Supporting Information for: Structural Properties of Protein-Detergent Complexes from SAXS and MD Simulations

Po-chia Chen[*,†,‡] and Jochen S. Hub[*,†]

†*Institute for Microbiology and Genetics, Georg-August-University Göttingen, Justus-von-Liebig weg 11, 37077 Göttingen, Germany*

‡ *Present address: Institut des Sciences Analytiques, UMR 5280, CNRS, Université de Lyon, 5 Rue de la Doua, 69100 Villeurbanne, France*

E-mail: po-chia.chen@univ-lyon1.fr; jhub@gwdg.de

Primer on included SAXS prediction tools

A small-angle X-ray scattering (SAXS) pattern of a biomolecules is obtained by measuring the difference in scattered intensities at an scattering angle 2θ, between the sample solution $I_{\text{solute}}(q)$ and its corresponding pure-buffer solution $I_{\text{buf}}(q)$. This difference intensity $I(q)$ isolates the signal contributed by the molecule, minus an excluded volume in the buffer. SAXS patterns are reported using $q = 4\pi \sin(\theta)/\lambda_{X-ray}$, rather than θ directly.

An expected SAXS pattern can be calculated for any given molecular structure, excluding influences from solute fluctuations. Explicit-solvent approaches must first hydrate the molecule, then calculate the average distribution of nearby water molecules as well as a corresponding background buffer simulation. Implicit-solvents approaches circumvent this by imposing models of buffer-subtraction and solvation-layer density, thereby requiring at least two scaling factors to parametrise (1) net solute volume and (2) excess solvation layer density. We term these C1 and C2 respectively, after AquaSAXS conventions.

We give here an executive summary of the $I(q)$ calculation and meaning of C1 and C2 parameters in CRYSOL, AquaSAXS, and FoXS, converting the source nomenclature where necessary to enable comparisons. (*E.g.*, ρ_w represents bulk density of water at $334\,\mathrm{e\,nm^{-3}}$.) Interested readers are directed to respective publications for full details.

- CRYSOL (version 2.x).[1] Intensities are calculated according to spherical multi-pole expansions of the orientational average

$$\begin{aligned} I(q) &= \int d\Omega D(\mathbf{q}) \\ &= \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{\text{solute},lm}(q) - \rho_w F_{\text{buf},lm}(q) + \delta\rho F_{\text{layer},lm}(q)|^2, \end{aligned}$$

  where $A_{\text{solute}}(q)$, $F_{\text{buf}}(q)$ and $F_{\text{layer}}(q)$ are multi-pole functions depending on the form factors of the molecule, the buffer excluded volume, and the hydration layer, respectively. All form-factors $f_i(q)$ represent expected scattering magnitude from an

atom/element as a function of q.

C1: The background buffer underlying $F_{\mathrm{buf}}(q)$ is modelled as a set of solvent particles each centered around an atom of the molecule, with Gaussian form-factors whose width and volume depend on an *effective atomic radius* r_m corresponding to the atomic group in the molecule being considered. The parameters Ra associated with the scaling the radii r_m and the "Total Excluded Volume" (the latter unique to CRYSOL) are then used to fit the calculated to the experimental curves.

C2: The hydration layer underlying $F_{\mathrm{layer}}(q)$ is modelled as a uniform shell 3 Å–thick around the solute, using $\delta\rho$ to represent excess density in $\mathrm{e\,\AA^{-3}}$.

- FoXS.[2] Intensities are calculated following the Debye formula

$$\begin{aligned} I(q) &= \sum_i \sum_j f_i(q) f_j(q) \frac{\sin(q r_{ij})}{q r_{ij}}, \\ f_i(q) &= f_i^v(q) - F_{\mathrm{buf}}(c1, q) f_i^s(q) + c2 s_i f^w(q) \end{aligned}$$

C1: $F_{buf}(c1, q)$ is calculated equivalently to CRYSOL, with $0.95 < c1 < 1.05$ representing scaling of the effective atomic radius r_m (fixed in FoXS to the average atomic radius of the molecule).

C2: Instead of evaluating a separate $F_{layer}(q)$, additional *water* form-factors $f^w(q)$ are placed upon solvent-exposed atoms of the molecule, scaled according to solvent exposure ratio s_i. The overall scaling factor $-2.0 < c2 < 4.0$ is modulated such that $c2 = 0$ corresponds to $334\,\mathrm{e\,nm^{-3}}$ (bulk electron density of water), and $c2 = 1$ corresponds to $347.5\,\mathrm{e\,nm^{-3}}$.

- AquaSAXS.[3] Intensities are calculated following an explicit orientational average via

a cubature formulae

$$\begin{aligned} I(q) &= \int d\Omega A(\mathbf{q}) A^*(\mathbf{q}), \\ A(\mathbf{q}) &= A_{\text{solute}}(\mathbf{q}) - \rho_w F_{buf}(\mathbf{q}) + \rho_w F_{\text{layer}}(\mathbf{q}), \\ A_{\text{solute}}(\mathbf{q}) &= \sum_i f_j(q) e^{i\mathbf{q}.\mathbf{r}} \end{aligned}$$

C1: $F_{buf}(\mathbf{q})$ is calculated in an equivalent manner to FoXS. The AquaSAXS $0.9 < C1 < 1.12$ is the equivalent of $c1$ in FoXS.

C2: By default, solvent-shell density is modelled in $F_{\text{layer}}(\mathbf{q})$ by a solvent-density map derived by the Poisson-Boltzmann formalism using an accompanying program AquaSol. Non-zero excess densities (relative to bulk ρ_w) are summed and scaled with parameter $0.0 < C2 < 1.4$, such that $C2 = 1$ represents the ideal solvent-shell contribution expected for a protein.

