STRUCTURAL BIOLOGY

ISSN 2059-7983

Ambiguities in and completeness of SAS data analysis of membrane proteins: the case of the sensory rhodopsin II–transducer complex

Yury L. Ryzhykau,[a,b] Alexey V. Vlasov,[a,b] Philipp S. Orekhov,[a,c] Maksim I. Rulev,[d,e,f] Andrey V. Rogachev,[a,b] Anastasia D. Vlasova,[a] Alexander S. Kazantsev,[a] Dmitry P. Verteletskiy,[a] Vadim V. Skoi,[a,b] Martha E. Brennich,[g] Petra Pernot,[d] Tatiana N. Murugova,[a,b] Valentin I. Gordeliy[a,e,f,h]* and Alexander I. Kuklin[a,b]*

[a]Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny 141700, Russian Federation, [b]Frank Laboratory of Neutron Physics, Joint Institute for Nuclear Research, Dubna 141980, Russian Federation, [c]Faculty of Biology, Lomonosov Moscow State University, Moscow 119991, Russian Federation, [d]Structural Biology Group, European Synchrotron Radiation Facility, 38000 Grenoble, France, [e]Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich, 52425 Jülich, Germany, [f]JuStruct: Jülich Center for Structural Biology, Forschungszentrum Jülich, 52425 Jülich, Germany, [g]Synchrotron Crystallography Team, EMBL Grenoble Outstation, 38042 Grenoble, France, and [h]Institut de Biologie Structurale Jean-Pierre Ebel, Université Grenoble Alpes–Commissariat à l'Energie Atomique et aux Energies Alternatives–CNRS, 38027 Grenoble, France. *Correspondence e-mail: valentin.gordeliy@ibs.fr, kuklin@nf.jinr.ru

Received 6 June 2021
Accepted 14 September 2021

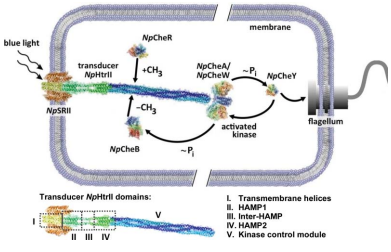
Edited by M. Czjzek, Station Biologique de Roscoff, France

Keywords: membrane proteins; two-component systems; small-angle scattering; *ab initio* modelling; detergent belt; oligomerization state; sensory rhodopsin II–transducer complex; *Natronomonas pharaonis*.

SASBDB references: *Np*SRII–*Np*HtrII at 0.15 *M* NaCl, SAXS data, SASDJ69; at 0.15 *M* NaCl, SANS data, SASDJ79; at 1.4 *M* NaCl, SANS data, SASDJ89; at 2.8 *M* NaCl, SANS data, SASDJ99; at 4.0 *M* NaCl, SANS data, SASDJA9

Supporting information: this article has supporting information at journals.iucr.org/d

Membrane proteins (MPs) play vital roles in the function of cells and are also major drug targets. Structural information on proteins is vital for understanding their mechanism of function and is critical for the development of drugs. However, obtaining high-resolution structures of membrane proteins, in particular, under native conditions is still a great challenge. In such cases, the low-resolution methods small-angle X-ray and neutron scattering (SAXS and SANS) might provide valuable structural information. However, in some cases small-angle scattering (SAS) provides ambiguous *ab initio* structural information if complementary measurements are not performed and/or *a priori* information on the protein is not taken into account. Understanding the nature of the limitations may help to overcome these problems. One of the main problems of SAS data analysis of solubilized membrane proteins is the contribution of the detergent belt surrounding the MP. Here, a comprehensive analysis of how the detergent belt contributes to the SAS data of a membrane-protein complex of sensory rhodopsin II with its cognate transducer from *Natronomonas pharaonis* (*Np*SRII–*Np*HtrII) was performed. The influence of the polydispersity of *Np*SRII–*Np*HtrII oligomerization is the second problem that is addressed here. It is shown that inhomogeneity in the scattering length density of the detergent belt surrounding a membrane part of the complex and oligomerization polydispersity significantly impacts on SAXS and SANS profiles, and therefore on 3D *ab initio* structures. It is described how both problems can be taken into account to improve the quality of SAS data treatment. Since SAS data for MPs are usually obtained from solubilized proteins, and their detergent belt and, to a certain extent, oligomerization polydispersity are sufficiently common phenomena, the approaches proposed in this work might be used in SAS studies of different MPs.



1. Introduction

Membrane proteins are key players in most biochemical processes in living organisms. They are responsible for intercellular contacts and many other functions, also providing cells with information about their surroundings. Most microorganisms use membrane proteins for taxis, transport and bioenergetic needs (Okamoto *et al.*, 1998; Terlau & Kirchhoff, 2006; Koebnik *et al.*, 2000; Shi & Massagué, 2003). Membrane

proteins often serve as targets for different drugs, *i.e.* antibacterial and anticancer drugs, drugs against neurodegenerative diseases *etc.* Structural studies of membrane proteins are important for pharmacological and medical applications (Tanford & Reynolds, 1976; Arnold & Linke, 2007; Seddon *et al.*, 2004; von Heijne, 2006; Verchère *et al.*, 2017; Vlasov *et al.*, 2020).

One family of proteins of interest are two-component signalling systems (TCSs; Jacob-Dubuisson *et al.*, 2018; Buschiazzo & Trajtenberg, 2019; Gushchin & Gordeliy, 2018). TCSs allow microorganisms to communicate with the environment. They are present in almost all domains of life and are the most abundant of nature's signalling systems. A TCS consists of a histidine kinase and a response regulator. The first component is usually represented by a transmembrane receptor (histidine kinases, chemoreceptors and photoreceptors) with a similar modular structure (Jacob-Dubuisson *et al.*, 2018; Buschiazzo & Trajtenberg, 2019). Photoreceptors are similar to chemoreceptors (Hoff *et al.*, 1997; Klare *et al.*, 2008; Gushchin & Gordeliy, 2018).

The multistate conformation of the receptors and a lack of information on the tertiary structure of the proteins, their oligomeric functional state (a trimers of dimers in the case of chemoreceptors and photoreceptors; Akkaladevi *et al.*, 2018; Stalla *et al.*, 2019; Orban-Glass *et al.*, 2015) and the dependence of the structure on the surroundings are major limitations in attempts to obtain full-length high-resolution structures by X-ray crystallography and cryo-EM.

Applying small-angle scattering (SAS) to studies of the membrane-protein complex of sensory rhodopsin and its cognate transducer (see Fig. 1) from *Natronomonas pharaonis* (*Np*SRII–*Np*HtrII; a representative of the TCS sensors), we faced several methodological problems which initially made us question the reliability of the SAS *ab initio* structural model of the complex. This study continues the investigation of the structure of the *Np*SRII–*Np*HtrII complex that was started in our previous work (Ryzhykau *et al.*, 2021), in which we approximated SAS data by molecular models of the complexes based on crystal structures of the fragments and their modifications. A full description of this work is given in Ryzhykau *et al.* (2021). While working on this problem, we recognized that the difficulties we encountered were in common with those that occur when performing SAS with many other MPs. Here, we describe these problems in detail and how they were overcome. We performed SAS data analysis of particular membrane proteins; however, the results of our work might be applicable to SAS studies of different MPs. The pipeline to obtaining a membrane-protein structure can have a bottleneck at the membrane-protein crystallization step. In the case of difficulties, SAS structural information can be of great value. Moreover, even when a high-resolution structure is available, low-resolution SAS data obtained with the protein in solution, and not confined in the crystal, are considered to be complementary information.

When an atomic structure is available for some of the subunits of a protein complex, a mosaic model can be built and checked for consistency with an experimentally obtained *ab initio* low-resolution structure. In our work, we created a high-resolution model combining the crystal structures of fragments of *Np*SRII–*Np*HtrII using computer modelling. The

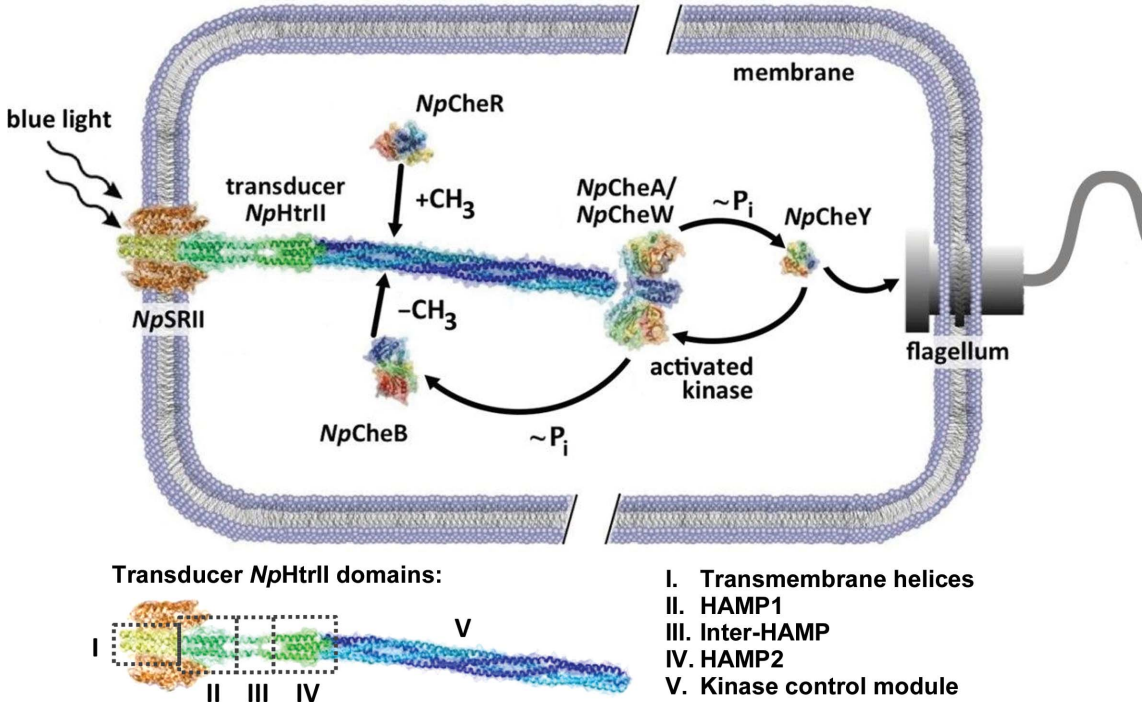


Figure 1
TCS-based signal transduction in the case of chemotaxis/phototaxis. Top: signalling pathway in the case of the *Np*SRII–*Np*HtrII complex responsible for negative phototaxis in *N. pharaonis* (Orekhov *et al.*, 2015). Bottom: domain architecture of the transducer *Np*HtrII, which includes TM helices bound to sensory rhodopsin II, two HAMP domains (connected by a helical inter-HAMP region; Koch *et al.*, 2008) and a kinase control module (KCM).

Table 1
SAS experimental details and data-evaluation summary (Trewhella *et al.*, 2017; Brennich *et al.*, 2017).

(*a*) Sample details. Parameters were calculated for the dimer of the *Np*SRII–*Np*HtrII complex without a detergent belt.

	*Np*SRII–*Np*HtrII at 0.15 *M* NaCl	*Np*SRII–*Np*HtrII at 1.4 *M* NaCl	*Np*SRII–*Np*HtrII at 2.8 *M* NaCl	*Np*SRII–*Np*HtrII at 4 *M* NaCl
Description of sequence	Complex of sensory rhodopsin II (UniProt ID P42196) with its cognate transducer (UniProt ID P42259) from *N. pharaonis*			
Extinction coefficient ε (M^{-1} cm^{-1})	118720 (280 nm), 91000 (498 nm)			
Partial specific volume $\bar{\nu}$ (cm^3 g^{-1})	0.6422			
Mean solute and solvent SLD (10^{-6} Å^{-2})	SAXS, 14.17, 9.465; SANS, 2.124, 6.404	SANS, 2.124, 6.334	SANS, 2.124, 6.259	SANS, 2.124, 6.187
Mean scattering contrast $\Delta\bar{\rho}$ (10^{-6} Å^{-2})	SAXS, 4.700; SANS, −4.281	SANS, −4.211	SANS, −4.136	SANS, −4.064
Molecular mass (Da)	168970			
Sample concentration (mg ml^{-1})	SAXS, 0.47–0.57; SANS, 0.51	SANS, 0.33	SANS, 0.31	SANS, 0.40
Solvent composition†	0.5 m*M* EDTA, 0.05% DDM pH 8.0			
	150 m*M* NaCl, 25 m*M* sodium phosphate	1400 m*M* NaCl, 49 m*M* sodium phosphate	2800 m*M* NaCl, 77 m*M* sodium phosphate	4000 m*M* NaCl, 0.1 *M* sodium phosphate

(*b*) SAS data-collection parameters.

