CHAPTER FIFTEEN

Predicting solution scattering patterns with explicit-solvent molecular simulations

Leonie Chatzimagas and Jochen S. Hub*
Theoretical Physics and Center for Biophysics, Saarland University, Saarbrücken, Germany
*Corresponding author: e-mail address: jochen.hub@uni-saarland.de

Contents

Abstract

Small-angle X-ray or neutron scattering (SAXS/SANS/SAS) is widely used to obtain structural information on biomolecules or soft-matter complexes in solution. Deriving a molecular interpretation of the scattering signals requires methods for predicting SAS patterns from a given atomistic structural model. Such SAS predictions are nontrivial because the patterns are influenced by the hydration layer of the solute, the excluded

solvent, and by thermal fluctuations. Many computationally efficient methods use simplified, implicit models for the hydration layer and excluded solvent, leading to some uncertainties and to free parameters that require fitting against experimental data. SAS predictions based on explicit-solvent molecular dynamics (MD) simulations overcome such limitations at the price of an increased computational cost. To rationalize the need for explicit-solvent methods, we first review the approximations underlying implicit-solvent methods. Next, we describe the theory behind explicit-solvent SAS predictions that are easily accessible via the WAXSiS web server. We present the workflow for computing SAS pattern from a given molecular dynamics trajectory. The calculations are available via a modified version of the GROMACS simulations software, coined GROMACS-SWAXS, which implements the WAXSiS method. Practical considerations for running routine explicit-solvent SAS predictions are discussed.
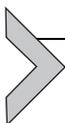
1. Introduction

Small- and wide-angle X-ray scattering (SWAXS) has originally been used to probe the overall shape of biomolecules or soft-matter complexes in solution. In recent decades, SWAXS has developed into an increasingly quantitative probe, primarily due to improved sample preparation, more brilliant light sources, and single-photon counting detectors (Graewert & Svergun, 2013; Koch, Vachette, & Svergun, 2003; Putnam, Hammel, Hura, & Tainer, 2007). Thanks to the advent of free-electron lasers, time-resolved SWAXS (TR-SWAXS) measurements are capable of tracking ultrafast dynamics of biomolecules or small molecules in solution (Arnlund et al., 2014; Brinkmann & Hub, 2016; Levantino et al., 2015). Complementary, small-angle neutron scattering (SANS) has remained popular because it allows for contrast-variation experiments by changing the D_2O concentration of the buffer (Gabel, 2015). Henceforth, we use the term SAS (small-angle scattering) when referring to both SWAXS and SANS.

Analysis of the SAS curve gives information on the overall shape and macroscopic properties of the probe such as the radius of gyration, maximum particle size, the structural order (globular vs unfolded), molecular mass, and particle volume. In contrast, an atomistic interpretation cannot be drawn from the data alone owing to the low information content of the SAS curves, but instead requires methods for theoretically predicting SAS intensities from a given atomistic structure. Such SAS curve predictions, so-called forward models, enable the validation or the refinement of structural models or ensembles against SAS data.

The physical properties that give rise to a solution scattering curve $I(q)$ are well understood. Namely, the intensities are given by

$$I(q) = \langle |F(\mathbf{q})|^2 \rangle_\Omega, \tag{1}$$

where $\langle \cdot \rangle_\Omega$ denotes the orientational average in reciprocal space, and $F(\mathbf{q})$ is the Fourier transform of the electron density contrast $\Delta\rho(\mathbf{r})$ between the solution and the pure-buffer system:

$$F(\mathbf{q}) = \int \Delta\rho(\mathbf{r}) \, e^{-i\mathbf{q} \cdot \mathbf{r}} \, \mathrm{d}^3 r \tag{2}$$

However, computing $I(q)$ from a given structural model such as a crystal structure or a molecular dynamics (MD) simulation of a biomolecule is complicated by several aspects:

(i) Since SAS detects the electron density contrast in solution, computing $I(q)$ requires knowledge of the volume of solvent that is displaced by the solute. For a biomolecule with internal disorder and a rough surface, the displaced volume is far from obvious. In addition, as discussed below, the volume taken by atoms of a certain chemical element depends on the chemical environment, suggesting that tabulated atomic volumes are subject to marked uncertainties.

(ii) The density of the hydration layer of biomolecules differs from the density of bulk solvent, thus contributing to the density contrast $\Delta\rho$. The hydration layer is influenced by properties such as the biomolecule's charge and geometry, the amino acid composition of the surface (anionic, cationic, hydrophobic, hydrophilic neutral), or the salt type and concentration of the buffer. For instance, highly charged biomolecules such as DNA/RNA were shown to exhibit a tight hydration layer (Pollack, 2011). For biomolecules with large surface-to-volume ratio, such as intrinsically disordered proteins, the hydration layer is expected to strongly contribute to the SAS signal. However, how such properties determine the hydration layer is not yet understood on a quantitative level (Kim et al., 2016).

(iii) SAS curves are influenced by thermal fluctuations of the biomolecule. Using analytic models, Moore (2014) showed that including Debye–Waller factors or correlated atomic fluctuations influences the calculated SAS curves at relatively small scattering angles of $q \gtrsim 0.2 \, \text{Å}^{-1}$. These findings are in line with earlier experimental observations (Tiede, Zhang, & Seifert, 2002) and with MD simulations, which revealed

alterations of SAXS curves at $q \gtrsim 0.25$ Å^{-1} upon modulating atomic fluctuations during the simulations (Chen & Hub, 2014).

In the last three decades, a wide range of methods have been developed for predicting SAS intensities from atomistic structures, reflecting the increasing importance of SAS experiments for biomolecular research and the need for a physically founded interpretation of the data (Bardhan, Park, & Makowski, 2009; Chen & Hub, 2014; Grishaev, Guo, Irving, & Bax, 2010; Grudinin, Garkavenko, & Kazennov, 2017; Knight & Hub, 2015; Köfinger & Hummer, 2013; Liu, Morris, Hexemer, Grandison, & Zwart, 2012; Merzel & Smith, 2002 a, 2002 b; Nguyen, Pabit, Meisburger, Pollack, & Case, 2014; Oroguchi, Hashimoto, Shimizu, Sato, & Ikeguchi, 2009; Oroguchi & Ikeguchi, 2012; Park, Bardhan, Roux, & Makowski, 2009; Poitevin, Orland, Doniach, Koehl, & Delarue, 2011; Putnam, Weiner, Woetzel, Lowe, & Meiler, 2015; Ravikumar, Huang, & Yang, 2013; Schneidman-Duhovny, Hammel, & Sali, 2010; Schneidman-Duhovny, Hammel, Tainer, & Sali, 2013; Stovgaard, Andreetta, Ferkinghoff-Borg, & Hamelryck, 2010; Svergun, Barberato, & Koch, 1995; Tjioe & Heller, 2007; Virtanen, Makowski, Sosnick, & Freed, 2011; Yang, Park, Makowski, & Roux, 2009). The methods mainly differ by

- the method for modeling the hydration layer, involving explicit all-atom models, increased atomic form factors for solvent-exposed solute atoms, layers of uniform electron density, implicit models with internal structure, or several others;
- modeling of the excluded solvent based on tabulated atomic volumes up to explicit all-atom models;
- mathematical methods for computing the orientational average involving the Debye equation, a spherical harmonics expansion, or numerical quadrature;
- the spatial resolution, that is, whether the SAS curves are computed from individual atoms or from coarse-grained beads.

