

# Combined Small-Angle X-ray and Neutron Scattering Restraints in Molecular Dynamics Simulations

Po-chia Chen,<sup>\*,†,∇</sup> Roman Shevchuk,<sup>‡,∇</sup> Felix M. Strnad,<sup>‡</sup> Charlotte Lorenz,<sup>§,#</sup> Lukas Karge,<sup>||</sup> Ralph Gilles,<sup>||</sup> Andreas M. Stadler,<sup>§,#</sup> Janosch Hennig,<sup>†</sup> and Jochen S. Hub<sup>\*,⊥</sup>

<sup>†</sup>Structural and Computational Biology Unit, EMBL Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany

<sup>‡</sup>Institute for Microbiology and Genetics, Georg-August-Universität Göttingen, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

<sup>§</sup>Jülich Centre for Neutron Science (JCNS-1) and Institute for Complex Systems ICS (ICS-1), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

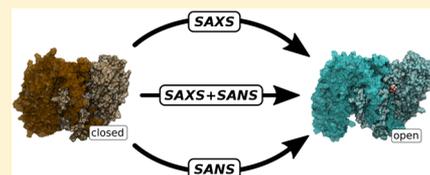
<sup>||</sup>Heinz Maier-Leibnitz Zentrum, Technische Universität München, Lichtenbergstrasse 1, 85748 Garching, Germany

<sup>⊥</sup>Theoretical Physics and Center for Biophysics, Saarland University, Campus E2 6, 66123 Saarbrücken, Germany

<sup>#</sup>Institute of Physical Chemistry, RWTH Aachen University, Landoltweg 2, 52056 Aachen, Germany

## Supporting Information

**ABSTRACT:** Small-angle X-ray and small-angle neutron scattering (SAXS/SANS) provide unique structural information on biomolecules and their complexes in solution. SANS may provide multiple independent data sets by means of contrast variation experiments, that is, by measuring at different D<sub>2</sub>O concentrations and different perdeuteration conditions of the biomolecular complex. However, even the combined data from multiple SAXS/SANS sets is by far insufficient to define all degrees of freedom of a complex, leading to a significant risk of overfitting when refining biomolecular structures against SAXS/SANS data. Hence, to control against overfitting, the low-information SAXS/SANS data must be complemented by accurate physical models, and, if possible, refined models should be cross-validated against independent data not used during the refinement. We present a method for refining atomic biomolecular structures against multiple sets of SAXS and SANS data using all-atom molecular dynamics simulations. Using the protein citrate synthase and the protein/RNA complex Sxl–Unr–msl2 mRNA as test cases, we demonstrate how multiple SAXS and SANS sets may be used for refinement and cross-validation, thereby excluding overfitting during refinement. For the Sxl–Unr–msl2 complex, we find that perdeuteration of the Unr domain leads to a unique, slightly compacted conformation, whereas other perdeuteration conditions lead to similar solution conformations compared to the nondeuterated state. In line with our previous method for predicting SAXS curves, SANS curves were predicted with explicit-solvent calculations, taking atomic models for both the hydration layer and the excluded solvent into account, thereby avoiding the use of solvent-related fitting parameters and solvent-reduced neutron scattering lengths. We expect the method to be useful for deriving and validating solution structures of biomolecules and soft-matter complexes, and for critically assessing whether multiple SAXS and SANS sets are mutually compatible.



## INTRODUCTION

In studies of biomolecular complexes, small-angle X-ray and neutron scattering (SAXS/SANS) techniques serve a vital role in the determination of global structure, subunit composition, structural diversity, and intrinsic disorder.<sup>1–6</sup> While SAXS and SANS rely upon similar physical principles, they both offer distinct advantages. Synchrotron SAXS measurements provide high-throughput data collection at low statistical noise for both small- and wide-angle regimes.<sup>7</sup> SANS is restricted to small scattering angles, but it leverages the collection of multiple independent small-angle data sets (i) by modulating the contrast via the D<sub>2</sub>O concentration of the buffer, and (ii) by measuring at different (per)deuteration conditions of the subunits of the biomolecule. In contrast-matching experiments, modulating the D<sub>2</sub>O concentration may render certain subunits invisible to SANS, thereby providing subunit-specific

structural information in multiprotein complexes, as well as subdomain-specific distances and orientations in multidomain proteins.<sup>8–12</sup> To maximize the amount of available structural information, both techniques are often combined, with a history in applications in soft-matter systems,<sup>13,14</sup> protein complexes,<sup>15,16</sup> heterogeneous protein/detergent systems,<sup>17</sup> and, more recently, biomolecular hydration.<sup>18</sup> Henceforth, we use the term “small-angle scattering” (SAS) when referring to both SAXS and SANS. SAS techniques can be further complemented with information derived from nuclear magnetic resonance (NMR), fluorescence spectroscopy, etc.<sup>19–24</sup> Such integration of diverse structural information

Received: March 21, 2019

Published: June 28, 2019

empowers research groups to tackle current challenges in molecular biology and soft-matter physics.

Combining multiple independent SAXS and SANS data sets obviously increases the amount of information that may be used to guide algorithms during structure refinement. Even more critically, and as emphasized in this work, multiple SAXS/SANS sets open the route for cross-validation to control against overfitting, which remains a central problem in SAS-based structural modeling.<sup>25,26</sup> Cross-validation has become routine in crystallography<sup>27</sup> and, more recently, in cryo-electron microscopy (EM)<sup>28,29</sup> but has remained underexplored in the SAS field.

The integration of data from independent sources, however, first requires that their mutual compatibility is validated—in other words, that the measurements indeed report the same structural state. Particular to SANS, the strong dependence of hydrogen-bond energies<sup>30</sup> on deuteration can lead to significant changes in complexation and aggregation behavior. Thus, SAXS curves are often measured in addition to SANS at different deuteration conditions to provide an independent reference during structural modeling.

Structure validation and refinement against SAS curves requires a “forward model”, i.e., an algorithm for computing SAS curves from a given structural model. Algorithms for SAXS curve prediction can be grouped into implicit-solvent and explicit-solvent methods.<sup>26</sup> Implicit-solvent methods use simplified models for the hydration layer and excluded solvent, hence requiring multiple solvent-related fitting parameters.<sup>31–34</sup> Such methods are computationally efficient, yet with the cost of losing some structural information to the free fitting parameters.<sup>35,36</sup> Further, implicit-solvent methods typically employ “reduced atomic form factors”, where the atomic form factors are corrected by the scattering of the displaced solvent. However, computing reduced form factors requires knowledge of atomic volumes, but atomic volumes are only approximately known and may depend upon the local chemical environment, leading to uncertainties during SAS predictions.<sup>26</sup> Explicit-solvent methods overcome such limitations at an increased computational cost by accurately modeling the hydration layer and excluded solvent from an atomistic molecular dynamics (MD) simulation.<sup>37–40</sup> Explicit-solvent methods were primarily used for the prediction of SAXS, and only very recently for the prediction of SANS curves.<sup>14</sup>

Refining protein structures against experimental data requires, apart from a forward model, also a spatial and energetic representation of the biomolecule, i.e., a physical model.<sup>26</sup> For refinement against SAS data, the physical model is particularly critical because the information in the data is by far insufficient to define all degrees of freedom of a biomolecule, suggesting that the information in the physical model is required to avoid overfitting during structure refinement. Relatively simple physical models such as rigid-body or normal-model representations have been used extensively in SAXS and SANS modeling.<sup>41–44</sup> Alternatively, accurate and computationally more expensive physical models based on all-atom MD simulations have been used to refine atomic structures and ensembles against SAXS.<sup>45–47</sup> Given that modern and well-validated force fields are used in such methods, the force field is capable of restraining the biomolecule in reasonable conformations of low free-energy, thereby greatly reducing the risk of overfitting. Our group has established a method for MD-based refinement against SAXS

data, termed “SAXS-driven MD”,<sup>45,48</sup> which uses explicit-solvent SAXS predictions as the forward model.<sup>40</sup> This initial method utilized a differentiable energetic penalty function to drive the simulation toward a distinct conformation that is compatible with the data. The method was recently improved by adopting a Bayesian framework.<sup>49</sup> Notably, the explicit-solvent SAXS predictions are available to nonexperts via the web server WAXSiS.<sup>50</sup>

In this work, our previously published SAXS-driven MD<sup>45</sup> was augmented to also incorporate SANS data, with the aim to (i) refine structures simultaneously against SAXS and SANS data, and (ii) to cross-validate SAS-refined structures against other SAS data sets that were not used during refinement. The method was applied to two systems: the ternary protein–RNA complex between Sex-lethal (Sxl), Upstream-of-N-Ras (Unr), and *male-specific lethal 2* (*msl2*) mRNA (together referred to as SUM),<sup>51</sup> and the closed-to-open transition of the enzyme citrate synthase.

