

Low Resolution Structure Determination of Detergent Solubilised Membrane Proteins from Small Angle Scattering Data.

A. Koutsioubas

Abstract

Despite the ever increasing usage of small-angle scattering as a valuable complementary method in the field of structural biology, applications concerning membrane proteins remain elusive mainly due to experimental challenges and the relative lack of theoretical tools for the treatment of scattering data. This fact adds up to general difficulties encountered also by other established methods (crystallography, NMR) for the study of membrane proteins. Here following the general paradigm of "ab initio methods" for low-resolution restoration of soluble protein structure from small angle scattering data, we construct a general multiphase model with a set of physical constraints which together with an appropriate minimisation procedure, gives direct structural information concerning the different components (protein, detergent molecules) of detergent solubilised membrane protein complexes. Assessment of the method's precision and robustness is evaluated by performing shape restorations from simulated data of a tetrameric α -helical membrane channel (Aquaporin-0) solubilized by n-Dodecyl β -D-Maltoside and from previously published small angle neutron scattering (SANS) experimental data of the FhaC β -barrel protein transporter solubilised by n-Octyl β -D-glucopyranoside. It is shown that the acquisition of SANS data at two different solvent contrasts together with an estimation of detergent aggregation number around the protein, permits the reliable reconstruction of the shape of membrane proteins without the need for any prior structural information.

1 Introduction

Membrane proteins perform their biological role that includes signalling, molecular transport, cell-cell recognition and energy production, by being partially embedded in the cellular membrane. Despite that knowledge of their structure is of paramount importance for better understanding a multitude of biological processes, they are heavily underrepresented in the Protein Data Bank in comparison to soluble globular proteins, with less than 700 unique membrane structures resolved (<http://blanco.biomol.uci.edu/mpstruc/>). The severe difficulties in their study by high resolution structural techniques such as crystallography and nuclear magnetic resonance (NMR) is on the one hand related to the laborious biochemical expression and purification steps needed for the preparation of appropriate samples, and on the other hand to the need of special strategies for their extraction and solubilisation.

In the endeavour to keep membrane proteins folded and active outside their natural membrane environment different approaches have been developed including the use of detergent micelles (1), amphipols (2), nanodiscs (3, 4) and bicelles (5). In all cases these molecular assemblies attempt to shield the hydrophobic transmembrane surface of the protein, that would otherwise lead to the aggregation and denaturation of the proteins in a solution. The use of detergents (small amphiphilic molecules) above their critical micelle concentration (cmc), represents the easiest and more widespread method, having the relative drawback of its tendency to destabilise certain proteins, something that can be partially remedied by the development of new 'milder' detergents or by the use of mixed lipid/detergent micelles (bicelles) that provide a more native environment to the proteins. In turn short amphiphilic polymers named amphipols despite their strong stabilising properties, tend to be much larger than the protein that they encompass. This disadvantage together with difficulties in sample preparation and protein reconstitution complexity is shared by the nanodisc system. Nanodiscs are non-covalent assemblies of a phospholipid bilayer that is surrounded by a genetically engineered lipoprotein dimer and probably represent the best mimic of the membrane environment that can at the same time solubilise membrane proteins.

Success in keeping membrane proteins stable and active in solution gives the ability to implement various experimental techniques for the study of membrane protein structure as, NMR (6), circular dichroism (7) and small angle scattering (8), while for flash-frozen proteins, cryo-electron microscopy represents a quite promising technique with many recent breakthroughs (9). However in all cases special attention should be paid for the usually non-trivial subtraction of the contribution of detergents, amphipols etc. which coexist together with the protein in solution. This is especially true in the case of small angle neutron (SANS) and x-ray (SAXS) scattering which offer the ability to acquire direct information concerning the low-resolution ($> 1\text{nm}$) structure of biological molecules (10, 11) for a wide range of solution conditions and molecular sizes.

In a small-angle scattering experiment of detergent solubilised membrane proteins, the measurable signal may contain contributions from both free detergent micelles and from the protein detergent complex (PDC) (12). Even when the scattering intensity belonging only to the PDC can be extracted, calculations regarding scattering invariants or application of structure recovery tools (13, 14) that have been developed for soluble proteins do not apply in a direct manner, thus hindering the

widespread use of the technique for this class of systems. From the experimental point of view isolation of the scattering curve corresponding only to the PDC involves the careful subtraction of buffer solution having the same free-micelle concentration as the PDC solution. This can be accomplished either by collecting reference buffer that is eluted before the PDC during a size exclusion chromatography (SEC) step right before the scattering experiment (15), or by performing extensive dialysis of the PDC solution against detergent containing buffer (16). Another proposed method based on singular value decomposition involves the acquisition of scattering data for many different protein-detergent stoichiometries (17). Alternatively the use of on-line SEC-SAXS has proven as a very efficient way for the proper subtraction of micelle contributions (18, 19).

The determination of the scattering curve corresponding to the PDC provides information about the overall size and shape of the system. In previous SAXS and SANS investigations where a high-resolution structure of the protein molecule or a candidate model was available, modelling and molecular simulations (18, 20–22) provided a better understanding of detergent organisation around the protein molecule. The ability to model adequately the detergent belt around a known protein structure (23), opens the road for experimental investigations where conformational transitions or interaction of membrane proteins with other molecules need to be resolved (24, 25).

A more ambitious approach to the whole problematic aims to exploit the big advantage of the SANS technique ie. contrast variation, in an effort to remove contributions to the scattering from detergent molecules (26–30). In principle by the use of appropriate D_2O/H_2O ratios as solvent and the full or partial deuteration of detergent molecules, one may obtain the same scattering length density (sld) for the solvent and the detergents, thus rendering the detergent belt around the protein and free-micelles 'invisible' to neutrons (31). However due to the chemically heterogeneous nature of the hydrophilic and hydrophobic part of detergents, the elimination of their scattering contribution over the entire q -range is difficult to achieve, and has been demonstrated only for special cases of detergent molecules (32). Such combinations of detergent and solvent with homogeneous matching, provided that they are able to solubilise the protein under study, permit easy sample preparation and the acquisition of scattering curves that correspond only to the protein. This same approach has been shown to be feasible also for the nanodisc system (33).

Since the requirement for simultaneous membrane protein solubilisation and detergent contrast matching in a SANS experiment is quite strict and is expected to apply in limited cases, it would be interesting to explore different approaches for the determination of membrane protein structure from small-angle scattering data. Recently, Kynde et al. (34) developed a sophisticated hybrid ab-initio approach with the aim to recover the low-resolution structure of a protein embedded in a nanodisc system from a combination of SANS and SAXS data. It was reported that for the case of bacteriorhodopsin embedded in a nanodisc, the amount of experimental information was not enough for this approach to be successful. The same method was shown to give better results for human cytochrome P450 anchored in a nanodisc environment (35).

The goal of the present work is to explore the potential of an ab-initio strategy for the general case of detergent solubilised membrane proteins. In this context we develop a multi-phase coarse-grained bead model with a set of appropriate constraints and a minimisation algorithm where with no prior knowledge of the shape of the protein we attempt to recover

the low-resolution structure of the entire PDC. The required input to the algorithm consists of a set of scattering curves, the molecular characteristics of the detergent and its aggregation number around the protein that can be independently estimated by other experimental techniques (e.g. analytical ultracentrifugation or refractometry). We explicitly show that a single scattering curve contains insufficient amount of information for obtaining a unique and physically meaningful solution. On the contrary, for a set of at least two SANS curves at different solvent contrasts, the concurrent fit of the data, leads to stable solutions where the obtained molecular envelopes are representative of the structure of each PDC component. The efficiency of the presented methodology is evaluated using two different sets of data a) scattering curves of a tetrameric α -helical membrane channel (Aquaporin-0) solubilized by n-Dodecyl β -D-Maltoside (dDM) calculated from all-atom molecular dynamics trajectories of the system and b) SANS experimental curves of the Fhac β -barrel protein transporter solubilized by n-Octyl β -D-glucopyranoside (OG_D) taken from the work of Gabel et al. (22).

2 Methods

2.1 Model and minimisation procedure

Following the methodology that was initially proposed by Chacon (13), further improved and popularised by Svergun (14, 36) for the low-resolution restoration of biomolecule shape from small angle scattering data, and also based on previous work (20, 23, 37, 38), we have developed an algorithm where a set of different types of beads, participate in a simulated annealing procedure that aims to arrange the beads in a way that is compatible with the experimentally measured scattering curves and also with the general anticipated organisation of detergent molecules around a membrane protein.

