

## Article

## Interpretation of Solution X-Ray Scattering by Explicit-Solvent Molecular Dynamics

Po-chia Chen<sup>1</sup> and Jochen S. Hub<sup>1,\*</sup><sup>1</sup>Institute for Microbiology and Genetics, Georg-August-University Göttingen, Göttingen, Lower Saxony, Germany

**ABSTRACT** Small- and wide-angle x-ray scattering (SWAXS) and molecular dynamics (MD) simulations are complementary approaches that probe conformational transitions of biomolecules in solution, even in a time-resolved manner. However, the structural interpretation of the scattering signals is challenging, while MD simulations frequently suffer from incomplete sampling or from a force-field bias. To combine the advantages of both techniques, we present a method that incorporates solution scattering data as a differentiable energetic restraint into explicit-solvent MD simulations, termed SWAXS-driven MD, with the aim to direct the simulation into conformations satisfying the experimental data. Because the calculations fully rely on explicit solvent, no fitting parameters associated with the solvation layer or excluded solvent are required, and the calculations remain valid at wide angles. The complementarity of SWAXS and MD is illustrated using three biological examples, namely a periplasmic binding protein, aspartate carbamoyltransferase, and a nuclear exportin. The examples suggest that SWAXS-driven MD is capable of refining structures against SWAXS data without foreknowledge of possible reaction paths. In turn, the SWAXS data accelerates conformational transitions in MD simulations and reduces the force-field bias.

## INTRODUCTION

Small- and wide-angle x-ray scattering (SWAXS) has been gaining popularity as a structural probe for biomolecules in solution (1,2). Whereas classical SAXS experiments detect the overall shape of the molecules and large-scale conformational transitions at resolutions down to 20 Å, WAXS is also sensitive to smaller rearrangements in tertiary and secondary structures (3). Hence, SWAXS is capable of detecting a wide range of structural transitions, triggered by processes such as ligand-binding (3–6), photodissociation (7), or photon absorption (8,9). Improvements in light sources and detectors have paved the way for further application of SWAXS in many applications such as structure determination (10–12) and molecular docking (13,14). Recent developments include the probing of transient structures in time-resolved measurements (7,8,15,16) and progress toward the characterization of heterogeneous ensembles (17–19).

While procedures for obtaining experimental (1) and calculated (20–26) scattering curves have steadily improved over the years, the structural interpretation of the curves has remained challenging (27). The number of independent data points contained in a SAXS curve is typically given by the number of independent Shannon channels, expressed as  $N_{\text{indep}} \approx q_{\text{max}} D_{\text{max}} / \pi$ , where  $q_{\text{max}}$  is the maximum scattering vector, and  $D_{\text{max}}$  the maximum diameter of the solute (1). These independent data points are insufficient to derive a molecular interpretation of the data. Instead, prior physical

knowledge needs to be added to avoid overfitting of the data. Prior knowledge is also required to interpret structural data from other techniques such as x-ray crystallography or NMR. However, because SWAXS curves contain much less information, more physical knowledge is required or, equivalently, a tighter prior distribution of possible conformations should be imposed. A possible strategy for the interpretation of data with limited information content is to restrain molecular dynamics (MD) simulations to conformations that are compatible with the experimental data, while the force field adds the required prior knowledge. This approach is routine for the refinement of structures against crystallographic or NMR data, or for the fitting of structures into electron microscopy maps (28–30).

Devising a related methodology for SWAXS, however, remains a technical challenge. Coupling of MD simulations to experimental SWAXS curves requires on-the-fly calculations of SWAXS curves from simulation frames. Scattering intensities at small angles are typically computed based on static structures and by using implicit solvent models. However, their applicability is limited when the internal structure of water and thermal fluctuations becomes important at wide angles (22,26,31). In addition, implicit solvent models require fitting of free parameters associated with the solvation layer and the excluded solvent, thereby reducing the amount of available information. These issues are being addressed by explicit-solvent approaches that are capable of accurately describing the solvation shell or excluded solvent (25,26,32). However, a direct coupling of explicit-solvent MD to SWAXS data has remained elusive.

