N112 and D251, and the last one is moved to the SBC2 near L75 and R109 at the inner extracellular part of the protein (Fig. 2f). Upon sodium binding and formation of the ‘compact’ state L74 side chain also flips simultaneously with the N112 in order to avoid the steric conflict of these two residues. Our mutational analysis indicated that L74A substitution dramatically decreases pumping activity of the protein (Supplementary Fig. 7). Hence, this additionally supports the importance of the ‘compact’ conformation for Na<sup>+</sup> pumping by KR2.

Interestingly, the location of the sodium-binding site in the O-state of KR2 is similar to that of the chloride ion-binding sites in the ground state of light-driven chloride pumps (Supplementary Fig. 8). Namely, in a chloride-pumping rhodopsin from *Nonlabens marina* S1-08 (CIR)<sup>23,24</sup> the anion is coordinated by the N98 and T102 of the NTQ motif, which are analogous to the N112 and D116 of the NDQ motif of light-driven sodium pumps (Supplementary Fig. 8).

**Conformational switches guide Na<sup>+</sup> uptake and release.** The similarity of protein conformation in the O-state to the ‘compact’ is intriguing, however, could easily be explained. Indeed, relative location of the RSBH<sup>+</sup> and Na<sup>+</sup>–D116<sup>-</sup> pair makes the distribution of the charges in the central part of the protein nearly identical to that of the KR2 with protonated D116 at acidic pH<sup>2–4</sup>. The structures of KR2 in the ‘compact’ conformation are also observed only at low pH<sup>3–5</sup>. It was thus suggested that the ‘compact’ conformation may appear in response to the D116 neutralization<sup>4</sup>. To understand better the nature of conformational switches in KR2 and the influence of D116 protonation on the protein conformation, we produced and crystallized KR2-D116N at pH 8.0, which mimics the WT protein with fully

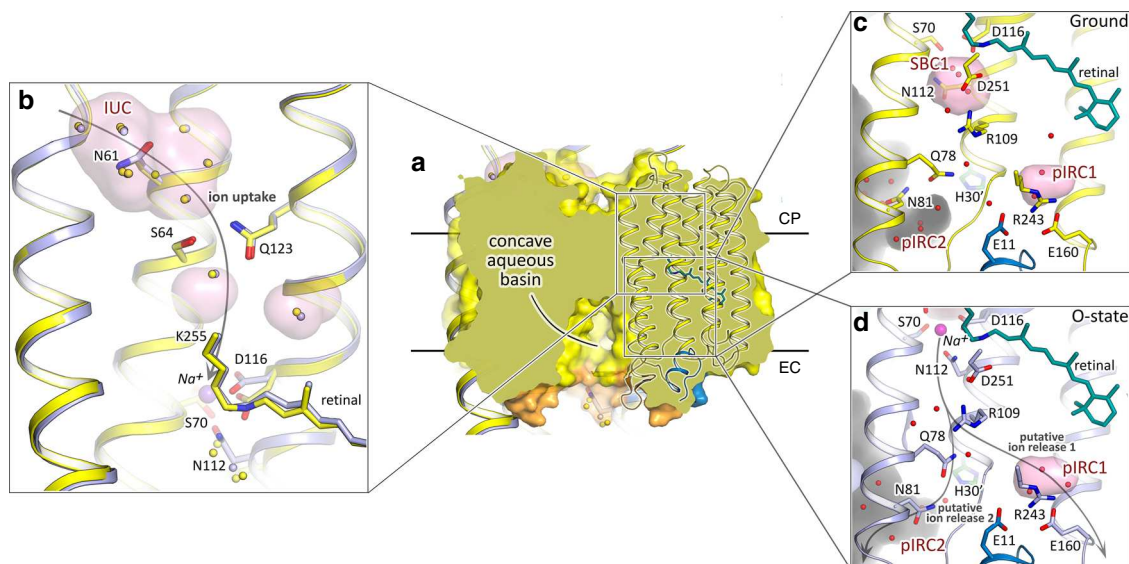
protonated D116 and solved its structure in the pentameric form at 2.35 Å.

Confirming our hypothesis, the structure shows that introduction of asparagine at the position of D116 led to the flip of the side chains of N112 and L74 in comparison to the ground state of the wild type (WT) protein and disappearance of the SBC1, characteristic for the ‘expanded’ conformation (Supplementary Fig. 6). Overall, the structure of D116N is very similar to that of the O-state (RMSD 0.2 Å) and also to the ‘compact’ conformation (RMSD 0.3 Å) of the KR2-WT (Supplementary Fig. 6). However, Na<sup>+</sup> is absent inside the protomers and the relative orientation of the RSBH<sup>+</sup> and N116 is altered (Supplementary Fig. 6). Particularly, the RSBH<sup>+</sup> forms two alternative conformations, and hydrogen bond between the RSBH<sup>+</sup> and N116 is absent (see Supplementary material). We also observed that D116 protonation destabilizes the pentameric assembly of KR2 (see Supplementary material). Thus, we suggest that neutralization of D116 is the key determinant of the formation of the ‘compact’ conformation, which explains their structural similarity.

The structures of KR2 O-state and D116N mutant, together with previously described pH dependence of the KR2 organization<sup>4</sup>, allow us to conclude that the ‘compact’ conformation and, particularly, N112 flip towards D116, stabilizes neutralized RSB counterion and correspondingly neutral transiently formed Na<sup>+</sup>–D116<sup>-</sup> pair during the photocycle.

It was suggested previously that the SBC1 in the ‘expanded’ conformation surrounded by R109, N112, W113, D116, and D251 might be a transient Na<sup>+</sup>-binding site in an intermediate state of the protein photocycle<sup>4,12</sup>. However, the present work shows that Na<sup>+</sup> binds far from R109 and D251 (Fig. 2f). Since the release of Na<sup>+</sup> occurs upon the O-to-ground state transition, which structurally corresponds to ‘compact’-to-‘expanded’ switch, we suggest that the ‘expanded’ conformation is also important for the ion release to the extracellular space. Importantly, Na<sup>+</sup> uptake and release are guided by the switches from the ‘expanded’ to the ‘compact’ and then again back to the ‘expanded’ conformations, respectively.

**Sodium translocation pathway.** Although the structure of the KR2 protomer near the RSB is altered in the O-state, the organization of both putative ion uptake and ion release regions remains the same to those in the ground state (Fig. 3). It is not surprising when we consider the cytoplasmic part of the protein. In the ground state Na<sup>+</sup> does not penetrate to the inside of KR2. At the same time, in the O-state Na<sup>+</sup> should not have a way to return back to the cytoplasm. Therefore, the pathway, formed upon transition from the ground to the O-state, connecting the ion uptake cavity (IUC) with the RSB region should be blocked in both ground and the O-states. Consequently, the restoration in the O-state of the initial conformation of S64–Q123 pair, separating the IUC from the RSB environment, is expected. On the other hand, the same organization of the E11–E160–R243 cluster and putative ion release cavity (pIRC1) in the ground and O-states may seem quite surprising. Previously, these residues were suggested to line the pathway of Na<sup>+</sup> release to the extracellular bulk during the direct O-to-ground transition<sup>3–5</sup>. The absence of the disturbance in this region means either that the energy stored in the distorted all-*trans* retinal in the O-state is enough to relocate Na<sup>+</sup> directly from the core of the protein to the bulk without any transient-binding sites or/and that there might be another ion release pathway in KR2. Indeed, the second hypothesis is supported by the mutational analysis, which showed that substitution of E11, E160, or R243 to alanines or polar non-charged residues does not abolish Na<sup>+</sup>-pumping activity, however, affects the stability of the proteins<sup>2,3,5</sup>. The other



**Fig. 3 Ion uptake and release pathways of KR2.** **a** Section view of KR2 pentamer in the membrane. Concave aqueous basin facing the extracellular space is indicated by the black line. Only one protomer is shown in cartoon representation. Membrane core boundaries were calculated using PPM server<sup>56</sup> and are shown with black lines. **b** Structural alignment of the cytoplasmic parts of the ground (yellow) and O- (blue) states of KR2. Water molecules are shown with yellow and blue spheres for the ground and O-state, respectively. **c** Detailed view of the extracellular side of KR2 in the ground state. **d** Detailed view of the extracellular side of KR2 in the O-state. Cavities inside the protein are calculated using HOLLOW<sup>57</sup> shown in pink and marked with red labels. Protein surface concavity from the aqueous basin at the extracellular side is colored gray. Retinal cofactor is colored teal. Water molecules are shown with red spheres. Sodium ion is shown with purple sphere. N-terminal  $\alpha$ -helix is colored blue. BC loop is colored orange. H30' of adjacent protomer is colored with dark-green. Helices A, F, and G are hidden for clarity. Gray arrows identify putative ion uptake and two ion release pathways.

(alternative) putative way for  $\text{Na}^+$  release goes from the inner extracellular part of the protein through the elongated in the O-state SBC2 to the bulk near the  $\text{Na}^+$  bound at the surface of KR2 in both the ground and O-states. These channel-like pathway is constricted with the only side chain of Q78 residue (Figs. 2, 3d). The pathway propagates from the inner region between Q78, N106, and R109 to the relatively large cavity (pIRC2) between helices B and C, BC loop and helix A' of adjacent protomer at the extracellular side (Figs. 2, 3 and Supplementary Fig. 9). The cavity proceeds further to a concave aqueous basin facing the extracellular solution, formed in the central pore of the KR2 pentamer at the extracellular side and is surrounded by Q78, N81, S85, D102, Y108, and Q26' residues and filled with water molecules in both the ground and O-states (Fig. 3). Notable displacements of these residues and waters occur in the O-state, such as the flip of N81 towards H30' of the adjacent protomer (Supplementary Fig. 10). Consequently, in the O-state additional water molecule appears in the pIRC2, which is coordinated by hydrogen bonds with E26', H30', and N81 (Supplementary Fig. 10). The positions of Q78 and Y108 are also altered in the O-state (Supplementary Fig. 10).

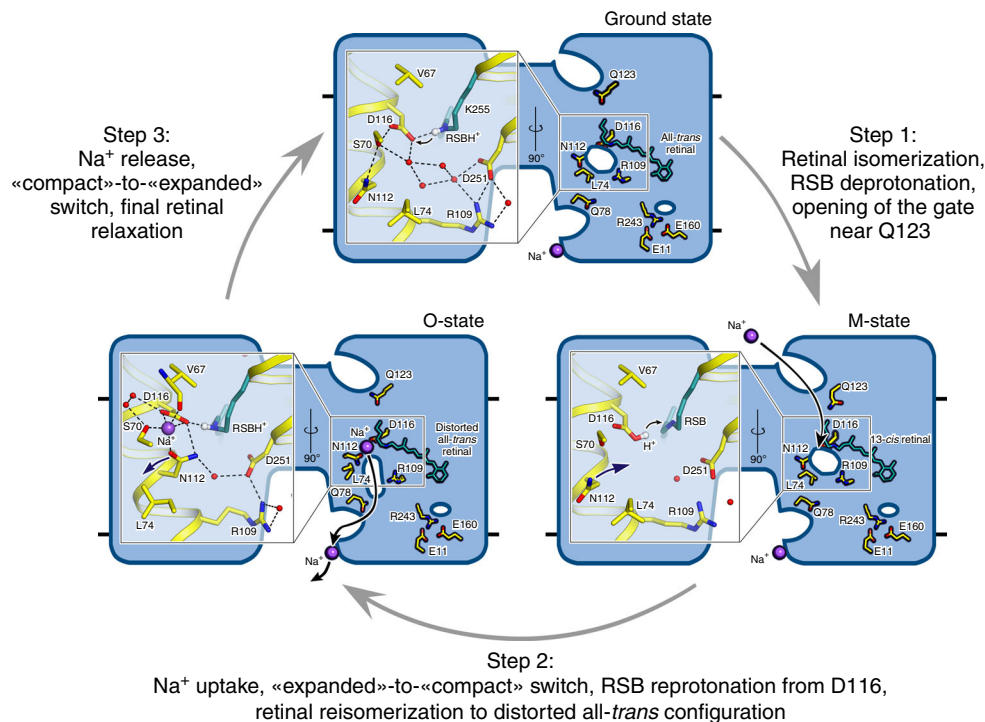
To probe the possible ion release pathways, we conducted 10 short metadynamics simulations starting from the sodium-bound conformation (Supplementary Fig. 11 and Supplementary Movie 1). The simulations revealed that the R109 sidechain forms a barrier for sodium exiting SCB1, and changes its position to allow sodium passage. Upon passing R109,  $\text{Na}^+$  either exits the protein via pIRC2 (8 simulations out of 10), or proceeds towards pIRC1 (two simulations). In the first scenario, the ion is quickly released towards the aqueous basin in the middle of KR2 pentamer in the vicinity of another ion found at the interface between the protomers, and is sometimes observed to replace it (Supplementary Movie 1). In the second scenario, the ion samples different locations around the E11–E160–R243 triad and is later released via N106 and Q157 on the outer side of the pentamer.

Importantly, the organization of the pIRC2 region is identical in the O-state and KR2-D116N (Supplementary Fig. 12A). It means that the rearrangements on the surface of the KR2 occur not directly in response to the retinal isomerization upon photon absorption, but rather due to the redistribution of charges in the central inner part of the protein protomer. Such long-distance interactions between the RSB-counterion and pentamer surface were already studied for the WT and H30A variant of KR2<sup>14,25</sup>. Recently, it was also shown that there is an allosteric communication between the interprotomer  $\text{Na}^+$ -binding site and the RSB hydrogen bond already in the ground state<sup>26</sup>. Thus, structural rearrangements of the RSB-counterion pair upon  $\text{Na}^+$  release and corresponding 'compact'-to-'expanded' conformational switch may allosterically affect the interprotomer  $\text{Na}^+$ -binding site, promoting  $\text{Na}^+$  unbinding from the site, observed in metadynamics simulations.

To gather more details about long-distance interactions between the protein core and surface, we solved the structure of pentameric form of KR2-H30A at pH 8.0 at 2.2 Å.

Overall, the structure of this mutant is nearly the same as that of the ground state of WT protein (RMSD 0.15 Å) (Supplementary Figs. 5, 13). The organization of their inner cytoplasmic, central, and extracellular parts is identical. Surprisingly, in contradiction to the earlier FTIR experiments,  $\text{Na}^+$  bound at the oligomerization interface of the WT protein is also present in H30A<sup>2</sup>. However, the region of H30 is altered (Supplementary Fig. 12). Particularly, H30 side chain is replaced by two additional water molecules ( $w_B'$  and  $w_B''$ ) (Supplementary Fig. 12). Moreover, the H30A mutation leads to appearance of second alternative conformation of Y108, not identified in other structures of KR2 or its variants (Supplementary Fig. 12). It was shown that H30A is more selective to sodium and almost does not pump protons<sup>2</sup>. As the only existing differences in the structures of the WT and the mutant occur near Q78, N81, Y108, H30' and pIRC2, we suggest that this region is important for





**Fig. 4 Proposed Na<sup>+</sup> pumping mechanism.** Schematic side section view of the KR2 pentamer is shown. Membrane core boundaries are shown with black lines. Cavities are demonstrated as white ellipses. Enlarged view of the RSB region is shown at the left part of the pentamer. The 13-*cis* configuration of the retinal cofactor is modeled manually for schematic representation. Na<sup>+</sup> is shown with violet spheres. Black arrows indicate proposed Na<sup>+</sup> uptake and release pathways. Violet arrows indicate rearrangement of N112 side chain during “expanded”-to-“compact” and back “compact”-to-“expanded” switches. Small gray arrows indicate the translocation of the hydrogen from the Schiff base to D116 during the formation of the M-state and following reprotonation of the Schiff base from the D116 in the M-to-O transition. Retinal cofactor is colored teal. Waters are shown with red spheres. Hydrogen bonds in the RSB region are shown with black dashed lines.

cation selectivity, which additionally supports the hypothesis that this region is a part of ion translocation pathway.

To verify the suggested ion-release pathway, we performed functional studies of KR2 mutants in *E. coli* cells suspension, similar to previous works<sup>3,27</sup> (Supplementary Fig. 7). The results showed that Q78 is the key residue, which is likely to act as a gate for sodium, flipping upon sodium passage (Q78L mutant remains almost fully functional). The blocking of Q78 motion (Q78Y,W mutations) resulted in dramatic decrease of the pumping activity. In Q78A mutant, similar to Q123A, the sodium pumping activity is retained, however decreased notably in comparison to the WT protein. Another interesting finding was that Y108A mutant almost fully lost its pumping ability.

Hence, we suggest that Na<sup>+</sup> translocation pathway propagates from IUC to pIRC2 via a chain of polar inner cavities, which are modified during photocycle. IUC and pIRC2 are separated from the inside of KR2 by two weak gates near Q123 and Q78, respectively. This makes the KR2 ion pathway similar to that of the channelrhodopsin 2<sup>28</sup> (CrChR2) (Supplementary Fig. 14). However, unlike in CrChR2, KR2 has the Na<sup>+</sup>-binding site in the central region near the RSB.

**Mechanism of sodium pumping.** Crystal structures of the O-state and D116N and H30A mutants of KR2 together with available literature data allow us to suggest a mechanism of protein functioning (Fig. 4).

Step 1: In the ground state KR2 is in the ‘expanded’ conformation with the SBC1 filled with four water molecules. The RSBH<sup>+</sup> is hydrogen bonded to its counterion D116<sup>-</sup>. The Q123-S64 gate separates the IUC and the RSB. With the

absorption of the light photon the retinal isomerizes from all-*trans* to 13-*cis* configuration and the red-shifted K-state appears in nanoseconds, followed by the formation of the L/M intermediate in about 30 μs. The proton is translocated from the RSBH<sup>+</sup> to the D116<sup>-</sup> with the formation of the M-state and the hydrogen bond between them is absent in this intermediate. The Q123-S64 gate also opens in this step.

Step 2: With the rise of the O-state Na<sup>+</sup> passes the gate and deprotonated RSB, and binds between S70, N112, and D116 belonging to helices B and C. Na<sup>+</sup> uptake results in the proton translocation from D116 back to the RSB with the restoration of the RSB-counterion hydrogen bond, thus preventing the Na<sup>+</sup> backflow to the cytoplasmic side. It also causes the flip of N112 side chain for the stabilization of Na<sup>+</sup>-D116<sup>-</sup> pair, therefore, the ‘compact’ state appear at this step. Retinal is still in the distorted all-*trans* conformation in the O-state.

Step 3: With the decay of the O-state retinal returns completely to its ground configuration and the ‘expanded’ conformation of KR2 occurs with the N112 flip back to the pentamerization interface, thus opening the way for Na<sup>+</sup> release. In the O-state the release pathway is prepared as it is evidenced from the O-state structure. We suggest that the release preferentially involves Q78 and proceeds towards the cavity, formed by N81, Y108, and H30’ of adjacent protomer (pIRC2). We also suggest that four water molecules, filling the SBC1 in the dark state are involved in the Na<sup>+</sup> hydration during its transitions inside the protein. The Na<sup>+</sup> release might proceed using the relay mechanism, when the ion, released from the KR2 protomer, replaces the ion, bound at the protein oligomerization interface. The proposed relay mechanism allows lowering the energy barriers for facilitation of the Na<sup>+</sup>

release. The E11–E160–R243 cluster and the cavity near it (pIRC1), suggested earlier to be involved in Na<sup>+</sup> release, thus may be involved mainly in protein stabilization, rather than in ion translocation pathway.

Last but not least, in the absence of Na<sup>+</sup> KR2 acts as a proton pump with the significantly altered photocycle and the absence of the pronounced O-state. The long living L/M/O-like state decays slowly in the proton-pumping mode<sup>2</sup>. We note that this is in agreement with the suggested mechanism of Na<sup>+</sup> pumping. We suggest that the formation of the K-, L- and M-states is the same for the Na<sup>+</sup>- and H<sup>+</sup>-pumping modes, however, in the absence of Na<sup>+</sup> the ion does not flow diffusely to the central region when the RSB is neutral, and RSB is presumably reprotonated from the cytoplasmic side through the IUC and no rearrangements occur in the region of D116. Then the slow relaxation of the retinal to all-*trans* configuration triggers the proton release from the D116 to the extracellular side and the protein returns to the ground state.

## Discussion

The presented here structure of the key functional O intermediate state of the KR2 rhodopsin allowed us to propose the pathway and molecular mechanism of the active light-driven Na<sup>+</sup> transport. We suggest that the pathway of Na<sup>+</sup> connects two aqueous concavities in the cytoplasmic and extracellular parts of the KR2, respectively, via several polar cavities inside the protein, which are separated from each other by weak gates. From that side, the organization is very similar to that of CrChR2<sup>28</sup>. The principal difference is that in opposite to CrChR2, KR2 has the tight transient Na<sup>+</sup>-binding site near the Schiff base, separating the cytoplasmic and extracellular parts of the protein. Judging from the arrangement of the cavities in the O-state, Na<sup>+</sup> transport in KR2 appears rather like in a perfectly outwardly directed Na<sup>+</sup> channel than in a pump, which we interpret as a chimera between the two fundamental mechanisms. Similar concept was suggested earlier<sup>10</sup>, however, it mostly concerned the diffusion Na<sup>+</sup> uptake from the cytoplasmic part. The presence of the positively charged R109 side chain at the extracellular part of the protein presumably prevents Na<sup>+</sup> uptake from that side in the M-state, as also evidenced by a recent finding that R109Q mutation converts KR2 into a potassium channel<sup>9</sup>. Presented here model of the O-state together with metadynamics simulations and mutational analysis, as well as the absence of any intermediate states during O-to-ground transition in KR2, strongly suggest that Na<sup>+</sup> release proceeds also on the path of least resistance without any notable structural rearrangements of the protein. This is supported by the fact that the retinal in the O-state is almost completely relaxed to its initial configuration. We also speculate that for the facilitation of the Na<sup>+</sup> release (lowering of the energy barrier for the Na<sup>+</sup> release against electrochemical gradient), the protein use the relay mechanism, in which the released ion replaces the ion bound at the oligomerization interface. The fact was also evidenced by the metadynamics simulations, described in the present manuscript and by the recently reported existence of the allosteric communications between the RSB and interprotomer ion-binding site<sup>26</sup>. Current understanding of the Na<sup>+</sup> release from the KR2 is summarized in the Supplementary Movie 1.

## Methods

**Protein expression and purification.** *E. coli* cells of strain SE1 (Staby™Codon T7, Eurogentec, Belgium) were transformed with the KR2 expression plasmid. Transformed cells were grown at 37 °C in shaking baffled flasks in an auto-inducing medium ZYP-5052<sup>29</sup> containing 100 mg/L ampicillin. When glucose level in the growing bacterial culture dropped below 10 mg/L, 10 μM all-*trans*-retinal (Sigma-Aldrich, Germany) was added, the incubation temperature was reduced to 20 °C and incubation continued overnight. Collected cells were disrupted in M-110P Lab Homogenizer (Microfluidics, USA) at 25,000 psi in a buffer containing

20 mM Tris–HCl pH 8.0, 5% glycerol, 0.5% Triton X-100 (Sigma-Aldrich, USA) and 50 mg/L DNase I (Sigma-Aldrich, USA). The membrane fraction of cell lysate was isolated by ultracentrifugation at 90,000 × g for 1 h at 4 °C. The pellet was resuspended in a buffer containing 50 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> pH 8.0, 0.1 M NaCl and 1% DDM (Anatrace, Affymetrix, USA) and stirred overnight for solubilization. Insoluble fraction was removed by ultracentrifugation at 90,000 × g for 1 h at 4 °C. The supernatant was loaded on Ni-NTA column (Qiagen, Germany) and KR2 was eluted 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, 0.5 M imidazole and 0.1% DDM. The eluate was subjected to size-exclusion chromatography on 125 ml Superdex 200 PG column (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, 0.05% DDM. Protein-containing fractions with the minimal A<sub>280</sub>/A<sub>525</sub> absorbance ratio were pooled and concentrated to 60 mg/ml for crystallization.

**Measurements of pumping activity in *E. coli* cells.** *E. coli* cells of strain C41 (DE3) (Lucigen, USA) were transformed with the KR2 expression plasmid. Transformed cells were grown at 37 °C in shaking baffled flasks in an autoinducing medium, ZYP-5052<sup>29</sup> containing 100 mg/L ampicillin, and were induced at optical density OD<sub>600</sub> of 0.7–0.9 with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 10 μM all-*trans*-retinal. 3 h after induction, the cells were collected by centrifugation at 4000 × g for 10 min and were washed three times with unbuffered salt solution (100 mM NaCl, and 10 mM MgCl<sub>2</sub>) with 30-min intervals between the washes to allow exchange of the ions inside the cells with the bulk. After that, the cells were resuspended in 100 mM NaCl solution and adjusted to an OD<sub>600</sub> of 8.0. The measurements were performed on 3 ml of stirred cell suspension kept at 1 °C. The cells were illuminated for 5 min with a halogen lamp (Intralux 5000-1, VOLPI) and the light-induced pH changes were monitored with a pH meter (LAB 850, Schott Instruments). Measurements were performed with the addition of 30 μM of protonophore carbonyl cyanide 3-chlorophenylhydrazone (CCCP).

**Crystallization details and crystals preparation.** The crystals were grown using the in meso approach<sup>30,31</sup>, similarly to our previous work<sup>27,32,33</sup>. 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 25 mg/ml. The crystals of monomeric (for D116N mutant) and pentameric (for WT protein and H30A mutant) forms were grown using the precipitate 1.0 M sodium malonate pH 4.6 and 1.2 M sodium malonate pH 8.0, respectively (Hampton Research, USA). Crystallization probes were set up using the NT8 robotic system (Formulatrix, USA). The crystals were grown at 22 °C and appeared in 2–4 weeks. Before harvesting, crystallization drop was opened and covered with 3.4 M sodium malonate solution, pH 8.0, to avoid dehydration. All crystals were harvested using micromounts (MiTeGen, USA) and were flash-cooled and stored in liquid nitrogen for further crystallographic analysis.

**Time-resolved visible absorption spectroscopy.** The laser flash photolysis was performed similar to that described by Chizhov et al.<sup>34,35</sup> with minor differences. The excitation system consisted of Nd:YAG laser Q-smart 450mJ with OPO Rainbow 420–680 nm range (Quantel, France). For the experiments the wavelength of the laser was set 525 nm. Microcrystals of KR2 in the lipidic cubic phase were plastered on the 4 × 7 mm cover glass. The thickness of the slurries was adjusted in order to give sufficient signal. The glass with crystal slurries was placed into 5 × 5 mm quartz cuvette (Starna Scientific, China) filled with the buffer solution containing 3.4 M sodium malonate pH 8.0 and thermostabilized via sample holder qpod2e (Quantum Northwest, USA) and Huber Ministat 125 (Huber Kältemaschinenbau AG, Germany). The detection system beam emitted by 150 W Xenon lamp (Hamamatsu, Japan) housed in LSH102 universal housing (LOT Quantum Design, Germany) passed through pair of Czerny–Turner monochromators MSH150 (LOT Quantum Design). The received monochromatic light was detected with PMT R12829 (Hamamatsu). The data recording subsystem represented by a pair of DSOX4022A oscilloscopes (Keysight, USA). The signal offset signal was measured by one of oscilloscopes and the PMT voltage adjusted by Agilent U2351A DAQ (Keysight). The time-resolved visible absorption spectroscopy data was processed and fitted using OriginPro8.5 software.

**Accumulation of the intermediate state in KR2 crystals.** Absorption spectra of KR2 in solution were collected using the UV-2401PC spectrometer (Shimadzu, Japan) using the UVProbe 2.33 software. The spectroscopic characterization of O-state build-up in KR2 crystals was performed at the iCOS Lab located at the ESRF<sup>36</sup>. The same set up was established at the P14 beamline of the PETRAIII synchrotron source (Hamburg, Germany) for accumulation of the O-state in crystals for X-ray diffraction data collection. Also the same accumulation procedure was applied to crystals at iCOS and P14 beamline. Briefly, UV–visible absorption spectra were measured using as a reference light that of a DH-200-BAL deuterium-halogen lamp (Ocean Optics, Dunedin, FL) connected to the incoming objective via a 200 μm diameter fiber, resulting in a 50 μm focal spot on the sample, and a QE65 Pro spectrometer (Ocean Optics, Dunedin, FL) connected to the outgoing objective via a 400 μm diameter fiber. The actinic light comes from a 532 nm laser (CNI Laser, Changchun, P.R. China) coupled to a 1000 μm diameter fiber which is connected to the third objective whose optical axis is perpendicular to those of the ingoing and

outgoing objectives. Ground states spectra (100 ms acquisition time averaged 20×) were collected on crystals flash-cooled in liquid nitrogen and kept under a cold nitrogen stream at 100 K. In order to maximize the population of the O-state, a crystal was put under constant laser illumination at 100 K, the nitrogen stream was then blocked for 2 s, then the laser was switched off once the crystal is back at 100 K. For the accumulation of the O-state laser power density of 7.5 mW/cm<sup>2</sup> at the position of the sample was used. The mean size of the crystals was 200 × 100 × 30 μm<sup>3</sup> (Supplementary Fig. 22). The plate-like crystals were oriented so that the largest plane (200 × 100 μm<sup>2</sup>) was as perpendicular to the laser beam. The laser beam was focused to the size of 500 × 500 μm<sup>2</sup> (1/e<sup>2</sup>). A UV–visible absorption spectrum was then recorded to show the red-shifted absorption maximum characteristic of the O-state of KR2. The crystals with accumulated intermediate state were then stored in liquid nitrogen and transported to the PETRAIII, Hamburg, Germany for the X-ray experiments and showed the same structure as that obtained using crystals with the O-state, accumulated directly at the P14 beamline of PETRAIII.

**Acquisition and treatment of diffraction data.** X-ray diffraction data of D116N and H30A mutants were collected at the beamlines ID23-1 and ID29 of the ESRF, Grenoble, France, using a PILATUS 6M and EIGER 16M detectors, respectively. The data collection was performed using MxCube2 software. X-ray diffraction data of the KR2 O-state at 100 and 293 K (room temperature, RT) was collected at the P14 beamline of the PETRAIII, Hamburg, Germany, using EIGER 16M detector. For the collection of the X-ray diffraction data at RT the crystals in the cryoloop were placed on the goniometer of P14 beamline and maintained in the stream of the humid air (85% humidity). The stream of humid air was provided by the HC humidity controller (ARINAX, France). For activation of the proteins in crystals and obtaining the structure of the O-state at RT the laser flash was synchronized with the X-ray detector. The laser was illuminating the crystal only during X-ray data collection to avoid drying and bleaching. The crystals were rotated during the data collection and laser illumination. Diffraction images were processed using XDS<sup>37</sup>. The reflection intensities of the monomeric form of D116N mutant were scaled using the AIMLESS software from the CCP4 program suite<sup>38</sup>. The reflection intensities of all the pentameric forms were scaled using the Staranisio server<sup>39</sup>. There is no possibility of twinning for the crystals. For the both structures of KR2 at RT diffraction data from three crystals was used (Supplementary Table 1). In all other cases, diffraction data from one crystal was used. The data statistics are presented in Table S1.

**Serial crystallography data collection and processing.** Serial millisecond crystallography data of the O-state of KR2 was obtained at RT at BL13-XALOC beamline of ALBA (Barcelona, Catalunya) using a PILATUS 6M detector working at 12.5 Hz and a 60 × 40 μm FWHM (HxV) sized beam at 12.6 keV (1.4 × 10<sup>12</sup> ph/s). For that purpose, an LCP stream of protein microcrystals was injected into the focus region using a LCP injector<sup>40</sup>, placed at 45° of the diffractometer table, with the help of an ÅKTA pump flowing at 1 μl/min and a constant helium supply (10–14 psi) yielding an extrusion speed of 30 nl/min for 100 μm capillary. A Roithner laser source (RLTMLL-532-100-5) working at 20 mW was used for protein activation.

A total number of 1,208,640 detector images were collected and processed with CrystFEL (version 0.8.0)<sup>41</sup> without any additional modification. Among all images collected, 350,862 were identified as potential crystal hits with more than 30 Bragg peaks with min-snr = 3.5, threshold = 12, highres = 2.5 parameters using peakfinder8 algorithm as implemented in CrystFEL, corresponding to an average hit rate of 29%. The overall time of data collection from a sample with a total volume of 120 μl was about 36 h.

Data were processed using CrystFEL. For peak finding, we used peakfinder8 and min-snr = 3.5, threshold = 12. For indexing, indexers dirax<sup>42</sup>, xds<sup>37</sup>, taketwo<sup>43</sup> (in that order) were used, with --multi option enabled. Integration was performed using int-radius = 3,4,5. Data were merged using process\_hkl with min-res = 3.3, push-res = 3.0, and symmetry = mmm. This yielded a dataset with 131,872 indexed images with 136,656 crystals, corresponding to 39% average indexing rate. Among these images, 38,761 were merged together after process\_hkl rejection. Among indexers, dirax was the most successful one, providing 77,681 indexed crystals. Initial geometry, provided by the beamline staff, was optimized with detector-shift and geoptimiser, as described in ref. 44.

**Structure determination and refinement.** Initial phases for the pentameric structures were successfully obtained in the C222<sub>1</sub> space group by molecular replacement (MR) using MOLREP<sup>45</sup> using the 6REW structure as a search model. Initial phases for monomeric KR2-D116N were successfully obtained in the I222 space group by MR using the 4XTL structure as a search model. The initial MR models were iteratively refined using REFMAC5<sup>46</sup>, PHENIX<sup>47</sup>, and Coot<sup>48</sup>.

**Molecular dynamics simulations.** The simulation system consisted of a KR2 pentamer in the O-state with sodium ions in SBC2 and cocrystallized water molecules. The proteins was embedded in a POPC bilayer (256 lipids) and then solvated with TIP3P water with a Na<sup>+</sup>/Cl<sup>-</sup> concentration of 150 mM using the CHARMM-GUI web-service<sup>49</sup>. The simulation box contained 94,817 atoms in

total. All ionizable amino acids were modeled in their standard ionization state at pH 8, including D116 and D251 which were modeled charged.

The CHARMM-GUI recommended protocols were followed for the initial energy minimization and equilibration of the system. The atoms of protein and lipids in the system were subjected to a harmonic positional restraint and 5000 steps of steepest descent minimization followed by two 25 ps equilibration steps in the NVT ensemble using Berendsen thermostat and one 25 ps and three 50 ps equilibration steps in the NPT ensemble using Berendsen thermostat and barostat. During all equilibration steps, the force constants of the harmonic positional restraints were gradually reduced to zero. The system was further equilibrated for 10 ns in the NPT ensemble with Nose–Hoover thermostat and Parrinello–Rahman barostat, which were also used for the further production simulations. The temperature and pressure were set to 303.3 K and 1 bar with temperature and pressure coupling time constants  $\tau_T = 1.0 \text{ ps}^{-1}$  and  $\tau_P = 0.5 \text{ ps}^{-1}$ , respectively. All MD simulations were performed with GROMACS version 2018.1<sup>50</sup>. The time step of 2 fs was used for all the simulations except for the early steps of equilibration. The CHARMM36 force field<sup>51</sup> was used for the protein, lipids, and ions. Parameters for retinal bound to lysine were adapted from ref. 52.

In order to investigate the putative sodium translocation pathways, metadynamics approach (metaMD) was employed<sup>53</sup>. This method is based on biasing of the potential surface via addition of repulsive functions (“hills”, typically Gaussians) which force the investigated molecular system to explore its configurational space broader and faster than in a regular unbiased MD simulation. Since the correct sampling of protein degrees of freedom relevant to the ion release at the late states of the photocycle appears problematic, the performed metadynamics simulations allowed us to reveal possible sodium unbinding pathways rather than quantitatively assess the free energies changes associated with them. We used the PLUMED plugin for GROMACS to perform metaMD simulations<sup>54</sup>. The projections of the vector connecting the sodium ion and its original position onto  $x$ ,  $y$ , and  $z$  directions were used as the collective variables (i.e., 3 CVs were used). In order to prevent sodium passage back to the cytoplasmic side we applied a flat-bottom potential ( $k = 1000 \text{ kJ/mol/nm}^2$ ) in the normal to the membrane direction, which discouraged ion moving towards the cytoplasmic side. Also, harmonic restraints were applied to all protein C $\alpha$  atoms above the C $\alpha$  atom of K255 (in the direction of CP) to prevent the overall motion of the protein complex. The deposition rate for hills was 0.5 ps; the width and height of deposited hills were equal to 0.05 nm and 1 kJ/mol, respectively. The simulations were continued until the exit of ion from the protein interior was observed (typically, during 5–20 ns). We have carried out 10 metaMD runs in total, 2 replicates for each of the 5 protomers of the KR2 pentamer.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Data supporting the findings of this manuscript are available from the corresponding author upon reasonable request. A reporting summary for this Article is available as a Supplementary Information file. The source data underlying Fig. 1b, c, e and Supplementary Figs. 7 and 21 are provided as a Source Data file. Crystallographic data that support the findings of this study have been deposited in the Protein Data Bank (PDB) with the accession codes: PDB: 6XYT (the O-state of KR2 at 100K), PDB: 6YBY (D116N mutant in monomeric form), PDB: 6YBZ (D116N mutant in pentameric form), PDB: 6YC0 (steady-state-SMX activated state of KR2 at RT), PDB: 6YC1 (H30A mutant in pentameric form), PDB: 6YC2 (dark-state of KR2 at RT), PDB: 6YC3 (dark-state of KR2 at 100K) PDB: 6YC4 (steady-state activated state of KR2 at RT). Serial crystallography data corresponding to the 6YC0 model have been deposited to the Coherent X-ray Imaging Data Bank<sup>55</sup> with CXIDB ID 141. These include detector.cbf files without any additional treatment for images with more than 30 diffraction peaks, as described in “Methods” section.

Received: 31 January 2020; Accepted: 3 April 2020;  
Published online: 01 May 2020

## References

- Gushchin, I. & Gordeliy, V. Microbial rhodopsins. 19–56, [https://doi.org/10.1007/978-981-10-7757-9\\_2](https://doi.org/10.1007/978-981-10-7757-9_2) (2018).
- Inoue, K. et al. A light-driven sodium ion pump in marine bacteria. *Nat. Commun.* **4**, 1678 (2013).
- Gushchin, I. et al. Crystal structure of a light-driven sodium pump. *Nat. Struct. Mol. Biol.* **22**, 390–396 (2015).
- Kovalev, K. et al. Structure and mechanisms of sodium-pumping KR2 rhodopsin. *Sci. Adv.* **5**, eaav2671 (2019).
- Kato, H. E. et al. Structural basis for Na<sup>+</sup> + transport mechanism by a light-driven Na<sup>+</sup> + pump. *Nature* **521**, 48–53 (2015).



