Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Interpreting solution X-ray scattering data using molecular simulations

Jochen S Hub

Small-angle and wide-angle X-ray scattering in solution (SAXS, WAXS, SWAXS) is an increasingly accurate method for obtaining information on biomolecular structures, ensembles, and time-resolved dynamics at near-native conditions. However, the interpretation of the solution scattering data by computational methods is complicated by the low information content of the data, by scattering contributions from the hydration layer, and by unknown systematic errors. In the light of available computational methods, we first review the main computational challenges with the interpretation of SWAXS data. Molecular dynamics (MD) simulations may help to overcome these challenges and guide the interpretation of SWAXS in multiple ways. The physical information in atomistic force fields complements the low-information SWAXS data; explicit-solvent MD may be used to predict solvent scattering, and the MD-related sampling methods may guide the structure refinement against SWAXS data.

Address
University of Goettingen, Institute for Microbiology and Genetics, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

Corresponding author: Hub, Jochen S (jhub@gwdg.de)

Current Opinion in Structural Biology 2018, **49**:18–26

This review comes from a themed issue on **Theory and simulation**

Edited by **Robert Best** and **Kresten Lindorff-Larsen**

https://doi.org/10.1016/j.sbi.2017.11.002

Introduction: combining experimental data and computation

The discovery of biomolecular structures and ensembles requires joint experimental and computational efforts. This is because the information content of common experimental data is too low to define all the degrees of freedom of a biomolecule. Instead, the data must be complemented with additional physical, stereo-chemical or structural information to avoid overfitting during structure determination and refinement.

A wide range of physical models have been used to complement experimental data. Popular physical models vary greatly by the number of degrees of freedom, ranging from rigid-body up to flexible all-atom models. The energy terms of the physical models (force fields) vary by accuracy, predictive power, and computational cost, ranging from simple volume-exclusion restraints up to all-atom molecular-dynamics force fields. An appropriate choice for a physical model is largely determined by the information content of the experimental data. Generally, the lower the information content of the data, the more predictive physical models are required to avoid overfitting.

Algorithms for combining experimental data with physical models may be roughly grouped into three families: first, a trivial 'algorithm' is to compare the experimental data with the data back-calculated from an unbiased simulation. Obviously, this approach is restricted to physical models that are sufficiently accurate to propose a good structural model. Second, unbiased Boltzmann sampling may be used to propose an approximate structural ensemble, and the ensemble may be reweighted *a posteriori* such that the reweighted ensemble agrees with the experimental data. Third, the experimental data may be integrated into a physical simulation as an experiment-derived energetic bias, thereby restraining the simulation into conformations that meet both the physical model and the experimental data [1].

Several excellent reviews summarized recent developments in structural modeling, with a focus on crystallography, NMR, cryo-EM, and integrative modeling, including perspectives on Bayesian methods and maximum entropy-based ensemble refinement ([2–4,5••,6••,7] and references therein). The concepts developed in these reviews are well transferrable to the interpretation of solution scattering data, so they shall not be repeated here. Instead, this review focuses on challenges specific to solution scattering, and we outline recent developments based on molecular dynamics (MD) simulation to overcome such challenges.

Small-angle and wide-angle X-ray scattering (SAXS, WAXS, or SWAXS) has developed to an increasingly accurate and predictive tool for obtaining structural information on biomolecules in solution [8–13]. The quality of SWAXS data has greatly benefited from better light sources, single-photon counting detectors, and from setups coupled to size-exclusion chromatography (SEC-SAXS) [14], which have led to reduced statistical noise and reduced systematic errors due to particle aggregation.

The risk of systematic errors has been further reduced by established standards for sample preparation, quality control, and data deposition [15,16]. Additional improvements are expected with the increasing availability of in-house SWAXS beamlines, which help to optimize samples before beamtime at the synchrotron. The sister method, small-angle neutron scattering (SANS), remains popular owing to the possibility of contrast variation experiments.

In practice, the structural interpretation from SWAXS and SANS is often conducted using computationally efficient algorithms and simplified physical models (such as simulated annealing, rigid-body modeling, and so on.), allowing modeling calculations on the laptop [17]. Since solution scattering data are often of low spatial resolution and of low information content, and since the data used to be noisy and at times biased by systematic errors, there was limited demand for elaborate and computationally expensive analysis methods (such as Bayesian methods, all-atom MD, etc.). Only in recent years, thanks to increasing availability of SWAXS data with low noise and reduced systematic errors, the need for elaborate methods that are capable of harvesting all the structural information in the data is more and more recognized [18].

Challenges with the interpretation of SWAXS data

To motivate the need for MD-based methods for SWAXS interpretation, we first review some of the challenges related to SWAXS data analysis. Recent MD-based method developments discussed further down are primarily an answer so these challenges.

Challenge 1: low information content of SWAXS curves

Because SWAXS curves are smooth and one-dimensional (1D), they contain quite a limited amount of information. How the information is distributed over the q-range is a matter of ongoing research [19•], but it is generally accepted that experimental SWAXS curves do not contain more than 10–30 independent data points [8,9]. Hence, the number of backbone angles of biomolecules exceeds the number of independent data points of SWAXS curves by roughly two orders of magnitude. This precludes any straightforward fitting of protein structures against SWAXS data, but instead it leads to a high risk of overfitting.

One approach to reduce the risk of overfitting structural models is to constrain nearly all degrees of freedom of the biomolecule except for a few collective modes, leading to methods such as rigid-body modeling and normal mode refinement [20–22]. Alternatively, accurate physical and chemical knowledge may be added to the low-information SWAXS data, as provided by the potential energy function of modern biomolecular force fields.