Submitted October 7, 2014, and accepted for publication March 9, 2015.

\*Correspondence: [jhub@gwdg.de](mailto:jhub@gwdg.de)

Editor: Michael Sattler.

© 2015 by the Biophysical Society  
0006-3495/15/05/2573/12 \$2.00



Here, we close this gap and present a method to refine atomistic biomolecular structures against experimental SWAXS data. The calculations are fully based on explicit solvent models, thereby avoiding fitting of the solvation shell and excluded solvent, and allowing the interpretation of wide-angle data. We sample the conformational space using all-atom MD simulations, thus referring to the method as SWAXS-driven MD. Below, using three biomolecular examples, we illustrate how SWAXS-driven MD induces conformational transitions without previous knowledge of potential reaction paths. The derivation of conformations consistent with given experimental and simulated SWAXS data demonstrates the capabilities of this approach, and enables future extension toward wide-angle and time-resolved applications. An implementation of the algorithms presented here is available from the authors upon request.

## Theory

### Coupling potential

The coupling of the MD simulation to an (experimental) target SWAXS curve  $I_e(q)$  was implemented by a hybrid energy

$$E_{\text{hybrid}} = E_{\text{FF}} + E_{\text{SWAXS}}, \quad (1)$$

where  $E_{\text{FF}}$  denotes the energy from the MD force field, while  $E_{\text{SWAXS}}$  adds an energetic penalty if this simulation structure is incompatible with the target curve. Here, we applied two different functional forms for  $E_{\text{SWAXS}}$  to quantify the deviation between the experimental SWAXS curve and the SWAXS curve calculated from the simulation frames.

First, a nonweighted set of harmonic potentials on a logarithmic scale was applied,

$$E_{\text{SWAXS}}^{(\log)}(t) = \alpha(t)k_c \frac{k_B T}{n_q} \sum_{i=1}^{n_q} [\log I_c(q_i, t) - \log I_e(q_i)]^2, \quad (2)$$

where  $I_c$  is the SWAXS curve calculated on-the-fly from the simulation frames and  $n_q$  is the number of intensity points spread over the scattering vector  $q$ . The parameter  $k_c$  is a constant that determines the weight of the SWAXS data compared to the force field. The value  $\alpha(t)$  is a time-dependent function that allows a gradual introduction of the SWAXS-derived potential at the beginning of the simulation. The nonweighted coupling potential (Eq. 2) turned out to be numerically stable and to require little optimization of  $k_c$ , but it does not explicitly account for statistical uncertainties.

Therefore, second, to account for experimental and calculated uncertainties, we used

$$E_{\text{SWAXS}}^{(w)}(t) = \alpha(t)k_c \frac{k_B T}{n_q} \sum_{i=1}^{n_q} \frac{[I_c(q_i, t) - I_e(q_i)]^2}{\sigma^2(q_i)}. \quad (3)$$

Here, the error  $\sigma(q_i)$  accounts for 1) experimental errors  $\sigma_e$ , 2) statistical calculated errors  $\sigma_c$ , and 3) a systematic error

$\sigma_{\text{buf}}$  that originates from the uncertainty of the buffer density. Because those three sources of error are independent, we have

$$\sigma^2(q_i) = \sigma_e^2(q_i) + \sigma_c^2(q_i) + \sigma_{\text{buf}}^2(q_i). \quad (4)$$

### Calculation of SWAXS curves during MD simulations

SWAXS curves are typically reported after subtracting the intensity curve of the buffer. Accordingly, our algorithm requires on-the-fly calculations of the difference in scattering intensity between the solute simulation (system A) and a second simulation containing purely buffer (system B), that is,  $I_c(q, t) = I_A(q, t) - I_B(q)$ . Because we rely on explicit water models,  $I_c$  cannot be computed from an individual simulation frame. Instead, fluctuations of the solvent must be temporally averaged to compute a converged SWAXS curve, approximating a local ensemble of states in phase space. In addition, the curves represent an orientational average of the scattering intensities because the solutes are randomly oriented in the experiment. Following the nomenclature of Chen and Hub (26) and Park et al. (33), we thus calculated the net intensity at simulation time  $t$ , as

$$I_c(q, t) = \left\langle \left\langle |\tilde{A}(\mathbf{q})|^2 \right\rangle_{t;\tau}^{(\omega)} - \left\langle |\tilde{B}(\mathbf{q})|^2 \right\rangle_{\Omega}^{(\omega)} \right\rangle. \quad (5)$$