	*Np*SRII–*Np*HtrII at 0.15 *M* NaCl	*Np*SRII–*Np*HtrII at 1.4 *M* NaCl	*Np*SRII–*Np*HtrII at 2.8 *M* NaCl	*Np*SRII–*Np*HtrII at 4 *M* NaCl
Instrument	SAXS, BM29 at ESRF; SANS, YuMO at IBR-2			
Wavelength (Å)	SAXS, 0.9918; SANS, 0.5–8.0			
Beam geometry (size, sample-to-detector distance) and sample configuration	SAXS, 700 × 700 µm, 2.864 mm, 1.8 mm diameter quartz capillary; SANS, diameter 14 mm, 4.5/12.97 m, 20 × 50 × 1 mm quartz cell			
q measurement range (Å^{-1})	SAXS, 0.008–0.5; SANS, 0.007–0.5			
Absolute scaling method	SAXS, comparison with scattering from pure H_2O; SANS, vanadium standard (Kuklin *et al.*, 2017; Ostanevich, 1988)			
Basis for normalization to constant counts	SAXS, to transmitted intensity by direct beam counter; SANS, vanadium standard (Kuklin *et al.*, 2017; Ostanevich, 1988)			
Exposure time	SAXS, 16 s; SANS, 2.0 h	SANS, 3.5 h	SANS, 3.5 h	SANS, 2.0 h
Sample temperature (°C)	20			

(*c*) Software employed for SAS data reduction, analysis and interpretation.

SAS data averaging, subtraction, merging	*PRIMUSqt* from *ATSAS* 2.8.4 (Manalastas-Cantos *et al.*, 2021)
Calculation of ε from sequence	*ProtParam* (Gasteiger *et al.*, 2005; https://web.expasy.org/protparam/)
Calculation of $\bar{\nu}$ values from chemical composition	*Peptide Property Calculator* (Kibbe, 2007; http://biotools.nubic.northwestern.edu/proteincalc.html)
Calculation of $\bar{\rho}$ values from chemical composition	*SLD calculator Web* (https://sld-calculator.appspot.com/)
Guinier, $P(r)$	*GNOM* (Svergun, 1991) from *ATSAS*
Shape/bead modelling/validation	*DAMMIF* (Franke & Svergun, 2009), *DAMAVER* (Volkov & Svergun, 2003), *GASBOR* (Svergun *et al.*, 2001) and *SASRES* (Tuukkanen *et al.*, 2016) from *ATSAS*
Atomic structure modelling	*Memprot* 2.2 (Pérez & Koutsioubas, 2015), *CRYSOL* 3.0, *CRYSON* (Svergun *et al.*, 1995, 1998)
Molecular graphics	*VMD* 1.9.3 (Humphrey *et al.*, 1996)

(*d*) Structural parameters.

	*Np*SRII–*Np*HtrII at 0.15 *M* NaCl	*Np*SRII–*Np*HtrII at 1.4 *M* NaCl	*Np*SRII–*Np*HtrII at 2.8 *M* NaCl	*Np*SRII–*Np*HtrII at 4 *M* NaCl
Guinier analysis				
R_g (Å)	SAXS, 76.0 ± 1.6; SANS, 88 ± 26	SANS, 86 ± 126	SANS, 90 ± 16	SANS, 104 ± 134
qR_g range	SAX, 0.98–1.23; SANS, 0.94–1.45	SANS, 0.79–1.04	SANS, 0.74–1.42	SANS, 0.95–1.17
$P(r)$ analysis				
R_g (Å)	SAXS, 101.6 ± 0.4; SANS, 113 ± 11	SANS, 114 ± 10	SANS, 111 ± 23	SANS, 94 ± 21
d_{max} (Å)	SAXS, 410; SANS, 390	SANS, 390	SANS, 390	SANS, 390
q-range (Å^{-1})	SAX, 0.0129–0.3660; SANS, 0.0102–0.2246	SANS, 0.0092–0.2196	SANS, 0.0082–0.2142	SANS, 0.0092–0.2196
Total quality estimate (*GNOM*)	SAXS, 0.6232; SANS, 0.2323	SANS, 0.376	SANS, 0.3585	SANS, 0.4611

(*e*) Shape/bead modelling results (*DAMMIF*, *DAMAVER*, *SASRES*).

	SAXS: *Np*SRII–*Np*HtrII, 0.15 *M* NaCl			SANS: *Np*SRII–*Np*HtrII, 0.15 *M* NaCl
q-range for fitting	0.0129–0.3660			0.0102–0.2246
Anisotropy assumptions	Prolate			Prolate
Assumed symmetry	*P*1	*P*2	*P*3	*P*1
No. of *DAMMIF* calculations	20	20	20	20
χ^2 values	1.3–1.6	1.5–1.7	1.4–2.0	0.5–0.6
Average excluded volume (nm^3)	155.6	158.8	158.2	240.8
Model resolution (*SASRES*) (Å)	54 ± 4	51 ± 4	61 ± 4	52 ± 4

Table 1 (continued)

(*f*) Atomistic modelling.

	SAXS: *Np*SRII–*Np*HtrII, 0.15 *M* NaCl	SANS: *Np*SRII–*Np*HtrII, 0.15 *M* NaCl
Method	*CRYSOL* 3.0	*CRYSON*
q-range for fitting	0.0129–0.3499	0.0102–0.2246
Any measures of model precision	Pseudo-atomic detergent-belt models were generated using *Memprot*	
χ^2 value	5.0	1.3

(*g*) Shape/bead modelling results (*GASBOR*, intensity mode).

	SANS: *Np*SRII–*Np*HtrII, 0.15 *M* NaCl	SANS: *Np*SRII–*Np*HtrII, 4.0 *M* NaCl
q-range for fitting	0.0102–0.2246	0.00913–0.2196
Anisotropy/symmetry assumptions	*P*1/prolate	*P*3/prolate
χ^2 value	0.56	0.72
Model volume (nm^3)	255	765
Model resolution (from q_{max}) (Å)	28	29

(*h*) Shape/bead modelling results from simultaneous fitting of four SANS curves (*GASBOR*, mix mode).

	*Np*SRII–*Np*HtrII at 0.15 *M* NaCl	*Np*SRII–*Np*HtrII at 1.4 *M* NaCl	*Np*SRII–*Np*HtrII at 2.8 *M* NaCl	*Np*SRII–*Np*HtrII at 4 *M* NaCl
q-range for fitting	0.0102–0.2246	0.0092–0.2196	0.0082–0.2142	0.0102–0.2246
Anisotropy/symmetry assumptions	*P*3/prolate			
Any measures of model precision	*GASBOR* model obtained for *Np*SRII–*Np*HtrII at 0.15 *M* NaCl assuming *P*1 symmetry (see Table 1*g*) was used as an initial dummy-residue model			
χ^2 value	0.79	0.37	0.37	0.36
Model volume (nm^3)	422.96			
Volume fraction of the trimers of dimers	0.000 (fixed)	0.071	0.100	0.147

(*i*) Data and model deposition IDs.

	*Np*SRII–*Np*HtrII at 0.15 *M* NaCl	*Np*SRII–*Np*HtrII at 1.4 *M* NaCl	*Np*SRII–*Np*HtrII at 2.8 *M* NaCl	*Np*SRII–*Np*HtrII at 4 *M* NaCl
SASDBD code	SASDJ69, SASDJ79	SASDJ89	SASDJ99	SASDJA9

† Buffers for SANS and SAXS measurements were prepared with D_2O and H_2O, respectively.

resulting models were used as a reference to check the reliability of a SAS *ab initio* low-resolution structure of the complex.

Initially, we found considerable differences between the reference and SAS models due to ambiguities in and incompleteness of the SAS data analysis. 3D *ab initio* structures are usually built with the assumption of a homogeneous sample with equal scattering density at every point of the investigated object. This assumption is often valid in the case of water-soluble proteins. However, it is usually not the case for membrane proteins and the resulting 3D *ab initio* structures might be incorrect.

In the case of solubilized MPs and MP complexes, a lipid or detergent belt surrounds their hydrophobic part. This implies different scattering densities for the protein and its surroundings, and consequently different contributions to the scattering profile. An additional problem is that the scattering intensity $I(q)$ is the squared sum of the amplitudes of the studied object; therefore, in the case of MPs the contribution of the detergent belt to the scattering intensity cannot be eliminated in a trivial way.

In the literature, approaches for SAS data treatment for MPs have been developed and high accuracy has been demonstrated using hybrid models combining an atomic model of the membrane proteins with pseudo-atomic or non-atomic models of membrane-mimicking systems: liposomes, nanodiscs, detergent belts *etc.* (Calcutta *et al.*, 2012; Berthaud *et al.*, 2012; Pérez & Koutsioubas, 2015; Koutsioubas, 2017; Rubinson *et al.*, 2013; Ryzhykau *et al.*, 2015; Hunt *et al.*, 1997). In the pioneering papers (Calcutta *et al.*, 2012; Berthaud *et al.*, 2012), the SAXS data were well approximated using hybrid models that include atomistic membrane-protein models and pseudo-atomic detergent-belt models. The next step in this work (Berthaud *et al.*, 2012) was the creation of the *Memprot* program (Pérez & Koutsioubas, 2015). *Memprot* requires an atomic protein model and SAXS data as input, and outputs the optimum detergent-belt parameters. For this purpose, protein atomistic models can be obtained from crystal structures of a protein or its fragments, as well as from homology modelling (Shtykova *et al.*, 2017; Waterhouse *et al.*, 2018; Kelley *et al.*, 2015). However, the approaches usually considered MPs that have a relatively small soluble part or small MPs without the nonmembrane part of the protein.

In this study, we show that the influence of the detergent belt surrounding a membrane protein, for example in the *Np*SRII–*Np*HtrII membrane-protein complex, is sufficiently significant in the case of MPs with large soluble parts to be able to corrupt the *ab initio*-derived model of the low-resolution 3D structure. It seems that for proteins in which the soluble part is much larger than the membrane part and the

detergent belt, the detergent belt would make a negligible contribution to the SAXS profile and, consequently, to the 3D *ab initio* structure. However, the inhomogeneity in the scattering length density of the detergent belt (it has both positive and negative values of contrast with solvent for polar and apolar moieties, respectively) implies a characteristic local maximum at $q \simeq 0.1$–0.2 Å^{-1} (see Fig. 2*a*). A similar but more prominent local maximum is observed for pure detergent micelles (Jensen *et al.*, 2013; Ivanović *et al.*, 2020). The presence of the local maximum at $q \simeq 0.1$–0.2 Å^{-1} has a significantly impact when recovering the 3D *ab initio* structure; for example, it adds correlations of densities which correspond to distances of $2\pi/q \simeq 30$–60 Å. This might result in a misleading interpretation of the data. For example, in our case the obtained *ab initio* 3D model had a structure similar to that of a trimer of dimers (Figs. 2*b* and 2*c*); however, in fact it was a dimer with a detergent belt (see the hybrid *Memprot* models in Fig. 3). Besides, if the centre of mass of a protein lies outside its transmembrane (TM) part, as in most cases, the presence of a detergent belt surrounding the TM part shifts the centre of mass and therefore changes the R_g value, thus affecting the obtained form factor and 3D structure.

In addition, we built *ab initio* models of the membrane-protein complex *Np*SRII–*Np*HtrII using SAXS data in order to check the sensitivity of the models to the polydispersity of the oligomeric state of the complex, which occurs at low salt concentrations, and using SANS data to check the oligomeric state when varying the salinity of the solvent. The models were built under the assumption of a homogeneous sample with equal scattering density at every point of the investigated object. We show that these assumptions do not guarantee a reliable 3D *ab initio* structure.

Our study suggests that additional information about the sample (for example a detergent belt and sample inhomogeneity) is necessary in order to build reliable 3D *ab initio* models in the case of large MP complexes. Only when all of the above are taken into account is the SAS data treatment of MPs adequate.

