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 $2\theta$, 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 $q = 4\pi \sin(\theta)/\lambda_{\text{X-ray}}$, rather than $\theta$ 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., $\rho_w$ represents bulk density of water at 334 e nm$^{-3}$.) Interested readers are directed to respective publications for full details.

- CRYSOL (version 2.x).\textsuperscript{1} Intensities are calculated according to spherical multi-pole expansions of the orientational average

\[
I(q) = \int d\Omega D(q) = \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{\text{solute},lm}(q) - \rho_w F_{\text{buf},lm}(q) + \delta \rho F_{\text{layer},lm}(q)|^2;
\]

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 $F_{\text{buf}}(q)$ 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 $F_{\text{layer}}(q)$ is modelled as a uniform shell 3 Å–thick around the solute, using $\delta\rho$ to represent excess density in e Å$^{-3}$.

- FoXS.$^2$ Intensities are calculated following the Debye formula

$$
I(q) = \sum_i \sum_j f_i(q)f_j(q)\frac{\sin(qr_{ij})}{qr_{ij}},
$$

$$
f_i(q) = f_i^w(q) - F_{\text{buf}}(c1,q)f_i^s(q) + c2s_if_i^w(q)
$$

C1: $F_{\text{buf}}(c1,q)$ 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 $F_{\text{layer}}(q)$, additional *water* form-factors $f_i^w(q)$ 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 e nm$^{-3}$ (bulk electron density of water), and $c2 = 1$ corresponds to 347.5 e nm$^{-3}$.

- AquaSAXS.$^3$ Intensities are calculated following an explicit orientational average via
a cubature formulae

\[ I(q) = \int d\Omega A(q)A^*(q), \]
\[ A(q) = A_{\text{solute}}(q) - \rho_w F_{\text{buf}}(q) + \rho_w F_{\text{layer}}(q), \]
\[ A_{\text{solute}}(q) = \sum_i f_j(q)e^{i\mathbf{q} \cdot \mathbf{r}} \]

C1: \( F_{\text{buf}}(q) \) 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 \( F_{\text{layer}}(q) \) by a solvent-density map derived by the Poisson-Boltzmann formalism using an accompanying program AquaSol. Non-zero excess densities (relative to bulk \( \rho_w \)) 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 \( q \) distributed along a spiral on a sphere of radius \( q \).

\[ I(q) = \int d\Omega D(q) \]
\[ D(q) = \langle |A_{\text{solute}}(q)|^2 \rangle^{(\omega)} - \langle |A_{\text{buf}}(q)|^2 \rangle^{(\omega)} \]
\[ + 2Re \left[ -\langle A_{\text{buf}}^*(q) \rangle^{(\omega)} \langle A_{\text{solute}}(q) - \rho_w F_{\text{buf}}(q) \rangle^{(\omega)} \right] \]
\[ A(q) = \sum_i f_j(q)e^{i\mathbf{q} \cdot \mathbf{r}} \]

The conformational averages \( \langle . \rangle \) is taken in the solute reference orientation (\( \omega \)), and backbone-restrained MD simulations are used to sample both solvent and side-chain fluctuations. For \( A_{\text{solute}} \), 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. $A_{\text{buf}}$ 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 $I_{\text{exp, fit}} = f I_{\text{exp}} + c$. 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 $\chi$. 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 $\delta \rho = 0.033 \, e\, \text{Å}^{-3}$.

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 eÅ$^{-1}$.

AquaSAXS calculations were carried out using solvent maps generated by AquaSol, with arguments $q_{\text{max}} = 0.5 \text{ Å}$ and $q_{\text{step}} = 0.005 \text{ Å}$. A parameter file for $\beta$DDM 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 as follows: $nFib = 19$, $q_{\text{minfit}} = 0.02$, $q_{\text{maxfit}} = 0.24$. 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: $q_{\text{max}} = 0.5 \text{ Å}$, 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 $\beta$DDM 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, ScAtter was also briefly considered, but its Java platform was found not suitable for automation of workflow necessary to tackle ∼10³ structures.