Challenge 2: prediction of SWAXS curves from structural models

The interpretation of experimental data in terms of biomolecular structures requires accurate 'forward models', that is, theoretical predictions of the data from a given structural model. The physical basis of the SWAXS intensity is well understood; namely, the intensity is given by the orientationally averaged Fourier transform of the correlation function of the electron density contrast (compared to the buffer).

Since the quality of the experimental SWAXS data is improving, increasingly accurate forward models are required to harvest the structural information encoded in finer details of the data. Three aspects complicate the prediction of SWAXS curves of biomolecules:

(i) the density of the hydration layer of biomolecules differs from the density of bulk solvent, thus contributing the density contrast [23,24]. Simplified methods for SWAXS prediction do not predict the hydration layer density and hence require fitting of the hydration layer against the data (Table 1) [25–30]. This fitting may simultaneously adjust the radius of gyration R_g, a key quantity extracted from SWAXS experiments. In consequence, the fitting parameter may absorb smaller modulations of R_g [31].

(ii) Computing the density contrast requires accurate knowledge of the volume of solvent that is displaced by the solute. Many methods for SWAXS prediction rely on 'reduced atomic form factors', which incorporate the volume displaced by each atom into the atomic form factors. The displaced atomic volumes are often taken from tables derived from high-resolution crystal structures [32] or densitometric data [33,34], and subsequently fitted against the data. This protocol may add uncertainty since estimates for atomic volumes greatly differ [25,32,35]. For instance, the volume for the backbone N–H group reported by Fraser *et al.*, as adopted by many SWAXS tools [34], is roughly half the volume estimated from Voronoi tessellation [35]. In addition, effective atomic volumes could depend on the type of solute and atomic packing density (rigid vs. disordered protein, nucleotide, or detergent), posing additional uncertainty on the volume estimates.

(iii) Fluctuations of atoms and larger groups may influence the SWAXS curve [19•,36–38], but are often neglected during the interpretation of the data, adding uncertainty to the interpretation of the data.

Challenge 3: overfitting and lack of a cross validation set

Because of the low information content of SWAXS data, overfitting is a major problem for structural modeling [39]. A rigorous strategy to avoid overfitting would be to

Table 1

Incomplete list of methods for predicting SWAXS curves from structural models: Fitting of hydration layer required ($\delta\rho_{fit}$, including method that ignore the hydration layer), using tabulated reduced form factors (f_{red}), resolution [atomistic or coarse grained (CG)], fluctuations included, free availability [Download (D), web server (W)]. Additional software is listed in Refs. [63,64]

ID	Name/authors	Year	$\delta\rho_{fit}/f_{red}$	Resol.	Fluct.	Avail.	Refs.
Implicit solvent methods							
1	CRYSOL	1995	Yes/yes	atom.	–	D/W	[25]
2	ORNL-SAS	2007	Yes/yes	atom.	–	D	[65]
3	SoftWAXS	2009	Yes/–	atom.	–	D	[66]
4	Fast-SAXS-pro	2009	Yes/yes	CG	Yes	D/W	[30,36]
5	FoXS	2010	Yes/yes	atom.	–	D/W	[67,29]
6	PHAISTOS	2010	Yes/yes	CG	–	D	[68]
7	AquaSAXS/AquaSol	2011	Yes/yes	atom.	–	W	[27]
8	SASbtx/Zernike	2012	Yes/–	atom.	–	W	[69]
9	RISM-SAXS	2014	–/yes	atom.	–	D	[70]
10	BCL::SAXS	2015	Yes/yes	atom.	–	D	[71]
11	Pepsi-SAXS	2017	yes/yes	atom.	–	D	[72•]
Explicit solvent methods							
12	SASSIM/Sassena	2002	–/yes	atom.	Yes	D	[73]
13	MD-SAXS	2009	–/–	atom.	Yes	–	[74,75]
14	AXES	2010	Yes/–	atom.	–	W	[26]
15	HyPred	2011	–/–	atom.	–	W	[76]
16	Park et al.	2009	–/–	atom.	–	–	[77]
17	Köfinger &Hummer	2013	–/–	atom.	Yes	D	[78]
18	WAXSiS	2014	–/–	atom.	Yes	D/W	[38,79]

decompose the data into statistically independent training and test sets, as has become routine in crystallography and, more recently, in cryo-EM [40–42]. Accordingly, the structural model is fitted using only the training set, while overfitting is excluded by cross-validation against the test set. However, this strategy requires that the information in the test set is not used during model fitting — indeed, problems due to the implicit use of the test set, for instance by human supervision [43] or in consequence of non-crystallographic symmetry, have been discussed [44].

Since it is unclear how the structural information of SWAXS curves is distributed over the q-range, SWAXS curves were not decomposed into independent training and test sets. Instead, overfitting must be avoided by the use of sufficiently accurate physical models, thereby requiring only little model optimization against the data [45]. Further, the χ_{free} measure has been proposed in analogy to the R_{free} value in crystallography [40,46], yet the significance of χ_{free} has remained controversial [47].

In principle, Bayesian methods for model refinement should avoid overfitting. If the information in the data together with the physical model is insufficient to define the biomolecular structure, Bayesian methods would yield merely wide posterior distributions for the refined model, indicating large confidence intervals. Such methods have been used in NMR refinement since long [48], but related approaches have only recently been considered for SWAXS-based structure refinement [49•,50•].

Challenge 4: account for systematic errors

Common sources for systematic errors in SWAXS are aggregation, inter-particle repulsion, and poor buffer matching. Because modern photon-counting detectors allow data collection with low statistical errors, the overall uncertainty of the data is by now often dominated by systematic errors. This development asks for a statistically founded treatment of systematic errors. Bayesian methods that simultaneously estimate the biomolecular structures and systematic errors provide a route to accomplish this [50•], as originally proposed in the context of NMR refinement [48,51].