Small angle scattering curves of membrane protein/detergent complexes contain three different contrast contributions from the protein, detergent heads and tails. In this respect the constructed model consists of three types of beads (dummy atoms) corresponding to the membrane protein, detergent hydrocarbon tails and detergent hydrophilic heads, with different contrasts $\delta\rho_M$, $\delta\rho_T$ and $\delta\rho_H$ relative to the solvent. These beads of packing radius r are placed on a close-packed face centered cubic (fcc) lattice thus having a maximum of 12 nearest neighbours. The fcc lattice constant L is equal to $4r/\sqrt{2}$. The search space available for the placement of beads, is not bounded.

The scattering curve $I(q)$ vs q (scattered intensity vs wavevector transfer) of each bead configuration at each solvent contrast is calculated using the Debye equation (39)

$$I(q) = \sum_{i=1}^N \sum_{j=1}^N f_i(q) f_j(q) \frac{\sin qr_{ij}}{qr_{ij}} \quad (1)$$

where N is the total number of beads, $f_i(q)$ is the form factor of each bead given by $V_i \delta\rho_i$ with $V_i = L^3/4$ the effective bead volume, and r_{ij} is the distance between the i_{th} and j_{th} bead.

The target of the simulated annealing procedure is to obtain a compact configuration that reproduces the (M) experimentally measured scattering curves for a given number of detergent molecules (N_{det}) around the protein. The algorithm proceeds

by selecting random positions of the fcc lattice and by performing trial additions, removals and changes of the type of beads in an effort to obtain an overall bead configuration (X) that minimises the score function

$$f(X) = \sum_{l=1}^M R_l(I_{exp,l}, X)^2 + \sum_{k=1}^4 \alpha_k P_k(X) \quad (2)$$

The main contribution in $f(X)$ comes from the sum for each different contrast (l) of the q^2 -weighted discrepancies $R(I_l, X)^2$ between the scaled intensity $I_l(X)$ (scaling factor is calculated according to eq. 16 in ref. (40)) corresponding to the configuration X and the experimental scattering intensity $I_{exp,l}$, given by

$$R_l(I_l, X)^2 = \frac{\sum_q [I_l(X, q)q^2 - I_{exp,l}(q)q^2]^2}{\sum_q [I_{exp,l}(q)q^2]^2} \quad (3)$$

Since small-angle scattering curves are decaying rapidly as a function of q , the choice of q^2 -weighting ensures that experimental points at intermediate- q (where most of the shape information is contained) will contribute substantially to the score function (41). Note that especially in the case of SANS curves where resolution effects have to be considered, $I_{exp,l}$ represents the desmeared curve calculated according to the iterative method proposed by Lake (42), taking into account the beam divergence $\Delta\theta$ and wavelength spread $\Delta\lambda/\lambda$.

Additional terms in the score function are related to various penalty weights α_k and penalties P_k having to do with proteins's compactness P_1 , detergent belt compactness P_2 , detergent aggregation number P_3 and overall radius of gyration ($R_{g,l}$) penalty P_4 . Penalties P_1 and P_2 are calculated as described by Svergun (14) for the ensemble of protein and tail beads respectively. Penalty $P_3(X)$ ensures that the number of head and tail beads corresponds to the given detergent aggregation number N_{det} , and is defined as

$$P_3(X) = \frac{(N_{det} - N_t L^3 / 4V_t)^2}{N_{det}^2} + \frac{(N_{det} - N_h L^3 / 4V_h)^2}{N_{det}^2} \quad (4)$$

where N_h and N_t represent the number of head and tail beads in the configuration X , and V_h, V_t is the known molecular volume of detergent heads and tails respectively.

Penalty $P_4(X)$ is given by

$$P_4(X) = \sum_{l=1}^M \frac{(R_{g,l} - R_{gIFT,l})^2}{R_{gIFT,l}^2} \quad (5)$$

where $R_{g,l}$ is the radius of gyration of the configuration X for the contrast l , and $R_{gIFT,l}$ the corresponding radius of gyration that is estimated from the indirect Fourier transform of the experimental curve (see also next subsection). Since the search space is not bounded, this penalty prohibits large deviations of the size of the complex from the expected values, especially in the beginning of the minimisation procedure.

In order to guide the algorithm towards physically meaningful arrangements, it is was found essential to impose an additional set of constraints related to the anticipated general properties of detergent organisation around the hydrophobic transmembrane surface of the proteins. These are a) the prohibition of any hydrophobic bead contacts with empty lattice positions that represent the solvent b) hydrophobic detergent tail beads are separated from the solvent by a single layer of hydrophilic head beads c) the thickness of the hydrophobic tail bead core along the z-axis is bounded and cannot exceed the limit $2L_{hb-max}$ d) tail beads should always stay connected and should also encircle the z-axis and e) protein beads cannot be disconnected.

The first two constraints reflect the requirement that hydrophobic tails should avoid contact with water. The third and fourth constraint is implemented in order to promote "doughnut-like" detergent belt structures around the protein transmembrane hydrophobic surface. Note that the z-axis is assumed to coincide with the approximate axis of the protein's transmembrane part . Bounding the thickness of the detergent structure (along the z-axis) is related to the properties of both detergent and proteins. For compact lipid packings we expect that there is a maximum detergent hydrocarbon tail extension (L_{hb-max}) given by the Tanford formula (43) $L_{hb-max} = 1.5 + 1.265n_c \text{ \AA}$, where n_c is the number of carbons of the detergent tail. However for more disordered detergent arrangements, a more relaxed limit can be used, by arguing that the maximum thickness of the tail structure ($2L_{hb-max}$) should be roughly equal to the the typical hydrocarbon core thickness of cell membranes ($\approx 40\text{ \AA}$) where the transmembrane part of proteins is embedded and performs its function. Finally the last constraint is simply implemented for keeping the protein beads interconnected, as it is expected for a protein molecule.

From the definition of the model it can be seen that the thickness of the hydrophilic detergent head layer depends on the selected size of the beads and is equal to two times the their packing radius $2r$. However the real thickness of this layer t as can be determined by fitting small angle scattering data of detergent micelles in solution (44), might be different than $2r$. Since the contribution of this "thin" layer to the scattering depends primarily on the product of layer thickness and layer contrast (45), the effective contrast of the head beads ρ'_H is adjusted so that $2r\delta\rho'_H = t\delta\rho_H$. Additionally a modified molecular volume V'_h is used in the second term of eq. 4 following the relation $V'_h = (2r/t)V_h$.

A spherical configuration of tail beads, covered with a single layer of head beads, together with an embedded string of protein beads along the z-axis is constructed. This configuration respects all the constraints described above and serves as the initial state for the minimisation of the score function $f(X)$. The simulated annealing minimisation involves the definition of an effective annealing temperature T and proceeds as proposed in (14) (36). We begin at a high temperature where the majority of trials is accepted. Then random trials producing modified bead configurations X' (that always respect the defined constraints) are attempted and accepted if the score function is reduced or if $f(X') > f(X)$ with a probability $e^{[f(X)-f(X')]/T}$. For the results presented in this work, an annealing schedule of $T' = 0.90T$ was used, and $100N$ trials at each temperature are attempted. Note that since only one bead is modified in each trial, the recalculation of only N cross terms of eq. 1 are needed, thus avoiding the computationally intensive recalculation of the whole double sum (37). The annealing procedure is stopped when no further reduction of the score function is observed. The following values of the penalty weights were used $a_1 = 0.1$,

$a_2 = 0.1$, $a_3 = 1$ and $a_4 = 1$, that are expected to provide in the general case a balance between compactness and flexibility in the obtained PDC shapes. The used bead packing radius during the runs was in the range $2.5 - 3.0 \text{ \AA}$. In all reconstructions, scattering data up to $q = 0.2 \text{ \AA}^{-1}$ are considered, since in this q -range the scattering contribution from the shape of the PDC dominates (14). For smaller protein molecules where a higher q -range needs to be considered and internal inhomogeneities from each phase might contribute to the scattered intensity, an appropriate constant has to be subtracted from the scattering data, so that shape scattering curves are used with a q^{-4} behaviour at high q .

2.2 Scattering Data and initial treatment

Two different sets of scattering data were considered for validating the efficiency and robustness of the presented algorithm. The first set, concerns theoretically calculated SAXS and SANS curves from all-atom molecular dynamics (MD) trajectories of n-Dodecyl β -D-Maltoside/Aquaporin-0 (dDM/AQ0) complexes, while the second set consists of experimental SANS curves of n-Octyl β -D-glucopyranoside (OG_D) solubilised Fhac protein transporter in 90% D_2O and 42% D_2O , taken from the elegant work of Gabel et al. (22).