6. Kato, Y., Inoue, K. & Kandori, H. Kinetic analysis of H<sup>+</sup>-Na<sup>+</sup> selectivity in a light-driven Na<sup>+</sup>-pumping rhodopsin. *J. Phys. Chem. Lett.* <https://doi.org/10.1021/acs.jpcclett.5b02371> (2015).
7. Abe-Yoshizumi, R., Inoue, K., Kato, H. E., Nureki, O. & Kandori, H. Role of Asn112 in a light-driven sodium ion-pumping rhodopsin. *Biochemistry* **55**, 5790–5797 (2016).
8. Shevchenko, V. et al. 6 Sodium and Engineered Potassium Light-Driven Pumps. *Optogenetics: From Neuronal Function to Mapping and Disease Biology* (ed Krishnarao Appasani) 79–92 (Cambridge University Press, 2017).
9. Vogt, A. et al. Engineered passive potassium conductance in the KR2 sodium pump. *Biophys. J.* <https://doi.org/10.1016/j.bpj.2019.04.001> (2019).
10. Kandori, H., Inoue, K. & Tsunoda, S. P. Light-driven sodium-pumping rhodopsin: a new concept of active transport. *Chem. Rev.* <https://doi.org/10.1021/acs.chemrev.7b00548> (2018).
11. Nishimura, N., Mizuno, M., Kandori, H. & Mizutani, Y. Distortion and a strong hydrogen bond in the retinal chromophore enable sodium-ion transport by the sodium-ion pump KR2. *J. Phys. Chem. B* <https://doi.org/10.1021/acs.jpcc.9b00928> (2019).
12. Gushchin, I. et al. Structure of the light-driven sodium pump KR2 and its implications for optogenetics. *FEBS J.* **283**, 1232–1238 (2016).
13. Shibata, M. et al. Oligomeric states of microbial rhodopsins determined by high-speed atomic force microscopy and circular dichroic spectroscopy. *Sci. Rep.* **8**, 8262 (2018).
14. Kaur, J. et al. Solid-state NMR analysis of the sodium pump *Krokinobacter* rhodopsin 2 and its H30A mutant. *J. Struct. Biol.* <https://doi.org/10.1016/j.jsb.2018.06.001> (2018).
15. Gerwert, K., Freier, E. & Wolf, S. The role of protein-bound water molecules in microbial rhodopsins. *Biochim. Biophys. Acta—Bioenerg.* <https://doi.org/10.1016/j.bbabi.2013.09.006> (2014).
16. Oesterhelt, D. & Stoekenius, W. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nat. New Biol.* **233**, 149–152 (1971).
17. Gordeliy, V. I. et al. Molecular basis of transmembrane signalling by sensory rhodopsin II–transducer complex. *Nature* **419**, 484–487 (2002).
18. Moukhametdzhanov, R. et al. Development of the signal in sensory rhodopsin and its transfer to the cognate transducer. *Nature* **440**, 115–119 (2006).
19. Weinert, T. et al. Proton uptake mechanism in bacteriorhodopsin captured by serial synchrotron crystallography. *Science* <https://doi.org/10.1126/science.aaw8634> (2019).
20. Liebschner, D. et al. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. Sect. D* <https://doi.org/10.1107/S2059798316018210> (2017).
21. Chen, H. F. et al. Time-resolved FTIR study of light-driven sodium pump rhodopsins. *Phys. Chem. Chem. Phys.* <https://doi.org/10.1039/c8cp02599a> (2018).
22. Balashov, S. P. et al. Light-driven Na<sup>+</sup> pump from *Gillisia limnaea*: a high-affinity Na<sup>+</sup> binding site is formed transiently in the photocycle. *Biochemistry* <https://doi.org/10.1021/bi501064n> (2014).
23. Yoshizawa, S. et al. Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.1403051111> (2014).
24. Yun, J. H. et al. Non-cryogenic structure of a chloride pump provides crucial clues to temperature-dependent channel transport efficiency. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.RA118.004038> (2019).
25. Shigetani, A. et al. Long-distance perturbation on Schiff base–counterion interactions by His30 and the extracellular Na<sup>+</sup>-binding site in *Krokinobacter* rhodopsin 2. *Phys. Chem. Chem. Phys.* **20**, 8450–8455 (2018).
26. Otomo, A., Mizuno, M., Inoue, K., Kandori, H. & Mizutani, Y. Allosteric communication with the retinal chromophore upon ion binding in a light-driven sodium ion-pumping rhodopsin. *Biochemistry* <https://doi.org/10.1021/acs.biochem.9b01062> (2020).
27. Shevchenko, V. et al. Inward H<sup>+</sup> pump xenorhodopsin: mechanism and alternative optogenetic approach. *Sci. Adv.* **3**, e1603187 (2017).
28. Volkov, O. et al. Structural insights into ion conduction by channelrhodopsin 2. *Science* **358**, ean8862 (2017).
29. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* <https://doi.org/10.1016/j.pep.2005.01.016> (2005).
30. Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc. Natl Acad. Sci. USA* **93**, 14532–14535 (1996).
31. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**, 706–731 (2009).
32. Gushchin, I. et al. Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. *Proc. Natl Acad. Sci. USA* **110**, 12631–12636 (2013).
33. Gordeliy, V. I. et al. Molecular basis of transmembrane signalling by sensory rhodopsin II–transducer complex. *Nature* **419**, 484–487 (2002).
34. Chizhov, I. & Engelhard, M. Temperature and halide dependence of the photocycle of halorhodopsin from *Natronobacterium pharaonis*. *Biophys. J.* **81**, 1600–1612 (2001).
35. Chizhov, I. et al. The photophobic receptor from *Natronobacterium pharaonis*: temperature and pH dependencies of the photocycle of sensory rhodopsin II. *Biophys. J.* **75**, 999–1009 (1998).
36. von Stetten, D. et al. *In crystallo* optical spectroscopy (icOS) as a complementary tool on the macromolecular crystallography beamlines of the ESRF. *Acta Crystallogr. Sect. D* <https://doi.org/10.1107/S139900471401517X> (2015).
37. Kabsch, W. XDS. *Acta Crystallogr. D* <https://doi.org/10.1107/S0907444909047337> (2010).
38. Winn, M. D. et al. Overview of the CCP 4 suite and current developments. *Acta Crystallogr. Sect. D* <https://doi.org/10.1107/S0907444910045749> (2011).
39. Tickle, I. J. et al. STARANISO (Global Phasing Ltd., Cambridge, 2018).
40. Weierstall, U. et al. Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5**, 3309 (2014).
41. White, T. A. et al. CrystFEL: a software suite for snapshot serial crystallography. *J. Appl. Crystallogr.* **45**, 335–341 (2012).
42. Duisenberg, A. J. M. Indexing in single-crystal diffraction with an obstinate list of reflections. *J. Appl. Crystallogr.* **25**, 92–96 (1992).
43. Ginn, H. M. et al. TakeTwo: an indexing algorithm suited to still images with known crystal parameters. *Acta Crystallogr. Sect. D* **72**, 956–965 (2016).
44. White, T. A. Processing serial crystallography data with CrystFEL: a step-by-step guide. *Acta Crystallogr. Sect. D* **75**, 219–233 (2019).
45. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D* <https://doi.org/10.1107/S0907444909042589> (2010).
46. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
47. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D* <https://doi.org/10.1107/S0907444909052925> (2010).
48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D* **60**, 2126–2132 (2004).
49. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
50. Abraham, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
51. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
52. Zhu, S., Brown, M. F. & Feller, S. E. Retinal conformation governs pKa of protonated Schiff base in rhodopsin activation. *J. Am. Chem. Soc.* **135**, 9391–9398 (2013).
53. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl Acad. Sci. USA* **99**, 12562–12566 (2002).
54. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
55. Maia, F. R. N. C. The coherent X-ray imaging data bank. *Nat. Methods* <https://doi.org/10.1038/nmeth.2110> (2012).
56. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
57. Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).

## Acknowledgements

We acknowledge the Structural Biology Group of the European Synchrotron Radiation Facility (ESRF) and The European Molecular Biology Laboratory (EMBL) unit in Hamburg at Deutsche Elektronen-Synchrotron (DESY) for granting access to the synchrotron beamlines. Special thanks are given to BL13-XALOC technician, ALBA floor coordinators and, most specially, to J.M. Martín-García (Centre for Applied Structural Discovery-ASU, US), for their contribution to adapt the injector to XALOC Beamline specifications. This work was supported by the common program of Agence Nationale de la Recherche (ANR), France and Deutsche Forschungsgemeinschaft, Germany (ANR-15-CE11-0029-02), Ministry of Science and Higher Education of the Russian Federation (075-00337-20-03/FSMG-2020-0003) and by funding from Frankfurt Cluster of Excellence Frankfurt Macromolecular Complexes (to E.B.) by the Max Planck Society (to E.B.) and by the Commissariat à l’Energie Atomique et aux Energies Alternatives (Institut de Biologie Structurale)–Helmholtz–Gemeinschaft Deutscher Forschungszentren (Forschungszentrum Jülich) Special Terms and Conditions 5.1 specific agreement. This work used the icOS and HTX platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) within the Grenoble Partnership for Structural Biology (PSB). Platform access was supported by FRISBI (ANR-10-INBS-05-02) and GRAL, a project of the University Grenoble Alpes graduate school (Ecoles Universitaires de



Recherche) CBH-EUR-GS (ANR-17-EURE-0003). Data collection, data treatment and structure solution were supported by RFBR (19-29-12022).

### Author contributions

C.B., S.V., and D.Z. expressed and purified the proteins; T.B. supervised the expression and purification; R.A. and K.K. crystallized the proteins; D.Z. helped with crystallization; K.K. collected absorption spectra from crystals and performed cryo-trapping of the intermediate; A.R. and P.C. supervised the absorption spectra collection; D.V. performed flash photolysis experiments on KR2 crystals and processed the data; K.K. helped with the flash photolysis experiments; K.K. collected the diffraction data with the help of R.A., D.Z., and solved the structures; X.C. and R.B. performed serial crystallography data acquisition at XALOC; E.Z. and E.M. processed the serial crystallography data; I.G. supervised structure refinement and analysis; G.B. helped with data collection; P.O. performed the molecular dynamics simulations; I.S. prepared the systems for simulations; M.R. performed oligomerization analysis of the proteins; A.A., N.M., V.B., E.B., and G.B. helped with data analysis, V.G. supervised the project; K.K. and V.G. analyzed the results and prepared the manuscript with input from all the other authors.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-16032-y>.

**Correspondence** and requests for materials should be addressed to V.G.

**Peer review information** *Nature Communications* thanks Steffen Wolf and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.




**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Article

# Crystal Structure of the N112A Mutant of the Light-Driven Sodium Pump KR2

Nina Maliar <sup>1,†</sup> , Kirill Kovalev <sup>1,2,3,4,5,\*,†</sup>, Christian Baeken <sup>3,4</sup>, Taras Balandin <sup>3,4</sup>, Roman Astashkin <sup>1,2</sup>, Maksim Rulev <sup>3,4,6</sup>, Alexey Alekseev <sup>1,3,4,5</sup>, Nikolay Ilyinsky <sup>1</sup>, Andrey Rogachev <sup>1,7</sup>, Vladimir Chupin <sup>1</sup>, Dmitry Dolgikh <sup>8</sup>, Mikhail Kirpichnikov <sup>8,9</sup> and Valentin Gordeliy <sup>1,2,3,4,\*</sup>

<sup>1</sup> Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, 141700 Dolgoprudny, Russia; malyar@phystech.edu (N.M.); astashkin.r@gmail.com (R.A.); alekseyka93@gmail.com (A.A.); ilinsky\_nick@mail.ru (N.I.); andrey.v.rogachev@gmail.com (A.R.); vvchupin@gmail.com (V.C.)

<sup>2</sup> Institut de Biologie Structurale (IBS), Université Grenoble Alpes, CEA, CNRS, 38000 Grenoble, France

<sup>3</sup> Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich GmbH, 52428 Jülich, Germany; c.baeken@fz-juelich.de (C.B.); taras.balandin@gmail.com (T.B.); rulevmaksim@gmail.com (M.R.)

<sup>4</sup> JuStruct: Jülich Center for Structural Biology, Forschungszentrum Jülich GmbH, 52428 Jülich, Germany

<sup>5</sup> Institute of Crystallography, University of Aachen (RWTH), 52062 Aachen, Germany

<sup>6</sup> European Synchrotron Radiation Facility (ESRF), 38000 Grenoble, France

<sup>7</sup> Frank Laboratory of Neutron Physics, Joint Institute for Nuclear Research, 141980 Dubna, Russia

<sup>8</sup> M. M. Shemyakin–Yu. A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117997 Moscow, Russia; dolgikh@nmr.ru (D.D.); kirpichnikov@inbox.ru (M.K.)

<sup>9</sup> Biological Faculty, Lomonosov Moscow State University, Leninskie Gory 1, 119991 Moscow, Russia

\* Correspondence: kirill.kovalev@ibs.fr (K.K.); valentin.gordeliy@ibs.fr (V.G.)

† These authors contributed equally to the work.

Received: 8 May 2020; Accepted: 5 June 2020; Published: 8 June 2020



**Abstract:** The light-driven sodium pump KR2, found in 2013 in the marine bacteria *Krokinobacter eikastus*, serves as a model protein for the studies of the sodium-pumping microbial rhodopsins (NaRs). KR2 possesses a unique NDQ (N112, D116, and Q123) set of the amino acid residues in the functionally relevant positions, named the NDQ motif. The N112 was shown to determine the Na<sup>+</sup>/H<sup>+</sup> selectivity and pumping efficiency of the protein. Thus, N112A mutation converts KR2 into an outward proton pump. However, no structural data on the functional conversions of the light-driven sodium pumps are available at the moment. Here we present the crystal structure of the N112A mutant of KR2 in the ground state at the resolution of 2.4 Å. The structure revealed a minor deflection in the central part of the helix C and a double conformation of the L74 residue in the mutant. The organization of the retinal Schiff base and neighboring water molecules is preserved in the ground state of KR2-N112A. The presented data provide structural insights into the effects of the alterations of the characteristic NDQ motif of NaRs. Our findings also demonstrate that for the rational design of the KR2 variants with modified ion selectivity for optogenetic applications, the structures of the intermediate states of both the protein and its functional variants are required.

**Keywords:** microbial rhodopsin; retinal; ion transport; mutation; NDQ motif; sodium pump; X-ray crystallography

## 1. Introduction

Microbial rhodopsins (MRs) are light-sensitive membrane proteins, which use a retinal molecule as a cofactor. All MRs share a similar seven transmembrane (TM)  $\alpha$ -helical fold [1]. The helices (named

from A to G) are interconnected via extracellular and cytoplasmic loops, in many cases containing  $\beta$ -strands [2–4] or small helical regions [5–8]. The retinal is covalently bound to the highly-conserved lysine residue in the TM helix G via a protonated Schiff base (RSB).

The first MR, bacteriorhodopsin from the archaeon *Halobacterium salinarum* (HsBR), was discovered in 1971 [9]. HsBR is a light-driven proton pump, translocating proton outside of the cell and thus creating a transmembrane electrochemical gradient for the ATP synthesis [10]. Since 1971, many MRs with diverse functions were found and characterized [11,12]. The most significant discoveries were archaeal light-driven chloride pumps (halorhodopsins, HRs) [13] and sensory rhodopsins (SRs) [14], bacterial proton pumps (proteorhodopsins, PRs) [15], cation and anion channelrhodopsins (ChRs) [16–20], bacterial sodium and chloride pumps (NaRs and ClRs) [21,22], native inward proton pumps (xenorhodopsins, XeRs and schizorhodopsins, SzRs) [23–26], heliorhodopsins (HeRs) [27], rhodopsin genes in giant viruses [6,8,28,29], enzymorhodopsins [30,31], etc. It was recently shown the MRs are the major solar energy capturing proteins in the ocean [32]. The discovery of new MRs and the growth of the family led to the development of their biotechnological applications, where the most significant one is the optogenetics – able to control living tissues, such as neural or muscle, with light [33,34].

Despite the increasing number of MRs with novel properties, there is only one known group of the active non-proton cation transporters. They can pump  $\text{Li}^+$  and  $\text{Na}^+$  ions outside of the cell but are converted into the outward proton pumps in the presence of larger cations [21] or at low pH [35,36]. It was shown that under physiological conditions in the ocean (pH 8.0, ~500 mM NaCl), the proteins act almost exclusively as light-driven sodium pumps [36]; therefore, they were denoted as sodium-pumping rhodopsins (NaRs). First, described in 2013, a light-driven NaR from the marine bacteria *Krokinobacter eikastus* (KR2) [21], has yet become the most characterized representative of this group of MRs. KR2 remains the only NaR, for which the high-resolution structures are available [2,35,37].

Upon light illumination, KR2 undergoes a photocycle, comprised of the distinct K, L, M, and O intermediate states [21]. The L and M states of KR2 photocycle coexist, with a minor fraction of the M-state and the significant portion of the L. The sodium is not bound inside the protein in the ground state. The sodium uptake occurs with the decay of the M-state as it was also reported for another NaR from *Gillisia limnaea* (GLR) [38]. Thus, the late red-shifted O-state is the only one, where the ion is bound in the core of the KR2 protomer [21,38].

KR2, like all other NaRs, possesses a characteristic NDQ motif (N112, D116, and Q123 residues at the functionally relevant positions, corresponding to D85, T89, and D96 belonging to the helix C in HsBR) [4,39]. Some light was shed on the particular role of each of these amino acid residues. For instance, it was shown that D116 is a proton acceptor from the RSB and forms a direct hydrogen bond with it in the ground, K, L, and O-states [2,35,37,40]. The bond is absent only in the M-state when the proton is translocated from the RSB to the D116 side chain [40]. Recently it was shown that D116 is also a part of the transient sodium binding site near the RSB, formed in the O-state of KR2 photocycle [41]. It was demonstrated that the substitution of the D116 with other residues lead to the loss of the sodium-pumping function of KR2 [21]. The Q123 is a part of a gate at the cytoplasmic side of the protein, separating a large ion-uptake cavity from the RSB region in the ground state [2,35,37]. However, Q123A/V mutants retain sodium-pumping activity, although it is much lower than that of the wild type (WT) protein [37,42]. Thus, it was suggested that the role of Q123 is in the optimization of the sodium pathway inside the protein, and in the creation of the proper inter-helical interactions in the protein [43].

The role of the N112 was investigated by an extensive mutational analysis [44]. It was shown that N112D/G/S/T mutants are able to pump sodium, while N112A/C/P/V/E/Q/L/I/M/F/W mutations convert KR2 into an outward proton pump. N112H/K/Y/R substitutions lead to the loss of any pumping activity of the protein [44]. At the same time, Fourier-transformed infrared (FTIR) spectroscopy data on the N112A mutant of KR2 showed no alterations of the RSB region in the ground and the K-state of the

mutant [45]. The N112 is a part of the transient sodium binding site in the O intermediate state [41]. Thus, the N112 of the NDQ motif is believed to determine the ion selectivity of KR2. However, no structural information on the N112X variants of KR2 is available at the moment, which limits our understanding of the N112 role in functional conversions of NaRs.

The goal of this study is to add important structural information on the functional Na<sup>+</sup>-to-H<sup>+</sup> pump conversion of KR2. For that we present a crystal structure of the N112A mutant of KR2 in the ground state at 2.4 Å resolution. We show that N112A mutation changes neither the oligomeric form nor the internal organization of the protein. However, in KR2-N112A, the helix C is slightly deflected, and the side chain of the L74 residue adopts two alternative conformations. The lack of asparagine at the 112th position of KR2 leads to the enlargement of the polar cavity near the RSB. The structure provides insights into the rearrangements in the NaRs upon alteration of the characteristic NDQ motif. It might also contribute to the rational design of the KR2 variants with modified ion selectivity.

## 2. Materials and Methods

### 2.1. Cloning, Protein Expression, and Purification

Cloning, protein expression, and purification were performed as described previously [2,35]. KR2 gene, synthesized commercially (Eurofins Scientific, Luxembourg), was subcloned into the pET15bmod (Apr) expression plasmid using the restriction sites for *Xba*I and *Xho*I enzymes (Thermo Fisher Scientific, Waltham, MA, USA). A point N112A mutation was introduced into the KR2-gene-containing plasmid by PCR mutagenesis. *E. coli* cells of strain C41 (DE3) (Lucigen, Middleton, WI, USA) were transformed with the KR2-N112A expression plasmid. Transformed cells were grown at 37 °C in shaking baffled flasks in an autoinducing medium (ZYP-5052 [46]) containing 100 mg/L ampicillin and were induced at optical density OD<sub>600</sub> of 0.7–0.9 with 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and 10 μM *all-trans*-retinal. 3 h after induction, the cells were collected by centrifugation at 4000× *g* for 30 min. Harvested cells were disrupted in M-110P Lab Homogenizer (Microfluidics, Westwood, MA, USA) at 25,000 psi in a buffer containing 20 mM Tris-HCl pH 8.0, 5% glycerol, 0.5% Triton X-100 (Sigma-Aldrich, St. Louis, MI, USA) and 50 mg/L DNase I (Sigma-Aldrich, St. Louis, MI, USA). The membrane fraction of the cell lysate was isolated by ultracentrifugation at 90,000× *g* for 1 h at 4 °C. The pellet was resuspended in a buffer containing 50 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> pH 8.0, 0.1 M NaCl and 1% n-Dodecyl β-D-maltoside (DDM) (Anatrace, Affymetrix, Maumee, OH, USA) and stirred overnight for solubilization at 4 °C. The insoluble fraction was removed by ultracentrifugation at 90,000× *g* for 1 h at 4 °C. The supernatant was loaded on the Ni-NTA column (Qiagen, Hilden, Germany) and KR2 was eluted 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, 0.5 M imidazole and 0.1% DDM. The eluate was subjected to size-exclusion chromatography on 24 mL Superdex 200i column (GE Healthcare Life Sciences, Marlborough, MA, 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, 0.05% DDM. Protein-containing fractions with the minimal A<sub>280</sub>/A<sub>525</sub> absorbance ratio were pooled and concentrated to 60 mg/mL for crystallization.

### 2.2. Crystallization

The crystals were grown using the *in meso* approach [47], similar to our previous works [35,48]. The solubilized protein in the crystallization buffer was added to the monooleoyl-formed lipidic phase (Nu-Chek Prep, Elysian, MN, USA). The best crystals were obtained using the protein concentration of 20 mg/mL. The crystals were grown using the precipitate 1.2 M sodium malonate pH 8.0 (Hampton Research, Aliso Viejo, CA, USA). Crystallization probes were set up using the NT8 robotic system (LCP version, Formulatrix, Bedford, MA, USA). The crystals were grown at 22 °C and appeared in 2–4 weeks. Before harvesting, the crystallization drop was opened and covered with 3.4 M sodium malonate solution, pH 8.0, to avoid dehydration. All crystals were harvested using micromounts

(MicroLoops HT, MiTeGen, Ithaca, NY, USA) and were flash-cooled and stored in liquid nitrogen for further crystallographic analysis.

### 2.3. Acquisition and Treatment of Diffraction Data

The diffraction data were collected at 100 K at the Swiss Light Source (SLS, Villigen, Switzerland) beamline X06SA (PXI) equipped with an EIGER X 16M detector (Dectris, Baden-Daettwil, Baden, Switzerland). The data collection statistics are reported in Table 1. The diffraction was anisotropic as determined by decay of the  $CC_{1/2}$  values in  $20^\circ$  cones along the reciprocal cell directions. Diffraction images were processed using XDS [49] (version from 15 March 2019). XSCALE (version from 15 March 2019) was used to merge different datasets. Staraniso server [50] was used to merge, scale, assess the quality, convert intensities to structure factor amplitudes, and generate Free-R labels.

**Table 1.** Crystallographic data collection and refinement statistics. R.m.s.: root mean square.

Data Collection	
Mutant	N112A
Protein Data Bank ID	6YT4
Space group	C222 <sub>1</sub>
Cell dimensions	-
<i>a</i> , <i>b</i> , <i>c</i> (Å)	129.84, 239.70, 134.58
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90
Wavelength (Å)	0.9762
Resolution (Å)	47.85–2.40 (2.455–2.400) *
<i>R</i> <sub>merge</sub> (%)	21.6 (273.4) *
<i>R</i> <sub>pim</sub> (%)	3.5 (43.6) *
$\langle I/\sigma I \rangle$	19.6 (1.8) *
<i>CC</i> <sub>1/2</sub> (%)	94.3 (82.1) *
Completeness, spherical (%)	94.0 (72.1) *
Completeness, ellipsoidal (%)	97.5 (99.2) *
Multiplicity	40.7 (40.2) *
Unique reflections	77,168 (3858) *
Refinement	
Resolution (Å)	20.00–2.40
No. reflections	73,219
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	18.6/21.7
No. atoms	-
Protein	10,791
Retinal	100
Water	286
Lipids	510
Sodium	5
Average <i>B</i> factors (Å <sup>2</sup> )	-
Protein	44
Retinal	42
Water	44
Lipids	73
Sodium	33
R.m.s. deviations	-
Protein bond lengths (Å)	0.0022
Protein bond angles (°)	1.0395
Ramachandran analysis	-
Favored (%)	96.3
Outliers (%)	0.4

\* The data for the highest resolution shell is shown in parenthesis.

#### 2.4. Structure Determination and Refinement

The structure was solved using molecular replacement with MOLREP [51] (version 11.7.02) and the structure of the wild type KR2 in the pentameric form (Protein Data Bank Identifier 6YC3 [41]) as a search model. The model was refined manually using Coot [52] (version 0.8.9.2) and REFMAC5 [53] (version 5.8.0258) from the CCP4 suite [54]. The refinement statistics are summarized in Table 1.

#### 2.5. Modeling of the O-State Structures of KR2-N112X Mutants

The hypothetical models of the O-state of N112A and N112G mutants were created for illustration of the proposed mechanism of functional conversions in KR2 using PyMOL (version 2.3.1). For that, the N112 of the 6XYT model (the O-state of the WT KR2) was substituted with either Ala or G residues. In the case of N112G mutant, the putative water molecule was added at the position of the OD1 atom of N112.

### 3. Results and Discussion

#### 3.1. N112A Mutation Does Not Alter the Pentameric Assembly of KR2

We have expressed and purified KR2-N112A in *E. coli* strain C41 cells. Following the previously reported data, the purified protein has a spectrum with the characteristic maximum absorption wavelength of 522 nm [42,44].

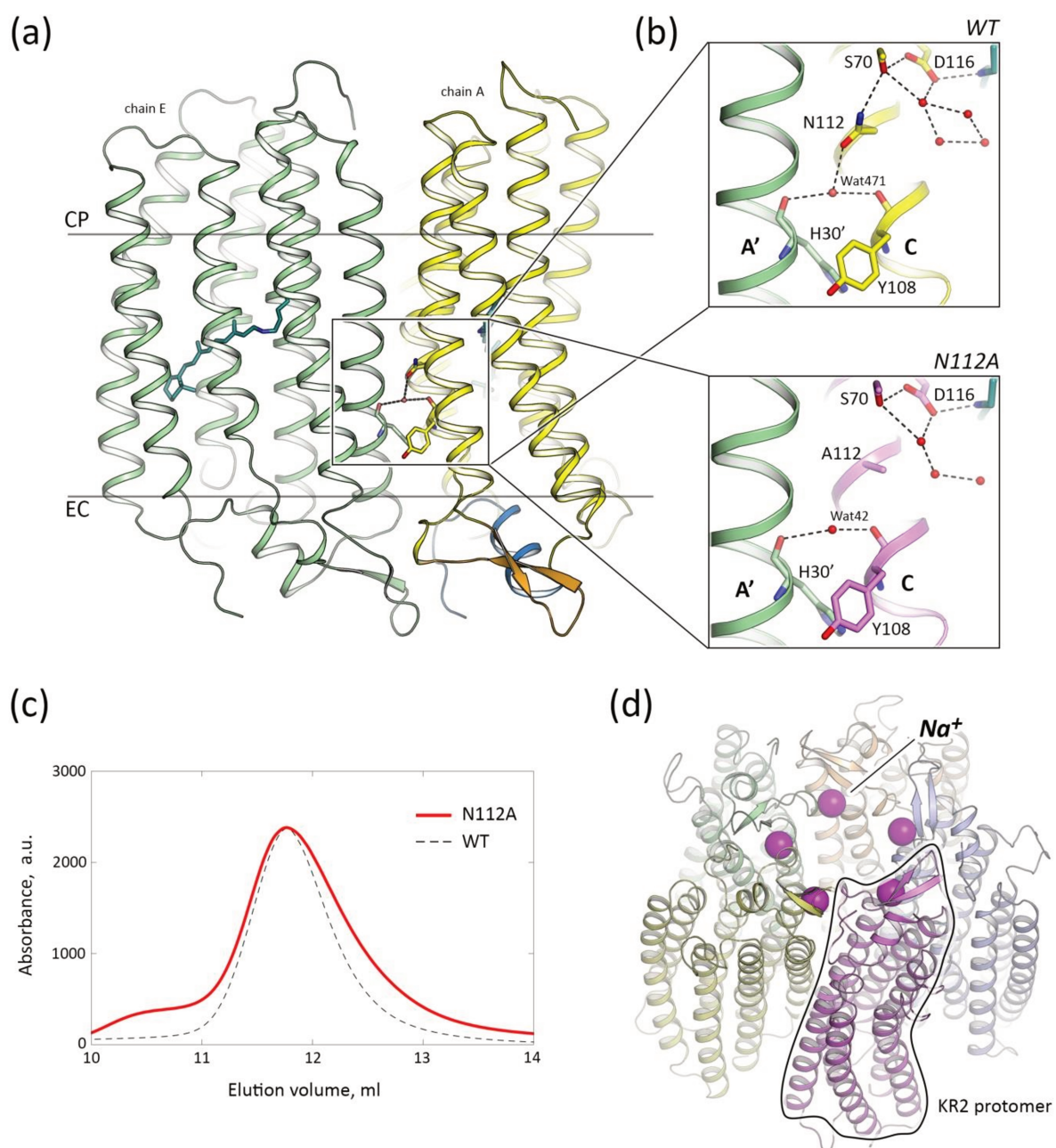
It was shown that KR2 forms pentamers when solubilized in the DDM at pH higher than 6.0 or reconstituted in the lipids at both acidic and neutral pH values [35,55]. KR2 is also in the pentameric form in the crystals grown using in meso approach [47] at the pH as low as 4.9 [2,35,56]. Importantly, KR2 pumps sodium, only being in the pentameric form [35,55]. Nothing is known on the functionality of the monomeric form of the protein. Therefore, the analysis of the oligomeric assembly of the protein and its functional mutants could provide valuable insights into the basis of the pentamerization of KR2 and its functional role. For instance, a recent analysis of the size-exclusion chromatography (SEC) profile of the D116N mutant of KR2 demonstrated that this mutation destabilizes the protein pentamers [41].

The oligomeric state of the KR2-N112A was shown to be pentameric in lipids [55]. However, there is a great penalty in terms of free energy for the protein oligomer to dissociate being inside the lipid membrane. Thus, to probe the stability of the KR2-N112A pentamers, its oligomeric state should be studied in other environments, which allow the disruption of the protein complex. The more detailed analysis of the KR2-N112A is also of specific interest, as the N112 side chain is pointed towards the oligomerization interface of KR2 and is hydrogen-bonded via a water molecule (Wat471 in the case of the chain A of the 6YC3 model [41]) to the main chain oxygen of H30 residue of the nearby protomer, additionally stabilizing the pentameric assembly of the protein (Figure 1a,b).

In the present study, we performed SEC for the purification and analysis of the KR2-N112A. The SEC profile is similar to that of the WT protein (Figure 1c). The slight broadening of the peak in the KR2-N112A SEC profile may originate from the different viscosity of the samples, likely due to the fluctuating detergent concentration (Figure 1a). We cannot exclude that the tailing of the SEC profile peak may also come from other differing properties of the samples. Nevertheless, the KR2-N112A did not pass through the 100 kDa filter when concentrated for the crystallization, which is characteristic for the pentameric WT protein, but not for smaller oligomers of, for instance, H30K and Y154F mutants of KR2 [35]. As will be described later, KR2-N112A also forms pentamers in crystals, similar to those of the WT protein (Figure 1d).

Therefore, we conclude that the interactions of the N112 side chain are not critical for the pentamerization of KR2, and the water molecule, hydrogen-bonded to the N112 in the ground state, is likely to appear for the stabilization of the polar side chain of the residue.





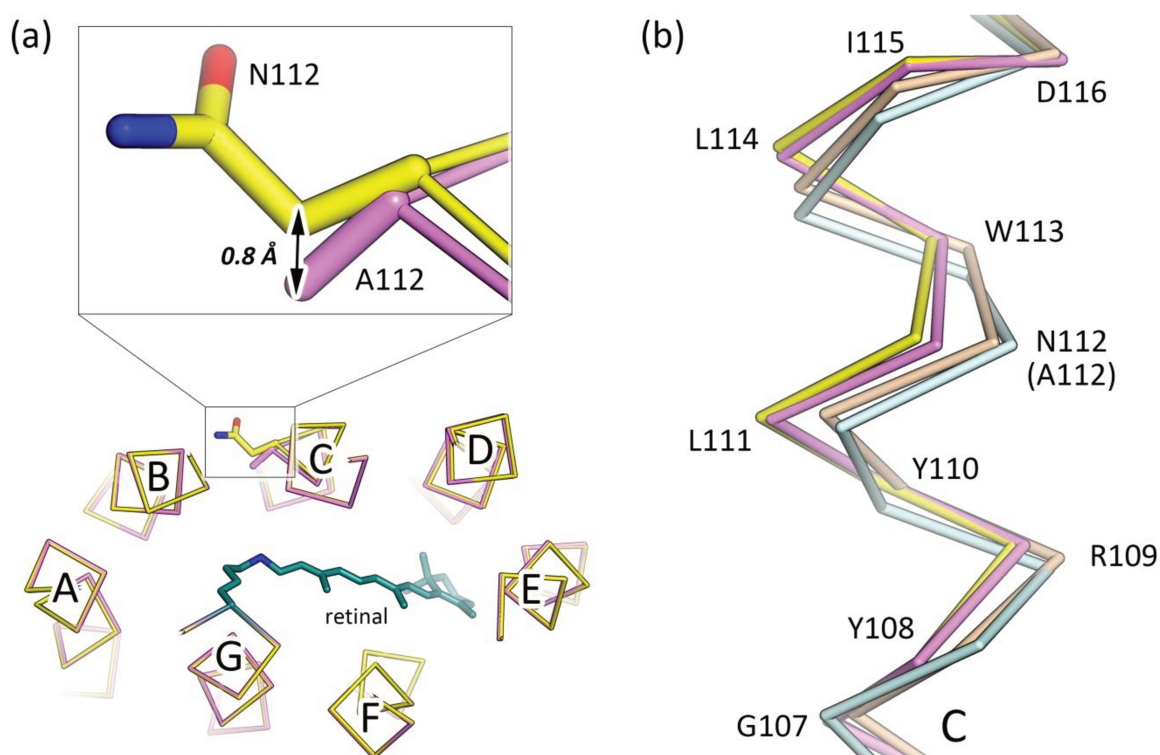
**Figure 1.** Pentameric organization of KR2-N112A. (a) Side view of two neighboring protomers (chain A, yellow and chain E, green of the 6YC3 model) of KR2 pentamer in the membrane. N-terminal  $\alpha$ -helix of chain A is colored blue. BC loop of the chain A is colored orange. Cytoplasmic (CP) and extracellular (EC) sides of the membrane are marked. (b) Detailed view of the inter-protomers contacts in the region of the N112 of the WT (top part) and A112 of the N112A mutant (bottom part). Neighboring protomer is colored green. The retinal cofactor is colored teal. Helices are marked with bold capital letters. (c) Size-exclusion chromatography (SEC) profiles of the N112A mutant (red line) and WT KR2 (black dashed line). The position of the absorption peak is the same for the WT KR2 and N112A variant ( $\sim 11.8$  mL when using 24 mL Superdex 200i column). The absorption was measured at 525 and 522 nm for the WT [21] and N112A [42] KR2, respectively. The small left peak of the KR2-N112A profile likely corresponds to the fraction of the aggregates. (d) KR2-N112A pentamer, view from the cytoplasmic side. Sodium ions, located at the oligomerization interface, are shown with purple spheres. One of five KR2 protomers is contoured for clarity.



### 3.2. Crystal Structure of the N112A Protomer

The N112A mutant was concentrated to ~60 mg/mL and crystallized using *in meso* approach [47]. The crystals appeared within a month and diffracted to 2.4 Å resolution. We obtained the KR2-N112A structure in its biologically relevant pentameric form at the resolution of 2.4 Å. As the diffraction of the KR2-N112A crystals was anisotropic, we merged the X-ray diffraction data from 6 crystals to improve the quality of the electron density maps and the corresponding model (Table 1, Figure S1).

The crystal packing, as well as the overall fold of the mutant, is very similar to that of the WT protein (Figure 1d). The root-mean-square deviation (RMSD) between the ground state structures of the N112A mutant and WT KR2 (PDB ID: 6YC3 [41]) is 0.15 Å. At the same time, 109–115 residues of the helix C are slightly shifted closer to the core of the protein (Figure 2a). Thus, the helix C of KR2-N112A is deflected in the same manner as it appears in the «compact» conformation of pentameric KR2 [2,56] and also in the monomeric form of the protein [2,35] (Figure 2b). However, in the case of the mutant, the shift of the helix is much smaller (Figure 2b).