Challenge 5: confidence intervals of structural models

Students in natural science are taught that reporting quantities without at least approximate confidence intervals (CIs) is pointless. In SWAXS-based modeling, methods for the calculation of statistically founded CIs for structural models are still underdeveloped. Calculating CIs is generally complicated by unknown systematic errors, unknown errors of the physical model, and by the unclear information content of SAXS data, but also by the lack of computational tools for propagating the errors of the data and of the physical model into the error of the fitted structural model.

An important step toward such error analysis was recently suggested for *ab initio* low-resolution modeling against SWAXS data, based on the variability of repeated simulated annealing calculations [52]. However, it should be noted that the spread of repeated 'best fits' is not equivalent to CIs. Only very recently, a route to computing CIs based on Bayesian statistics has been shown in the

context of structure refinement against SAXS data [50•]. The error of the physical model is so far neglected. Interesting ideas for propagating the error in the physical model by using 'distributions of force fields' instead of using a single force field were proposed, but the ideas were to the best of our knowledge not yet implemented [5••].

Challenge 6: interpretation of SWAXS data of heterogeneous ensembles

Algorithms for ensemble reconstructions were reviewed frequently, hence we mention these only briefly [6••,39,53–55]. The majority of work has focused on NMR data, partly complemented by SWAXS, whereas less work has focused on pure SWAXS data [21,45,56,57]. Because of the lower information content of SWAXS compared to NMR data, reconstructing the ensembles from SWAXS data is more challenging and more prone to overfitting, suggesting that the data should, if possible, be complemented by accurate physical models. Indeed, it was pointed out that reproducing the data with a fitted ensemble does by far not guarantee that the ensemble is correct; hence, great care has been advised for interpreting ensembles fitted to SWAXS data [6••].

Most algorithms for SAXS-based ensemble modeling follow an 'ensemble reweighting' approach [7,58], which requires exhaustive sampling of the conformational space in free a simulation. Therefore, reweighing approaches were mainly used with simplified physical models, such as coarse-grained or rigid-body models [21,57], or using an accelerated sampling scheme that disturbs the Boltzmann distribution [58]. Theoretical frameworks for refining ensembles by energetic restraints to experimental data have been developed, yet a previous work focused on test applications of reduced complexity and simplified force fields [49•,59,60,61,62••]. As such, the refinement of ensembles of complex biological systems based on SWAXS data and accurate physical models, if possible including the calculation of CIs, has remained challenging.

Method developments for SWAXS data interpretation with atomistic MD simulations

Prediction of SWAXS curves

In the pioneering SWAXS prediction method by Merzel and Smith, the hydration layer was taken for the first time from an explicit-solvent MD simulation, revealing an increased hydration layer density of lysozyme; however, the method modeled the buffer subtraction using reduced form factors (hence using tabulated atomic volumes) [24,73]. Inversely, the AXES method avoids the use of 'reduced form factors' by using an explicit-solvent representation for the excluded solvent; however, the hydration layer is not optimized by an MD simulation and hence requires fitting against the experimental data [26]. Recently, a range of methods have been proposed for SWAXS prediction based on explicit-solvent MD simulations of both the hydration layer *and* the excluded solvent [38,74,75,77,78]. Here, Oroguchi and coworkers built upon earlier work by Seki *et al.* [80], while Chen and Hub built upon the method by Park *et al.* [77], making it applicable to flexible MD simulations, and correcting it for small yet significant force field and finite-size artifacts [38], similar to thoughts by Köfinger and Hummer [78]. The method was also implemented in the web server WAXSiS [79], which runs an explicit-solvent MD simulation with an uploaded biomolecule, making the SWAXS predictions accessible to non-experts. Notably, computationally efficient approximations for modelling the hydration layer with atomic resolution were suggested, either based on integral equations (RISM) [70], or based on protein-water radial distributions functions parameterized against MD simulations (HyPred) [76].

Compared to implicit-solvent methods (see Table 1), SWAXS predictions based on explicit-solvent MD are computationally more expensive but hold a number of advantages: firstly they naturally reproduce the increased radius of gyration due to the hydration layer [38], thus avoiding a fitting parameter for the hydration layer. This further avoids that the radius of gyration is fitted against the experimental data. Secondly by avoiding the use of 'reduced form factors', they circumvent uncertainties and fitting parameters due to the atomic volumes. Thirdly the SWAXS predictions remain valid at wide angles, where the internal structure of water may be reflected in the signals, and finally most of the methods account for thermal fluctuations of the biomolecule (see Table 1). For small solutes such as short peptides, however, explicit-solvent SWAXS predictions may converge slowly, in consequence of a weak contrast between solute and solvent. Hence, careful convergence checks are generally advised. The applied water force field hardly influences the SAXS predictions if (and only if) inaccurate bulk densities of certain water models are corrected. In the WAXS regime around the water scattering peak, in contrast, the water force field strongly influences the predictions, as expected from the fact that radial distributions functions of water are quite force field-dependent [38,81].