## THEORY

Refinement simulations were conducted by coupling the MD simulations to experimental data with a hybrid potential

$$E_{\text{hybrid}}(\mathbf{r}) = E_{\text{ff}}(\mathbf{r}) + \sum_s E_s(\mathbf{r}) \quad (1)$$

where  $\mathbf{r}$  denotes the atomic coordinates of the biomolecule and the solvent,  $E_{\text{ff}}(\mathbf{r})$  is the force field energy, and the experiment-derived energies were taken as

$$E_s(\mathbf{r}) = k_s \beta^{-1} n_s^{-1} \sum_{i=0}^{n_s} \frac{[I_{\text{calc},s}(q_i; \mathbf{r}) - (f_s I_{\text{exp},s}(q_i) + c_s)]^2}{\sigma_i^2} \quad (2)$$

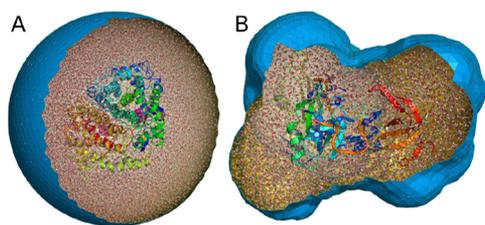
Here, the index  $s$  runs over the available SAXS and SANS data sets including SANS data sets at different D<sub>2</sub>O concentrations and different perdeuteration conditions of the biomolecule.  $k_s$  are the force constants,  $\beta$  is the inverse temperature, and  $n_s$  is the number of data points along  $q$  within set  $s$  used for coupling the simulation to the experimental data.  $I_{\text{calc},s}$  and  $I_{\text{exp},s}$  denote the calculated and experimental SAXS/SANS intensities, respectively. At each update step for  $I_{\text{calc},s}$  the two fitting parameters  $f_s$  and  $c_s$  were adjusted so as to minimize  $E_s$ :  $f_s$  adjusts the overall scale of  $I_{\text{exp},s}$  while  $c_s$  absorbs contributions from incoherent neutron scattering as well as uncertainties owing to buffer subtraction. The uncertainties  $\sigma_i$  include contributions from both statistical errors and estimated systematic errors. Where simulations were refined to match SAXS or SANS data alone, the respective force constants  $k_s$  for the excluded data sets were set to zero.

As a side note, given that the force constant is taken as  $k_s = N_{\text{indep},s}/2$ , where  $N_{\text{indep},s}$  is the number of independent data points in SAS curve  $s$ , the ensemble obtained by running a simulation with the energy  $E_{\text{hybrid}}(\mathbf{r})$  may be interpreted as the posterior distribution obtained by Bayesian inference.<sup>49,52</sup> In such a Bayesian framework, choosing  $f_s$  and  $c_s$  by minimizing  $E_s$  is equivalent with considering  $f_s$  and  $c_s$  as (initially unknown) nuisance parameters that were analytically marginalized out.<sup>49</sup> As such,  $f_s$  and  $c_s$  are not fixed to empirical ad hoc values, but instead all possible combinations of the unknown  $f_s$  and  $c_s$  are taken into account during structure refinement. Because the posterior quantifies our state of knowledge in the light of the data and the force field, the width of the posterior as given by the SAS-derived ensemble provides a rigorous confidence

interval for the refined biomolecular structures. In this work, however, we did not further emphasize the possibility of a Bayesian interpretation of SAS-restrained simulations. Instead, we here chose  $k_s$  empirically and focused on driving conformational transitions guided by SAS data.

## IMPLEMENTATION

**SANS Predictions.** The SAXS and SANS intensities were computed using the explicit-solvent algorithms first implemented via the WAXSiS method.<sup>50</sup> Here, we expanded WAXSiS capabilities to also tackle SANS prediction by replacing the X-ray atomic form factors with neutron scattering lengths, utilizing previously described formalisms.<sup>40</sup> Scattering length values were taken from ref 53. Due to the explicit-solvent description, all SAXS and SANS predictions employ an atomistic model for both the hydration layer and the excluded solvent. This approach provides several benefits, as it requires no solvent-related fitting parameters nor reduced form factors, and it also incorporates contributions from fluctuations of protein and solvent. To ensure a numerically accurate buffer subtraction,<sup>40</sup> we corrected the density of all water models to the experimental standard value of  $334 \text{ e nm}^{-3}$ . A fixed spatial envelope around the solute was defined to include the contribution of the hydration layer (Figure 1), while ensuring



**Figure 1.** Simulation snapshots of (A) citrate synthase and (B) Sxl–Unr–*msl2* mRNA. Explicit water molecules (red/white sticks) inside the envelopes (blue surface) contribute to the calculated SAXS and SANS curves. Water molecules present in the simulation system but outside the envelope, as well as some water in front of the biomolecules, are not shown for clarity.

that solvent behaviors at and beyond the envelope surface are sufficiently bulk-like. This results in cancellation of solvent contributions outside the envelope within the explicit-solvent formalism.<sup>38,40</sup> Our implementation of all methods described above is available via the authors' Web site (<https://biophys.uni-saarland.de/software.html>) as an in-house extension of the MD software GROMACS. We note that the net performance penalty was estimated to be 10–30%, relative to an unconstrained GROMACS simulation of the same version, with dependence on system size, number of constraints, and hardware acceleration.

**Deuteration of Biomolecule and Buffer, Isotope-Dependent Incoherent Scattering.** Biomolecules in deuterated buffer experience hydrogen–deuterium exchange at solvent exposed sites. We thus adopt a convention<sup>54</sup> similar to ATASAS that assumes only polar hydrogen atoms are exchangeable, with a reduced exchange ratio (90%) for backbone amide groups.<sup>55</sup> For the purposes of defining exchangeable hydrogens, our current implementation considers hydrogen atoms as polar if they are bound to a heavy atom of element symbol O, N, S, or P.

We implemented two protocols for computing the SANS signal of a biomolecule in partly deuterated buffer of deuterium

fraction  $w_{D_2O}$  ( $0 \leq w_{D_2O} \leq 1$ ). First, we assigned the mean neutron scattering length to all potentially deuterated hydrogen atoms, that is, to hydrogen atoms of water and to labile hydrogen atoms of the biomolecule:

$$b = (1 - f_{bb} w_{D_2O}) b_H + f_{bb} w_{D_2O} b_D \quad (3)$$

Here,  $b_H$  and  $b_D$  denote the scattering lengths of hydrogen and deuterium, respectively. The factor  $f_{bb}$  is 0.9 for backbone amide hydrogen atoms and unity for all other hydrogen atoms. This protocol does not account for any contributions from incoherent scattering, which arises if the scattering length of an element is not constant but varies due to (i) different isotopes or (ii) different orientations between nuclear spins and neutron spins.

Second, to account for isotope-dependent incoherent scattering, we implemented a stochastic deuteration protocol, where scattering lengths of  $b_H$  and  $b_D$  were randomly assigned to all potentially deuterated hydrogen atoms of buffer and biomolecule. This protocol models the experimental conditions in which individual hydrogen atoms at each time point are either fully deuterated or not but are obviously not partially deuterated. Here, at each update step for  $I_{calc,s}$ , each deuteratable hydrogen atom obtained a scattering length of  $b_D$  with the probability  $f_{bb} w_{D_2O}$ , and a scattering length of  $b_H$  otherwise. As may be easily confirmed with the Debye equation for the scattering intensity,  $I(q) = \sum_{ij} b_i b_j \sin(qr_{ij}) / qr_{ij}$ , where  $q$  is the momentum transfer and  $r_{ij}$  is the distance between atoms  $i$  and  $j$ , the two protocols lead to scattering intensities that differ on average only by a constant owing to the self-scattering contribution of the atom (terms with  $i = j$ ). Hence, averaging the atomic scattering lengths prior to computing the intensity (first protocol) leads only to a constant offset in the intensity curve. However, the second protocol accounts purely for isotope-dependent incoherent scattering but not for spin-dependent incoherent scattering. In this work, because (i) we fitted a constant offset to the experimental curve in any case to absorb contributions from spin-dependent incoherent scattering, and (ii) curves using stochastic deuteration converge more slowly, we only report results following the first protocol.

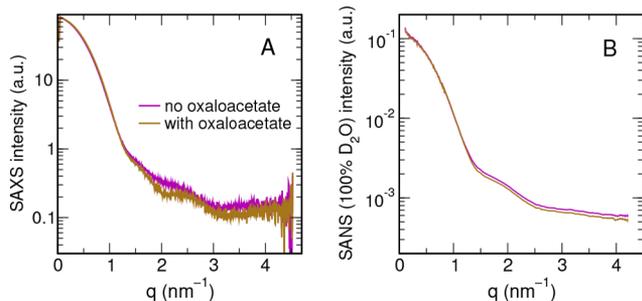
**SANS-Derived Forces.** The intensity gradients, as required to compute SANS-derived forces during refinement simulations, were computed similar to our previously established SAXS-driven MD method, again by replacing atomic form factors with neutron scattering lengths.<sup>45</sup> At large  $D_2O$  concentrations, however, the biomolecule may impose a negative contrast with respect to the buffer. Consequently, as the biomolecule moves in one direction, the contrast moves in the opposite direction, which must be accounted for during the calculation of the intensity gradients with respect to the biomolecule's atomic coordinates. For the present study, we implemented two schemes to account for negative contrast. First, we simply scaled the SAS-derived force by a factor of  $-1$  if the overall contrast of the biomolecule was negative. Second, we scaled all forces by the relative contrast  $(\rho_{solute} - \rho_{solvent}) / \rho_{solute}$ . Both procedures allowed us to refine biomolecular structures against SANS data at high  $D_2O$  and, hence, were sufficient for the simulations presented here. A more rigorous procedure, which we keep for future work, will be the use of solvent-reduced scattering lengths for computing intensity gradients (but not for the intensity itself), which would

likewise ensure the correct direction of SANS-derived forces at negative contrast.

**Note on D<sub>2</sub>O Force Field Parameters.** To our knowledge, there are no publicly available force field parameters for D<sub>2</sub>O that have been validated to be compatible with common biomolecular force fields. Thus, all simulations presented here were carried out in pure H<sub>2</sub>O, with assignment of deuterium scattering lengths as necessary for SANS computations.