Here,  $\tilde{A}(\mathbf{q})$  and  $\tilde{B}(\mathbf{q})$  denote the Fourier transforms of the instantaneous electron densities of the systems A and B, respectively. The symbol  $\langle \cdot \rangle_{\Omega}$  is the orientational average, while the superscript  $\langle \cdot \rangle^{(\omega)}$  indicates a temporal average at a fixed solute orientation  $\omega$ . The temporal average over solute and solvent fluctuations was achieved through a weighted average over frames of previous simulation, using weights that decay exponentially into the past:

$$\left\langle |\tilde{A}(\mathbf{q})|^2 \right\rangle_{t;\tau}^{(\omega)} = \mathcal{N}(t)^{-1} \int_0^t |\tilde{A}(\mathbf{q}, t')|^2 e^{(t-t')/\tau} dt'. \quad (6)$$

Here,

$$\mathcal{N}(t) = \int_0^t \exp[(t-t')/\tau] dt'$$

is a normalization constant, and  $\mathcal{N}(t) \rightarrow \tau$  as  $t \gg \tau$ .

The average in Eq. 6 is thus dominated by recent conformations, where the memory time  $\tau$  determines the extent of fluctuations that are represented by  $I_c$ . This procedure permits dynamic updating of the calculated SWAXS intensity in response to structural transitions carried out in the simulation. In addition, the exponential weight allows for a memory-efficient implementation (see below). We used values of  $\tau$  between 1 and 2.5 ns to include solvent and protein side-chain fluctuations. In contrast, for the pure-buffer system (system B), all averages  $\langle \cdot \rangle^{(\omega)}$  were equally weighted.

The evaluation of Eq. 5 based on a spatial envelope around the solute (Fig. S1 in the Supporting Material), the approximation of Eq. 6 by a discrete sum, and the calculation of the uncertainties, is each described below and in Chen and Hub (26).

## MATERIALS AND METHODS

### Computational details

#### On-the-fly calculation of scattering curves

The net scattering intensity  $I_c(q, t)$  in Eq. 5 was computed following the methods described previously (26), which were found to yield excellent agreement to experimental SWAXS curves. Accordingly, we constructed a spatial envelope that encloses the protein with sufficient distance, such that the solvent at the envelope surface is bulklike. Calculating  $I_c(q, t)$  required two averaging steps: 1) orientational averages representing all solute rotations and (2) weighted averages over simulation frames representing conformation sampling (26,33),

$$I_c(q, t) = \langle D(\mathbf{q}, t) \rangle_{\Omega}, \quad (7)$$

$$D(\mathbf{q}, t) = \left| \langle \tilde{A}_i(\mathbf{q}) \rangle_{t;\tau}^{(\omega)} - \langle \tilde{B}_i(\mathbf{q}) \rangle \right|^2 + \left[ \left\langle \left| \tilde{A}_i(\mathbf{q}) \right|^2 \right\rangle_{t;\tau}^{(\omega)} - \left| \langle \tilde{A}_i(\mathbf{q}) \rangle_{t;\tau}^{(\omega)} \right|^2 \right] - \left[ \left\langle \left| \tilde{B}_i(\mathbf{q}) \right|^2 \right\rangle^{(\omega)} - \left| \langle \tilde{B}_i(\mathbf{q}) \rangle^{(\omega)} \right|^2 \right]. \quad (8)$$