However, since only few studies compared implicit-solvent with explicit-solvent methods, it is not well understood under which conditions explicit-solvent models are mandatory for correctly interpreting the data, and under which conditions implicit-solvent methods offer an acceptable and computationally efficient approximation. SAXS curves computed from ensembles of disordered proteins agreed well at small angles between implicit-solvent and explicit-solvent methods [82•,83]. In the same study, in contrast, only an explicit-solvent method provided accurate fitting-free SAXS curves of the ring-shaped PCNA protein [83]. Another study investigated the effect of solvent-related fitting parameters, when

simulations of a protein-detergent complex, conducted with increasing detergent numbers N_{det}, are validated against SEC-SAXS data. [14,31]. The authors found that implicit-solvent methods tend to overfit the hydration layer density, absorbing the variations of N_{det} into the fitting parameter, and thereby loosing the ability to differentiate between the right and wrong N_{det}. Explicit-solvent calculations did not face these problems. Further, the study revealed that an ensemble average of the protein-detergent complex agreed better with experiment than any single structure, highlighting the importance of conformational fluctuations even for a relatively well-defined system such as a protein-detergent complex. Clearly, more work is required to identify strengths and limitations of available SWAXS prediction methods.

Structure refinement by energetic restraints to SWAXS data

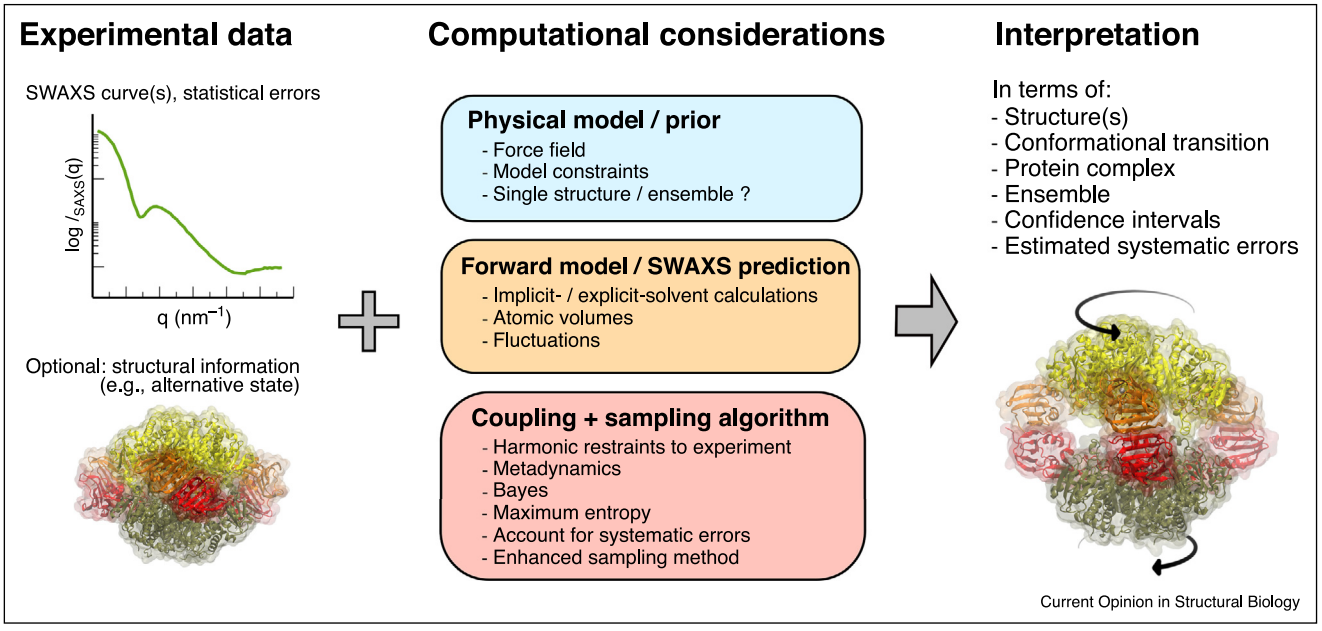
Multiple MD-based methods were recently published that aim at the refinement of biomolecular structures against SWAXS data using experiment-derived energetic restraints [50•,84•,85•,86]. Forces derived from energetic restraints may 'drive' the biomolecule into conformations that satisfy the experimental data, thereby, for instance, guiding the simulation over energy barriers associated with large-scale conformational transitions. As such, simulations are not only a tool to interpret the data; instead, SAXS data may, in turn, help the simulation to overcome limitations due to sampling problems and force field inaccuracies. This property is in contrast to methods based on 'reweighting' of states that require the sampling of the relevant states in a free, unbiased simulation, thereby restricting the methods to simplified and computationally efficient physical models [45,57].

Energetic restraints to the SWAXS data have been implemented with a harmonic biasing potential [85•,86] using metadynamics [84•], or following a Bayesian approach [50•]. Further, the SAXS curve prediction (or forward model) used to evaluate the energetic restraints vary between the methods: certain methods used the Debye formula, similar to pioneering work by Grishaev *et al.* on combined SAXS/NMR refinement [84•,86,87]. These two methods neglected scattering contributions from the hydration layer [84•,86]. Other methods accurately accounted for the hydration layer and excluded solvent by using explicit-solvent descriptions [50•,85•].

Interpretation of time-resolved SWAXS data

Time-resolved SWAXS (TR-SWAXS) experiments were used to track conformational dynamics of photoactive proteins in solution. Synchrotron-based TR-SWAXS experiments have covered the time regime between milliseconds and ~100 ps [88–90]. Conducted at X-ray free electron lasers, such experiments have reached sub-picosecond time resolution, providing a glimpse on ultra-fast protein dynamics [91,92]. The interpretation of such signals in terms of real-space dynamics is complicated by

Figure 1



Scheme and considerations for the interpretation of SWAXS data. For the interpretation of SWAXS data, MD simulations and MD force fields may provide an accurate physical model (prior), be used for accurate SWAXS predictions (forward model), and provide algorithms for conformational sampling.

the fact that the signals reflect non-equilibrium conditions; in the picosecond regime, the signals even reflect highly dissipative processes. To interpret such data, methods for the refinement of equilibrium structures and ensembles may not be applicable, but instead realistic non-equilibrium simulations, which reproduce the data without an experiment-derived bias, are needed.