The AQ0 dataset was calculated as follows. In previous computational works (20, 46) it was shown that all-atom molecular dynamics simulations of dDM/AQ0 complexes can reliably reproduce SEC-SAXS experimental data (18) of this system. Following the exact same methodology as in Koutsioubas et al. (20) and by placing 270 dDM molecules around the backbone restrained structure of AQ0 (PDB code 2B6P (47)), we have sampled after the proper equilibration of the system, the molecular dynamics trajectory over a time frame of 5ns. Using the programs CRY SOL v.2.8.2 (40) and CRYSON v.2.7 (45) (applying default parameters including a $3 \text{ \AA}$, 10% contrast hydration layer), we calculated the corresponding SAXS and SANS curves at various solvent contrasts (100% D_2O , 42% D_2O and 0% D_2O). Averaging of scattering curves corresponding to the various time frames (over a simulation time of 5ns), yielded the final scattering curves that were further smeared (in the neutron case) by assuming a beam divergence of $\Delta\theta = 0.002 \text{ rad}$ and $\Delta\lambda/\lambda = 0.1$, which are typical for a neutron instrument.

All simulated and experimental scattering curves were treated using the program GNOM v.4.6 (48), which performs an indirect transform of the data and among other parameters, provides as output a regularised scattering curve I_{exp} and also an estimation of the protein/detergent complex radius of gyration R_{gIFT} that are in turn used as input for our simulated annealing algorithm.

2.3 Post-processing of molecular envelopes

For each shape reconstruction, 10 independent runs were executed. Using the programs DAMAVER (49) we aligned the obtained protein molecular envelopes, compared all models with each other and found the most probable one. Additionally the mean normalised spatial discrepancy (NSD) is calculated that represents a measure of the similarity between the obtained envelopes. For ideally superimposed objects, NSD tends to 0 while it exceeds 1 if the objects are systematically different.

Additionally the standard deviation of the NSD is indicative of the overall stability of the solution. The found most probable envelope is in turn aligned with the crystallographic structure of the protein using the program SUPCOMB (50). The structural similarity between the ab-initio shape and the crystallographic structure is quantified as follows: using the program pdb2sax (SITUS package) (51), the all-atom structure of the protein is mapped to a close-packed sphere model on a hexagonal lattice, with the same lattice parameter as the ab-initio bead model. Subsequently the NSD_{cryst} between the ab-initio shape and the mapped crystallographic structure is calculated using SUPCOMB. Visualisation of the PDC envelopes was performed with the program PyMOL (52).

3 Results

3.1 First example: dDM solubilized Aquaporin-0 channel

The generation of simulated scattering data concerning a well characterised system as AQ0, gave us the ability to explore different combinations of data input to our algorithm and also understand the required amount of experimental information that is needed for obtaining stable and realistic shape reconstructions. In this respect ab-initio runs from a single SAXS curve and from two or three SANS curves at different solvent contrast were executed for the 270 dDM/AQ0 system. Additionally for each of these input data combinations, two sets of runs are performed, one with no assumed symmetry and another one with imposed 4-fold axial symmetry around the z-axis. The used sld or electron densities ρ for each component of the system and detergent molecular volumes V are given in table 1. Furthermore, the thickness of the detergent hydrophilic heads t is taken to be equal to 6Å as it is found by core-shell model fits of experimental free-micelle scattering curves (44).

In all ab-initio runs, the algorithm was able to fit the MD generated input curves. In figure 1, PDC shapes for the SAXS and 2 curve SANS (100% D₂O and 0% D₂O) reconstructions are depicted, while in table 2 we give a summary of the mean NSD values and their corresponding standard deviation for each set of shape reconstructions. Also the calculated NSD_{cryst} between the most probable shape and the AQ0 tetramer crystallographic structure is presented. It is found that for a single SAXS curve as input without symmetry, the obtained NSD values are quite high indicating shape variability in the final solution. By comparing the most probable shape with the crystallographic structure we observe systematic deviations. Only when we impose a 4-fold symmetry, NSD_{cryst} tends close to 1, but still the obtained shape possesses clear differences from the AQ0 tetramer structure. We also note that the detergent belt around the protein is not quite compact as it is expected from the MD trajectories. The same conclusions are reached in the case where a single SANS curve is used as input.

In contrast, when 2 or 3 SANS curves at different solvent contrast are used as input, then the ab-initio run is much more constrained and the algorithm is guided towards stable solutions ($NSD < 1$) while as can be seen in fig. 2 the resemblance of ab-initio shapes with the crystallographic structure is evident ($NSD_{cryst} \approx 0.7$) and all major structural features of the tetramer are well reproduced. The application of symmetry restrictions improves solution stability and leads to marginally better agreement with the crystallographic structure. We also note that the inclusion of a third solvent contrast (42% D₂O)

or of the SAXS curve as input, does not lead to lowered NSD and NSD_{crist} values comparing to the 2-curve SANS runs, while identical final shapes are obtained. Additionally from a visual inspection of fig. 1 we may notice that especially for the 2 SANS curve run with 4-fold symmetry, the detergent belt is highly compact and its size and shape identical to the one of the MD trajectories.

All above results were obtained by assuming a fixed detergent aggregation number $N_{det} = 270$. However since in real cases this quantity has to be estimated experimentally and might be prone to error, we have tested the stability of the ab-initio reconstructions against small variations of the parameter N_{det} . We have found that a 10% variation of N_{det} leads to small degradation of the obtained NSD_{crist} values, while visual inspection of the protein envelopes show only minor differences comparing to the runs with $N_{det} = 270$. In more detail for $N_{det} = 240$, NSD_{crist} is equal to 0.83 and for $N_{det} = 300$, NSD_{crist} is equal to 0.72.

bead phase	$\rho(e^-/\text{\AA}^3)$	$sld_{100\%D_2O} (10^{-6}\text{\AA}^{-2})$	$sld_{42\%D_2O} (10^{-6}\text{\AA}^{-2})$	$sld_{0\%D_2O} (10^{-6}\text{\AA}^{-2})$	$V(\text{\AA}^3)$
dDM detergent heads	0.52	3.93	2.72	1.85	350
dDM detergent tails	0.28	-0.41	-0.41	-0.41	340
AQ0 protein	0.44	2.90	2.26	1.80	-
solvent	0.33	6.40	2.36	-0.56	-

Table 1: Scattering length densities (sld) at different contrasts, electron densities (ρ) and molecular volumes (V) of protein, detergent parts and solvent, that were used during the ab-initio runs for the AQ0/dDM system. Molecular volume and sld values for the detergents were taken from the work of Breyton et al. (31). Electron densities were taken from the work of Oliver et al. (44). Note that owing to the H/D exchange occurring in solvents containing a certain ratio of D_2O (53), the sld of a protein is adjusted accordingly for each solvent contrast. The related calculations are performed using the contrast calculator of the program SASSIE (54)

3.2 Second example: OG_D solubilized Fhac protein transporter

The second application of the algorithm concerned an actual set of SANS experimental data (fig. 3) from the OG_D solubilised outer-membrane β barrel transporter of the *Bordetella pertussis* filamentous hemagglutinin adhesin (Fhac) (22) which consists of a 16-stranded β barrel that is occluded by an N-terminal α helix and an extracellular loop and a periplasmic module

ab-initio run	NSD	σ_{NSD}	NSD_{crist}
single SAXS, no symmetry	1.082	0.056	1.64
single SAXS, 4-fold symmetry	1.035	0.161	1.10
two SANS, no symmetry	0.668	0.077	0.73
two SANS, 4-fold symmetry	0.457	0.048	0.70
three SANS, no symmetry	0.715	0.041	0.75
three SANS, 4-fold symmetry	0.454	0.018	0.73

Table 2: Summary of mean NSD, its standard deviation and NSD_{crist} values for ab-initio runs of the 270 dDM/AQ0 system with different number of curve inputs and imposed symmetries.