## RESULTS AND DISCUSSION

**Citrate Synthase. Conformational Transitions upon Ligand Binding.** Citrate synthase (CS) is a well-characterized homodimeric enzyme involved in the first step of the citric acid cycle. It catalyzes the reaction of oxaloacetate with acetyl coenzyme A to citric acid. Each monomer contains two domains: a large domain mediating the monomer contacts, and a more flexible small domain. Binding of ligands into the cleft between the large and small domains leads to a closure of the cleft and, hence, to increased contacts between the large and small domains. Two conformations have been identified by crystallography, which provided atomic models for (i) a partially open state bound to citric acid (CIT) but in the absence of coenzyme A (CoA, PDB 1CTS) and (ii) for the closed state in the presence of both CIT and CoA (PDB 2CTS).<sup>56</sup> We collected SAXS data of CS at beamline BM29 of ESRF (Grenoble)<sup>57</sup> and at SWING beamline of SOLEIL (Paris),<sup>58</sup> as well as SANS data at 100% D<sub>2</sub>O at instrument SANS-1 of Heinz Maier-Leibnitz Zentrum (Garching).<sup>59,60</sup> SAXS and SANS data were collected in both the presence and absence of oxaloacetate (Figure 2 and Figure S1).



**Figure 2.** (A) SAXS curves collected with H<sub>2</sub>O buffer at BM29 of ESRF and (B) SANS curves at 100% D<sub>2</sub>O, in oxaloacetate-bound state (tan) and oxaloacetate-unbound state (purple).

Experimental SAXS and SANS data consistently demonstrated a small closure of CS structure upon binding with oxaloacetate, as measured using radius of gyration ( $R_g$ ) values computed from Guinier fitting of SAS curves (Table 1). For SAXS, a comparison of holo versus apo states reveals  $R_g$  decreases between 0.4 and 1.0 Å upon ligand addition. For SANS, this  $R_g$  decrease was 0.6 Å. Although a comparison of H<sub>2</sub>O solvent and D<sub>2</sub>O solvent at SOLEIL did not show any deuteration effects on the  $R_g$  of apo-CS, D<sub>2</sub>O appears to reduce the changes exerted by oxaloacetate. This suggests an influence of deuteration by decreasing ligand affinity. The SAXS data measured at ESRF reported a slightly larger  $R_g$  and a smaller change upon oxaloacetate binding, which is due to a residual presence of aggregated CS (Figure S2). In addition, we computed SAXS and SANS curves of the 1CTS and 2CTS

crystal structures, using backbone-restrained MD simulations and explicit-solvent SAXS/SANS predictions (Table 1). Both SAXS and SANS calculations reflect the closure of CS upon binding of CoA (in addition to citric acid), as is evident from a decrease of  $R_g$  by 1.3 Å.

**Effect of Hydration Layer on SAXS and SANS Data.** Critical to this study, the explicit-solvent SAXS/SANS formalism naturally reproduces the effect of the hydration layer density upon  $R_g$ , as found for many proteins.<sup>37,40,55</sup> A comparison of  $R_g$  values between solution and vacuum conditions showed that the pure H<sub>2</sub>O SAXS  $R_g$  was ~0.9 Å larger than  $R_g$  directly computed from protein coordinates alone (Table 1). In contrast, the 100% D<sub>2</sub>O SANS  $R_g$  was ~0.5 Å smaller. These  $R_g$  modulations reflect an increased density of the hydration layer compared to bulk water; for SANS, the hydration layer manifests in a reduced  $R_g$  owing to the negative contrast of the protein in 100% D<sub>2</sub>O. As such, hydration effects are naturally included in the explicit-solvent SAS predictions, without the need for fitting the hydration layer against the experimental data.

We further quantified the impact of the hydration layer by exploring the difference between  $R_g$  values from SAXS and SANS at 100% D<sub>2</sub>O, defining  $\Delta R_g = R_g^{\text{SAXS/D}_2\text{O}} - R_g^{\text{SANS/D}_2\text{O}}$ . SAXS and SANS in D<sub>2</sub>O suggested the value  $\Delta R_g \approx 2$  Å, whereas MD simulations in H<sub>2</sub>O found a slightly smaller value of  $\Delta R_g \approx 1.4$  Å (Table 1). This difference is compatible with the results of Piana et al.<sup>61</sup> and Best et al.,<sup>62</sup> who found that common water models underestimate dispersion interactions with the protein. We hypothesize that the underestimation manifests in simulation as a slightly thinner hydration layer, and hence as a slightly smaller  $\Delta R_g$  than expected from experiment. However, also different packing of D<sub>2</sub>O versus H<sub>2</sub>O on the protein surface as well as small fractions of aggregated protein may partially explain a small difference of  $\Delta R_g$ . We expect that SAXS and SANS data coupled to size-exclusion chromatography (SEC-SAXS, SEC-SANS) as well as additional simulations in a future study will be required to rigorously compare the hydration layer between simulation and experiment.

**Structure Refinement against SAXS and SANS Data.** To demonstrate the feasibility of structure refinement using multiple SAS data sets, we refined CS starting from the closed state against SAS data of the open apo state, following three protocols: (i) coupling to SAXS data and cross-validation against SANS data; (ii) coupling to SANS data and cross-validation against SAXS data; and (iii) coupling to both SAXS and SANS data simultaneously (Figure 3C–H). Here, SAXS data collected at BM29 was used. For each protocol, five independent simulations were carried out. The restraints were gradually turned on during the simulation time interval between 1 and 2 ns.

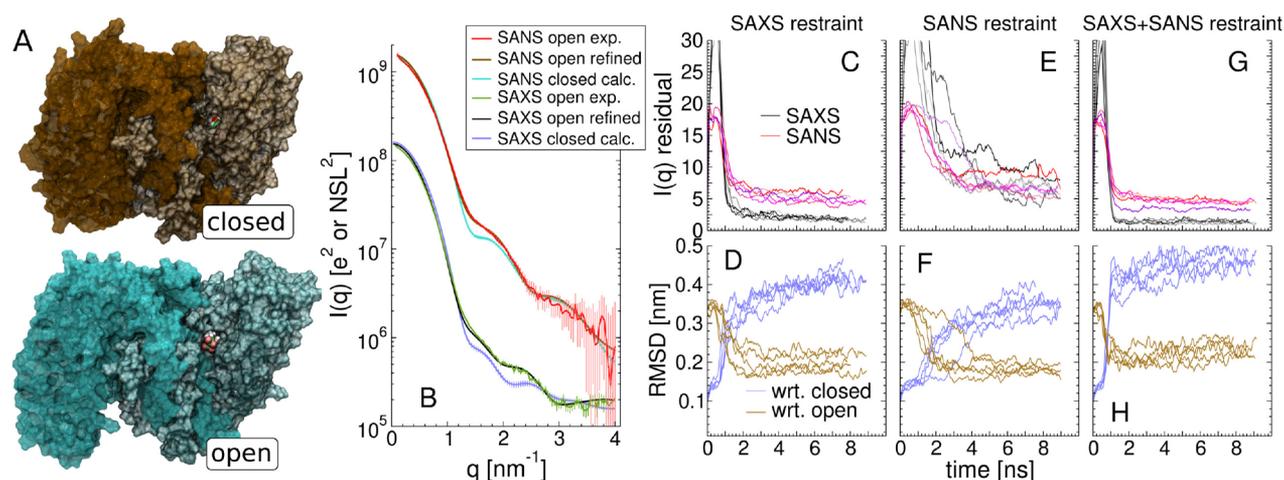
Regardless of the protocol, we observed a simultaneous decrease of both SAXS and SANS  $I(q)$  residuals (Figure 3C,E,G). This finding suggests that the SAXS and SANS data are mutually compatible; i.e., refinement against SAXS data does not violate the SANS data and vice versa. In turn, the compatibility between the SAXS and SANS sets allows us to employ them as training and test sets, that is, to allow for cross-validating the refined structure against SAS data not used during refinement.

However, on rare occasions during the opening simulations, only one of the two monomers would open and “overshoot” so as to match the overall experimental  $R_g$ . This was evident from

**Table 1.** Radius of Gyration  $R_g$  (Å) Values from Experimental Data and from Explicit-Solvent SAXS/SANS Predictions with Crystal Structures 1CTS and 2CTS<sup>56a</sup>

method	experiment/MD	solvent	ligand		
			none	CIT/OA	CIT + CoA
SAXS	SWING/SOLEIL	H <sub>2</sub> O	29.35 ± 0.06	28.34 ± 0.1	
SAXS	SWING/SOLEIL	D <sub>2</sub> O	29.31 ± 0.05	28.76 ± 0.1	
SAXS	BM29/ESRF	H <sub>2</sub> O	29.70 ± 0.01	29.31 ± 0.01	
SANS	SANS-1/MLZ	D <sub>2</sub> O	27.3 ± 0.2	26.7 ± 0.2	
SAXS	backbone-restrained MD	H <sub>2</sub> O		29.0	27.7
SANS	backbone-restrained MD	D <sub>2</sub> O		27.6	26.4
–	crystal structure	vacuum		28.1	26.9

<sup>a</sup>Experimental data with one ligand are based on CS bound to oxaloacetate (OA), with full details reported in Table S1. The backbone-restrained MD and the vacuum calculations are based on the 1CTS structure containing citric acid (CIT) or on the 2CTS structure containing CIT and coenzyme A (CoA).  $R_g$  from SAXS/SANS data were computed by Guinier analysis.  $R_g$  for the crystal structure in vacuum were computed purely from the atomic coordinates of the protein.