Synchrotron-based TR-SWAXS data were interpreted using available crystallographic structures, or via low-resolution bead-modeling [88,89,93]. Recently, MD simulations were used to interpret sub-picosecond TR-SWAXS data of myoglobin, providing an atomic view of the ultrafast dissipation of energy following CO-photodissociation [92,94•], a process coined 'protein quake' in the 1980s [95]. The simulations revealed a strong SAXS signal as the quake propagated across the hydration layer into bulk solvent. Hence, the SAXS data were dominated by solvent and not by protein dynamics, highlighting that explicit-solvent SAXS calculations were required to correctly interpret the data.

Notably, TR-SWAXS patterns are anisotropic [96,97]. The anisotropy leads to additional structural information compared to common orientationally averaged equilibrium SWAXS data. A method for computing anisotropic TR-SWAXS patterns has been proposed [98], but applications that harvest the additional information remain elusive.

Outlook

A reliable atomistic interpretation of SWAXS data builds on three computational pillars (Figure 1): (a) accurate physical models, that is, accurate prior distributions for conformational states; (b) accurate and predictive SWAXS curve calculations from atomic models (forward models); and (c) algorithms founded on probability theory for integrating available experimental data into conformational sampling, for deriving structures and ensembles. All three pillars are, we believe, equally important. Most of the previous works focused on one of the three pillars, while paying less attention to the others, which is a practical strategy at the stage of method developments. Since force fields, SAXS prediction methods, and sampling methods have seen exciting developments in recent years, future efforts should try to integrate developments from the three pillars. In addition, the credibility of structural models derived from SWAXS will strongly benefit from increased efforts in computing confidence intervals, as well as from cross validation against complementary data, for instance from SANS, FRET, or DEER. Finally, benchmark suites of high-precision experimental SWAXS data are urgently needed to reveal strengths and weaknesses of the wide range of SWAXS-related software. As high-precision experimental SWAXS data are becoming more abundant, such developments will pave the way to a truly quantitative use of SWAXS data in structural and unstructural biology [53].

Conflict of interest statement

Nothing declared.

Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (grant no. HU 1971/1-1, HU 1971/4-1).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Jack A, Levitt M: **Refinement of large structures by simultaneous minimization of energy and R factor**. *Acta Crystallogr A* 1978, **34**:931-935.

2. Schröder GF: **Hybrid methods for macromolecular structure determination: experiment with expectations**. *Curr Opin Struct Biol* 2015, **31**:20-27.

3. Allison JR: **Using simulation to interpret experimental data in terms of protein conformational ensembles**. *Curr Opin Struct Biol* 2017, **43**:79-87.

4. Schneidman-Duhovny D, Pellarin R, Sali A: **Uncertainty in integrative structural modeling**. *Curr Opin Struct Biol* 2014, **28**:96-104.

5. •• Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K: **Combining experiments and simulations using the maximum entropy principle**. *PLoS Comput Biol* 2014, **10**:e1003406.
This is an instructive perspective on using the maximum entropy principle for combining experimental data with physical models. The authors propose interesting options for propagating the error of the physical model into the error of the fitted structural model, by using 'distributions of force fields' instead of using a single force field.

6. •• Ravera E, Sgheri L, Parigi G, Luchinat C: **A critical assessment of methods to recover information from averaged data**. *Phys Chem Chem Phys* 2016, **18**:5686-5701.
An insightful review of methods for the fitting of structural ensembles against experimental data is being presented. Available methods are systematically classified and illustrated by a geometric model. Underlying assumptions and the correct interpretation of fitted ensembles are clarified.

7. Bonomi M, Heller GT, Camilloni C, Vendruscolo M: **Principles of protein structural ensemble determination**. *Curr Opin Struct Biol* 2017, **42**:106-116.

8. Koch MH, Vachette P, Svergun DI: **Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution, Quart**. *Rev Biophys* 2003, **36**:147-227.

9. Putnam CD, Hammel M, Hura GL, Tainer JA: **X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution, Quart**. *Rev Biophys* 2007, **40**:191-285 http://dx.doi.org/10.1017/S0033583507004635.

10. Graewert MA, Svergun DI: **Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS)**. *Curr Opin Struct Biol* 2013, **23**:748-754.

11. Rambo RP, Tainer JA: **Super-resolution in solution X-ray scattering and its applications to structural systems biology**. *Annu Rev Biophys* 2013, **42**:415-441.

12. Bizien T, Durand D, Roblina P, Thureau A, Vachette P, Pérez J: **A brief survey of state-of-the-art BioSAXS**. *Protein Pept Lett* 2016, **23**:217-231.

13. Trewhella J: **Small-angle scattering and 3D structure interpretation**. *Curr Opin Struct Biol* 2016, **40**:1-7.

14. Berthaud A, Manzi J, Pérez J, Mangenot S: **Modeling detergent organization around aquaporin-0 using small-angle X-ray scattering**. *J Am Chem Soc* 2012, **134**:10080-10088.

15. Skou S, Gillilan RE, Ando N: **Synchrotron-based small-angle X-ray scattering of proteins in solution**. *Nat Protoc* 2014, **9**:1727-1739.

16. Jeffries CM, Graewert MA, Blanchet CE, Langley DB, Whitten AE, Svergun DI: **Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments**. *Nat Protoc* 2016, **11**:2122-2153.

17. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI: **ATSAS 2.1, a program package for small-angle scattering data analysis**. *J Appl Crystallogr* 2006, **39**:277-286.

18. Boldon L, Laliberte F, Liu L: **Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application**. *Nano Rev* 2015, **6**:25661.

19. • Spill YG, Nilges M: **SAS profile correlations reveal SAS hierarchical nature and information content**. *PLOS ONE* 2017, **12**:e0177309.

Variations of the SAXS curve are correlated with fluctuations of the protein, suggesting that information on protein dynamics is not uniformly distributed over the *q*-range of SAXS curve.

20. Petoukhov MV, Svergun DI: **Global rigid body modeling of macromolecular complexes against small-angle scattering data**. *Biophys J* 2005, **89**:1237-1250.

21. Pelikan M, Hura GL, Hammel M: **Structure and flexibility within proteins as identified through small angle X-ray scattering**. *Gen Physiol Biophys* 2009, **28**:174.

22. Gorba C, Miyashita O, Tama F: **Normal-mode flexible fitting of high-resolution structure of biological molecules toward one-dimensional low-resolution data**. *Biophys J* 2008, **94**:1589-1599.

23. Svergun D, Richard S, Koch M, Sayers Z, Kuprin S, Zaccai G: **Protein hydration in solution: experimental observation by X-ray and neutron scattering**. *Proc Natl Acad Sci U S A* 1998, **95**:2267-2272.

24. Merzel F, Smith JC: **Is the first hydration shell of lysozyme of higher density than bulk water?** *Proc Natl Acad Sci U S A* 2002, **99**:5378-5383.

25. Svergun D, Barberato C, Koch MHJ: ***CRYSOL* — a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates**. *J Appl Crystallogr* 1995, **28**:768-773.

26. 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.

27. 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.

28. Liu H, Hexemer A, Zwart PH: **The Small Angle Scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering**. *J Appl Crystallogr* 2012, **45**:587-593.

29. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A: **Accurate SAXS profile computation and its assessment by contrast variation experiments**. *Biophys J* 2013, **105**:962-974.

30. Ravikumar KM, Huang W, Yang S: **Fast-SAXS-pro: a unified approach to computing SAXS profiles of DNA, RNA, protein, and their complexes**. *J Chem Phys* 2013, **138**:01B609.

31. Chen P-c, Hub JS: **Structural properties of protein-detergent complexes from SAXS and MD simulations**. *J Phys Chem Lett* 2015, **6**:5116-5121.

32. Gerstein M, Richards FM: *International Tables for Crystallography* . Chester: International Union of Crystallography; 2006 (Chapter 22.1).

33. Zamyatnin A: **Protein volume in solution**. *Prog Biophys Mol Biol* 1972, **24**:107-123.

34. Fraser R, MacRae T, Suzuki E: **An improved method for calculating the contribution of solvent to the X-ray diffraction pattern of biological molecules**. *J Appl Crystallogr* 1978, **11**:693-694.

35. Pontius J, Richelle J, Wodak SJ: **Deviations from standard atomic volumes as a quality measure for protein crystal structures**. *J Mol Biol* 1996, **264**:121-136.

36. Yang S, Park S, Makowski L, Roux B: **A rapid coarse residue-based computational method for X-ray solution scattering characterization of protein folds and multiple conformational states of large protein complexes**. *Biophys J* 2009, **96**:4449-4463.

37. Moore PB: **The effects of thermal disorder on the solution-scattering profiles of macromolecules**. *Biophys J* 2014, **106**:1489-1496.

38. Chen P, Hub JS: **Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data**. *Biophys J* 2014, **107**:435-447.

39. Hammel M: **Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS)**. *Eur Biophys J* 2012, **41**:789-799.

40. Brunger AT: **Free R value: a novel statistical quantity for assessing the accuracy of crystal structures**. *Nature* 1992, **355 (6359)**:472.

41. Falkner B, Schröder GF: **Cross-validation in cryo-EM-based structural modeling**. *Proc Natl Acad Sci U S A* 2013, **110**:8930-8935.

42. DiMaio F, Zhang J, Chiu W, Baker D: **Cryo-EM model validation using independent map reconstructions**. *Protein Sci* 2013, **22**:865-868.

43. Kleywegt GJ: **Separating model optimization and model validation in statistical cross-validation as applied to crystallography**. *Acta Crystallogr D* 2007, **63**:939-940.

44. Fabiola F, Korostelev A, Chapman MS: **Bias in cross-validated free R factors: mitigation of the effects of non-crystallographic symmetry**. *Acta Crystallogr D* 2006, **62**:227-238.

45. Rózycki B, Kim YC, Hummer G: **SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions**. *Structure* 2011, **19**:109-116.

46. Rambo RP, Tainer JA: **Accurate assessment of mass, models and resolution by small-angle scattering**. *Nature* 2013, **496 (7446)**:477.

47. Franke D, Jeffries CM, Svergun DI: **Correlation map, a goodness-of-fit test for one-dimensional X-ray scattering spectra**. *Nat Methods* 2015, **12**:419-422.

48. Rieping W, Habeck M, Nilges M: **Inferential structure determination**. *Science* 2005, **309(5732)**:303-306.

49. • Antonov LD, Olsson S, Boomsma W, Hamelryck T: **Bayesian inference of protein ensembles from SAXS data**. *Phys Chem Chem Phys* 2016, **18**:5832-5838.

A method for Bayesian inference on continuous protein ensembles from SAXS data is presented.

50. • Shevchuk R, Hub JS: **Bayesian refinement of protein structures and ensembles against saxs data using molecular dynamics**. *PLoS Comput Biol* 2017, **13**:e1005800 (in press).