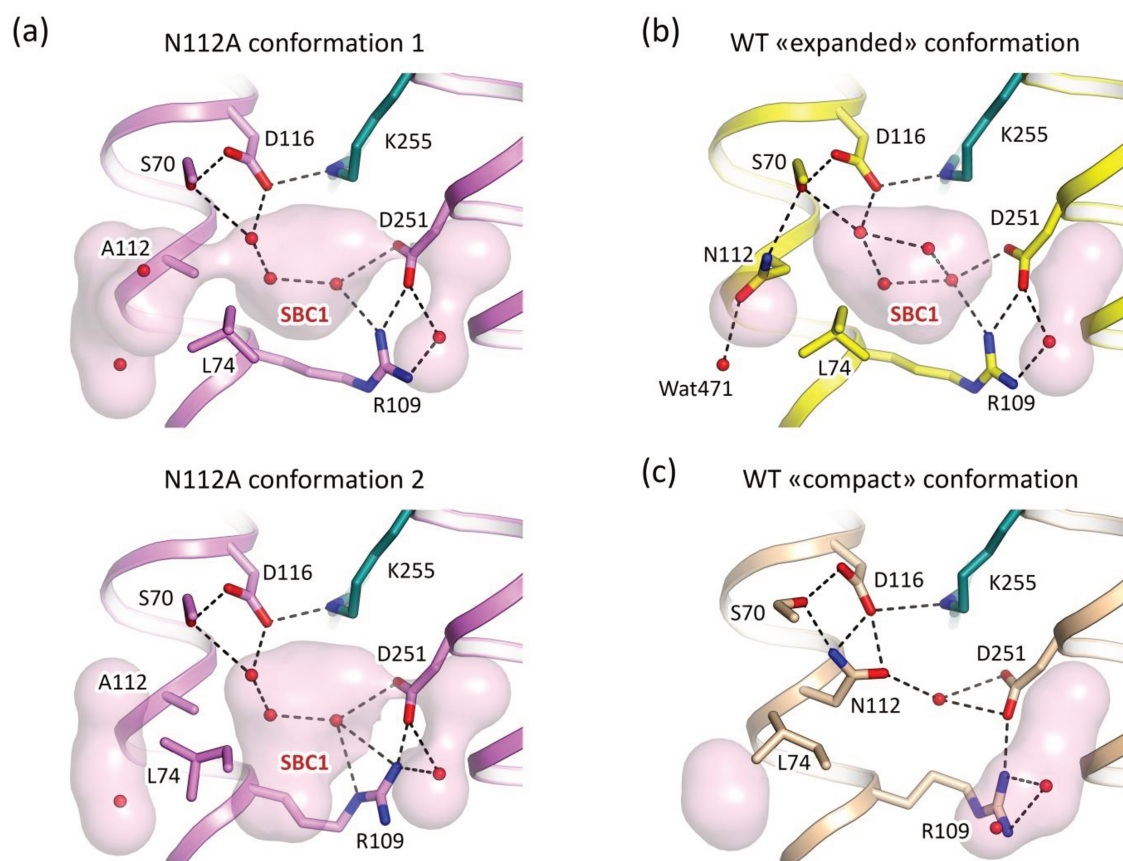
**Figure 2.** Deflection of the helix C in the N112A mutant of KR2. (a) Displacement of the C $\beta$  atoms of A112 and N112 of the N112A and WT KR2 structures, respectively. The 0.8 Å shift towards the inside of the KR2 protomer is shown with a black arrow. The retinal cofactor is colored teal. (b) Side view of the structural alignment of the helix C of the “expanded” conformation of the WT KR2 (yellow, PDB ID: 6YC3 [41]), KR2-N112A (violet, PDB ID: 6YT4, present work), “compact” conformation of the WT KR2 (light-orange, PDB ID: 4XTN, chain I [2]), and monomeric form of the WT KR2 (light-cyan, PDB ID: 4XTL [2]).

The sodium ion, found in the structures of the WT protein at the pentamerization interface, is also present in the model of KR2-N112A (Figure 1b). The organization of the binding site is the same as that in the WT protein, which is in contrast to the FTIR spectroscopy data on the N112A mutant of KR2 presented in [21]. Indeed, Inoue et al. found that KR2-N112A binds sodium in the ground state; however, the binding site is modified in the mutant [21]. Similar inconsistency of the FTIR spectroscopy data on KR2 with the high-resolution structural data was already found for the H30A mutant of the protein. Particularly, the same study [21] reported no sodium binding in the ground state of KR2-H30A.

By contrast, the binding site was revealed in the 2.2 Å resolution crystal structure of the mutant in its pentameric form (PDB ID: 6YC1) [41]. In this case, similarly to the N112A mutant, the site is identical to that of the WT protein. We suggest that such inconsistency may originate from the differences of the sample preparation, or from the different lipid environments of the protein in the spectroscopy experiments and in the crystals.

### 3.3. The Organization of the RSB Region of KR2-N112A

The WT KR2 may form two principal conformations, which were named «expanded» and «compact» (Figure 3b,c) [35,56]. In the «expanded» conformation, there is a large water-filled cavity near the RSB, called the Schiff base cavity 1 (SBC1) [35]. In this configuration, the N112 side chain is pointed outside of the KR2 protomer and faces the oligomerization interface (Figure 1a,b). It is stabilized by the hydrogen bond with the water molecule (Wat471 for the chain A of the 6YC3 model [41]) (Figure 3b). To avoid the steric conflict, L74 side chain orients inside the protomer. On the contrary, in the «compact» conformation, the SBC1 is absent, and N112 is turned to the D116 and S70 residues and is hydrogen bonded to these residues (Figure 3c). Consequently, L74 is flipped outside the protein to allow enough space for the N112 (Figure 3c).



**Figure 3.** Schiff base region of the N112A mutant. (a) Two conformations of the RSB region of the N112A mutant of KR2 (PDB ID: 6YT4, present work). The flip of the L74 side chain leads to the enlargement of the Schiff base cavity 1 (marked SBC1 in the Figure) in the conformation 2. (b) The “expanded” conformation of the WT KR2 (PDB ID: 6YC3 [41]). (c) The “compact” conformation of the WT KR2 (PDB ID: 4XTN, chain I [2]). Water molecules are shown with red spheres. Cavities were calculated using HOLLOW [57] and are shown with a pink surface. Hydrogen bonds are shown with black dashed lines. The side chain of the K255 residue and covalently bound retinal cofactor are colored teal.

In the ground state under physiological conditions, and being in the biologically relevant pentameric state, KR2 is in the «expanded» conformation [35] (Figure 3b). The conformation, similar to the «compact» one, appears transiently in the O intermediate state of the KR2 photocycle when the sodium ion is bound in the core of the protein [41]. Thus, the conformational «expanded»-to-«compact» and «compact»-to-«expanded» switches guide the uptake and release of the sodium ion by the protein [41].

Initially, the «compact» conformation was found in the ground state of the WT KR2 at acidic pH values [35] (Figure 3c). It was also demonstrated that KR2 pumps protons at pH 4.3 [35]. Therefore, the «compact» conformation was suggested to be a determinant of the proton-pumping mode of KR2 [35]. However, the proton-pumping ability of KR2 at low pH values remains speculative due to the small  $\Delta\text{pH}$  signals in the reported experiments [35]. Consequently, authors of ref [35] suggested that alternatively, at acidic pH, KR2 is in an inactive state with the protonated RSB counterion D116. If this is the case, the residual weak pumping activity might correspond to the remaining minor fraction of the protein with deprotonated D116 [35]. As N112A mutation converts KR2 into the proton pump, the study of the conformation of the ground state of KR2-N112A under physiological conditions might help to distinguish between the two proposed mechanisms of KR2 functional and structural switches upon pH decrease. It could also provide insights into the roles of the «expanded» and «compact» conformations of KR2.

Consequently, we next analyzed the organization of the RSB region of the N112A mutant of KR2. In the mutant, the asparagine side chain of 112th residue is absent, which gives more space for the L74. Following this, L74 adopts two alternative conformations, one of which is similar to that of the «expanded» form (Figure 3a, Figure S1). At the same time, the second is closer to that of the «compact» conformation of the WT protein (Figure 3a, Figure S1).

It should be noted that the R109 side chain adopts two conformations, similar to those in the monomeric model of the KR2 (PDB ID: 4XTL [2]). This is most likely caused by the minor deflection of the helix C (Figure 2b).

Although A112 is located closer to the water molecules in the SBC1 of the KR2 ground state ( $\text{C}\beta$  atom of A112 in the N112A mutant is shifted by 0.8 Å to the inside of the protein in comparison to the  $\text{C}\beta$  atom of N112 in the 6YC3 model), the positions of the water molecules in the cavity remain the same to those in the WT protein (Figures 2a and 3a). This is consistent with the recent FTIR spectroscopy data on KR2 and its N112A mutant, where the authors concluded from the N-D stretching vibrations shifts that the organization of the RSB region is nearly the same in the N112A and WT proteins [45]. Hence, the SBC1 is retained in KR2-N112A and is of the same size as that in the WT protein; it is even enlarged in the second conformation of the molecule when L74 is flipped outside the protomer.

Therefore, we conclude that KR2-N112A shapes the conformation, similar to the «expanded» one of the WT protein, however, with minor alterations. In other words, N112A mutation does not result in the appearance of the «compact» conformation in the ground state of the protein, which could be associated with the conversion of KR2 into a proton pump. This allows us to speculate that either the forming of the «compact» conformation is not the determinant of the switch from the sodium- to the proton-pumping mode, or there could be other mechanisms underneath the functional conversion of the KR2-N112A.

#### 3.4. Role of N112 in the Ion Selectivity of KR2

As it was demonstrated in the mutational study of KR2, N112 is (at least) one of the determinants of the  $\text{Na}^+/\text{H}^+$  selectivity of the protein [44]. As the N112 side chain is pointed away from the active center of the protein in the ground state and interacts neither with the water molecules in the SBC1 nor with the other critical residues of the putative sodium pathway, the molecular basis of the mechanism of this selectivity remains elusive.

Recently, the crystal structure of the O intermediate state of KR2 has been reported, which showed the transient sodium binding site in the core of the KR2 protomer (PDB ID: 6XYT) [41]. It was

shown that with the rise of the O-state, N112 flips towards D116 and S70 residues (Figure S2). Thus, the orientation of the N112 is similar to that in the «compact» conformation. In this configuration, the ion is coordinated by the side chains of S70, N112, and D116 residues. The geometry is adapted for the binding of sodium. Indeed, the cation is coordinated by five oxygen atoms with a mean sodium-oxygen distance of 2.3 Å (Figure S2). However, the role of N112 in the ion selectivity of KR2 was not discussed in [41].

The flipping motion of the N112 side chain could not occur in the KR2-N112A; thus, the «expanded»-to-«compact» conformational switch is also not possible in the mutant. At the same moment, it was shown that KR2-N112A does not form the O-state [42,44]; however, it forms all other intermediates, such as K, L, and M. Therefore, we suggest that N112 is perfectly optimized in KR2 (as well as the corresponding Asn in any other NaRs) for an accurate transient binding of the sodium ion in the O-state, and does not show any key structural and/or functional role in the ground state of the protein.

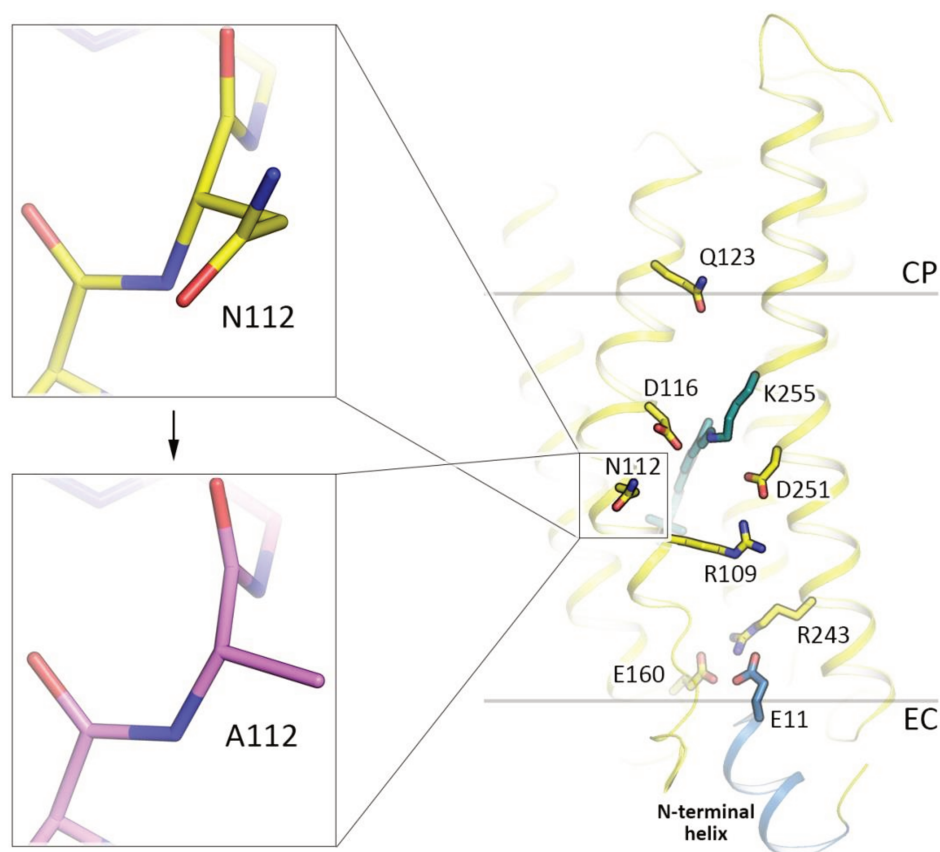
In this approximation, the fact that N112D/S/T/G mutants retain sodium-pumping activity [44] could be easily explained. Indeed, Asp often coordinates sodium ions in proteins and is similar to Asn in terms of geometry and size. Ser and Thr are also able to coordinate cations due to the presence of the oxygen atoms in their side chains. However, their size is smaller than that of the asparagine group, which may explain significantly lowered sodium-pumping activity of the N112S/T mutants. In the case of KR2-N112G, the retaining of the sodium pumping activity is surprising. At the same time, the transiently bound sodium ion in N112G mutant is likely coordinated by the additional water molecule, which may be located at the position of the N112 OD1 atom of the O-state of KR2 (PDB ID: 6XYT [41]) (See Section 2.5 of Materials and Methods, Figure S2). The appearance of the water molecule at this position is constricted in the case of N112A, due to the steric conflict with the Ala side chain, which may explain the absence of the sodium-pumping activity of the KR2-N112A, but not of KR2-N112G (Figure S2).

On the other hand, similar to N112A, N112C/P/V/E/Q/L/I/M/F/W mutations convert KR2 into a proton pump. This fact is easily understandable since these mutations cannot support efficient sodium binding inside the protein. Therefore, they also cannot provide the formation of the O-state, which is characteristic of the sodium-pumping mode of KR2. Particularly, nonpolar side chains of Pro, Val, Leu, Ile, Met, Phe, and Trp amino acid residues are not consistent with the sodium coordination. Cys was demonstrated to coordinate metal ions like manganese, iron, copper, cobalt, nickel, and zinc [58]. However, it does not participate in the binding of sodium, potassium, magnesium, and calcium [58]. At the same time, polar Glu and Gln residues are supposedly too bulky to allow enough space for sodium binding inside the KR2 protomer.

#### 4. Conclusions

In this work, we described the crystal structure of functionally important N112A mutant of the light-driven sodium pump KR2. The mutation converts the protein to an outward proton pump. N112 is a part of the NDQ motif, characteristic for all known NaRs (Figure 4). Our study provides a deeper understanding of the structural rearrangements, which occur in response to the variations of the critical functional residues of the protein. Although the overall fold and the retinal Schiff base region of the mutant are similar to the «expanded» conformation of the WT KR2, the SBC1 is enlarged in the KR2-N112A. The helix C is also slightly deflected in the mutant. The N112A mutation also does not affect the functionally important pentameric assembly of the protein; thus, the involvement of the N112 into the interprotomer interactions in the ground state of KR2 is not crucial for the stability of the oligomer.





**Figure 4.** Location of the N112A substitution site (yellow, the WT of KR2, PDB ID: 6YC3 [41]; magenta, the N112A variant of KR2, present work, PDB ID: 6YT4). Hydrophobic/hydrophilic membrane core boundaries were calculated with PPM server [59] and are shown with gray lines. Cytoplasmic (CP) and extracellular (EC) sides of the membrane are marked. N-terminal  $\alpha$ -helix of KR2 is colored blue. Helices A, B, and BC loop are hidden for clarity.

Hence, the reported here crystal structure of the KR2-N112A, together with the available functional and spectroscopy data, allows us to conclude that the N112 does not show any important structural and/or functional roles in the ground state of the protein. Therefore, our study shows that the analysis of the ground state of N112A cannot solely explain the  $\text{Na}^+$ -to- $\text{H}^+$  pump conversion of KR2. However, N112 is vital for the organization of the transient sodium binding site in the O-state of the KR2 photocycle. Thus, we speculate that the major role of N112 as a key functional amino acid residue is in the organization of the site. For the complete understanding of the functional conversions of NaRs, and the principles of the ion selectivity of the proteins, further investigations on KR2 are required. Particularly, the structures of the intermediate states of KR2 mutants are essential.

Finally, the engineering of KR2 variants with modified or optimized pumping efficiency and ion selectivity is of high importance for the development of KR2-based optogenetic tools. However, the crystal structure of the ground state of KR2-N112A is a beautiful demonstration of the fact that the rational design of such variants should be supported by the structures of the intermediate state's models of not only WT but functional mutants of the protein.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4352/10/6/496/s1>, Figure S1: Examples of the electron density maps of the KR2-N112A structure., Figure S2: Transient sodium binding site in the O-state of the KR2 photocycle.

**Author Contributions:** Conceptualization, K.K. and V.G.; investigation, N.M., K.K., C.B., T.B., R.A., A.R., N.I., A.A., M.R., V.G.; writing—original draft preparation, K.K. and N.M.; writing—review and editing, K.K. and V.G.; visualization, K.K., and N.M.; supervision, V.G.; project administration, N.M., K.K. and V.G.; funding acquisition, V.G, V.C., D.D., and M.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Russian Foundation for Basic Research (grants 17-00-00164, 17-00-00165, 17-00-00166, and 17-00-00167 komfi).

**Acknowledgments:** We acknowledge the Paul Scherrer Institut, Villigen, Switzerland, for the provision of synchrotron radiation beamtime at beamline X06SA (PXI) of the SLS and would like to thank Takashi Tomizaki for assistance. Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank under the accession code 6YT4.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

WT	Wild type
KR2	<i>Krokinobacter eikastus</i> rhodopsin 2
SEC	Size exclusion chromatography
RSB	Retinal Schiff base
SBC1	Schiff base cavity 1

## References

1. Gushchin, I.; Gordeliy, V. Microbial Rhodopsins. In *Membrane Protein Complexes: Structure and Function*; Springer: Singapore, 2018; pp. 19–56.
2. Gushchin, I.; Shevchenko, V.; Polovinkin, V.; Kovalev, K.; Alekseev, A.; Round, E.; Borshchevskiy, V.; Balandin, T.; Popov, A.; Gensch, T.; et al. Crystal structure of a light-driven sodium pump. *Nat. Struct. Mol. Biol.* **2015**, *22*, 390–396. [[CrossRef](#)]
3. Gushchin, I.; Chervakov, P.; Kuzmichev, P.; Popov, A.N.; Round, E.; Borshchevskiy, V.; Ishchenko, A.; Petrovskaya, L.; Chupin, V.; Dolgikh, D.A.; et al. Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12631–12636. [[CrossRef](#)]
4. Luecke, H.; Schobert, B.; Richter, H.T.; Cartailler, J.P.; Lanyi, J.K. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **1999**, *291*, 899–911. [[CrossRef](#)]
5. Kovalev, K.; Volkov, D.; Astashkin, R.; Alekseev, A.; Gushchin, I.; Haro-Moreno, J.M.; Chizhov, I.; Siletsky, S.; Mamedov, M.; Rogachev, A.; et al. High-resolution structural insights into the heliorhodopsin family. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 4131–4141. [[CrossRef](#)]
6. Zabelskii, D.; Alekseev, A.; Kovalev, K.; Oliviera, A.-S.; Balandin, T.; Soloviov, D.; Bratanov, D.; Volkov, D.; Vaganova, S.; Astashkin, R.; et al. Viral channelrhodopsins: Calcium-dependent Na<sup>+</sup>/K<sup>+</sup> selective light-gated channels. *bioRxiv* **2020**. [[CrossRef](#)]
7. Shihoya, W.; Inoue, K.; Singh, M.; Konno, M.; Hososhima, S.; Yamashita, K.; Ikeda, K.; Higuchi, A.; Izume, T.; Okazaki, S.; et al. Crystal structure of heliorhodopsin. *Nature* **2019**, *574*, 132–136. [[CrossRef](#)] [[PubMed](#)]
8. Needham, D.M.; Yoshizawa, S.; Hosaka, T.; Poirier, C.; Choi, C.J.; Hehenberger, E.; Irwin, N.A.T.; Wilken, S.; Yung, C.-M.; Bachy, C.; et al. A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 20574–20583. [[CrossRef](#)] [[PubMed](#)]
9. Oesterhelt, D.; Stoeckenius, W. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nat. New Biol.* **1971**, *233*, 149–152. [[CrossRef](#)] [[PubMed](#)]
10. Matsuno-Yagi, A.; Mukohata, Y. ATP synthesis linked to light-dependent proton uptake in a red mutant strain of *Halobacterium* lacking bacteriorhodopsin. *Arch. Biochem. Biophys.* **1980**, *199*, 297–303. [[CrossRef](#)]
11. Govorunova, E.G.; Sineshchekov, O.A.; Li, H.; Spudich, J.L. Microbial Rhodopsins: Diversity, Mechanisms, and Optogenetic Applications. *Annu. Rev. Biochem.* **2017**, *86*, 845–872. [[CrossRef](#)]
12. Ernst, O.P.; Lodowski, D.T.; Elstner, M.; Hegemann, P.; Brown, L.S.; Kandori, H. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chem. Rev.* **2014**, *114*, 126–163. [[CrossRef](#)] [[PubMed](#)]
13. Schobert, B.; Lanyi, J.K. Halorhodopsin is a light-driven chloride pump. *J. Biol. Chem.* **1982**, *257*, 10306–10313. [[CrossRef](#)] [[PubMed](#)]
14. Bogomolni, R.A.; Spudich, J.L. The photochemical reactions of bacterial sensory rhodopsin-I. Flash photolysis study in the one microsecond to eight second time window. *Biophys. J.* **1987**, *52*, 1071–1075. [[CrossRef](#)]

15. Béjà, O.; Spudich, E.N.; Spudich, J.L.; Leclerc, M.; DeLong, E.F. Proteorhodopsin phototrophy in the ocean. *Nature* **2001**, *411*, 786–789. [[CrossRef](#)] [[PubMed](#)]
16. Nagel, G.; Szellas, T.; Huhn, W.; Kateriya, S.; Adeishvili, N.; Berthold, P.; Ollig, D.; Hegemann, P.; Bamberg, E. Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13940–13945. [[CrossRef](#)]
17. Nagel, G. Channelrhodopsin-1: A Light-Gated Proton Channel in Green Algae. *Science* **2002**, *296*, 2395–2398. [[CrossRef](#)]
18. Govorunova, E.G.; Sineshchekov, O.A.; Spudich, J.L. Structurally Distinct Cation Channelrhodopsins from Cryptophyte Algae. *Biophys. J.* **2016**, *110*, 2302–2304. [[CrossRef](#)]
19. Oppermann, J.; Fischer, P.; Silapetere, A.; Liepe, B.; Rodriguez-Rozada, S.; Flores-Uribe, J.; Peter, E.; Keidel, A.; Vierock, J.; Kaufmann, J.; et al. MerMAIDs: A family of metagenomically discovered marine anion-conducting and intensely desensitizing channelrhodopsins. *Nat. Commun.* **2019**, *10*, 3315. [[CrossRef](#)]
20. Govorunova, E.G.; Sineshchekov, O.A.; Janz, R.; Liu, X.; Spudich, J.L. Natural light-gated anion channels: A family of microbial rhodopsins for advanced optogenetics. *Science* **2015**, *349*, 647–650. [[CrossRef](#)]
21. Inoue, K.; Ono, H.; Abe-Yoshizumi, R.; Yoshizawa, S.; Ito, H.; Kogure, K.; Kandori, H. A light-driven sodium ion pump in marine bacteria. *Nat. Commun.* **2013**, *4*, 1678. [[CrossRef](#)]
22. Kim, K.; Kwon, S.-K.; Jun, S.-H.; Cha, J.S.; Kim, H.; Lee, W.; Kim, J.F.; Cho, H.-S. Crystal structure and functional characterization of a light-driven chloride pump having an NTQ motif. *Nat. Commun.* **2016**, *7*, 12677. [[CrossRef](#)] [[PubMed](#)]
23. Inoue, K.; Ito, S.; Kato, Y.; Nomura, Y.; Shibata, M.; Uchihashi, T.; Tsunoda, S.P.; Kandori, H. A natural light-driven inward proton pump. *Nat. Commun.* **2016**, *7*, 13415. [[CrossRef](#)] [[PubMed](#)]
24. Shevchenko, V.; Mager, T.; Kovalev, K.; Polovinkin, V.; Alekseev, A.; Juettner, J.; Chizhov, I.; Bamann, C.; Vavourakis, C.; Ghai, R.; et al. Inward H<sup>+</sup> pump xenorhodopsin: Mechanism and alternative optogenetic approach. *Sci. Adv.* **2017**, *3*. [[CrossRef](#)] [[PubMed](#)]
25. Inoue, K.; Tsunoda, S.P.; Singh, M.; Tomida, S.; Hososhima, S.; Konno, M.; Nakamura, R.; Watanabe, H.; Bulzu, P.-A.; Banciu, H.L.; et al. Schizorhodopsins: A family of rhodopsins from Asgard archaea that function as light-driven inward H<sup>+</sup> pumps. *Sci. Adv.* **2020**, *6*, eaaz2441. [[CrossRef](#)]
26. Bulzu, P.-A.; Andrei, A.-Ş.; Salcher, M.M.; Mehrshad, M.; Inoue, K.; Kandori, H.; Beja, O.; Ghai, R.; Banciu, H.L. Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. *Nat. Microbiol.* **2019**, *4*, 1129–1137. [[CrossRef](#)]
27. Pushkarev, A.; Inoue, K.; Larom, S.; Flores-Uribe, J.; Singh, M.; Konno, M.; Tomida, S.; Ito, S.; Nakamura, R.; Tsunoda, S.P.; et al. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **2018**, *558*, 595–599. [[CrossRef](#)]
28. Yutin, N.; Koonin, E.V. Proteorhodopsin genes in giant viruses. *Biol. Direct* **2012**, *7*, 34. [[CrossRef](#)]
29. Bratanov, D.; Kovalev, K.; Machtens, J.-P.; Astashkin, R.; Chizhov, I.; Soloviov, D.; Volkov, D.; Polovinkin, V.; Zabelskii, D.; Mager, T.; et al. Unique structure and function of viral rhodopsins. *Nat. Commun.* **2019**, *10*, 4939. [[CrossRef](#)]
30. Luck, M.; Mathes, T.; Bruun, S.; Fudim, R.; Hagedorn, R.; Tran Nguyen, T.M.; Kateriya, S.; Kennis, J.T.M.; Hildebrandt, P.; Hegemann, P. A Photochromic Histidine Kinase Rhodopsin (HKR1) That Is Bimodally Switched by Ultraviolet and Blue Light. *J. Biol. Chem.* **2012**, *287*, 40083–40090. [[CrossRef](#)]
31. Yoshida, K.; Tsunoda, S.P.; Brown, L.S.; Kandori, H. A unique choanoflagellate enzyme rhodopsin exhibits light-dependent cyclic nucleotide phosphodiesterase activity. *J. Biol. Chem.* **2017**, *292*, 7531–7541. [[CrossRef](#)]
32. Gómez-Consarnau, L.; Raven, J.A.; Levine, N.M.; Cutter, L.S.; Wang, D.; Seegers, B.; Arístegui, J.; Fuhrman, J.A.; Gasol, J.M.; Sañudo-Wilhelmy, S.A. Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Sci. Adv.* **2019**, *5*, eaaw8855. [[CrossRef](#)]
33. Boyden, E.S.; Zhang, F.; Bamberg, E.; Nagel, G.; Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **2005**, *8*, 1263–1268. [[CrossRef](#)] [[PubMed](#)]
34. Deisseroth, K. Optogenetics: 10 years of microbial opsins in neuroscience. *Nat. Neurosci.* **2015**, *18*, 1213–1225. [[CrossRef](#)] [[PubMed](#)]
35. Kovalev, K.; Polovinkin, V.; Gushchin, I.; Alekseev, A.; Shevchenko, V.; Borshchevskiy, V.; Astashkin, R.; Balandin, T.; Bratanov, D.; Vaganova, S.; et al. Structure and mechanisms of sodium-pumping KR2 rhodopsin. *Sci. Adv.* **2019**, *5*, eaav2671. [[CrossRef](#)] [[PubMed](#)]



36. Kato, Y.; Inoue, K.; Kandori, H. Kinetic Analysis of H<sup>+</sup>–Na<sup>+</sup> Selectivity in a Light-Driven Na<sup>+</sup> -Pumping Rhodopsin. *J. Phys. Chem. Lett.* **2015**, *6*, 5111–5115. [[CrossRef](#)]
37. Kato, H.E.; Inoue, K.; Abe-Yoshizumi, R.; Kato, Y.; Ono, H.; Konno, M.; Hososhima, S.; Ishizuka, T.; Hoque, M.R.; Kunitomo, H.; et al. Structural basis for Na<sup>+</sup> transport mechanism by a light-driven Na<sup>+</sup> pump. *Nature* **2015**, *521*, 48–53. [[CrossRef](#)]
38. Balashov, S.P.; Imasheva, E.S.; Dioumaev, A.K.; Wang, J.M.; Jung, K.-H.; Lanyi, J.K. Light-Driven Na<sup>+</sup> Pump from *Gillisia limnaea*: A High-Affinity Na<sup>+</sup> Binding Site Is Formed Transiently in the Photocycle. *Biochemistry* **2014**, *53*, 7549–7561. [[CrossRef](#)]
39. Gerwert, K.; Freier, E.; Wolf, S. The role of protein-bound water molecules in microbial rhodopsins. *Biochim. Biophys. Acta Bioenerg.* **2014**, *1837*, 606–613. [[CrossRef](#)]
40. Nishimura, N.; Mizuno, M.; Kandori, H.; Mizutani, Y. Distortion and a Strong Hydrogen Bond in the Retinal Chromophore Enable Sodium-Ion Transport by the Sodium-Ion Pump KR2. *J. Phys. Chem. B* **2019**, *123*, 3430–3440. [[CrossRef](#)]
41. Kovalev, K.; Astashkin, R.; Gushchin, I.; Orekhov, P.; Volkov, D.; Zinovev, E.; Marin, E.; Rulev, M.; Alekseev, A.; Royant, A.; et al. Molecular mechanism of light-driven sodium pumping. *Nat. Commun.* **2020**, *11*, 2137. [[CrossRef](#)]
42. Inoue, K.; Konno, M.; Abe-Yoshizumi, R.; Kandori, H. The Role of the NDQ Motif in Sodium-Pumping Rhodopsins. *Angew. Chem. Int. Ed.* **2015**, *54*, 11536–11539. [[CrossRef](#)]
43. Tomida, S.; Ito, S.; Inoue, K.; Kandori, H. Hydrogen-bonding network at the cytoplasmic region of a light-driven sodium pump rhodopsin KR2. *Biochim. Biophys. Acta Bioenerg.* **2018**, *1859*, 684–691. [[CrossRef](#)] [[PubMed](#)]
44. Abe-Yoshizumi, R.; Inoue, K.; Kato, H.E.; Nureki, O.; Kandori, H. Role of Asn112 in a Light-Driven Sodium Ion-Pumping Rhodopsin. *Biochemistry* **2016**, *55*, 5790–5797. [[CrossRef](#)]
45. Tomida, S.; Ito, S.; Mato, T.; Furutani, Y.; Inoue, K.; Kandori, H. Infrared spectroscopic analysis on structural changes around the protonated Schiff base upon retinal isomerization in light-driven sodium pump KR2. *Biochim. Biophys. Acta Bioenerg.* **2020**, *1861*, 148190. [[CrossRef](#)]
46. Studier, F.W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **2005**, *41*, 207–234. [[CrossRef](#)]
47. Landau, E.M.; Rosenbusch, J.P. Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 14532–14535. [[CrossRef](#)] [[PubMed](#)]
48. Gordeliy, V.I.; Labahn, J.; Moukhametzianov, R.; Efremov, R.; Granzin, J.; Schlesinger, R.; Büldt, G.; Savopol, T.; Scheidig, A.J.; Klare, J.P.; et al. Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* **2002**, *419*, 484–487. [[CrossRef](#)] [[PubMed](#)]
49. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 125–132. [[CrossRef](#)]
50. Tickle, I.J.; Flensburg, C.; Keller, P.; Paciorek, W.; Sharff, A.; Vonrhein, C.; Bricogne, G. *STARANISO*; Global Phasing Ltd.: Cambridge, UK, 2018.
51. Vagin, A.; Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 22–25. [[CrossRef](#)]
52. Emsley, P.; Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60*, 2126–2132. [[CrossRef](#)]
53. Murshudov, G.N.; Skubák, P.; Lebedev, A.A.; Pannu, N.S.; Steiner, R.A.; Nicholls, R.A.; Winn, M.D.; Long, F.; Vagin, A.A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355–367. [[CrossRef](#)] [[PubMed](#)]
54. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.W.; McCoy, A.; et al. Overview of the CCP 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 235–242. [[CrossRef](#)] [[PubMed](#)]
55. Shibata, M.; Inoue, K.; Ikeda, K.; Konno, M.; Singh, M.; Kataoka, C.; Abe-Yoshizumi, R.; Kandori, H.; Uchihashi, T. Oligomeric states of microbial rhodopsins determined by high-speed atomic force microscopy and circular dichroic spectroscopy. *Sci. Rep.* **2018**, *8*, 8262. [[CrossRef](#)] [[PubMed](#)]
56. Gushchin, I.; Shevchenko, V.; Polovinkin, V.; Borshchevskiy, V.; Buslaev, P.; Bamberg, E.; Gordeliy, V. Structure of the light-driven sodium pump KR2 and its implications for optogenetics. *FEBS J.* **2016**, *283*, 1232–1238. [[CrossRef](#)]

57. Ho, B.K.; Gruswitz, F. HOLLOW: Generating Accurate Representations of Channel and Interior Surfaces in Molecular Structures. *BMC Struct. Biol.* **2008**, *8*, 49. [[CrossRef](#)]
58. Harding, M.M.; Nowicki, M.W.; Walkinshaw, M.D. Metals in protein structures: A review of their principal features. *Crystallogr. Rev.* **2010**, *16*, 247–302. [[CrossRef](#)]
59. Lomize, M.A.; Pogozheva, I.D.; Joo, H.; Mosberg, H.I.; Lomize, A.L. OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **2012**, *40*, D370–D376. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

ARTICLE



<https://doi.org/10.1038/s41467-020-19457-7>

OPEN

# Viral rhodopsins 1 are an unique family of light-gated cation channels

Dmitrii Zabelskii  et al.<sup>#</sup>

Phytoplankton is the base of the marine food chain as well as oxygen and carbon cycles and thus plays a global role in climate and ecology. Nucleocytoplasmic Large DNA Viruses that infect phytoplankton organisms and regulate the phytoplankton dynamics encompass genes of rhodopsins of two distinct families. Here, we present a functional and structural characterization of two proteins of viral rhodopsin group 1, OLPVR1 and VirChR1. Functional analysis of VirChR1 shows that it is a highly selective, Na<sup>+</sup>/K<sup>+</sup>-conducting channel and, in contrast to known cation channelrhodopsins, it is impermeable to Ca<sup>2+</sup> ions. We show that, upon illumination, VirChR1 is able to drive neural firing. The 1.4 Å resolution structure of OLPVR1 reveals remarkable differences from the known channelrhodopsins and a unique ion-conducting pathway. Thus, viral rhodopsins 1 represent a unique, large group of light-gated channels (viral channelrhodopsins, VirChR1s). In nature, VirChR1s likely mediate phototaxis of algae enhancing the host anabolic processes to support virus reproduction, and therefore, might play a major role in global phytoplankton dynamics. Moreover, VirChR1s have unique potential for optogenetics as they lack possibly noxious Ca<sup>2+</sup> permeability.

<sup>#</sup>A list of authors and their affiliations appears at the end of the paper.

Microbial and animal rhodopsins (type-1 and 2 rhodopsins, respectively) comprise a superfamily of heptahelical (7-TM) transmembrane proteins covalently linked to a retinal chromophore<sup>1,2</sup>. Type-1 rhodopsins are the most abundant light-harvesting proteins that have diverse functions, such as ion pumping, ion channeling, sensory, and enzymatic activities<sup>3–11</sup>. The discovery, in 2000, of the light-driven pump proteorhodopsin (PR) in marine microbes triggered an extensive search of metagenomes for light-activated proteins<sup>12</sup>. As a result, about 10,000 rhodopsin genes have been identified in archaea, bacteria, unicellular eukaryotes, and viruses, although the biological functions of most of these proteins remain elusive. The study of microbial rhodopsins culminated in the discovery of channelrhodopsins<sup>13</sup> yielding the development of optogenetics, the revolutionary method for controlling cell behavior in vivo using light-gated channels and light-driven pumps<sup>13,14</sup>. Currently, major efforts are being undertaken to identify rhodopsins with novel functions and properties that could be harnessed to enhance optogenetic applications<sup>15–18</sup>.

In 2012, bioinformatic analysis of proteins encoded by nucleocytoplasmic large DNA viruses (NCLDV) resulted in the identification of highly-diverged PR homologs (hereafter, viral rhodopsins) in Organic Lake phycodnavirus and *Phaeocystis globosa* viruses, which belong to the extended Mimiviridae family<sup>19</sup>. Phylogenetic analysis shows that, within the rhodopsin superfamily, viral rhodopsins form a monophyletic family that consists of two distinct groups, VR1 and VR2<sup>20</sup>. Recently, a DTS-rhodopsin from the VR1 group (VirR<sub>DTS</sub>) has been reported to pump protons when expressed in *E. coli* plasma membrane<sup>21</sup>. Almost at the same time, the OLPVR1 protein from the VR2 group has also been shown to have a proton-pumping capacity, although forming unusual pentamers in lipid membrane<sup>22</sup>. The broad representation of a distinct family of rhodopsins in virus genomes implies an important light-dependent function in virus-host interactions, but the nature of this function remains uncertain. Given that NCLDV play a major role in marine algae population dynamics, elucidation of the virus-host interaction mechanisms would make an important contribution to a better understanding of the impact of viruses on global ecology and climate<sup>23,24</sup>.

Here we present the results of a comprehensive structure-function study of two homologous proteins (61% sequence similarity) from the VR1 group, OLPVR1 (ADX06642.1), and VirChR1 (TARA-146-SRF-0.22-3-C376786\_1). We show that unlike previously reported data<sup>21</sup>, viral rhodopsins of the VR1 group demonstrate sodium/potassium-selective channelrhodopsin activity when expressed in human neuroblastoma cells. Upon light absorption, VirChR1 depolarizes cell membranes to a level sufficient to drive neural firing. This finding, together with the fact that, in contrast to the previously characterized channelrhodopsins, VirChR1 is not permeable for calcium ions, suggested that viral rhodopsins of the VR1 group (VirChR1s) could become invaluable optogenetic tools for the remote control of Ca<sup>2+</sup>-dependent processes in the cell without Ca<sup>2+</sup>-induced noxious side effects. To verify this, we expressed VirChR1 in neurons and showed directly that the channel is able to elicit firing. Following functional characterization, we crystallized and solved multiple structures of OLPVR1 that revealed unique structure-function features of VirChR1s.