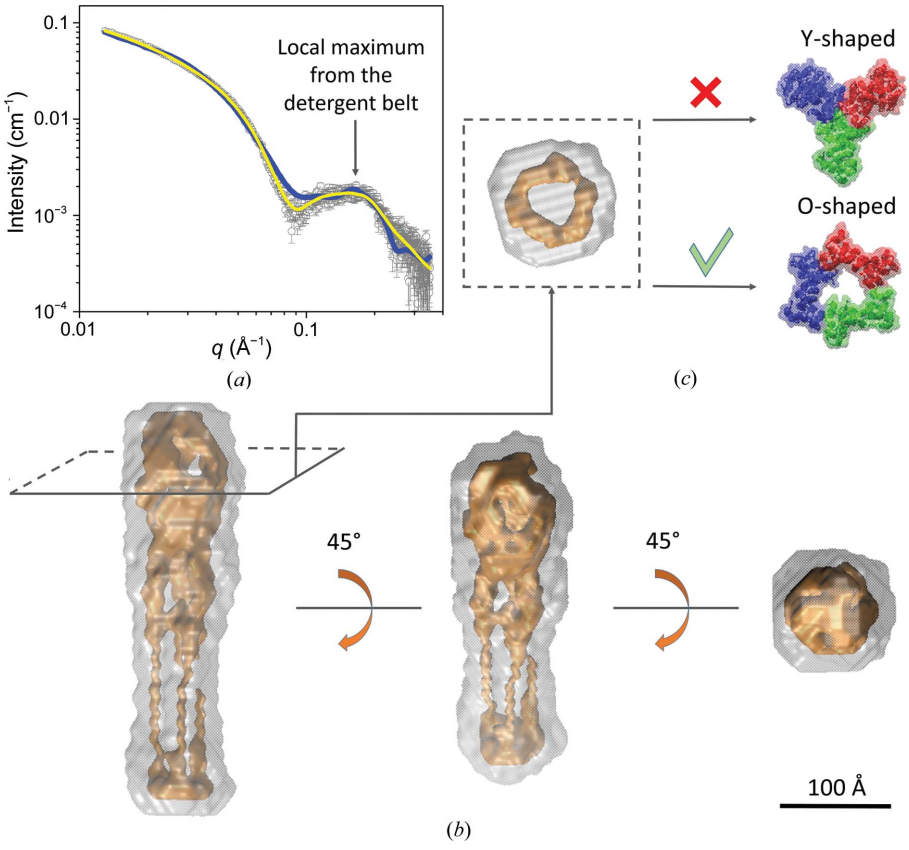


Figure 2
An example of the misinterpretation of SAXS data related to the bead *ab initio* models obtained assuming *P*3 symmetry. (*a*) SAXS experimental data (grey hollow circles) for *Np*SRII–*Np*HtrII at 150 m*M* NaCl; the corresponding indirect Fourier transform (IFT) fit obtained during $P(r)$ calculation (see details in Table 1) and *Memprot* fit obtained with a hybrid model (see Fig. 3*a*) are shown as yellow and blue lines, respectively. (*b*) Three views of the average *ab initio* structure (transparent, silver) obtained by *DAMAVER* and the most probable volume (opaque, orange) obtained by *DAMFILT* from 20 *ab initio* models obtained from SAXS data assuming *P*3 symmetry using *DAMMIF* (see Table 1 and Section 2 for details). (*c*) Comparison of TM regions corresponding to the obtained *ab initio* models with Y-shaped and O-shaped conformations of the trimer of dimers of the *Np*SRII–*Np*HtrII complex. The presence of a cavity in the centre of the expected TM region in the model favours the O-shaped conformation.

2. Materials and methods

2.1. Cloning, expression and purification

Genes optimized for expression in *Escherichia coli* (Raab *et al.*, 2010) for *Np*SRII (UniProt ID P42196) and *Np*HtrII (UniProt ID P42259) were cloned into the co-expression vector pET27bmod (Klostermeier *et al.*, 1998) via BglII–NotI and NotI–BlpI restriction sites, respectively. A Strep-tag II was added to the C-terminus of *Np*SRII and a 6×His-tag was added to the C-terminus of *Np*HtrII. The *Np*SRII–*Np*HtrII complex was co-expressed in *E. coli* strain BL21(DE3); expression conditions were the same as described by Bratanov *et al.* (2015) for the expression of *Np*SRII. After expression, the cells were pelleted and resuspended in a buffer consisting of 150 m*M* NaCl, 25 m*M* sodium phosphate buffer pH 8.0, 1 m*M* PMSF and lysed using an M-110P microfluidizer (Microfluidics, Massachusetts, USA). The cell membranes were pelleted by centrifugation. The protein was solubilized in 1% (DDM) and purified via IMAC (HisTrap HP 5 ml column, GE Healthcare, Illinois, USA) and SEC (Superose 6 10/300 GL column, 24 ml, GE Healthcare, USA). The Superose 6 column was calibrated for molecular-weight estimation (Fig. 4) using the following proteins: myoglobin,

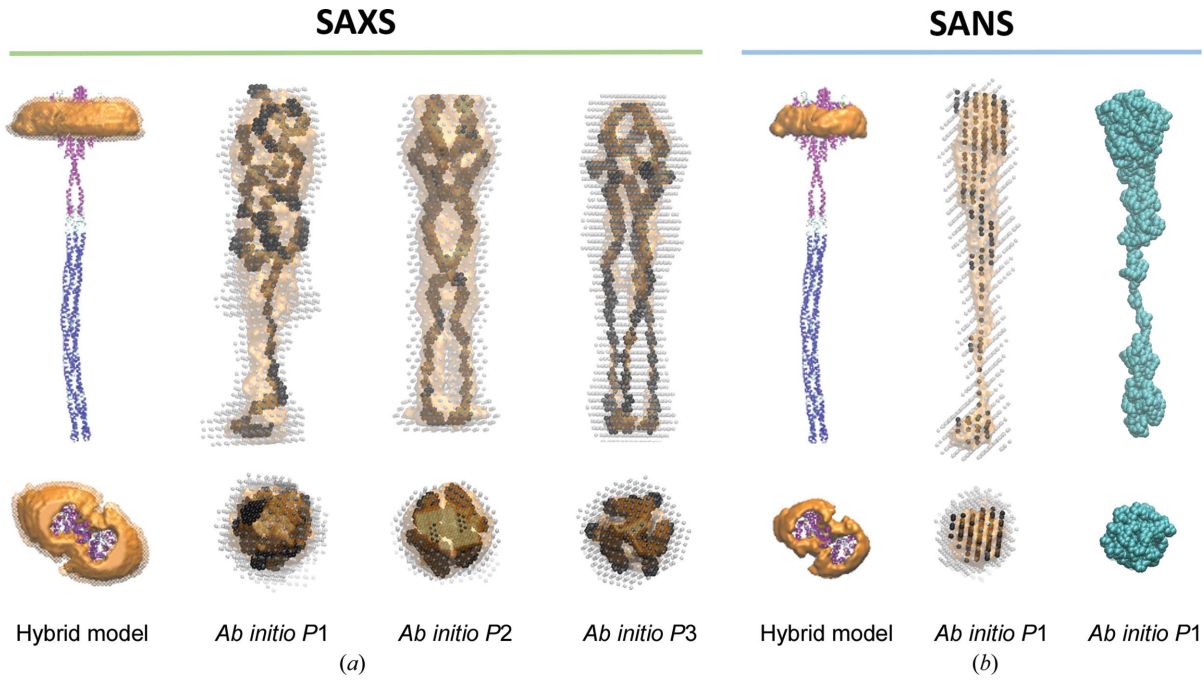


Figure 3
Comparison of *ab initio* and hybrid models of the *Np*SRII–*Np*NtrII complex obtained by SAXS and SANS methods under different symmetry assumptions. (*a*) Comparison of structures corresponding to approximations of SAXS data (from left to right: hybrid *Memprot* model of the *Np*SRII–*Np*HtrII dimer, examples of bead *ab initio* models (transparent, black; obtained with *DAMMIF*), average structures (transparent, silver; obtained with *DAMAVER*) and the most probable volumes (transparent, orange; obtained with *DAMFILT*; see Section 2 for details) obtained assuming *P*1, *P*2 and *P*3 symmetry. All three of these *ab initio* models are incorrect. (*b*) Comparison of structures corresponding to approximations of SANS data (from left to right: hybrid *Memprot* model of the *Np*SRII–*Np*HtrII dimer, average structure and the most probable volume obtained assuming *P*1 symmetry (with the same designations as in the left part), and *ab initio* reconstruction of the protein structure by a chain-like ensemble of dummy residues obtained assuming *P*1 symmetry in *GASBOR*.

ovalbumin, γ-globulin and thyroglobulin (Gel Filtration Standard; Bio-Rad, catalogue No. 151-1901), bovine serum albumin (Sigma–Aldrich, CAS No. 9048-46-8, catalogue No. A2153) and apoferritin from equine spleen (Sigma–Aldrich, CAS No. 9013-31-4, catalogue No. A3660). The described protocol for co-expression and co-purification was also used in previous SAS studies of *Np*SRII–*Np*HtrII (Ryzhykau *et al.*, 2017, 2018, 2021).

2.2. Molecular modelling

High-resolution crystal structures of methyl-accepting chemotaxis protein from *Thermotoga maritima* (PDB entry 2ch7; Li *et al.*, 2013), the *E. coli* NarQ receptor (PDB entry 5jeq; Gushchin *et al.*, 2017) and a truncated *Np*SRII–*Np*HtrII complex (PDB entry 1h2s; Gordeliy *et al.*, 2002) were used as templates for the KCM, HAMP and TM domains to generate an initial *Np*SRII–*Np*HtrII model. Template-based homology modelling was performed using *SWISS-MODEL* (Waterhouse *et al.*, 2018). The inter-HAMP region (Hayashi *et al.*, 2007) and other missing fragments were modelled as ideal helices. A 100 ns MD simulation was performed for the initial all-atom model of the full-length *Np*SRII–*Np*HtrII dimer in order to equilibrate it. The simulation protocol was similar to that employed previously (Orekhov *et al.*, 2015). The obtained optimized model of the *Np*SRII–*Np*HtrII dimer was further used to simulate SAS data. *VMD* (version 1.9.3) was used for

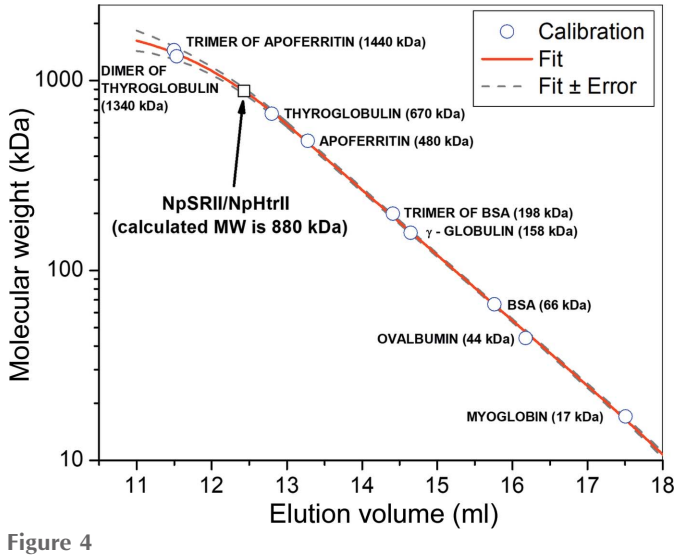


Figure 4
Analytical size-exclusion chromatography results (see Section 2 for details). The calculated MW is 880 kDa. MW estimation was performed using size-exclusion chromatography with a Superose 6 column (see Section 2 for details). The calculated MW for *Np*SRII–*Np*HtrII at 150 m*M* NaCl is 880 kDa, which corresponds to the MW of a trimer of dimers of the *Np*SRII–*Np*HtrII complex (504 kDa) with a detergent belt including about 700 DDM molecules. However, detailed analysis of the SAXS and SANS data shows that the protein under study forms dimers under these conditions, not trimers of dimers. The reason for this inconsistency is that the MW was calculated using a calibration obtained with globular proteins. *Np*SRII–*Np*HtrII is an anisotropic protein with an elongated shape.

the visualization of molecular PDB structures and *ab initio* bead and chain-like models (Humphrey *et al.*, 1996).