- WAXSiS.[4] Intensities are calculated following an explicit orientational average, numerically over a set of scattering angles $\mathbf{q}$ distributed along a spiral on a sphere of radius q.

$$\begin{aligned} I(q) &= \int d\Omega D(\mathbf{q}) \\ D(\mathbf{q}) &= \left\langle |A_{\text{solute}}(\mathbf{q})|^2 \right\rangle^{(\omega)} - \left\langle |A_{\text{buf}}(\mathbf{q})|^2 \right\rangle^{(\omega)} \\ &+ 2Re\left[- \left\langle A^*_{\text{buf}}(\mathbf{q}) \right\rangle^{(\omega)} \left\langle A_{\text{solute}}(\mathbf{q}) - A_{\text{buf}}(\mathbf{q}) \right\rangle^{(\omega)} \right] \\ A(\mathbf{q}) &= \sum_i f_j(q) e^{i\mathbf{q}.\mathbf{r}} \end{aligned}$$

The conformational averages $\langle . \rangle$ is taken in the solute reference orientation (ω), and backbone-restrained MD simulations are used to sample both solvent and side-chain fluctuations. For A_{solute}, explicit solvent up to a preselected distance from the solute is included (7Å by default), defined by a spatial envelope. Because the MD force field yields a physically correct model for the hydration layer, no fitting parameter

associated with the hydration layer is needed. A_{buf} is computed from an explicit-solvent inside the same envelope, taken from an pure-water MD simulation. Likewise, no fitting of the excluded solvent is required. A small correction to the solvent density ensures that the bulk density in the solute and solvent simulations match exactly with the experimental value of $334\,e\,\text{nm}^{-3}$ but this correction is not used to fit the calculated against the experimental SAXS curve.[5] Instead, only a constant offset c to the experimental SAXS curve is fitted to account for small uncertainties in the buffer subtraction, via $I_{\text{exp, fit}} = fI_{\text{exp}} + c$. Hence, there is no correspondence for C1 and C2 in WAXSiS.

- Ensemble MD.[5] The formulation in WAXSiS is adopted, *sans* constraints on backbone atoms.

Supplementary methods

SAXS prediction methods examined and parameters used Atom naming conventions from the CHARMM forcefield were first standardised to follow PDB conventions as closely as possible. CRYSOL2.8 calculations were run with arguments:

```
/lm 30 /fb 19 /sm 0.5 /ns 101 /un 1 /err /kp /eh
```

The use of constant background subtraction `/cst` resulted in a minor but consistent improvement of χ. These results are not shown in this work. We note in particular that an increased number of spherical harmonics (option `/lm`) was required to reflect intensities beyond $\sim 3\text{Å}^{-1}$, and hydrogens must be included to correctly reflect lipid-tail intensities (option `/eh`). Calculations with fixed hydration-shell densities are carried out at $\delta\rho = 0.033$ e Å^{-3}.

FoXS calculations were run with arguments:

```
--max_q=0.5 --profile_size=100 --hydrogens --offset
```

In particular, hydrogens must be included to correctly reflect lipid-tail intensities. Calculations with fixed hydration-shell densities are carried out at $c2 = 2.4$, corresponding roughly to $0.0324\,\mathrm{e\,\AA^{-1}}$.

AquaSAXS calculations were carried out using solvent maps generated by AquaSol, with arguments $q_{max} = 0.5\,\text{Å}$ and $q_{step} = 0.005\,\text{Å}$. A parameter file for βDDM was added to specify correct scattering types. Due to the large size of PDC complexes, grid dimensions of 65 and 129 points (corresponding to a resolution of 2.6 and 1.3 Å respectively,) were tested with other parameters set at server defaults. For AquaSol, ion concentrations were set at 100 mM, and solvent dipoles were set at 2.8 Debye and 55 molar. The two grid resolutions resulted in qualitatively similar curves with an offset of fitted parameters values. For brevity, only the results for 65 points have been presented. Calculations with fixed hydration-shell densities are carried out at $C2 = 1.0$.

The default AXES server settings are modified follows: $nFib = 19$, $q_{\mathrm{minfit}} = 0.02$, $q_{\mathrm{maxfit}} = 0.24$. No buffer curves were given to AXES because the experimental data was provided as a pre-subtracted curve.

Default WAXSiS server settings are modified as follows: $q_{max} = 0.5\,\text{Å}$, 8 Å solvation layer thickness, total buffer scattering subtracted, thorough convergence, and new random seeds.

Other SAXS programs Initial tests with AXES led to strong artefacts at the detergent–detergent length scale, which we discussed with Dr. Grishaev.[6] Further examination revealed that the values of excluded-volume radii used for protein atom types were too small when transferred to βDDM lipid tails, resulting in significant penetration of water into the lipid interior. Thus, application of AXES to lipids require a significant reparametrisation effort not available at the time of writing.

Although SoftWAXS reproduced both the prominent minima and twin-peak characteristics of Aqp0-βDDM complexes, we found that the relative magnitudes do not match with other methods when default parameters values were used.[7] Therefore, SoftWAXS was not

further considered for this study. Finally, ScÅtter[8] was also briefly considered, but its Java platform was found not suitable for automation of workflow necessary to tackle $\sim 10^3$ structures.

Fitting details χ agreement with the experiment were calculated using the methodology provided by the SAXS predictor, where available. Although this precluded direct comparisons, it provided a more faithful test of the authors' methods. SAXS calculations based on MD ensembles were fitted as follows:

$$\chi^2 = \sum_{i}^{N_q} \Big[\frac{\log I_{com}(q_i) - \log(f\, I_{exp}(q_i) + c)}{\sigma_i / I_{exp}} \Big]^2, \tag{S1}$$

where N_q of experiment is interpolated to match computed q-points.