The orientational average

$$\langle D(\mathbf{q}, t) \rangle_{\Omega} := (4\pi)^{-1} \int d\Omega_{\mathbf{q}} D(\mathbf{q}, t)$$

was evaluated numerically, using  $J = 1500$   $\mathbf{q}$ -vectors per absolute value of  $q$ , which were chosen by the spiral method. The conformational averages are denoted with  $\langle \cdot \rangle_{t;\tau}^{(\omega)}$ , indicating averages at time  $t$  with the memory time constant  $\tau$ , at fixed reference orientation  $\omega$  of the solute. The values  $\tilde{A}_i(\mathbf{q})$  and  $\tilde{B}_i(\mathbf{q})$  denote the scattering amplitudes from electron densities inside the envelope for the solute simulation system-A and the buffer simulation system-B, respectively. The pure buffer simulations were conducted before running SWAXS-driven MD. Envelopes for the three proteins studied are shown in Fig. S1.

The scattering amplitude of a single frame was given by

$$\tilde{A}_i(\mathbf{q}) = \sum_{j=1}^{N_A} f_j(q) e^{i\mathbf{q} \cdot \mathbf{r}_j}, \quad (9)$$

where  $N_A$  is the number of atoms within the envelope,  $f_j(q)$  is the atomic form factor, and  $\mathbf{r}_j$  is the position of atom  $j$ . The value  $\tilde{B}_i(\mathbf{q})$  was calculated analogously, but over atoms of the buffer system enclosed by the same envelope. The atomic form factors were approximated by  $f(q) = c + \sum_{k=1}^4 a_k \exp[-b_k(q/4\pi)^2]$ , where the values  $a_k$ ,  $b_k$ , and  $c$  are the Cromer-Mann parameters (34). Atomic form factors for water were corrected to account for electron-withdrawing effects (35). The density of the solvent was corrected to match the experimental value of  $334 \text{ e nm}^{-3}$  using the density correction scheme described in previous work (26).

### Exponential averaging

The exponentially weighted averages at time  $t$  in Eq. 6 were approximated by a discrete sum over previous simulation frames,

$$\langle X \rangle_{t;\tau} = \mathcal{N}_n^{-1} \sum_{k=0}^n e^{-k\delta t/\tau} X(t - k \delta t). \quad (10)$$

Here,  $t = n\delta t$  is this time and  $\delta t$  is an update interval, set at  $>0.5$  ps, in order to average over frames with statistically independent solvent configurations, where  $\mathcal{N}_n = \sum_{k=0}^n \exp[-k\delta t/\tau]$  is the normalization constant.

The weighted average is thus dominated by contributions within the previous  $\sim 2\tau$ , whereas older frames hardly contribute. Equation 10 was further rearranged into a cumulative formulation  $\langle X \rangle_{t;\tau} = \mathcal{N}_n^{-1} [X(t) + (\mathcal{N}_n - 1) \langle X \rangle_{t-\delta t;\tau}]$ , with  $\mathcal{N}_n = 1 + \exp(-\delta t/\tau) \mathcal{N}_{n-1}$ . This drastically reduces memory footprint of the algorithm, but also requires double precision to avoid accumulation of numerical errors. We note that the unweighted average adopted for the buffer system is recovered as  $\tau \rightarrow \infty$ .

### SWAXS-derived forces

MD simulations require forces and, hence, they require analytic derivatives of  $E_{\text{SWAXS}}$  with respect to atomic coordinates. In this work, we applied SWAXS-derived forces only to solute atoms. According to the potential form in Eq. 2, the force on atom  $k$  is

$$\mathbf{F}_k(t) = -\nabla_k E_{\text{SWAXS}}^{(\log)}(t) = -2 \frac{\alpha(t) k_c k_B T}{n_q} \sum_{i=1}^{n_q} \frac{\log I_c(q, t) - \log I_e(q)}{I_c(q)} \nabla_k I_c(q, t), \quad (11)$$

where  $\nabla_k$  denotes the gradient with respect to the position of atom  $k$ . A similar expression for the forces can be obtained if the uncertainty-weighted  $E^{(w)}_{\text{SWAXS}}$  is applied instead. Using Eqs. 7 and 8, as well as the fact that the derivatives of  $\tilde{B}_i(\mathbf{q})$  vanish, the gradients of the intensity are given by