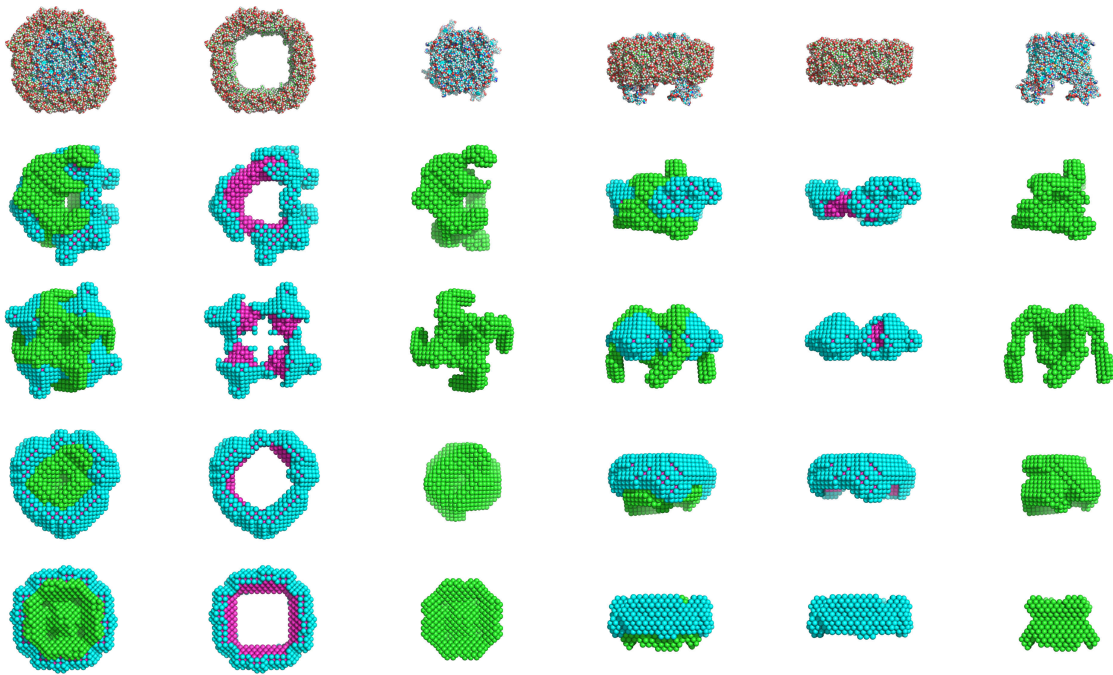


Figure 1: Ab-initio shape reconstructions of the 270 dDM/AQ0 tetramer. First row, MD structure. Second row, single SAXS curve with no imposed symmetry. Third row, single SAXS curve with imposed 4-fold axial symmetry. Fourth row, two SANS curves with no imposed symmetry. Fifth row, two SANS curves with imposed 4-fold axial symmetry. In the first three columns the top view of the different PDC components (entire PDC, detergent belt and protein respectively) are depicted, while in the last three columns the side view. The bead colours correspond to: green - protein, cyan - detergent hydrophilic heads, magenta - detergent hydrophobic tails.

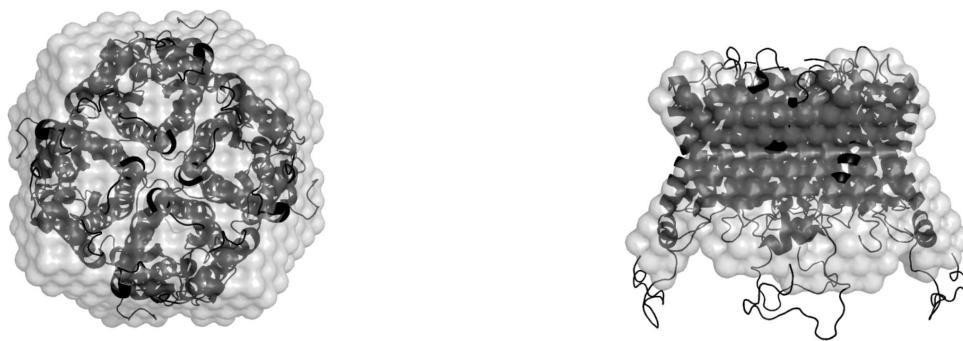


Figure 2: Comparison of most probable ab-initio envelope (light gray) with the crystallographic structure of the AQ0 tetramer PDB code 2B6P (47) (black). The reconstruction belongs to the ab-initio run performed with two SANS curves as input and with imposed 4-fold symmetry.

bead phase	sld _{90%D₂O} (10 ⁻⁶ Å ⁻²)	sld _{42%D₂O} (10 ⁻⁶ Å ⁻²)	V(Å ³)
OG _D heads	3.79	2.72	190
OG _D tails	7.28	7.28	230
Fhac protein	3.07	2.43	-
solvent	5.70	2.36	-

Table 3: Scattering length densities (sld) at different contrasts and molecular volumes (V) of protein, detergent parts and solvent, that were used during the ab-initio runs for the Fhac/OG_D system. Molecular volume and sld values for the detergents were taken from the work of Breyton et al. (31). Protein sld is adjusted for each solvent contrast (54).

composed of two aligned polypeptide-transport-associated (POTRA) domains (55). The used detergent had its hydrophobic tail part fully deuterated thus presenting high contrast in relation to the protein. The two SANS measurements were performed at 90% D₂O and 42% D₂O solvent contrast, which represent the matching point for the overall detergent molecule and protein respectively. As reported, by the use of a dialysis step against solvents with the same detergent concentrations, the free-micelle contribution was subtracted and the SANS curves belonging only to the PDC were obtained. Additionally for this particular system, the detergent aggregation number around the protein N_{det} was determined by analytical ultracentrifugation and was found to be equal to 142 ± 10 OG_D molecules.

For the ab-initio runs using both SANS contrasts, the applied sld for each component of the system and OG_D molecular volumes V are given in table 3. N_{det} is fixed at the value 140 and the thickness of the detergent hydrophilic heads t is taken to be equal to 3.2Å (44). The resulting fit of the experimental results is presented in figure 3 and the corresponding PDC shapes in figure 4. The obtained NSD of the ab-initio runs is equal to 0.657 ± 0.046 suggesting a stable set of solutions, while from the comparison with the crystallographic structure we get $NSD_{cryst} = 0.75$. Inspection of the crystallographic structure aligned with the most probable shape (fig. 5) reveals that the ab-initio envelope is representative of the overall molecular shape and that the β -barrel, extra-membrane periplasmic POTRA domains and their relative orientation are nicely reproduced. It is interesting to note that while the detergent belt encompasses the transmembrane part of the protein, its shape is somewhat disordered and not highly compact, exposing a small part of the protein β -barrel surface to the solvent. This result will be further discussed below in relation to previous MD modelling investigations (22).

As in the case of the AQ0 tetramer, we have checked the effect of detergent aggregation variation on the final solution, where we again find that for N_{det} in the range 125-155, the determined envelopes are very similar and NSD_{cryst} values remain almost constant (for $N_{det} = 125$ we find $N_{cryst} = 0.77$ and for $N_{det} = 155$, $N_{cryst} = 0.74$).

4 Discussion

In general, the different chemical nature of detergent heads and tails, makes it difficult to contrast match their contribution over the entire q-range in a small-angle scattering experiment. That means that in order to directly access protein structural information from SAXS or SANS measurements where detergents bound to the protein contribute to the scattering, we need

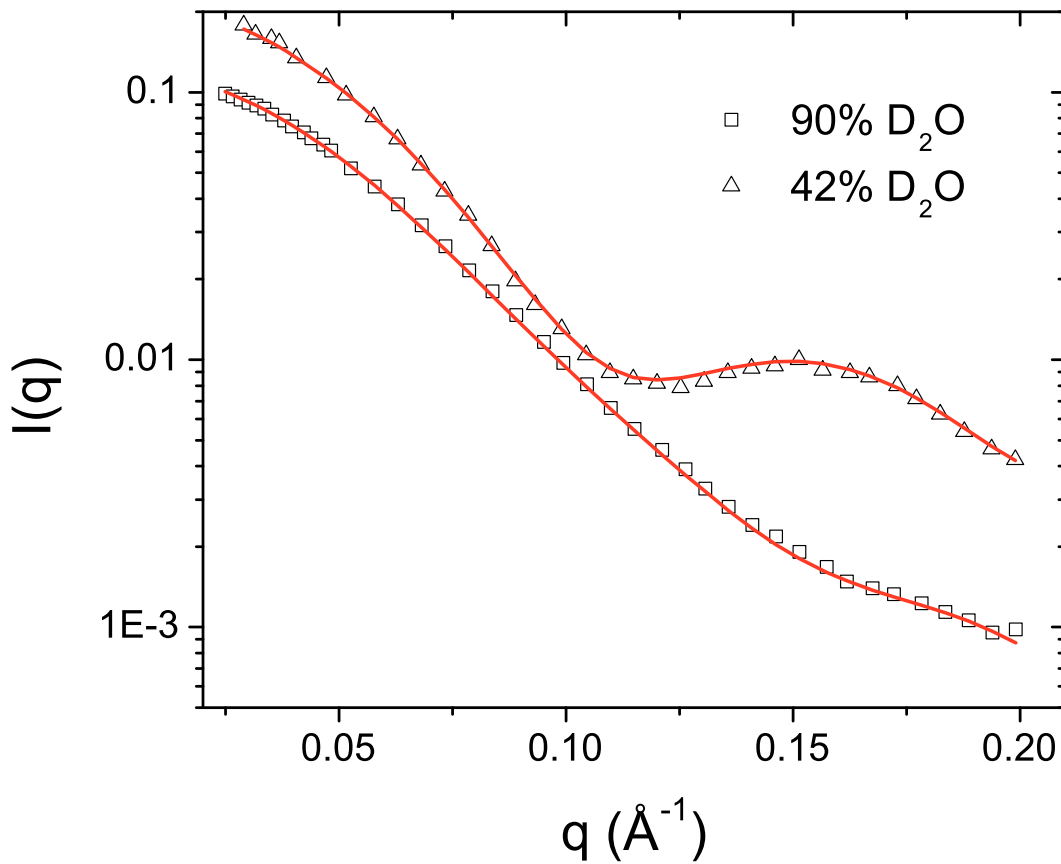


Figure 3: Fit of the two Fhac/OG_D SANS experimental curves by the ab-initio algorithm.

to rely on modelling. We note that since their introduction almost two decades ago, ab-initio structure recovery algorithms for soluble globular proteins have become increasingly popular in the literature, assuming a complementary role in integrative structural studies of proteins and nucleic acids. What makes ab-initio approaches quite attractive despite their low resolution, is that no prior structural information is required, concerning the biomolecule under study.