Contributions of pure-DDM micelles to the PDC SAXS curve

The experimental setup coupling size–exclusion chromatography (SEC) to the SAXS measurement filters out pre-existing pure-DDM micelles in the sample solution, leaving contributions from spontaneously reforming micelles between SEC column exit and SAXS measurement ($\sim$20 sec. according to private communications with Dr. Pérez). The spontaneously-forming micelles from the elution buffer are further subtracted by measuring SAXS curves of the empty column or fractions prior to the eluted PDCs. This leaves only contributions due to detergent movement between eluted PDCs and background DDM-micelles.

We tested for the presence of such contaminations in the experimental SAXS curve, using the DDM micelle data from Lipfert *et al.*,[9] and taking the SAXS curve at 150 mM as a guide (Fig. S1). The fit-equation (S1) is modified to add a second scaling factor to account for the component of pure DDM added/subtracted and its scaling to the computed SAXS in

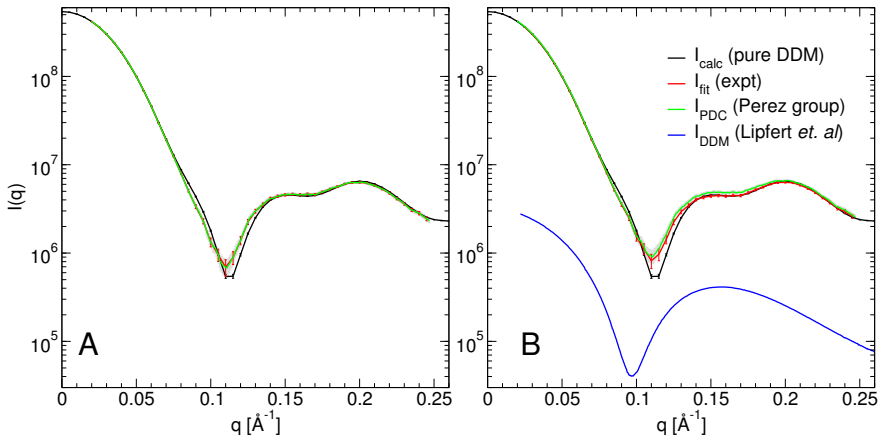

Figure S1. Optimised $I(q)$ fit to the 290-ensemble using only the experimental. PDC SAXS curve (A), or both the PDC and pure-DDM SAXS curve (B). Colours as follows: (black) predicted SAXS curve of a 290-PDC ensemble, (red) optimised fit of experimental curves, (green) scaled experimental PDC SAXS curve after fit with errorbars in grey, and (blue) scaled experimental pure-DDM SAXS curve after it.

absolute scale:

$$\chi^2 = \sum_i^{N_q} \Big[\frac{\log I_{\mathrm{PDC,com}}(q_i) - \log(f_{\mathrm{PDC}}\, I_{\mathrm{PDC}}(q_i) + f_{\mathrm{DDM}}\, I_{\mathrm{DDM}}(q_i) + c)}{\sigma_{\mathrm{PDC}}(q_i)/I_{\mathrm{PDC}} + \sigma_{\mathrm{DDM}}(q_i)/I_{\mathrm{DDM}}} \Big]^2, \tag{S2}$$

The results in the table and figure below shows that our best-fit 290-ensemble is further improved in χ from 2.17 to 2.01, by subtracting a small amount of pure DDM from the experimental PDC measurement (Fig. S1 and Table S1). While the ratios of f_{DDM}:f_{PDC} cannot be quantified into pure-DDM:PDC ratios without exact knowledge of experimental $I(0)$, the converted absolute $I(q)$ suggests an upper-bound of of $< 1\%$ pure-DDM in the reported experimental curve.

However, we warn that an alternative explanation in the form of over-fitting must also be considered. Other $N_{\beta\mathrm{DDM}}$ fits also show measurable improvement in χ, and a large variation in calculated f_{DDM} (Table S1). This supports the over-fitting hypothesis, which cannot excluded for the 290-ensemble result. Thus, it is unlikely that residual pure-DDM signals will contribute significantly to the SAXS pattern.

Table S1. Optimised scaling and fit parameters for SAXS-fit, using the 5×90–100 ns ensemble. Two fitting models are presented, with just the experimental PDC SAXS curve, or with both PDC and pure-DDM SAXS curves.

N_{DDM}	χ	f_{PDC}	f_{DDM}	c
250	11.49	2.27×10^{9}		-5.9×10^{5}
	10.76	2.26×10^{9}	6.9×10^{5}	-9.8×10^{5}
270	5.78	2.39×10^{9}		-3.8×10^{5}
	5.65	2.41×10^{9}	-4.4×10^{5}	-2.0×10^{5}
290	2.17	2.50×10^{9}		-2.5×10^{5}
	2.01	2.52×10^{9}	-5.7×10^{5}	-2.7×10^{4}
310	5.44	2.59×10^{9}		-1.7×10^{5}
	4.13	2.65×10^{9}	-1.8×10^{6}	4.7×10^{5}
330	9.92	2.70×10^{9}		-1.5×10^{5}
	7.86	2.82×10^{9}	-3.4×10^{6}	8.7×10^{5}