**Figure 3.** Summary of citrate synthase (CS) SAS and structure refinement via 10 ns SAS-restrained simulations. (A) Simulation snapshots of the closed (brown) and open (cyan) CS dimers. The two domains are distinguished by color intensity. The bound citric acid is depicted as spheres colored by element. (B) Collated SAS profiles: calculated SAXS and SANS curves of closed CS (light blue and cyan, respectively), calculated SAXS and SANS curves of refined open CS (black and brown, respectively), and experimental SAXS and SANS curves obtained in the absence of CoA (green and red, respectively). Experimental curves were fitted to the refined open curves by adjusting the absolute scale and a constant offset. SAXS and SANS curves are plotted in units of the squared unit charge  $e$  and the squared neutron scattering length (NSL). (C, D) Five opening trajectories each restrained to SAXS and cross-validated against SANS, (E, F) restrained to SANS data and cross-validated against SAXS, and (G, H) restrained to SAXS and SANS simultaneously. (C, E, G) Mean residual between calculated and experimental SAS curve plotted versus simulation time. Black, residual of SAXS curves; red, residual of SANS curves. (D, F, H) Backbone RMSD during refinement simulation, shown with respect to the open crystal structure (brown, PDB code 1CTS) and closed crystal structure (blue, PDB code 2CTS).<sup>56</sup>

increased root-mean-square deviations (RMSDs) of some simulation with respect to the open 1CTS structure (Figure S3). To prevent this overshooting behavior, we applied center-of-mass (COM) restraints to the CS subunits by matching the COM of the two small subunits in each monomer with the COM of the two large subunits (see Methods). This also loosely preserved the enzyme's twofold symmetry. As a consequence, the refined RMSD versus open 1CTS was greatly decreased and we then observed simultaneous opening of both monomers in all restrained simulations (Figure 3D,F,H).

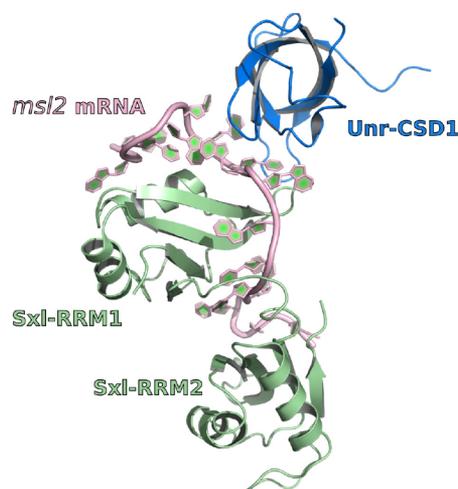
Overall, the analysis shows that MD simulations in conjunction with explicit-solvent SAXS/SANS predictions are capable of refining atomic models of biomolecules. However, SAXS/SANS data contain no symmetry information and may be insufficient alone to unambiguously direct conformational transitions. For symmetric proteins such as CS, additional symmetry restraints should be applied to enable consistent,

unambiguous refinement simulations. This symmetry-related requirement echoes results known in general SAXS modeling.

**Solution Refinement of Partial Sxl–Unr–*msl2* mRNA Ribonucleoprotein Complex.** The Sxl–Unr–*msl2* mRNA ternary complex (SUM) plays an important role in female *Drosophila* flies to maintain equal expression levels of X chromosome linked genes between the sexes. In this species, a dosage compensation mechanism occurs in males via a twofold increase in the transcription of X-linked genes achieved by the MSL complex. The *msl2* mRNA encodes an essential component of the protein complex, which is suppressed in females via the female-specific expression of Sex-lethal (Sxl) protein. The binding of Sxl to regulatory regions of *msl2* recruits Upstream-of-N-ras (Unr), forming an extended ternary complex. This complex formation is essential but not sufficient for preventing translation of *msl2* mRNA, thus leading to normal transcription levels of the female X chromosomes.

The core ternary complex has been revealed by combined crystallography with NMR and SAS validation (PDB ID

4QQB).<sup>51</sup> This structure contains two RNA recognition motif (RRM) domains of Sxl and the first cold shock domain (CSD1) of Unr, both bound to an 18-nucleotide-long stretch of *msl2* (Figure 4). The crystallographic conformation has



**Figure 4.** Cartoon representation 4QQB structure of Sxl–Unr–*msl2* mRNA.<sup>51</sup> Sxl domain shown in green, labeled by the first and second RNA recognition motifs (RRM1 and RRM2); first CSD domain of Unr shown in blue; and *msl2* mRNA in pink.

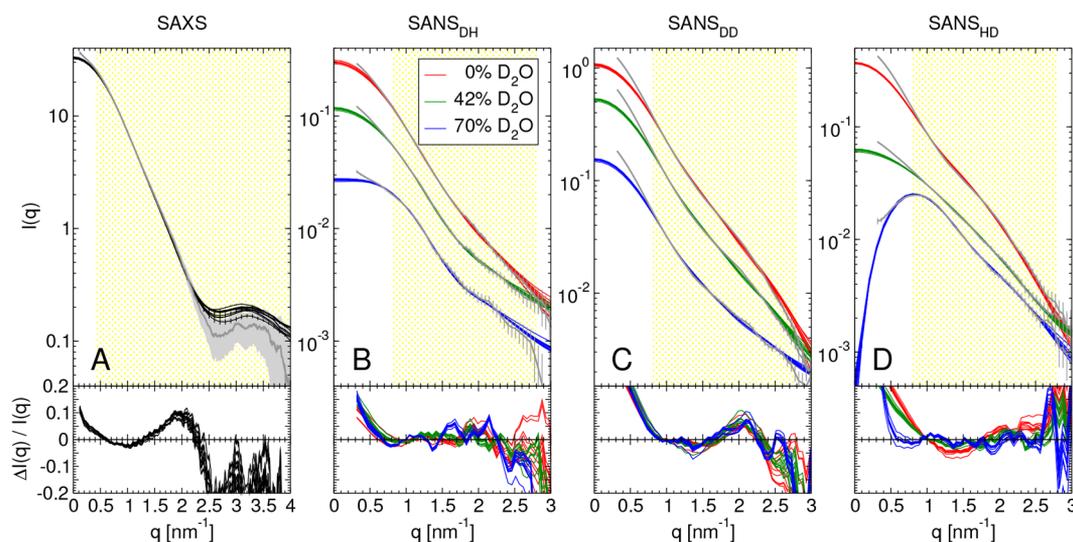
been validated in solution utilizing multiple SAXS/SANS data together with NMR restraints, and also in a separate study utilizing residual dipolar couplings alone.<sup>63</sup> During this validation process SANS curves were measured at nine deuteration conditions, several of which however have been excluded due to significant aggregation. In particular, increased solute aggregation in SAS curves was observed in experiments where both Sxl and Unr were perdeuterated, relative to conditions where only one of the proteins was perdeuterated. These observations raise questions as to (i) the actual degree

of consistency between multiple SAS data obtained via contrast variation and protein perdeuteration, and (ii) whether protein conformations may be influenced by the protein perdeuteration.

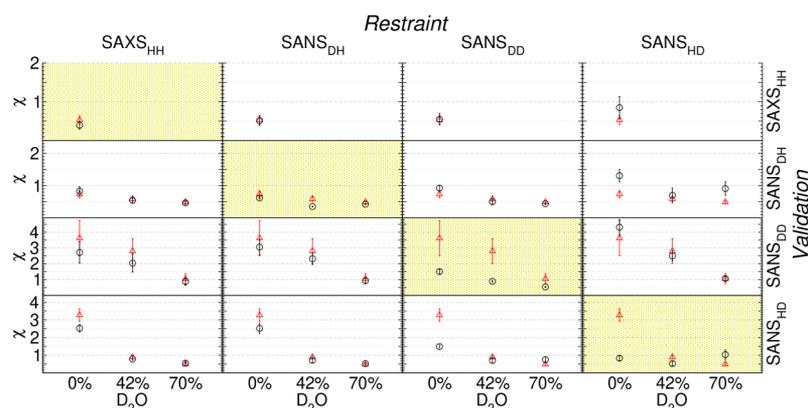
**Cross-Validation Studies.** Ten SAS curves at nine conditions have been taken from the published Sxl study,<sup>51</sup> comprised of one SAXS curve of the nondeuterated SUM at 0% D<sub>2</sub>O and nine SANS curves at multiple protein and solvent deuterations (Figure 5). For brevity, we will denote the protein deuteration state with two-letter subscripts and the solvent deuteration state with superscripts. For instance, the individual SANS curve containing deuterated Sxl and nondeuterated Unr in 42% D<sub>2</sub>O will be denoted as SANS<sub>DH</sub><sup>42%</sup> (Figure 5B, green). The three SANS curves with identical protein deuterations but different D<sub>2</sub>O contents will be denoted without superscripts, e.g. SANS<sub>DH</sub> (Figure 5B).

To confirm overall agreement of the 4QQB structure with the SAS curves, we first back-calculated the expected SAS patterns for the SUM complex based on conformations sampled in free MD simulations (Figure 5, black and colored lines). All experimental conditions are found to be adversely affected by aggregation, as evidenced by positive deviations (toward larger  $I(q)$ ) from the Guinier behavior at small  $q$  (Figure 5, gray lines). The extent of aggregation artifacts depends on deuteration conditions: The nondeuterated SAXS curve exhibited the smallest degree of aggregation, whereas the SANS<sub>HD</sub><sup>0%</sup> curve exhibited the largest (Figure 5D, compare red and gray lines). The degree of aggregation depended on solvent deuteration for SANS<sub>HD</sub>, but not for the other data sets. Overall, Figure 5 demonstrates that the free MD based on published structures is broadly consistent with the majority of SAS data but the agreement between calculated and experimental curves is compromised at low  $q$  owing to aggregation.