2.3. SAXS and SANS measurements

SAXS measurements were performed on the BM29 beamline at ESRF, Grenoble, France (Pernot *et al.*, 2013; Round *et al.*, 2015). Initial SAXS data processing was performed automatically using the *EDNA* pipeline (Brennich *et al.*, 2016; Incardona *et al.*, 2009). SANS experiments were performed using aYuMO SANS spectrometer in two-detector mode (Kuklin *et al.*, 2005) located on beamline 4 of the IBR-2 pulsed reactor at the Joint Institute for Nuclear Research, Dubna, Russia (Kuklin *et al.*, 2018). A vanadium standard was used for absolute cross-section normalization (Kuklin *et al.*, 2017; Ostanevich, 1988). The q-value range was determined using the TOF method, and the diffraction position of silver behenate was checked (Nyam-Osor *et al.*, 2012). Primary data treatment was performed using the *SAS* program (Soloviev *et al.*, 2017). Temperature control and sample environment were realized as described in Kuklin *et al.* (2011). See Table 1 for other details of SAS measurements.

2.4. SAS data processing

SAS data were processed using the *ATSAS* (Manalastas-Cantos *et al.*, 2021) and *BioXTAS RAW* (Hopkins *et al.*, 2017) software suites. During data processing, the possible influence of the structural factor was neglected, since the protein concentrations were sufficiently low (Murugova *et al.*, 2015; Vlasov *et al.*, 2014). Primary manipulations with 1D scattering profiles (averaging, subtraction and merging) were performed using *PRIMUS* and/or *PRIMUSqt*. Indirect Fourier transform (IFT) approximations and calculations of distance distribution functions $P(r)$ and regularized intensity profiles $I(q)$ were performed using *GNOM* (Svergun, 1991). *CRYSOL* and *CRYSON* were used to evaluate the solution scattering from available PDB structures of macromolecules and to fit it to experimental small-angle scattering curves (Svergun *et al.*, 1995, 1998). For the calculation of ε, the molecular mass, $\bar{\nu}$ and SLD from the sequence, *ProtParam* (Gasteiger *et al.*, 2005; https://web.expasy.org/protparam/), *Peptide Property Calculator* (Kibbe, 2007; http://biotools.nubic.northwestern.edu/proteincalc.html) and *SLD calculator Web* (https://sld-calculator.appspot.com/) were used (see Table 1). *DAMMIF* was used for *ab initio* bead modelling based on SAS data (Franke & Svergun, 2009). 20 runs of the program were performed for every data set and every symmetry assumption. *DAMAVER* (Volkov & Svergun, 2003) was used to align the 20 obtained *DAMMIF ab initio* structures and calculate an average bead model. *DAMFILT* filtered an averaged *DAMAVER* model, removing low-occupancy and loosely connected beads, given the frequency map calculated by *DAMAVER* and the cutoff volume value. *DAMFILT* output a PDB file of the compact model of the most probable areas (*i.e.* the 'most probable volume'). The *SASRES* tool (Tuukkanen *et al.*, 2016), which uses the Fourier shell correlation-based approach, was used to evaluate the resolution of an *ab initio* reconstruction using an ensemble of models. *GASBOR* was used for *ab initio* reconstruction of the protein structure as a chain-like ensemble of dummy residues (Svergun *et al.*, 2001). *GASBOR* can perform the fitting of the intensity in reciprocal space (GASBORI, or 'intensity' mode) or the fitting of the $P(r)$ function in real space (GASBORP). In our work, we used only the former. A reciprocal-space version of *GASBOR* is also available to work with an oligomeric equilibrium (*GasborMX*, or 'mix' mode). In the case of *GasborMX*, an *ab initio* model of a symmetric oligomer is built with the assumption of possible polydispersity caused by the presence of some fraction of monomers in solution.

Memprot (Pérez & Koutsioubas, 2015) was used to generate a pseudo-atomic model of the detergent belt surrounding the TM part of the protein, which was primarily oriented using the *PPM* web server (Lomize *et al.*, 2012). To take into account the different electron densities ρ of the hydrophobic and hydrophilic parts of the detergent belt, *Memprot* generated CH_3 and NH_3 pseudo-atoms (the numbers of electrons n_{e^-} is 9 and 10, respectively) and placed them at the nodes of two densely packed cubic networks with different network spacings. These network-spacing parameters were calculated in accordance with electron-density values of 0.275 and 0.515 e Å^{-3} for the hydrophobic and hydrophilic areas of the detergent corona. DDM molecules consist of two principally different parts: hydrophobic tails ($C_{12}H_{25}$, n_{e^-} = 97, neutron scattering length b_n = −13.72 fm) and hydrophilic heads ($C_{12}H_{21}O_{11}$, n_{e^-} = 181, b_n = 65.07 fm). The hydrophilic head of DDM has seven exchanging hydrogens, therefore in D_2O solution (as in our SANS experiments) the chemical formula of the head is $C_{12}H_{14}D_7O_{11}$ (b_n = 137.9 fm). In accordance with the number of electrons, the X-ray scattering lengths of the CH_3 and NH_3 pseudo-atoms correspond to 9/87 and 10/181 for the DDM tail and head, respectively. For the correct *CRYSON* calculation of the theoretical SANS curves for the membrane protein with the DDM belt, it is necessary to assign the neutron scattering lengths of pseudo-atoms in accordance with the mentioned ratios. For this purpose, a perdeuteration parameter was used in the *CRYSON* settings: if this option is applied to the CH_3 pseudo-atoms, the chemical formula of a pseudo-atom can be presented as $CH_{3-3x}D_{3x}$, where x is a perdeuturation parameter (which takes values in the range from 0 to 1). Therefore, the neutron scattering lengths of these hydrophobic pseudo-atoms are given by the expression

$$b_n(CH_{3-3x}D_{3x}) = b_n(C) + 3b_n(H) + 3x[b_n(D) - b_n(H)]. \quad (1)$$

Solving the equation $b_n(CH_{3-3x}D_{3x}) = 9b_n(C_{12}H_{25})/87$ for the unknown variable x, we obtain x = 0.1058. We then set the perdeuteration parameter in the *CRYSON* settings to this value for the chain corresponding to the hydrophobic part of the detergent belt. For the NH_3 pseudo-atoms, the described procedure does not work, so we replaced the NH_3 pseudo-atoms generated by *Memprot* by CH_3 pseudo-atoms, separating them into a chain in the modified PDB file. Solving the equation $b_n(CH_{3-3x}D_{3x}) = 10b_n(C_{12}H_{14}D_7O_{14})/181$ for the unknown variable x, we obtain x = 0.3903, which was used as the perdeuteration parameter for the chain corresponding to

the hydrophilic part of the detergent belt. Neutron scattering length values were taken from the web tool of the NIST Center for Neutron Research (https://www.ncnr.nist.gov/resources/n-lengths/), which is based on the data reported in Sears (1992).

For other details of SAS data treatment, see Table 1, which is in accordance with Trewhella *et al.* (2017) and Brennich *et al.* (2017). The SAS data were deposited in SASBDB (http://sasbdb.org; Kikhney *et al.*, 2020): SAXS data for *Np*SRII–*Np*HtrII at 0.15 *M* NaCl were deposited with accession code SASDJ69; SANS data were deposited with accession codes SASDJ79, SASDJ89, SASDJ99 and SASDJA9 for *Np*SRII–*Np*HtrII at 0.15, 1.4, 2.8 and 4.0 *M* NaCl, respectively.

3. Results

3.1. The influence of the detergent belt on the quality of the *ab initio* structural model

In this section, we describe the following sequence of results and conclusions. Firstly, we present the SAXS data obtained for the *Np*SRII–*Np*HtrII complex solubilized in DDM at low salt (150 m*M* NaCl). Secondly, we show an example of the misinterpretation of these SAXS data related to the *ab initio* models obtained assuming *P*3 symmetry. We show that the misinterpretation is caused by fitting the SAXS curve containing the peak related to the presence of the detergent belt. Thirdly, we show the correct model based on a molecular model of the *Np*SRII–*Np*HtrII dimer combined with a pseudo-atomic representation of the detergent belt surrounding the TM part of the protein and demonstrate the incorrect models obtained in the cases of different symmetry assumptions while using *ab initio* approaches. Fourthly, we show a comparison of the models obtained by both methods in the case of SANS data (combining a molecular MP model with a pseudo-atomic model of the detergent belt and an *ab initio* approach) and show that in this case the obtained results are self-consistent.

SAXS structural investigations of the *Np*SRII–*Np*HtrII complex solubilized in detergent (DDM) were performed. We used *DAMMIF* from the *ATSAS* software suite for *ab initio* shape determination, which works well for soluble proteins (Franke & Svergun, 2009). *DAMMIF* is also used in an automated data-processing and analysis pipeline (Molodenskiy *et al.*, 2020) for transmembrane proteins in detergent solutions in cases when additional information (protein atomic model, detergent chemical composition *etc.*) is not given. The pair distance distribution function $p(r)$ (Fig. 5*a*) was used to determine the shape of the protein in *DAMMIF*. Analogously to chemoreceptors, the standard physiologically functional oligomeric state of the photoreceptor is a trimer of heterodimers (Stalla *et al.*, 2019; Li *et al.*, 2011; Li & Hazelbauer, 2011). Therefore, for *ab initio* modelling *a priori* information that the particle is prolate and has *P*3 symmetry was chosen as an initial assumption because of the expected physiological oligomerization.

Two models for the formation of trimers of dimers have been proposed (Orekhov, 2016; Orekhov *et al.*, 2017): O-shaped and Y-shaped. The obtained *ab initio* structure at first could have been interpreted as the O-shaped conformation. This assumption becomes clearer when comparing the bottom view of the *ab initio* models with the bottom view of the models of the Y-shaped and O-shaped conformations (see Fig. 2*c*).

The results of *ab initio* modelling at first showed an O-shaped trimer of dimers under the assumption of *P*3 symmetry. It even corresponded to the SEC data (Fig. 6), which showed a molecular mass of 880 kDa that corresponds to a trimer of dimers of *Np*SRII–*Np*HtrII with a detergent belt including approximately 700 DDM molecules.

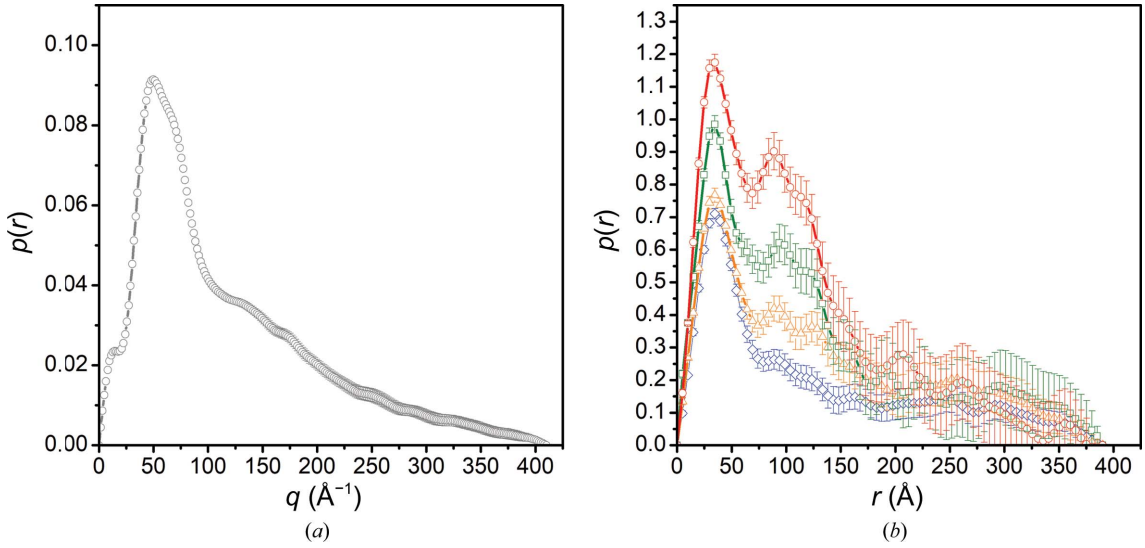


Figure 5
Distance distribution functions $p(r)$. (*a*) $p(r)$ calculated from SAXS experimental data. (*b*) $p(r)$ calculated from SANS experimental data (with the same designations as in Fig. 6*a*). The pair distance distribution functions $p(r)$ were calculated using *GNOM* (Svergun, 1991) from the *ATSAS* software suite and were used to determine the shapes of the protein in *DAMMIF* and *GASBOR*.