Supporting Information Figures

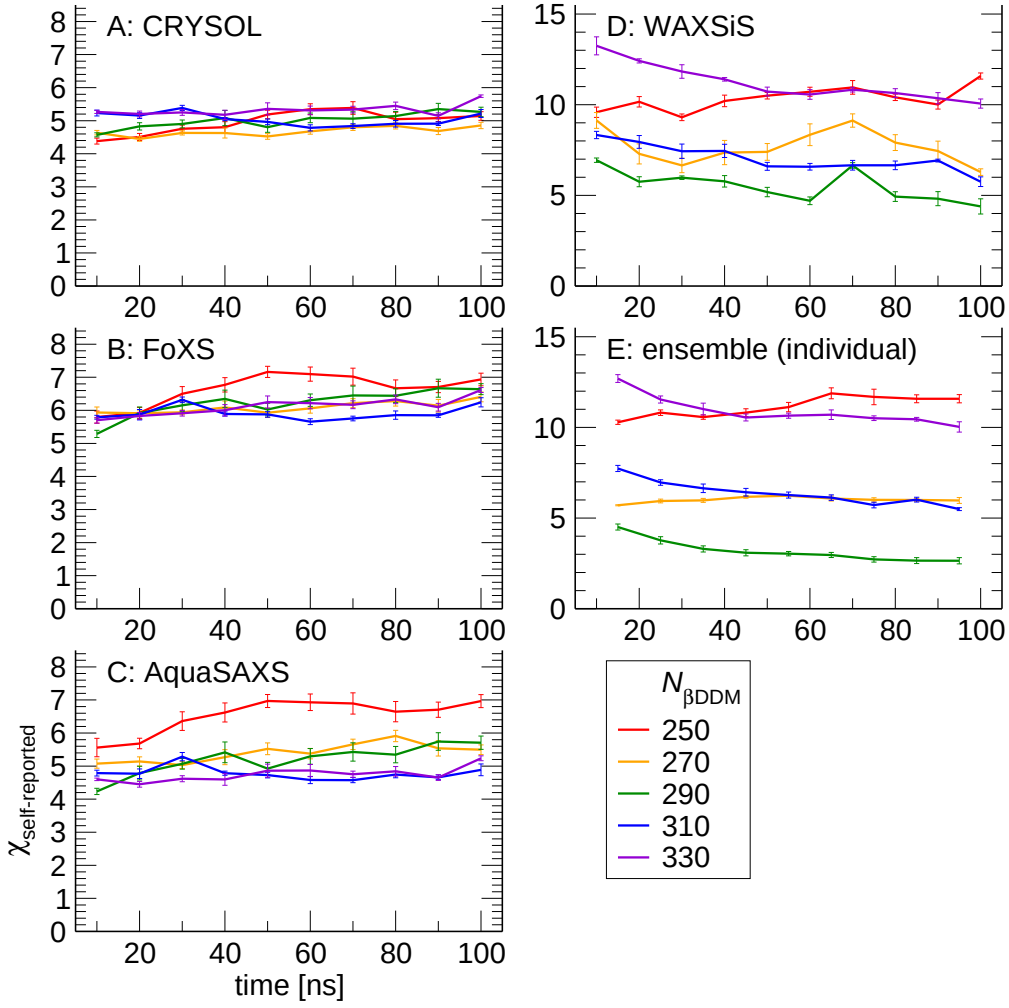

Figure S2. SAXS predictions based on PDC structures at different simulation times, with χ-valued averaged over 5 replica trajectories. (A-D) Single-structure predictions methods. SAXS software used are as labelled on the figure. (E) Ensemble SAXS predictions calculated over 10 ns chunks. χ values stabilise after ~70ns.

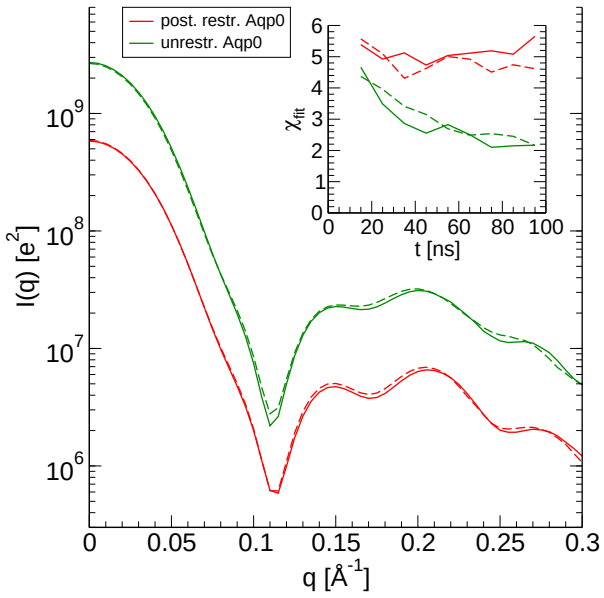

Figure S3. Influence of backbone position restraints upon SAXS predictions. Main figure: Two SAXS patterns of Aqp0-βDDM complexes, using individual trajectories between 90 and 100 ns, with backbone restraints of 1000 kcal mol^{-1} Å^{-2} (red), and without (green). Green curves have been offset vertically for visual clarity. Inset: χ-agreement with the experiment SAXS curve, using 10 ns chunks of individual trajectory data as done for Fig. S2E