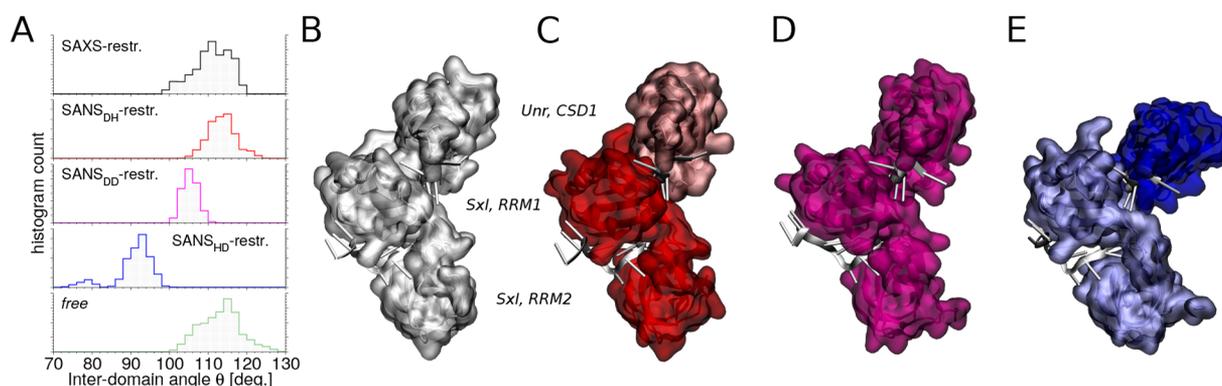
To examine whether different protein deuteration conditions affect the conformation of the complex, we divided the target data into four sets—SAXS<sub>HH</sub>, SANS<sub>DD</sub>, SANS<sub>DH</sub>, and



**Figure 5.** Comparison of SAS profiles from ten 50 ns free simulations of the SUM ternary complex with experimental scattering, with  $q$ -ranges used for fitting indicated as light yellow shading. In each panel, experimental SAS profiles are shown as gray lines and computed SAS curves are shown as black or colored lines. Deviations between the experimental and the calculated curve are shown in the bottom row. (A) SAXS profile for the nondeuterated SUM complex. (B–D) SANS profiles grouped by protein deuteration state, where each panel contains scattering data at three D<sub>2</sub>O concentrations (colored lines). See main text for explanation of protein deuteration nomenclature.



**Figure 6.** Four-way cross-validation of SAS-restrained MD ensembles, testing for dependence on protein deuteration. Each column reports data from a discrete set of restraints SAXS, SANS<sub>DH</sub>, SANS<sub>DD</sub>, and SANS<sub>HD</sub>, where all solvent deuterations have been targeted simultaneously. To guide the eye, blocks reporting on agreement with the training set have been shaded yellow. Agreement is reported as average and  $\sigma$  ( $N = 10$ ) of reduced  $\chi$ -values (black circles), computed from the last 2 ns of each replicate against experiment using experimental error over the  $q$ -range  $0.8\text{--}2.8\text{ nm}^{-1}$ . Equivalent calculations from unrestrained MD are shown for comparison (red triangles).



**Figure 7.** Structural variation of refined SUM complexes and representative conformations shown as nucleic acid backbones and protein surfaces. Darker shades and lighter shades of protein domains indicate respectively deuterated and nondeuterated forms. (A) Pinching of the SUM complex according to the interdomain angle  $\theta$ , as defined by the centers-of-mass of the three distinct domains: RRM2 and RRM1 domains of Sxl, and CSD1 domain of Unr. Distributions of  $\theta$  are shown for simulations restrained to (from top to bottom) SAXS, SANS<sub>DH</sub>, SANS<sub>DD</sub>, or SANS<sub>HD</sub>, as well as for free simulations (green). Histogram data taken from the final 2 ns of all trajectories. (B–E) Representative conformations after SAXS refinement (gray), SANS<sub>DH</sub> refinement (red), SANS<sub>DD</sub> refinement (pink), and SANS<sub>HD</sub> refinement (blue).

SANS<sub>HD</sub>—where different solvent deuterations were combined as simultaneous restraints. Each set of data was used in 10 SAS-restrained simulations of  $\sim 20$  ns, and the final 2 ns of each restrained simulation was used to compute SAS curves for comparison with experimental curves. The reduced  $\chi$ -values from 10 repeated simulations were averaged to produce a quantitative measure (i) of the agreement with SAS data used as restraint (training set) and (ii) for cross-validation (Figure 6).

The simultaneous SANS restraints across solvent deuterations produced universal improvements of reduced  $\chi$ -values, except for a marginal deterioration in SANS<sub>HD</sub><sup>70%</sup> (Figure 6, yellow background). Here, SANS<sub>DD</sub>- and SANS<sub>HD</sub>-restrained ensembles exhibited the largest margin of improvement, indicating significant conformational adjustments upon restraining the simulation to the data. In contrast, SAXS- and SANS<sub>DH</sub>-restrained simulations exhibit minor improvements across all data, indicating that these restraints merely eliminate the conformations that are inconsistent with experimental data. In terms of absolute reduced  $\chi$ -values, the SAXS<sub>HH</sub> and SANS<sub>DH</sub> structures are mutually compatible, as is evident from very small  $\chi$ -values in the SAXS<sub>HH</sub>/SANS<sub>DH</sub> block of Figure 6

(top left four blocks). In contrast, the SANS<sub>HD</sub> restraints result in poorer agreement with SAXS and SANS<sub>DH</sub>, suggesting a distinct conformation when only Unr-CSD1 was deuterated.

**Dependence of SUM Complex Configuration upon Deuteration of Unr-CSD1.** To explain the changes of reduced  $\chi$ -values as a function of protein deuteration, we compared the SUM conformation between different SAS-restrained ensembles in terms of their relative domain arrangements. Figure 7A shows conformations extracted from the final 2 ns of restrained and free MD trajectories as a function of the angle  $\theta$  between three centers of mass: of (i) the Sxl RRM2 domain, (ii) the Sxl RRM1 domain, and (iii) the Unr CSD1 domain. Representative conformations are shown in Figure 7B–E. We found that  $\theta$  values refined by SAXS, SANS<sub>DD</sub>, and SANS<sub>DH</sub> data lie between  $100$  and  $125^\circ$ , covering the range of angles also observed in free MD simulations (Figure 7A). Indeed, restraints from SAXS and SANS<sub>DH</sub> have only a small effect on  $\theta$ , leading to refined ensembles similar to the free simulation and in excellent agreement with the crystallographic conformation at  $\theta = 113^\circ$  (Figure 7A, black and red). Conformations refined against SANS<sub>DD</sub> are slightly pinched toward lower  $\theta$ , yet reasonably

consistent with the simulations restrained to SAXS and SANS<sub>DH</sub> (Figure 7A, pink).

The sole outlier is given by the SANS<sub>HD</sub>-optimized ensemble, whose  $\theta$ -distribution is shifted to far smaller angles compared to all the other conditions (Figure 7A, blue). Visually, this latter ensemble depicts a Unr CD1 domain at  $\theta \approx 90^\circ$ . The distinct conformation of the SANS<sub>HD</sub>-restrained ensemble is corroborated by  $C_\alpha$ -RMSD distributions, revealing the largest deviation from the reference crystallographic conformation (Figure S4). The numerical agreement in reduced  $\chi$ -values further suggests that the SANS<sub>HD</sub><sup>0%</sup> is uniquely responsible for this trend (Figure 6, bottom row), since the  $\chi$ -values of SANS<sub>HD</sub><sup>42%</sup> and SANS<sub>HD</sub><sup>70%</sup> are small and hardly change upon introduction of the SANS<sub>HD</sub> restraints.

## CONCLUSIONS

We established a method for the simultaneous refinement of biomolecular structures against SAXS and SANS data based on all-atom MD simulations. In this method, the MD force field complements the SAXS and SANS data by maintaining physically reasonable structures of low free energy. In other words, the physicochemical information in the force field is added to the low-information experimental data, thereby allowing the refinement of atomic models of many degrees of freedom without overfitting the model. We caution, however, that the current implementation based on a single simulation will not correctly reproduce the full diversity of systems that possess either distinct coexisting conformations or significant disorder, as for instance found for intrinsically disordered proteins. However, an extension of our protocols has recently been implemented to couple disordered systems to SAS data, with commitment to the maximum entropy principle. This will be published elsewhere.<sup>64</sup>

In line with our previous work on purely SAXS-based refinement,<sup>45,48,49</sup> all SAXS and SANS predictions presented here were based on explicit-solvent calculations, which take atomic models for the hydration layer and excluded solvent into account, thereby avoiding solvent-related fitting parameters. Indeed, the  $R_g$  analysis of CS showed that the calculations naturally account for hydration layer effects on the  $R_g$  detected by a Guinier analysis. As expected,  $R_g$  detected by SAXS was increased compared to the  $R_g$  of the bare protein owing to an increased hydration layer density compared to bulk solvent. In contrast, the  $R_g$  detected by SANS at 100% D<sub>2</sub>O was decreased, which is a consequence of (i) increased hydration layer density at (ii) overall negative contrast between protein and fully deuterated water. The modulations of  $R_g$  due to the hydration layer were slightly smaller in simulation than in experiment, which may be taken as an indication that the applied force field slightly underestimates protein–water interactions. However, additional studies are needed to confirm this trend.