In this chapter, we focus on the calculation of SAS curves from atomistic explicit-solvent MD simulations as implemented in GROMACS-SWAXS, an extension of the GROMACS simulation software (https://gitlab.com/cbjh/gromacs-swaxs) (Chen & Hub, 2014; Chen et al., 2019). The method is easily accessible via the WAXSiS web server at https://waxsis.uni-saarland.de (Knight & Hub, 2015). As a rationale behind the explicit-solvent methods, we first review the approximations and limitations underlying widely used implicit-solvent methods, with a focus on the role of the hydration layer and excluded solvent for obtaining the density contrast. Next, we briefly describe the theory of explicit-solvent SAS curve predictions from

explicit-solvent MD simulations as used by GROMACS-SWAXS and WAXSiS. Finally, the workflow of SAS calculations from an MD trajectory and practical considerations are discussed.
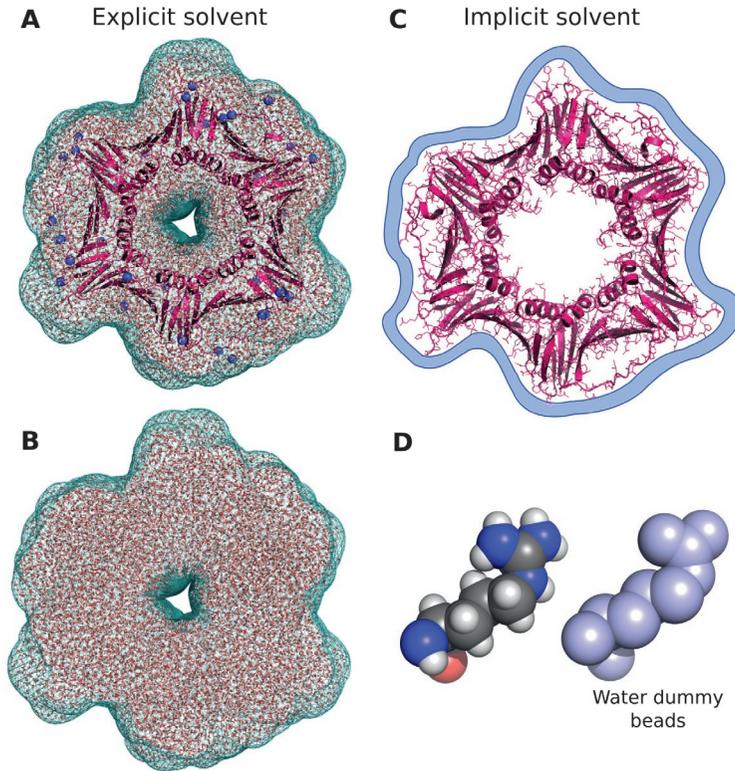
2. Implicit-solvent methods

Computationally efficient methods for computing SAS curves, as implemented in CRYSOL, FoXS, PepsiSAXS, or several other tools, use an implicit representation of the hydration layer and excluded solvent. To rationalize the benefits of explicit-solvent SAS calculations, we discuss the limitation of implicit-solvent methods in the following.

2.1 Displaced solvent

Many methods represent the displaced (or excluded) solvent with the help of water dummy beads, which are typically placed at the positions of the biomolecule's heavy atoms (see Fig. 1D) (Grudinin et al., 2017; Svergun et al., 1995). The electron density of the bead is essentially taken as $\rho_d(r) = \rho_s \exp\left(-\pi r^2/v^{2/3}\right)$, where ρ_s is the electron density of the solvent and v is the displaced volume such that the bead carries $\rho_s v$ electrons, as desired. The atomic form factor of the bead is given by the Fourier transform of the electron density, $f_d(q) = \rho_s v \exp\left(-v^{2/3}q^2/4\pi\right)$. In the implicit-solvent description, the buffer subtraction is carried out by reducing the atomic form factor of the biomolecule's heavy atom with the form factor of the dummy bead $f_d(q)$, leading to so-called "reduced form factors."

This method requires knowledge of the displaced volumes v for all atoms. However, estimates for the displaced volume differ substantially within the literature, leading to some uncertainty in the predicted SAS curve (Hub, 2018). Most SAXS prediction methods use and cite the displaced volumes reported by Fraser, MacRae, and Suzuki (1978). These data originally date back to a publication in 1895 by Isidor Traube, which was based on the volume change of water in response to the solvation of various organic compounds (Traube, 1895). Table 1 presents the atomic volumes reported by Fraser et al. in Å^3 as well as the original values by Traube in ccm/mol for several atoms or chemical groups. Traube and Fraser values are identical since ccm/mol $= 1.66\ \text{Å}^3$.

More recently, atomic volumes have been obtained by Pontius et al. using Voronoi tessellation applied to the core region of high-resolution protein structures (Pontius, Richelle, & Wodak, 1996). The volumes by Traube and Fraser greatly differ from the volumes obtained from crystal structures

Fig. 1 (A/C) Hydration layer (B/D) displaced solvent, as represented by (A/B) explicit-solvent and (C/D) several implicit-solvent methods. (A) A spatial envelope is defined (blue mesh) that includes the biomolecule and the hydration layer, here shown for the protein PCNA (pdb code 4D2G) (De Biasio et al., 2015). (B) The same spatial envelope is used to define the displaced solvent. (C) Implicit-solvent methods use a simplified description of the hydration layer, often modeled as a layer of uniform density, while (D) the displaced solvent is typically described by water dummy beads with estimated volumes taken from a publication by Traube (1895) and fitted against the data.

(Table 1, columns 2 and 3). Compared to the Pontius values, the Fraser values greatly overestimate the volumes of CH and CH_2 groups, and they underestimate the volumes of N and O atoms and of NH or OH groups. Notably, Traube reported volumes of N or O atoms for alternative chemical environments such as nitro or (today outdated) "pentavalent" environments as well as for hydroxyl groups; however, these were not cited by Fraser, which may explain why they are not used in SAS calculations today (Table 1, column 4).

Besides the uncertainty owing to the choice of tabulated atomic volumes, additional uncertainty may arise because the atomic volumes depend

Table 1 Displaced volumes from Voronoi tessellation of crystal structure cores by Pontius et al. (1996), as well as from densitometric data by Traube (1895), as cited by Fraser et al. (1978).

Atomic group	Pontius et al. (1996) [Å³]	Fraser et al. (1978) [Å³]	Traube (1895) [ccm/mol]
H		5.15	3.1
C		16.44	9.9
N	8.8 (0.8)	2.49	1.5 (trivalent)
			10.7 (pentavalent)
			8.5–10.7 (in nitro compound)
O	22.3 (0.4)	9.13	5.5 (carbonyl oxygen)
			2.3 or 0.4 (hydroxy oxygen)
OH	23.9 (0.9)	14.28	
NH	14.1 (0.3)	7.64	
CH	11.8 (0.6)	21.59	
CH$_2$	20.9 (1.8)	26.74	
CH$_3$	33.9 (1.2)	31.89	

Pontius values represent averages and standard deviations from all amino acids. Traube and Fraser values are identical, using ccm/mol = 1.66Å³. Fraser volumes of chemical groups represent sums of single-atom volumes.

on the chemical environment. Whereas the volumes by Pontius et al. represent the atomic packing in the core of stable proteins, it seems unlikely that the volumes also hold for more flexible protein surfaces, for detergent micelles, or for nucleotides. For n-hexadecane, for instance, the Pointius and Fraser volumes would imply a molecular volume of 360.4 and 438.1 Å^3, respectively, whereas the experimental value is 486.6 Å^3.