The recent hypothesis of proton pumping by VirR<sub>DTS</sub> (from the same phylogenetic VR1 group) was based on pH measurements in suspensions of *E. coli* cells expressing VirR<sub>DTS</sub> in their plasma membrane<sup>21</sup>. Taking into account that such experiments cannot definitively prove the absence of ion channel properties of rhodopsin, here we directly addressed the hypothesis that VirR<sub>DTS</sub> is a light-gated channel. We expressed VirR<sub>DTS</sub> in

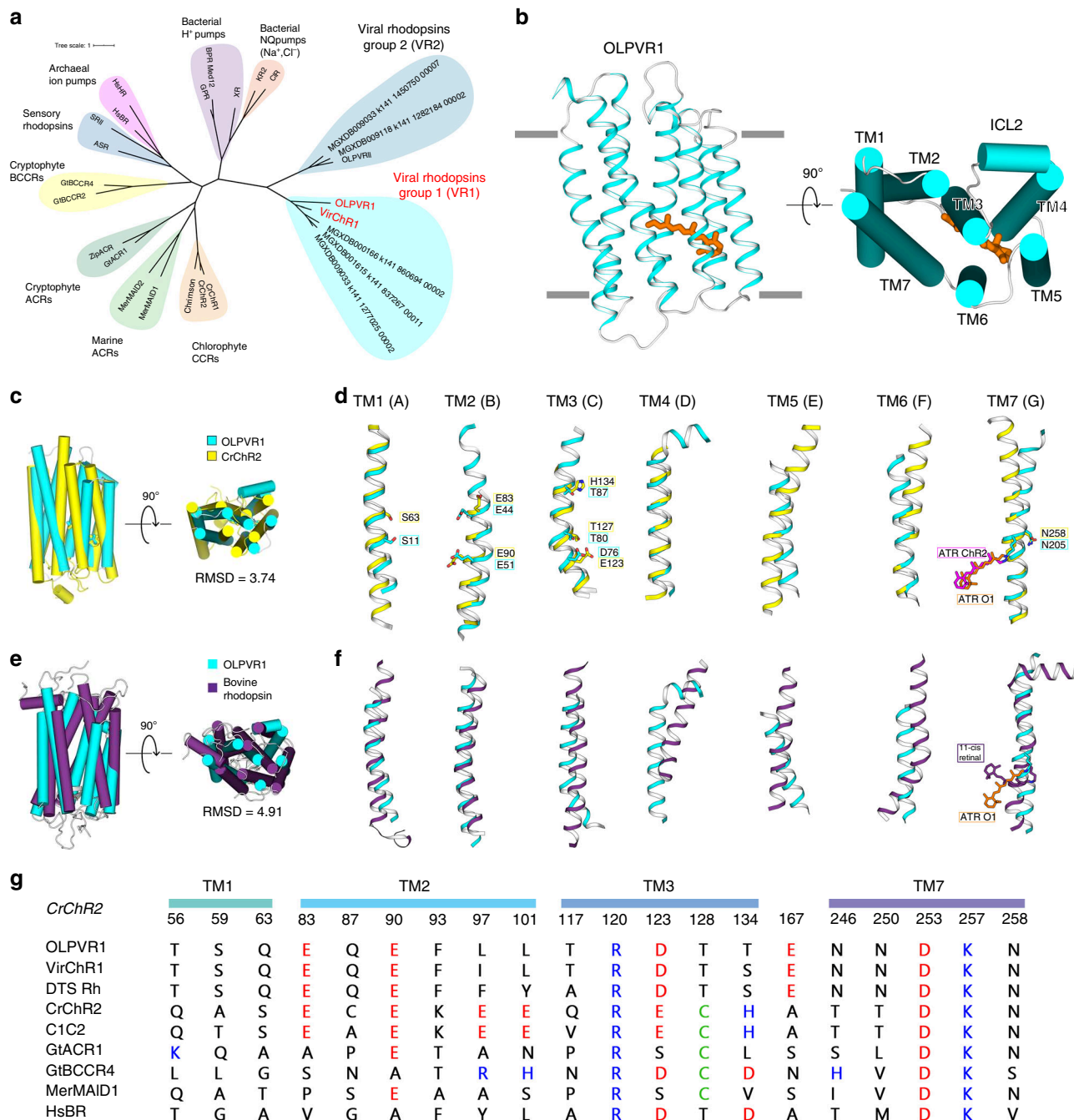
HEK293 cells, conducted voltage-clamp measurements, and demonstrated that VirR<sub>DTS</sub> is also a light-gated ion channel.

Thus, taking into account all the data, high homology of VirChR1s, high conservation of functionally key amino acids, we suggest that the VirChR1s (including VirR<sub>DTS</sub> from giant PgV virus infecting *Phaeocystis*, an important phytoplankton component) form a so far undescribed group of light-gated channels different from the known channelrhodopsins.

## Results

**Metagenomic search for viral rhodopsins genes and sequence analysis.** To obtain a comprehensive set of rhodopsins in the vast metagenomic sequence database produced by the *Tara Ocean* project, we compared 36 rhodopsin sequences representative of the previously identified major groups to the sequences of all open reading frames from *Tara Ocean* contigs. This search retrieved 2584 Type 1 rhodopsins, of which 385 belonged to VR1 and 172 belonged to VR2 as confirmed by phylogenetic analysis that also supported the monophyly of viral rhodopsins family (Fig. 1a and Supplementary Fig. 1). Consistent with the monophyly of viral rhodopsins and the separation of the VR1 and VR2 groups, the examination of sequence alignments detected several amino acid motifs that partially differed between the two groups. The amino acid triad implicated in proton exchange with the retinal Schiff base (residues 85, 89, and 96 in the reference bacteriorhodopsin<sup>25,26</sup>) had the form DTS/DTT in the VR1 group and DTT/DSV in the VR2 group. The members of the VR1 group are characterized by several fully conserved residues, such as S11, Q15, E51, K192, N193, N197, and N205 (annotated with OLPVR1 numbering, Supplementary Fig. 2) that are mainly located in proximity to the retinal Schiff base. Despite the overall low structural similarity with chlorophyte cation-conducting channelrhodopsins (Fig. 1d), VR1 rhodopsins retain the two highly conservative glutamates in TM2 (E44 and E51 in OLPVR1 corresponding to E83 and E90 in CrChR2) that have been shown to be critical for CrChR2 ion channelling<sup>27,28</sup>. In addition, as it will be shown later, the VR1 group has a signature topological feature, namely, an extended TM4 helix that consists of its transmembrane (TM4) and membrane-associated parts (ICL2) and has not been previously observed in characterized microbial rhodopsins (Fig. 1e and Supplementary Fig. 3).

**Spectroscopic characterization of VirChR1s.** To characterize photochemical characteristics of viral channelrhodopsins, we expressed C-terminally his-tagged OLPVR1 and VirChR1 proteins in *E. coli* and purified them using combination of Ni-NTA and size-exclusion chromatography methods (see “Methods” for details). VirChR1 protein was additionally supplemented with BRIL protein on the N-terminus of the protein, to improve expression level of the protein<sup>29</sup>. Both OLPVR1 and VirChR1 show absorption spectra sensitive to blue light with  $\lambda_{\max}$  of 500 nm and 507 nm, respectively at pH 7.5 (Fig. 2a), which is consistent with the fact that blue light penetrates deep throughout the seawater photic zone<sup>30</sup>. Similar to VirR<sub>DTS</sub> rhodopsin<sup>21</sup>, HsBR<sup>31</sup> and proteorhodopsins<sup>32</sup>, OLPVR1 and VirChR1 undergo a ~30 nm spectral red-shift under acidic conditions, associated with protonation of retinal chromophore counterion (Fig. 2b, c). The Schiff base region of VirChR1s is reminiscent of those in light-driven proton pumps, such as HsBR, suggesting that D76 in OLPVR1 (D80 in VirChR1) acts as counterion, as in HsBR (Fig. 2d). In order to estimate the pKa values we plotted the absorption maximum values against buffer pH and fitted the data by the Henderson-Hasselbalch equation with a single pKa (Fig. 2e). The resulting pKa values for OLPVR1 (pKa = 4.8) and VirChR1 (pKa = 3.5) are in a good agreement with pKa = 3.6,



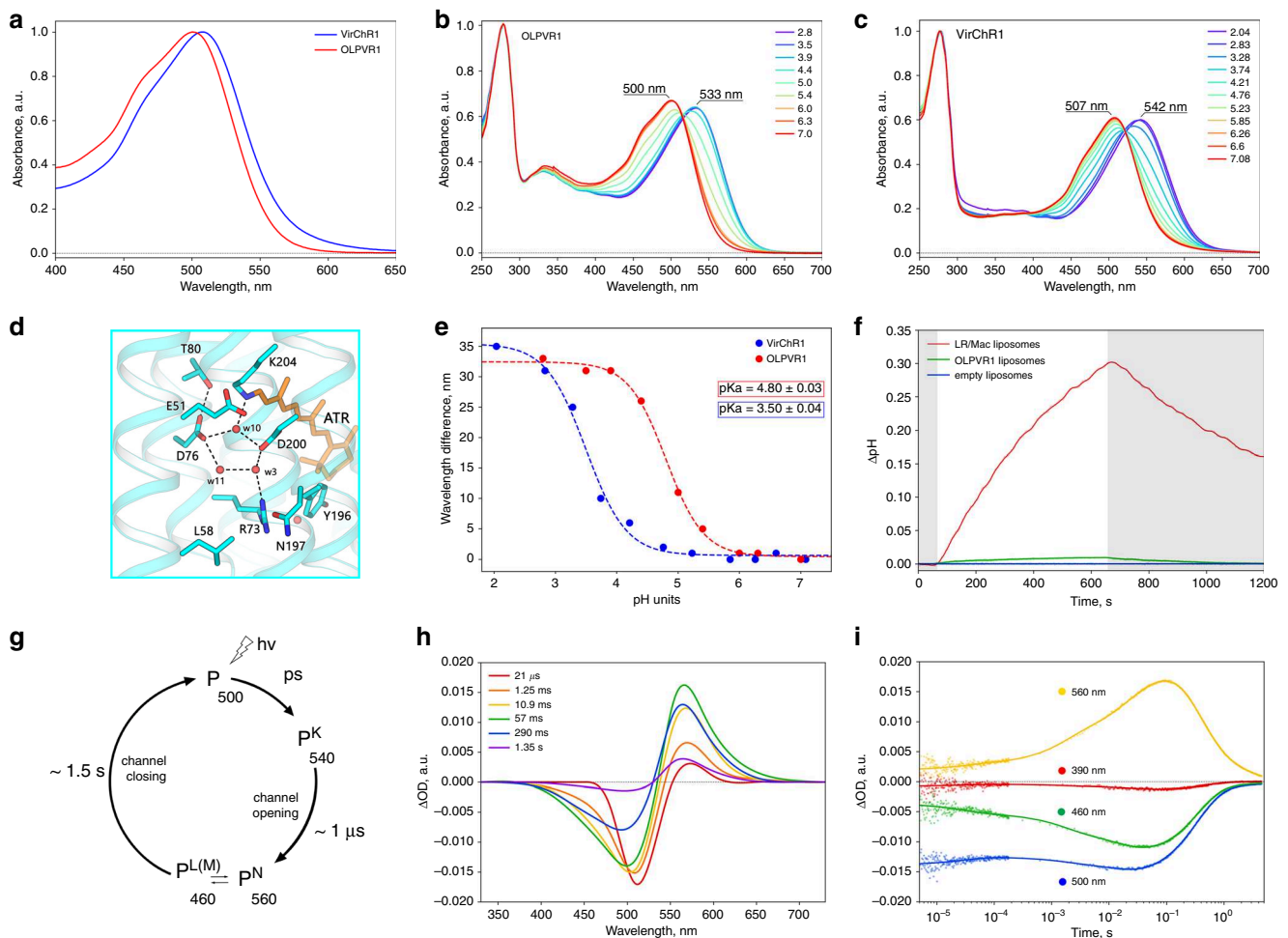
**Fig. 1** Phylogenetic and structural overview of the viral channelrhodopsin group. **a** Unrooted phylogenetic tree of the channelrhodopsin superfamily, including viral channelrhodopsin representatives. Scale bar indicates the average number of amino acid substitutions per site. CCR, cation-conducting channelrhodopsin, ACR, anion-conducting channelrhodopsin. OLPVR1 and VirChR1 proteins are additionally indicated in red. Rhodopsins were named according to the previous works<sup>18,101,102</sup>. **b** Crystal structure of OLPVR1 protein, viewed parallel to membrane (left), and from the extracellular side (right). All-trans retinal (ATR) is depicted with orange sticks. The hydrophobic membrane boundaries were calculated with the PPM server and are shown by gray lines<sup>103</sup>. **c** Structure alignment of OLPVR1 and CrChR2 (PDB ID: 6EID<sup>104</sup>) structures colored cyan and yellow, respectively. RMSD, root mean square deviation of atomic positions. **d** Individual TM helices are shown after the superimposition of the OLPVR1 and CrChR2 rhodopsins. **e** Structure alignment of OLPVR1 and bovine rhodopsin (PDB ID: 1U19<sup>43</sup>) structures colored cyan and purple, respectively. **f** Individual TM helices are shown after the superimposition of the OLPVR1 and bovine rhodopsin. **g** Alignments of functionally important residues of transmembrane helices 1, 2, 3, and 7 of representative proteins from channelrhodopsin subfamilies. Positively and negatively charged residues are highlighted blue and red; cysteine residues are highlighted green.

previously reported for VirR<sub>DTS</sub> rhodopsin<sup>21</sup>. The one-unit difference between OLPVR1 and VirChR1 pKa values might be possibly explained by difference in relative positions of the TM1–3 and TM7 helices and difference in neighboring to

counterion residues, such as I50 and L79 in OLPVR1, which are replaced with V53 and I83 in VirChR1 (Supplementary Fig. 2).

To elucidate photocycle kinetics of viral channelrhodopsins, we performed transient absorption measurements with OLPVR1-





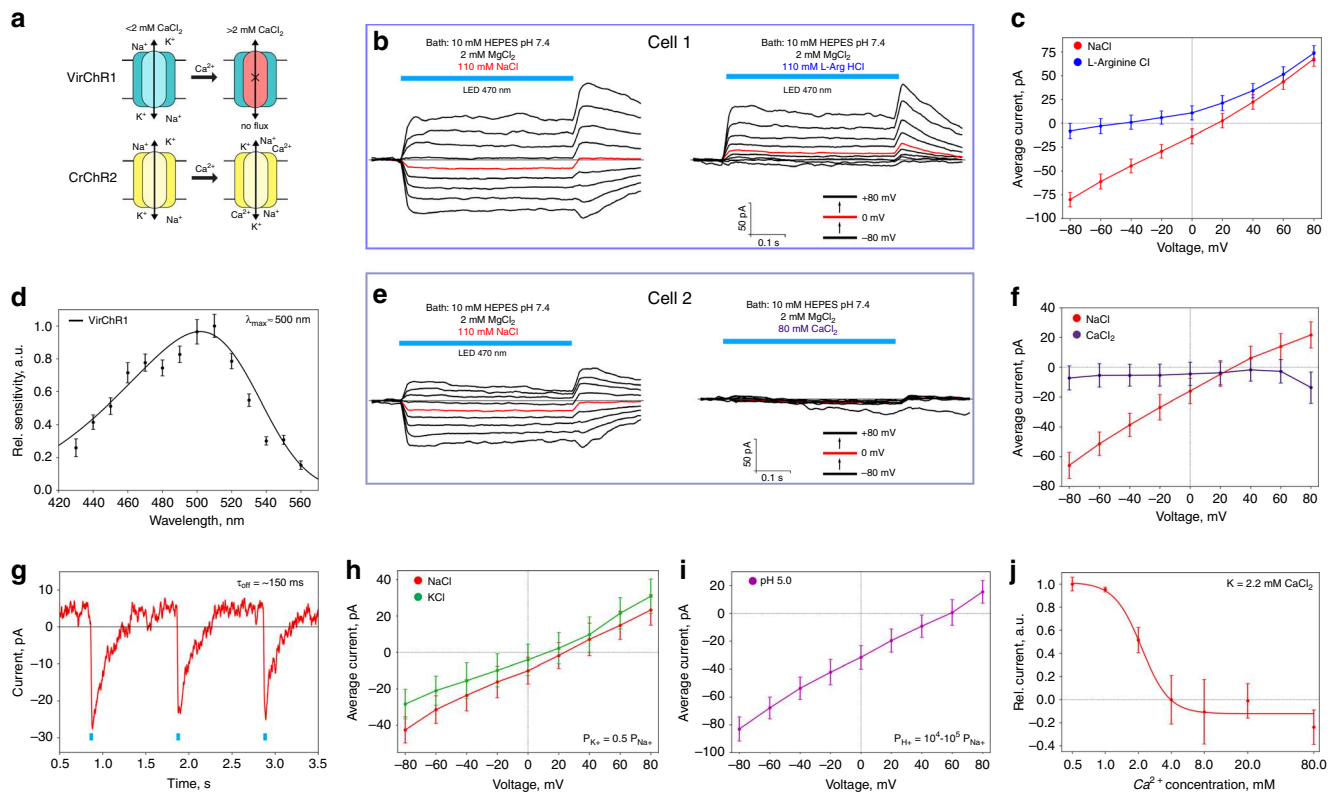
**Fig. 2 Spectroscopic characterization of viral channelrhodopsins.** **a** Normalized absorption spectra of OLPVR1 and VirChR1 proteins at neutral pH (pH 7.5). **b, c** Absorption spectra of OLPVR1 and VirChR1 at acidic (pH 2.1–7.1) pH range, normalized for absorption at 280 nm. **d** Schiff base region of OLPVR1 protein, key residues and water molecules are shown as sticks and spheres; hydrogen bonds are shown as dashed lines. **e** Absorption spectra of OLPVR1 at acidic (pH 2.8–7.0) pH range, normalized for absorption at 280 nm. **e** Red shift of UV-visible absorption spectrum and protonation of counterion of OLPVR1 and VirChR1. Wavelength maximum values are shown as circles. Sigmoidal curve fits are presented as dashed lines. The  $pK_a$  values were calculated using a sigmoidal fit. **f** Ion-transport activity assay of OLPVR1-containing proteoliposomes in 100 mM NaCl salt. The onset of illumination is indicated with white (light on) and gray (light off) background color, pH was adjusted to pH 6.0 prior to measurements. LR/Mac-containing liposomes and empty liposomes were used as positive and negative controls, respectively. **g** Schematic model of viral rhodopsins photocycle. **h** Transient absorption spectra and **i** time traces at specific wavelengths of wild type OLPVR1 protein at pH 7.5.

containing nanodiscs that revealed three distinct intermediate states of OLPVR1 photocycle (Fig. 2g). An early decaying K-like state ( $\lambda_{\max} = 540$  nm), followed by major accumulation of L-like ( $\lambda_{\max} = 460$  nm) and N-like ( $\lambda_{\max} = 560$  nm) states that live for about 1.5 s (Fig. 2h, i). Unlike other channelrhodopsins, OLPVR1 lacks a detectable M-state that is generally associated with the ion-conducting state of the protein (Fig. 2c, red curve). At the same time, VirR<sub>DTS</sub> also forms similar intermediates and lacks the M-like state<sup>21</sup>. Although the equilibrium between the L-like and N-like states is the major candidate for the ion-conducting state in VirChR1s, further investigations are required for its identification.

**Functional analysis of OLPVR1.** To investigate the functional properties of the VR1 group, we performed the measurements of the pH changes in the suspension of proteoliposomes containing the viral rhodopsin, upon continuous light illumination. This method allows determination of the pumping activity and is often used for the characterization of microbial rhodopsins<sup>7,33–35</sup>. We did not observe any significant ion-translocation ability of the

viral rhodopsin OLPVR1 in pH change experiments with proteoliposomes (Fig. 2g). Under continuous bright light illumination, OLPVR1-containing liposomes did not show any substantial pH change of the external solvent (Fig. 2f). The maximum pH shift of the proteoliposomes suspension containing OLPVR1 (0.03 pH units), is about 10 times less than of that containing the eukaryotic light-driven proton pumping rhodopsin from *Leptospira maculans* (LR/Mac<sup>36</sup>) (Fig. 2f). Taking into account the known fact that microbial rhodopsins in liposomes are usually oriented in the opposite direction to that in the cell membrane<sup>7,33,34</sup>, we conclude that OLPVR1 possesses a weak outward proton pumping activity. The outward proton pumping has also been shown for another member of the VR1 group, VirR<sub>DTS</sub><sup>21</sup>. However, in the case of OLPVR1, the pumping activity was much lower than that of VirR<sub>DTS</sub>. Importantly, although the pH recordings upon light illumination using a reduced system, such as proteoliposomes suspension, are suitable for the validation of the light-driven ion pumps (bacteriorhodopsin, proteorhodopsin, halorhodopsin and bacterial ion pumps), they cannot demonstrate the ion-channeling activity of





**Fig. 3 Ion selectivity and physiological features of VirChR1.** **a** Schematic comparison of VirChR1 and CrChR2 ion channeling activity under different calcium concentrations, membrane boundaries are shown schematically as black horizontal lines. **b** Voltage-clamp records from  $n = 1$  representative SH-SY5Y cell, expressing VirChR1 (with) 10 mM HEPES pH 7.4, 110 mM NaCl, 2 mM MgCl<sub>2</sub> and (right) 110 mM L-arginine hydrochloride replacing NaCl in bath. Pipette solution during experiments was: 10 mM HEPES pH 7.4, 110 mM NaCl, 2 mM MgCl<sub>2</sub>, 10 mM EGTA, illumination by LED (470 nm) lamp is indicated with light blue color. **c** Current-voltage dependences for  $n = 1$  representative SH-SY5Y cell in 110 mM NaCl (red) and 110 mM L-arginine hydrochloride (blue). Currents are reproducible and typical to those in  $n = 9$  experiments with other cells (and  $n = 21$  experiments under slightly different NaCl concentrations varied from 110 mM to 140 mM). **d** Action spectrum of VirChR1 measured using equal photon fluxes (Sample size,  $n = 18$ –20). **e** Voltage-clamp records from  $n = 1$  representative SH-SY5Y cell expressing VirChR1 in bath solution (left) 10 mM HEPES pH 7.4, 110 mM NaCl, 2 mM MgCl<sub>2</sub> and (right) in 80 mM CaCl<sub>2</sub> replacing NaCl in bath solution. **f** Current-voltage dependences for  $n = 1$  representative SH-SY5Y cell in 110 mM NaCl (red) and 80 mM CaCl<sub>2</sub> (indigo) solutions. **g** Excitation recovery of photocurrent after a short pulse of nanosecond laser (500 nm) activation. Tau-off was measured in  $n = 5$  independent experiments. Current-voltage dependences for  $n = 1$  cell for different bath/pipette solution. Traces are shown for **h** bath solutions: 110 mM NaCl (red) and 110 mM KCl (green) (pipette solution is standard) and **i** pipette solution 110 mM L-arginine hydrochloride salt solution of pH 5.0 (bath solution is standard). Estimation of relative conductivities for different ions was done by fitting traces with Goldman-equation. **j** Current dependence on calcium concentration in bath solution measured at +80 mV (inflection point is at -2.2 mM of calcium). For all electrophysiological recordings at  $n = 1$  cell currents were reproducible in  $n = 3$ –10 independent experiments with other cells. No current averaging between cells was done, since different cells have significantly different protein expression levels. Data are presented as mean values  $\pm$  SEM of current value under illumination in the cell measured.

the protein. To test the possible light-gated ion channeling by the VR1 group, we conducted electrophysiological studies of these proteins.

**Electrophysiology of VirChR1, a light-gated cation channel.** In order to identify the possible ion-conducting activity, we expressed human codon-optimized OLPVR1 and VirChR1 genes in SH-SY5Y human neuroblastoma cell line in the presence of all-*trans* retinal. Despite both proteins expressed well, they showed strong retention in the cytosol according to the fluorescence microscopy and electrophysiology data. To improve membrane trafficking and localization, we supplemented the proteins with C-terminal p2A self-cleavage peptide prior to fluorescent tag (see Methods for full details). This modification helped with VirChR1 localization and enabled us to analyze its photocurrents, however, OLPVR1 did not show significant improvements with this approach. Therefore, we characterized the ion-channeling activity of VirChR1 as a representative of the VR1 group 1.

Given the high sequence similarity of OLPVR1 and VirChR1 (Fig. 1b and Supplementary Fig. 2), we hereafter refer to the function of viral channelrhodopsins based on the data obtained for VirChR1. Whole-cell patch-clamp recordings revealed photocurrents generated by VirChR1 (Fig. 3b). Photocurrents were observed in a bath solution of 10 mM HEPES pH 7.4, 110 mM NaCl, 2 mM MgCl<sub>2</sub> and a pipette solution of 10 mM HEPES pH 7.4, 110 mM NaCl, 2 mM MgCl<sub>2</sub>, 10 mM EGTA (hereafter both denoted standard). In response to continuous illumination by LED light ( $\lambda_{\text{max}} = 470$  nm). Measuring photocurrents in different cells under standard conditions, we did not detect changes in kinetics or shifts of the reversal potential. For one representative neuroblastoma cell, photocurrent stabilized at 50 pA at 80 mV (Fig. 3b). The photocurrents varied in amplitude for different cells depending on the size of the cell and protein expression level, but the overall pattern remained the same. The photocurrent density in the neuroblastoma cells was  $1.4 \pm 0.2$  pA/pF (mean  $\pm$  std.dev,  $n = 9$ ), which is comparably low compared to

the known channelrhodopsins, which might be due to a lower single channel conductance or due to less efficient expression of active molecules in the plasma membrane. The photocurrents showed a reversal potential,  $U_{rev}$  of  $25 \pm 6$  mV (mean  $\pm$  std.dev,  $n = 9$ ), indicating that light triggers a passive ion conductance by VirChR1. VirChR1 exhibits an action spectrum similar to those of typical rhodopsins, with the maximum sensitivity observed close to 500 nm (Fig. 3d). Tau-off for VirChR1  $\tau_{off} = 155 \pm 5$  ms (mean  $\pm$  std.dev.,  $n = 5$ ) was directly determined using single-exponential fit of photocurrent recovery (Fig. 3g). Next, we performed ion substitution experiments to discriminate between possible cation and anion conductivity of the VirChR1 channel. First, we replaced the standard bath solution with 63 mM  $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$  pH 7.4 and 2 mM  $\text{MgCl}_2$ , and found that this modification changed neither the magnitude nor the reversal potential of the photocurrent. By contrast, when replacing 110 mM NaCl in bath solution with 110 mM L-arginine hydrochloride, the inward current became immeasurably low resulting in a dramatic change in the current–voltage dependency (Fig. 3c). These results indicate that viral channelrhodopsins possess only cation-conducting activity.

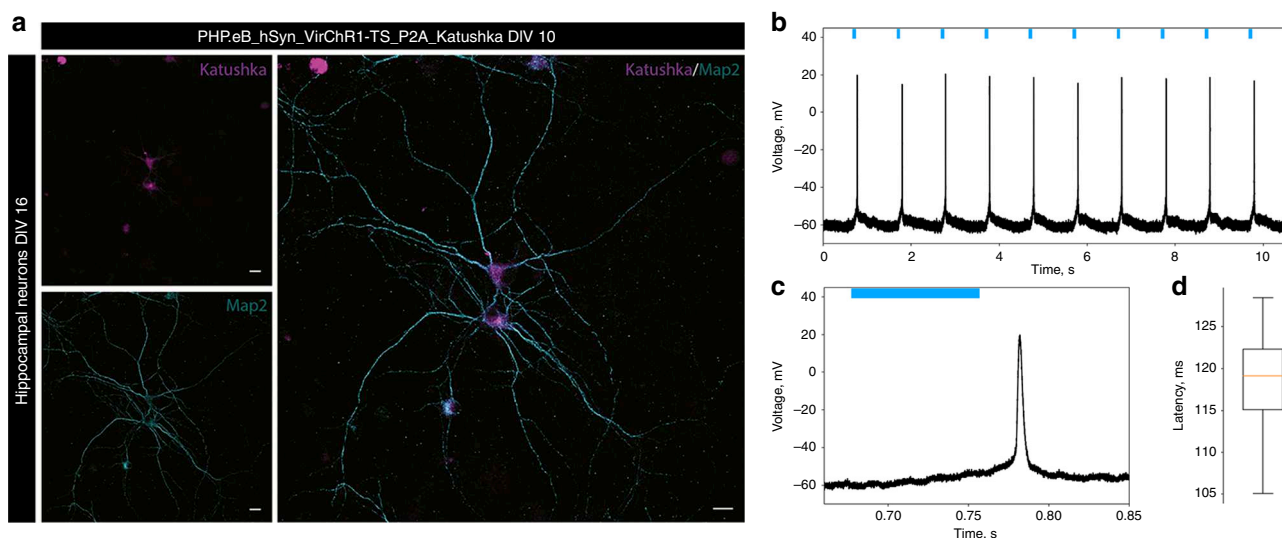
**Unusual  $\text{Ca}^{2+}$  sensitivity of VirChR1.** To evaluate the conductance of different cations by viral channelrhodopsins, we measured the voltage dependence of the photocurrent in different bath solutions. Replacing  $\text{Na}^+$  with  $\text{K}^+$  cations (Fig. 3h) in the bath solution yields an estimate of potassium permeability,  $P_{\text{K}^+} \approx 0.5 \cdot P_{\text{Na}^+}$ . To estimate  $\text{H}^+$  permeability, we replaced the standard bath solution with 10 mM citric acid pH 5.0, 110 mM L-arginine hydrochloride, 2 mM  $\text{MgCl}_2$  (Fig. 3i). Under these conditions, we observed full suppression of the photocurrent, which occurred, presumably, due to the protonation of the Schiff base proton acceptor (Fig. 2c). Therefore, instead of changing the standard bath solution, we replaced the pipette solution with 10 mM Citric Acid pH 5.0, 110 mM L-arginine hydrochloride, 2 mM  $\text{MgCl}_2$ , 10 mM EGTA, which preserved the photocurrent at a measurable level. Fitting the photocurrent data with the Goldman-Hodgkin-Katz equation estimated the  $\text{H}^+$  permeability,  $P_{\text{H}^+} \approx 10^4\text{--}10^5 \cdot P_{\text{Na}^+}$ . Overall, VirChR1 shows ion selectivity comparable to those of CrChR2, namely,  $P_{\text{K}^+} \approx 0.5 \cdot P_{\text{Na}^+}$  and  $P_{\text{H}^+} \approx 10^6 \cdot P_{\text{Na}^+}$ . Thus, group 1 of viral rhodopsins and chlorophyte cation channels are nearly similar with respect to the conductivity of monovalent ions.

Next, we tested whether VirChR1 was permeable for divalent cations, such as  $\text{Ca}^{2+}$ , similarly to CrChR2<sup>13</sup>. Strikingly, replacement of 110 mM NaCl in bath solution for 80 mM  $\text{CaCl}_2$  completely abolished the photocurrent (Fig. 3e). The current–voltage dependences of VirChR1 in the presence and absence of  $\text{Ca}^{2+}$  indicate that VirChR1 is completely impermeable for  $\text{Ca}^{2+}$  cations (Fig. 3f), in contrast to the high  $\text{Ca}^{2+}$  conductivity of CrChR2 (Supplementary Fig. 4). Importantly, VirChR1 is fully blocked for both inward and outward ionic fluxes when the concentration of  $\text{Ca}^{2+}$  exceeds a certain threshold. To characterize the phenomena of VirChR1 inhibition by  $\text{Ca}^{2+}$  ions, we measured the dependence of the average photocurrent at +80 mV voltage on the  $\text{CaCl}_2$  concentration (Fig. 3j). We observed a sigmoid-like dependence with an inflection point at  $K_{\text{Ca}^{2+}} = 2.2$  mM  $\text{CaCl}_2$ , which is surprisingly close to the  $\text{Ca}^{2+}$  concentration in the world ocean<sup>37</sup>, and thus, suggestive of a functional role of viral rhodopsin inhibition by  $\text{Ca}^{2+}$  ions. Taken together, our findings suggest that VirChR1 is a light-gated cation channel that conducts exclusively monovalent ions ( $\text{H}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ) and is completely inhibited by divalent ions ( $\text{Ca}^{2+}$ ) (Fig. 3a), with characteristic enzyme-substrate kinetics.

**VirChR1s are able to drive neural activity.** In order to verify the potential of viral channelrhodopsins for optogenetic stimulation, we transduced rat hippocampal neurons with adenoassociated virus (AAV-PHP.eB) carrying the VirChR1 gene under the control of the human synapsin I promoter. We used VirChR1 gene C-terminally fused to the Kir2.1 membrane trafficking signal, followed by a p2A self-cleavage peptide and Katushka fluorescent protein (see “Methods” for details). First, the experiments showed that VirChR1 protein was robustly expressed in hippocampal neurons (Fig. 4a). We found that VirChR1 with the N-terminal HA-FLAG tag, which we had used in the above experiments, caused major neuronal death, which made them impossible to measure with patch-clamp. In contrast, VirChR1 without the HA-FLAG tag expressed well and neurons were still viable enough for electrophysiological measurements. Second, patch-clamp experiments demonstrated firing in VirChR1-expressing neurons upon optical stimulation at 1 Hz (Fig. 4b). Due to the low photocurrent densities (in hippocampal neurons  $0.5 \pm 0.2$  pA/pF measured at  $-75$  mV  $n = 11$ , compared to  $2.3 \pm 0.5$  pA/pF for CrChR2 and  $8.0 \pm 0.4$  pA/pF for CatCh<sup>38</sup> (mean  $\pm$  std.dev)) spikes were elicited with long latencies (Fig. 4c, d, latency of action potential peak is  $119 \pm 6$  ms (mean  $\pm$  std.dev) for this particular neuron). Optogenetic stimulation at higher rates failed to elicit spikes (Supplementary Fig. 5). However, neurons with higher photocurrents showed shorter spike latencies ( $50 \pm 10$  ms, Supplementary Fig. 6). Furthermore, we observed action potentials arising after turning off the light, potentially due to the slow closing kinetics of the rhodopsin (Figs. 3g, 4c), which also led to double-spiking in other neurons (Supplementary Fig. 6). In conclusion, these results show that virus-mediated expression VirChR1s enables light-driven firing in neurons.

**Crystal structure of the viral rhodopsin OLPVR1.** To decipher the molecular mechanism of ion channeling, the structure of viral channelrhodopsin from group 1 is essential. One crystal structure of the VirR<sub>DTS</sub> (PDB ID: 6J00<sup>21</sup>) was recently reported<sup>21</sup>. However, from the only available structure, it is not possible to distinguish the features of the entire group. Moreover, the ion-channeling activity of VirR<sub>DTS</sub> was not demonstrated in the original work<sup>21</sup>; thus structural insights of the channel of VirChR1s were not described. Here, we present a high-resolution structure of another VR1 protein, OLPVR1, at 1.4 Å resolution. It was crystallized with an *in meso* approach<sup>39</sup> similar to that used in our previous studies<sup>3</sup>. We obtained three different types of crystals. Type A rhombic crystals were grown at pH 8.0 using the monopalmitolein (MP) host lipid matrix and have the P2<sub>1</sub>2<sub>1</sub>2 space group with one protein molecule in the asymmetric unit. Type B hexagonal crystals were grown at pH 7.0 using a monoolein (MO) lipid matrix, have the P1 space group, and contain two protein molecules in the asymmetric unit. Type A and type B crystals resulted in the obtaining of the OLPVR1 structures at the resolution of 1.4 and 1.6 Å, respectively (Supplementary Table 1). OLPVR1 molecules are nearly identical in both structures (root mean square deviation (RMSD) less than 0.2 Å), so hereafter, we refer to the structure from type A crystals as it has the highest resolution. The crystal packing and examples of the electron density maps are shown in Supplementary Figs. 7 and 8.

The structure of the OLPVR1 protomer is composed of 7 transmembrane helices (TM1–7), connected by three intracellular and three extracellular loops (Figs. 1c, 5a). The OLPVR1 protein (residues 2–223), all-*trans* retinal (ATR) covalently bound to K204 (K257 in CrChR2<sup>8</sup>), 9 lipid molecules, and 107 water molecules are clearly resolved in the electron density maps (Supplementary Fig. 8). Despite the fact that only one OLPVR1



**Fig. 4 Neuronal application of VirChR1 for optogenetic activation.** **a** AAV transduction of primary hippocampal neurons at DIV10, intrinsic fluorescence of Katushka (magenta), and neuronal marker Map2 (cyan). Scale bars 10  $\mu\text{m}$ . VirChr1 enables light-driven neuron spiking. Data presented in this figure refers to one representative neuron. The results showed in the microphotograph reproduced in  $n = 6$  independent experiments. **b** Voltage trace showing depolarization and spikes of the neuron in response to the 1 Hz light pulse series, with 80 ms light pulses (green bars). **c** Expanded single spike induced by VirChR1 photoresponse. **d** Latencies distributions, when the neuron is illuminated with 80 ms light pulses. The box extends from lower quartile to upper quartile. The whiskers show the range of the latencies, green line is median latency, sample size,  $n = 10$ .

protomer is present in the asymmetric unit, the crystal packing of the protein shows that OLPVR1 could be organized into homodimers, similar to those of CrChR2<sup>8,40</sup>. These dimers might reflect the oligomeric state of the viral channelrhodopsin in the cell membrane (Supplementary Fig. 9). The interfacial interaction in the putative dimer occurs mainly in the TM4 helix and involves several non-conservative residues, namely E108-E108', Y111-Y111', F122-F122' with low surface interaction area (429  $\text{\AA}^2$ ) (Supplementary Fig. 9). An alternative configuration of OLPVR1 dimer predicted with GalaxyHomomer server<sup>41</sup> showed stronger interfacial interaction (1861  $\text{\AA}^2$ ). Therefore, the orientation of the OLPVR1 protomers in solution is unclear at the moment (Supplementary Fig. 9). Overall, the OLPVR1 backbone is tolerantly superimposed with that of the *Med12* proteorhodopsin (PDB ID: 4JQ6<sup>42</sup>) with RMSD of 2.1  $\text{\AA}$ , whereas the alignments with the *HsBR* (PDB ID: 1C3W<sup>26</sup>) and *CrChR2* (PDB ID: 6EID<sup>8</sup>) structures gives RMSD values of 4.3 and 3.7  $\text{\AA}$ , respectively (Fig. 1d, e and Supplementary Fig. 10). Interestingly, unlike other microbial rhodopsins, OLPVR1 architecture closely resembles the architecture of G-protein coupled receptors with TM3 helix protruding to the center of the protein. In particular, OLPVR1 aligns with bovine rhodopsin (PDB ID: 1U19<sup>43</sup>) with RMSD of 4.91  $\text{\AA}$  (Fig. 1f, g), with a high similarity among helices forming ion-conducting pathway (TM1–3 and TM7).