$$\nabla_k I_c(q) = \frac{1}{4\pi} \int d\Omega_{\mathbf{q}} \nabla_k D(\mathbf{q}), \quad (12)$$

$$\nabla_k D(\mathbf{q}) = 2 \text{Re} \left[ \left\langle \tilde{A}_i(\mathbf{q}) \nabla_k \tilde{A}_i^*(\mathbf{q}) \right\rangle_{t;\tau}^{(\omega)} - \left\langle \nabla_k \tilde{A}_i^*(\mathbf{q}) \right\rangle_{t;\tau}^{(\omega)} \langle \tilde{B}_i(\mathbf{q}) \rangle^{(\omega)} \right]. \quad (13)$$

Here,  $\text{Re}[\cdot]$  denotes the real part. As with the intensity calculations (Eqs. 7 and 8), the forces were computed in the reference orientation  $\omega$  after superimposing the solute onto a reference structure. Subsequently, the  $\mathbf{F}_k$  were rotated into this orientation of the solute. The new averages in Eq. 13 impose significant memory requirements in the order of  $N_{\text{solute}} \times n_q \times J$ , where  $N_{\text{solute}}$  is the number of solute atoms. In practice, a reasonable choice for  $n_q$  is limited by the information content of the SWAXS curve, while  $J$  should be chosen according to 1) the maximum diameter of the envelope, and 2) the maximum scattering angle  $q$  (26,36). Depending on the size of the solute and the maximum  $q$ , several gigabytes of memory may be required.

We note that the SWAXS-derived forces are not energy-conservative because they depend not only on present but also on previous coordinates, which would lead to an energy drift in an NVE ensemble simulation. However, using a tight stochastic-dynamics temperature coupling scheme, we did not observe significant drifts in the temperature. In addition, the  $\mathbf{F}_k$

may impose a small net force or net torque on the entire solute. Because, in this article, we are only interested in internal motions of the solute, such net forces and torques were removed.

To make sure that  $I_c(q,t)$  is sufficiently converged before applying SWAXS-derived forces, the switch function  $\alpha(t)$  was set to zero at  $t < \tau$ , then switched on gradually following  $[1 - \cos(\pi(t - \tau)/\tau)]/2$ , and set to unity at  $t > 2\tau$ .

### Calculation of uncertainties

The statistical uncertainty  $\sigma_c(q)$  of the calculated intensities  $I_c(\mathbf{q})$  was computed from the SD of the set  $D(\mathbf{q}_i)$  ( $|\mathbf{q}_i| = q$ ,  $i = 1, \dots, J$ ), then divided by the square-root of the number  $N_{\text{indep}}(q)$  of independent  $D(\mathbf{q}_i)$  at a given  $q$  value. Note that, for small  $q$ ,  $D(\mathbf{q})$  is a smooth function of the direction of  $\mathbf{q}$  (at fixed  $q$ ), whereas  $D(\mathbf{q})$  becomes increasingly rough at increasing  $q$ . Hence,  $N_{\text{indep}}(q)$  increases with  $q$ . Here, we estimated  $N_{\text{indep}}(q)$  from the autocorrelation function of  $D(\mathbf{q}_i)$  on the solid angle in  $\mathbf{q}$  space. As expected (36),  $N_{\text{indep}}(q)$  was found to increase quadratically with  $q$ .