2.2 Hydration layer (HL)

Several tools model the hydration layer (HL) as a layer with a predefined thickness and a uniform excess density (Fig. 1C). For instance, CRYSOL constructs a homogeneous 3 Å-wide layer described by a two-dimensional angular function leading to a simplistic representation of the HL, which may not be suitable for molecules with cavities or nonglobular shapes (Svergun et al., 1995). CRYSOL3 aims to overcome this limitation by incrusting the surface (Franke et al., 2017)) similar to the grid-based density employed by PepsiSAXS (Grudinin et al., 2017). In contrast, FoXS models

the HL by increasing atomic form factors of solvent–exposed atoms (Schneidman–Duhovny et al., 2010).

The density of HLs of biomolecules likely depends on the type of the water–surface interactions and, therefore, on the chemical composition of the surface. For instance, Kim et al. (2016) found that acidic amino acids (Glu/Asp) exhibit a tighter HL as compared to basic amino acids (Arg/Lys), an effect that our group recently observed in MD simulations (unpublished data). Likewise, it seems plausible that the HL of maltoside or sodium dodecyl sulfate (SDS) head groups of detergent micelles exhibit different densities as compared to the HL of charged or hydrophobic amino acids. Such chemical specificity of the HL is not captured by implicit solvent methods, explaining why they require fitting of the HL against experimental data.

## 2.3 Risk of overfitting solvent-related parameters

Owing to the simplified description of the HL and excluded solvent, implicit-solvent methods require two or three parameters that are adjusted upon fitting the calculated curve to the experimental curve. One parameter adjusts the excess density of the HL, while another parameter adjusts the volumes of the dummy beads that represent the excluded solvent. Some methods adjust in addition the overall displaced volume. The parameters are typically fitted by minimizing the residuals between the calculated and experimental SAS curves. Whether the fitted parameters reflect the physical situation such as the true HL density, or whether they merely absorb errors in the structural model or in the experimental data, is typically not clear. Hence, fitting solvent-related parameter may hide structurally relevant information.

A possible consequence of fitting parameters is illustrated in Fig. 2, which presents the validation of MD simulations of a protein-detergent complex against SAXS data (Berthaud et al., 2012; Chen & Hub, 2015 b). The complexes were composed of an aquaporin-0 tetramer solvated in 250 to 330 βDDM detergent molecules. Computing SAXS curves from the simulations with implicit-solvent methods and fitting the curves to experimental data leads to χ-values that hardly depend on the number of βDDM molecules, suggesting that alterations in the structural model have been absorbed into the fitting parameters (Fig. 2D). Consequently, the experimental number of βDDM molecules cannot be obtained by comparison with the data. This example illustrates that structurally relevant information may be hidden upon fitting free parameters. In contrast, when using explicit-solvent SAXS calculations described below, the χ-agreement of models with different βDDM numbers with the data can be distinguished because no
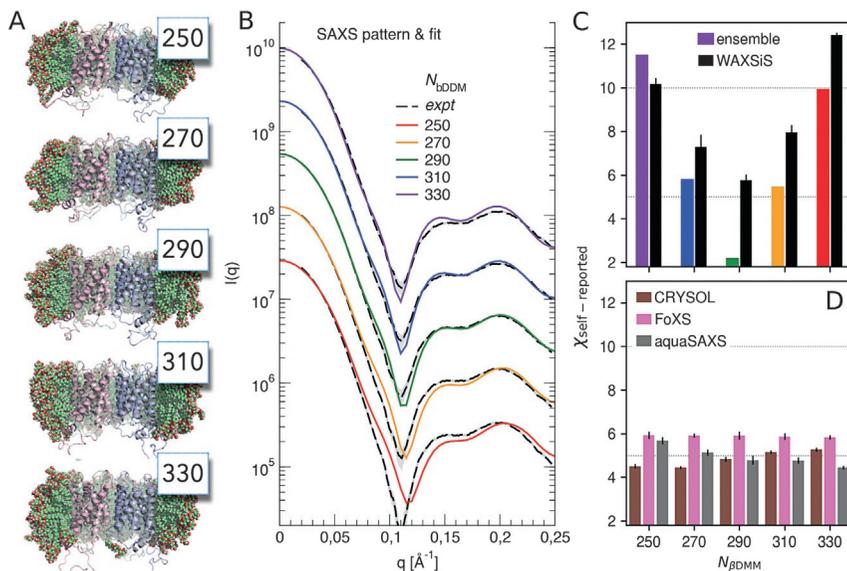
Fig. 2 SAXS curve prediction and snapshots of a protein-detergent complex. (A) MD simulation systems of aquaporin-0 solubilized in 250 to 330 βDDM detergent molecules simulated with the Charmm36 force field. (B) Explicit-solvent SAXS curve predictions for different number of βDDM using an aggregated ensemble of 90 ns to 100 ns MD trajectories (*colored lines*). Experimental data (*black dashed line*) from Berthaud, Manzi, Pérez, and Mangenot (2012). (C) χ-agreement between experimental SAXS curves and explicit-solvent SAXS calculations from free MD simulations (*colored bars*) or from the WAXSiS webserver (*black bars*). WAXSiS restrains the DDM molecules in a short simulation, rationalizing the increased χ-values. $N_{\beta DDM} = 290$ reveals the best agreement with experiment. (D) χ-agreement by CRYSOL, FoXS, and AquaSAXS, revealing nearly constant χ due to the adjustment of free parameters. *Reprinted with permission from Chen, P.-c., & Hub, J. S. (2015b). Structural properties of protein-detergent complexes from SAXS and MD simulations. The Journal of Physical Chemistry Letters, 6, 5116–5121, Copyright 2020 American Chemical Society.*

solvent–related parameters are fitted, thereby revealing that the model with 290 βDDM exhibits the best agreement with the data (Fig. 2C).
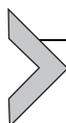
## 3. Explicit-solvent SAS predictions with the WAXSiS method

Several methods for predicting SAS curves have been presented that model both the excluded solvent and the HL with explicit–solvent MD simulations (Chen & Hub, 2014; Köfinger & Hummer, 2013; Oroguchi et al., 2009; Park et al., 2009). Here, we focus on the WAXSiS method, which is based on the formalism by Park et al. (2009). However, whereas the

implementation by Park et al. required a constrained biomolecule, the WAXSiS method allows SAS predictions from MD simulations of flexible biomolecules by defining the SAS-contributing solvent with a spatial envelope. In addition, the WAXSiS method corrects (i) imprecise densities of water models and (ii) small density mismatches between the biomolecule and the pure-solvent simulation systems, both of which may influence the SAS prediction significantly. A careful comparison of different implicit and explicit SAXS prediction methods has been presented by Bernetti and Bussi (2021).

In the WAXSiS method, a spatial envelope is defined that encloses the biomolecule and the heterogeneous density of the HL (Figs. 1A and 3A). Hence, the HL structure is fully defined by the force field and does not require a free fitting parameter. The buffer subtraction is carried out by computing the scattering intensity difference between (a) the biomolecule with the envelope-enclosed solvent and (b) an explicit-solvent pure-buffer simulation, from which an identical volume is taken with the help of the same envelope (Fig. 1B). Thereby, the displaced solvent is modeled with atomic detail and without need of knowing atomic volumes of solute atoms.