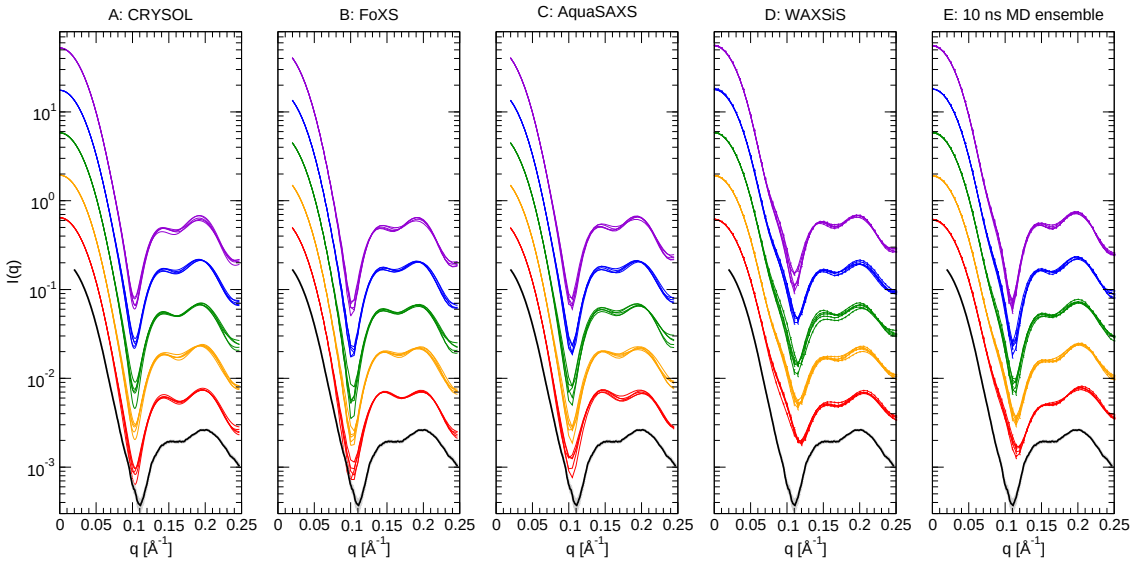

Figure S4. The fitted SAXS intensities used for χ statistics in Fig. 2A-E of the main text. (A) CRYSOL, (B) FoXS, (C) AquaSAXS, (D) WAXSiS, and (E) MD ensembles. Colors indicate aggregation number (red: 250, orange: 270, green: 290, blue: 310, violet: 330, and black: experimental SAXS as reference). Curves have been offset vertically for visual clarity.

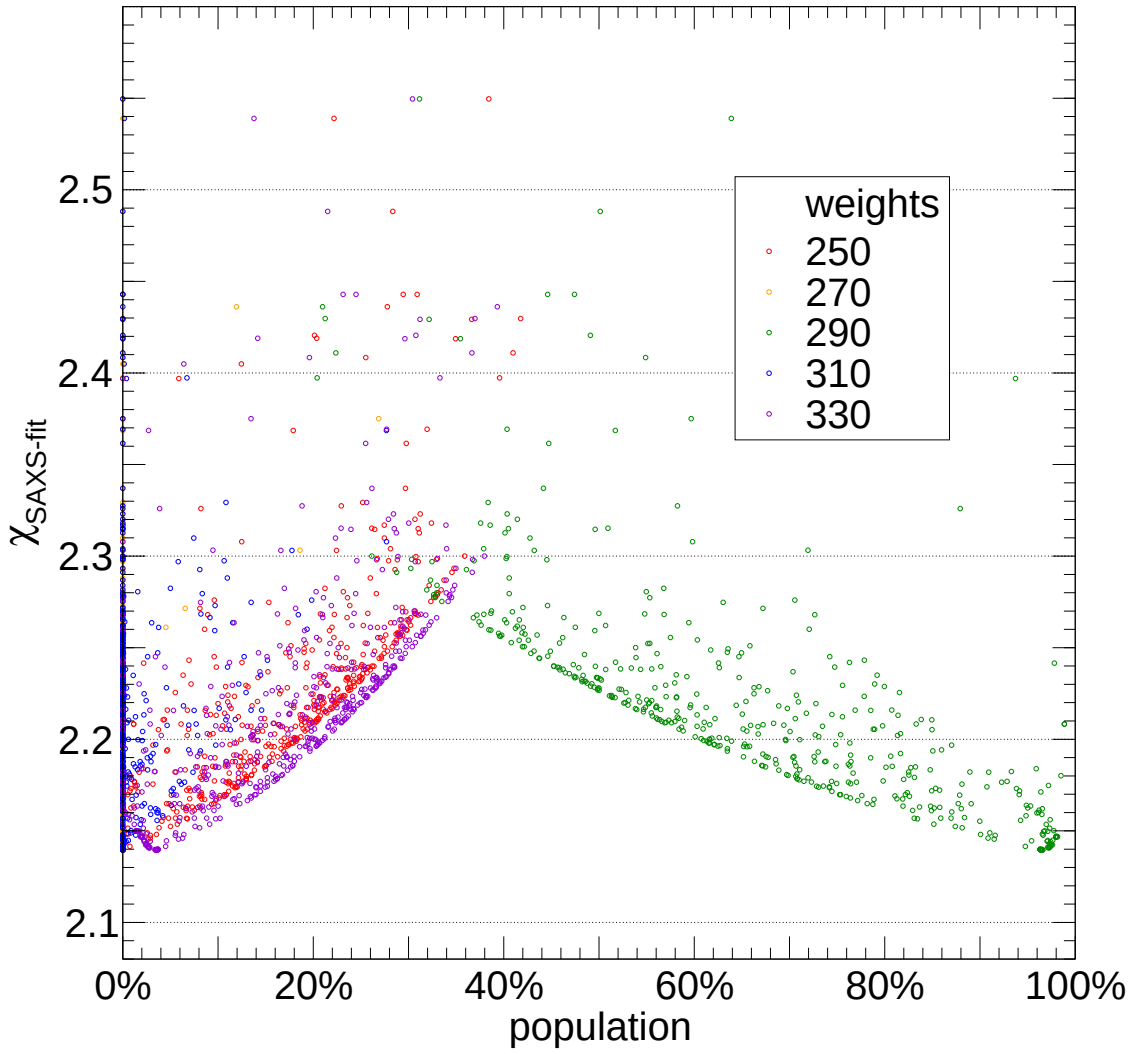

Figure S5. Population modelling of aggregation number $N_{\beta\mathrm{DDM}}$ distribution against SAXS data. A single aggregate I(q) per $N_{\beta\mathrm{DDM}}$ was calculated by collating the 5×10–100ns trajectory segments, corresponding to Fig. 2G in the main text. Using the metric $\chi_{fit} = \Sigma_{N_{\mathrm{DDM}}} w_{\mathrm{DDM}} I_{\mathrm{DDM}}(q) - f I_{exp} q + c$, 500 minimisation trials were conducted over six parameters (w_{250}, w_{270}, w_{310}, w_{330}, f, and c), while fixing w_{290} to normalise total weight. Initial values of w_{DDM} were assigned by random, while $f = I_{total}(0)$ and $c = 0$. The results of each trial is plotted by 5 circles according to final χ_{fit} and w_{DDM}.