MD simulations are employed for refining a single protein structure or an ensemble of two states against SAXS data. Bayesian inference is used to simultaneously estimate structures, structural weights, and systematic errors, providing statistically founded confidence intervals for structures and weights. SAXS predictions are based on explicit-solvent calculations.

51. Habeck M, Rieping W, Nilges M: **Weighting of experimental evidence in macromolecular structure determination**. *Proc Natl Acad Sci U S A* 2006, **103**:1756-1761.

52. Tuukkanen AT, Kleywegt GJ, Svergun DI: **Resolution of ab initio shapes determined from small-angle scattering**. *IUCrJ* 2016, **3**:440-447.

53. Tompa P: **Unstructural biology coming of age**. *Curr Opin Struct Biol* 2011, **21**:419-425.

54. Fisher CK, Stultz CM: **Constructing ensembles for intrinsically disordered proteins**. *Curr Opin Struct Biol* 2011, **21**:426-431.

55. Cordeiro TN, Herranz-Trillo F, Urbanek A, Esta na A, Cortés J, Sibille N, Bernaó P: **Small-angle scattering studies of intrinsically disordered proteins and their complexes**. *Curr Opin Struct Biol* 2017, **42**:15-23.

56. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI: **Structural characterization of flexible proteins using small-angle X-ray scattering**. *J Am Chem Soc* 2007, **129**:5656-5664.

57. Yang S, Blachowicz L, Makowski L, Roux B: **Multidomain assembled states of Hck tyrosine kinase in solution**. *Proc Natl Acad Sci U S A* 2010, **107**:15757-15762.

58. Bowerman S, Rana AS, Rice AM, Pham GH, Strieter ER, Wereszczynski J: **Determining atomistic SAXS models of tri-ubiquitin chains from Bayesian analysis of accelerated molecular dynamics simulations**. *J Chem Theory Comput* 2017, **13**:2418-2429.

59. Roux B, Weare J: **On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method**. *J Chem Phys* 2013, **138**:084107.

60. White AD, Voth GA: **Efficient and minimal method to bias molecular simulations with experimental data**. *J Chem Theory Comput* 2014, **10**:3023-3030.

61. White AD, Dama JF, Voth GA: **Designing free energy surfaces that match experimental data with metadynamics**. *J Chem Theory Comput* 2015, **11**:2451-2460.

62. •• Hummer G, Köfinger J: **Bayesian ensemble refinement by replica simulations and reweighting**. *J Chem Phys* 2015, **143** 12B634_1.

A framework for Bayesian inference of continuous ensembles is presented, based on a reference distribution and experimental data. The relation to maximum entropy refinement and ensemble refinement with a fixed number of states is clarified.

63. Schneidman-Duhovny D, Kim SJ, Sali A: **Integrative structural modeling with small angle X-ray scattering profiles**. *BMC Struct Biol* 2012, **12**:17.

64. Putnam DK, Lowe EW Jr, Meiler J: **Reconstruction of SAXS profiles from protein structures**. *Comput Struct Biotechnol J* 2013, **8**:e201308006.

65. Heller WT, Tjioe E: **ORNL_SAS: software for calculation of small-angle scattering intensities of proteins and protein complexes**. *J Appl Crystallogr* 2007, **40**:782-785.

66. 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.

67. Schneidman-Duhovny D, Hammel M, Sali A: **FoXS: a web server for rapid computation and fitting of SAXS profiles**. *Nucl Acids Res* 2010, **38**:W540-W544.

68. Stovgaard K, Andreetta C, Ferkinghoff-Borg J, Hamelryck T: **Calculation of accurate small angle X-ray scattering curves from coarse-grained protein models**. *BMC Bioinform* 2010, **11**:429.

69. Liu H, Morris RJ, Hexemer A, Grandison S, Zwart PH: **Computation of small-angle scattering profiles with three-dimensional Zernike polynomials**. *Acta Crystallogr A* 2012, **68**:278-285.

70. Nguyen HT, Pabit SA, Meisburger SP, Pollack L, Case DA: **Accurate small and wide angle X-ray scattering profiles from atomic models of proteins and nucleic acids**. *J Chem Phys* 2014, **141**:D508.

71. Putnam DK, Weiner BE, Woetzel N, Lowe EW, Meiler J: **BCL:: SAXS: GPU accelerated Debye method for computation of small angle X-ray scattering profiles**. *Proteins Struct Funct Bioinform* 2015, **83**:1500-1512.

72. • S. Grudinin, M. Garkavenko, A. Kazennov, Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles, Acta. Cryst. D 73 (5). A very fast implicit-solvent SWAXS prediction software is introduced, using an adaptive multipole expansion depending on solute size and maximum scattering angle, with similar accuracy compared to established implicit-solvent methods.

73. Merzel F, Smith JC: **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 Crystallogr D* 2002, **58**:242-249.

74. Oroguchi T, Hashimoto H, Shimizu T, Sato M, Ikeguchi M: **Intrinsic dynamics of restriction endonuclease *eco* O109I studied by molecular dynamics simulations and X-ray scattering data analysis**. *Biophys J* 2009, **96**:2808-2822.

75. Oroguchi T, Ikeguchi M: **MD-SAXS method with nonspherical boundaries**. *Chem Phys Lett* 2012, **541**:117-121.

76. Virtanen JJ, Makowski L, Sosnick TR, Freed KF: **Modeling the hydration layer around proteins: applications to small-and wide-angle X-ray scattering**. *Biophys J* 2011, **101**:2061-2069.