Compared to implicit-solvent methods, the explicit-solvent calculations are computationally far more expensive as they require MD simulations, but they have several crucial advantages:
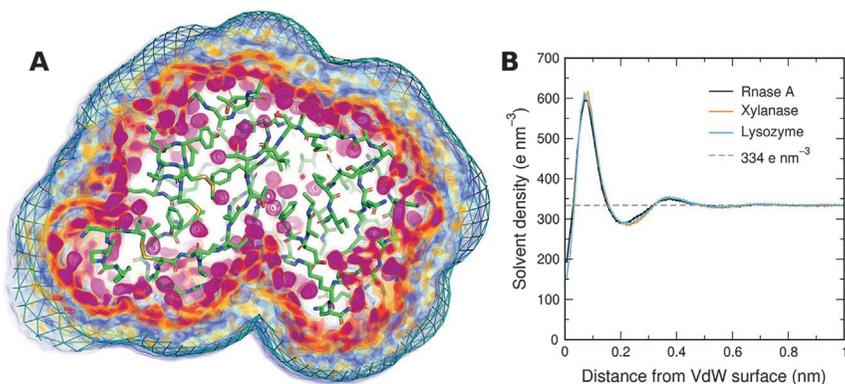
Fig. 3 (A) Solvent electron density around Rnase A, taken from an explicit-solvent MD simulations with position restraints on all heavy atoms, revealing the marked structure of the hydration layer. Only solvent density inside an envelope (*blue mesh*) is shown, where the envelope was constructed at a 7 Å distance from the protein surface. (B) Solvent density as function of the distance from the Van der Waals surface for three different proteins (see legend), averaged over the protein surface. Simulations were carried out with the Amber99sb-ildn/Tip3p force fields. Here, the solvent density was scaled with a small constant factor to match the experimental bulk density of $334\ e\ nm^{-3}$, as implemented in GROMACS-SWAXS to correct imprecise densities of some water models.

(i) The calculations do not require any free solvent-related parameters that must be fitted to experimental data. Thus, the amount of structural information, which can be extracted from the low-information SAS data, is not further reduced. By circumventing the fitting of the HL density, the radius of gyration is not adjusted against the data.

(ii) The use of "reduced form factors" is avoided and, with this, the uncertainties subject to atomic volumes. This may explain the better accuracy compared to implicit-solvent methods for predicting the SAS intensities of heterogeneous systems (such as protein–detergent complexes) and for molecules with nonglobular geometries (Ivanović, Bruetzel, Lipfert, & Hub, 2018; Ivanović, Hermann, Wójcik, Pérez, & Hub, 2020).

(iii) The explicit-solvent description remains valid at wide scattering angles ($q >1.5$ Å), where the internal structure of water becomes relevant (Park et al., 2009).

(iv) By computing the SAS patterns from an MD trajectory, the method naturally accounts for thermal fluctuations, which influence the SAS patterns at remarkably small scattering angles of $q \gtrsim 0.25$ Å^{-1} (Chen & Hub, 2014, 2015 b; Moore, 2014; Tiede et al., 2002).

As an example, Fig. 4A compares the experimental SAXS curve of the ring–shaped protein PCNA with a curve computed with the WAXSiS
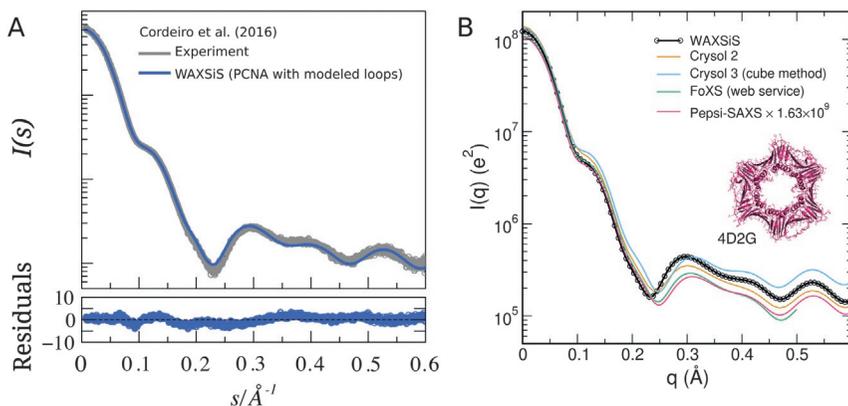
Fig. 4 (A) Experimental SAXS data of the ring-shaped protein PCNA (*gray*) and SAXS curve computed with the WAXSiS model from a PCNA model with modeled loops ($\chi^2 = 1.35$) (Cordeiro et al., 2017). (B) Comparison of SAXS curves of the PCNA crystal structure (pdb code 4D2G) computed with different methods with default settings (see legend). *Panel A: Adapted and reused with permission from Cordeiro, T. N., Chen, P.-c., De Biasio, A., Sibille, N., Blanco, F. J., Hub, J. S., …Bernadó, P. (2017). Disentangling polydispersity in the PCNA-p15PAF complex, a disordered, transient and multivalent macromolecular assembly. Nucleic Acids Research, 45, 1501–2015.*

method. The WAXSiS curve was computed from the 4D2G crystal struc-
ture of PCNA, to which missing termini were added (Cordeiro et al., 2017),
and the computed curve was henceforth scaled with a constant factor
$(I_{scaled}(q) = f I_{WAXSiS}(q))$. No other parameters were adjusted. Fig. 4B com-
pares SAXS curves computed from the 4D2G structure (without termini)
computed with WAXSiS and with several implicit-solvent methods.
All methods used default settings, and the calculated curves were not fitted
against the experiment. Evidently, the predicted SAXS curves differ signif-
icantly, presumably owing to different modeling of the hydration layer and
excluded solvent.
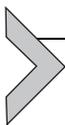
4. Theory

The experimental SAS intensity is given by the difference between the
scattering intensity of the solution $I_{sample}(q)$ and the solvent $I_{solvent}(q)$:

$$I(q) = I_{sample}(q) - I_{solvent}(q) \tag{3}$$

To calculate this excess intensity $I(q)$ from MD trajectories, the WAXSiS
method follows the formalism by Park et al. (2009). The low-dilution limit
is considered, such that correlations between different solute molecules can
be neglected and the scattering experiment can be modeled by a single solute
molecule in a solvent bath. A spatial envelope is constructed around the sol-
ute including the hydration layer (Fig. 1A). The instantaneous electron den-
sity of the solute system $A(\mathbf{r})$ and of the solvent system $B(\mathbf{r})$ is divided into the
electron density inside and outside of the envelope, indicated by subscripts i
and o, respectively,

$$A(\mathbf{r}) = A_i(\mathbf{r}) + A_o(\mathbf{r}), \tag{4}$$
$$B(\mathbf{r}) = B_i(\mathbf{r}) + B_o(\mathbf{r}). \tag{5}$$