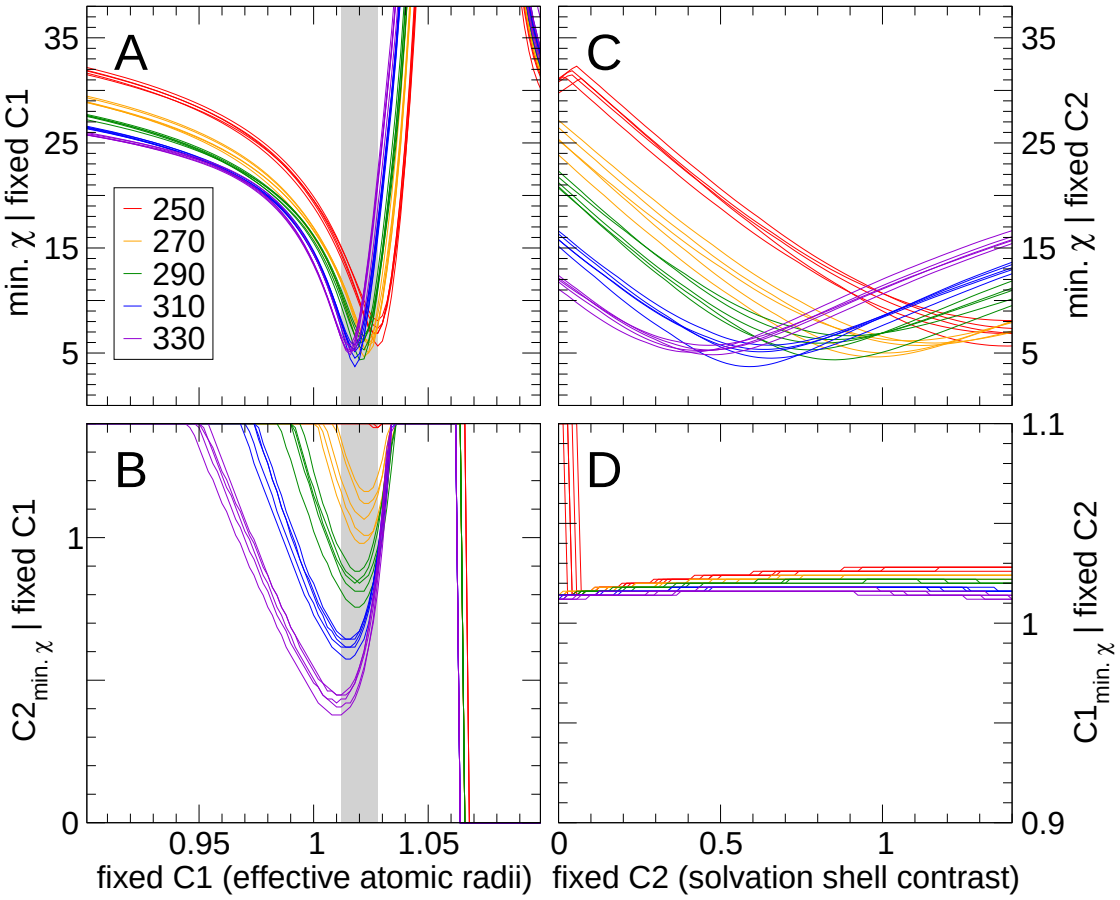

Figure S6. Dependence of χ_{AquaSAXS} upon $N_{\beta\mathrm{DDM}}$, C1, and C2, using AquaSAXS fitting of PDCs after 100 ns of MD simulations. Colours indicate $N_{\beta\mathrm{DDM}}$, and given in the legends. (A) minimum χ given a fixed C1, (B) the optimised C2 value corresponding to minimum χ, (C) minimum χ given a fixed C2, and (D) the optimised C1 value corresponding to minimum χ. To guide the eye, a grey area in (A) and (B) is shown, indicating the range of C1 values found in (D).