The OLPVR1 protomer has short extracellular loops, which sharply differentiates it from other channelrhodopsins that typically have large N- and C-terminal domains. Unlike in other microbial rhodopsins, helices TM3 and TM4 of OLPVR1 are connected by the loop containing the membrane-associated helix (ICL2 helix), which is composed of hydrophilic residues (Fig. 1c). Strikingly different from other rhodopsins, the intracellular parts of TM6 and TM7 helices of OLPVR1 are significantly moved apart from each other far enough to form a pore (Supplementary Fig. 11). The pore is located at about 8  $\text{\AA}$  from the cytoplasmic side of the lipid membrane border (Supplementary Fig. 11) and connects the inside of the protein with the groove on its surface, which leads further to the intracellular bulk. Surprisingly, in our structure, the groove and a part of the pore are occupied with a fragment of the lipid molecule, a host lipid of the crystallization

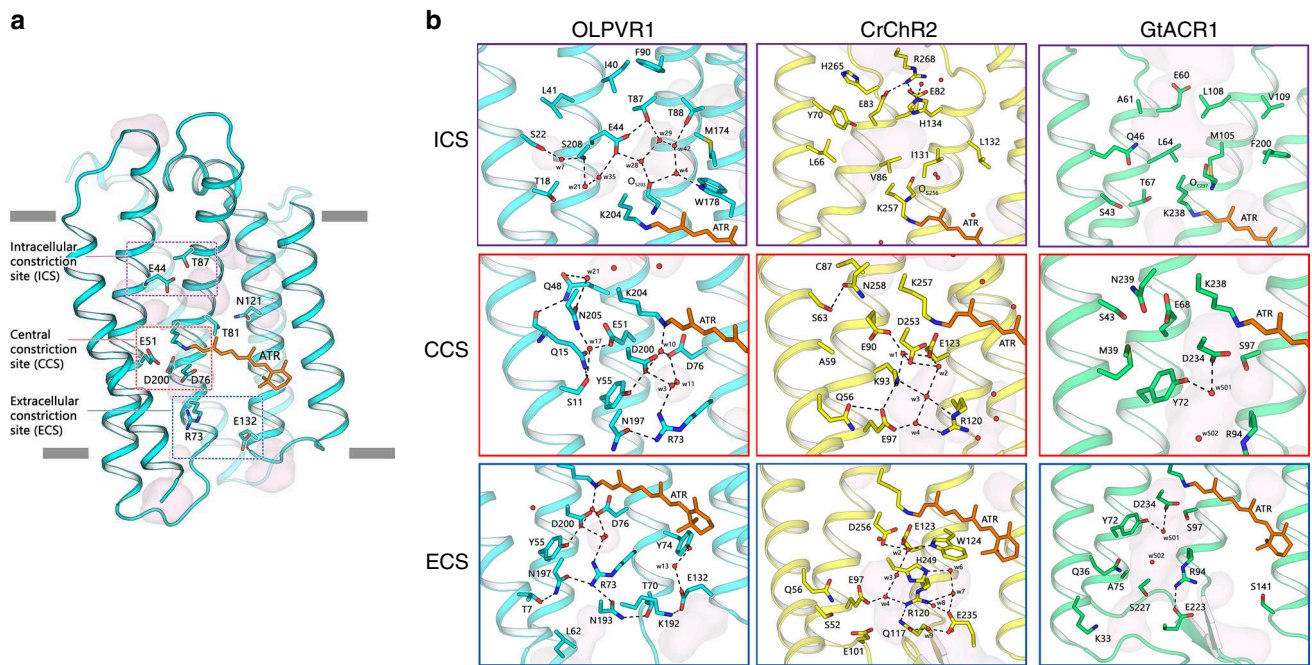
matrix (MP and MO in case of 1.4 and 1.6  $\text{\AA}$  structures, respectively), which is likely to be a crystallization artifact and is discussed in more details in the Supplementary Notes (Supplementary Fig. 11).

#### Structure of the retinal binding pocket and Schiff base region.

The retinal cofactor is covalently attached to the conserved K204 residue in OLPVR1.  $2F_o - F_c$  electron density maps at 1.4  $\text{\AA}$  reveal two alternative conformations of the retinal in the region of the  $\beta$ -ionone ring. At the same time, the configuration near the Schiff base in both of them is all-*trans* 15-*anti* (Supplementary Fig. 8). The retinal Schiff base (RSB) region of OLPVR1 is very similar to that of *HsBR* (Supplementary Fig. 12c). Notably, the D76 and D200 side chains and water molecules w3, w10 and w11 in OLPVR1 (corresponding to D85, D212, w401, w402, and w406 in *HsBR*, respectively) form an almost identical to *HsBR* pentagon-like hydrogen bonds structure, which is important for proton transfer in light-driven proton pumps<sup>44</sup>. Moreover, the pentagon is similarly stabilized by T80 and R73 (T89 and R82 in *HsBR*, respectively). In contrast, the pair of RSB counterions in *CrChR2* is composed of E123 and D253, which, together with water molecules positioned in this region, presumably results in notably different stabilization of the RSB as compared to OLPVR1 (Supplementary Fig. 12c). The walls of the retinal pocket of OLPVR1 around the polyene chain are composed of several aromatic amino acids similar to those in *HsBR*, namely W77, W178, and Y181 (Supplementary Fig. 12a). However, there are several important changes near the  $\beta$ -ionone ring, particularly L142, G182, and F185 instead of W138, P186, and W189 in *HsBR*. These amino acids can potentially be candidates for mutation scanning to obtain red-shifted versions of VirChR1s, more suitable for practical applications, considering that OLPVR1 and *HsBR* have retinal absorption maxima at 500 nm and 560 nm, respectively.

A characteristic feature of the retinal binding pocket of OLPVR1 (and, presumably, all viral channelrhodopsins) is the presence of highly-conserved and directly hydrogen-bonded residues T81 and N121 at the positions of the corresponding





**Fig. 5 Organization of ion pathway constriction sites (CSs) in OLPVR1.** **a** Three CSs and cavities are forming the putative ion-conductive pathway in viral CCRs and highly conservative residues of OLPVR1. TM6 and TM7 helices are omitted for clarity. Membrane core boundaries were calculated using PPM server<sup>103</sup> and are shown with gray lines. **b** Magnified view of the CSs in OLPVR1 (left, present work), CrChR2 (middle, PDB ID: 6EID<sup>8</sup>) and (right, PDB ID: 6CSM<sup>10</sup>) structures, colored cyan, yellow and green, respectively. Water accessible cavities were calculated using HOLLOW<sup>105</sup> and are presented as a pink surface.

residues T80 and D115 in *HsBR*, and C128 and D156 in *CrChR2* (Supplementary Fig. 12b). These pairs connect the middle parts of TM3 and TM4. Importantly, in the case of *CrChR2*, where C128 and D156 are interconnected by hydrogen bonds via a water molecule, the alteration of the pair dramatically affects the kinetics of the protein, and it was suggested that the pair is involved in the RSB reprotonation during photocycle<sup>8,45,46</sup>.

#### Organization of the OLPVR1 ion-conducting pathway.

Detailed analysis of the amino acid conservation in the VR1 group (Fig. 6a, b) shows that the majority of the conserved residues forms the interior of the protein differ from VR2 and PR groups (Fig. 6c) and are predicted to contribute to ion-channeling of VirChR1s. The structure suggests that the ion-conducting pathway of OLPVR1 is formed by TM1–3 and TM7 helices and is lined with several water-accessible cavities (Fig. 5a). A similar organization of VirR<sub>DTS</sub> from the same VR1 group also supports this hypothesis (see Supplementary Note 1 and Supplementary Fig. 13). Unlike other channelrhodopsins, OLPVR1 lacks any prominent cavities in the extracellular part of the protein. Instead, it has a pore in the intracellular part, which ends up with a relatively large hydrophilic cavity inside the protein near the retinal (Fig. 5a, b). The cavity is filled with four water molecules (w4, w28, w29, w42) and surrounded by polar residues E44, T87, T88, and W178. Water molecules, together with the backbone oxygen of S203 residue, form a hydrogen bond pentagon (Fig. 5b) and may play a role in the hydration of cation during its translocation. A dense hydrogen bonding network involving water molecules and polar/charged residues protrude from the cavity almost to the extracellular bulk, only breaking in the central region near E51 residue. The putative ion pathway includes three constriction sites inside the protein (Fig. 5b). Each site (described in detail below) is comprise highly-conserved residues (Supplementary Fig. 2). The regions around the constriction sites are

almost identical in OLPVR1 and VirR<sub>DTS</sub><sup>21</sup> (see Supplementary Notes and Supplementary Fig. 13), and, therefore, we consider these to be a characteristic feature of VirChR1s and conjecture that they are also essential for the ion channeling of the entire VR1 group.

The intracellular constriction site (ICS) of the OLPVR1 is formed by the E44 side chain (Fig. 5b). It separates the large intracellular cavity from a polar region near T18, S22, Q48, N205, and S208 in the middle part of the protein, also containing three water molecules w7, w21 and w35. The E44 side chain is pointed towards the retinal, similarly to E122 in C1C2<sup>40</sup> (Supplementary Fig. 12), and is stabilized by hydrogen bonds with T87 and water molecules w20 and w35. Interestingly, unlike in *CrChR2* and Chrimson<sup>47</sup>, where the intracellular constriction sites (intracellular gates) are almost 14 Å far from the RSB and separated from it by the hydrophobic cavity, the ICS of OLPVR1 is located closer (9 Å) to the active center and is connected by extended hydrogen bonding network to the central constriction site (CCS) (Fig. 5b and Supplementary Fig. 14). Moreover, in *CrChR2* and Chrimson, the ICSs are comprise tightly connected charged amino acids, completely blocking the flow of ions in the resting state (Fig. 5b and Supplementary Fig. 14). In these terms, the lack of compaction in the cytoplasmic region of OLPVR1 and the existence of the pore between TM6 and TM7 make the organization of the intracellular part of the protein closer to that of anion channel *GtACR1*<sup>10,48</sup>, where the pore protrudes from the intracellular bulk almost down to the retinal without any constrictions (Fig. 5b). This might mean a different gating mechanism in the cytoplasmic part of OLPVR1 from other channelrhodopsins (Supplementary Fig. 15).

The CCS of the OLPVR1 includes the S11, Q15, E51, and N205 residues that are fully conserved among the members of the VR1 group and likely hinder the ion translocation pathway in the resting state (Fig. 6a, b). The core of the CCS is comprises S11–E51–N205 (S–E–N) triad, which is similar to S63–E90–





## Discussion

**Electrophysiology of VirChR1 and potential of VirChR1s for optogenetic applications.** Do all members of the VR1 function as ion channels? Taking into account all our data, high homology of VirChR1s, high conservation of functionally key amino acids, the high similarity of the structure of OLPVR1 to the recently described structure of VirR<sub>DTS</sub><sup>21</sup>, another representative of the VR1, we suggest that this rhodopsin group (including VirR<sub>DTS</sub>) should all comprise a distinct group of light-gated channels different from the known channelrhodopsins. It was recently reported that VirR<sub>DTS</sub> is a proton pump, based on the experiments with the *E. coli* plasma membrane expressing VirR<sub>DTS</sub>, where the authors measured the pH changes in the *E. coli* suspension upon continuous light illumination<sup>21</sup>. We should note that such experiments are able to detect, in some cases, a proton-pumping activity; however, they are unable to identify a rhodopsin channel activity. Besides, only the observation of proton-pumping activity does not mean that the protein functions as an ion pump. Indeed, it has been shown that CrChR2 also pumps protons; it is the so-called “leaky proton pump”<sup>50</sup>. Moreover, this also is definitely the case of OLPVR1, studied in the present work (Fig. 2g). However, the experimental approach ( $\Delta$ pH measurements) used in ref. <sup>21</sup> cannot prove the existence of ion channel properties. Therefore, by the direct electrophysiology approach, we detected and studied the ion channel properties, with HEK cells expressing VirR<sub>DTS</sub>. The experiments showed that VirR<sub>DTS</sub> is also a light-gated channel (see Supplementary Notes and Supplementary Fig. 13). Indeed, predictably, the data showed photocurrents, which reverse their direction at approximately 0 mV, which is characteristic of rhodopsin ion channels but not of the ion pumps (Supplementary Fig. 14e).

Thus, taking into account all the data, high homology of VirChR1s, high conservation of functionally key amino acids, we suggest that the VirChR1s (including VirR<sub>DTS</sub> from giant PgV virus infecting the algae *Phaeocystis*) likely form a distinct group of compact light-gated Ca<sup>2+</sup>-blocked channels, different from the known channelrhodopsins.

To the best of our knowledge, the exclusive conductivity for monovalent cations and its regulation by divalent cations have not been reported for any rhodopsin before. Recently, the cryptophyte alga *Guillardia Theta* has been reported to encompass genes of two different groups of channelrhodopsins, namely, anion-conducting channelrhodopsins with ~50% sequence identity to chlorophyte channelrhodopsins<sup>9</sup>, and also a group with an architecture unconventional for channelrhodopsins, and containing the DTD motif that is characteristic of the archaeal proton pumps<sup>18</sup>. Consequently, these proteins were denoted BR-like cation channelrhodopsins (BCCRs). It has been shown that cation permeability of one of such proteins (Gt\_CCR4) is reduced under high concentrations of Ca<sup>2+</sup> (more than 40 mM CaCl<sub>2</sub>)<sup>51</sup>. However, such Ca<sup>2+</sup> concentrations are far beyond the physiological conditions. Moreover, even at 40 mM of Ca<sup>2+</sup>, the channel permeability of Gt\_CCR4 is not blocked and is about 20% of that at 2 mM Ca<sup>2+</sup><sup>51</sup>. It is unclear whether these results are biologically relevant. By contrast, VirChR1 is impermeable to divalent cations, and its channeling activity for monovalent cations is suppressed completely in the presence of a few mM of Ca<sup>2+</sup>.

Despite the fact that CrChR2 seemingly exceeds VirChR1 performance in terms of optogenetics, the Ca<sup>2+</sup> impermeability is an important feature that separate VirChR1s from direct competition with chlorophyte channelrhodopsins. At the moment application of VirChR1 in optogenetics is limited by low photocurrent densities, poor plasma membrane localization and relatively slow kinetics. Nevertheless, we expect that VirChR1s may be useful for optogenetic applications because

VR1 family comprise more than 300 potential channelrhodopsins, some of which are likely to have improved plasma membrane localization and faster kinetics. Besides that, because VirChR1s would not interfere with important native Ca<sup>2+</sup>-dependent processes in the cells, Ca<sup>2+</sup>-impermeable channelrhodopsins could become valuable tools for Ca<sup>2+</sup>-sensitive applications, for example, in cellular organelles like mitochondria, or in muscle cells and also to study processes in the brain, where optogenetic manipulation of synapses is advantageous and profitable<sup>52–54</sup>. In some cases, the activation of the slow light-gated channels can result in activation of voltage-gated calcium channels, that might be an issue for VirChR1 protein, however, faster VirChR1s would be able to overcome these limitations<sup>55</sup>.

Another feature of VirChR1 is a non-zero negative photocurrent under symmetrical conditions at 0 mV (Fig. 3d, it also results in positive reversal potential). The same negative photocurrent has also been found when the channel activity was blocked by calcium (Fig. 3f). This suggests that inward-pumping activity could be responsible for this current. However, this explanation contradicts the results of pH measurements with OLPVR1 reconstituted to lipid vesicles. Additional work is required to resolve this discrepancy.

Also, the photocurrent of VirChR1 has an overshooting feature after turning the light off (Fig. 3b, e). Moreover, the photocurrent in the overshooting state tends to have reversal potential shifted towards zero. Although the causes of this effect remain unknown, we suggest that it might be explained by second-photon absorption during the measurements upon continuous illumination. The reversal potential shift in its turn can be explained by the change in the channeling-pumping ratio during the redistribution of proteins between photocycle states during overshooting.

**Distinct structural features of VirChR1s.** The comparison of the high-resolution structures of OLPVR1 and VirR<sub>DTS</sub> (see Supplementary Notes and Supplementary Fig. 13), together with the phylogenetic analysis of VR1, helped us to identify the distinctive structural features of the VirChR1s.

VirChR1s have proteorhodopsin-like architecture with short extracellular loops and share several structural features, such as membrane-associated ICL2 helix and an unconventional TM6-TM7 orientation. VirChR1s also share a set of highly-conserved residues that encompass the ion-conducting pathway and, by analogy with CrChR2, possess three consecutive constriction sites that are likely to be displaced in the open state of the channel (Supplementary Figs. 2, 15). Notably, VirChR1s lack the DC amino acid pair (C128 and D156 in CrChR2), which is replaced by an NT pair (T88 and N121 pair in OLPVR1) that is conserved in all VR1.

In contrast to VR2, VR1 members share the highly-conserved S–E–N triad (S11, E51, and N205) in the core of the protein, which likely plays the key role in the ion-conducting mechanisms of VirChR1s. At the same time, similar triads have been found in most of the known channelrhodopsins (Fig. 6d). In the case of OLPVR1 and VirChR1s in general, the S–E–N triad might also play a role in the inhibition of the protein activity by Ca<sup>2+</sup>. We hypothesize that Ca<sup>2+</sup> binds near the triad and thus blocks the permeation of monovalent cations. This suggestion is based on the recently reported Na<sup>+</sup>-bound O-state structure of the light-driven sodium pump KR2, described in<sup>56</sup>. Indeed, Na<sup>+</sup> has been shown to bind transiently in the core of the protein and is coordinated by the S70–N112–D116 triad, similar to the S11–E51–N205 triad of the CSS of OLPVR1. Given that Ca<sup>2+</sup> has similar coordination geometry to that of Na<sup>+</sup>, it cannot be ruled out that Ca<sup>2+</sup> binds in the CSS of OLPVR1. Similarly to



KR2, where  $\text{Na}^+$  binds to the protein only in the intermediate O-state,  $\text{Ca}^{2+}$  binding also might occur not in the ground state of OLPVR1 but in the course of the protein photocycle. To clarify the details of  $\text{Ca}^{2+}$  binding, as well as to elucidate the gating mechanism of VirChR1s, additional experiments should be performed; in particular, the structures of the intermediate states of OLPVR1 should be solved.

## Discussion

In this work, we demonstrate that VR1 rhodopsins are  $\text{Na}^+/\text{K}^+$  selective light-gated ion channels that are inhibited by divalent cations. Viral channelrhodopsins are encoded together with retinal biosynthesis genes and widely distributed across the photic zone of World Ocean<sup>21,57,58</sup>. Large and especially giant viruses possess many auxiliary metabolic genes (AMG) that enhance the host metabolic functions and hence promote virus reproduction, without direct involvement in the virus replication processes<sup>59–61</sup>. By analogy with CrChR1 and CrChR2 channelrhodopsins from the chlorophyte alga *Chlamydomonas reinhardtii*<sup>62</sup>, VirChR1s could be involved in the sensory activity of their hosts. They could also play the role of an additional ion channel to supplement and augment the host phototaxis systems. Photoexcitation of the channelrhodopsin receptors generates photoreceptor currents and membrane depolarization followed by activation of voltage-gated  $\text{Ca}^{2+}$  channels triggering flagellar motion<sup>62–65</sup>. Due to their impermeability to  $\text{Ca}^{2+}$  cations, viral channelrhodopsins could activate secondary  $\text{Ca}^{2+}$ -channels only by membrane depolarization, but not via biochemical amplification, as has been suggested for CrChR1 and CrChR2<sup>62</sup>. Thus, although the details of the function of viral channelrhodopsins remain to be explored, it appears likely that they enhance the light-induced motility of the host, and so boost anabolic processes required for virus reproduction.

However, some of the hosts infected by giant viruses apparently lack any light-sensory structures, and therefore, viral rhodopsins might be involved in other processes<sup>21</sup>. Many viruses encode small hydrophobic molecules, viroporins, that function as ion channels and size-limited pores and are able to permeabilize cellular membranes<sup>66,67</sup>. Viroporins can function as ion-conducting channels that open in either a voltage-dependent or in a voltage-independent manner, promoting virion assembly. Some of the viral channelrhodopsins might similarly augment virus budding using light energy<sup>68,69</sup>. Many phycodnaviruses, large viruses that belong to the NCLDV, encode channels, in particular, voltage-gated  $\text{K}^+$ -channels<sup>70,71</sup>. The channel proteins are essential for virus reproduction, but their specific functions have not been established. Thus, the role of rhodopsins in the reproduction of the viruses that encode them is part of the more general theme of the functions of diverse virus-encoded transport proteins that remain to be experimentally characterized<sup>72,73</sup>.

## Methods

All experiments involving mouse materials were done in compliance with all relevant ethical regulations for animal testing and research. The study received ethical approval from the Lower Saxony State Office for Consumer Protection and Food Safety (LAVES). No experiments were randomized or blinded in this study. Sample sizes were determined based on prior literature and best practices in the field; no statistical methods were used to predetermine sample size.

**Metagenomic analysis.** Rhodopsins of the VR1 group were retrieved from metagenomic assembled contigs through combining similarity search, protein clustering, and Hidden Markov Models. Briefly, the first dataset of *bona fide* rhodopsins was retrieved from TARA Ocean metagenome-assembled contigs that were downloaded from ENA (<https://www.ebi.ac.uk>). In addition, a new assembly was performed for each sample starting by raw metagenomic reads. Sequencing reads were pre-processed using Trimmomatic<sup>74</sup> in order to remove low-quality bases (Phred quality score was set at 20, sliding windows of 4), and assembled using MEGAHIT<sup>75</sup> and the default parameters for generic metagenome assemblies.

Coding DNA sequences were predicted from contigs longer than 2 Kb using Prodigal<sup>76</sup>, and annotated against the NR database of NCBI using Diamond<sup>77</sup>. *Bona fide* rhodopsins were selected by screening for different keywords related to rhodopsins that must be contained in the annotation, and filtered according to 7 transmembrane domains that were predicted using Phobius<sup>78</sup>. Selected proteins were next aligned using the R package Decipher<sup>79</sup>, and alignments were used to infer a phylogenetic tree through FastTree 2<sup>80</sup> and using default parameters. Phylogenetic distances between nodes on the tree topology were considered for clustering *bona fide* rhodopsins into distinct clades, each of which was used to train a Hidden Markov Model (HMM). All HMMs were finally queried against TARA Ocean assembled contigs using HMMER version v3.1b2 (<http://hmmmer.org>) and setting an e-value threshold of  $1e-5$ . Proteins identified through HMMs were clustered at 100% identity using CD-HIT suite<sup>81</sup> to remove redundancy, and reduced to a total of 2584 Type-1 rhodopsins that were further analyzed.

The dataset of group I viral rhodopsins was constructed by searching the NCBI non-redundant protein sequence databases along with the TARA metagenomic sequences using BLSTP and TBLASTN. For the sake of clarity, for Supplementary Fig. 1, we used a reduced number of sequences. To obtain a representative set of 16 TARA metagenomic sequences with OLPVR1 and VirChR1 sequences included, we used the CD-HIT suite with default parameters and 60% identity cut-off.

**Sequence alignment and phylogenetic analysis.** Rhodopsin sequences were aligned using MUSCLE using UGENE software<sup>82</sup> with the default parameters. Type-1 rhodopsins were named according to their names in literature. The sequence alignment was created using ESPript3 online server<sup>83</sup>. Phylogenetic tree reconstruction was conducted by PHYLIP Neighbor Joining method using UGENE software<sup>82</sup> with the following parameters: Jones–Taylor–Thornton model, transition/transversion ratio = 2.0, no gamma distribution applied. Tree visualization was done using iTOL server<sup>84</sup>. GenBank accession numbers are additionally indicated.

**Cloning, expression, and purification.** The *E. coli* codon-optimized OLPVR1 and VirChR1 genes were synthesized commercially (Eurofins). The nucleotide sequence was optimized for *E. coli* expression using the GeneOptimizer software (Life Technologies). The gene, together with the 5' ribosome-binding sites and the 3' extensions coding additional LEHHHHHHH\* tag, was introduced into the pEKT expression vector (Novagen) via NdeI and XhoI restriction sites and verified by sequencing. VirChR1 protein was additionally supplemented with BRIL protein on the N-terminus of the protein, to improve protein folding and expression level<sup>29</sup>. Full plasmid and gene constructs, as well as primers used in this study can be found in Supplementary Tables 2–4. The proteins were expressed as described previously<sup>7</sup> with further modifications. *E. coli* cells of strain C41 (StabyCodon T7, Eurogentec, Belgium) were transformed with the expression plasmid. Transformed cells were grown in shaking baffled flasks in an autoinducing medium ZYP-5052 containing 50 mg/L kanamycin at 37 °C. When the  $\text{OD}_{600}$  in the growing bacterial culture is 0.8–1.0 (glucose level < 10 mg/L), 10  $\mu\text{M}$  all-*trans*-retinal (Sigma-Aldrich), and 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside were added, the incubation temperature was reduced to 20 °C and incubated for 18 h. After incubation, cells were collected by centrifugation (5000  $\times$  g, 30 min) and disrupted in an M-110 P Lab Homogenizer (Microfluidics) at 20,000 p.s.i. in a buffer containing 20 mM Tris-HCl, pH 8.0 with 50 mg/L DNase I (Sigma-Aldrich). The membrane fraction of the cell lysate was isolated by ultracentrifugation at 90,000  $\times$  g for 1 h at 4 °C (Type 70 Ti Fixed-Angle Titanium Rotor, Beckmann). The pellet was resuspended in a buffer containing 20 mM  $\text{Na}_2\text{HPO}_4/\text{Na}_2\text{HPO}_4$ , pH 8.0, 0.1 M NaCl, and 1% n-dodecyl  $\beta$ -D-maltoside (DDM, Anatrace, Affymetrix) and stirred for 18 h for solubilization. The insoluble fraction was removed by ultracentrifugation at 90,000  $\times$  g for 1 h at 4 °C. The supernatant was loaded on a Ni-NTA column (Qiagen), and washed with a buffer containing 10 mM  $\text{Na}_2\text{HPO}_4/\text{Na}_2\text{HPO}_4$ , 150 mM NaCl, 30 mM imidazole, and 0.05% DDM buffer (pH 8.0). Elution of the protein was done in a buffer containing 10 mM  $\text{Na}_2\text{HPO}_4/\text{Na}_2\text{HPO}_4$ , 150 mM NaCl, 300 mM imidazole and 0.05% DDM (pH 8.0). The eluate was subjected to size-exclusion chromatography on a 20 ml Superdex 200i 10/300 GL column (GE Healthcare Life Sciences) in a buffer containing 10 mM  $\text{Na}_2\text{HPO}_4/\text{Na}_2\text{HPO}_4$ , pH 8.0, 150 mM NaCl, and 0.05% DDM. Protein-containing fractions with an  $A_{280}/A_{500}$  absorbance ratio (peak ratio, p.r.) of lower than 1.5 were pooled and dialyzed against 100 volumes of 10 mM  $\text{Na}_2\text{HPO}_4/\text{Na}_2\text{HPO}_4$ , 150 mM NaCl, and 0.05% DDM (pH 8.0) buffer twice for 2 h to dispose of imidazole. The purified protein was concentrated for 40 mg/ml for crystallization.

**Reconstitution of the protein into lipid-based systems.** Phospholipids (azolectin from soybean, Sigma-Aldrich) were dissolved in  $\text{CHCl}_3$  (Chloroform ultrapure, Applichem Panreac) and dried under a stream of  $\text{N}_2$  in a glass vial. The solvent was removed by overnight incubation under vacuum. The dried lipids were resuspended in 100 mM NaCl buffer supplemented with 2% (w/v) sodium cholate. The mixture was clarified by sonication at 4 °C, and OLPVR1 was added at a protein/lipid ratio of 1:20 (w/w). The detergent was removed by 2 days stirring with detergent-absorbing beads (Amberlite XAD 2, Supelco). The mixture was dialyzed against 100 mM NaCl, (pH 7.0) buffer at 4 °C for 8 h to adjust the desired pH. The obtained liposomes were used for the measurement of pump activity with pH

electrode. The OLPVR1-containing nanodiscs were assembled using a standard protocol described elsewhere<sup>85</sup>. 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC, Avanti Polar Lipids, USA) and an MSP1D1 version of apolipoprotein-1 were used as a lipid and scaffold protein, respectively. The molar ratio during assembly was DMPC:MSP1D1:OLPVR1 = 100:2:3. The protein-containing nanodiscs were dialyzed against 100 volumes of 10 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>, 100 mM NaCl (pH 7.5) buffer twice, and then subjected to size-exclusion chromatography on a 20 ml Superdex 200i 10/300 GL column (GE Healthcare Life Sciences) for detergent removal.

**Ion-trafficking assay with protein-containing liposomes.** The measurements were performed on 2 ml of stirred proteoliposomes suspension at 0 °C. OLPVR1- and LR/Mac- containing liposomes were prepared following the protocol described above. Liposomes were illuminated for 10 min with a halogen lamp (Intralux 5000-1, VOLPI) and then were kept in the dark for another 10 min. Changes in pH were monitored with a pH-meter (LAB 850, Schott Instruments). Some of the measurements were repeated in the presence of 30 μM of carbonyl cyanide m-chlorophenyl hydrazone (CCCP, Sigma-Aldrich) under the same conditions. We used purified LR protein incorporated in POPC:POPS liposomes (3:1) as a positive control. The codon-optimized *Leishmania tarentolae* LR gene was synthesized commercially (Eurofins). Full length LR protein was expressed in LEXSY expression system using protocols described for expression of CrChR2. Full protocol details are described in<sup>8</sup>.

**pH titration.** To investigate the pH dependence of the absorption spectra of OLPVR1, about 6 μM protein was suspended in the titration buffer (10 mM citrate, 10 mM MES, 10 mM HEPES, 10 mM MOPS, 10 mM CHES and 10 mM CAPS). Then, the pH was changed by the addition of diluted or concentrated HCl or NaOH to obtain 0.5–0.7 pH change. The absorption spectra were measured with a UV-visible spectrometer (V-2600PC, Shimadzu).

**VirChR1 expression in mammalian cell lines.** The human codon-optimized OLPVR1, VirChR1 and DTS rhodopsin genes were synthesized commercially (Eurofins). The gene was cloned into the pcDNA3.1(–) vector bearing an additional membrane trafficking signal and YFP fluorescent protein (pcDNA3.1\_VirChR1\_TS\_YFP). The modified version of the proteins included a P2A self-cleaving peptide and Katushka fluorescent protein at the C-terminal part of the gene, and Hemagglutinin and Flag-tag peptides at the N-terminal part of the gene<sup>86–88</sup> (pcDNA3.1\_HF\_VirChR1\_TS\_P2A\_Katushka). Full plasmid and gene constructs, as well as primers used in this study can be found in Supplementary Tables 2–4. The SH-SY5Y human neuroblastoma cells at a confluency of 80–90% were transfected with the plasmid and Lipofectamine LTX according to the manufacturer's protocol (Thermo Fisher Scientific). The cells were incubated under 5% CO<sub>2</sub> at 37 °C. After transfection (16–24 h), electrophysiological experiments were performed.

**Electrophysiological recordings.** For the electrophysiological characterization of VirChR1, whole-cell patch-clamp recordings were performed (Scientifica LASU, Axon Digidata 1550A, Multiclamp 700B). Horizontal puller (Sutter Instrument CO, Model P-2000) was used for the fabrication of patch pipettes (borosilicate glass GB150F-8P, 3–6 MΩ). Experiments were conducted using SH-SY5Y cell line. Photocurrents were measured in response to LED light pulses with saturating intensity  $\lambda = 470 \pm 20$  nm (~3 mW/mm<sup>2</sup>), and the corresponding photon flux was  $7 \times 10^{15}$  photons/s/mm<sup>2</sup>, assuming wavelength of emitted light to be 470 nm). For the action spectra, ultrashort nanosecond light pulses were generated by Brilliant Quantel using OPO Opotek MagicPrism for different wavelengths.

**Virus preparation.** For virus purification and transduction of the primary hippocampal neurons we used VirChR1 gene without N-terminal HA-FLAG tag (pAAV\_hSyn\_VirChR1\_TS\_P2\_Katushka-WPRE\_bGH), which was generated by site-directed mutagenesis using QuikChange II XL Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's instructions. Here, P2A is a self-cleaving peptide, TS is trafficking signal to the plasma membrane from inwardly rectifying potassium channel subunit Kir2.1<sup>89</sup>, WPRE represents Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element, and bGH is polyadenylation element. The template for this mutagenesis was the original N-terminally HA-FLAG tagged pAAV construct similar to those used for SH-SY5Y experiments. AAVs were generated in HEK-293 T cells (ATCC) using polyethylenimine (25,000 MW, Polysciences, USA) transfection. Briefly, triple transfection of HEK-293T cells was performed using pHelper plasmid (TaKaRa/Clontech), plasmid providing viral capsid AAV-PHP.eB (PHP.eB was a gift from Viviana Gradinaru (Addgene plasmid # 103005; <http://n2t.net/addgene:103005>; RRID:Addgene\_103005)), and plasmid providing the VirChR1 gene. The cell line was regularly tested for mycoplasma. We harvested viral particles 72 h after transfection from the medium and 120 h after transfection from cells and the medium. Viral particles from the medium were precipitated with 40% polyethylene glycol 8000 (Acros Organics, Germany) in 500 mM NaCl for 2 h at 4 °C and then after centrifugation at 4000 × g for 30 min combined with cell pellets for processing. The cell pellets were suspended in 500 mM NaCl, 40 mM Tris, 2.5 mM MgCl<sub>2</sub>, pH 8, and 100 U/ml of salt-

activated nuclease (Arcticzymes, USA) at 37 °C for 30 min. Afterward, the cell lysates were clarified by centrifugation at 2000 × g for 10 min and then purified over iodixanol (Optiprep, Axis Shield, Norway) step gradients (15, 25, 40, and 60%) at 320,006 × g for 2.25 h. Viruses were concentrated using Amicon filters (EMD, UFC910024) and formulated in sterile phosphate buffered saline (PBS) supplemented with 0.001% Pluronic F-68 (Gibco, Germany). Virus titers were measured using AAV titration kit (TaKaRa/Clontech) according to the manufacturer's instructions by determining the number of DNase I-resistant vg using qPCR (StepOne, Applied Biosystems). Purity of produced viruses was routinely checked by silver staining (Pierce, Germany) after gel electrophoresis (NovexTM 4–12% Tris-glycine, Thermo Fisher Scientific) according to the manufacturer's instruction. The presence of viral capsid proteins was positively confirmed in all virus preparations. Viral stocks were kept at 80 °C until experimental day. The AAV production protocol is the same as previously described<sup>90</sup>. The titer for AAV-PHP.eB\_hSyn\_VirChR1\_TS\_P2A\_Katushka\_WPRE\_bGH was  $1.8 \times 10^{13}$  GC/ml.

**Hippocampal neuronal cultures, electrophysiology, and immunocytochemistry.** Primary cultured hippocampal neurons were prepared from embryonic E18/E19 Wistar rat pups. Isolated hippocampi were digested with Trypsin 0.25% and plated onto glass coverslips precoated with poly-D-lysine (25,000 cells per cm<sup>2</sup>). Hippocampal cultures were transduced at 10 DIV with AAV-PHP.eB virus carrying the VirChR1 gene (~10<sup>10</sup> GC/ml). Whole-cell patch-clamp measurements were performed at 16–17 DIV. The intracellular solution contained 129 mM K-gluconate, 10 mM HEPES, 10 mM KCl, 4 mM MgATP, 0.3 mM Na<sub>3</sub>GTP, pH 7.2. Extracellular solution consisted of 140 mM NaCl, 1 mM MgCl<sub>2</sub>, 2.5 mM KCl, 10 mM HEPES, 1.5 mM CaCl<sub>2</sub>, 15 mM glucose, pH 7.4. 10 μM of NBQX and 50 μM of AP5 were added to the extracellular solution to block synaptic transmission. The illumination of cells was performed using 473 nm laser with saturation intensity (150 mW/cm<sup>2</sup>). All experiments were conducted at room temperature. Current-clamp recordings were performed with zero current injection. The electrophysiology data was processed using in-house Python 3 scripts (Jupyter Notebook). Hippocampal neurons used for immunocytochemistry were transduced at DIV 10 (final concentration  $9 \times 10^9$  –  $1.8 \times 10^6$  gc) and after 5–6 days of incubation fixed by 4% paraformaldehyde (PFA) and subjected further to immunostaining. Microtubule associated protein (Map2) was used as neuronal marker and immunolabelled by primary rabbit polyclonal Map2 antibody (Abcam, ab32454, 1:750). Secondary antibody against rabbit Map2 was Alexa Fluor goat anti-rabbit 647 IgG (Invitrogen, A21244, 1:750). VirChR1 was indirectly identified by intrinsic fluorescence of Katushka. All images were obtained with Leica scanning confocal microscope SP5 and processed using FIJI software. Final images were assembled in Adobe Illustrator software.

**Time-resolved absorption spectroscopy.** Excitation/detection systems were composed as such: Brilliant B laser with OPO Rainbow (Quantel Inc.) was used, providing pulses of 4-ns duration at 530-nm wavelength and an energy of 2 mJ per pulse. Samples (5 × 5-mm spectroscopic quartz cuvette; Hellma GmbH & Co.) were placed in a thermostated house between two collimated and mechanically coupled monochromators (LOT MSH150). The probing light (xenon arc lamp, 75 W, Hamamatsu) passed the first monochromator sample and arrived after a second monochromator at a photomultiplier tube (PMT) detector (R12829, Hamamatsu). The current-to-voltage converter of the PMT determines the time resolution of the measurement system of ca. 50 ns (measured as an apparent pulse width of the 5-ns laser pulse). Two digital oscilloscopes (Keysight DSO-x4022A) were used to record the traces of transient transmission changes in two overlapping time windows. The maximal digitizing rate was 10 ns per data point. Transient absorption changes were recorded from 10 ns after the laser pulses until full completion of the phototransformation. At each wavelength, 25 laser pulses were averaged to improve the signal-to-noise ratio. The quasilogarithmic data compression reduced the initial number of data points per trace (~32,000) to ~850 points evenly distributed in a log time scale giving ~100 points per time decade. The wavelengths were varied from 330 to 700 nm in steps of 10 nm using a computer-controlled step motor. Absorption spectra of the samples were measured before and after each experiment on a standard spectrophotometer (Avantes Avaspec 2048 L). Obtained datasets were independently analyzed using the multiexponential least-squares fitting by MEXFIT software<sup>91</sup>. The number of exponential components was incremented until the SD of weighted residuals did not further improve. After establishing the apparent rate constants and their assignment to the internal irreversible transitions of a single chain of relaxation processes, the amplitude spectra of exponents were transformed to the difference spectra of the corresponding intermediates in respect to the spectrum of the final state.