The main goal of the present work is to develop an ab-initio shape recovery method for detergent solubilised membrane proteins without the need for special contrast matched detergents, so that sample preparation can be more flexible and focussed on the proper solubilisation of the protein in a folded and active state. Following the paradigm of ab-initio methods, we constructed a four-phase bead model (protein, detergent heads and tails, solvent) together with a set of geometric constraints that reflect the anticipated structure of the detergent belt around the protein which should resemble a core-shell doughnut-like structure. The importance of applying constraints may be also illustrated by performing ab-initio reconstructions using the multiphase program MONSA (14) where no detergent belt related constraints are implemented and only bead connectivity

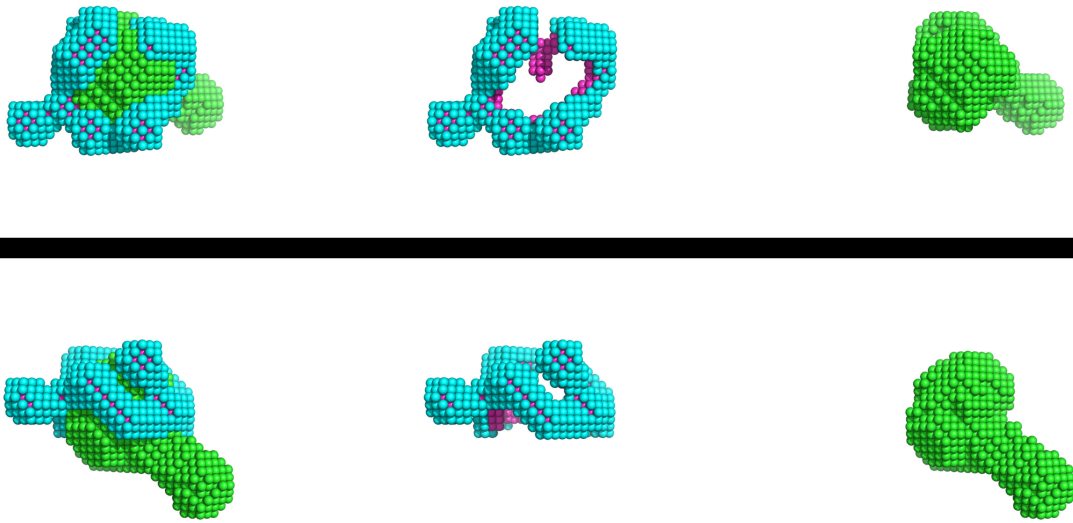


Figure 4: Most probable ab-initio shape reconstruction of the 140 OG_D/Fhac PDC. First row top view. Second row side view. In each column different PDC components (entire PDC, detergent belt and protein respectively) are depicted. The bead colours correspond to: green - protein, cyan - detergent hydrophilic heads, magenta - detergent hydrophobic tails.

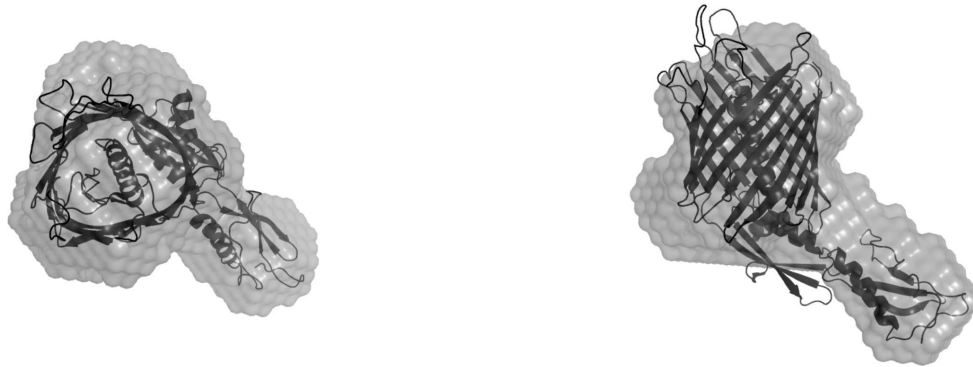


Figure 5: Comparison of most probable ab-initio envelope (light grey) with the crystallographic structure of the Fhac transporter PDB code 2QDZ (55) (black). Top view (left), side view (right).

and compactness is considered. Related results for the Fhac/OG_D system correspond to non-physical final shapes, while reconstructions for the dDM/AQ0 system using 3 SANS curves as input (fig. S1) reveal that in the final models, phases of similar contrast (protein, detergent heads) cannot be easily differentiated and obtained protein shape possesses systematic differences from the crystallographic structure.

Ideally an additional bead phase corresponding to the hydration layer that surrounds the entire PDC has to be included in our model (56), however since only a marginal improvement of the obtained shapes is expected (38) and due to the increased

complexity of such an approach, we opted not to implement it in the present algorithm. We stress that no particular geometry (for example a torus) is assumed for the detergent belt. Two extra restraints that are imposed are related a) to the maximum thickness of the belt that cannot exceed the typical cellular membrane thickness and b) to the total number of detergent molecules that are bound to the protein (aggregation number, N_{det}). By fixing N_{det} and by taking into account the molecular volume of detergent heads and tails, we essentially fix the corresponding number of beads in the final solution.

All the geometric constraints together with the requirement for model compactness and the discrepancy between the calculated and input scattering curves contribute to the defined score function $f(X)$. Successful minimisation of the score function by simulated annealing leads to a bead arrangement that respects all the described constraints and also fits the small angle scattering curves. The final bead arrangement represents the molecular envelope for each of the PDC components. The major finding of the minimisations of the MD generated dDM/AQ0 scattering data, is that a single curve even with appropriate axial symmetry applied, cannot lead to a stable solution, while for obtaining molecular envelopes that are representative of the PDC shape we need to provide at least two SANS curves at different solvent contrasts as input (fig. 1). Addition of a third contrast does not improve the shape reconstructions. Enforcement of 4-fold axial symmetry, as it is expected for the AQ0 tetramer, improves the agreement with the known crystallographic structure.

In an effort to better understand these results, we can think of the PDC system as having two principal contributions to the scattering. One coming from the protein and a second from the hydrophobic part of the detergent belt. The hydrophilic head layer that covers the belt is indeed contributing to the scattering, however by design, this structural feature is taken into account inherently in our model, since hydrophobic beads are separated from the solvent by a layer of hydrophilic beads that reproduce the scattering contribution of detergent head layers as determined in measurements of detergent micelles (31, 44). Consequently it is not surprising that two scattering curves at different contrast can lead to stable solution and to an objective shape for both the protein and the detergent belt. Here we note that the term "two different contrasts" refers to well separated contrasts (> 30% D₂O content). Additional tests have revealed that shape reconstruction quality is improved when one of the two contrasts is close to the matching point of the hydrophobic part of detergent micelles. Consequently during experimental planning and in the case of two SANS measurements, one should consider acquiring scattering data either close to the overall matching points of detergent tails and protein or at the most distant possible contrasts (0% and 100% D₂O). In the latter case one of the two measurements will always be close to the matching point of detergent tails both for deuterated or hydrogenated hydrocarbon chains.

Having established that two contrast separated SANS curves provide enough information for the ab-initio reconstruction, we tested the algorithm using previously published SANS data of the Fhac transporter in two different solvent contrasts. The final obtained molecular shape of the protein is highly similar to the crystallographic structure (fig 5) with the β barrel and POTRA domains easily distinguishable. It is interesting that the resulting detergent belt seems somewhat disordered and not covering the entire β barrel surface (fig. 4). Inspection PDC shapes belonging to different ab-initio runs (fig. S2) reveals detergent belt shapes with gaps and in some cases protrusions. The same topological variability was present in the all-atom

molecular dynamics models that gave the best fits to the scattering data in the work of Gabel et al. (22), which indicates on the one hand the relative ambiguity of small-angle data derived models and on the other hand the dynamic nature of the PDC. In this respect we note that a compact detergent belt is not a universal characteristic of solubilised membrane proteins, something that has to be carefully considered in modelling approaches.