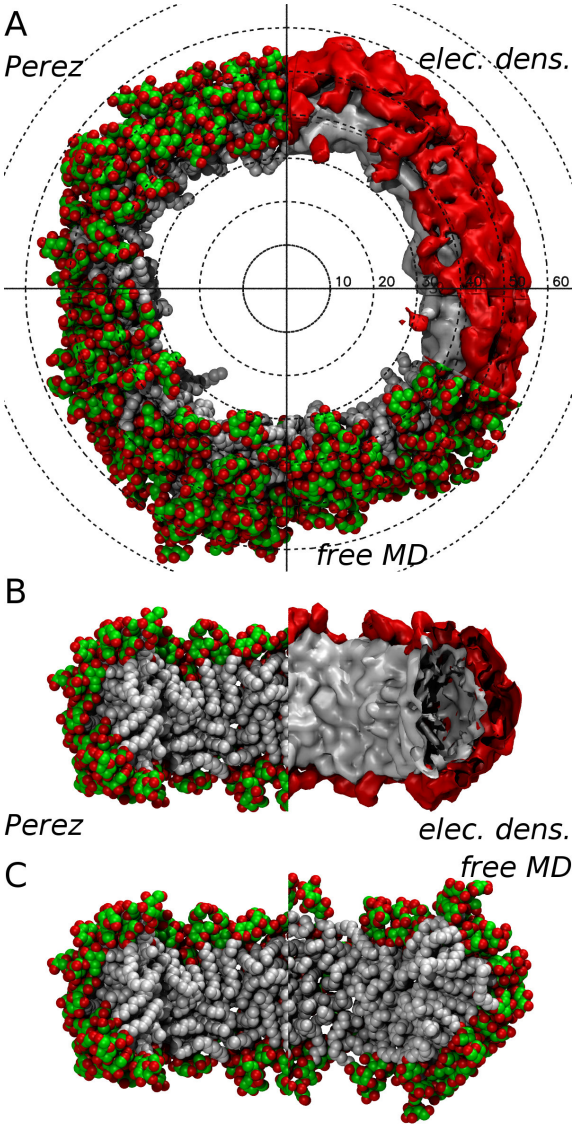

Figure S7. Comparison of corona structure between a free-MD snapshot, the ensemble electron density, and the proposed modelled structure by Pérez *et al.*[10] $N_{\beta\mathrm{DDM}}$=290 for this work, and $N_{\beta\mathrm{DDM}}$=270 for the previous model. Aq0 has been removed for clarity. Isosurfaces of electron densities have seen set at 0.277 (lipid tails) and 0.520 (detergent head group) e Å^{-3} based on shell models in pure micelles.[11] (A) View from top, units in Å. (B) Side-view of previous model (left) versus derived electron density (right). (C) Side-view of previous model (left) versus our free-MD snapshot (right).

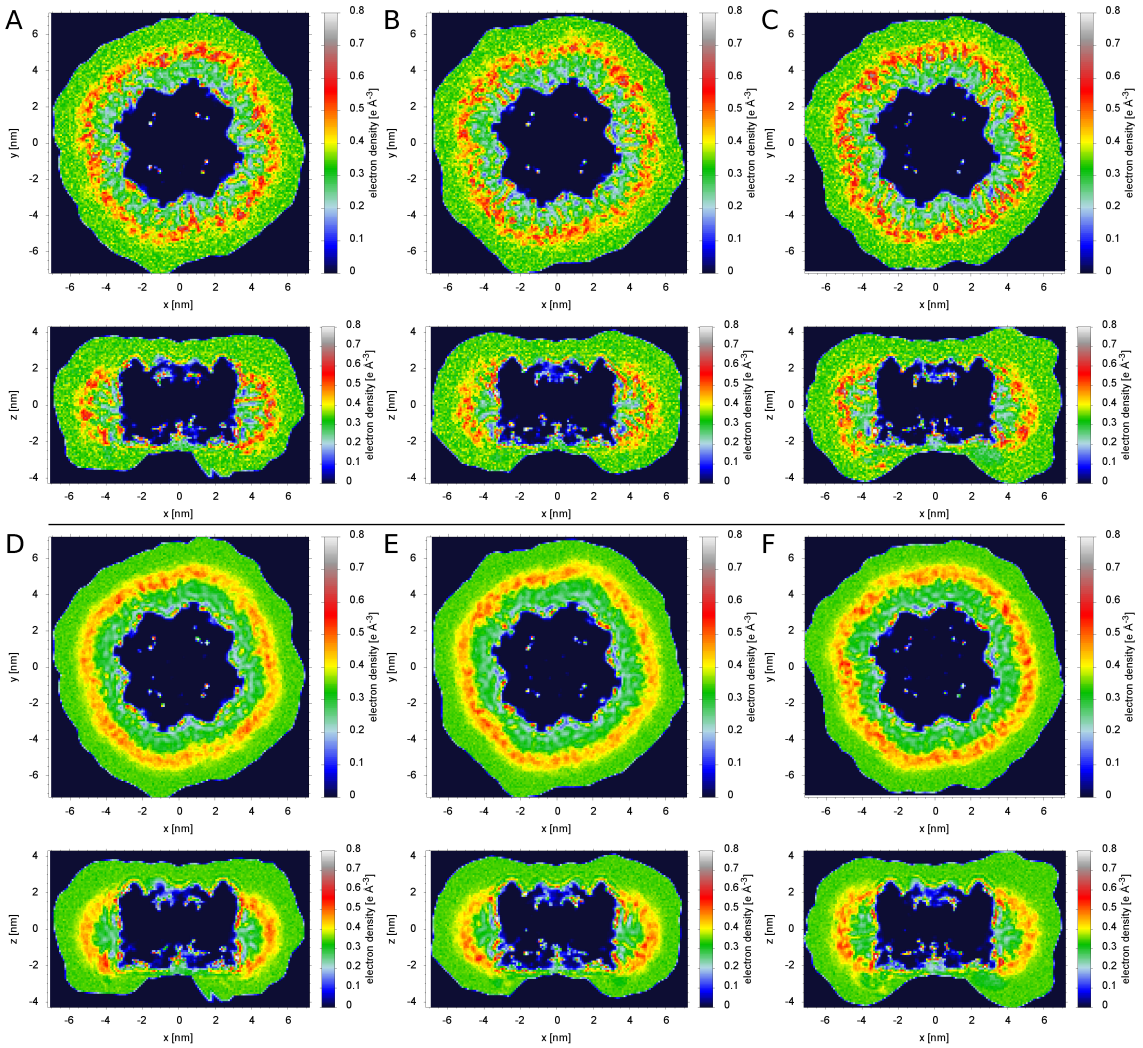

Figure S8. Average electron density of the βDDM corona and solvent around Aqp0 during MD simulations, measured in a 1 Å-thick cross section across the center of Aqp0 transmembrane helices, parallel and transverse to the channel axes. The yellow/red bands indicate densities of DDM head groups, while their encapsulated volumes within indicate the lipid tail densities between Aqp0 (black) and DDM head groups.
(A-C) Combined data from 5×10-20 ns trajectories, for $N_{\beta\mathrm{DDM}}$ of 270, 290 and 310. (D-F) As (A-C), but with additional data from 5×10-100 ns trajectories. Detergent structures are partially masked in the 90 ns average, but are visible in the 10 ns average.