Assuming that the envelope is large enough such that density correlations
between the inside and the outside of the envelope are due to bulk water,
the excess scattering intensity only requires knowledge of the Fourier trans-
forms of the electron densities *inside* the envelope $\tilde{A}_i(\mathbf{q})$ and $\tilde{B}_i(\mathbf{q})$:

$$I(q) = \langle D(\mathbf{q}) \rangle_{\Omega}, \tag{6}$$

where

$$D(\mathbf{q}) = \left\langle |\tilde{A}_i(\mathbf{q})|^2 \right\rangle^{(\omega)} - \left\langle |\tilde{B}_i(\mathbf{q})|^2 \right\rangle^{(\omega)} + 2\,\mathrm{Re}\left[-\left\langle \tilde{B}_i^*(\mathbf{q}) \right\rangle^{(\omega)} \left\langle \tilde{A}_i(\mathbf{q}) - \tilde{B}_i(\mathbf{q}) \right\rangle^{(\omega)} \right].$$

(7)

Here, $\langle \ldots \rangle_\Omega$ denotes the orientational average in $\mathbf{q}$-space, $\langle \ldots \rangle^{(\omega)}$ the average over solute and solvent fluctuations at fixed solute orientation ω, and Re is the real part. In practice, $\langle \ldots \rangle^{(\omega)}$ represents the average over MD frames after superimposing the solute onto a reference structure of orientation ω. $\tilde{A}_i(\mathbf{q})$ and $\tilde{B}_i(\mathbf{q})$ are calculated using the atomic form factors $f_j(q)$ and coordinates $\mathbf{r}_j$ of atom j,

$$\tilde{A}_i(\mathbf{q}) = \sum_{j=1}^{N_A} f_j(q) e^{-i\mathbf{q}\cdot\mathbf{r}_j},$$

(8)

where N_A is the number of atoms within the envelope in the respective MD frame (Fig. 1A). $\tilde{B}_i(\mathbf{q})$ is calculated analogously over the solvent atoms within the envelope (Fig. 1B). The form factors are approximated using the Cromer–Mann parameters a_k, b_k, and c of atom j (Cromer & Mann, 1968),

$$f_j(q) = \sum_{k=1}^{4} a_k e^{-b_k(q/4\pi)^2} + c.$$

(9)

The formalism is identical for SANS calculations, except that the atomic form factors $f_j(q)$ are replaced by the coherent neutron scattering lengths b_j of the scattering atoms (Chen et al., 2019; Dias–Mirandela et al., 2018).

4.1 Comparing to experimental data

To compare a predicted curve with experimental data, the experimental curve $I_{\exp}$ can be fitted to the calculated curve I_c by minimizing

$$\chi^2 = N_q^{-1} \sum_{i=1}^{N_q} \frac{\left[I_c(q_i) - \left(f I_{\exp}(q_i) + c \right) \right]^2}{\sigma_i^2},$$

(10)

where N_q denotes the number of q-points, and σ_i are the experimental errors. Besides the overall scale f, an offset c may be fitted to account for small experimental uncertainties from the buffer subtraction. Here, the experimental curves are fitted instead of the calculated curves, because the calculated curves are predicted without any free parameters. In

practice, it is advisable to fit also without the offset c and, thereby, to test whether c absorbs only a small constant offset, as desired, or whether fitting the offset would hide a structurally relevant discrepancy between I_{exp} and I_c.
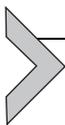
5. Workflow

We present a brief workflow for computing a SAS curve from an explicit–solvent MD simulation (Fig. 5). Because the SAS calculations are implemented into an extension of GROMACS, it is convenient (but not required) to run the MD simulation with GROMACS as well. This
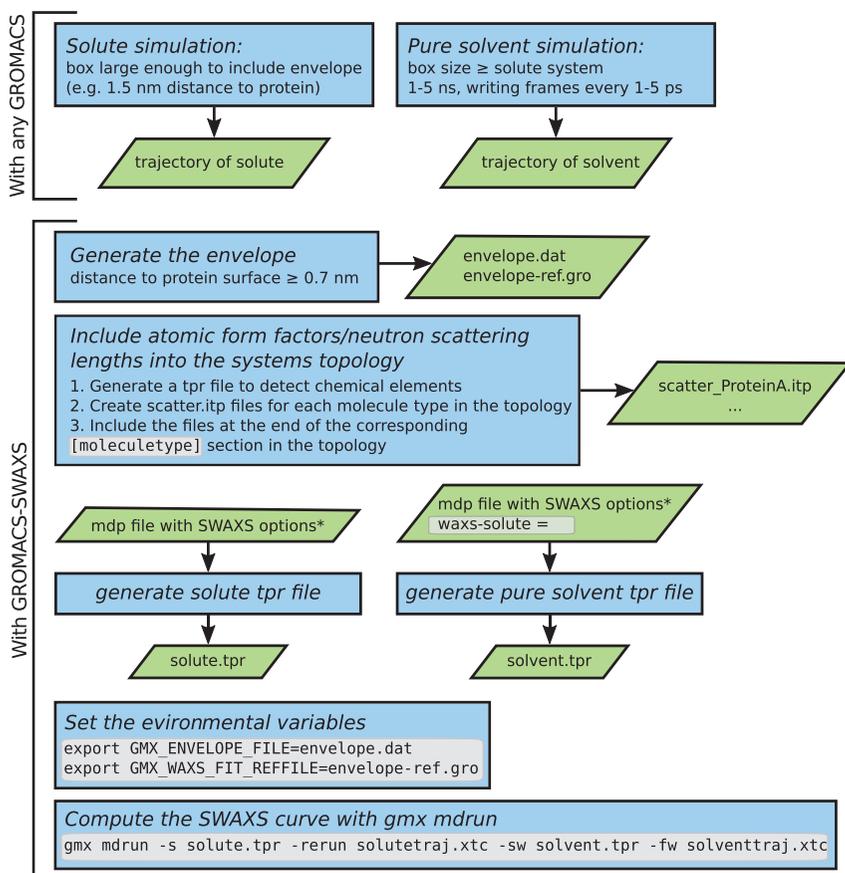

Fig. 5 Workflow for SAS curve predictions with GROMACS-SWAXS. *Suggested mdp parameters are shown in Listing 1.

workflow is automated also on the web server WAXSiS (https://waxsis.uni-saarland.de). However, the MD simulations on WAXSiS are carried out with the Yasara software which allows fully automated MD simulations even in presence of ligands or modified amino acids (Krieger & Vriend, 2015). Several additional tutorials and more detailed documentation for GROMACS-SWAXS are available at https://cbjh.gitlab.io/gromacs-swaxs-docs.

5.1 A: With any GROMACS

1. Run an MD simulation of your biomolecule in solution with either GROMACS-SWAXS or any other GROMACS version. Choose the simulation box large enough, such that the envelope (see below) will fit into the box. To generate the box use, for example,

```
gmx editconf -d 1.5 -f protein.gro -bt dodecahedron -o box.gro
```

where for the -d option a distance between 1.5 and 2.0 is usually sufficient.
2. Run an MD simulation of pure solvent for the buffer subtraction. Use the same water model and MD parameter (mdp) file as used for the solute simulations to ensure that solute and solvent simulations exhibit identical bulk solvent densities, as required for an accurate buffer subtraction. The solvent simulation box should be at least as large as your solute simulation box. Produce 1 to 5 ns of simulation, writing the simulation frames every 1 to 5 ps, that is, $\sim$1000 frames.