77. Park S, Bardhan JP, Roux B, Makowski L: **Simulated X-ray scattering of protein solutions using explicit-solvent models**. *J Chem Phys* 2009, **130**:134114 http://dx.doi.org/10.1063/1.3099611.

78. Köfinger J, Hummer G: **Atomic-resolution structural information from scattering experiments on macromolecules in solution**. *Phys Rev E* 2013, **87**:052712.

79. Knight CJ, Hub JS: **WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics**. *Nucleic Acids Res* 2015, **43**:W225-W230.

80. Seki Y, Tomizawa T, Khechinashvili NN, Soda K: **Contribution of solvent water to the solution X-ray scattering profile of proteins**. *Biophys Chem* 2002, **95**:235-252.

81. Sorenson JM, Hura G, Glaeser RM, Head-Gordon T: **What can X-ray scattering tell us about the radial distribution functions of water?** *J Chem Phys* 2000, **113**:9149.

82. • Zheng W, Borgia A, Buholzer K, Grishaev A, Schuler B, Best RB: **Probing the action of chemical denaturant on an intrinsically disordered protein by simulation and experiment**. *J Am Chem Soc* 2016, **138**:11702-11713.

A refined force field is used to characterize the denaturant-induced unfolding of a protein. The unbiased simulations reveal good agreement with FRET and SAXS data.

83. Cordeiro TN, Chen P-c, De Biasio A, Sibille N, Blanco FJ, Hub JS, Crehuet R, Bernadó P: **Disentangling polydispersity in the PCNA-p15PAF complex, a disordered, transient and multivalent macromolecular assembly**. *Nucleic Acids Res* 2017, **45**:1501-2015.

84. • Kimanius D, Pettersson I, Schluckebier G, Lindahl E, Andersson M: **SAXS-guided metadynamics**. *J Chem Theory Comput* 2015, **11**:3491-3498.

Metadynamics is suggested as a sampling algorithm for driving simulations into agreement with SAXS data.

85. • Chen P, Hub JS: **Interpretation of solution X-ray scattering by explicit-solvent molecular dynamics**. *Biophys J* 2015, **108**:2573-2584.

A method for structure refinement is introduced, by coupling MD simulations to SWAXS data with a differentiable energetic restraint. The SWAXS prediction (forward model) is fully based on explicit-solvent calculations.

86. Björling A, Niebling S, Marcellini M, van der Spoel D, Westenhoff S: **Deciphering solution scattering data with experimentally guided molecular dynamics simulations**. *J Chem Theory Comput* 2015, **11**:780-787.

87. Grishaev A, Wu J, Trewhella J, Bax A: **Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data**. *J Am Chem Soc* 2005, **127**:16621-16628.

88. Andersson M, Vincent J, van der Spoel D, Davidsson J, Neutze R: **A proposed time-resolved X-ray scattering approach to track local and global conformational changes in membrane transport proteins**. *Structure* 2008, **16**:21-28.

89. Cammarata M, Levantino M, Schotte F, Anfinrud PA, Ewald F, Choi J, Cupane A, Wulff M, Ihee H: **Tracking the structural dynamics of proteins in solution using time-resolved wide-angle X-ray scattering**. *Nat Methods* 2008, **5**:881-886 http://dx.doi.org/10.1038/nmeth.1255.

90. Cho HS, Dashdorj N, Schotte F, Graber T, Henning R, Anfinrud P: **Protein structural dynamics in solution unveiled via 100-ps time-resolved X-ray scattering**. *Proc Natl Acad Sci U S A* 2010, **107**:7281-7286.

91. Arnlund D, Johansson LC, Wickstrand C, Barty A, Williams GJ, Malmerberg E, Davidsson J, Milathianaki D, DePonte DP, Shoeman RL *et al.*: **Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser**. *Nat Methods* 2014, **11**:923-926.

92. Levantino M, Schiro G, Lemke HT, Cottone G, Glownia JM, Zhu D, Chollet M, Ihee H, Cupane A, Cammarata M: **Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser**. *Nat Commun* 2015, **6**:6772.

93. Kim TW, Lee JH, Choi J, Kim KH, van Wilderen LJ, Guerin L, Kim Y, Jung YO, Yang C, Kim J *et al.*: **Protein structural dynamics of photoactive yellow protein in solution revealed by pump-probe X-ray solution scattering**. *J Am Chem Soc* 2012, **134**:3145-3153.

94. • Brinkmann LUL, Hub JS: **Ultrafast anisotropic protein quake propagation after CO-photodissociation in myoglobin**. *Proc Natl Acad Sci U S A* 2016, **113**:10565-10570.
It is shown that explicit-solvent SWAXS calculations are mandatory for the interpretation of femtosecond time-resolved SWAXS data. The first molecular movie of a protein quake is presented.

95. Ansari A, Berendzen J, Bowne SF, Frauenfelder H, Iben I, Sauke TB, Shyamsunder E, Young RD: **Protein states and proteinquakes**. *Proc Natl Acad Sci U S A* 1985, **82**:5000-5004.

96. Kim J, Kim KH, Kim JG, Kim TW, Kim Y, Ihee H: **Anisotropic picosecond X-ray solution scattering from photoselectively aligned protein molecules**. *J Phys Chem Lett* 2011, **2**:350-356.

97. Cho HS, Schotte F, Dashdorj N, Kyndt J, Anfinrud PA: **Probing anisotropic structure changes in proteins with picosecond time-resolved small-angle X-ray scattering**. *J Phys Chem B* 2013, **117**:15825-15832.

98. Brinkmann LUL, Hub JS: **Anisotropic time-resolved solution X-ray scattering patterns from explicit-solvent molecular dynamics**. *J Chem Phys* 2015, **143**:104108.