**Crystallization.** The crystals of OLPVR1 and O1O2 proteins were grown with an *in meso* approach<sup>39</sup>, similar to that used in our previous works<sup>3</sup>. In particular, the solubilized protein (40 mg/ml) in the crystallization buffer was mixed with pre-melted at 50 °C monoolein (MO, Nu-Chek Prep) or monopalmitolein (MP, Nu-Chek Prep) in 3:2 ratio (lipid:protein) to form a lipidic mesophase. The mesophase was homogenized in coupled syringes (Hamilton) by transferring the mesophase from one syringe to another until a homogeneous and gel-like material was formed. 150 nl drops of a protein-mesophase mixture were spotted on a 96-well LCP glass

sandwich plate (Marienfeld) and overlaid with 400 nL of the precipitant solution by means of the NT8 crystallization robot (Formulatrix). The best crystals of OLPVR1 were obtained with a protein concentration of 20 mg/ml and 10 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>, 24% PEG 6000, 100 mM Tris (pH 8.2) for MP lipid and 10 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>, 24% PEG 550, 100 mM Tris (pH 8.2) for MO lipid (Hampton Research). The best crystals of O1O2 were obtained with a protein concentration of 20 mg/ml and 1.8 M Na<sub>2</sub>HPO<sub>4</sub>/KH<sub>2</sub>PO<sub>4</sub> (pH 4.6). The crystals were grown at 22 °C and appeared in 1 to 4 weeks. Once crystals reached their final size, crystallization wells were opened, and drops containing the protein-mesophase mixture were covered with 100 µl of the respective precipitant solution. For data collection harvested crystals were incubated for 5 min in the respective precipitant solutions.

**Acquisition and treatment of diffraction data.** X-ray diffraction data of OLPVR1 were collected at the beamlines ID30b and ID23-1 of the ESRF, Grenoble, France, using a PILATUS 6M detector. The data collection at ESRF was performed using MxCube2 software. X-ray diffraction data of the O1O2 chimera were collected at the X06SA beamline of the SLS, Villigen, Switzerland, using EIGER 16M detector. Diffraction images were processed using XDS<sup>92</sup>. The reflection intensities were scaled using the AIMLESS software from the CCP4 program suite<sup>93</sup>. The reflection intensities of the highest-resolution data on OLPVR1 (1.4 Å) were also scaled using the Staranis server<sup>94</sup> for the validation of the electron density maps quality. There is no possibility of twinning for the crystals. In all cases, diffraction data from one crystal was used. The data treatment statistics are presented in Supplementary Table 1.

**Structure determination and refinement.** Initial phases for the OLPVR1 structures were successfully obtained in the P2<sub>1</sub>2<sub>1</sub>2 and P1 space groups by molecular replacement (MR) using MOLREP<sup>95</sup> using the chain A of the 6SQG structure (OLPVR1 protein) as a search model. Initial phases for O1O2 chimera were successfully obtained in the I121 space group by MR using the obtained structure of OLPVR1 as a search model. The initial MR models were iteratively refined using REFMAC5<sup>96</sup>, PHENIX<sup>97</sup> and Coot<sup>98</sup>. The structure refinement statistics are presented in Supplementary Table 1.

**Molecular dynamics.** We used the refined 1.4 Å resolution crystallographic OLPVR1 structure for the initial conformation. All non-protein atoms except water were removed from the structure and the all-trans retinal molecule connected to the Lys were renamed using the retinol and retinal parameters for Charmm36 force field. The system then was prepared using Charmm GUI<sup>99</sup> input generator using the POPC lipid membrane and Tip3P water model. The resulting amount of lipids was 132, amount of water molecules –8901, amount of sodium ions –26, chlorine ions –23, overall system size was 48,350 atoms. Energy minimization and equilibration were performed in several steps with the gradual removal of spatial atomic constraints. The resulting simulation time was 1 µs (current time is 0.75 µs). Simulations were performed using velocity-rescale thermostat at 303.15 K and Parrinello-Rahman semi isotropic barostat with Gromacs 2018.4<sup>100</sup>.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Data supporting the findings of this manuscript are available from the corresponding author upon reasonable request. A reporting summary for this Article is available as a Supplementary Information file. The protein coordinates and atomic structure factors have been deposited in the Protein Data Bank (PDB) under accession number PDB 7AKW (O1O2 mutant), PDB 7AKX (OLPVR1 in P1 space group), and PDB 7AKY (OLPVR1 in P21212 space group), respectively. Source data are provided with this paper.

Received: 22 July 2020; Accepted: 7 October 2020;

Published online: 11 November 2020

## References

- Ernst, O. P. et al. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163 (2014).
- Gushchin, I. & Gordeliy, V. Microbial rhodopsins. *Subcell. Biochem.* [https://doi.org/10.1007/978-981-10-7757-9\\_2](https://doi.org/10.1007/978-981-10-7757-9_2) (2018).
- Gordeliy, V. I. et al. Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* **419**, 484–487 (2002).
- Gushchin, I. et al. Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. *Proc. Natl Acad. Sci. USA* **110**, 12631–12636 (2013).
- Kato, H. E. et al. Structural basis for Na<sup>+</sup> transport mechanism by a light-driven Na<sup>+</sup> pump. *Nature* **521**, 48–53 (2015).
- Gushchin, I. et al. Crystal structure of a light-driven sodium pump. *Nat. Struct. Mol. Biol.* **22**, 390–396 (2015).
- Shevchenko, V. et al. Inward H<sup>+</sup> pump xenorhodopsin: mechanism and alternative optogenetic approach. *Sci. Adv.* **3**, e1603187 (2017).
- Volkov, O. et al. Structural insights into ion conduction by channelrhodopsin 2. *Science* **358**, eaan8862 (2017).
- Govorunova, E. G., Sineshchekov, O. A., Janz, R., Liu, X. & Spudich, J. L. Natural light-gated anion channels: a family of microbial rhodopsins for advanced optogenetics. *Science* <https://doi.org/10.1126/science.aaa7484> (2015).
- Kim, Y. S. et al. Crystal structure of the natural anion-conducting channelrhodopsin GtACR1. *Nature* <https://doi.org/10.1038/s41586-018-0511-6> (2018).
- Mukherjee, S., Hegemann, P. & Broser, M. Enzymorhodopsins: novel photoregulated catalysts for optogenetics. *Curr. Opin. Struct. Biol.* <https://doi.org/10.1016/j.sbi.2019.02.003> (2019).
- Beja, O. et al. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
- Nagel, G. et al. Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.1936192100> (2003).
- Zhang, F., Aravanis, A. M., Adamantidis, A., De Lecea, L. & Deisseroth, K. Circuit-breakers: optical technologies for probing neural signals and systems. *Nat. Rev. Neurosci.* <https://doi.org/10.1038/nrn2192> (2007).
- Berndt, A., Lee, S. Y., Ramakrishnan, C. & Deisseroth, K. Structure-guided transformation of channelrhodopsin into a light-activated chloride channel. *Science* **344**, 420–424 (2014).
- Berndt, A. & Deisseroth, K. Expanding the optogenetics toolkit. *Science* <https://doi.org/10.1126/science.aac7889> (2015).
- Kato, H. E. et al. Structural mechanisms of selectivity and gating in anion channelrhodopsins. *Nature* <https://doi.org/10.1038/s41586-018-0504-5> (2018).
- Sineshchekov, O. A., Govorunova, E. G., Li, H. & Spudich, J. L. Bacteriorhodopsin-like channelrhodopsins: alternative mechanism for control of cation conductance. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.1710702114> (2017).
- Yutin, N. & Koonin, E. V. Proteorhodopsin genes in giant viruses. *Biol. Direct* <https://doi.org/10.1186/1745-6150-7-34> (2012).
- López, J. L. et al. Microbial and viral-like rhodopsins present in coastal marine sediments from four polar and subtropical regions. *FEMS Microbiol. Ecol.* <https://doi.org/10.1093/femsec/fiw216> (2017).
- Needham, D. M. et al. A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.1907517116> (2019).
- Bratanov, D. et al. Unique structure and function of viral rhodopsins. *Nat. Commun.* **10**, 4939 (2019).
- Short, S. M. The ecology of viruses that infect eukaryotic algae. *Environ. Microbiol.* <https://doi.org/10.1111/j.1462-2920.2012.02706.x> (2012).
- Gómez-Consarnau, L. et al. Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Sci. Adv.* <https://doi.org/10.1126/sciadv.aaw8855> (2019).
- Oesterheld, D. & Stoekenius, W. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nat. N. Biol.* **233**, 149–152 (1971).
- Luecke, H., Schobert, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291**, 899–911 (1999).
- Wietek, J. et al. Conversion of channelrhodopsin into a light-gated chloride channel. *Science* **344**, 409–412 (2014).
- Kuhne, J. et al. Unifying photocycle model for light adaptation and temporal evolution of cation conductance in channelrhodopsin-2. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.1818707116> (2019).
- Chun, E. et al. Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20**, 967–976 (2012).
- Jorgensen, B. B., Cohen, Y. & Des Marais, D. J. Photosynthetic action spectra and adaptation to spectral light distribution in a benthic cyanobacterial mat. *Appl. Environ. Microbiol.* <https://doi.org/10.1128/aem.53.4.879-886.1987> (1987).
- Balashov, S. P. Protonation reactions and their coupling in bacteriorhodopsin. *Biochim. Biophys. Acta* **1460**, 75–94 (2000).
- Huber, R. et al. pH-Dependent photoisomerization of retinal in proteorhodopsin †. *Biochemistry* **44**, 1800–1806 (2005).
- Li, H., Sineshchekov, O. A., Da Silva, G. F. Z. & Spudich, J. L. In vitro demonstration of dual light-driven Na<sup>+</sup>/H<sup>+</sup> pumping by a microbial rhodopsin. *Biophys. J.* **109**, 1446–1453 (2015).
- Huang, K. S., Bayley, H. & Khorana, H. G. Delipidation of bacteriorhodopsin and reconstitution with exogenous phospholipid. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.77.1.323> (1980).
- Kovalev, K. et al. Structure and mechanisms of sodium-pumping KR2 rhodopsin. *Sci. Adv.* **5**, eaav2671 (2019).



36. Waschuk, S. A., Bezerra, A. G., Shi, L. & Brown, L. S. Leptosphaeria rhodopsin: Bacteriorhodopsin-like proton pump from a eukaryote. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.0409659102> (2005).
37. Tsunogai, S., Yamahata, H., Kudo, S. & Saito, O. Calcium in the Pacific Ocean. *Deep-Sea Res. Oceanogr. Abstr.* [https://doi.org/10.1016/0011-7471\(73\)90087-9](https://doi.org/10.1016/0011-7471(73)90087-9) (1973).
38. Kleinlogel, S. et al. Ultra light-sensitive and fast neuronal activation with the Ca<sup>2+</sup>-permeable channelrhodopsin CatCh. *Nat. Neurosci.* **14**, 513–518 (2011).
39. Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc. Natl Acad. Sci. USA* **93**, 14532–14535 (1996).
40. Kato, H. E. et al. Crystal structure of the channelrhodopsin light-gated cation channel. *Nature* <https://doi.org/10.1038/nature10870> (2012).
41. Baek, M., Park, T., Heo, L., Park, C. & Seok, C. GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkx246> (2017).
42. Ran, T. et al. Cross-protomer interaction with the photoactive site in oligomeric proteorhodopsin complexes. *Acta Crystallogr. D. Biol. Crystallogr.* <https://doi.org/10.1107/S0907444913017575> (2013).
43. Okada, T. et al. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* <https://doi.org/10.1016/j.jmb.2004.07.044> (2004).
44. Freier, E., Wolf, S. & Gerwert, K. Proton transfer via a transient linear water-molecule chain in a membrane protein. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.1104735108> (2011).
45. Berndt, A., Yizhar, O., Gunaydin, L. A., Hegemann, P. & Deisseroth, K. Bistable neural state switches. *Nat. Neurosci.* **12**, 229–234 (2009).
46. Nack, M. et al. The DC gate in channelrhodopsin-2: crucial hydrogen bonding interaction between C128 and D156. *Photochem. Photobiol. Sci.* <https://doi.org/10.1039/b9pp00157c> (2010).
47. Oda, K. et al. Crystal structure of the red light-activated channelrhodopsin Chrimson. *Nat. Commun.* <https://doi.org/10.1038/s41467-018-06421-9> (2018).
48. Li, H. et al. Crystal structure of a natural light-gated anion channelrhodopsin. *eLife* <https://doi.org/10.7554/eLife.41741> (2019).
49. Gerwert, K., Freier, E. & Wolf, S. The role of protein-bound water molecules in microbial rhodopsins. *Biochim. Biophys. Acta.* <https://doi.org/10.1016/j.bbabi.2013.09.006> (2014).
50. Feldbauer, K. et al. Channelrhodopsin-2 is a leaky proton pump. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.0905852106> (2009).
51. Shigemura, S., Hososhima, S., Kandori, H. & Tsunoda, S. P. Ion channel properties of a cation channelrhodopsin. *Gt\_CCR4. Appl. Sci. Switz.* <https://doi.org/10.3390/app9173440> (2019).
52. Lin, J. Y. A user's guide to channelrhodopsin variants: features, limitations and future developments. *Exp. Physiol.* <https://doi.org/10.1113/expphysiol.2009.051961> (2011).
53. Lin, J. Y. et al. Optogenetic inhibition of synaptic release with chromophore-assisted light inactivation (CALI). *Neuron* <https://doi.org/10.1016/j.neuron.2013.05.022> (2013).
54. Tritsch, N. X., Granger, A. J. & Sabatini, B. L. Mechanisms and functions of GABA co-release. *Nat. Rev. Neurosci.* <https://doi.org/10.1038/nrn.2015.21> (2016).
55. Zhang, Y.-P. & Oertner, T. G. Optical induction of synaptic plasticity using a light-sensitive channel. *Nat. Methods* **4**, 139–141 (2007).
56. Kovalev, K. et al. Molecular mechanism of light-driven sodium pumping. *Nat. Commun.* **11**, 2137 (2020).
57. Uitz, J., Claustre, H., Gentili, B. & Stramski, D. Phytoplankton class-specific primary production in the world's oceans: seasonal and interannual variability from satellite observations. *Glob. Biogeochem. Cycles* <https://doi.org/10.1029/2009GB003680> (2010).
58. Yau, S. et al. Virophage control of antarctic algal host-virus dynamics. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.1018221108> (2011).
59. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* <https://doi.org/10.1038/nature04111> (2005).
60. Sharon, I. et al. Photosystem I gene cassettes are present in marine virus genomes. *Nature* <https://doi.org/10.1038/nature08284> (2009).
61. Filée, J. Giant viruses and their mobile genetic elements: the molecular symbiosis hypothesis. *Curr. Opin. Virol.* <https://doi.org/10.1016/j.coviro.2018.07.013> (2018).
62. Sineshchekov, O. A., Jung, K. H. & Spudich, J. L. Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.122243399> (2002).
63. Kianianmomeni, A., Stehfest, K., Nematollahi, G., Hegemann, P. & Hallmann, A. Channelrhodopsins of volvox carteri are photochromic proteins that are specifically expressed in somatic cells under control of light, temperature, and the sex inducer. *Plant Physiol.* <https://doi.org/10.1104/pp.109.143297> (2009).
64. Jékely, G. Evolution of phototaxis. *Philos. Trans. R. Soc. B: Biol. Sci.* <https://doi.org/10.1098/rstb.2009.0072> (2009).
65. Böhm, M. et al. Channelrhodopsin-1 phosphorylation changes with phototactic behavior and responds to physiological stimuli in chlamydomonas. *Plant Cell* <https://doi.org/10.1105/tpc.18.00936> (2019).
66. Gonzalez, M. E. & Carrasco, L. Viroporins. in: *FEBS Lett.* [https://doi.org/10.1016/S0014-5793\(03\)00780-4](https://doi.org/10.1016/S0014-5793(03)00780-4) (2003).
67. Nieva, J. L., Madan, V. & Carrasco, L. Viroporins: Structure and biological functions. *Nat. Rev. Microbiol.* <https://doi.org/10.1038/nrmicro2820> (2012).
68. Fischer, W. B. & Hsu, H. J. Viral channel forming proteins—modeling the target. *Biochim. Biophys. Acta.* <https://doi.org/10.1016/j.bbamem.2010.05.014> (2011).
69. Rossman, J. S., Jing, X., Leser, G. P. & Lamb, R. A. Influenza Virus M2 Protein Mediates ESCRT-Independent Membrane Scission. *Cell* <https://doi.org/10.1016/j.cell.2010.08.029> (2010).
70. Kang, M. et al. Small potassium ion channel proteins encoded by chlorella viruses. *Proc. Natl Acad. Sci. USA* **101**, 5318–5324 (2004).
71. Siotto, F. et al. Viruses infecting marine picoplankton encode functional potassium ion channels. *Virology* <https://doi.org/10.1016/j.virol.2014.05.002> (2014).
72. Thiel, G., Greiner, T., Dunigan, D. D., Moroni, A. & Van Etten, J. L. Large dsDNA chloroviruses encode diverse membrane transport proteins. *Virology* **479–480**, 38–45 (2015).
73. Greiner, T., Moroni, A., Van Etten, J. L. & Thiel, G. Genes for membrane transport proteins: not so rare in viruses. *Viruses* <https://doi.org/10.3390/v10090456> (2018).
74. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
75. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
76. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* <https://doi.org/10.1186/1471-2105-11-119> (2010).
77. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* <https://doi.org/10.1038/nmeth.3176> (2014).
78. Käll, L., Krogh, A. & Sonnhammer, E. L. L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* <https://doi.org/10.1016/j.jmb.2004.03.016> (2004).
79. Wright, E. S. DECIPHER: Harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinforma.* <https://doi.org/10.1186/s12859-015-0749-z> (2015).
80. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0009490> (2010).
81. Li, W. & Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btl158> (2006).
82. Okonechnikov, K. et al. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/bts091> (2012).
83. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDScript server. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gku316> (2014).
84. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkw290> (2016).
85. Ritchie, T. K. et al. Chapter 11 reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Methods Enzymol.* [https://doi.org/10.1016/S0076-6879\(09\)64011-8](https://doi.org/10.1016/S0076-6879(09)64011-8) (2009).
86. Gradinaru, V. et al. Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* **141**, 154–165 (2010).
87. Shcherbo, D. et al. Bright far-red fluorescent protein for whole-body imaging. *Nat. Methods* <https://doi.org/10.1038/nmeth1083> (2007).
88. Wang, Y., Wang, F., Wang, R., Zhao, P. & Xia, Q. 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*. *Sci. Rep.* **5**, 16273 (2015).
89. Hofherr, A., Fakler, B. & Klöcker, N. Selective Golgi export of Kir2.1 controls the stoichiometry of functional Kir2.x channel heteromers. *J. Cell Sci.* <https://doi.org/10.1242/jcs.02322> (2005).
90. Keppeler, D. et al. Ultrafast optogenetic stimulation of the auditory pathway by targeting-optimized Chronos. *EMBO J.* <https://doi.org/10.15252/emj.201899649> (2018).
91. Chizhov, I. et al. Spectrally silent transitions in the bacteriorhodopsin photocycle. *Biophys. J.* **71**, 2329–2345 (1996).
92. Kabsch, W. XDS. *Acta Crystallogr. Biol. Crystallogr.* <https://doi.org/10.1107/S0907444909047337> (2010).

93. Winn, M. D. et al. Overview of the CCP 4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* <https://doi.org/10.1107/S0907444910045749> (2011).
94. Tickle, I. J. et al. STARANISO. (Global Phasing Ltd, Cambridge, United Kingdom, 2018).
95. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D. Biol. Crystallogr.* <https://doi.org/10.1107/S0907444909042589> (2010).
96. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* <https://doi.org/10.1107/S0907444911001314> (2011).
97. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* <https://doi.org/10.1107/S0907444909052925> (2010).
98. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
99. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
100. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
101. Klapoetke, N. C. et al. Independent optical excitation of distinct neural populations. *Nat. Methods* **11**, 338–346 (2014).
102. Govorunova, E. G. et al. The expanding family of natural anion channelrhodopsins reveals large variations in kinetics, conductance, and spectral sensitivity. *Sci. Rep.* <https://doi.org/10.1038/srep43358> (2017).
103. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
104. Volkov, O. et al. Structural insights into ion conduction by channelrhodopsin 2. *Science* **358**, eaan8862 (2017).
105. Ho, B. K. & Gruswitz, F. HOLLOW: Generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).
106. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* <https://doi.org/10.1101/gr.849004> (2004).
107. Fenalti, G. et al. Molecular control of  $\delta$ -opioid receptor signalling. *Nature* <https://doi.org/10.1038/nature12944> (2014).

## Acknowledgements

We thank C. Baeken for technical assistance. We thank D. Gerke for technical assistance during virus purification and S. Langer for technical assistance during hippocampal culture preparation. We acknowledge the Structural Biology Group of the European Synchrotron Radiation Facility (ESRF) for granting access to the synchrotron beamlines. We also acknowledge the Paul Scherrer Institut, Villigen, Switzerland for provision of synchrotron radiation beamtime at beamline X06SA of the SLS and would like to thank Dr. Anuschka Pauluhn for assistance. Electrophysiological experiments were supported by HGF-RSF grant (Helmholtz—RSF Joint Research Groups grant (RSF No. 19-44-06302)), the common program of Agence Nationale de la Recherche (ANR), France and Deutsche Forschungsgemeinschaft (DFG), Germany (ANR-15-CE11-0029-02/FA 301/11-1), by the DFG Research Unit FOR 2518 (*Dynlon*, project P4 to JPM, MA 7525/1-1), and by funding from Frankfurt: Cluster of Excellence Frankfurt Macromolecular Complexes by the Max Planck Society (to E.B.) and by the Commissariat à l’Énergie Atomique et aux Énergies Alternatives (Institut de Biologie Structurale)—Helmholtz—Gemeinschaft Deutscher Forschungszentren (Forschungszentrum Jülich) Special Terms and Conditions 5.1 specific agreement. This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) within the Grenoble Partnership for Structural Biology (PSB). Platform access was supported by FRISBI (ANR-10-INBS-05-02) and GRAL, a project of the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). Crystallographic experiments were funded by RFBR and CNRS according to the research project № 19-52-15017. This work was supported by the project ANR-19-CE11-0026. FRV was supported by grant VIREVO CGL2016-76273-P [AEI/FEDER, EU], (cofounded with FEDER funds). DZ was supported by the Russian Foundation for Basic Research project numbers

17-00-00164 and 20-34-90009. A.G., K.S. and M.K. were supported by Russian Foundation for Basic Research project numbers 17-00-00166 and 17-00-00167. V.G. was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement # 075-00337-20-03, project FSMG -2020-0003). Work in Göttingen was funded by the European Research Council through the Advanced Grant ‘OptoHear’ to T. M. under the European Union’s Horizon 2020 Research and Innovation program (grant agreement No. 670759) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - EXC 2067/1- 390729940.

## Author contributions

D.Z., A.A., and K.K. contributed equally and either has the right to list himself first in bibliographic documents. D.Z., T.B., D.B., and S.V. did molecular cloning, expressed and purified the proteins. D.V., D.S., and I.C. measured the photocycle kinetics and analyzed the corresponding data. D.Z. and R.A. crystallized the proteins. K.K. and D.Z. harvested crystals, processed the data and solved the structures with the help of A.P. A.A., and A.S.\_O. performed electrophysiology experiments under E.B., T.Ma., and M.V. supervision. E.S. performed electrophysiological experiments with DTS protein under A.O. supervision. T.Ma. directly supervised isolation of the hippocampal neurons. A.A. and E.P. performed site-directed mutagenesis and produced AAVs under direct supervision of V.R. V.R. and E.P. performed immunocytochemistry of hippocampal neurons and processed the images. D.Z., M.R., T.R., and Y.A. carried out the functional tests. G.A. and K.S. did the molecular dynamics experiments. R.R. and F.R.V. performed metagenomic search and identified new viral rhodopsin sequences. D.Z., N.Y., and E.K. analyzed the viral rhodopsin sequences and their possible biological role. V.G. initiated, designed and supervised the project and preparation of the manuscript. V.G., G.B., E.K., M.K., T.Mo., and E.B. planned and guided the work. D.Z., A.A., K.K., and V.G. analyzed the data and prepared the manuscript with the input from all other authors. D.W. and A.R. helped with preparation of the revised version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19457-7>.

Correspondence and requests for materials should be addressed to V.G.

Peer review information *Nature Communications* thanks Satoshi Tsunoda and other, anonymous, reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Dmitrii Zabelskii<sup>1,2,3,26</sup>, Alexey Alekseev<sup>1,2,3,4,26</sup>, Kirill Kovalev<sup>1,2,3,4,5,26</sup>, Vladan Rankovic<sup>6,7</sup>, Taras Balandin<sup>1,2</sup>, Dmytro Soloviov<sup>3,8,9</sup>, Dmitry Bratanov<sup>1,2</sup>, Ekaterina Savelyeva<sup>10,11,12</sup>, Elizaveta Podolyak<sup>3</sup>, Dmytro Volkov<sup>1,2</sup>, Svetlana Vaganova<sup>1,2</sup>, Roman Astashkin<sup>3,5</sup>, Igor Chizhov<sup>13</sup>, Natalia Yutin<sup>14</sup>,

Maksim Rulev<sup>1,2,15</sup>, Alexander Popov<sup>15</sup>, Ana-Sofia Eria-Oliveira<sup>5</sup>, Tatiana Rokitskaya<sup>16</sup>, Thomas Mager<sup>6,17</sup>, Yuri Antonenko<sup>16</sup>, Riccardo Rosselli<sup>18,19</sup>, Grigoriy Armeev<sup>20</sup>, Konstantin Shaitan<sup>20,21</sup>, Michel Vivaudou<sup>5,22</sup>, Georg Büldt<sup>3</sup>, Andrey Rogachev<sup>3,8</sup>, Francisco Rodriguez-Valera<sup>3,18</sup>, Mikhail Kirpichnikov<sup>20,23</sup>, Tobias Moser<sup>6,7,17</sup>, Andreas Offenhäusser<sup>10</sup>, Dieter Willbold<sup>1,2,24</sup>, Eugene Koonin<sup>14</sup>, Ernst Bamberg<sup>3,25</sup> & Valentin Gordeliy<sup>1,2,3,4,5</sup>✉

<sup>1</sup>Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich GmbH, Jülich, Germany. <sup>2</sup>JuStruct: Jülich Center for Structural Biology, Forschungszentrum Jülich GmbH, Jülich, Germany. <sup>3</sup>Research Center for Molecular Mechanisms of Aging and Age-related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia. <sup>4</sup>Institute of Crystallography, University of Aachen (RWTH), Aachen, Germany. <sup>5</sup>Institut de Biologie Structurale (IBS), Université Grenoble Alpes, CEA, CNRS, Grenoble, France. <sup>6</sup>Institute for Auditory Neuroscience and InnerEarLab, University Medical Center Göttingen, Göttingen, Germany. <sup>7</sup>Auditory Neuroscience and Optogenetics Laboratory, German Primate Center, Göttingen, Germany. <sup>8</sup>Joint Institute for Nuclear Research, Dubna, Russia. <sup>9</sup>Institute for Safety Problems of Nuclear Power Plants, NAS of Ukraine, Kyiv 03680, Ukraine. <sup>10</sup>Institute of Biological Information Processing (IBI-3: Bioelectronics), Forschungszentrum Jülich GmbH, Jülich, Germany. <sup>11</sup>Laboratory of Functional Materials and Devices for Nanoelectronics, Moscow Institute of Physics and Technology, Dolgoprudny, Russia. <sup>12</sup>Center of Shared Research Facilities, Moscow Institute of Physics and Technology, Dolgoprudny, Russia. <sup>13</sup>Institute for Biophysical Chemistry, Hannover Medical School, Hannover, Germany. <sup>14</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. <sup>15</sup>European Synchrotron Radiation Facility, Grenoble, France. <sup>16</sup>Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia. <sup>17</sup>Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, Göttingen, Germany. <sup>18</sup>Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, Spain. <sup>19</sup>Department of Marine Microbiology and Biogeochemistry, Royal Netherland Institute for Sea Research (NIOZ), and Utrecht University, Den Burg, The Netherlands. <sup>20</sup>Biological Faculty, M. V. Lomonosov Moscow State University, Moscow 119991, Russia. <sup>21</sup>N. N. Semenov Institute of Chemical Physics, Russian Academy of Sciences, Moscow 119991, Russia. <sup>22</sup>Laboratories of Excellence, Ion Channel Science and Therapeutics, 06560 Valbonne, France. <sup>23</sup>M. M. Shemyakin-Yu. A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow 117997, Russia. <sup>24</sup>Institut für Physikalische Biologie, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>25</sup>Max Planck Institute of Biophysics, Frankfurt am Main, Germany. <sup>26</sup>These authors contributed equally: Dmitrii Zabelskii, Alexey Alekseev, Kirill Kovalev. ✉email: [valentin.gordeliy@ibs.fr](mailto:valentin.gordeliy@ibs.fr)





OPEN

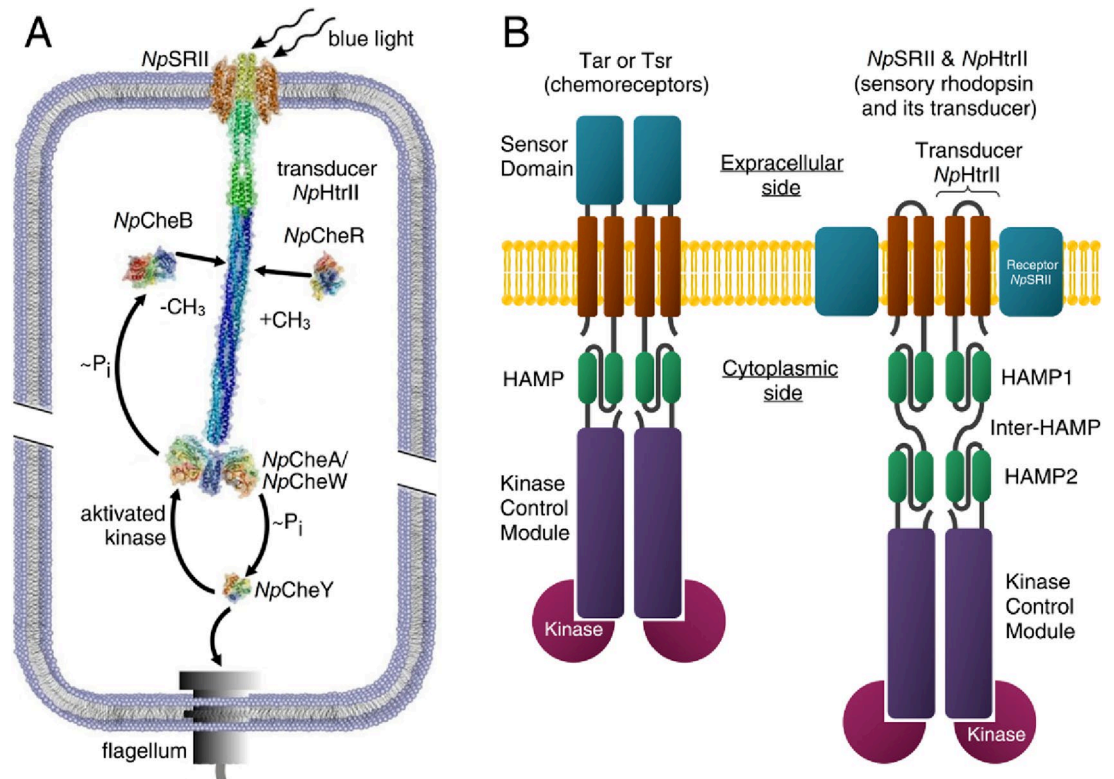
## Molecular model of a sensor of two-component signaling system

Yury L. Ryzhykau<sup>1,11</sup>, Philipp S. Orekhov<sup>1,11</sup>, Maksim I. Rulev<sup>2,3,4,11</sup>, Alexey V. Vlasov<sup>1,5</sup>, Igor A. Melnikov<sup>1,2</sup>, Dmytro A. Volkov<sup>3,4</sup>, Mikhail Yu. Nikolaev<sup>1,3</sup>, Dmitrii V. Zabelskii<sup>1,3,4</sup>, Tatiana N. Murugova<sup>1,6</sup>, Vladimir V. Chupin<sup>1</sup>, Andrey V. Rogachev<sup>1,6</sup>, Andrey Yu. Gruzinov<sup>7</sup>, Dmitri I. Svergun<sup>7</sup>, Martha E. Brennich<sup>8</sup>, Ivan Yu. Gushchin<sup>1</sup>, Montserrat Soler-Lopez<sup>1,2</sup>, Arne Bothe<sup>9</sup>, Georg Büldt<sup>1</sup>, Gordon Leonard<sup>1,2</sup>, Martin Engelhard<sup>9</sup>, Alexander I. Kuklin<sup>1,6</sup>✉ & Valentin I. Gordeliy<sup>1,3,4,10</sup>✉

Two-component systems (TCS) are widespread signaling systems present in all domains of life. TCS typically consist of a signal receptor/transducer and a response regulator. The receptors (histidine kinases, chemoreceptors and photoreceptors) are often embedded in the membrane and have a similar modular structure. Chemoreceptors were shown to function in highly ordered arrays, with trimers of dimers being the smallest functional unit. However, much less is known about photoreceptors. Here, we use small-angle scattering (SAS) to show that detergent-solubilized sensory rhodopsin II in complex with its cognate transducer forms dimers at low salt concentration, which associate into trimers of dimers at higher buffer molarities. We then fit an atomistic model of the whole complex into the SAS data. The obtained results suggest that the trimer of dimers is "tripod"-shaped and that the contacts between the dimers occur only through their cytoplasmic regions, whereas the transmembrane regions remain unconnected.

Two-component systems (TCS) are present in all domains of Life. They are the most common signaling systems in prokaryotes but are absent in mammals, making them potential antimicrobial drug targets<sup>1</sup>. A recent work indicated that TCS networks expressed in mammalian cells might pave the way for orthogonal signaling<sup>2</sup>. However, this will require more atomic level detail concerning their structure/function relationships. In this regard, detailed information has been obtained for chemotaxis and photoreceptor TCS of motile enteric bacteria and archaea, respectively<sup>3,4</sup>. This family of TCS consist of a transmembrane receptor, which interacts with a His-kinase (CheA) and adaptor proteins (CheW) attached at their cytoplasmic tips. The response regulators CheY and CheB are phosphorylated by CheA and function either as a switch factor for the flagellar motor (CheY) or as an adaptation component (CheB)<sup>5</sup> (Fig. 1A). Chemotaxis receptors of motile bacteria form homodimers comprised of extracellular sensor domain and transmembrane regions, and a long rod-shaped cytoplasmic domain. For photoreceptors, the sensor domain constitutes a microbial rhodopsin which forms a 2:2 complex with its cognate transducer (Htr). Transducers have a long cytoplasmic domain, similarly to chemoreceptors, with which they share a high level of homology<sup>4,6</sup> (Fig. 1B). In literature, the term "sensor" usually means the first component of the two-component system. In the case of chemotaxis, it is a chemoreceptor, which has the so-called sensory

<sup>1</sup>Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, 141700 Dolgoprudny, Russia. <sup>2</sup>Structural Biology Group, European Synchrotron Radiation Facility, 71 Avenue des Martyrs, 38000 Grenoble, France. <sup>3</sup>Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich, 52425 Jülich, Germany. <sup>4</sup>JuStruct: Jülich Center for Structural Biology, Forschungszentrum Jülich, 52428 Jülich, Germany. <sup>5</sup>Institute of Crystallography, University of Aachen (RWTH), Jaegerstrasse 17-19, 52056 Aachen, Germany. <sup>6</sup>Frank Laboratory of Neutron Physics, Joint Institute for Nuclear Research, 141980 Dubna, Russia. <sup>7</sup>European Molecular Biology Laboratory, Hamburg Outstation, 22607 Hamburg, Germany. <sup>8</sup>Synchrotron Crystallography Team, EMBL Grenoble Outstation, 71 Avenue des Martyrs, 38042 Grenoble, France. <sup>9</sup>Department Structural Biochemistry, Max Planck Institute of Molecular Physiology, 44227 Dortmund, Germany. <sup>10</sup>Institut de Biologie Structurale Jean-Pierre Ebel, Université Grenoble Alpes-Commissariat à l'Energie Atomique et aux Energies Alternatives-CNRS, 38027 Grenoble, France. <sup>11</sup>These authors contributed equally: Yury L. Ryzhykau, Philipp S. Orekhov and Maksim I. Rulev. ✉email: kuklin@nf.jinr.ru; g.valentin@fz-juelich.de



**Figure 1.** Signal transduction pathway in case of the two-component phototaxis system of *Natronomonas pharaonis*<sup>5</sup> and domain architecture of membrane chemo- and photoreceptors of TCS. (A) Light activated sensory rhodopsin II (*NpSR11*) induces conformational and/or dynamical changes in the transducer (*NpHtrII*), which are converted by two HAMP domains and conveyed along the 200 Å long transducer to the tip region. Activated by the transducer histidine kinase CheA (bound to the adapter protein CheW) undergoes auto-phosphorylation and further transfers the phosphate group to the response regulators CheY or CheB. CheY affects the rotational bias of the flagellar motor, while the methyltransferase CheR controls the adaptation mechanism. (B) Cartoon representations of the chemoreceptor dimer (Tar and Tsr in complex with kinases) from *E. coli* and of the photosensor dimer of the complex of the sensory rhodopsin II with its cognate transducer *NpHtrII* and kinases from *N. pharaonis*.

domain. In the case of a photoreceptor, the sensory domain is represented by a sensory rhodopsin. Thus, the complex of a sensory rhodopsin with the transducer plays the role of a sensor of TCS. We use the term "sensor" in our work meaning that it is the first part of the phototaxis TCS.

The cytoplasmic domain of Htr has a modular design comprising a transmembrane region, two HAMP domains (connected by a helical inter-HAMP region<sup>7</sup>) and an extended adaptation module bearing reversible methylation sites and containing a Gly-hinge. The adjacent signaling region harbors the binding sites for the adaptor protein CheW and the histidine kinase CheA. Interestingly, chemoreceptors from enteric bacteria possess only one HAMP-domain<sup>8</sup> in contrast to archaeal transducers<sup>9</sup>.

Despite a wealth of biochemical data, the molecular mechanism of the signal transduction by TCSs is still to be fully revealed. A common understanding concerns the alternating dynamics of the cytoplasmic domains. Both chemoreceptors and Htrs display different dynamics in adjoining modules, which has been correlated with the signal transfer along the cytoplasmic rod<sup>5,10</sup>. A possible role of hinges between the modules could be the steric accommodation of the dimers into trimers of dimers and oligomers of higher order<sup>11,12</sup>. These trimers of dimers constitute the functional unit capable of activating CheA<sup>13</sup>. Together with CheA and CheW, trimers form the core unit of the extended signaling arrays<sup>14</sup>, compact membrane-attached assemblages of signaling units responsible for amplification of the incoming stimulus<sup>15–17</sup>.