5.2 B: With GROMACS-SWAXS

When the simulations of solute and pure solvent have finished, we continue with preparing and running the SAS curve calculation. Once the steps below are understood, they can be scripted to automate the SAS curve calculation.
1. Download and compile GROMACS-SWAXS (https://gitlab.com/cbjh/gromacs-swaxs), and load it into your path:

```
source /path/to/gromacs-swaxs/bin/GMXRC
```

5.2.1 Specify the atomic form factors/neutron scattering lengths
2. Generate a run–input (tpr) file using `gmx grompp` of GROMACS-SWAXS to detect the chemical elements, which are used below for

assigning the Cromer–Mann parameters or neutron scattering lengths. Chemical elements are recognized using atom names and masses to ensure that a C_α carbon atom (C) is not confused with calcium (Ca), Fluor (F) not with Iron (Fe), etc. Atomic masses are reliably provided by a tpr file. An empty mdp file may be used for this step:

```
rm -f anymdp.mdp; touch anymdp.mdp
gmx grompp -f anymdp.mdp -p topol.top -o tmp.tpr
```

3. Now, the previously created tpr file is used to generate a `scatter.itp` file for each molecule, which contains the Cromer–Mann parameters or the neutron scattering lengths. If the solute contains nondefault groups, such as chromophore, a heme group, or similar, first prepare an index file with the *entire* solute. Then run

```
gmx genscatt -s tmp.tpr [-n index.ndx] [-vsites]
```

Select your *complete* solute, such as Protein, Protein_HEME or Protein_ Chromophore. If you use virtual sites, use `-vistes` and select Prot–Masses (or a group with heme, the chromophore, etc.). The `-visites` option instructs `gmx genscatt` that atomic masses deviate from the physical masses, as common when using virtual sites. `gmx genscatt` writes one itp file for each molecule type. These must be added to each moleculetype definition in the topology, similar to the `#include "posres.itp"` line for defining position restraints.

5.2.2 Generating the envelope

4. Generate the envelope from the trajectory of the solute simulations

```
gmx genscatt -s tmp.tpr -f solutetraj.xtc -d 0.7
```

The output `envelope.dat` lists inner and outer radii of the envelope surface, whereas `envelope-ref.gro` is a reference structure with orientation ω used to superimpose the solute frames onto the envelope (see Theory). The envelope may be visualized with PyMol with the file `envelope.py`.

5.2.3 Generating the solute and solvent run-input files

5. Generate tpr files with `gmx grompp` and with an mdp file that contains all SAS-specific parameters, as presented in .

```
gmx grompp -f rerun.mdp -c solutesystem.gro
\[-n index.ndx] -o solute.tpr
```

In addition, a tpr file of the pure-solvent system is required, prepared with a mdp file with an empty solute entry (`waxs-solute =`)

LISTING 1 A typical set of SAS-specific mdp parameters for calculating a SWAXS/SANS curve with GROMACS-SWAXS. The calculation of SANS curves is optional, as indicated by square brackets.

```
; read scattering info from topology
define = -DSCATTER
; turn on SAXS calc. and, optionally, multiple SANS calc.
scatt-coupl = xray [neutron neutron ...]
; solute group
waxs-solute = Protein ; or Protein-Masses
; solvent group
waxs-solvent = Water_and_ions
; rotational fit group as used with gmx genenv
waxs-rotfit = C-alpha
; or use a PBC atom near the geometric center,
; as suggested by gmx genenv
waxs-pbcatom = -1
; for uniform average over all simulation frames
waxs-tau = 0
; number of q-points
waxs-nq = 101
; qmin and qmax in nm^(-1)
waxs-startq = 0
waxs-endq = 10
; number of q-vectors for spherical average,
; use ~0.2*(D*qmax)^2
waxs-nsphere = 1500
; experimental solvent density in e/nm3, for solvent density
; correction
waxs-solvdens = 334
; Use I(sample)-I(buffer) as buffer subtraction scheme
waxs-correct-buffer = no
; D2O concentrations of SANS calculations
waxs-deuter-conc = [0.42 1]
```

```
gmx grompp -f solvent.mdp -c solventsystem.gro -o solvent.tpr
```

5.2.4 Compute the SAS curve

6. Finally, compute the SAS curves from the solute simulations with the rerun functionality of `gmx mdrun`. The envelope files are specified with environment variables, as follows:

```
export GMX_ENVELOPE_FILE=envelope.dat
export GMX_WAXS_FIT_REFFILE=envelope-ref.gro
gmx mdrun -s solute.tpr -rerun solutetraj.xtc
          \-sw solvent.tpr -fw solventtraj.xtc
```

The SAS calculations strongly benefit from the use of a GPU or from using multiple CPU cores, for instance using `gmx mdrun -ntomp 16 -gpu_id 0` .... The SAS curves are written to the files `waxs_final.xvg`.

6. Practical considerations

6.1 Convergence

The number of simulation frames required to receive a converged SAS curve depends on the contrast between solute and solvent, where a lower contrast leads to slower convergence. The contrast is lower if the hydration layer is large compared to the biomolecule, that is, if the envelope contains mostly water. Specifically, small biomolecules or intrinsically disordered proteins (IDPs) require a larger number of simulation frames as compared to larger or globular biomolecules. As a reference, for a smaller protein such as the GB3 domain, approximately 300 frames are required to obtain a well-converged SAS curve. A larger protein such as glucose isomerase requires approximately 70 frames. The statistical errors listed in `waxs_final.xvg` are computed with error propagation, assuming that each simulation frame provides an independent solvent structure, which is fulfilled when using frames with time spacing $\gtrsim$ 1 ps.

For IDPs, the convergence should be carefully assessed, for instance by dividing the trajectory of *both* the biomolecule and the buffer into ten independent blocks, and then comparing the ten SAS curves computed from the blocks. The standard error over the blocks provides a rigorous measure for statistical uncertainty.

6.2 Orientational average

The parameter J (mdp option waxs-nsphere) determines the number of **q**-vectors used to take the orientational average. It should be taken as $J \approx \alpha(Dq_{max})^2$. Where D is the maximum diameter of the solute, q_{max} is the maximum momentum transfer (mdp option waxs-endq), and the constant α determines the accuracy of the orientational average. For most biomolecules, $\alpha = 0.2$ is sufficient, while for excessively elongated rod–like structures, a value of $\alpha = 0.5$ may be required.

6.3 Envelope size

The distance of the envelope from the biomolecular surface (specified with gmx genenv –d) should be at least 0.6 to 0.7 nm, thereby containing the density modulations along the hydration layers. For highly charged biomolecules, a larger envelope may be used to account for the entire counter ion cloud. We found previously that a distance of ~3 times the Debye length is required to include most effects from the counter ion cloud on the radius of gyration of glucose isomerase (Ivanović, Bruetzel, Shevchuk, Lipfert, & Hub, 2018). At a salt concentration of 100 mM, this may require an envelope distance of ~3 nm and, hence, large simulation systems.

6.4 Solvent density correction

A solvent density correction is necessary for two reasons: (i) Due to finite-size effects, the bulk solvent densities and correlations may slightly differ between the solute simulation and the pure-solvent simulation (Köfinger & Hummer, 2013). (ii) The density of certain water models, such as the popular Tip3p model, deviates from the experimental water densities. The parameter for solvent density correction should be chosen to match the experimental electron density of the solvent (mdp option `waxs-solvdens`).

6.5 Atomic fluctuations

Structural fluctuations influence the SAS profile. Depending on the scientific question, position restraints may or may not be applied to the biomolecule, for instance with the aim to obtain the SAS curve for structure with given backbone coordinates or for a flexible biomolecular ensemble, respectively. To resolve the structure of the HL, it may be useful to restrain all heavy atoms of the biomolecule to avoid that side chain fluctuations smear out the HL structure, as used to obtain the solvent densities shown in Fig. 3.