**Fitting details**  \( \chi \) 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:

\[
\chi^2 = \sum_i \left[ \frac{\log I_{\text{com}}(q_i) - \log(f I_{\text{exp}}(q_i) + c)}{\sigma_i / I_{\text{exp}}} \right]^2,
\]

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 (∼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., 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.

The results in the table and figure below shows that our best-fit 290-ensemble is further improved in $\chi$ 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 $N_{DDM}$ fits also show measurable improvement in $\chi$, 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.

<table>
<thead>
<tr>
<th>$N_{DDM}$</th>
<th>$\chi$</th>
<th>$f_{PDC}$</th>
<th>$f_{DDM}$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>11.49</td>
<td>2.27×10^9</td>
<td>-5.9×10^5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.76</td>
<td>2.26×10^9</td>
<td>6.9×10^5</td>
<td>-9.8×10^5</td>
</tr>
<tr>
<td>270</td>
<td>5.78</td>
<td>2.39×10^9</td>
<td>-3.8×10^5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.65</td>
<td>2.41×10^9</td>
<td>-4.4×10^5</td>
<td>-2.0×10^5</td>
</tr>
<tr>
<td>290</td>
<td>2.17</td>
<td>2.50×10^9</td>
<td>-2.5×10^5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.01</td>
<td>2.52×10^9</td>
<td>-5.7×10^5</td>
<td>-2.7×10^4</td>
</tr>
<tr>
<td>310</td>
<td>5.44</td>
<td>2.59×10^9</td>
<td>-1.7×10^5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.13</td>
<td>2.65×10^9</td>
<td>-1.8×10^6</td>
<td>4.7×10^5</td>
</tr>
<tr>
<td>330</td>
<td>9.92</td>
<td>2.70×10^9</td>
<td>-1.5×10^5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.86</td>
<td>2.82×10^9</td>
<td>-3.4×10^6</td>
<td>8.7×10^5</td>
</tr>
</tbody>
</table>
Figure S2. SAXS predictions based on PDC structures at different simulation times, with $\chi$-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. $\chi$ values stabilise after $\sim$70 ns.
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: \(\chi\)-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 \(\chi\) 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 $N_{βDDM}$ distribution against SAXS data. A single aggregate $I(q)$ per $N_{βDDM}$ was calculated by collating the $5\times10^{10}$ns trajectory segments, corresponding to Fig. 2G in the main text. Using the metric $\chi_{fit} = \Sigma w_{DDM} I_{DDM}(q) - f I_{exp} q + c$, 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 $\chi_{fit}$ and $w_{DDM}$. 
Figure S6. Dependence of $\chi_{\text{AquaSAXS}}$ upon $N_{\text{DDM}}$, C1, and C2, using AquaSAXS fitting of PDCs after 100 ns of MD simulations. Colours indicate $N_{\text{DDM}}$, and given in the legends. (A) minimum $\chi$ given a fixed C1, (B) the optimised C2 value corresponding to minimum $\chi$, (C) minimum $\chi$ given a fixed C2, and (D) the optimised C1 value corresponding to minimum $\chi$. 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.\textsuperscript{10} $N_{\beta\text{DDM}}=290$ for this work, and $N_{\beta\text{DDM}}=270$ 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) eÅ\textsuperscript{−3} based on shell models in pure micelles.\textsuperscript{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 $5 \times 10^{-20}$ ns trajectories, for $N_{\beta\text{DDM}}$ of 270, 290 and 310. (D-F) As (A-C), but with additional data from $5 \times 10^{-10}$-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, $4 \times 10^3$-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 Koutsoubas, 2015, Figure 6, row 6)\textsuperscript{12} is also visualised, with parameters $a = 2.92$ nm, $b = 3.51$ nm, $t = 0.58$ nm, $e = 1.11$, $\rho_{\text{tail}} = 290$ e nm$^{-3}$, and $\rho_{\text{head}} = 500$ 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 $a \sim 3.4$ nm, $b \sim 3.0$ nm, $t = 1.5$ nm, $e = 1.08$, $\rho_{\text{tail}} \sim 290$ e nm$^{-3}$ and $\rho_{\text{head}} \sim 400$ 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