Given that the number of protein bound detergents N_{det} is an important input parameter in our method and since its estimation by experimental methods like analytical ultra-centrifugation (22), refractometry or absolute scale small-angle scattering (18) may contain systematic or statistical errors, it was relevant to investigate the stability of the shape reconstructions to small variations of N_{det} . Our findings suggest that an error of the order of $\pm 10\%$ is not appreciably affecting the shape reconstructions. This observation gives confidence that the typical uncertainty that characterises the aforementioned experimental techniques should not affect the accuracy of our method.

The design of the PDC bead model was performed by having mainly in focus systems of integral membrane proteins solubilised by detergents. However our effort to not heavily constrain the detergent belt shape gives the ability to straightforwardly extend the applicability of the current method to bicelle / protein systems (5). Provided that the contrast and structural parameters of 'empty' bicelles are known from previous investigations, the bicelle belt can be modelled in the same manner as in the case of detergents. Furthermore the method in its current formulation seems also appropriate for the study of peripheral membrane proteins that have only a small fraction of their amino-acid sequence anchored inside the detergent micelle.

From the experimental and sample preparation point of view, it is absolutely essential to properly subtract free-micelle contributions from the scattering data in order to obtain curves corresponding solely to the PDC. Extensive dialysis against detergent containing solutions is a viable method, however the in-situ combination of SEC and SAXS has proved as a very efficient technique for this task (18). In this context and since our methodology relies on multi-contrast SANS, the recent development of an in-situ SEC-SANS platform at the Institute Laue-Langevin (57) presents special interest in relation to the acquisition of high quality scattering data from detergent solubilised membrane proteins. Such setups coupled with upcoming high flux diffractometers at future neutron sources like the European Spallation Source (ESS), may bring down the requirements for protein sample amounts thus rendering the method as a medium-throughput option for membrane protein structure determination.

The executable code of the presented ab-initio algorithm (program DANVILLE) for all major software platforms is freely available for academic use from the author upon request. A typical shape reconstruction with no imposed symmetry takes 5-20 hours on the single core of modern processors.

5 Conclusions

In summary we have presented a method for the determination of the low-resolution structure of detergent solubilised membrane proteins from small-angle neutron scattering datasets. Measurements at two different solvent contrasts together with an

independent estimation of the detergent aggregation number around the protein appear to provide reliable shape reconstructions. The generality of the model permits the treatment of data from detergent or bicelle solubilised integral or peripheral membrane proteins where their membrane embedded part is surrounded by a detergent belt of arbitrary shape. Given the extreme difficulties in the study of membrane protein structure, the present work offers a powerful tool that allows the general use of small-angle scattering for detailed studies of membrane proteins in solution without the need for any prior structural information. Related ab-initio approaches for soluble globular proteins have proved increasingly popular in the literature in recent years. In this respect, we expect that the proposed methodology coupled with the preparation of high quality samples and the implementation of adapted experimental protocols, will allow the investigation of challenging systems and will also complement current advances in other fronts like cryo-electron microscopy.

Acknowledgments

The author thanks Dr. Javier Pérez for his introduction in the general problematic of the use of small angle scattering for the study of membrane protein structure.

References

1. Seddon, A. M., P. Curnow, and P. J. Booth, 2004. Membrane proteins, lipids and detergents: not just a soap opera. Biochimica et Biophysica Acta (BBA) - Biomembranes 1666:105 – 117. Lipid-Protein Interactions.
2. Tribet, C., R. Audebert, and J.-L. Popot, 1996. Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. Proceedings of the National Academy of Sciences 93:15047–15050.
3. Bayburt, T. H., Y. V. Grinkova, and S. G. Sligar, 2002. Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. Nano Letters 2:853–856.
4. Knowles, T. J., R. Finka, C. Smith, Y.-P. Lin, T. Dafforn, and M. Overduin, 2009. Membrane Proteins Solubilized Intact in Lipid Containing Nanoparticles Bounded by Styrene Maleic Acid Copolymer. Journal of the American Chemical Society 131:7484–7485.
5. Dos Santos Morais, R., O. Delalande, J. Prez, L. Mouret, A. Bondon, A. Martel, M.-S. Appavou, E. Le Rumeur, J.-F. Hubert, and S. Combet, 2017. Contrast-Matched Isotropic Bicelles: A Versatile Tool to Specifically Probe the Solution Structure of Peripheral Membrane Proteins Using SANS. Langmuir 33:6572–6580.
6. Kang, C., and Q. Li, 2011. Solution NMR study of integral membrane proteins. Current Opinion in Chemical Biology 15:560 – 569. Next Generation Therapeutics.
7. Miles, A. J., and B. A. Wallace, 2016. Circular dichroism spectroscopy of membrane proteins. Chem. Soc. Rev. 45:4859–4872.
8. Roy, A. L., K. Wang, B. Schaack, P. Schuck, C. Breyton, and C. Ebel, 2015. AUC and Small-Angle Scattering for Membrane Proteins. Methods in Enzymology 562:257 – 286. Analytical Ultracentrifugation.
9. Rawson, S., S. Davies, J. D. Lippiat, and S. P. Muench, 2016. The changing landscape of membrane protein structural biology through developments in electron microscopy. Molecular Membrane Biology 33:12–22.

10. Koch, M. H. J., P. Vachette, and D. I. Svergun, 2003. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Quarterly Reviews of Biophysics 36:147227.
11. Jacques, D. A., and J. Trehwella, 2010. Small-angle scattering for structural biology Expanding the frontier while avoiding the pitfalls. Protein Science 19:642–657.
12. Lipfert, J., and S. Doniach, 2007. Small-Angle X-Ray Scattering from RNA, Proteins, and Protein Complexes. Annual Review of Biophysics and Biomolecular Structure 36:307–327.
13. Chacon, P., F. Moran, J. Diaz, E. Pantos, and J. Andreu, 1998. Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. Biophysical Journal 74:2760–2775.
14. Svergun, D. I., 1999. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. Biophysical journal 76:2879–2886.
15. Cooley, R. B., J. P. O'Donnell, and H. Sondermann, 2016. Coincidence detection and bi-directional transmembrane signaling control a bacterial second messenger receptor. eLife 5:e21848.
16. Zimmer, J., D. A. Doyle, and J. G. Grossmann, 2006. Structural Characterization and pH-Induced Conformational Transition of Full-Length KcsA. Biophysical Journal 90:1752 – 1766.
17. Lipfert, J., L. Columbus, V. B. Chu, and S. Doniach, 2007. Analysis of small-angle X-ray scattering data of protein–detergent complexes by singular value decomposition. Journal of Applied Crystallography 40:s235–s239.
18. Berthaud, A., J. Manzi, J. Perez, and S. Mangenot, 2012. Modeling detergent organization around aquaporin-0 using small-angle X-ray scattering. Journal of the American Chemical Society 134:10080–10088.
19. Chen, P.-c., and J. S. Hub, 2015. Structural Properties of ProteinDetergent Complexes from SAXS and MD Simulations. The Journal of Physical Chemistry Letters 6:5116–5121.
20. Koutsioubas, A., A. Berthaud, S. Mangenot, and J. Perez, 2013. Ab initio and all-atom modeling of detergent organization around Aquaporin-0 based on SAXS data. The Journal of Physical Chemistry B 117:13588–13594.
21. Dvling Kaspersen, J., C. Moestrup Jessen, B. Stougaard Vad, E. Skipper Srensen, K. Kleiner Andersen, M. Glasius, C. L. Pinto Oliveira, D. E. Otzen, and J. S. Pedersen, 2014. Low-Resolution Structures of OmpADDM ProteinDetergent Complexes. ChemBioChem 15:2113–2124.
22. Gabel, F., M. F. Lensink, B. Clantin, F. Jacob-Dubuisson, V. Villeret, and C. Ebel, 2014. Probing the conformation of FhaC with small-angle neutron scattering and molecular modeling. Biophysical journal 107:185–196.
23. Pérez, J., and A. Koutsioubas, 2015. Memprot: a program to model the detergent corona around a membrane protein based on SEC–SAXS data. Acta Crystallographica Section D: Biological Crystallography 71:86–93.
24. Gao, Y., G. Westfield, J. W. Erickson, R. A. Cerione, G. Skinotis, and S. Ramachandran, 2017. Isolation and Structure-function Characterization of a Signaling-active Rhodopsin-G protein Complex. Journal of Biological Chemistry .
25. Wojtowicz, H., A. Prochnicka-Chalufour, G. Cardoso de Amorim, O. Roudenko, C. Simenel, I. Maki, G. Pehaud-Arnaudet, F. Gubellini, A. Koutsioubas, J. Perez, P. Delepelaire, M. Delepierre, R. Fronzes, and N. Izadi-Pruneyre, 2016. Structural basis of the signaling through a bacterial membrane receptor HasR deciphered by an integrative approach. Biochemical Journal .
26. Clifton, L. A., C. L. Johnson, A. S. Solovyova, P. Callow, K. L. Weiss, H. Ridley, A. P. Le Brun, C. J. Kinane, J. R. P. Webster, S. A. Holt, and J. H. Lakey, 2012. Low Resolution Structure and Dynamics of a Colicin-Receptor Complex Determined by Neutron