A small uncertainty of the buffer density  $\delta\rho_{\text{buf}}$  translates into a systematic error  $\sigma_{\text{buf}}(q)$  of the calculated intensities, which is dominant at small  $q$ . The error  $\sigma_{\text{buf}}(q)$  was computed as follows: The uncertainty  $\delta\rho_{\text{buf}}$  leads to errors in the scattering amplitudes of  $\delta A_i(\mathbf{q}) = \delta\rho_{\text{buf}} \mathcal{F}_s(\mathbf{q})/\rho_{\text{buf}}$ , where  $\mathcal{F}_s(\mathbf{q})$  denotes the Fourier transform of the solvent density within the envelope of the solute simulation, and  $\delta B_i(\mathbf{q}) = \delta\rho_{\text{buf}} \mathcal{F}_e(\mathbf{q})$ , where  $\mathcal{F}_e(\mathbf{q})$  is the Fourier transform of a unit density within the envelope. Inserting these relations into Eq. 8, we get

$$\delta D_{\text{buf}}(\mathbf{q}) = 2\delta\rho_{\text{buf}} \text{Re} \left[ \left\langle \left\langle \tilde{A}_i^*(\mathbf{q}) \right\rangle_{t;\tau}^{(\omega)} - \left\langle \tilde{B}_i^*(\mathbf{q}) \right\rangle \right\rangle \times \left( \rho_{\text{buf}}^{-1} \mathcal{F}_s(\mathbf{q}) - \mathcal{F}_e(\mathbf{q}) \right) \right]. \quad (14)$$

Then,  $\sigma_{\text{buf}}(q)$  is given by the spherical average of  $\delta D_{\text{buf}}(\mathbf{q})$ . In this study, we used  $\delta\rho_{\text{buf}} = 0.01\rho_{\text{buf}}$ .

### Choice of target SWAXS data and restrained scattering angles

For leucine-binding protein (LBP), target data was generated from unbiased simulations. ATCase (Aspartate carbamoyltransferase) experimental data was transcribed from Fetler et al. (37) as a buffer-subtracted curve without errors, and Chromosomal Maintenance 1, i.e., exporting 1 (CRM1) experimental data was made available by the group of R. Ficner (see Acknowl-

edgments) as raw sample and buffer curves. For SWAXS-driven MD, the experimental intensities were smoothed by a 20-point running average (see Supporting Details in the Supporting Material). Shannon information theory implies that the intensity of nearby scattering angles  $q$  is largely redundant (12,38). Thus, there is no need to compute and restrain the entire SWAXS curve during SWAXS-driven MD. Instead, the number of restrained  $q$  points  $n_q$  was approximately chosen according to the number of Shannon channels (see Table S2).

### Simulation details

Details on initial coordinates, simulation parameters, analysis, and rigid-body modeling are provided in the Supporting Material. The algorithms were implemented into an in-house version of GROMACS 4.6 (39), available from the authors upon request.

## RESULTS AND DISCUSSION

Applications of SWAXS-driven MD to three different proteins are presented below. As a test case, conformational transitions of LBP are presented using different functional forms of  $E_{\text{SWAXS}}$  (Eqs. 2 and 3), and the refined structures are compared to models derived by rigid-body modeling. For the LBP case only, MD simulations were coupled to theoretically computed SWAXS curves. In addition, using experimental SAXS data, solution states of ATCase and of the nuclear exportin CRM1 were derived. For those three proteins, we found that conformational transitions significantly affect the scattering intensities mainly in the SAXS but not in the far WAXS regime (Fig. S2). Hence, structures were here refined only using SAXS or near-WAXS data up to  $8 \text{ nm}^{-1}$ . However, because the explicit-solvent methods employed to compute intensities are accurate up to wide angles (26), we believe that SWAXS-driven MD is also suitable to interpret wide-angle data in a future study.

The three examples illustrate that SWAXS-driven MD is capable of refining structures against SWAXS data, which we consider as the primary application of the method. In turn, from a molecular-dynamics perspective, all three