We validated combined SAXS- and SANS-driven MD using two examples, with the protein citrate synthase and the protein/RNA complex Sxl–Unr–msl2. The stimulated opening of CS demonstrated that the SAXS and SANS data sets were mutually compatible, as shown by increased agreements with all experimental SAS curves regardless of restraints. On the other hand, SAS-based refinement of the SUM complex revealed the effect of deuteration on the optimized structure in several SANS contrast variation conditions. Specifically, the refinement simulations suggested that SUM takes a more compacted conformation between the three protein domains

when only Unr-CSD1 was deuterated, relative to all the other deuteration conditions. The mechanisms underlying this pinched structure with deuterated Unr-CSD1 remains unclear. We excluded relative deuteration of polar hydrogen bonds, as the polar H-binding sites in SUM are exposed to solvent and will likely exchange to match solvent deuteration. Hence, canonical hydrogen bonds would be modulated by the D<sub>2</sub>O in the buffer and not by perdeuteration of the solute, as observed here. The angular ranges of the SANS restraint also exclude regions below  $q < 0.8 \text{ nm}^{-1}$ , which eliminated the majority of aggregation effects observed between experiment and computed curves (see Figure S5). Therefore, we speculate that the modified conformation may instead be a consequence of modified hydrophobic effects, stemming from the altered carbon–deuterium bonds of deuterated CSD1. It is worth noting that repeating the measurements using SEC–SAS setups will yield additional information from the excluded lowest angle region, which may help to clarify the relationship between deuteration and structure.

In conclusion, we have expanded the capabilities of previously established SAXS-driven MD simulations to also incorporate SANS data collected at arbitrary D<sub>2</sub>O concentrations and arbitrary perdeuteration conditions of the solute. Because SANS provides subunit-specific structural information by means of contrast variation experiments, the new method enables MD simulations to validate and to refine the relative arrangements of subunits in large multisubunit complexes. In addition, as emphasized in this work, SAXS and SANS may provide independent structural information, hence allowing one to cross-validate the structure after refinement. We expect the method to be useful for deriving novel solution conformations of biomolecules or soft-matter systems consistent with experimental SAS data, and for providing hypotheses of structural alteration as a consequence of protein or solvent deuteration.

## METHODS

### SAXS and SANS Experiments of Citrate Synthase.

Experimental details for the eight SAXS/SANS citrate synthase (CS) scattering measurements are described below. To ease comparison, Table S1 is also provided according to 2017 community guidelines.<sup>65</sup> CS from porcine heart was obtained commercially from Sigma-Aldrich. The protein was dissolved in either D<sub>2</sub>O (99.9% D<sub>2</sub>O) or H<sub>2</sub>O buffer (50 mM TRIS, 50 mM NaCl, at pH 7.6) without further purification. The ligand oxaloacetate was added to one buffer solution at a concentration of 40 mM. Protein solutions in D<sub>2</sub>O buffer were dialyzed against the 100 excess D<sub>2</sub>O buffer volume to remove exchangeable protons.

SANS was measured of all CS solutions and corresponding buffers on SANS-1 operated by TUM and HZG at the MLZ, Garching.<sup>59,60</sup> Two concentration series of the protein without ligand (2.6, 5.3, 10.4, 51.2 mg/mL) and with ligand saturated (2.5, 4.9, 9.7, 47.4 mg/mL) were measured by SANS. All solutions were measured in 1 mm thick quartz cuvettes at room temperature. Two neutron wavelengths of 4.5 and 6 Å were used. Sample-to-detector distances and corresponding collimation distances were 2, 8, and 20 m. Buffer subtracted and concentration normalized SANS data of the protein solutions were linearly extrapolated to infinite dilution in order to determine the ligand-free and ligand-bound form factors of CS. Guinier analysis was carried out by fitting to the  $q$ -range 0.18–0.5 nm<sup>−1</sup>.

SAXS was measured on the BioSAXS beamline BM29<sup>57</sup> at the ESRF, Grenoble, using an X-ray wavelength of 0.992 Å. Protein concentration was 2.5 mg/mL in both the absence and presence of oxaloacetate. A sample volume of 75  $\mu\text{L}$  was purged continuously through a quartz capillary during a SAXS measurement on BM29, and 10 X-ray exposure frames were collected per sample. The X-ray exposure time on BM29 was 2 s per frame. In addition, to test if H<sub>2</sub>O versus D<sub>2</sub>O solvent could influence the degree of opening of CS, SAXS was measured on the SWING beamline at SOLEIL. The X-ray wavelength was 1.03 Å. Three different protein concentrations at 1.25, 2.5, and 5 mg mL<sup>-1</sup> were measured for hydrogenated and deuterated solvents in the absence and presence of oxaloacetate. In total, between 20 and 60 frames were recorded of each sample with an X-ray exposure time of 0.1 s per frame.

The individual SAXS frames were checked for the absence of radiation damage. No radiation damage was detected, and all frames were merged. The buffer was measured before and after each protein solution, and all data frames of the buffer were merged as well. The scattering contribution of the buffer was subtracted from the merged data sets of the protein solutions. The dilution series measured on SWING verified the absence of protein aggregation and interparticle interference, and the SAXS data of the different protein concentrations measured on SWING were finally merged by averaging replicate curves after concentration matching. Guinier analysis was carried out by fitting to the  $q$ -range 0.125–0.43 nm<sup>-1</sup>.

**Simulation Setup and Parameters.** *Simulation and Visualization Software.* Equilibration and free MD simulation were performed with GROMACS (version 5.12).<sup>66</sup> SAS-restrained simulations were conducted using an in-house modification of GROMACS-4.6.2,<sup>45,67</sup> augmented in this work to incorporate neutron-scattering and contrast-variation information. The implementation is available on the authors' Web site at <https://biophys.uni-saarland.de>. All biomolecular figures have been rendered in VMD.<sup>68</sup> Line-based graphs have been rendered in xmgrace and histograms in Gnuplot-5.

*System Preparation and Equilibration.* The crystal structures of the closed and open states of citrate synthase (CS) were taken from the Protein Data Bank (PDB codes 1CTS and 2CTS<sup>56</sup>), where the coenzyme A molecules in the closed structure were removed. The parametrization of citric acid in the Amber scheme was conducted as follows: atomic charges were fitted to quantum-mechanical potentials as calculated by Gaussian 09,<sup>69</sup> while bonded and van der Waals interaction parameters were assigned by Antechamber.<sup>70</sup> The resulting files in Amber format were transformed to GROMACS format using ACPYPE.<sup>71</sup> Protein interactions were described by the Amber99sb-ILDN force field,<sup>72</sup> and the TIP3P water model was used.<sup>73</sup> The structures were placed in a periodic rhombic dodecahedron box with a margin of 1.5 nm, solvated, and neutralized by the addition of two sodium ions. Energy minimization was done by steepest descent over 2000 steps. Subsequently, the system was equilibrated with position restraints on the backbone atoms for 10 ns. Electrostatic interactions were computed with the particle mesh Ewald method.<sup>74</sup> Dispersion interactions and short-range repulsion were described by a Lennard-Jones potential with a cutoff at 1 nm. Dispersion corrections for energy and pressure were applied. The temperature was controlled at 300 K with a stochastic dynamics integration scheme,<sup>75</sup> and the pressure was controlled at 1 bar with the weak coupling scheme.<sup>76</sup> An integration time step of 2 fs was used. The geometry of water

molecules was constrained with the SETTLE algorithm,<sup>77</sup> and the bond lengths of all other molecules were constrained with LINCS.<sup>78</sup>

The experimental structure of SUM was taken from the Protein Data Bank (PDB code 4QQB)<sup>51</sup> and placed in a periodic rhombic dodecahedron box with the primary dimension of 10.52 nm. The box was then solvated and neutralized to 150 mM NaCl. Water molecules intruding into the macromolecule were removed. Interactions of the biomolecule were described by the Amber14sb force field.<sup>73,79</sup> Energy minimization was done by steepest descent over 2500 steps. Subsequently, the system was equilibrated over 2 ns with gradually relaxing position restraints, first over the side chains and then over the backbones. The temperature was controlled at 300 K by velocity rescaling.<sup>80</sup> All other parameters were chosen identical to the CS simulations.

*SAS-Restrained MD.* For CS, SAXS data in the  $q$ -range [0.1, 3.4] nm<sup>-1</sup> and SANS data in [0.4, 3.4] nm<sup>-1</sup> was applied during SAS-restrained simulations and for fitting to experimental data. SUM simulations were sparsely restrained at every 0.4 nm<sup>-1</sup> to maintain  $\sim 1$  restraint per Shannon channel, using  $D_{\text{max}}$  of 7.2 to determine the width of a Shannon channel  $\pi/D_{\text{max}}$ . The  $q$ -range that removes the majority of aggregation contributions was decided by plotting the  $\chi_{\text{reduced}}$  agreement between unrestrained SAS predictions versus experiment (Figure S5), looking for minimum  $q$  that results in plateauing of  $\chi_{\text{reduced}}$  values. This resulted in  $q$ -ranges of [0.4, 4.0] nm<sup>-1</sup> and [0.8, 2.8] nm<sup>-1</sup> for SAXS and SANS, respectively, while maintaining parity between the restrained  $q$ -points. SAS restraints were gradually introduced between simulation times of 1 and 2 ns. The memory times  $\tau$  for on-the-fly SAS predictions were set to 300 ps for CS and to 200 ps for SUM, sufficient to capture solvent, side chain, and fast backbone motions. The force constant was scaled to 10 for both SAXS and SANS restraints for SUM. For CS, we used force constants of 3 and 7.5 for SAXS and SANS restraints, respectively. All restraints were weighted by  $1/\sigma_i^2$ , where  $\sigma_i$  is the overall uncertainty of the  $i$   $q$ -point.  $\sigma_i$  was computed by taking into account experimental statistical errors, computed statistical errors, and a systematic error modeled by an uncertainty of the solvent density.<sup>45</sup> Here, systematic uncertainties were modeled assuming an uncertainty of the solvent density of 1%.<sup>45</sup> The average electron density of the bulk solvent was corrected to 334 e nm<sup>-3</sup> to remove artifacts arising from the water model density.