However, it was further shown that under these conditions (150 m*M* NaCl) the *Np*SRII–*Np*HtrII complex forms dimers instead of trimers of dimers. To prove this fact, we had to directly take the presence of DDM into account. This was performed by building a hybrid model: the protein was represented by a full-atom model of the dimer (see Section 2) and the detergent was represented by pseudo-atoms generated by *Memprot* (Pérez & Koutsioubas, 2015). Comparison of the fit showed the dimer model [see the fit in Fig. 2(*a*) (blue curve) and the hybrid model in Fig. 3(*a*)].

Additional confirmation of the dimeric state was obtained from SANS data for the complex at 150 m*M* NaCl. When measuring SANS at 100% D_2O, in contrast to SAXS, where the contrast of the scattering electron density changes sign between hydrophobic and hydrophilic parts of the detergent belt, in the case of SANS the contrast of the nuclear density has the same sign for protein and the detergent belt (both its hydrophobic and hydrophilic parts). Therefore, the assumption of homogeneous density distribution works satisfactorily and allows satisfactory *ab initio* models to be obtained.

Ab initio structures obtained from SANS data have a shape similar to the hybrid model obtained using the maximum *a priori* information about the protein (see Fig. 3*b*); they have a widening at one end, like the hybrid model (they differ in detail; however, this is expected for low-resolution structures). In accordance with the values of molecular weight (MW) and partial specific volume $\bar{\nu}$ of the *Np*SRII–*Np*HtrII dimer (see Table 1*a*), it has a volume of 180 nm^3; taking into account the detergent belt, the total volume is larger and equals 242 nm^3. The average excluded volume for the bead *ab initio DAMMIF* models is 240.8 nm^3 (see Table 1*e*), which is in excellent agreement with the volume of the hybrid model.

Ab initio structures obtained from SAXS (unlike SANS) data have much less similarity to the hybrid model. The reason for the differences lies in the presence of a scattering maximum in the range 0.1–0.2 $Å^{-1}$, which is associated with the detergent belt (detergent micelles also have an intensity maximum in this range of scattering vectors; Jensen *et al.*, 2013; Ivanović *et al.*, 2020). This peak corresponds to correlation distances of 30–60 Å. Realization of these correlation distances in *ab initio* structures leads to artefacts such as incorrect thicknesses and volumes of the protein and false-positive confirmations of symmetry assumptions when the correlation distances of 30–60 Å become the major distance between monomers (asymmetric parts). The latter is especially evident when trying to set symmetry that differs from *P*1 (see Figs. 2*b* and 3*a*). For this reason, a false-positive confirmation was obtained for the O-shaped model of the trimer of dimers, while in fact the oligomeric state corresponds only to a dimer and not to a trimer of dimers; moreover, the protein monomers in these dimers have contacts along their entire length and not separated by 30–60 Å, as in the case of the *ab initio* model obtained under the assumption of *P*2 symmetry (see Fig. 3*a*).

One can note that the size of the detergent belt corresponding to the optimal fit is different in the cases of SAXS and SANS data. The main reason for this discrepancy is a difference in the contrast of the hydration shell: water molecules strongly bonded to the detergent/lipid heads form additional densities at the periphery of the belt. In the case of SAXS, these additional densities effectively increase the resulting sizes. In contrast, in the case of SANS, where 100% D_2O buffers are used, these additional densities effectively decrease the observed sizes.

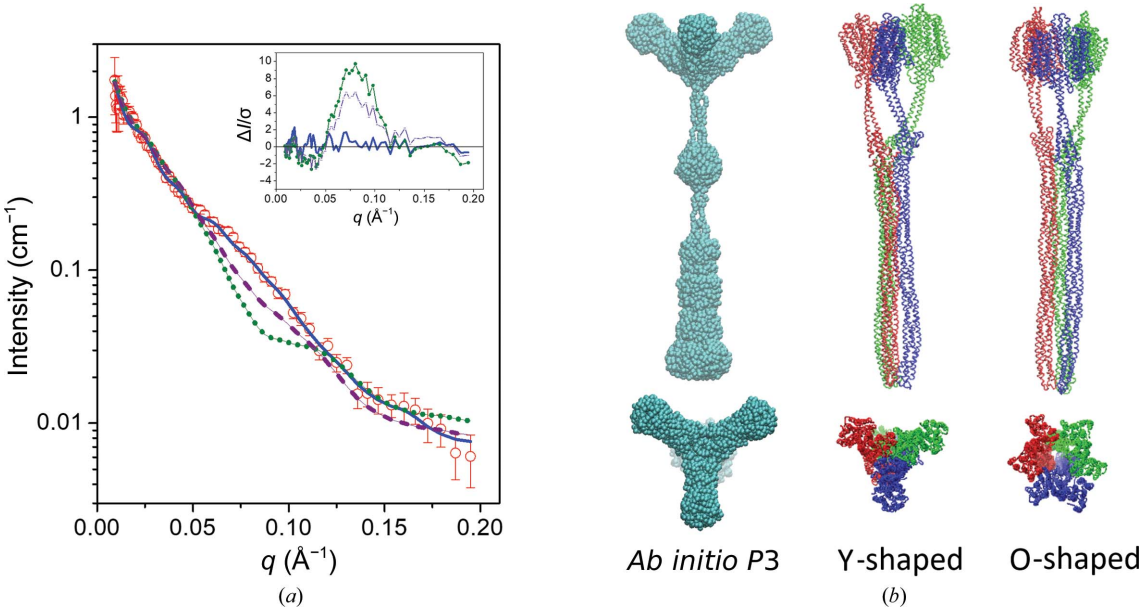


Figure 6
Interpretation of the *ab initio* model obtained from SANS data at 4.0 *M* NaCl assuming *P*3 symmetry. (*a*) SANS data at 4 *M* NaCl and the corresponding approximations obtained using *GASBOR* (blue curve) and *CRYSOL* (purple dashed curve and green dotted curve for the Y-shaped and O-shaped models, respectively). Relative differences of the approximations are given in the inset (with the same designations). (*b*) Comparison of the *GASBOR ab initio* structure (obtained assuming *P*3 symmetry) with the Y-shaped and O-shaped models of the trimer of dimers of *Np*SRII–*Np*HtrII.

3.2. The influence of the oligomerization polydispersity on the quality of the *ab initio* structural models

In this section, we describe the following sequence of results and conclusions. Firstly, we present the SANS data obtained for the *Np*SRII–*Np*HtrII complex solubilized in DDM at high salt (4 *M* NaCl pH 8.0), which is close to the optimal growth conditions. Secondly, we show an example of the misinterpretation of these SANS data related to the chain-like *ab initio* models obtained assuming *P*3 symmetry. The misinterpretation is that the obtained *ab initio* model has similarities to a Y-shaped model and is caused by fitting the SANS curve with the assumption of monodispersity, while the sample could be a mixture of proteins in two oligomeric states. In the case of the *Np*SRII–*Np*HtrII complex, these are dimers and hexamers (trimers of dimers). Thirdly, we compare the theoretical SANS curves obtained for the O- and Y-shaped models of the trimer of dimers of *Np*SRII–*Np*HtrII with the SANS experimental data (4 *M* NaCl) and conclude that these two models are not suitable to describe the shape of the studied object. Fourthly, we include an analysis of the SANS curves obtained at other salt concentrations (0.15, 1.4 and 2.8 *M* NaCl) and show that the sample can be represented by a mixture of dimers and trimers of dimers. This allows us to obtain a chain-like *ab initio* model of the trimer of dimers that differs from both the O-shaped and Y-shaped models but is in good agreement with the tripod-shaped model proposed in our previous work (Ryzhykau *et al.*, 2021).

Before SAXS and SANS experiments showed that *Np*SRII–*Np*HtrII usually forms dimers in solutions with low ionic strength, the salinity of the solvents was not taken into account. To take the effect of salt into account, the SANS curve obtained for the *Np*SRII–*Np*HtrII complex at 4.0 *M* NaCl was analyzed (see Fig. 6*a*). In contrast to the low-salt conditions (150 m*M* NaCl) discussed in the previous paragraph, high-salt conditions such as 4.0 *M* NaCl correspond to completely physiological conditions, since *N. pharaonis* lives in saturated salt solutions (Soliman & Trueper, 1982) and grows optimally at 3.5 *M* NaCl (Falb *et al.*, 2005). In accordance with the similarity of photoreceptors and chemoreceptors (Hoff *et al.*, 1997), for which the trimer of dimers is a core signalling unit (Li & Hazelbauer, 2011), and the EPR data that are available for the photoreceptor (Orban-Glass *et al.*, 2015), it is reasonable to expect that at 4.0 *M* NaCl the folding and oligomerization of *N. pharaonis* proteins correspond to their native folding and oligomerization, and in particular that *Np*SRII–*Np*HtrII forms trimers of dimers. However, from general considerations it is difficult to predict whether the dimers will assemble into trimers of dimers completely or only partially.

To check for the presence or absence of oligomerization polydispersity, we performed calculations of the chain-like *ab initio* structures of the *Np*SRII–*Np*HtrII trimer of dimers in two cases: under the assumption of a monodisperse system (using *GASBOR* in 'intensity' mode; see Section 2 for details) and assuming the possibility of the simultaneous presence of two components (using *GASBOR* in 'mix' mode; see Section 2 for details), dimers and trimers of dimers. It is expected that if all dimers assemble into trimers of dimers at 4 *M* NaCl and the system is monodisperse, the results of *ab initio* modelling in both cases have to show similar results. The pair distance distribution functions $p(r)$ (Fig. 7*b*) were used to determine the shape of the protein complexes by *GASBOR*.

The *ab initio* chain-like model obtained using *GASBOR* (in 'intensity' mode) assuming *P*3 symmetry is in best agreement with the Y-shaped model of the trimer of dimers (see Fig. 6*b*). The larger size of the membrane part in comparison with the atomic structure can be explained by the presence of a detergent environment. It is interesting to note that the thin section of the *ab initio* structure obtained in the region near the TM part corresponds to the localization of the inter-HAMP regions, which are also the thinnest parts of the complex.

However, additional comparison of the theoretical SANS curves obtained for the Y- and O-shaped models of trimers of dimers of *Np*SRII–*Np*HtrII with the SANS experimental data (Fig. 6) shows poor metrics of approximation quality (χ^2 is 4.9 and 11.3 for the Y- and O-shaped models, respectively). The inclusion of a pseudo-atomic detergent belt in the model only impairs the quality of the approximations (χ^2 is 10.0 and 12.6 for the Y- and O-shaped models, respectively; the data are not shown). In both the Y- and O-shaped models, the plot of relative differences (Fig. 6*a*) has a maximum at $q \simeq 0.08$ Å^{-1}, which corresponds to $r \simeq 2\pi/q = 84$ Å. In the chain-like *ab initio* model (Fig. 6*b*), this characteristic distance of ∼84 Å correlates with the distances between thinner parts of the model. All this allows us to conclude that the Y- and O-shaped models are not suitable for describing the shaped of the studied object, and that the visual similarity of the *ab initio* structure and the Y-shaped model is just an artefact of the data treatment.

Given that the protein forms dimers at 150 m*M* NaCl and trimers of dimers at 4.0 *M* NaCl, then at intermediate salt concentrations the coexistence of these two oligomeric states can be observed. Moreover, there is essentially no guarantee that all *Np*SRII–*Np*HtrII molecules are 100% in the trimer-of-dimers state at 4.0 *M* NaCl. For chemoreceptors, it is assumed that the protein can coexist in the forms of dimers and trimers of dimers. The trimers of dimers of chemoreceptors (or, in the case of *Np*SRII–*Np*HtrII, a photoreceptor) together with the kinases CheA and CheW form a core unit of chemotaxis signalling complexes (Sourjik, 2004; Li & Hazelbauer, 2011). The presence of CheA and CheW is crucial for the formation of extended arrays consisting of a large number of trimers of dimers (Boukhvalova *et al.*, 2002; Piñas *et al.*, 2016; Briegel *et al.*, 2014). The possibility of the coexistence of different oligomeric states does not contradict the functionality, since the main factor affecting cell viability is the presence of the mentioned receptor-kinase membrane arrays, which can lose/acquire dimers and/or trimers of dimers only along the perimeter without losing functionality as a whole.

Since the kinases CheA and CheW were not used in our experiments, the complex can only be detected in either the dimer or the trimer-of-dimers state, or with the coexistence of these oligomeric states.