- Scattering. Journal of Biological Chemistry 287:337–346.
27. Nogales, A., C. Garca, J. Prez, P. Callow, T. A. Ezquerra, and J. Gonzalez-Rodriguez, 2010. Three-dimensional Model of Human Platelet Integrin IIB3 in Solution Obtained by Small Angle Neutron Scattering. Journal of Biological Chemistry 285:1023–1031.
 28. Johs, A., M. Hammel, I. Waldner, R. P. May, P. Laggner, and R. Prassl, 2006. Modular Structure of Solubilized Human Apolipoprotein B-100: LOW RESOLUTION MODEL REVEALED BY SMALL ANGLE NEUTRON SCATTERING. Journal of Biological Chemistry 281:19732–19739.
 29. Cardoso, M. B., D. Smolensky, W. T. Heller, and H. O'Neill, 2009. Insight into the Structure of Light-Harvesting Complex II and Its Stabilization in Detergent Solution. The Journal of Physical Chemistry B 113:16377–16383.
 30. Tang, K.-H., V. S. Urban, J. Wen, Y. Xin, and R. E. Blankenship, 2010. SANS Investigation of the Photosynthetic Machinery of *Chloroflexus aurantiacus*. Biophysical Journal 99:23982407.
 31. Breyton, C., F. Gabel, M. Lethier, A. Flayhan, G. Durand, J.-M. Jault, C. Juillan-Binard, L. Imbert, M. Moulin, S. Ravaut, M. Härtlein, and C. Ebel, 2013. Small angle neutron scattering for the study of solubilised membrane proteins. The European Physical Journal E 36:71.
 32. Compton, E. L. R., E. Karinou, J. H. Naismith, F. Gabel, and A. Javelle, 2011. Low Resolution Structure of a Bacterial SLC26 Transporter Reveals Dimeric Stoichiometry and Mobile Intracellular Domains. Journal of Biological Chemistry 286:27058–27067.
 33. Maric, S., N. Skar-Gislinge, S. Midtgaard, M. B. Thygesen, J. Schiller, H. Frielinghaus, M. Moulin, M. Haertlein, V. T. Forsyth, T. G. Pomorski, and L. Arleth, 2014. Stealth carriers for low-resolution structure determination of membrane proteins in solution. Acta Crystallographica Section D 70:317–328.
 34. Kynde, S. A. R., N. Skar-Gislinge, M. C. Pedersen, S. R. Midtgaard, J. B. Simonsen, R. Schweins, K. Mortensen, and L. Arleth, 2014. Small-angle scattering gives direct structural information about a membrane protein inside a lipid environment. Acta Crystallographica Section D 70:371–383.
 35. Skar-Gislinge, N., S. A. R. Kynde, I. G. Denisov, X. Ye, I. Lenov, S. G. Sligar, and L. Arleth, 2015. Small-angle scattering determination of the shape and localization of human cytochrome P450 embedded in a phospholipid nanodisc environment. Acta Crystallographica Section D 71:2412–2421.
 36. Franke, D., and D. I. Svergun, 2009. *DAMMIF*, a program for rapid *ab-initio* shape determination in small-angle scattering. Journal of Applied Crystallography 42:342–346.
 37. Koutsioubas, A., and J. Perez, 2013. Incorporation of a hydration layer in the dummy atom *ab initio* structural modelling of biological macromolecules. Journal of Applied Crystallography 46:1884–1888.
 38. Koutsioubas, A., S. Jaksch, and J. Pérez, 2016. DENFERT version 2: extension of *ab initio* structural modelling of hydrated biomolecules to the case of small-angle neutron scattering data. Journal of applied crystallography 49:690–695.
 39. Debye, P., 1915. Zerstreuung von Röntgenstrahlen. Annalen der Physik 351:809–823.
 40. Svergun, D., C. Barberato, and M. H. J. Koch, 1995. *CRY SOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. Journal of Applied Crystallography 28:768–773.
 41. Sokolova, A. V., V. V. Volkov, and D. I. Svergun, 2003. Prototype of a database for rapid protein classification based on solution scattering data. Journal of Applied Crystallography 36:865–868.
 42. Lake, J., 1967. An iterative method of slit-correcting small angle X-ray data. Acta crystallographica 23:191–194.

43. Tanford, C., 1980. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes* 2d Ed. J. Wiley.
44. Oliver, R. C., J. Lipfert, D. A. Fox, R. H. Lo, S. Doniach, and L. Columbus, 2013. Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group. PLOS ONE 8:1–10.
45. Svergun, D., S. Richard, M. Koch, Z. Sayers, S. Kuprin, and G. Zaccai, 1998. Protein hydration in solution: experimental observation by x-ray and neutron scattering. Proceedings of the National Academy of Sciences 95:2267–2272.
46. Chen, P.-c., and J. S. Hub, 2015. Structural Properties of Protein–Detergent Complexes from SAXS and MD Simulations. The journal of physical chemistry letters 6:5116–5121.
47. Gonen, T., Y. Cheng, P. Sliz, Y. Hiroaki, Y. Fujiyoshi, S. C. Harrison, and T. Walz, 2005. Lipid–protein interactions in double-layered two-dimensional AQP0 crystals. Nature 438:633–638.
48. Svergun, D., 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of applied crystallography 25:495–503.
49. Volkov, V. V., and D. I. Svergun, 2003. Uniqueness of *ab initio* shape determination in small-angle scattering. Journal of Applied Crystallography 36:860–864.
50. Kozin, M. B., and D. I. Svergun, 2001. Automated matching of high- and low-resolution structural models. Journal of Applied Crystallography 34:33–41.
51. Wriggers, W., and P. Chacón, 2001. Using *Situs* for the registration of protein structures with low-resolution bead models from X-ray solution scattering. Journal of Applied Crystallography 34:773–776.
52. Schrödinger, LLC, 2015. The PyMOL Molecular Graphics System, Version 1.8.
53. Jacrot, B., 1976. The study of biological structures by neutron scattering from solution. Reports on Progress in Physics 39:911.
54. Sarachan, K. L., J. E. Curtis, and S. Krueger, 2013. Small-angle scattering contrast calculator for protein and nucleic acid complexes in solution. Journal of Applied Crystallography 46:1889–1893.
55. Clantin, B., A.-S. Delattre, P. Rucktooa, N. Saint, A. C. Méli, C. Loch, F. Jacob-Dubuisson, and V. Villeret, 2007. Structure of the Membrane Protein FhaC: A Member of the Omp85-TpsB Transporter Superfamily. Science 317:957–961.
56. Rambo, R. P., and J. A. Tainer, 2013. Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology. Annual Review of Biophysics 42:415–441.
57. Jordan, A., M. Jacques, C. Merrick, J. Devos, V. T. Forsyth, L. Porcar, and A. Martel, 2016. SEC-SANS: size exclusion chromatography combined *in situ* with small-angle neutron scattering. Journal of Applied Crystallography 49:2015–2020.