Solvent atoms contributing to the SAS predictions were defined by a spatial envelope enclosing the biomolecule. For CS, we here used a spherical envelope with a radius of 6.27 nm. For SUM, a shaped envelope was created, such that the distance between the envelope and all protein atoms in a 50 ns trajectory was 7 Å (Figure 1). During SAS-restrained MD, the temperature was controlled at room temperature using a stochastic dynamics integration scheme ( $\tau = 1$  ps),<sup>75</sup> and the pressure was maintained at 1 bar with the Parrinello–Rahman barostat.<sup>81</sup> All other MD parameters were identical to the equilibration simulations.

The approximate symmetry of the CS was maintained during refinement simulations as follows: The center of mass (COM) of the two large (inner) subunits  $R_L$  was computed from the COM of the backbone atoms of residues 55–121, 178–239, and 273–401. The COM of the two small (outer) subunits  $R_S$  was computed from the COM of the backbone atoms of residues 41–54, 122–177, 240–272, and 402–426.

Then, the approximate symmetry was maintained by applying the harmonic potential  $V = k(\mathbf{R}_L - \mathbf{R}_S)^2/2$ , where  $k = 2000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  is the force constant.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00292.

Additional CS SAXS measurements at SWING, SOLEIL; comparison of apo-CS SAXS measurements at SOLEIL and ESRF;  $I(q)$  residuals and RMSDs of SAS-restrained CS simulations without symmetry restraints;  $C_\alpha$ -RMSD histogram of restrained SUM conformations; dependence of  $\chi_{\text{reduced}}$  on minimum  $q$  cutoff when fitting unrestrained SUM simulations versus experimental SAXS and SANS data (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: pchen@embl.de.

\*E-mail: jochen.hub@uni-saarland.de.

### ORCID

Po-chia Chen: 0000-0003-2272-2082

Andreas M. Stadler: 0000-0003-2272-5232

Jochen S. Hub: 0000-0001-7716-1767

### Author Contributions

<sup>†</sup>P.C. and R.S.: These authors contributed equally to this work.

### Funding

R.S. and J.S.H. were supported supported by the Deutsche Forschungsgemeinschaft (HU 1971/3-1, HU 1971/4-1). P.C. was supported by the EIPOD postdoctoral programme cofunded by the European Molecular Biology Laboratory (EMBL) and the Marie Curie Actions Cofund grant MSCA-COFUND-FP No. 664726. J.H. gratefully acknowledges the EMBL and the German Research Council (Deutsche Forschungsgemeinschaft, DFG) for support via an Emmy-Noether Fellowship and the Priority Programme SPP1935. This work was supported by DFG grant STA 1325/2-1 to A.M.S. C.L. acknowledges the support of the International Helmholtz Research School of Biophysics and Soft Matter (BioSoft).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We acknowledge the ESRF and SOLEIL for provision of synchrotron radiation facilities (proposal numbers ESRF, MX1493; and SOLEIL, 20181927). We thank Drs. Martha Brennich and Javier Pérez for assistance in using beamlines BM29 and SWING. This work is based upon experiments performed at the SANS-1 instrument operated by TUM and HZG at the Heinz Maier-Leibnitz Zentrum (MLZ), Garching, Germany (Proposal No. 7668).

## ■ REFERENCES

(1) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution. *Q. Rev. Biophys.* **2007**, *40*, 191–285.

(2) Bernadó, P.; Svergun, D. I. Structural Analysis of Intrinsically Disordered Proteins by Small-Angle X-Ray Scattering. *Mol. BioSyst.* **2012**, *8*, 151–167.

(3) Schneidman-Duhovny, D.; Kim, S. J.; Sali, A. Integrative Structural Modeling with Small Angle X-Ray Scattering Profiles. *BMC Struct. Biol.* **2012**, *12*, 17.

(4) Graewert, M. A.; Svergun, D. I. Impact and Progress in Small and Wide Angle X-Ray Scattering (SAXS and WAXS). *Curr. Opin. Struct. Biol.* **2013**, *23*, 748–754.

(5) Rambo, R. P.; Tainer, J. A. Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology. *Annu. Rev. Biophys.* **2013**, *42*, 415–441.

(6) Trehwella, J. Small-Angle Scattering and 3D Structure Interpretation. *Curr. Opin. Struct. Biol.* **2016**, *40*, 1–7.

(7) Blanchet, C. E.; Spilotros, A.; Schwemmer, F.; Graewert, M. A.; Kikhney, A.; Jeffries, C. M.; Franke, D.; Mark, D.; Zengerle, R.; Cipriani, F.; Fiedler, S.; Roessle, M.; Svergun, D. I. Versatile Sample Environments and Automation for Biological Solution X-Ray Scattering Experiments at the P12 Beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **2015**, *48*, 431–443.

(8) Gabel, F.; Bicout, D.; Lehnert, U.; Tehei, M.; Weik, M.; Zaccari, G. Protein dynamics studied by neutron scattering. *Q. Rev. Biophys.* **2002**, *35*, 327–367.

(9) Heller, W. T. Small-Angle Neutron Scattering and Contrast Variation: A Powerful Combination for Studying Biological Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 1213–1217.

(10) Whitten, A. E.; Trehwella, J. Small-angle scattering and neutron contrast variation for studying bio-molecular complexes; *Micro and Nano Technologies in Bioanalysis*; Foote, R., Lee, J., Eds.; Methods in Molecular Biology 544; Humana Press: 2009; pp 307–323..

(11) Johansen, D.; Jeffries, C. M.; Hammouda, B.; Trehwella, J.; Goldenberg, D. P. Effects of macromolecular crowding on an intrinsically disordered protein characterized by small-angle neutron scattering with contrast matching. *Biophys. J.* **2011**, *100*, 1120–1128.

(12) Sonntag, M.; Jagtap, P. K. A.; Simon, B.; Appavou, M.-S.; Geerlof, A.; Stehle, R.; Gabel, F.; Hennig, J.; Sattler, M. Segmental, Domain-Selective perdeuteration and Small-Angle Neutron Scattering for Structural Analysis of Multi-Domain Proteins. *Angew. Chem., Int. Ed.* **2017**, *56*, 9322–9325.

(13) Zhou, Z.; Li, Z.; Ren, Y.; Hillmyer, M. A.; Lodge, T. P. Micellar shape change and internal segregation induced by chemical modification of a tryptich block copolymer surfactant. *J. Am. Chem. Soc.* **2003**, *125*, 10182–10183.

(14) Dias-Mirandela, G.; Tamburrino, G.; Ivanović, M. T.; Strnad, F. M.; Byron, O.; Rasmussen, T.; Hoskisson, P. A.; Hub, J. S.; Zachariae, U.; Gabel, F.; Javelle, A. Merging in-solution X-ray and neutron scattering data allows fine structural analysis of membrane-protein detergent complexes. *J. Phys. Chem. Lett.* **2018**, *9*, 3910–3914.

(15) Whitten, A. E.; Jacques, D. A.; Hammouda, B.; Hanley, T.; King, G. F.; Guss, J. M.; Trehwella, J.; Langley, D. B. The Structure of the KinA-Sda Complex Suggests an Allosteric Mechanism of Histidine Kinase Inhibition. *J. Mol. Biol.* **2007**, *368*, 407–420.

(16) Lapinaite, A.; Simon, B.; Skjaerven, L.; Rakwalska-Bange, M.; Gabel, F.; Carlomagno, T. The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* **2013**, *502*, 519–523.

(17) Franklin, J. M.; Surampudi, L. N.; Ashbaugh, H. S.; Pozzo, D. C. Numerical Validation of IFT in the Analysis of Protein–Surfactant Complexes with SAXS and SANS. *Langmuir* **2012**, *28*, 12593–12600.

(18) Kim, H. S.; Martel, A.; Girard, E.; Moulin, M.; Härtlein, M.; Madern, D.; Blackledge, M.; Franzetti, B.; Gabel, F. SAXS/SANS on supercharged proteins reveals residue-specific modifications of the hydration shell on supercharged proteins reveals residue-specific modifications of the hydration shell. *Biophys. J.* **2016**, *110*, 2185–2194.

(19) Grishaev, A.; Tugarinov, V.; Kay, L. E.; Trehwella, J.; Bax, A. Refined Solution Structure of the 82-kDa Enzyme Malate Synthase G from Joint Nmr and Synchrotron SAXS Restraints. *J. Biomol. NMR* **2008**, *40*, 95–106.