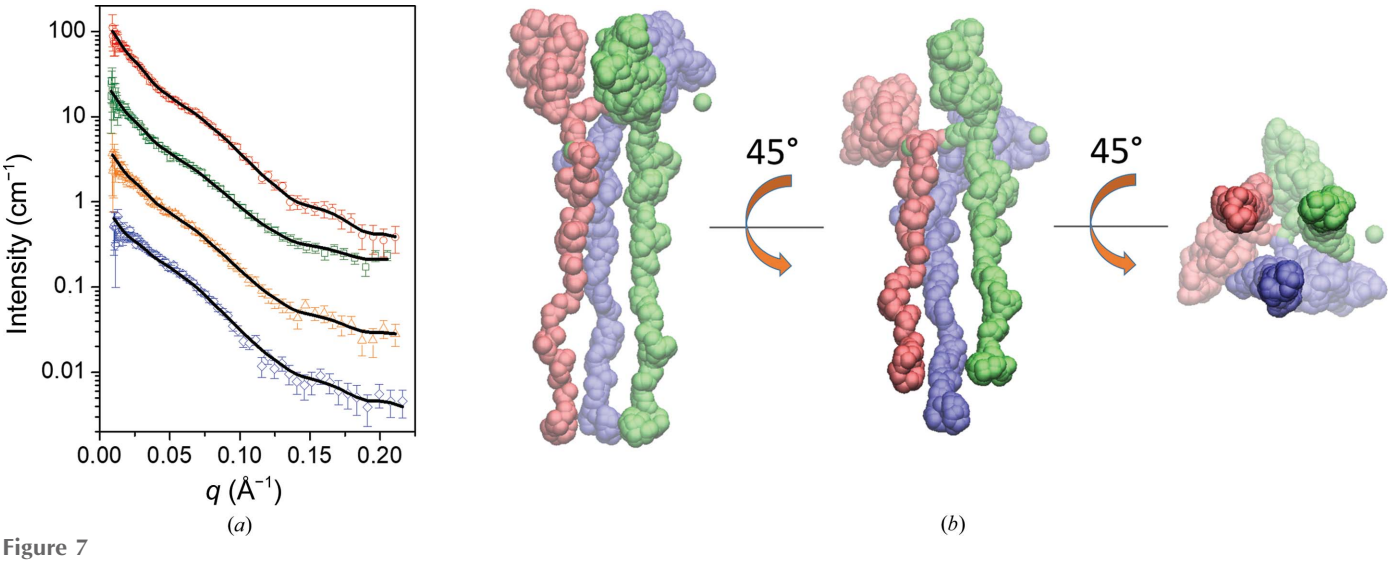


Figure 7
Obtaining the *ab initio* model from a set of SANS curves at 0.15, 1.4, 2.8 and 4.0 *M* NaCl assuming a monomer/trimer mixture (where 'monomer' corresponds to the *Np*SRII–*Np*HtrII dimer). (*a*) SANS data obtained at 0.15, 1.4, 2.8 and 4.0 *M* NaCl and the corresponding approximations obtained using *GASBOR* in 'mix' mode. For a clearer presentation of the data, intensities were multiplied by 4, 16 and 64 for 1.4, 2.8 and 4.0 *M* NaCl, respectively. (*b*) Three views of the chain-like *ab initio* model of the trimer of dimers of *Np*SRII–*Np*HtrII obtained using *GASBOR* (in 'mix' mode).

Considering the abovementioned, we have to question the correctness of the *ab initio* model shown in Fig. 6(*b*), which was obtained under the assumption of a monodisperse system. Firstly, to obtain the correct shape of the trimer of dimers of *Np*SRII–*Np*HtrII, it is necessary to consider the possibility of the coexistence of two oligomeric states. Secondly, using additional SANS data at intermediate salt concentrations of 1.4 and 2.8 *M* NaCl (Fig. 7*a*), along with the curves obtained at 150 m*M* and 4.0 *M* NaCl, can help to improve the statistical confidence of the *ab initio* structure of the trimer of dimers. For this purpose, *GASBOR* (Manalastas-Cantos *et al.*, 2021) from the *ATSAS* suite (Svergun *et al.*, 2001) was used in 'mix' mode. The resulting volume fractions of *P*3-assumed oligomer are 0.0% (fixed), 7.1%, 10.0% and 14.7% for 0.15, 1.4, 2.8 and 4.0 *M* NaCl, respectively.

The obtained *ab initio* model of the trimer of dimers *Np*SRII–*Np*HtrII (Fig. 7*b*) demonstrates another configuration, which differs from both the O- and Y-shaped models (see Fig. 6*b*). The main difference is that the contact is not along the whole protein but is only from one of the termini. To understand the biological meaning of this result, one has to build atomic models and to verify them using SAXS/SANS data and other available biophysical and biochemical data (as was performed for the sensor of a two-component signalling system in Ryzhykau *et al.*, 2021); however, this is beyond the scope of this paper. The main conclusion is that data treatment of the oligomeric protein mixture with the assumption of a homogeneous system may provide misleading proof of models that are theoretically possible, but however are not biologically relevant. Since the structure shown in Fig. 7(*b*) was obtained in the more general case, it can be considered to be more correct than that shown in Fig. 6(*b*); the latter was obtained in the more particular case of assuming monodispersity. If the assumption of monodispersity is correct, the results of *ab initio* modelling in both cases (with and without the assumption of monodispersity) have to lead to similar structures; however, this was not observed.

4. Discussion

In this work, ambiguities in SAS data analysis of membrane proteins were shown using the example of the *Np*SRII–*Np*HtrII photoreceptor complex solubilized in detergent.

As we have shown, the reasons for these ambiguities are as follows. Firstly, there is the influence of the scattering from the detergent belt on the resulting SAXS intensity profile. SAXS data for pure DDM solutions demonstrate the presence of a scattering maximum in the range 0.1–0.2 Å^{-1} (Jensen *et al.*, 2013; Ivanović *et al.*, 2020). Solubilized membrane proteins also demonstrate a local maximum in their SAXS profiles in the same range of scattering-vector modulus, which is caused by the detergent belt but not detergent micelles (this observation was reproduced in the case of SEC–SAXS data; Pérez & Koutsioubas, 2015; Berthaud *et al.*, 2012). If one uses *ab initio* modelling directly from SAXS data, the detergent-belt-created peak is approximated by correlation distances of 30–60 Å, which are undoubtedly artefacts (see Figs. 2*b*, 2*c*, 3*a* and 8*a*). Then, our results show that *ab initio* modelling directly from SAXS data cannot always be successful, especially with symmetry assumptions that differ from *P*1, because of the strong inhomogeneity of the electron density in the detergent belt, which cannot be neglected even in the case of large membrane proteins. The best way to interpret SAXS data is to use hybrid models, which combine protein atomic structure and a pseudo-atomic model of the detergent belt. The advantage of this approach is the possibility of optimizing the data approximation by fitting geometric parameters of the detergent belt. In the case of *Np*SRII–*Np*HtrII, this approach helps to define a correct oligomerization state of the protein at low-salt conditions, which equals two (*i.e.* *Np*SRII–*Np*HtrII

forms dimers). It should be taken into account that in the case of SANS data treatment, the given sizes of highly hydrated molecules will be smaller than those obtained from SAXS data processing.

In contrast to SAXS data, SANS profiles are not so sensitive to the detergent belt. This is caused by a different proportion of scattering length density (SLD) contrasts in the cases of electron and nuclear densities. In the case of the 100% D_2O buffers used in SANS experiments, the SLD of the solvent (6.4×10^{-6} Å^{-2}) is higher than the SLD values of all compounds in the particle ($\sim 2 \times 10^{-6}$, 4×10^{-6} and -0.4×10^{-6} Å^{-2} for the protein and the detergent heads and tails, respectively). Therefore, an approximation to homogeneous particle density could work satisfactory, resulting in adequate *ab initio* models. In case of water solutions of solubilized membrane proteins, the SAXS SLD value (which is proportional to the electron density) of the solvent (0.334 e Å^{-3}) is between the SLD values for the compounds in the particle ($\sim$0.42, 0.275 and 0.515 e Å^{-3} for protein, detergent heads and detergent tails, respectively). In general, the complementary use of SAXS and SANS helps to avoid the artefacts that arise when using one of the methods.

The second thing to discuss is the influence of the presence of protein oligomers on the resulting SAS profiles and on the structural parameters obtained from these data. Oligomerization of MPs is a sufficiently common phenomenon. In general, there are at least three possible reasons for proteins to experience oligomerization. The first is a physiologically important oligomerization, where proteins function as dimers [chemoreceptors and photoreceptors (Gordeliy *et al.*, 2002) and histidine kinases (Gushchin *et al.*, 2017)], trimers (*Gloeobacter* rhodopsin; Borshchevskiy *et al.*, 2010) and oligomers with higher oligomeric numbers [the pentameric sodium pump KR2 (Kovalev *et al.*, 2020), *N*(8–17)-meric c-rings of ATP synthases (Vlasov *et al.*, 2019) *etc.*], and could not function if the oligomeric state is lower than the necessary value. Alternatively, oligomerization can allow physiologically relevant cooperativity (Changeux & Edelstein, 2005). The second reason is protein aggregation, which can lead to dimerization and/or the formation of larger aggregates (Vlasov *et al.*, 2017; Böttcher & Gräber, 2000). In the case of misfolded peptides/proteins (Selivanova *et al.*, 2018), aggregation is often associated with disease. The third reason is a nonspecific protein interaction that may not result in aggregation. It can be dependent on the buffer composition (Borshchevskiy *et al.*, 2010; Petsev *et al.*, 2000) and/or, in the case of membrane proteins, on the lipid environment (Nikolaev *et al.*, 2017; Borshchevskiy *et al.*, 2010; Ishchenko *et al.*, 2017) and correspond to crystal contacts (Marchenkova *et al.*, 2016; Kovalchuk *et al.*, 2016) and/or intermediate variants of

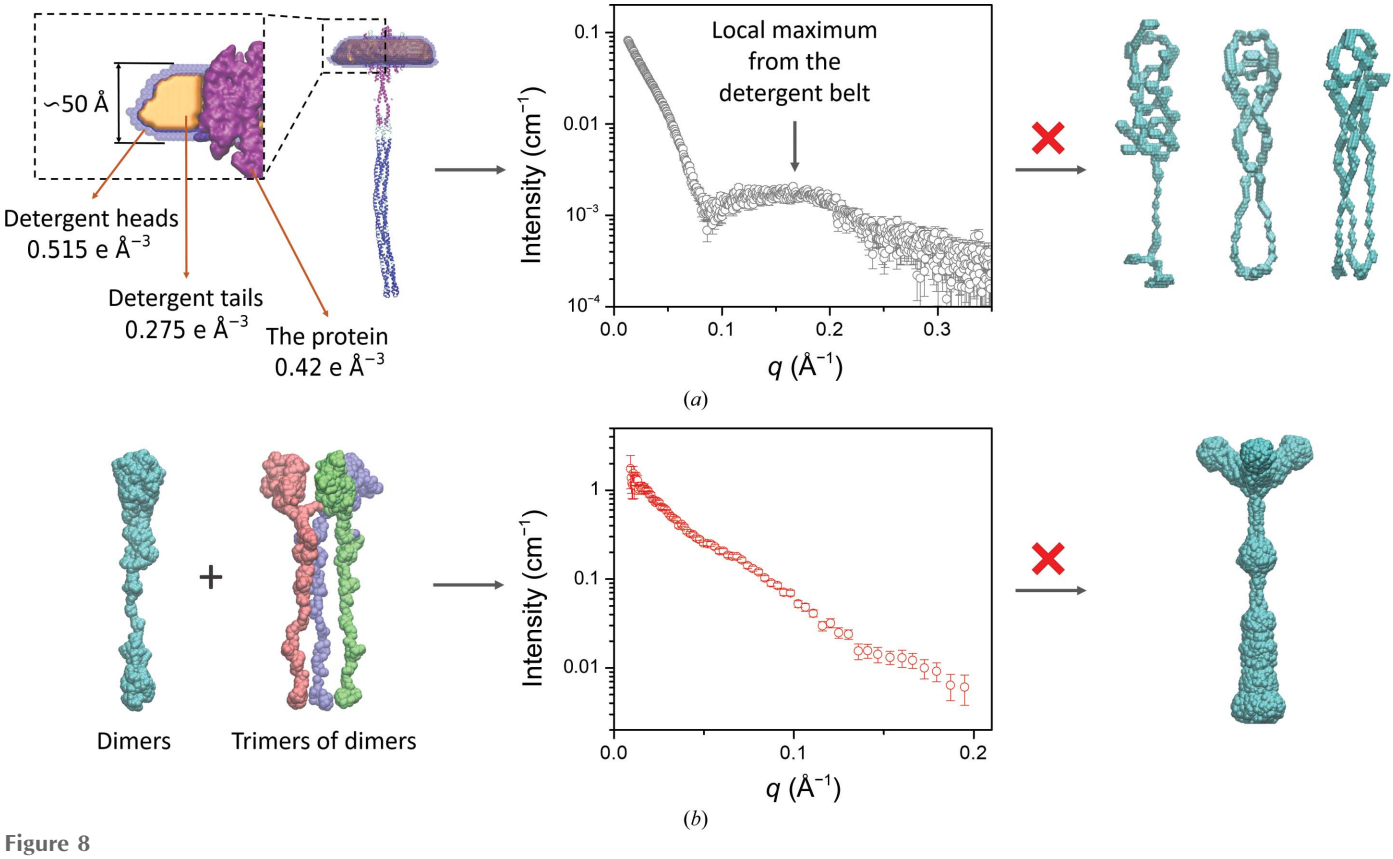


Figure 8
Schemes demonstrating the ambiguities in and completeness of SAS data analysis of membrane proteins solubilized in a detergent. (*a*) Scheme showing how a detergent belt surrounding MPs influences the local maximum in the SAXS $I(q)$ profiles, which could result in artefacts (correlation distances of 30–60 Å) if *ab initio* modelling is applied directly to the SAXS data. (*b*) Scheme showing the influence of oligomerization polydispersity on the resulting SAS curve and the corresponding *ab initio* structural model.

protein interaction during protein crystallization (Petsev *et al.*, 2000; Kazantsev *et al.*, 2018).