6.6 Water models

We validated previously that most popular water force fields lead to nearly identical SAS curves at moderate scattering angles, suggesting that most water models exhibit similar packing at the biomolecular surface (Chen & Hub, 2014). Only at very wide angle of $q \gtrsim 2 \text{ Å}^{-1}$, where the water scattering is dominant, large effects from the water models are visible. However, we recently observed that water models with increased dispersion interactions such as the Tip4p2005s or Tip4p-D may impose tighter packing on the protein surface, leading to slightly increased radii of gyration (unpublished data) (Best, Zheng, & Mittal, 2014; Piana, Donchev, Robustelli, & Shaw, 2015). These observations open a novel route for validating protein–water interactions by modern force fields against experimental SAS data.

6.7 Computational costs

Run times of SAS predictions depend on the number of atoms, on the applied convergence criteria, and on the hardware. Typical costs may be illustrated by the execution time for different proteins on the WAXSiS webserver at https://waxsis.uni-saarland.de, which is currently (by August 2022) equipped with 16-core AMD Ryzen 9 processor and an Nvidia RTX 3070Ti graphics card. Using the convergence setting "normal," which is suitable for tests, SAS predictions of lysozyme (PDB id 1LYS), glucose isomerase (PDB id 1MNZ), and RNA polymerase II (PDB id 1TWF) required 2.7 min, 4.8 min, and 15.2 min, respectively. Using the convergence setting "thorough," which is recommended for publications, the execution times for the same biomolecules increased to 10.8 min, 14.2 min, and 36.3 min, respectively. These numbers may change as the hardware improves or as the convergence criteria are adapted in future releases of the WAXSiS webserver.

7. Summary

Both implicit- and explicit-solvent SAS prediction methods are highly valuable for the structural interpretation of experimental SAS methods. Implicit-solvent methods are computationally efficient, thus allowing high-throughput calculations on a laptop. They are unequivocally capable of validating the overall shape of structural models against SAS data, and to reveal effects from larger conformational transitions. To harvest structural details, as encoded in high-precision SAS data collected from modern

SEC-SAXS or SEC-SANS experiments, explicit-solvent methods are needed. They are (i) based on an accurate representation of the hydration layer and the excluded solvent, thereby avoiding solvent–related fitting parameters, (ii) are accurate for inhomogeneous systems since they do not require tabulated atomic volumes, and (iii) include effects from structural fluctuations on SAS curves. The WAXSiS method presented here may be further used for structure refinement simulations by restraining an MD simulation to experimental SAS data (Chen & Hub, 2015 a; Chen et al., 2019). More details on structure refinement are provided in a chapter of Part B of this monograph.

Running a SAS prediction on the WAXSiS webserver requires between 3 and 40 min including the MD simulation. On a fast workstation or a computer cluster, tens to hundreds of explicit-solvent calculations are feasible within days, suggesting that the cost for explicit-solvent SAS calculations is negligible as compared to the costs for experimental sample preparation or for data collection at the beamline. Therefore, we expect that the methods presented here will become a routinely used tool for SAS-based structural biology.

Acknowledgments

References

Arnlund, D., Johansson, L. C., Wickstrand, C., Barty, A., Williams, G. J., Malmerberg, E., … Neutze, R. (2014). Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. *Nature Methods*, *11*(9), 923–926.

Bardhan, J., Park, S., & Makowski, L. (2009). SoftWAXS: A computational tool for modeling wide-angle X-ray solution scattering from biomolecules. *Journal of Applied Crystallography*, *42*(5), 932–943.

Bernetti, M., & Bussi, G. (2021). Comparing state-of-the-art approaches to back-calculate SAXS spectra from atomistic molecular dynamics simulations. *European Physical Journal B*, *94*(9), 180. https://doi.org/10.1140/epjb/s10051-021-00186-9.

Berthaud, A., Manzi, J., Pérez, J., & Mangenot, S. (2012). Modeling detergent organization around Aquaporin-0 using small-angle X-ray scattering. *Journal of the American Chemical Society*, *134*(24), 10080–10088.

Best, R. B., Zheng, W., & Mittal, J. (2014). Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *Journal of Chemical Theory and Computation*, *10*(11), 5113–5124. https://doi.org/10.1021/ct500569b.

Brinkmann, L. U. L., & Hub, J. S. (2016). Ultrafast anisotropic protein quake propagation after CO-photodissociation in myoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 10565–10570.

Chen, P.-c., & Hub, J. S. (2014). Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data. *Biophysical Journal*, *107*, 435–447.

Chen, P.-c., & Hub, J. S. (2015a). Interpretation of solution X-ray scattering by explicit-solvent molecular dynamics. *Biophysical Journal*, *108*, 2573–2584.

Chen, P.-c., & Hub, J. S. (2015b). Structural properties of protein-detergent complexes from SAXS and MD simulations. *The Journal of Physical Chemistry Letters*, *6*, 5116–5121.

Chen, P.-c., Shevchuk, R., Strnad, F. M., Lorenz, C., Karge, L., Gilles, R., … Hub, J. S. (2019). Combined small-angle X-ray and neutron scattering restraints in molecular dynamics simulations. *Journal of Chemical Theory and Computation*, *15*(8), 4687–4698. https://doi.org/10.1021/acs.jctc.9b00292.

Cordeiro, T. N., Chen, P.-c., De Biasio, A., Sibille, N., Blanco, F. J., Hub, J. S., … Bernadó, P. (2017). Disentangling polydispersity in the PCNA-p15PAF complex, a disordered, transient and multivalent macromolecular assembly. *Nucleic Acids Research*, *45*, 1501–2015.

Cromer, D. T., & Mann, J. B. (1968). X-ray scattering factors computed from numerical Hartree-Fock wave functions. *Acta Crystallographica. Section A, Foundations of Crystallography*, *24*(2), 321–324.

De Biasio, A., de Opakua, A. I., Mortuza, G. B., Molina, R., Cordeiro, T. N., Castillo, F., … Blanco, F. J. (2015). Structure of p15PAF–PCNA complex and implications for clamp sliding during DNA replication and repair. *Nature Communications*, *6*(1), 6439. https://doi.org/10.1038/ncomms7439.

Dias-Mirandela, G., Tamburrino, G., Ivanović, M. T., Strnad, F. M., Byron, O., Rasmussen, T., … Javelle, A. (2018). Merging in-solution X-ray and neutron scattering data allows fine structural analysis of membrane-protein detergent complexes. *The Journal of Physical Chemistry Letters*, *9*, 3910–3914.

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., … Svergun, D. I. (2017). *ATSAS 2.8*: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*, *50*(4), 1212–1225. https://doi.org/10.1107/S1600576717007786.

Fraser, R. D. B., MacRae, T. P., & Suzuki, E. (1978). An improved method for calculating the contribution of solvent to the X-ray diffraction pattern of biological molecules. *Journal of Applied Crystallography*, *11*(6), 693–694.

Gabel, F. (2015). Small-angle neutron scattering for structural biology of protein–RNA complexes. In *Methods in enzymology: Vol. 558* (pp. 391–415). Elsevier.

Graewert, M. A., & Svergun, D. I. (2013). Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). *Current Opinion in Structural Biology*, *23*(5), 748–754.

Grishaev, A., Guo, L., Irving, T., & Bax, A. (2010). Improved fitting of solution X-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. *Journal of the American Chemical Society*, *132*, 15484–15486.