An important *sine qua non* of a full understanding of the mechanism of signal transfer along the trimers of dimers is a high-resolution structural information for a full-length TCS sensor. However, three-dimensional, near atomic-resolution X-ray structures have only been published for fragments of histidine kinases<sup>9,18–20</sup> and chemoreceptors<sup>21–23</sup>, and for a truncated phototaxis-transducer complex from *N. pharaonis*<sup>24</sup>. For this latter, NMR data are also available<sup>25</sup>. Also, several atomistic models of *E. coli* chemoreceptors were built based on low-resolution EM data, particularly, the homodimer of the aspartate receptor Tar<sup>11</sup>, the trimers-of-dimers of the cytoplasmic domains of serine chemoreceptors in different signaling states<sup>26</sup>, and the chemosensory array core signaling unit formed by a mixture of receptors with different adaptational modifications<sup>27</sup>. This insufficiency of high-resolution descriptions of the structures of full-length TCS sensors is probably due to the inherent flexibility of their cytoplasmic domains<sup>5,10,11</sup>. However, small-angle scattering (SAS) is a low-resolution method, which

can be highly efficient when combined with available high-resolution structures of protein fragments using computer modeling<sup>28,29</sup>. An important advantage of SAS is that structural data can be collected for solubilized protein complexes in conditions close to their native environment.

*NpSR*II/*NpHtr*II mediates negative phototaxis in halobacterial *N. pharaonis* and represents a model TCS sensor and, similarly to chemoreceptors<sup>30</sup>, the *NpSR*II/*NpHtr*II complex forms trimers of dimers in the *N. pharaonis* membrane<sup>31</sup>. *N. pharaonis* grows optimally at 3.5 M NaCl<sup>32</sup>. It has been shown that structure and oligomerization state of the *NpSR*II/*NpHtr*II strongly depend on salt concentration<sup>33</sup>. Here, we describe the full-length structure of a sensory rhodopsin II/transducer complex (*NpSR*II/*NpHtr*II) by integrating published high-resolution structural data with SAS measurements. Structural information was gained for *NpSR*II in complex with the transmembrane domain of *NpHtr*II, which show a dimeric organization in different activated states<sup>24,34,35</sup>. For the rest of the *NpHtr*II, homologues were chosen including the HAMP domains of Aer2 cytoplasmic sensor<sup>23</sup> and for the hypothetical transmembrane receptor AF1503<sup>36</sup>, sensor histidine kinase NarQ<sup>18–20,37</sup>, kinase control modules including the methyl-accepting chemotaxis protein I (Tsr) from *E. coli*<sup>22</sup> and the chemotaxis protein from *Thermatoga maritima*<sup>38,39</sup>. Our work both confirms the dependence of oligomeric state on ionic strength and provides molecular models of the dimeric and "trimer of dimers" forms of the full-length complex. In our manuscript, we used the term "molecular model" meaning a model, which is constructed by molecular modeling with the high-resolution structures of the fragments of full-length protein (transmembrane domain, HAMP1 and HAMP2 domains and Kinase control module) under low resolution SAS constraints and verified with available biochemical and biophysical literature data. The results obtained allow us to suggest a "tripod"-shaped model for the full-length *NpSR*II/*NpHtr*II trimer of dimers in which the dimers associate solely through contacts between their cytoplasmic domains.

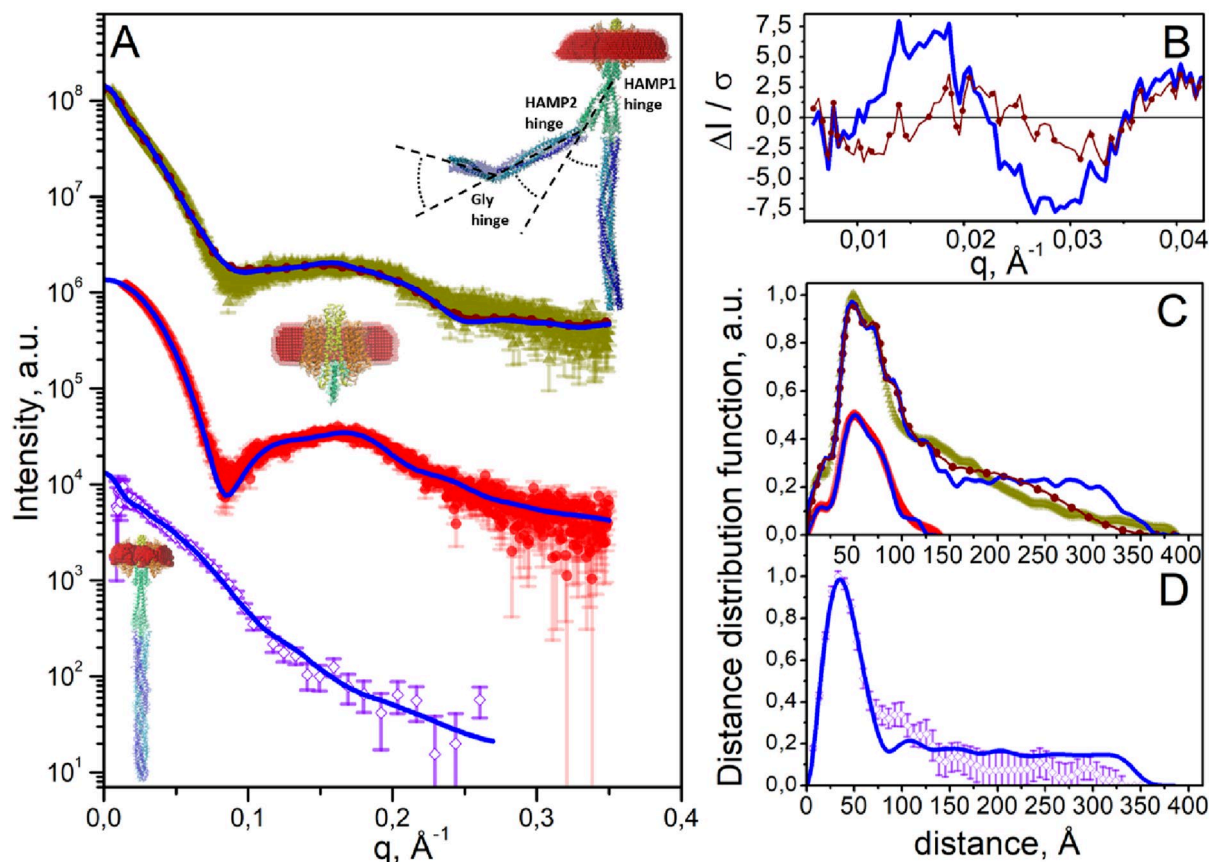
### Ionic strength dependence of the oligomerization state of the *NpSR*II/*NpHtr*II complex

In order to investigate conditions, under which trimers of dimers are formed, we performed SAXS (small angle X-ray scattering), SANS (small angle neutron scattering) and CD (circular dichroism) experiments (Fig. S1) at different salt concentrations on solutions of detergent solubilized [n-dodecyl- $\beta$ -D-maltopyranoside (DDM)] full-length *NpSR*II/*NpHtr*II as well as with truncated *NpSR*II/*NpHtr*II<sub>137</sub>. *NpHtr*II<sub>137</sub> comprises the transmembrane region (1–83 a. a.) and the HAMP1 domain (84–136 a. a.) of *NpHtr*II.

At low salt conditions SAXS scattering curves for the truncated complex were fitted well ( $\chi^2 = 1.5$ ) using a model of a dimer of *NpSR*II/*NpHtr*II<sub>137</sub> surrounded by a detergent belt (see "Methods" for the details) (Fig. 2A, middle, Table S1), strongly suggesting that, under the conditions employed, interactions between the membrane and/or HAMP1 domains of *NpHtr*II<sub>137</sub> are sufficient to induce dimerization.

A fit based on an optimized dimer model (see "Methods" for the modeling details) containing a "straight" adaptational domain (Fig. 2A, bottom) showed reasonable ( $\chi^2 = 0.7$ ) agreement with experimental low salt condition SANS data (this also applies to the matching of the distance distribution functions, see Fig. 2D). However, complementary SAXS experiments, performed as for *NpSR*II/*NpHtr*II<sub>137</sub> at a synchrotron source, suggested that the optimized model featuring the "straight" adaptation domain does not fully describe the overall structure of a full-length *NpSR*II/*NpHtr*II dimer. Here, a fit between the experimental scattering curve and that calculated from the "straight transducer" model yields  $\chi^2 = 5.1$ . This can be improved ( $\chi^2 = 3.4$ ) using a calculated scattering curve obtained by combining a series of models in which bends of between  $-90^\circ$  and  $90^\circ$  (with the step of  $30^\circ$ ) were induced in the HAMP1-, HAMP2- and Gly-regions (see "Methods", Fig. 2, Fig. S2 and Supplementary Text document S1 for details). This procedure produced a noticeably improved fit of the distance distribution functions calculated from the experimental and theoretical scattering curves, particularly in the 160–370 Å region (Fig. 2C). This range of distances corresponds in reciprocal space to the range of  $q < 0.04 \text{ \AA}^{-1}$ , in which there is a significant improvement in agreement with experimental SAXS data when the flexible hinges model is used (see Fig. 2B). This observation suggests presence of multiple conformational states of full-length *NpSR*II/*NpHtr*II dimers that can be described as structural polydispersity of the model that is in line with EM studies<sup>11</sup> showing that bending of Tar chemoreceptor dimers occurs at flexible HAMP and/or Gly hinges. Moreover, it has also been suggested that this bending may play a crucial role in the formation of the hexagonal arrays of membrane receptors in complex with kinases CheA and CheW<sup>12</sup>. Nevertheless, we cannot rule out that such conformational flexibility may be a result of low ionic strength in which the transducer domain is partially disordered. Indeed, in line with the results of other studies<sup>33,40</sup>, CD-spectra of the full-length *NpSR*II/*NpHtr*II complex under different salt conditions (Fig. S1) suggest partial disordering of *NpSR*II/*NpHtr*II dimers.

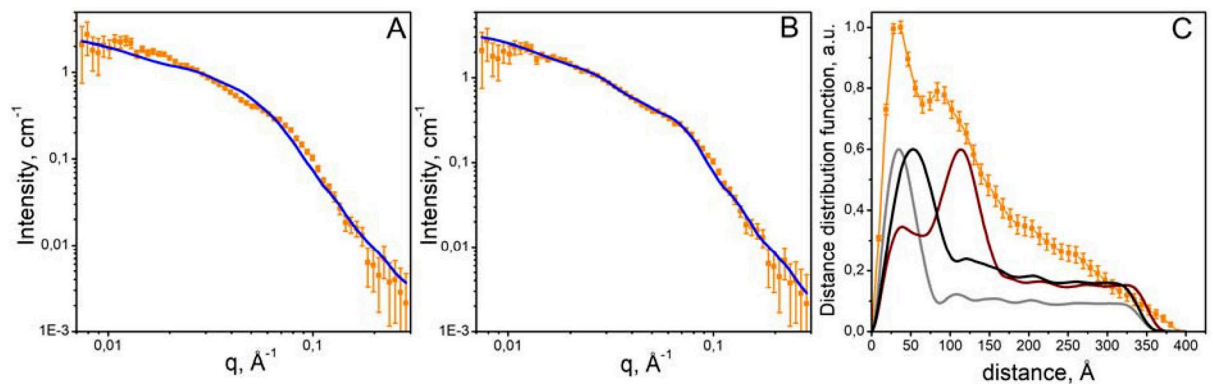
While both SAXS and SANS experiments clearly suggest that *NpSR*II/*NpHtr*II mainly forms dimers at low ionic strength, this is not the case at increasing salt concentrations. Figure 3A shows the SANS scattering curve for the full-length *NpSR*II/*NpHtr*II measured in a buffer containing 4.0 M NaCl. This cannot be satisfactorily matched by a theoretical curve calculated from the polydisperse dimers described above and to obtain a reasonable fit it was necessary to calculate a theoretical scattering curve based upon a mixture *NpSR*II/*NpHtr*II dimers and trimers of dimers. Two different forms of trimers of dimers were constructed for use in our calculations. Firstly, similar to a previously proposed model of the trimer of dimers<sup>5</sup>, the direction of bending angles at the HAMP and Gly hinges was such that inter-dimer contacts are induced both between the transmembrane regions of dimers and their cytoplasmic tips (see Fig. 4A, Fig. S3). For the second model, the direction of bending angles at the HAMP and Gly hinges was introduced such that the transmembrane regions at the base of the trimer of dimers adopt a tripod-type disposition and inter-dimer contacts are formed only between the tips of the cytoplasmic domains of each dimer (Fig. 4B). In both cases, the detergent belt was applied as described for the SANS analysis at the low salt (see above, Fig. 2A, bottom). Fitting the experimental SANS curve obtained for *NpSR*II/*NpHtr*II at 4.0 M NaCl to a theoretical curve based on a mixture of *NpSR*II/*NpHtr*II dimers and trimers of



**Figure 2.** Dimerization of *NpSR II/NpHtr II*<sub>137</sub> and of full-length *NpSR II/NpHtr II* at low salt concentration. (A) Bottom—experimental SANS curve for *NpSR II/NpHtr II* (purple hollow rhombus) and CRYSON fit ( $\chi^2 = 0.7$ , corresponding blue line) based on a model (shown on the left of the curves) of an *NpSR II/NpHtr II* dimer with a detergent corona. Middle—experimental SAXS curve for the *NpSR II/NpHtr II*<sub>137</sub> dimer (red circles) and MEMPROT fit ( $\chi^2 = 1.5$ , corresponding blue line) based on a model (shown near the corresponding curve) of the *NpSR II/NpHtr II*<sub>137</sub> dimer with the detergent corona. Top—experimental SAXS curve for full-length *NpSR II/NpHtr II* (dark yellow triangles) and two theoretical approximations. The first (blue solid line) is a CRYSON fit ( $\chi^2 = 5.1$ ) based on a "straight" model of the *NpSR II/NpHtr II* dimer. The second (brown pointed line) is a fit ( $\chi^2 = 3.4$ ) based on a combination of modified models of the *NpSR II/NpHtr II* dimer (see "Methods", Fig. S2 and Text document S1 for details). While the difference between the two fits is not immediately apparent, the improved fit of the second approximation is evident when considering of the relative residuals of the fit in the region in reciprocal space  $q < 0.04 \text{ \AA}^{-1}$  (see B), that corresponds to the distances  $> 160 \text{ \AA}$  in a real space. For this range of distances, discrepancy between the experimental distance distribution function and theoretical one obtained for the "straight" model of the *NpSR II/NpHtr II* dimer is evident (see C). In the representations of the atomic models, the detergent belt is shown in red. (B) Relative residuals of theoretical approximations and experimental SAXS data obtained for full-length *NpSR II/NpHtr II* (A, top). Data related to the "straight" model is shown as blue solid line; data related to the combination of modified models is shown as brown pointed line. (C) Distance distribution functions calculated from the SAXS curves shown in (A) (designations are the same as in A). (D) Distance distribution functions calculated from SANS data shown in (A) (designations are the same as in A).

dimers, inter-dimer contacts in which are induced both between the transmembrane regions of dimers and their cytoplasmic tips (Fig. 4A), resulted in  $\chi^2 = 5.4$  (Fig. 3A). In contrast, fitting the same SANS curve to a theoretical curve based on a mixture of *NpSR II/NpHtr II* dimers and "tripod"-shaped trimers of dimers (Fig. 4B) resulted in  $\chi^2 = 1.3$  (Fig. 3B). Pair distance distribution function calculated from experimental SANS data obtained at 4.0 M NaCl has two peaks (see Fig. 3C). The position of the first peak ( $r \sim 30 \text{ \AA}$ ) allows one to interpret it as a peak from dimers, while the position of the second peak ( $r \sim 100 \text{ \AA}$ ) is in a good agreement with the position of the peak corresponding to the "tripod"-shaped model of the trimer of dimers (see Fig. 3C). The conformation of trimer of dimers of the *NpSR II/NpHtr II* SANS experiments at NaCl concentrations 1.4 M and 2.8 M (Fig. S4) produced similar results but indicated an increasing fraction of tripod-style trimers of dimers with increasing salt concentration. Resulting weight fractions of "tripod"-shaped trimers of dimers are  $18 \pm 3\%$ ,  $28 \pm 2\%$  and  $36 \pm 2\%$  for 1.4 M, 2.8 M and 4.0 M, respectively (Fig. S5, Table S2). These results support that the formation of *NpSR II/NpHtr II* trimers of dimers at high salt concentration is likely to occur through the contacts of the cytoplasmic tips of dimers with their transmembrane parts remaining unconnected (Fig. 4B).





**Figure 3.** SANS curve for *NpSRII/NpHtrII* at 4.0 M NaCl and corresponding distance distribution functions. (A) Experimental scattering curve for *NpSRII/NpHtrII* at 4.0 M NaCl (orange squares) fitted with  $\chi^2=5.5$  to a theoretical curve (blue line) calculated for a mixture of *NpSRII/NpHtrII* dimers and trimers of dimers which inter-dimer contacts are induced both between the transmembrane regions of dimers and their cytoplasmic tips (Fig. 4A). (B) Experimental scattering curve for *NpSRII/NpHtrII* at 4.0 M NaCl (orange squares) fitted with  $\chi^2=1.3$  to a theoretical curve (blue line) calculated for a mixture of *NpSRII/NpHtrII* dimers and "tripod"-shaped trimers of dimers (Fig. 4B). (C) Distance distribution function calculated from the experimental curve shown in (A,B) (orange squares), and theoretical distance distribution functions of the dimers (grey line), the "tripod"-shaped (Fig. 4B) trimers of dimers (brown line) and the "transmembrane-bound" conformation of trimer of dimers of the *NpSRII/NpHtrII* (Fig. 4A) (black line). For greater clarity, distance distribution functions were normalized to obtain maximum values of 1.0 for the experimental and 0.6 for the theoretical curves.

### A molecular model of the full-length *NpSRII/NpHtrII* complex

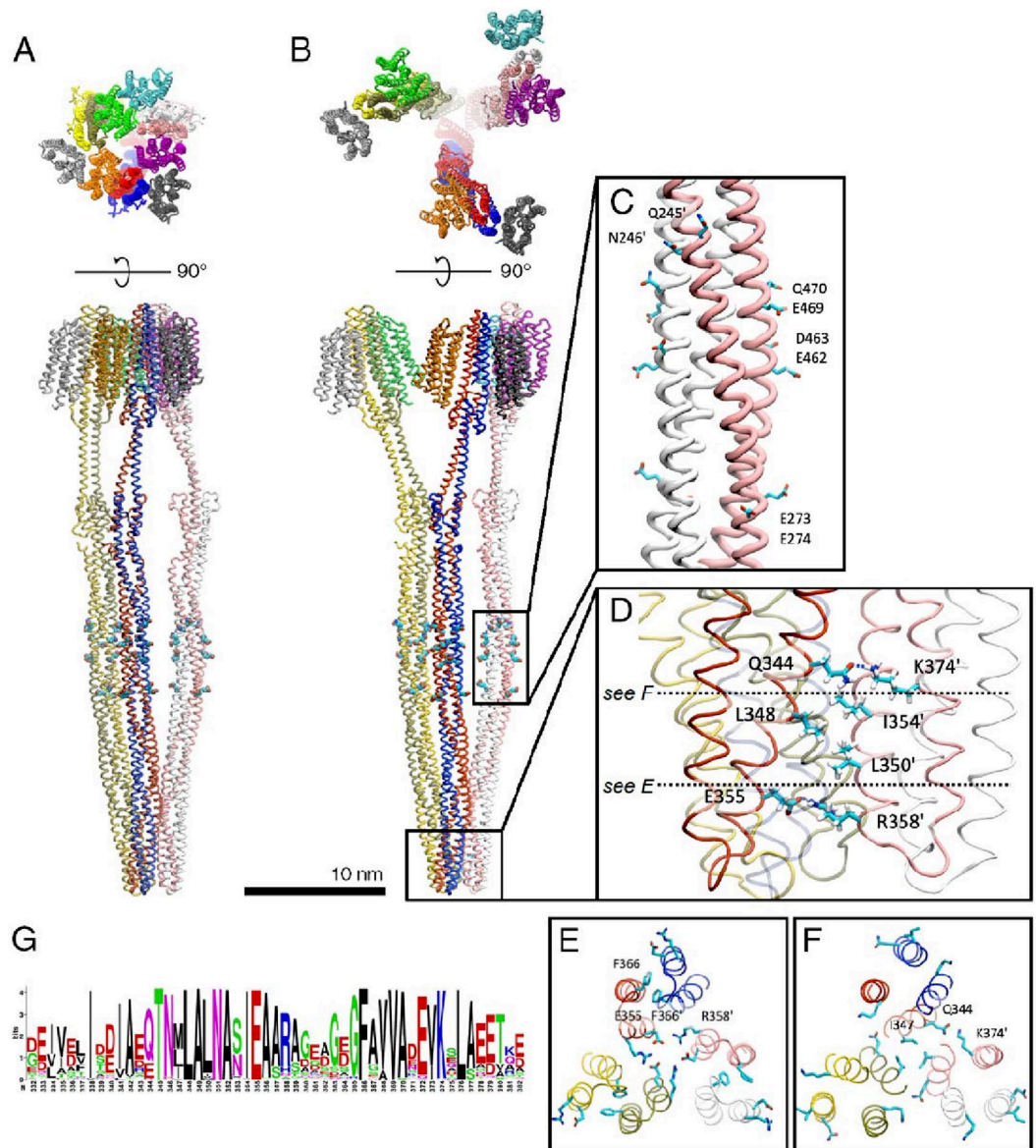
The tripod-like model of the structure of the *NpSRII/NpHtrII* trimer of dimers (Fig. 4B) which provides the best fit to the experimental SANS scattering curve obtained at 4.0 M NaCl (Fig. 3B) was constructed based on the high-resolution structures of individual fragments of the TCSs sensors from *E. coli*, *N. pharaonis*, and *T. maritima*<sup>18,22,24,34,35,37–39</sup>. The initial model obtained was further optimized using molecular dynamics simulations (see "Methods"). In the resulting model, the transmembrane domains of individual dimers are separated but their cytoplasmic tips form a tight interface. As expected, given that modelling of the cytoplasmic domain was based on the X-ray structure of the highly conserved interface of *E. coli* Tsr<sup>22</sup>, dimer-dimer interactions (Fig. 4D) are mediated by a number of inter-residue contacts involving amino acids conserved across homologs of *NpHtrII* (see logo plot in Fig. 4G). For example, E355 of one dimer forms a salt bridge with R358' of a neighboring dimer (Fig. 4E); a compact hydrophobic patch is formed by L348, I354' and L350'; hydrogen bonds are formed between Q344 and K374' (Fig. 4F). In addition, stacking between the aromatic rings of F366-F366' stabilizes the intra-dimer interface similar to bacterial chemoreceptors<sup>41</sup> (Fig. 4E).

### Discussion

In present study, the combination of SAS and molecular modeling has allowed us to propose a molecular model of the full-length phototaxis complex from the halophilic archaeon *N. pharaonis* and to reveal the effect of salt concentration on its oligomeric state and dynamics.

We verified our molecular model with available biochemical and biophysical literature data. The monomer–monomer interactions between transducers are in accordance with the following high-resolution structural data. The dimer of the transmembrane domain is based on its available structure (PDB code 1h2s<sup>24</sup>). Dimers of HAMP1 and HAMP2 domains are constructed according to homology with HAMP domain of NarQ from *E. coli* (PDB code 5jeq<sup>18</sup>). The cytoplasmic domain (kinase control module) is presented by the coiled-coil of two antiparallel helices connected by a "U-turn", as it was shown for cytoplasmic domains of TsrQ from *E. coli* (PDB code 1qu7<sup>22</sup>), MCP from *Termostoga maritima* (PDB code 2ch7<sup>38</sup>), etc. In these cases, coiled-coil structure stabilized by interactions between the extended interfaces consisting of hydrophobic amino acids, mainly leucines. Dimer–dimer interactions (Fig. 4D) are mediated by a number of inter-residue contacts involving amino acids conserved across homologs of *NpHtrII* (see logo plot, Fig. 4G).

Positions of highly conservative residues 345–355 in the cytoplasmic tip domain of *NpHtrII* in our molecular model of the *NpSRII/NpHtrII* trimer of dimers are in a good agreement with the EPR studies by Orban-Glaß et al.<sup>43</sup>. In that work, the dynamic and structural properties of the cytoplasmic tip domain of *NpHtrII* were investigated using site-directed spin labeling electron paramagnetic resonance spectroscopy. Distance distributions obtained by double electron–electron resonance typically have three peaks, which can be interpreted as three characteristic inter-spin distances that correspond to three groups of inter-residue distances (see Fig. S6A,B). This finding supports the fact that archaeal photoreceptor/transducer complexes form trimers of dimers analogous to methyl-accepting chemotaxis proteins at high salt concentrations. Comparison of the distance between the residues labeled with a spin label showed that the distance between the nearest isoleucines I347 decreases with an increase in the salt concentration from 500 mM to 2 M NaCl (Fig. S6C). This rearrangement indicates that



**Figure 4.** Optimized molecular models of the *NpSRII/NpHtrII* trimer of dimers. (A,B) Molecular models of the "transmembrane-bound" (A) and the "tripod"-shaped (B) trimer of dimers. Individual polypeptide chains are colored differently. Putative methylation sites are represented as spheres. (C) Putative methylation sites in the methyl-accepting region of a single dimer of the complex. (D) Inter-dimer contacts within the highly conservative tip region of the cytoplasmic domain of the "tripod"-shaped trimer of dimers. Key amino acid residues are shown with those belonging to the partnering dimer labeled by apostrophe. (E) Cross-section view highlighting the formation of ionic locks between E355-R358' (homologous to E385 and R388' from Tsr<sup>42</sup> and stacking between F366 (equivalent to F396 of Tsr) and F366'. (F) Cross section depicting contacts between conservative K374' and Q344 residues, and hydrophobic contacts between I347. (G) Logo plot showing conservation of residues in the tip region involved in the formation of the trimer contacts.

a trimer of dimers is formed with I347 facing inside the resulting structure (see Fig. 4E, Fig. S6A), playing an important role in the formation of inter-dimer contact due to hydrophobic interaction.

At low ionic strength (150 mM NaCl), our SAS data are most compatible with a dimeric conformation for both the truncated *NpSRII/NpHtrII*<sub>137</sub> and for full-length *NpSRII/NpHtrII*. In the latter, the very long cytoplasmic domain of the dimeric *NpHtrII* transducer appear to be highly flexible, corroborating the intrinsically dynamic nature of chemo- and photoreceptors described in some recent publications (e.g.<sup>10,40,44</sup>). Coupled with the conclusions of a previous study that the cytoplasmic domain of *NpHtrII*'s (*pHtrII*-cyt) does not form dimers at low ionic strength<sup>33</sup>, our results imply that only the transmembrane region is required for dimerization of *NpSRII/NpHtrII*. However, this observation is in contrast to that obtained for a similar construct (*NpSRII/NpHtrII*<sub>157</sub> containing HAMP1 and inter-HAMP (137–156 a. a.) domains). This might be due to a lower detergent concentration used in the current experiments.



While the minimum construct required for *NpSRII/NpHtrII* to form dimers may still be a matter for debate, our SAS data obtained at low ionic strength clearly suggest that full-length *NpSRII/NpHtrII* forms dimers under such conditions (see Fig. 2A). In order to improve the agreement between theoretical and experimental scattering curves data, we generated models of different possible conformations of the *NpSRII/NpHtrII* dimers which, while maintaining an elongated shape for the *NpHtrII* dimers, considered both bending at its flexible hinges<sup>11</sup> and partial unfolding<sup>33,40</sup> of this domain at low salt conditions. Our results confirm a highly dynamic nature of the transducer dimer at low salt: the root mean square deviations of bending angles from the zero value are in the range from 50° to 60° for all the three HAMP1-, HAMP2-, and Gly- hinges (see Fig. S2).

For the cytoplasmic domain of the *NpHtrII* (234–504 a. a.), Budyak et al.<sup>33</sup> also observed a partial random coil configuration at low salt concentrations. Due to the difficulty to unequivocally determine the amount of random coil structures, these authors argued in light of their SANS in favor of a highly dynamic helical rod at low salt conditions. This finding has been confirmed by subsequent results published in the work<sup>40</sup>, according to which *pHtrII*-cyt has a strong propensity for helix. Firstly, *pHtrII*-cyt has a high helix propensity inferred from the sequence; *pHtrII*-cyt is predicted to be a helical coiled-coil based on a homology to the Tst-cyt<sup>22</sup>. Secondly, despite that at low ionic strength deconvolution of CD-spectra the secondary structure values are 77% unstructured, 20% turns and  $\beta$ -sheets, and only 3%  $\alpha$ -helices, it was also shown by control experiments with helix-inducing solvent 2,2,2-trifluoroethanol (TFE) that a gradual rise in negative ellipticity at 222 nm occurs with increasing amounts of TFE. Thus, despite "random-coil" spectroscopic features, *pHtrII*-cyt can exist as a highly flexible, loosely packed but folded helical coiled-coil. The results of the comparison of CD spectra from the work<sup>40</sup> with and without TFE suggest that, despite the presence of  $\beta$ -sheets in the deconvolution of CD spectra (see Fig. S1B),  $\beta$ -sheets do not necessarily have to be present in the final molecular model. Moreover, numerous algorithms for the estimation of the secondary structure composition from the CD spectra often fail to provide acceptable results on  $\alpha/\beta$ -mixed or  $\beta$ -structure-rich proteins due to spectral variety and lower spectral amplitudes of the  $\beta$ -structures<sup>45,46</sup>. For these reasons, the secondary structure of the molecular model proposed in our work is predominantly  $\alpha$ -helical.

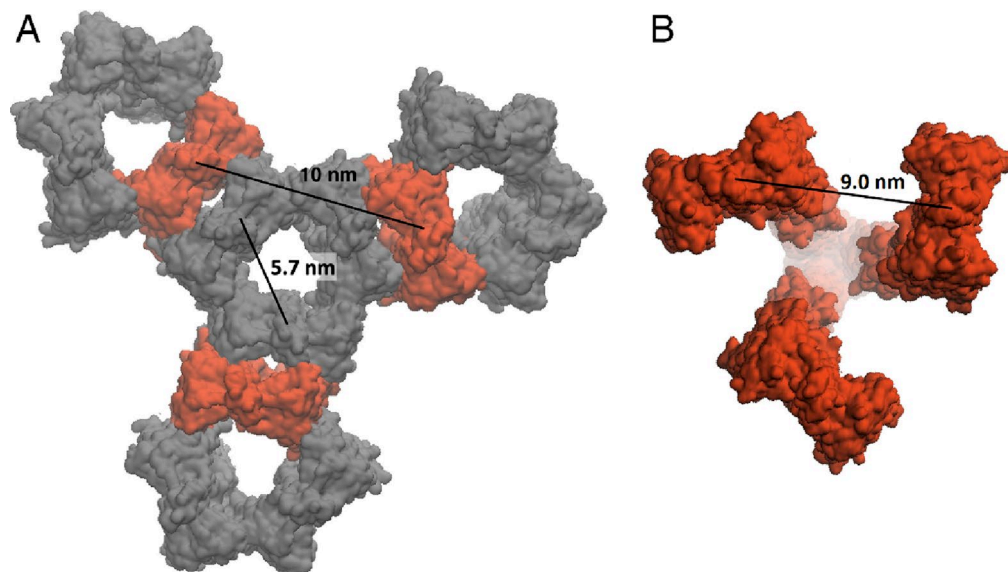
Despite the fact that isolated *pHtrII*-cyt is in monomeric form under conditions of low salt concentration<sup>33</sup>, this domain forms dimers under these conditions when the *NpSRII/NpHtrII* complex is full length. The coiled-coil does not disappear, which is confirmed by the results of work<sup>43</sup>, where the distances between residues for the cytoplasmic tip of the full-length *NpSRII/NpHtrII* studied by the EPR method at low salt concentration are the same as they should be in a dimer (see Fig. S6C). These are the reasons why we used the model where the protein fragments between the flexible hinges present elongated (not globular) structures. Additionally far-UV CD data (Fig. S1) obtained in our work under low salt conditions showed conformational disordering of *NpSRII/NpHtrII*. The Kratky plots (Fig. S7) also confirm this: the obtained curves have maximum values at  $q R_G \sim 4$  to 7, which are higher than expected for globular particles and are typical for elongated and/or flexible proteins<sup>47,48</sup>. After the maximum, the curves show a tendency to decrease to zero, which excludes complete unfolding of proteins<sup>49</sup>.

It has been shown that the chimeric proteins *NpSRII-NpHtrII-Tar* and *NpSRII-NpHtrII-Tsr*, which contain the transmembrane region and a truncated HAMP1 domain (a. a. 1–125) of the *NpSRII/NpHtrII* complex and the cytoplasmic domain of either *StTar* or *EcTsr*, can mediate phototaxis in *E. coli*<sup>50</sup>. This suggests that both chimeric complexes are able to transduce signal at low salt conditions. This, taken together with our results, implies that destabilization of the remaining fragments of the cytoplasmic domains of which are not included in these chimera is the main reason for blocking the formation of the trimers of dimers of native *NpSRII/NpHtrII* at low ionic strength. It also suggests that at high salt concentration corresponding to the physiological range of halophilic archaea, a reordering of the cytoplasmic domains takes place and this allows the dimers to assemble into trimers of dimers, implying that the transmembrane region of *NpSRII/NpHtrII* complex either is not sufficient to mediate formation of trimers-of-dimers or, consistent with our "tripod"-shaped model, is not involved into the trimerization of dimers.

The observed salt concentration-induced structural changes of the *NpSRII/NpHtrII* system may have two possible biological 'roles'. As pointed above, haloarchaeon *N. pharaonis* lives in the highly saline environment and the complex must be optimized to these conditions. If the phototaxis system would remain functional under the low salt conditions, it could drive the microorganisms to regions with optimal insolation regardless of the salt concentration that would ultimately lead to their death. On the other hand, we cannot also completely exclude that disordering (i.e. decreasing of secondary structure elements and increasing of conformational flexibility) of the complex may generate signal allowing the archaea to move towards higher salt concentration and avoid environments with low salinity. The salt-driven equilibrium between dynamic and compact conformations, which was observed for the *NpHtrII* HAMP1 domain by EPR<sup>44</sup>, supports this idea.

SANS scattering curves at higher salt concentrations (Fig. 3) clearly indicate the formation of a trimer of dimers, the fraction of which increases as does the ionic strength. Moreover, the best fit to the SANS data at higher ionic strength (Fig. 3B) is obtained by modelling a tripod-like shape, in which only the cytoplasmic tips of *NpHtrII* dimers are involved in inter-dimer contacts (Fig. 4B). This is in agreement with previous EPR studies suggesting that *NpSRII/NpHtrII* dimers may form oligomers of higher order (e.g., trimers of dimers) due to interactions between the transducer tips alone<sup>43</sup>. In our model, inter-dimer separations in the regions containing putative sites (Fig. 4C) of methylation/demethylation (determined by homology with *HsHtrII* of *Halobacterium salinarum*<sup>7</sup>, see alignment in Text document S2) vary between 30 and 40 Å (Fig. S8). Thus, they are all potentially accessible to modifications by the chemotaxis methyltransferase CheR and methyltransferase CheB (the radii of gyration calculated for CheR and CheB from *Salmonella typhimurium* (PDB IDs 1AF7 and 1A2O) being approximately 21 Å and 20 Å, respectively).

To the date, there is very little structural information available for full-length *E. coli* Tar/Tsr chemoreceptors or other chemoreceptors in either their demethylated (OFF) or methylated (ON) states. Our experiments provide a complete description of the demethylated complex (OFF) while a recent study by Burt et al.<sup>27</sup> describes a model



**Figure 5.** Inter-dimer distances between transmembrane parts of the *NpSRII/NpHtrII* dimers. (A) Characteristic distances between dimers in 2D-array proposed in<sup>31</sup>. (B) The "tripod"-shaped model of the trimer of dimers (Fig. 4B) demonstrating the inter-dimer distance of 9.0 nm.

based on a mixture of receptors with a wide range of adaptational modifications. In the later study, the authors were able to obtain the 3D cryo-ET map of the full-length *E. coli* chemoreceptor array together with the CheW/CheA baseplate in micelles with the local resolution varying between  $\sim 15$  and  $30 \text{ \AA}$ . They further constructed all-atom models of the *E. coli* CheA.P3.P4.P5 dimer, CheW monomer, and membrane-bound, full-length Tsr homodimer using coordinates from existing high-resolution crystallographic structures where available. The two models, share remarkable similarity in the tip region, where a number of key interactions appear concurrent despite the fact that they apparently correspond to different signaling states. This is very likely due to inherent bias of the both models towards the 1QU7 structure<sup>22</sup>, which was obtained for the QQQQ Tsr analogue (and thus it likely corresponds to the ON-state) which was used to model the trimeric contacts of the cytoplasmic tips of dimers in both studies. On the other hand, it stems from the high structural conservation of cytoplasmic receptor arrays between Bacteria and Archaea<sup>51</sup>. However, the lower local resolution of the periplasmic and transmembrane regions of the model report in<sup>27</sup> (apparently due to the large separation between the well-resolved cytoplasmic regions of the Tsr/Tar array and its periplasmic ligand-binding domains, combined with the relative flexibility of the cytoplasmic methylation helix bundle of the receptors) and the lack of homology between these fragments of bacterial chemoreceptors and the *NpSRII/NpHtrII* restrains further comparison of our results with the model of Burt et al.

Trimers of dimers are essential for the formation of large membrane signaling arrays of both photo- and chemoreceptors<sup>14</sup>. The absence of CheA/CheW in our experiments does not allow us to identify oligomers larger than trimers of dimers; however, our structure provides potential routes for the formation of the membrane arrays from the "tripod"-shaped trimers of dimers. One of the possibilities is that the transmembrane sensory domains of dimers within each trimer of dimers are able to form a trimeric contact in a native environment and are further packed with the neighboring trimers of dimers<sup>31</sup>. On the other hand, the "tripod"-shaped model for the trimer of dimers is also compatible with lattice models suggesting that trimeric contacts at the cytoplasmic part do not match trimeric contacts between the transmembrane domains (Fig. 5). Both scenarios seem feasible, taking into account the plasticity of dimers at the three flexible hinges discussed above, though the *summae* of evidences existing to the date supports the latter one. Further investigations should shed light on structural organization of photoreceptor arrays.