In the case of *Np*SRII–*Np*HtrII, physiologically important oligomerization takes place. Photoreceptors require dimerization for signal transduction through the cell membrane, and our results show that *Np*SRII–*Np*HtrII forms dimers even under low-salt conditions. The higher ordering of physiologically important oligomerization in the case of chemoreceptors and photoreceptors is the formation of a trimer of dimers. Even under physiological conditions, there is a dynamic equilibrium between dimers and trimers of dimers (see Fig. 8*b*). *Np*SRII–*Np*HtrII undergoes large structural changes due to light absorption. The results obtained in our work are crucial for the correct performance of subsequent investigations of TCS sensors, which may form a theoretical basis for the creation of a new generation of chimeric optogenetic instruments based on sensory rhodopsins and other biochemical tools and applications.

In this work, we show that if the possibility of the protein of interest having more than one oligomeric state is not taken into account when it is important, artefacts can be obtained, starting with incorrect values of structural parameters (R_g, d_{max} *etc.*) and ending with the generation of incorrect three-dimensional models. Our case of a misleading proof for the Y-shaped model of the trimer of dimers is a good example demonstrating the ambiguities in SAS data analysis and the possibilities for overcoming them.

Size-exclusion chromatography coupled with SAS (SEC–SAXS and SEC–SANS techniques) can be used to consider possible oligomerization polydispersity (Mathew *et al.*, 2004; David & Pérez, 2009; Zabelskii *et al.*, 2018; Jordan *et al.*, 2016). In addition, when several curves with variable proportions of oligomers are collected, one can extract each separate curve, as was performed in *GASBOR* by using the 'mix' mode. In our case, the proportions of dimers and trimers of dimers were fitting parameters. In general, independent information about the proportion of each component can be obtained by using dynamic light scattering (DLS), analytical gel filtration or any other protein-separation technique: one can use this information as an additional constraint when searching for an *ab initio* model in *GasborMX* or to check the consistency of different methods (if volumetric proportions of the components are not fixed in *GasborMX*).

We summarize all that has been mentioned in our work in the two schemes in Fig. 8.

Acknowledgements

We acknowledge the European Synchrotron Radiation Facility, Grenoble, France for granting access to the BM29 BioSAXS beamline and the staff of ESRF and EMBL Grenoble for assistance and support. We acknowledge the Frank Laboratory of Neutron Physics, Joint Institute for Nuclear Research at Dubna (Russia) for granting access to the YuMO (IBR-2) small-angle neutron scattering spectrometer. Open access funding enabled and organized by Projekt DEAL.

Funding information

The development of protocols for protein expression and purification, carrying out experiments with SAXS and SANS, and building a molecular model of the *Np*SRII–*Np*HtrII dimer were supported by the Russian Foundation for Basic Research (grant No. 20-54-12027 to Yury L. Ryzhykau and Andrey V. Rogachev), Deutsche Forschungsgemeinschaft (grant No. 430170559 to Maksim I. Rulev and Valentin I. Gordeliy) and the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-00958-21-05, project No. 730000F.99.1.BV10AA00006 to Andrey V. Rogachev and Anastasia D. Vlasova). The work on the calculation and analysis of *ab initio* models of the *Np*SRII–*Np*HtrII complex obtained from SAXS/SANS data was supported by the Russian Science Foundation (grant No. 21-64-00018 to Anastasia D. Vlasova, Valentin I. Gordeli and Alexander I. Kuklin). This study used the infrastructure of the Applied Genetics Resource Facility of MIPT (Support Grant 075-15-2021-684).

References

Akkaladevi, N., Bunyak, F., Stalla, D., White, T. A. & Hazelbauer, G. L. (2018). *J. Bacteriol.* **200**, e00593-17.
Arnold, T. & Linke, D. (2007). *Biotechniques*, **43**, 427–440.
Berthaud, A., Manzi, J., Pérez, J. & Mangenot, S. (2012). *J. Am. Chem. Soc.* **134**, 10080–10088.
Borshchevskiy, V., Moiseeva, E., Kuklin, A., Büldt, G., Hato, M. & Gordeliy, V. (2010). *J. Cryst. Growth*, **312**, 3326–3330.
Böttcher, B. & Gräber, P. (2000). *Biochim. Biophys. Acta*, **1458**, 404–416.
Boukhvalova, M. S., Dahlquist, F. W. & Stewart, R. C. (2002). *J. Biol. Chem.* **277**, 22251–22259.
Bratanov, D., Balandin, T., Round, E., Shevchenko, V., Gushchin, I., Polovinkin, V., Borshchevskiy, V. & Gordeliy, V. (2015). *PLoS One*, **10**, e0128390.
Brennich, M., Pernot, P. & Round, A. (2017). *Adv. Exp. Med. Biol.* **1009**, 47–64.
Brennich, M. E., Kieffer, J., Bonamis, G., De Maria Antolinos, A., Hutin, S., Pernot, P. & Round, A. (2016). *J. Appl. Cryst.* **49**, 203–212.
Briegel, A., Wong, M. L., Hodges, H. L., Oikonomou, C. M., Piasta, K. N., Harris, M. J., Fowler, D. J., Thompson, L. K., Falke, J. J., Kiessling, L. L. & Jensen, G. J. (2014). *Biochemistry*, **53**, 1575–1585.
Buschiazzo, A. & Trajtenberg, F. (2019). *Annu. Rev. Microbiol.* **73**, 507–528.
Calcutta, A., Jessen, C. M., Behrens, M. A., Oliveira, C. L. P., Renart, M. L., González-Ros, J. M., Otzen, D. E., Pedersen, J. S., Malmendal, A. & Nielsen, N. C. (2012). *Biochim. Biophys. Acta*, **1818**, 2290–2301.
Changeux, J.-P. & Edelstein, S. J. (2005). *Science*, **308**, 1424–1428.
David, G. & Pérez, J. (2009). *J. Appl. Cryst.* **42**, 892–900.
Falb, M., Pfeiffer, F., Palm, P., Rodewald, K., Hickmann, V., Tittor, J. & Oesterhelt, D. (2005). *Genome Res.* **15**, 1336–1343.
Franke, D. & Svergun, D. I. (2009). *J. Appl. Cryst.* **42**, 342–346.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.
Gordeliy, V. I., Labahn, J., Moukhametzianov, R., Efremov, R., Granzin, J., Schlesinger, R., Büldt, G., Savopol, T., Scheidig, A. J., Klare, J. P. & Engelhard, M. (2002). *Nature*, **419**, 484–487.
Gushchin, I. & Gordeliy, V. (2018). *Bioessays*, **40**, 1700197.