Grudinin, S., Garkavenko, M., & Kazennov, A. (2017). Pepsi-SAXS: An adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. *Acta Crystallographica. Section D, Structural Biology*, *73*(5), 449–464.

Hub, J. S. (2018). Interpreting solution X-ray scattering data using molecular simulations. *Current Opinion in Structural Biology*, *49*, 18–26.

Ivanović, M. T., Bruetzel, L. K., Lipfert, J., & Hub, J. S. (2018). Temperature-dependent atomic models of detergent micelles refined against small-angle X-ray scattering data. *Angewandte Chemie International Edition*, *57*, 5635–5639.

Ivanović, M. T., Bruetzel, L. K., Shevchuk, R., Lipfert, J., & Hub, J. S. (2018). Quantifying the influence of the ion cloud on SAXS profiles of charged proteins. *Physical Chemistry Chemical Physics*, *20*(41), 26351–26361. https://doi.org/10.1039/C8CP03080D.

Ivanović, M. T., Hermann, M. R., Wójcik, M., Pérez, J., & Hub, J. S. (2020). Small-angle X-ray scattering curves of detergent micelles: effects of asymmetry, shape fluctuations, disorder, and atomic details. *The Journal of Physical Chemistry Letters*, *11*(3), 945–951. https://doi.org/10.1021/acs.jpclett.9b03154.

Kim, H. S., Martel, A., Girard, E., Moulin, M., Härtlein, M., Madern, D., … Gabel, F. (2016). SAXS/SANS on supercharged proteins reveals residue-specific modifications of the hydration shell. *Biophysical Journal*, *110*(10), 2185–2194.

Knight, C. J., & Hub, J. S. (2015). WAXSiS: A web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic Acids Research*, *43*, W225–W230.

Koch, M. H. J., Vachette, P., & Svergun, D. I. (2003). Small-angle scattering: A view on the properties, structures and structural changes of biological macromolecules in solution. *Quarterly Reviews of Biophysics*, *36*(02), 147–227.

Köfinger, J., & Hummer, G. (2013). Atomic-resolution structural information from scattering experiments on macromolecules in solution. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *87*, 052712.

Krieger, E., & Vriend, G. (2015). New ways to boost molecular dynamics simulations. *Journal of Computational Chemistry*, *36*(13), 996–1007. https://doi.org/10.1002/jcc.23899.

Levantino, M., Schiro, G., Lemke, H. T., Cottone, G., Glownia, J. M., Zhu, D., … Cammarata, M. (2015). Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. *Nature Communications*, *6*, 6772.

Liu, H., Morris, R. J., Hexemer, A., Grandison, S., & Zwart, P. H. (2012). Computation of small-angle scattering profiles with three-dimensional Zernike polynomials. *Acta Crystallographica. Section A, Foundations of Crystallography*, *68*(2), 278–285.

Merzel, F., & Smith, J. C. (2002a). Is the first hydration shell of lysozyme of higher density than bulk water? *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 5378–5383.

Merzel, F., & Smith, J. C. (2002b). SASSIM: A method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins. *Acta Crystallographica. Section D, Biological Crystallography*, *58*, 242–249.

Moore, P. B. (2014). The effects of thermal disorder on the solution-scattering profiles of macromolecules. *Biophysical Journal*, *106*, 1489–1496.

Nguyen, H. T., Pabit, S. A., Meisburger, S. P., Pollack, L., & Case, D. A. (2014). Accurate small and wide angle X-ray scattering profiles from atomic models of proteins and nucleic acids. *The Journal of Chemical Physics*, *141*(22), 22D508.

Oroguchi, T., Hashimoto, H., Shimizu, T., Sato, M., & Ikeguchi, M. (2009). Intrinsic dynamics of restriction endonuclease EcoO109I studied by molecular dynamics simulations and X-ray scattering data analysis. *Biophysical Journal*, *96*, 2808–2822.

Oroguchi, T., & Ikeguchi, M. (2012). MD–SAXS method with nonspherical boundaries. *Chemical Physics Letters*, *541*, 117–121.

Park, S., Bardhan, J. P., Roux, B., & Makowski, L. (2009). Simulated X-ray scattering of protein solutions using explicit-solvent models. *The Journal of Chemical Physics*, *130*, 134114. https://doi.org/10.1063/1.3099611.

Piana, S., Donchev, A. G., Robustelli, P., & Shaw, D. E. (2015). Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *The Journal of Physical Chemistry B*, *119*(16), 5113–5123. https://doi.org/10.1021/jp508971m.

Poitevin, F., Orland, H., Doniach, S., Koehl, P., & Delarue, M. (2011). AquaSAXS: A web server for computation and fitting of SAXS profiles with non-uniformly hydrated atomic models. *Nucleic Acids Research*, *39*, W184–W189.

Pollack, L. (2011). Saxs studies of ion–nucleic acid interactions. *Annual Review of Biophysics*, *40*, 225–242.

Pontius, J., Richelle, J., & Wodak, S. J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. *Journal of Molecular Biology*, *264*(1), 121–136.

Putnam, C. D., Hammel, M., Hura, G. L., & Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*, *40*(3), 191–285. https://doi.org/10.1017/S0033583507004635.

Putnam, D. K., Weiner, B. E., Woetzel, N., Lowe, E. W., & Meiler, J. (2015). BCL::SAXS: GPU accelerated Debye method for computation of small angle X-ray scattering profiles. *Proteins: Structure, Function, and Bioinformatics*, *83*(8), 1500–1512.

Ravikumar, K. M., Huang, W., & Yang, S. (2013). Fast-SAXS-pro: A unified approach to computing SAXS profiles of DNA, RNA, protein, and their complexes. *The Journal of Chemical Physics*, *138*(2), 01B609.

Schneidman-Duhovny, D., Hammel, M., & Sali, A. (2010). FoXS: A web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Research*, *38*, W540–W544.

Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., & Sali, A. (2013). Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophysical Journal*, *105*(4), 962–974.

Stovgaard, K., Andreetta, C., Ferkinghoff-Borg, J., & Hamelryck, T. (2010). Calculation of accurate small angle X-ray scattering curves from coarse-grained protein models. *BMC Bioinformatics*, *11*(1), 429.

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

Tiede, D. M., Zhang, R., & Seifert, S. (2002). Protein conformations explored by difference high-angle solution x-ray scattering: Oxidation state and temperature dependent changes in cytochrome c. *Biochemistry*, *41*(21), 6605–6614.

Tjioe, E., & Heller, W. T. (2007). Ornl_sas: Software for calculation of small-angle scattering intensities of proteins and protein complexes. *Journal of Applied Crystallography*, *40*(4), 782–785.

Traube, J. (1895). Ueber Das Molekularvolumen. *Berichte der Deutschen Chemischen Gesellschaft*, *28*(3), 2722–2728. https://doi.org/10.1002/cber.18950280378.

Virtanen, J. J., Makowski, L., Sosnick, T. R., & Freed, K. F. (2011). Modeling the hydration layer around proteins: Applications to small-and wide-angle x-ray scattering. *Biophysical Journal*, *101*(8), 2061–2069.

Yang, S., Park, S., Makowski, L., & Roux, B. (2009). A rapid coarse residue-based computational method for X-ray solution scattering characterization of protein folds and multiple conformational states of large protein complexes. *Biophysical Journal*, *96*(11), 4449–4463.