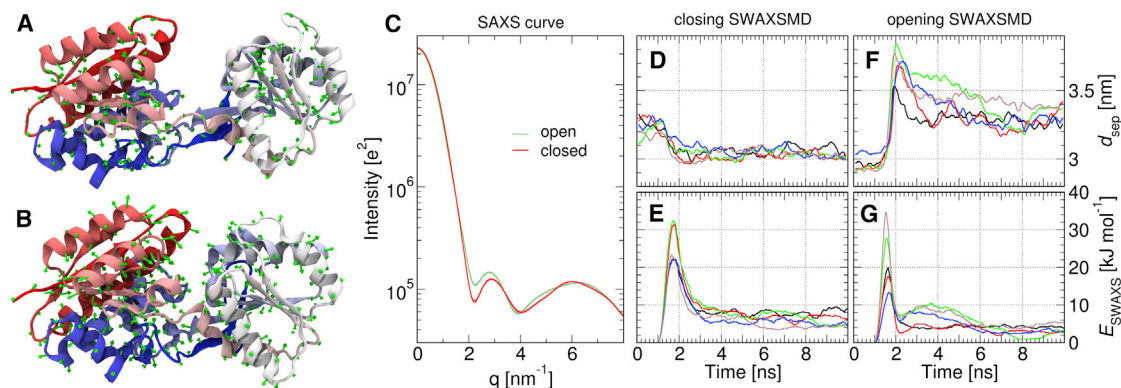


FIGURE 1 (A) Snapshot of LBP in open state and (B) near-closed state during SWAXS-driven MD. Imposed SWAXS-derived forces at this simulation time from  $C_{\alpha}$ -atoms (green arrows). (C) Computed SWAXS-curves for open (light green) and closed LBP (red), from 20 to 100-ns segments of equilibrium MD trajectories; these SWAXS curves were used as target curves for SWAXS-driven MD in (D–G). (D and F) Domain separation  $d_{\text{sep}}$  and (E and G)  $E_{\text{SWAXS}}$  over time for five closing (D and E) and five opening trajectories (F and G), using parameters of  $k_c = 1000$ ,  $\tau = 1$  ns. To see this figure in color, go online.

examples also highlight that SWAXS data help the simulations to overcome common limitations of MD due to limited sampling or inaccurate force fields.

## LBP

LBP is a bacterial periplasmic protein that contains a ligand recognition site positioned between two globular domains (Fig. 1, A and B). The transition from an open (apo) to a closed (holo) conformation occurs upon ligand binding and is observable in SAXS measurements (40). Because both functional states have been crystallized (41), we used LBP as a first test case to demonstrate the main principles of SWAXS-driven MD.

As a control, we first conducted multiple free simulations of LBP starting from the apo and holo crystal structures. These free simulations sampled their respective open and closed conformations (Fig. 2 A, green/red dots), but did not sample any complete transitions within 100 ns, regardless of the presence or absence of ligand at the binding interface (Fig. S3). This lack of sampling shows that MD simulations alone would be insufficient to predict the apo state of LBP given only the holo structure, or vice versa. Such problems are common, because most biologically relevant transitions occur on time scales inaccessible to MD.

## SWAXS-driven MD using a nonweighted SWAXS potential

In the following, we show rapid conformational transition of LBP by using information from computationally generated SWAXS curves. SWAXS curves of LBP were computed from apo and holo ensembles sampled in free simulations (Fig. 2, green/red dots), producing the target curves in Fig. 1 C. The radius of gyration  $R_{\text{gyr}}$  of these curves agree with SAXS measurements of a homolog LIVBP (78% identity), although it is significantly smaller than crystallographic values (Table 1). This suggests that our calculated LBP SWAXS curves are consistent with experimental solution conditions.

Multiple SWAXS-driven MD simulations were started from closed LBP and targeting the SWAXS curve of open LBP, and vice versa. SWAXS-driven MD based on a non-weighted coupling potential (Eq. 2) is presented in the following, and simulations based on a weighted potential (Eq. 3) are discussed further below. As shown in Fig. 1, D–G, coupling of simulations to target SWAXS curves of open and closed LBP induced rapid changes in the domain separation  $d_{\text{sep}}$  toward target values. At the end of the transition, the protein exhibited a reasonably low root-mean-square deviation (RMSD) to the cluster center of the open and closed ensembles, which were used to compute the target SWAXS curves (Fig. 2, B and C). In addition, and as expected,