- (20) Schwieters, C. D.; Suh, J.-Y.; Grishaev, A.; Ghirlando, R.; Takayama, Y.; Clore, G. M. Solution Structure of the 128 kDa Enzyme I Dimer from *Escherichia Coli* and Its 146 kDa Complex with HPr Using Residual Dipolar Couplings and Small- and Wide-Angle X-Ray Scattering. *J. Am. Chem. Soc.* **2010**, *132*, 13026–13045.
- (21) Hennig, J.; Wang, I.; Sonntag, M.; Gabel, F.; Sattler, M. Combining NMR and Small Angle X-Ray and Neutron Scattering in the Structural Analysis of a Ternary Protein-RNA Complex. *J. Biomol. NMR* **2013**, *56*, 17–30.
- (22) Karaca, E.; Bonvin, A. M. J. J. On the Usefulness of Ion-Mobility Mass Spectrometry and SAXS Data in Scoring Docking Decoys. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 683–694.
- (23) Göbl, C.; Madl, T.; Simon, B.; Sattler, M. NMR Approaches for Structural Analysis of Multidomain Proteins and Complexes in Solution. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *80*, 26–63.
- (24) Ibrahim, Z.; Martel, A.; Moulin, M.; Kim, H. S.; Härtle, M.; Franzetti, B.; Gabel, F. Time-resolved neutron scattering provides new insight into protein substrate processing by a AAA+ unfoldase. *Sci. Rep.* **2017**, *7*, 40948.
- (25) Hammel, M. Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). *Eur. Biophys. J.* **2012**, *41*, 789–799.
- (26) Hub, J. S. Interpreting solution X-ray scattering data using molecular simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 18–26.
- (27) Brunger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **1992**, *355*, 472.
- (28) Falkner, B.; Schröder, G. F. Cross-validation in cryo-EM-based structural modeling. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 8930–8935.
- (29) DiMaio, F.; Zhang, J.; Chiu, W.; Baker, D. Cryo-EM model validation using independent map reconstructions. *Protein Sci.* **2013**, *22*, 865–868.
- (30) Scheiner, S.; Čuma, M. Relative Stability of Hydrogen and Deuterium Bonds. *J. Am. Chem. Soc.* **1996**, *118*, 1511–1521.
- (31) Svergun, D.; Barberato, C.; Koch, M. H. J. CRYSOLO – a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
- (32) 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.
- (33) Poitevin, F.; Orland, H.; Doniach, S.; Koehl, P.; Delarue, M. AquaSAXS: A Web Server for Computation and Fitting of SAXS Profiles with Non-Uniformly Hydrated Atomic Models. *Nucleic Acids Res.* **2011**, *39*, W184–W189.
- (34) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. Accurate SAXS Profile Computation and Its Assessment by Contrast Variation Experiments. *Biophys. J.* **2013**, *105*, 962–974.
- (35) Chen, P.-c.; Hub, J. S. Structural Properties of Protein-Detergent Complexes from SAXS and MD Simulations. *J. Phys. Chem. Lett.* **2015**, *6*, 5116–5121.
- (36) Kim, H. S.; Gabel, F. Uniqueness of Models from Small-Angle Scattering Data: The Impact of a Hydration Shell and Complementary NMR Restraints. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2015**, *71*, 57–66.
- (37) Merzel, F.; Smith, J. C. Is the First Hydration Shell of Lysozyme of Higher Density than Bulk Water? *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5378–5383.
- (38) Park, S.; Bardhan, J. P.; Roux, B.; Makowski, L. Simulated X-Ray Scattering of Protein Solutions Using Explicit-Solvent Models. *J. Chem. Phys.* **2009**, *130*, 134114.
- (39) Köfinger, J.; Hummer, G. Atomic-Resolution Structural Information from Scattering Experiments on Macromolecules in Solution. *Phys. Rev. E* **2013**, *87*, No. 052712.
- (40) Chen, P.-c.; Hub, J. S. Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-Ray Scattering Data. *Biophys. J.* **2014**, *107*, 435–447.
- (41) Petoukhov, M. V.; Svergun, D. I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **2005**, *89*, 1237–1250.
- (42) 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.
- (43) Pelikan, M.; Hura, G.; Hammel, M. Structure and Flexibility within Proteins as Identified through Small Angle X-Ray Scattering. *Gen. Physiol. Biophys.* **2009**, *28*, 174–189.
- (44) Zheng, W.; Tekpinar, M. Accurate flexible fitting of high-resolution protein structures to small-angle x-ray scattering data using a coarse-grained model with implicit hydration shell. *Biophys. J.* **2011**, *101*, 2981–2991.
- (45) Chen, P.-c.; Hub, J. S. Interpretation of Solution X-Ray Scattering by Explicit-Solvent Molecular Dynamics. *Biophys. J.* **2015**, *108*, 2573–2584.
- (46) 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.
- (47) Kimanius, D.; Pettersson, I.; Schluckebier, G.; Lindahl, E.; Andersson, M. SAXS-guided metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 3491–3498.
- (48) Ivanović, M. T.; Bruetzel, L. K.; Lipfert, J.; Hub, J. S. Temperature-dependent atomic models of detergent micelles refined against small-angle X-ray scattering data. *Angew. Chem., Int. Ed.* **2018**, *57*, 5635–5639.
- (49) Shevchuk, R.; Hub, J. S. Bayesian refinement of protein structures and ensembles against SAXS data using molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005800.
- (50) Knight, C. J.; Hub, J. S. WAXSiS: A Web Server for the Calculation of SAXS/WAXS Curves Based on Explicit-Solvent Molecular Dynamics. *Nucleic Acids Res.* **2015**, *43*, W225.
- (51) Hennig, J.; Militti, C.; Popowicz, G. M.; Wang, I.; Sonntag, M.; Geerlof, A.; Gabel, F.; Gebauer, F.; Sattler, M. Structural Basis for the Assembly of the Sxl-Unr Translation Regulatory Complex. *Nature* **2014**, *515*, 287–290.
- (52) Rieping, W.; Habeck, M.; Nilges, M. Inferential Structure Determination. *Science* **2005**, *309*, 303–306.
- (53) Sears, V. F. Neutron scattering lengths and cross section. *Neutron News* **1992**, *3*, 26–37.
- (54) Breyton, C.; Gabel, F.; Lethier, M.; Flayhan, A.; Durand, G.; Jault, J.-M.; Juillan-Binard, C.; Imbert, L.; Moulin, M.; Ravaut, S.; Härtle, M.; Ebel, C. Small angle neutron scattering for the study of solubilised membrane proteins. *Eur. Phys. J. E* **2013**, *36*, 71.
- (55) Svergun, D. I.; Richard, S.; Koch, M. H. J.; 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.
- (56) Remington, S.; Wiegand, G.; Huber, R. Crystallographic Refinement and Atomic Models of Two Different Forms of Citrate Synthase at 2.7 and 1.7 Å Resolution. *J. Mol. Biol.* **1982**, *158*, 111–152.
- (57) Pernot, P.; Round, A.; Barrett, R.; De Maria Antolinos, A.; Gobbo, A.; Gordon, E.; Huet, J.; Kieffer, J.; Lentini, M.; Mattenet, M.; Morawe, C.; Mueller-Dieckmann, C.; Ohlsson, S.; Schmid, W.; Surr, J.; Theveneau, P.; Zerrad, L.; McSweeney, S. Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* **2013**, *20*, 660–664.
- (58) David, G.; Pérez, J. Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J. Appl. Crystallogr.* **2009**, *42*, 892–900.
- (59) Gilles, R.; Ostermann, A.; Schanzer, C.; Krimmer, B.; Petry, W. The concept of the new small-angle scattering instrument SANS-1 at the FRM-II. *Phys. B* **2006**, *385*, 1174–1176.
- (60) Mühlbauer, S.; Heinemann, A.; Wilhelm, A.; Karge, L.; Ostermann, A.; Defendi, I.; Schreyer, A.; Petry, W.; Gilles, R. The new small-angle neutron scattering instrument SANS-1 at MLZ—

characterization and first results. *Nucl. Instrum. Methods Phys. Res., Sect. A* **2016**, *832*, 297–305.

(61) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.

(62) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.

(63) Carlon, A.; Ravera, E.; Hennig, J.; Parigi, G.; Sattler, M.; Luchinat, C. Improved accuracy from joint X-ray and NMR refinement of a protein–RNA complex structure. *J. Am. Chem. Soc.* **2016**, *138*, 1601–1610.

(64) Hermann, M. R.; Hub, J. S. SAXS-restrained ensemble simulations of intrinsically disordered proteins with commitment to the principle of maximum entropy. Submitted to *J. Chem. Theory Comput.*

(65) Trewthella, J.; Duff, A. P.; Durand, D.; Gabel, F.; Guss, J. M.; Hendrickson, W. A.; Hura, G. L.; Jacques, D. A.; Kirby, N. M.; Kwan, A. H.; Pérez, J.; Pollack, L.; Ryan, T. M.; Sali, A.; Schneidman-Duhovny, D.; Schwede, T.; Svergun, D. I.; Sugiyama, M.; Tainer, J. A.; Vachette, P.; Westbrook, J.; Whitten, A. E. 2017 Publication Guidelines for Structural Modelling of Small-Angle Scattering Data from Biomolecules in Solution: An Update. *Acta Crystallogr. D* **2017**, *73*, 710–728.

(66) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(67) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(68) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(69) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 09*, revision D.01; Gaussian, Inc.: 2016.

(70) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

(71) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interface. *BMC Res. Notes* **2012**, *5*, 367.

(72) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins* **2010**, *78*, 1950–1958.

(73) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(74) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(75) Van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-Frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1988**, *1*, 173–185.

(76) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(77) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(78) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

(79) Gil-Ley, A.; Bottaro, S.; Bussi, G. Empirical Corrections to the Amber RNA Force Field with Target Metadynamics. *J. Chem. Theory Comput.* **2016**, *12*, 2790–2798.

(80) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.

(81) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.