Gushchin, I., Melnikov, I., Polovinkin, V., Ishchenko, A., Yuzhakova, A., Buslaev, P., Bourenkov, G., Grudinin, S., Round, E., Balandin, T., Borshchevskiy, V., Willbold, D., Leonard, G., Büldt, G., Popov, A. & Gordeliy, V. (2017). *Science*, **356**, eaah6345.
Hayashi, K., Sudo, Y., Jee, J., Mishima, M., Hara, H., Kamo, N. & Kojima, C. (2007). *Biochemistry*, **46**, 14380–14390.
Heijne, G. von (2006). *Nat. Rev. Mol. Cell Biol.* **7**, 909–918.
Hoff, W. D., Jung, K.-H. & Spudich, J. L. (1997). *Annu. Rev. Biophys. Biomol. Struct.* **26**, 223–258.
Hopkins, J. B., Gillilan, R. E. & Skou, S. (2017). *J. Appl. Cryst.* **50**, 1545–1553.
Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
Hunt, J. F., McCrea, P. D., Zaccaï, G. & Engelman, D. M. (1997). *J. Mol. Biol.* **273**, 1004–1019.
Incardona, M.-F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N. & Svensson, O. (2009). *J. Synchrotron Rad.* **16**, 872–879.
Ishchenko, A., Peng, L., Zinovev, E., Vlasov, A., Lee, S. C., Kuklin, A., Mishin, A., Borshchevskiy, V., Zhang, Q. & Cherezov, V. (2017). *Cryst. Growth Des.* **17**, 3502–3511.
Ivanović, M. T., Hermann, M. R., Wójcik, M., Pérez, J. & Hub, J. S. (2020). *J. Phys. Chem. Lett.* **11**, 945–951.
Jacob-Dubuisson, F., Mechaly, A., Betton, J. M. & Antoine, R. (2018). *Nat. Rev. Microbiol.* **16**, 585–593.
Jensen, G. V., Lund, R., Gummel, J., Monkenbusch, M., Narayanan, T. & Pedersen, J. S. (2013). *J. Am. Chem. Soc.* **135**, 7214–7222.
Jordan, A., Jacques, M., Merrick, C., Devos, J., Forsyth, V. T., Porcar, L. & Martel, A. (2016). *J. Appl. Cryst.* **49**, 2015–2020.
Kazantsev, A. S., Vlasov, A. V., Ryzhykau, Y. L., Zabelskii, D. V., Murugova, T. N., Ivankov, O. I. I., Rogachev, A. V., Zinovev, E. V., Kurbatov, N. M., Gordeliy, V. I. & Kuklin, A. I. (2018). *J. Bioenerg. Biomembr.* **50**, 548.
Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. (2015). *Nat. Protoc.* **10**, 845–858.
Kibbe, W. A. (2007). *Nucleic Acids Res.* **35**, W43–W46.
Kikhney, A. G., Borges, C. R., Molodenskiy, D. S., Jeffries, C. M. & Svergun, D. I. (2020). *Protein Sci.* **29**, 66–75.
Klare, J. P., Chizhov, I. & Engelhard, M. (2008). *Results Probl. Cell Differ.* **45**, 73–122.
Klostermeier, D., Seidel, R. & Reinstein, J. (1998). *J. Mol. Biol.* **279**, 841–853.
Koch, M. K., Staudinger, W. F., Siedler, F. & Oesterhelt, D. (2008). *J. Mol. Biol.* **380**, 285–302.
Koebnik, R., Locher, K. P. & Van Gelder, P. (2000). *Mol. Microbiol.* **37**, 239–253.
Koutsioubas, A. (2017). *Biophys. J.* **113**, 2373–2382.
Kovalchuk, M. V., Blagov, A. E., Dyakova, Y. A., Gruzinov, A. Y., Marchenkova, M. A., Peters, G. S., Pisarevsky, Y. V., Timofeev, V. I. & Volkov, V. V. (2016). *Cryst. Growth Des.* **16**, 1792–1797.
Kovalev, K., Astashkin, R., Gushchin, I., Orekhov, P., Volkov, D., Zinovev, E., Marin, E., Rulev, M., Alekseev, A., Royant, A., Carpentier, P., Vaganova, S., Zabelskii, D., Baeken, C., Sergeev, I., Balandin, T., Bourenkov, G., Carpena, X., Boer, R., Maliar, N., Borshchevskiy, V., Büldt, G., Bamberg, E. & Gordeliy, V. (2020). *Nat. Commun.* **11**, 2137.
Kuklin, A. I., Islamov, A. K. & Gordeliy, V. I. (2005). *Neutron News*, **16**(3), 16–18.
Kuklin, A. I., Ivankov, A. I., Soloviov, D. V., Rogachev, A. V., Kovalev, Y. S., Soloviev, A. G., Islamov, A. K., Balasoiu, M., Vlasov, A. V., Kutuzov, S. A., Sirotin, A. P., Kirilov, A. S., Skoi, V. V., Rulev, M. I. & Gordeliy, V. I. (2018). *J. Phys. Conf. Ser.* **994**, 012016.
Kuklin, A. I., Rogachev, A. V., Soloviov, D. V., Ivankov, O. I., Kovalev, Y. S., Utrobin, P. K., Kutuzov, S. A., Soloviev, A. G., Rulev, M. I. & Gordeliy, V. I. (2017). *J. Phys. Conf. Ser.* **848**, 012010.
Kuklin, A. I., Soloviov, D. V., Rogachev, A. V., Utrobin, P. K., Kovalev, Y. S., Balasoiu, M., Ivankov, O. I., Sirotin, A. P., Murugova, T. N., Petukhova, T. B., Gorshkova, Y. E., Erhan, R. V., Kutuzov, S. A., Soloviev, A. G. & Gordeliy, V. I. (2011). *J. Phys. Conf. Ser.* **291**, 012013.
Li, M. & Hazelbauer, G. L. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 9390–9395.
Li, M., Khursigara, C. M., Subramaniam, S. & Hazelbauer, G. L. (2011). *Mol. Microbiol.* **79**, 677–685.
Li, X., Fleetwood, A. D., Bayas, C., Bilwes, A. M., Ortega, D. R., Falke, J. J., Zhulin, I. B. & Crane, B. R. (2013). *Biochemistry*, **52**, 3852–3865.
Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. (2012). *Nucleic Acids Res.* **40**, D370–D376.
Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I. & Franke, D. (2021). *J. Appl. Cryst.* **54**, 343–355.
Marchenkova, M. A., Volkov, V. V., Blagov, A. E., Dyakova, Y. A., Ilina, K. B., Tereschenko, E. Y., Timofeev, V. I., Pisarevsky, Y. V. & Kovalchuk, M. V. (2016). *Crystallogr. Rep.* **61**, 5–10.
Mathew, E., Mirza, A. & Menhart, N. (2004). *J. Synchrotron Rad.* **11**, 314–318.
Molodenskiy, D. S., Mertens, H. D. T. & Svergun, D. I. (2020). *Sci. Rep.* **10**, 8081.
Murugova, T. N., Vlasov, A. V., Ivankov, O. I., Rogachev, A. V., Ryzhykau, Y. L., Soloviov, D. V., Zhigunov, A., Zinovev, E. V., Kovalev, Y. S., Round, A., Gordeliy, V. I. & Kuklin, A. I. (2015). *J. Optoelectron. Adv. Mater.* **17**, 1397–1402.
Nikolaev, M., Round, E., Gushchin, I., Polovinkin, V., Balandin, T., Kuzmichev, P., Shevchenko, V., Borshchevskiy, V., Kuklin, A., Round, A., Bernhard, F., Willbold, D., Büldt, G. & Gordeliy, V. (2017). *Cryst. Growth Des.* **17**, 945–948.
Nyam-Osor, M., Soloviov, D. V., Kovalev, Y. S., Zhigunov, A., Rogachev, A. V., Ivankov, O. I., Erhan, R. V. & Kuklin, A. I. (2012). *J. Phys. Conf. Ser.* **351**, 012024.
Okamoto, T., Schlegel, A., Scherer, P. E. & Lisanti, M. P. (1998). *J. Biol. Chem.* **273**, 5419–5422.
Orban-Glass, I., Voskoboynikova, N., Busch, K. B., Klose, D., Rickert, C., Mosslehy, W., Roder, F., Wilkens, V., Piehler, J., Engelhard, M., Steinhoff, H. J. & Klare, J. P. (2015). *Biochemistry*, **54**, 349–362.
Orekhov, P., Bothe, A., Steinhoff, H. J., Shaitan, K. V., Raunser, S., Fotiadis, D., Schlesinger, R., Klare, J. P. & Engelhard, M. (2017). *Photochem. Photobiol.* **93**, 796–804.
Orekhov, P. S. (2016). PhD thesis. University of Osnabrück, Germany. https://nbn-resolving.org/urn:nbn:de:gbv:700-2016081014821.
Orekhov, P. S., Klose, D., Mulkidjanian, A. Y., Shaitan, K. V., Engelhard, M., Klare, J. P. & Steinhoff, H.-J. (2015). *PLOS Comput. Biol.* **11**, e1004561.
Ostanevich, Y. M. (1988). *Makromol. Chem. Macromol. Symp.* **15**, 91–103.
Pérez, J. & Koutsioubas, A. (2015). *Acta Cryst.* D**71**, 86–93.
Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., Zerrad, L. & McSweeney, S. (2013). *J. Synchrotron Rad.* **20**, 660–664.
Petsev, D. N., Thomas, B. R., Yau, S. T. & Vekilov, P. G. (2000). *Biophys. J.* **78**, 2060–2069.
Piñas, G. E., Frank, V., Vaknin, A. & Parkinson, J. S. (2016). *Proc. Natl Acad. Sci. USA*, **113**, 3335–3340.
Raab, D., Graf, M., Notka, F., Schödl, T. & Wagner, R. (2010). *Syst. Synth. Biol.* **4**, 215–225.
Round, A., Felisaz, F., Fodinger, L., Gobbo, A., Huet, J., Villard, C., Blanchet, C. E., Pernot, P., McSweeney, S., Roessle, M., Svergun, D. I. & Cipriani, F. (2015). *Acta Cryst.* D**71**, 67–75.
Rubinson, K. A., Pokalsky, C., Krueger, S. & Prochaska, L. J. (2013). *Protein J.* **32**, 27–38.
Ryzhykau, Y. L., Nikolaev, M. Y., Soloviov, D. V., Ivankov, O. I., Kovalev, Y. S., Murugova, T. N., Zinovev, E. V., Vlasov, A. V.,

Rogachev, A. V., Borshchevskiy, V. I., Gordeliy, V. I. & Kuklin, A. I. (2015). *FEBS J.* **282**, Suppl. 1, 235.
Ryzhykau, Y. L., Nikolaev, M. Y., Zabelskii, D. V., Kuklin, A. I., Gordeliy, V. I. & Gordeliy, V. I. (2017). *FEBS J.* **284**, Suppl. 1, 154.
Ryzhykau, Y. L., Orekhov, P. S., Rulev, M. I., Vlasov, A. V., Melnikov, I. A., Volkov, D. A., Nikolaev, M. Y., Zabelskii, D. V., Murugova, T. N., Chupin, V. V., Rogachev, A. V., Gruzinov, A. Y., Svergun, D. I., Brennich, M. E., Gushchin, I. Y., Soler-Lopez, M., Bothe, A., Büldt, G., Leonard, G., Engelhard, M., Kuklin, A. I. & Gordeliy, V. I. (2021). *Sci. Rep.* **11**, 10774.
Ryzhykau, Y. L., Rulev, M. I., Zabelskii, D. V., Nikolaev, M. Y., Murugova, T. N., Soler-Lopez, M., Kuklin, A. I., Engelhard, M. & Gordeliy, V. I. (2018). *J. Bioenerg. Biomembr.* **50**, 577–578.
Sears, V. F. (1992). *Neutron News*, **3**(3), 26–37.
Seddon, A. M., Curnow, P. & Booth, P. J. (2004). *Biochim. Biophys. Acta*, **1666**, 105–117.
Selivanova, O. M., Surin, A. K., Ryzhykau, Y. L., Glyakina, A. V., Suvorina, M. Y., Kuklin, A. I., Rogachevsky, V. V. & Galzitskaya, O. V. (2018). *Langmuir*, **34**, 2332–2343.
Shi, Y. & Massagué, J. (2003). *Cell*, **113**, 685–700.
Shtykova, E. V., Bogacheva, E. N., Dadinova, L. A., Jeffries, C. M., Fedorova, N. V., Golovko, A. O., Baratova, L. A. & Batishchev, O. V. (2017). *Crystallogr. Rep.* **62**, 894–902.
Soliman, G. S. H. & Trueper, H. G. (1982). *Zentralbl. Bakteriol. Mikrobiol. Hyg. C*, **3**, 318–329.
Soloviev, A. G., Solovjeva, T. M., Ivankov, O. I., Soloviov, D. V., Rogachev, A. V. & Kuklin, A. I. (2017). *J. Phys. Conf. Ser.* **848**, 012020.
Sourjik, V. (2004). *Trends Microbiol.* **12**, 569–576.
Stalla, D., Akkaladevi, N., White, T. A. & Hazelbauer, G. L. (2019). *Int. J. Mol. Sci.* **20**, 2989.
Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Svergun, D. I. (1991). *J. Appl. Cryst.* **24**, 485–492.
Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**, 2946–2953.
Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 2267–2272.
Tanford, C. & Reynolds, J. A. (1976). *Biochem. Biophys. Acta*, **457**, 133–170.
Terlau, H. & Kirchhoff, F. (2006). *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, edited by D. Ganten & K. Ruckpaul, pp. 913–916. Springer: Berlin, Heidelberg.
Trewhella, J., Duff, A. P., Durand, D., Gabel, F., Guss, J. M., Hendrickson, W. A., Hura, G. L., Jacques, D. A., Kirby, N. M., Kwan, A. H., Pérez, J., Pollack, L., Ryan, T. M., Sali, A., Schneidman-Duhovny, D., Schwede, T., Svergun, D. I., Sugiyama, M., Tainer, J. A., Vachette, P., Westbrook, J. & Whitten, A. E. (2017). *Acta Cryst.* D**73**, 710–728.
Tuukkanen, A. T., Kleywegt, G. J. & Svergun, D. I. (2016). *IUCrJ*, **3**, 440–447.
Verchère, A., Broutin, I. & Picard, M. (2017). *Methods Mol. Biol.* **1635**, 259–282.
Vlasov, A., Murugova, T., Grudinin, S., Ivankov, O., Soloviov, D., Rogachev, A., Round, A., Ryzhykau, Y., Mishin, A., Balandin, T., Borschevskiy, V., Gordeliy, V. & Kuklin, A. (2014). *FEBS J.* **281**, Suppl. 1, 593.
Vlasov, A., Ryzhykau, Y., Gordeliy, V. & Kuklin, A. (2017). *FEBS J.* **284**, Suppl. 1, 87.
Vlasov, A. V., Maliar, N. L., Bazhenov, S. V., Nikelshparg, E. I., Brazhe, N. A., Vlasova, A. D., Osipov, S. D., Sudarev, V. V., Ryzhykau, Y. L., Bogorodskiy, A. O., Zinovev, E. V., Rogachev, A. V., Manukhov, I. V., Borshchevskiy, V. I., Kuklin, A. I., Pokorný, J., Sosnovtseva, O., Maksimov, G. V. & Gordeliy, V. I. (2020). *Crystals*, **10**, 38.
Vlasov, A. V., Kovalev, K. V., Marx, S., Round, E. S., Gushchin, I. Y., Polovinkin, V. A., Tsoy, N. M., Okhrimenko, I. S., Borshchevskiy, V. I., Büldt, G. D., Ryzhykau, Y. L., Rogachev, A. V., Chupin, V. V., Kuklin, A. I., Dencher, N. A. & Gordeliy, V. I. (2019). *Sci. Rep.* **9**, 18547.
Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.
Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R. & Schwede, T. (2018). *Nucleic Acids Res.* **46**, W296–W303.
Zabelskii, D. V., Vlasov, A. V., Ryzhykau, Y. L., Murugova, T. N., Brennich, M., Soloviov, D. V., Ivankov, O. I., Borshchevskiy, V. I., Mishin, A. V., Rogachev, A. V., Round, A., Dencher, N. A., Büldt, G., Gordeliy, V. I. & Kuklin, A. I. (2018). *J. Phys. Conf. Ser.* **994**, 012017.