List of Figures

- 1 Ab-initio shape reconstructions of the 270 dDM/AQ0 tetramer. First row, MD structure. Second row, single SAXS curve with no imposed symmetry. Third row, single SAXS curve with imposed 4-fold axial symmetry. Fourth row, two SANS curves with no imposed symmetry. Fifth row, two SANS curves with imposed 4-fold axial symmetry. In the first three columns the top view of the different PDC components (entire PDC, detergent belt and protein respectively) are depicted, while in the last three columns the side view. The bead colours correspond to: green - protein, cyan - detergent hydrophilic heads, magenta - detergent hydrophobic tails. 10
- 2 Comparison of most probable ab-initio envelope (light gray) with the crystallographic structure of the AQ0 tetramer PDB code 2B6P (47) (black). The reconstruction belongs to the ab-initio run performed with two SANS curves as input and with imposed 4-fold symmetry. 10
- 3 Fit of the two Fhac/OG_D SANS experimental curves by the ab-initio algorithm. 12
- 4 Most probable ab-initio shape reconstruction of the 140 OG_D/Fhac PDC. First row top view. Second row side view. In each column different PDC components (entire PDC, detergent belt and protein respectively) are depicted. The bead colours correspond to: green - protein, cyan - detergent hydrophilic heads, magenta - detergent hydrophobic tails. 13
- 5 Comparison of most probable ab-initio envelope (light grey) with the crystallographic structure of the Fhac transporter PDB code 2QDZ (55) (black). Top view (left), side view (right). 13

List of Tables

1	Scattering length densities (sld) at different contrasts, electron densities (ρ) and molecular volumes (V) of protein, detergent parts and solvent, that were used during the ab-initio runs for the AQ0/dDM system. Molecular volume and sld values for the detergents were taken from the work of Breyton et al. (31). Electron densities were taken from the work of Oliver et al. (44). Note that owing to the H/D exchange occurring in solvents containing a certain ratio of D ₂ O (53), the sld of a protein is adjusted accordingly for each solvent contrast. The related calculations are performed using the contrast calculator of the program SASSIE (54)	9
2	Summary of mean NSD, its standard deviation and NSD _{cryst} values for ab-initio runs of the 270 dDM/AQ0 system with different number of curve inputs and imposed symmetries.	9
3	Scattering length densities (sld) at different contrasts and molecular volumes (V) of protein, detergent parts and solvent, that were used during the ab-initio runs for the Fhac/OG _D system. Molecular volume and sld values for the detergents were taken from the work of Breyton et al. (31). Protein sld is adjusted for each solvent contrast (54).	11

Supporting Information:

Low Resolution Structure Determination of Detergent Solubilised Membrane Proteins from Small Angle Scattering Data.

Alexandros Koutsioubas*

*Jülich Centre for Neutron Science (JCNS) at Heinz Maier-Leibnitz Zentrum (MLZ),
Forschungszentrum Jülich GmbH, Lichtenbergstr. 1, 85748 Garching, Germany*

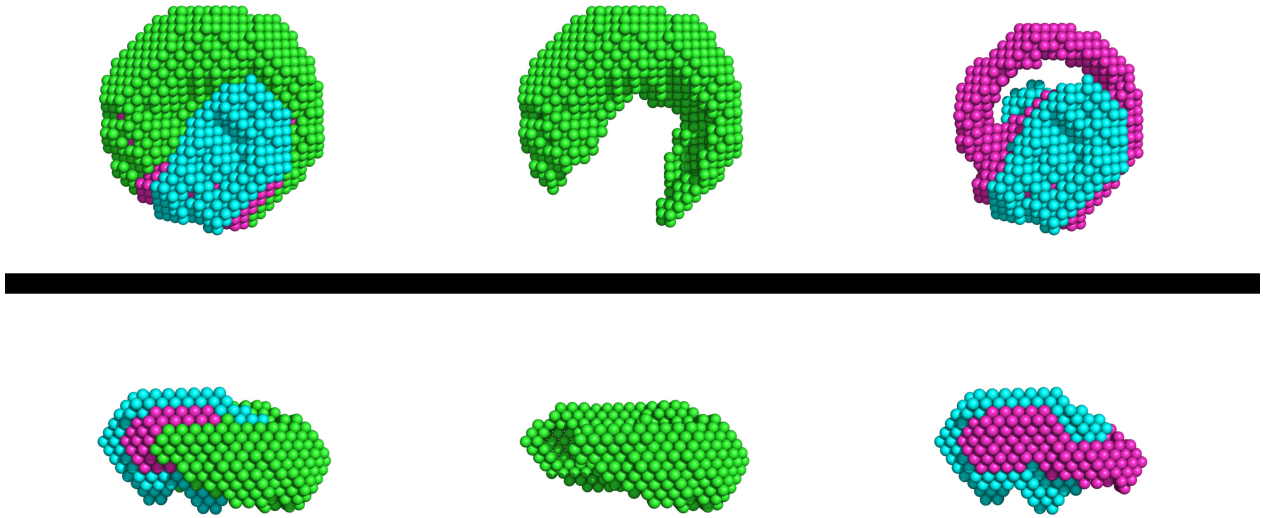


FIG. S1. Ab-initio shape reconstructions of the 270 dDM/AQ0 tetramer using the program MONSA [1], 3 SANS curves as input (0%, 42% and 100% D₂O) and default parameters concerning bead compactness and connectivity, volume fraction and center penalty. Bead colours correspond to: green-protein, cyan - detergent hydrophilic heads and magenta - detergent hydrophobic tails. The volumes of each different phase are provided as input. This is equivalent to providing the detergent aggregation number and head/tail molecular volumes, however the protein volume is an extra parameter that has to be provided to MONSA. The used scattering length densities are the same as those given in Table 1 of the manuscript. Note that due to the similar contrast of detergent heads and protein in all MONSA reconstructions the protein phase beads are positioned erroneously around the detergent tail phase while the detergent head beads form a core close to the centre of the structure. Even considering this core (cyan beads) as the structure representing the protein, we find systematic structural differences ($NSD_{cryst} > 1.0$) comparing to the Aquaporin-0 crystallographic structure. Similar MONSA reconstructions using the two OG_D/Fhac experimental SANS curves as input led always to non-physical final shapes for both the protein and the detergent belt.

* a.koutsioumpas@fz-juelich.de

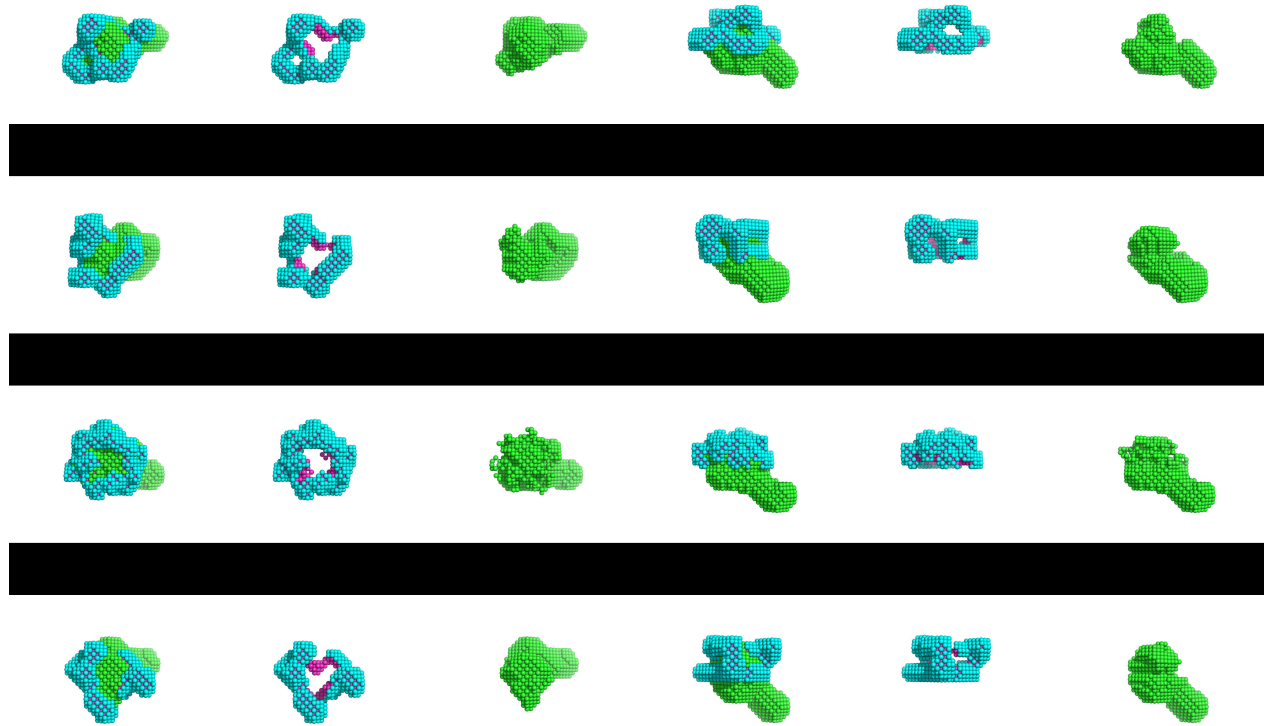


FIG. S2. Results of different ab-initio shape reconstructions of the 140 OG_D/Fhac PDC as the one presented in the manuscript's fig. 4. Each line represents a different ab-initio run. In the first three columns the top view of the different PDC components (entire PDC, detergent belt and protein respectively) are depicted, while in the last three columns the side view. The bead colours correspond to: green - protein, cyan - detergent hydrophilic heads, magenta - detergent hydrophobic tails.

-
- [1] Svergun, D. I., 1999. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophysical journal* 76:2879–2886.