Here a combination of SAXS and SANS allowed us to study conditions of the trimer-of-dimers formation from the individual *NpSRII/NpHtrII* dimers. SANS experiments allowed us to propose a "tripod"-shaped structural model for trimers of dimers of full-length *NpSRII/NpHtrII* in which contacts between dimers occur only at the tips of the cytoplasmic regions, leaving the membrane domains unhinged. In the case of SAXS the change of the salt concentration corresponds to a wide range of the scattering length density (SLD) of the buffer, which would result in a contrast variation simultaneously with the changes of the oligomeric state. Such data would be difficult to treat, also because of a detergent belt for which the contrast variation is a big challenge itself. An electron density of a buffer varies in a range from  $0.334 \text{ e/\AA}^3$  (pure  $\text{H}_2\text{O}$  or  $\text{D}_2\text{O}$ ) to  $\sim 0.37 \text{ e/\AA}^3$  (4 M NaCl in  $\text{H}_2\text{O}$ ), that means the changes of about  $\sim 11\%$ ; electron densities in a protein/detergent system are  $\sim 0.42$ ,  $0.275$  and  $0.515 \text{ e/\AA}^3$  for a protein, hydrophobic and hydrophilic parts of a detergent belt, respectively. It means that  $\Delta\text{SLD}$  values for different components of detergent/protein complexes have different sign and their ratio critically depends on SLD of the buffer, i.e. on salt concentration. This leads to changes in SAXS  $I(q)$  profiles caused by this salt concentration changes. In contrast to SAXS data, in the case of SANS the SLD of the solvent is slightly

changed with the increase of the salt concentration [see Table S2a). SLD of a D<sub>2</sub>O buffer changes from  $6.404 \times 10^{-6} \text{ \AA}^{-2}$  to  $6.187 \times 10^{-6} \text{ \AA}^{-2}$  at the increase of salinity from zero to 4 M (see the section (a) of the Table S2)], that means about 3%. The SLDs of a protein and the hydrophobic and hydrophilic parts of a detergent belt are about  $2.1 \times 10^{-6} \text{ \AA}^{-2}$ ,  $-0.4 \times 10^{-6} \text{ \AA}^{-2}$  and  $3.9 \times 10^{-6} \text{ \AA}^{-2}$ , respectively, i.e.  $\Delta \text{SLD} < 0$  and have the same sign. It allows treating SANS data with focusing on a detailed distribution of oligomeric state, which has a more significant influence on a SANS 1D scattering profile than the changes in SLD contrasts at different salt concentrations have.

Our small angle scattering experiments open up prospects for further successful use of this technique for studies of the TCS sensors. Some studies report huge conformational changes of the tertiary structure of the transducer after activation of the photosensor. They are manifested by a change of the diffusion coefficient (and, consequently, of the Stokes radius) of the protein complex and can be explained by unfolding of HAMP domains during signal transfer<sup>52,53</sup>. The changes in the Stokes radius, as well as in the radius of gyration, could be easily detected and characterized using SAS (direct measurements by SAXS). Furthermore, small-angle scattering experiments could provide detailed information about the shape of the molecule, including those arising from changes in tertiary and quaternary structure and/or conformational rearrangements. These changes could, for example, be monitored/characterized via standard SAS experiments with a continuously light-irradiated sample<sup>54</sup>, or using time-resolved SAXS technique<sup>55,56</sup> applied for the protein solution after the short flash of light. For the particular case studied here there are known mutations of the *NpSR*II (D75N) and in the *NpHtr*II (G83F), which can be used for this purpose<sup>57</sup> and such experiments may help to obtain detailed structural information concerning the mechanism of activation of *NpSR*II/*NpHtr*II trimer of dimers.

## Methods

**Cloning.** *NpSR*II (UniProt ID P42196) and *NpHtr*II (UniProt ID P42259) genes were optimized for expression in *Escherichia coli* using GeneArt services<sup>58</sup>. The Strep-tag II (WSHPQFEK with ENS-linker) and 6xHis-tag (with ENS-linker as well) were added to C-termini of the *NpSR*II and *NpHtr*II respectively. The modified *NpSR*II and *NpHtr*II genes were cloned into pSCodon1.2 vector using *Nhe*I-AccIII and *Nde*I-*Nhe*I restriction enzymes (FastDigest series, ThermoFisher, Massachusetts, USA), respectively. The gene of the truncated transducer *NpHtr*II<sub>137</sub> was obtained from *NpHtr*II by PCR. Positive clones were identified by restriction analysis and confirmed by sequencing. *NpSR*II and *NpHtr*II (or *NpHtr*II<sub>137</sub>) genes were then cloned in tandem into the co-expression vector pET27bmod<sup>59</sup> via *Bgl*II-*Not*I and *Not*I-*Blp*I restriction sites. *E. coli* strain Top10 (Invitrogen, Carlsbad, CA, USA) was used throughout.

**Co-expression of *NpSR*II and *NpHtr*II<sub>137</sub>.** The non-fused protein complex was co-expressed in *E. coli* cells, strain BL21(DE3) (Stratagene, La Jolla, CA, USA). Briefly, the culture was grown at 37 °C in Terrific Broth (TB) media with 100 mM of Na/K-Pi (pH 6.7), 25 mM of ammonium sulfate and 100 mg/L of kanamycin. Expression was induced with 1 mM isopropyl-β-D-galactopyranoside (IPTG), at OD<sub>600</sub> = 1.6–2.0. Simultaneously, a concentrated solution of all-*trans* retinal in ethanol was added to a final concentration 10 μM retinal. Cells were further cultivated for 4 h at 37 °C.

**Co-expression of *NpSR*II and full-length *NpHtr*II.** For SAXS measurements, the non-fused protein complex was co-expressed in *E. coli* strain BL21-AI (Invitrogen, Carlsbad, CA). The culture was grown at 37 °C in TB media with 70 mM of Na/K-Pi (pH 6.7) and 50 mg/L of kanamycin. Expression was induced with the mixture of 0.1% arabinose and 2 mM IPTG, at OD<sub>600</sub> = 1.6–2.0. Simultaneously, all-*trans* retinal solution in ethanol was added to a final concentration 10 μM, and cells were further cultivated 4 h at 37 °C. For SANS measurements, the non-fused protein complex was co-expressed in *E. coli* strain BL21(DE3). The culture was grown in TB-5052 media<sup>60</sup> with 100 mM of Na/K-Pi (pH 6.7), 25 mM of ammonium sulfate and 100 mg/L of kanamycin. The cells were incubated at 37 °C until OD<sub>600</sub> reached 1.0–1.2, when all-*trans* retinal solution in ethanol was added to final concentration 10 μM, and cells were further cultivated overnight at 20 °C as was described in<sup>61</sup> for expression of the *NpSR*II.

**Co-purification of *NpSR*II and full-length *NpHtr*II (or *NpHtr*II<sub>137</sub>).** After expression, cells were pelleted in 25 mM Na/Na-Pi buffer (pH 8.0) containing 150 mM NaCl, 1 mM PMSF and cOMplete protease inhibitor cocktail (Roche, Switzerland) and lysed using Microfluidizer M-110P (Microfluidics, Massachusetts, USA). After centrifugation, the pellet was solubilized in 1% of n-dodecyl-β-D-maltopyranoside (DDM). The solubilized protein mixture was then purified via Ni-NTA (HisTrap HP 5 ml column, GE Healthcare, Illinois, USA) and size-exclusion chromatography (Superose 6 10/300 GL column 24 ml, GE Healthcare, USA). A typical gel-filtration profile is shown in Fig. S9A. Laemmli 12% SDS-PAGE of purified *NpSR*II/*NpHtr*II is shown in Fig. S10 (PageRuler Plus Prestained Protein Ladder (Thermo Scientific, Catalog #26619) was used as a molecular weight standard). The protocols for co-expression and co-purification were based on procedures used in<sup>62,63</sup>. The procedure yielded about 5 mg and 1.5 mg per 1 L of cell culture for *NpSR*II/*NpHtr*II<sub>137</sub> and *NpSR*II/*NpHtr*II, respectively.

**Small-angle X-ray scattering measurements.** The majority of SAXS measurements were carried on BM29 beamline (ESRF, Grenoble, France)<sup>64</sup>. All measurements were performed with 100% of beam intensity at a wavelength of 0.9918 Å (12.5 keV). Initial data processing was performed automatically using the EDNA pipeline<sup>65,66</sup>. Additional SAXS experiments were done on the BioSAXS beamline P12 (PETRA III, DESY, Hamburg, Germany)<sup>67</sup>.

For the full-length *NpSR*II/*NpHtr*II and for the truncated complex *NpSR*II/*NpHtr*II<sub>137</sub>, SAXS profiles were obtained for the samples with protein concentrations of 0.57 and 0.78 mg/ml, correspondingly; exposure time was

7 and 10 s, respectively. For all SAXS measurements, peak fractions after gel-filtration were used<sup>68</sup>. See Table S1 for other details of SAXS measurements.

**Small-angle neutron scattering measurements.** SANS measurements were performed on the YuMO spectrometer (IBR-2, Dubna, Russia) with two-detector system<sup>69,70</sup>. Raw data were processed with program SAS<sup>71</sup>.

For SANS measurements, *NpSRII/NpHtrII* sample was divided after gel filtration into two parts (sample A and B). For each part, gel-filtration was used to replace H<sub>2</sub>O with D<sub>2</sub>O (Fig. S9A). The buffer for sample A contained 150 mM NaCl, 25 mM Na/Na-Pi (pD<sup>+</sup> = 8.0)<sup>72</sup>, 1 mM EDTA, 0.05% DDM. Sample B was dissolved in 4.0 M NaCl, 100 mM Na/Na-Pi (pD<sup>+</sup> = 8.0), 1 mM EDTA, 0.05% DDM.

67.5% (v/v) of the stock, containing protein in 150 mM NaCl and 32.5% (v/v) of the stock, containing protein in 4.0 M NaCl were taken to prepare the sample with the *NpSRII/NpHtrII* in 1400 mM NaCl; 31.2% (v/v) of the stock, containing protein in 150 mM NaCl and 68.8% (v/v) of the stock, containing protein in 4.0 M NaCl were taken to prepare the sample with the *NpSRII/NpHtrII* in 2800 mM NaCl.

The concentrations of the *NpSRII/NpHtrII* for the SANS measurements were 0.51, 0.33, 0.31, and 0.40 mg/ml in the buffers containing 150, 1400, 2800, and 4000 mM NaCl, correspondingly. Total exposure time was 2 h for heavy water solutions of the *NpSRII/NpHtrII* complex at 150 and 4000 mM NaCl and 3.5 h for solutions at 1400 and 2800 mM NaCl. See Table S2 for other details of SANS measurements.

**Molecular modeling.** Initial model of the full-length *NpSRII/NpHtrII* dimer was generated from the existing high resolution crystal structures of *T. maritima* MCP, PDB code 2ch7<sup>38</sup>, *E. coli* NarQ receptor, PDB code 5jeq<sup>18</sup> and truncated *N. pharaonis* *NpSRII/NpHtrII* complex, PDB code 1h2s<sup>24</sup> for the cytoplasmic domain, HAMP domains and transmembrane region, respectively, using template-based homology modeling in SWISS-MODEL<sup>73</sup>. The missing fragments (primarily, the inter-HAMP region, which was predicted to be an  $\alpha$ -helix<sup>74</sup>) were modeled ab initio as ideal helices.

The obtained all-atom model of the full-length *NpSRII/NpHtrII* dimer was a subject for further optimization by MD simulation during 100 ns with a number of constraints applied. The regions homologous to those resolved by X-ray crystallography were guided towards the initial coordinates by means of harmonic steering forces ( $k_{\text{constr}} = 150 \text{ kcal/mol/\AA}^2$ ). At each timestep, the RMS distance between the current coordinates and the target structure were computed after aligning the target structure to the current coordinates. The alignment and evaluation of the steering forces was independently done for the transmembrane region, HAMP domains and the cytoplasmic domain. In addition, the inter-HAMP region and the short fragments connecting the TM2 of *NpHtrII* and HAMP1 and AS2 of HAMP2 and the cytoplasmic domain were restrained in  $\alpha$ -helical conformation by means of harmonic dihedral restraints.

The initial model for trimer-of-dimers was built by aligning the optimized dimer model to the model of trimeric oligomer of the transmembrane region of *NpSRII/NpHtrII* predicted by the SymDock protein-protein docking web-service with the C3 symmetry constraints<sup>75</sup>. This model was further optimized using a 100 ns long MD simulation with the transmembrane region restrained at its initial coordinates and the cytoplasmic domain steered ( $k_{\text{steer}} = 200 \text{ kcal/mol/\AA}^2$ ) to the homology model of the highly conserved trimeric interface of *E. coli* Tsr receptor resolved by X-ray crystallography, PDB code 1qu7<sup>22</sup>.

Finally, in order to obtain the "tripod"-shaped conformation of the trimer-of-dimer we run another round of steered MD, in which the cytoplasmic tip of the complex was constrained to the crystallographic contacts similar to the previous simulation while the transmembrane regions of individual dimers were gradually repulsed from their joint center-of-masses using the *colvar* feature of NAMD ( $k_{\text{spring}} = 200 \text{ kcal/mol/\AA}^2$ ). The simulation time of this run was also 100 ns.

All all-atom MD simulations were carried out using NAMD 2.9<sup>76</sup> and CHARMM27 force field<sup>77</sup>. The simulations were run in NVT ensemble (maintained by the Langevin thermostat, T = 303.15 K) using Generalized Born implicit solvent model (GBIS). The ionic strength in the simulations were set to 4.0 M. The timestep of 2 fs was used. The production simulations were prefaced by energy minimization using the steepest descent (5000 steps). In all simulations of the trimer-of-dimers, the three-fold symmetry was maintained by the symmetry constraint as realized in NAMD 2.9.

**SAS data processing.** SAXS and SANS profiles  $I(q)$  were processed using ATSAS<sup>78</sup> and BioXTAS RAW<sup>79</sup> software suites. The protein concentrations were small, consequently the structural factor influence to scattering curves was negligible<sup>80,81</sup>. For calculation of values of  $\epsilon$ , molecular mass,  $\bar{v}$  and SLD from sequence, programs ProtParam<sup>82</sup>, Peptide Property Calculator<sup>83</sup> and SLD calculator web (<https://sld-calculator.appspot.com/>) were used (see Tables S1, S2). Distance distribution functions  $P(r)$  and regularized  $I(q)$  were obtained using GNOM program, which realizes the method of Indirect-Fourier Transform (IFT)<sup>84</sup>. Values of  $R_G$  and  $I(0)$  (Tables S1, S2) were calculated from  $P(r)$  and using Guinier approximations (see Fig. S11). CRY SOL and CRYSON programs were used for evaluating the solution scattering from macromolecules and fitting it to experimental small-angle scattering curves<sup>85,86</sup>. OLIGOMER program<sup>87</sup> from ATSAS software suite was used for a set of curves calculated for dimers and trimers of dimers using CRYSON to fit an experimental scattering curve from a two-component mixture of dimers and trimers of dimers of the *NpSRII/NpHtrII* to validate different variants of their molecular models and to find the volume fractions of each component in the mixture. MEMPROT software<sup>88</sup> was used to generate pseudo-atomic model of the detergent belt of the transmembrane part of the protein and to fit experimental SAXS curves using a model combined from the detergent belt pseudo-atomic model and protein atomic model. Before MEMPROT running, the center of the transmembrane part of the protein was placed at the origin (zero) and direction of the normal vector to the membrane plane was set along the z-axis using PPM



web server<sup>89</sup>. The MEMPROT settings included CRY SOL3 option and protein surface algorithm 2. See Supplementary Information (Tables S1, S2) for other details of SAS data treatment<sup>90</sup>.

**Pseudo-atomic model of the detergent belt.** Pseudo-atoms CH<sub>3</sub> и NH<sub>3</sub> generated by MEMPROT software simulate 9/97 of the X-ray scattering length of hydrophobic tail (C<sub>12</sub>H<sub>25</sub>) and 10/181 of the X-ray scattering length of hydrophilic head (C<sub>12</sub>H<sub>21</sub>O<sub>11</sub>) of DDM molecule correspondingly, proportionally to their numbers of electrons. To calculate theoretical SANS curves of the protein with the DDM belt, pseudo-atoms NH<sub>3</sub> generated by MEMPROT were renamed to CH<sub>3</sub>, and perdeuteration parameter in CRYSON settings were set to 0.1058 and 0.3903 for chains corresponding to hydrophobic and hydrophilic parts of the DDM belt respectively. Described procedure provides neutron scattering lengths – 1.27 fm (9/97 of C<sub>12</sub>H<sub>25</sub>) and 7.62 fm (10/181 of C<sub>12</sub>H<sub>14</sub>D<sub>7</sub>O<sub>11</sub>) for pseudo-atoms related to hydrophobic and hydrophilic parts of the detergent belt respectively, which corresponds to the same scattering length fractions as they are in case of X-ray scattering length. Here, it is taken into account that the DDM head has seven hydrogens exchanging to deuterium in D<sub>2</sub>O. These procedure provides neutron SLD values of – 0.388 × 10<sup>-6</sup> Å<sup>-2</sup> and 3.92 × 10<sup>-6</sup> Å<sup>-2</sup> for the hydrophobic core and hydrophilic face of the detergent belt, respectively.

**Fitting of the SAXS profile of the full-length complex using the model of the flexible hinges.** Firstly, SAXS data for the full-length complex *NpSR*II/*NpHtr*II (Fig. 2A, top) were approximated with the solution scattering evaluated from the atomistic model of the "straight" *NpSR*II/*NpHtr*II dimer with pseudo-atomic model of the detergent belt generated by MEMPROT software<sup>88</sup>, and  $\chi^2$  of the fit was 5.1. Secondly, we generated modified atomic models of the *NpSR*II/*NpHtr*II dimer with the bends at the HAMP1-, HAMP2- and Gly- hinges from – 90° to 90° each (with the step of 30°), and add pseudo-atoms imitating detergent belt obtained on previous step to them. Thirdly, we evaluated theoretical SAXS profiles from the modified *NpSR*II/*NpHtr*II dimer models with the detergent belt using CRY SOL3 software<sup>78</sup>. Fourthly, we approximated experimental SAXS profile for the full-length *NpSR*II/*NpHtr*II as a combination of the scattering profiles from the modified *NpSR*II/*NpHtr*II dimer models with the detergent belt using Tikhonov regularization method (see Text document S1 for the details). For fitting of this data, Wolfram Mathematica software<sup>91</sup> was used. Analogous analysis of the protein polydispersity based on SAS data was realized in works<sup>92,93</sup>. SAXS could be used in studying of highly polydisperse macromolecules, making it possible to obtain data that are in good agreement with other structural methods such as electron microscopy<sup>94</sup>.

**UV–VIS spectroscopy.** Absorption spectra for the protein solutions were acquired by using a UV-2450 UV–VIS Spectrophotometer (Shimadzu, Kyoto, Japan). Protein solutions were placed in a 1 mm quartz cell (Weiju, Lianyungang, China). To estimate concentration of the photoactive complex, the value  $\epsilon^{498\text{nm}} = 45,500 \text{ M}^{-1} \text{ cm}^{-1}$  of the *NpSR*II extinction coefficient was used, and the sensory rhodopsin and its transducer were assumed to be equimolar. Absorption spectrum of the sample of the full-length *NpSR*II/*NpHtr*II complex in D<sub>2</sub>O buffer with 150 mM NaCl is shown in Fig. S9B.

**CD spectroscopy.** CD spectra were acquired by using a J-1100 CD Spectrometer (Jasco, Easton, MD, USA). For CD measurements, the same samples were used as for SANS measurements. Protein solutions were placed in a 1 mm quartz cell (Weiju, Lianyungang, China). All spectra were recorded with a bandwidth of 1.0 nm, scan speed of 50 nm/min, and digital integration time of 1.0 s. For the protein at 150 and 4000 mM NaCl, five accumulations were averaged for each spectrum, for 1400 and 2800 mM of NaCl, averaging of eight accumulations was done. The quantification of secondary structures was analyzed by Dichroweb<sup>95,96</sup> using K2D program<sup>97</sup>.

### Data availability

The SAS data were deposited with SASBDB (<http://sasbdb.org>). SAXS data deposited with accession codes SASDKZ6 and SASDK27 for the *NpSR*II/*NpHtr*II<sub>137</sub> and the full-length *NpSR*II/*NpHtr*II at 150 mM NaCl, correspondingly. SANS data deposited with accession codes SASDK37, SASDK47, SASDK57, and SASDK67 for the full-length *NpSR*II/*NpHtr*II at 0.15 M, 1.4 M, 2.8 M and 4.0 M, respectively. Detailed data validation metrics related to SASBDB depositions are placed in the Table S3. PBD files for molecular models of *NpSR*II/*NpHtr*II<sub>137</sub> dimer and full-length *NpSR*II/*NpHtr*II dimer and trimer of dimers are available in the SASBDB depositions SASDKZ6, SASDK27, and SASDK67, respectively. Other data supporting the findings of this manuscript are available from the corresponding author upon reasonable request.

Received: 12 September 2020; Accepted: 21 April 2021

Published online: 24 May 2021

### References

1. Bem, A. E. *et al.* Bacterial histidine kinases as novel antibacterial drug targets. *ACS Chem. Biol.* **10**, 213–224 (2015).
2. Mazé, A. & Benenson, Y. Artificial signaling in mammalian cells enabled by prokaryotic two-component system. *Nat. Chem. Biol.* **16**, 179–187 (2020).
3. Hazelbauer, G. L., Falke, J. J. & Parkinson, J. S. Bacterial chemoreceptors: High-performance signaling in networked arrays. *Trends Biochem. Sci.* **33**, 9–19 (2008).
4. Klare, J. P., Chizhov, I. & Engelhard, M. Microbial rhodopsins: Scaffolds for ion pumps, channels, and sensors. *Results Probl. Cell Differ.* **45**, 73–122 (2008).
5. Orekhov, P. S. *et al.* Signaling and adaptation modulate the dynamics of the photosensory complex of *Natronomonas pharaonis*. *PLoS Comput. Biol.* **11**, e1004561 (2015).

6. Hoff, W. D., Jung, K. H. & Spudich, J. L. Molecular mechanism of photosignaling by archaeal sensory rhodopsins. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 223–258 (1997).
7. Koch, M. K., Staudinger, W. F., Siedler, F. & Oesterhelt, D. Physiological sites of deamidation and methyl esterification in sensory transducers of *Halobacterium salinarum*. *J. Mol. Biol.* **380**, 285–302 (2008).
8. Aravind, L. & Ponting, C. P. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol. Lett.* **176**, 111–116 (1999).
9. Gushchin, I. & Gordeliy, V. Transmembrane signal transduction in two-component systems: Piston, scissoring, or helical rotation?. *BioEssays* **40**, 1700197 (2018).
10. Bartelli, N. L. & Hazelbauer, G. L. Differential backbone dynamics of companion helices in the extended helical coiled-coil domain of a bacterial chemoreceptor. *Protein Sci* **24**, 1764–1776 (2015).
11. Akkaladevi, N., Bunyak, F., Stalla, D., White, T. A. & Hazelbauer, G. L. Flexible hinges in bacterial chemoreceptors. *J. Bacteriol.* **200**(200), e00593-e617 (2018).
12. Stalla, D., Akkaladevi, N., White, T. A. & Hazelbauer, G. L. Spatial restrictions in chemotaxis signaling arrays: A role for chemoreceptor flexible hinges across bacterial diversity. *Int. J. Mol. Sci.* **20**, 2989 (2019).
13. Li, M., Khursigara, C. M., Subramaniam, S. & Hazelbauer, G. L. Chemotaxis kinase CheA is activated by three neighbouring chemoreceptor dimers as effectively as by receptor clusters. *Mol. Microbiol.* **79**, 677–685 (2011).
14. Li, M. & Hazelbauer, G. L. Core unit of chemotaxis signaling complexes. *Proc. Natl. Acad. Sci. USA* **108**, 9390–9395 (2011).
15. Parkinson, J. S., Hazelbauer, G. L. & Falke, J. J. Signaling and sensory adaptation in *Escherichia coli* chemoreceptors: 2015 update. *Trends Microbiol.* **23**, 257–266 (2015).
16. Sourjik, V. & Armitage, J. P. Spatial organization in bacterial chemotaxis. *EMBO J.* **29**, 2724–2733 (2010).
17. Bray, D., Levin, M. D. & Morton-Firth, C. J. Receptor clustering as a cellular mechanism to control sensitivity. *Nature* **393**, 85–88 (1998).
18. Gushchin, I. *et al.* Mechanism of transmembrane signaling by sensor histidine kinases. *Science (80-)* **356**, 20 (2017).
19. Gushchin, I., Melnikov, I., Polovinkin, V., Ishchenko, A. & Gordeliy, V. Crystal structure of a proteolytic fragment of the sensor histidine kinase NarQ. *Curr. Comput. Aided Drug Des.* **10**, 149 (2020).
20. Gushchin, I. *et al.* Sensor histidine kinase NarQ activates via helical rotation, diagonal scissoring, and eventually piston-like shifts. *Int. J. Mol. Sci.* **21**, 3110 (2020).
21. Milburn, M. V. *et al.* Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand. *Science (80-)* **254**, 1342–1347 (1991).
22. Kim, K. K., Yokota, H. & Kim, S. H. Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. *Nature* **400**, 787–792 (1999).
23. Airola, M. V., Watts, K. J., Bilwes, A. M. & Crane, B. R. Structure of concatenated HAMP domains provides a mechanism for signal transduction. *Structure* **18**, 436–448 (2010).
24. Gordeliy, V. I. *et al.* Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* **419**, 484–487 (2002).
25. Etzkorn, M. *et al.* Complex formation and light activation in membrane-embedded sensory Rhodopsin II as seen by solid-state NMR spectroscopy. *Structure* **18**, 293–300 (2010).
26. Yang, W. *et al.* In situ conformational changes of the *Escherichia coli* serine chemoreceptor in different signaling states. *MBio* **10**, e00973-e1019 (2019).
27. Burt, A. *et al.* Complete structure of the chemosensory array core signalling unit in an *E. coli* minicell strain. *Nat. Commun.* **11**, 1–9 (2020).
28. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285 (2007).
29. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* **2**, 207–217 (2015).
30. Zhang, P., Khursigara, C. M., Hartnell, L. M. & Subramaniam, S. Direct visualization of *Escherichia coli* chemotaxis receptor arrays using cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA* **104**, 3777–3781 (2007).
31. Orekhov, P. *et al.* Sensory Rhodopsin I and sensory Rhodopsin II form trimers of dimers in complex with their cognate transducers. *Photochem. Photobiol.* **93**, 796–804 (2017).
32. Falb, M. *et al.* Living with two extremes: Conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome Res.* **15**, 1336–1343 (2005).
33. Budyak, I. L. *et al.* Shape and oligomerization state of the cytoplasmic domain of the phototaxis transducer II from *Natronobacterium pharaonis*. *Proc. Natl. Acad. Sci. USA* **103**, 15428–15433 (2006).
34. Moukhametzianov, R. *et al.* Development of the signal in sensory rhodopsin and its transfer to the cognate transducer. *Nature* **440**, 115–119 (2006).
35. Ishchenko, A. *et al.* New insights on signal propagation by sensory Rhodopsin II/transducer complex. *Sci. Rep.* **7**, 41811 (2017).
36. Ferris, H. U. *et al.* The mechanisms of HAMP-mediated signaling in transmembrane receptors. *Structure* **19**, 378–385 (2011).
37. Melnikov, I. *et al.* Fast iodide-SAD phasing for high-throughput membrane protein structure determination. *Sci. Adv.* **3**, e1602952 (2017).
38. Li, X. *et al.* The 3.2 Å resolution structure of a receptor: CheA:CheW signaling complex defines overlapping binding sites and key residue interactions within bacterial chemosensory arrays. *Biochemistry* **52**, 3852–3865 (2013).
39. Cassidy, C. K. *et al.* CryoEM and computer simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling. *Elife* **4**, e08419 (2015).
40. Budyak, I. L. *et al.* Flexibility of the cytoplasmic domain of the phototaxis transducer II from *Natronomonas pharaonis*. *J. Biophys.* **2008**, 1–11 (2008).
41. Ortega, D. R. *et al.* A phenylalanine rotameric switch for signal-state control in bacterial chemoreceptors. *Nat. Commun.* **4**, 1–8 (2013).
42. Gao, Q., Cheng, A. & Parkinson, J. S. Conformational shifts in a chemoreceptor helical hairpin control kinase signaling in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **116**, 15651–15660 (2019).
43. Orban-Glaß, I. *et al.* Clustering and dynamics of phototransducer signaling domains revealed by site-directed spin labeling electron paramagnetic resonance on SRII/HtrII in membranes and nanodiscs. *Biochemistry* **54**, 349–362 (2015).
44. Doeberl, M. *et al.* Salt-driven equilibrium between two conformations in the HAMP domain from *Natronomonas pharaonis*: The language of signal transfer?. *J. Biol. Chem.* **283**, 28691–28701 (2008).
45. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890 (2007).
46. Micsonai, A. *et al.* Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA* **112**, E3095–E3103 (2015).
47. Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod–Debye law. *Biopolymers* **95**, 559–571 (2011).
48. Hammel, M. Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). *Eur. Biophys. J.* **41**, 789–799 (2012).



49. Brennich, M., Pernot, P. & Round, A. How to analyze and present SAS data for publication. In *Advances in Experimental Medicine and Biology*, Vol 1009 47–64 (Springer, 2017).
50. Jung, K. H., Spudich, E. N., Trivedi, V. D. & Spudich, J. L. An archaeal photosignal-transducing module mediates phototaxis in *Escherichia coli*. *J. Bacteriol.* **183**, 6365–6371 (2001).
51. Briegel, A. *et al.* Structural conservation of chemotaxis machinery across Archaea and Bacteria. *Environ. Microbiol. Rep.* **7**, 414–419 (2015).
52. Inoue, K., Sasaki, J., Spudich, J. L. & Terazima, M. Laser-induced transient grating analysis of dynamics of interaction between sensory rhodopsin II D75N and the HtrII transducer. *Biophys. J.* **92**, 2028–2040 (2007).
53. Klose, D. *et al.* Light-induced switching of HAMP domain conformation and dynamics revealed by time-resolved EPR spectroscopy. *FEBS Lett.* **588**, 3970–3976 (2014).
54. Gapchenko, A. A. *et al.* Light-induced structure changes in Bacteriorhodopsin D96N by SAXS. *J. Bioenerg. Biomembr.* **50**, 540–540 (2018).
55. Cammarata, M. *et al.* Tracking the structural dynamics of proteins in solution using time-resolved wide-angle X-ray scattering. *Nat. Methods* **5**, 881–886 (2008).
56. Gustavsen, R. L. *et al.* Time resolved small angle X-ray scattering experiments performed on detonating explosives at the advanced photon source: Calculation of the time and distance between the detonation front and the X-ray beam. *J. Appl. Phys.* **121**, 105902 (2017).
57. Yang, C. S. & Spudich, J. L. Light-induced structural changes occur in the transmembrane helices of the *Natronobacterium pharaonis* HtrII transducer. *Biochemistry* **40**, 14207–14214 (2001).
58. Raab, D., Graf, M., Notka, F., Schödl, T. & Wagner, R. The GeneOptimizer algorithm: Using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. *Syst. Synth. Biol.* **4**, 215–225 (2010).
59. Klostermeier, D., Seidel, R. & Reinstein, J. Functional properties of the molecular chaperone DnaK from *Thermus thermophilus*. *J. Mol. Biol.* **279**, 841–853 (1998).
60. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
61. Bratanov, D. *et al.* An approach to heterologous expression of membrane proteins. The case of bacteriorhodopsin. *PLoS One* **10**, e0128390 (2015).
62. Ryzhykau, Y. L. *et al.* Trimers of dimers of SRII/HtrII full complex. Small angle scattering structural investigation. *FEBS J.* **284**, 154 (2017).
63. Ryzhykau, Y. L. *et al.* Small angle scattering structural study of the NpSRII/HtrII complex. *J. Bioenerg. Biomembr.* **50**, 577–578 (2018).
64. Pernot, P. *et al.* Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Rad.* **20**, 660–664 (2013).
65. Brennich, M. E. *et al.* Online data analysis at the ESRF bioSAXS beamline, BM29. *J. Appl. Crystallogr.* **49**, 203–212 (2016).
66. Incardona, M. F. *et al.* EDNA: A framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Radiat.* **16**, 872–879 (2009).
67. Blanchet, C. E. *et al.* Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **48**, 431–443 (2015).
68. Zabelskii, D. V. *et al.* Ambiguities and completeness of SAS data analysis: Investigations of apoferritin by SAXS/SANS EID and SEC-SAXS methods. *J. Phys. Conf. Ser.* **994**, 20 (2018).
69. Kuklin, A. I., Islamov, A. K. & Gordeliy, V. I. Scientific reviews: Two-detector system for small-angle neutron scattering instrument. *Neutron News* **16**, 16–18 (2005).
70. Kuklin, A. I. *et al.* New opportunities provided by modernized small-angle neutron scattering two-detector system instrument (YuMO). *J. Phys. Conf. Ser.* **291**, 012013 (2011).
71. Soloviev, A. G. *et al.* SAS program for two-detector system: Seamless curve from both detectors. *J. Phys. Conf. Ser.* **848**, 012020 (2017).
72. Covington, A. K., Paabo, M., Robinson, R. A. & Bates, R. G. Use of the glass electrode in deuterium oxide and the relation between the standardized pD (paD) scale and the operational pH in heavy water. *Anal. Chem.* **40**, 700–706 (1968).
73. Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
74. Hayashi, K. *et al.* Structural analysis of the phototactic transducer protein HtrII linker region from *Natronomonas pharaonis*<sup>Δ</sup>. *Biochemistry* **46**, 14380–14390 (2007).
75. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367 (2005).
76. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
77. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
78. Franke, D. *et al.* ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
79. Hopkins, J. B., Gillilan, R. E. & Skou, S. BioXTAS RAW: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.* **50**, 1545–1553 (2017).
80. Murugova, T. N. *et al.* Low resolution structural studies of apoferritin via SANS and SAXS: The effect of concentration. *J. Optoelectron. Adv. Mater.* **17**, 1397–1402 (2015).
81. Vlasov, A. *et al.* Protein structure and structural ordering versus concentration dependence. *FEBS J.* **281**, 593–593 (2014).
82. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExpASY server. In *The Proteomics Protocols Handbook* 571–607 (Humana Press, 2005).
83. Kibbe, W. A. OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, 20 (2007).
84. Svergun, D. I. Mathematical methods in small-angle scattering data analysis. *J. Appl. Crystallogr.* **24**, 485–492 (1991).
85. Svergun, D., Barberato, C. & Koch, M. H. CRYSOLE—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
86. Svergun, D. I. *et al.* Protein hydration in solution: Experimental observation by X-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA* **95**, 2267–2272 (1998).
87. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: A windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
88. Pérez, J. & Koutsoubas, A. Memprot: A program to model the detergent corona around a membrane protein based on SEC-SAXS data. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 86–93 (2015).
89. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
90. Trehwella, J. *et al.* 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: An update. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 710–728 (2017).
91. Wolfram Research, I. Mathematica, Version 12.0.
92. Mulato, M. & Chambouleyron, I. Small-angle X-ray and neutron scattering of polydisperse systems: Determination of the scattering-particle-size distribution. *J. Appl. Crystallogr.* **29**, 29–36 (1996).

93. Antonov, L. D., Olsson, S., Boomsma, W. & Hamelryck, T. Bayesian inference of protein ensembles from SAXS data. *Phys. Chem. Chem. Phys.* **18**, 5832–5838 (2016).
94. Selivanova, O. M. *et al.* To be fibrils or to be nanofilms? Oligomers are building blocks for fibril and nanofilm formation of fragments of A $\beta$  peptide. *Langmuir* **34**, 2332–2343 (2018).
95. Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668–W673 (2004).
96. Whitmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **89**, 392–400 (2008).
97. Andrade, M. A., Chacón, P., Merelo, J. J. & Morán, F. Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng. Des. Sel.* **6**, 383–390 (1993).

## Acknowledgements

We acknowledge European Synchrotron Radiation Facility at Grenoble (France) for granting access to the synchrotron BM29 Bio-SAXS beamline. We acknowledge Frank Laboratory of Neutron Physics (FLNP), JINR at Dubna (Russia) for granting access to the small-angle neutron scattering spectrometer YuMO (IBR-2). We acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for the provision of experimental facilities. Parts of this research were carried out at the EMBL PETRA III P12 BioSAXS beamline. We thank Anatoly Mikhailov and Pavel Kuzmichev for their help in measuring CD spectroscopy. We acknowledge Ivan Okhrimenko, Vitaliy Shevchenko, Taras Balandin, Andrii Ishchenko and Dmitriy Bratanov for fruitful discussions. We thank Alexander Kazantsev for the help with fitting of the SAXS experimental data and Dmitry Verteletskiy for the help with preparing the figures.

## Author contributions

V.I.G. designed and supervised the project, wrote the manuscript with strong contribution of Y.L.R., P.S.O., M.E., A.I.K., G.L. and also with the help of the others, analyzed the results. A.I.K. designed and supervised the project, made a considerable contribution to the writing of the manuscript, supervised SAXS and SANS experiments, analyzed the results and proposed "tripod"-shaped model of the trimer of dimers of the *NpSRII/NpHtrII*. Y.L.R. made a significant contribution to the writing of the manuscript, prepared the samples and participated in SAXS and SANS experiments, treated and analyzed the data. P.S.O. made a considerable contribution to the writing of the manuscript, produced MD simulations and computed atomic models of the dimer and trimer of dimers of the *NpSRII/NpHtrII* complex. M.I.R. optimized the protocol of protein co-expression, participated in SAXS experiments, and performed negative stained EM experiments to check the quality of the samples. A.V.R. wrote proposals and participated in SAXS experiments. I.A.M. performed experiments of checking the quality of the protein with negative stained electron microscopy. D.A.V. cloned the genes of *NpSRII* and *NpHtrII* optimized for expression in *E. coli* to a co-expression plasmid. M.Y.N. contributed to providing purified protein, participated in SAXS experiments and data treatment. A.V.V. contributed to writing the manuscript and participated in SAXS experiments. D.V.Z. contributed to providing purified protein and revised the manuscript. T.N.M. participated in SANS experiments and data treatment. V.V.C. contributed to analysis of CD spectra. A.Y.G. performed SAXS experiments at the EMBL PETRA III P12 BioSAXS beamline. D.I.S. supervised SAXS experiments at the EMBL PETRA III P12 BioSAXS beamline. M.E.B. performed SAXS experiments at the ESRF BM29 Bio-SAXS beamline. I.Yu.G. contributed to writing the manuscript. M.S.-L. supervised experiments on protein expression and purification and revised the manuscript. A.B. developed a plasmid for co-expression of *NpSRII* and *NpHtrII* in *E. coli*. G.B. contributed to data analysis. G.L. made a considerable contribution to the writing of the manuscript. M.E. made a significant contribution to the writing of the manuscript.

## Funding

The reported study was funded by Russian Foundation for Basic Research (project no. 20-54-12027) and Deutsche Forschungsgemeinschaft (project no. 430170559). A.V.R. acknowledges the support from the Ministry of Science and Higher Education of the Russian Federation (agreement # 075-00337-20-03, project FSMG-2020-0003). Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-89613-6>.

**Correspondence** and requests for materials should be addressed to A.I.K. or V.I.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021