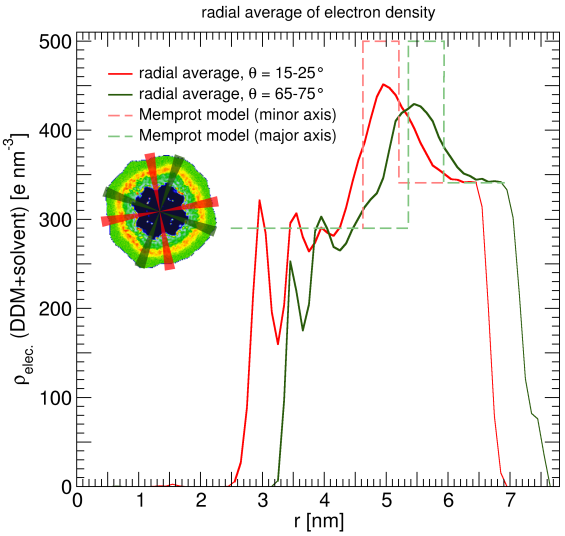

Figure S9. 1-D electron density profiles of the DDM corona, based on angular averaging of the transverse slice in Fig. S8E as shown in the inset. Here, $4\times10°$-sectors are averaged to obtain the profile of the corona at its broadest (solid green) and narrowest (solid red) extent so that an anisotropy parameter e can be obtained. For comparison, the parameters for a Memprot model (Pérez and Koutsioubas, 2015, Figure 6, row 6)[12] is also visualised, with parameters $a = 2.92\,\mathrm{nm}$, $b = 3.51\,\mathrm{nm}$, $t = 0.58\,\mathrm{nm}$, $e = 1.11$, $\rho_{\mathrm{tail}} = 290\,\mathrm{e\,nm^{-3}}$, and $\rho_{\mathrm{head}} = 500\,\mathrm{e\,nm^{-3}}$. Its narrowest (dashed red) and broadest (dashed green) extent corresponds to $b/e + a/2$ and $be + a/2$.

The MD data can be estimated via visual inspection to have parameters $a \sim 3.4\,\mathrm{nm}$, $b \sim 3.0\,\mathrm{nm}$, $t = 1.5\,\mathrm{nm}$, $e = 1.08$, $\rho_{\mathrm{tail}} \sim 290\,\mathrm{e\,nm^{-3}}$ and $\rho_{\mathrm{head}} \sim 400\,\mathrm{e\,nm^{-3}}$, noting that the MD result is 4-fold symmetric, while the Memprot model is 2-fold symmetric. The larger t captures the mobile and diffuse headgroup layer mixed in with the solvation shell, the latter of which is not explicitly included in Memprot. Since a expresses the thickness of the corona lipid-tail region along the transmembrane axis, and $b+a/2$ its transverse diameter, the MD-corona is thus significantly less disc-like than the Memprot-corona, with no visible cylindrical component b.

References

(1) Svergun, D.; Barberato, C.; Koch, M. H. J. CRYSOL –a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28*, 768–773.

(2) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophys. J.* **2013**, *105*, 962–974.

(3) Poitevin, F.; Orland, H.; Doniach, S.; Koehl, P.; Delarue, M. AquaSAXS: A Web Server for Computation and Fitting of SAXS Profiles with Non-Uniformally Hydrated Atomic Models. *Nucl. Acids Res.* **2011**, *39*, W184–W189, PMID: 21665925.

(4) Knight, C. J.; Hub, J. S. WAXSiS: A Web Server for the Calculation of SAXS/WAXS Curves Based on Explicit-Solvent Molecular Dynamics. *Nucl. Acids Res.* **2015**, gkv309, PMID: 25855813.

(5) Chen, P.-c.; Hub, J. S. Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data. *Biophys. J.* **2014**, *107*, 435–447.

(6) Grishaev, A.; Guo, L.; Irving, T.; Bax, A. Improved Fitting of Solution X-ray Scattering Data to Macromolecular Structures and Structural Ensembles by Explicit Water Modeling. *J. Am. Chem. Soc.* **2010**, *132*, 15484–15486.

(7) Bardhan, J.; Park, S.; Makowski, L. SoftWAXS: A Computational Tool for Modeling Wide-Angle X-Ray Solution Scattering from Biomolecules. *J. Appl. Crystallogr.* **2009**, *42*, 932–943, PMID: 21339902 PMCID: PMC3041499.

(8) Classen, S.; Hura, G. L.; Holton, J. M.; Rambo, R. P.; Rodic, I.; McGuire, P. J.; Dyer, K.; Hammel, M.; Meigs, G.; Frankel, K. A. et al. Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J. Appl. Crystallogr.* **2013**, *46*, 1–13.

(9) Lipfert, J.; Columbus, L.; Chu, V. B.; Lesley, S. A.; Doniach, S. Size and Shape of Detergent Micelles Determined by Small-Angle X-ray Scattering. *J. Phys. Chem. B* **2007**, *111*, 12427–12438.

(10) Koutsioubas, A.; Berthaud, A.; Mangenot, S.; Pérez, J. Ab Initio and All-Atom Modeling of Detergent Organization around Aquaporin-0 Based on SAXS Data. *J. Phys. Chem. B* **2013**, *117*, 13588–13594.

(11) Oliver, R. C.; Lipfert, J.; Fox, D. A.; Lo, R. H.; Doniach, S.; Columbus, L. Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group. *PLoS ONE* **2013**, *8*, e62488.

(12) Pérez, J.; Koutsioubas, A. *Memprot*: A Program to Model the Detergent Corona Around a Membrane Protein Based on SEC–SAXS Data. *Acta Crystallogr. D* **2015**, *71*, 86–93.