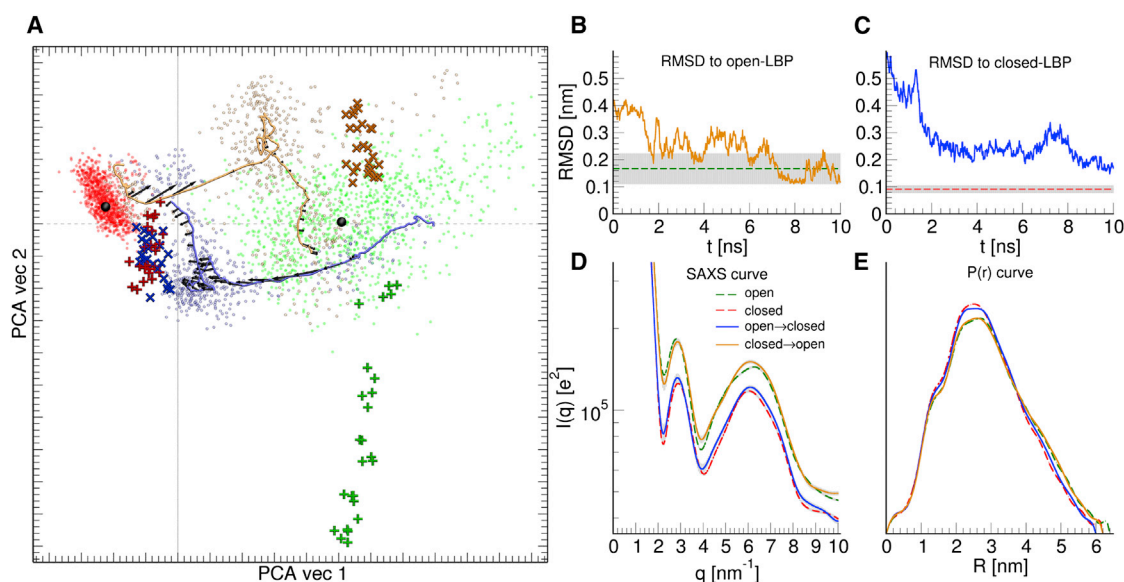


FIGURE 2 (A) PCA of unbiased simulations, SWAXS-driven MD, and rigid-body modeling. PCA vectors 1 and 2 approximately correspond to opening and twisting motions. (Red and green dots) Ensembles of unbiased simulations of closed (red) and open (green) LBP, which were used to compute the target SWAXS curves for SWAXS-driven MD (Fig. 1 C). (Black dots) Cluster center of closed and open ensembles used for the RMSD analysis in subplots (B and C). (Orange/blue dots) Conformations during SWAXS-driven MD in the opening (orange) and closing simulation (blue). (Orange and blue lines) To guide the eye, smoothed SWAXS-driven trajectories are shown. (Black arrows) Forces derived from  $E_{\text{SWAXS}}$ , mainly acting along the opening motion. (Plus and cross symbols) Conformations predicted by rigid-body modeling, using SASREF. Analogous colors to SWAXS-driven MD are adopted: (red) control SASREF runs of holo-LBP targeting holo spectra (holo  $\rightarrow$  holo); (green) apo-LBP targeting apo spectra (apo  $\rightarrow$  apo); (orange) holo  $\rightarrow$  apo; and (blue) apo  $\rightarrow$  holo. (B and C) SWAXS-driven trajectories in (A) plotted by RMSD to the cluster-center conformation. For reference, the RMSD of the target ensembles are indicated (green/red dashed lines) with errors (gray) shown in background. (D) Scattering-curves of SWAXS-driven trajectories at  $t = 10$  ns, with  $\tau = 1$  ns, as compared to the target curves from Fig. 1 C. (Orange and green curves) For clarity, these curves have been vertically offset. (E)  $P(r)$  curves of the spectra in (D) computed with GNOM (55). See the Supporting Material for details. To see this figure in color, go online.

**TABLE 1** Radius of gyration of apo and Phe-bound LBP

Structures	Crystallography	Simulation	LIVBP (SAXS)
Apo	2.80	2.28	2.33 ± 0.02
Holo	2.33	2.19	2.23 ± 0.02

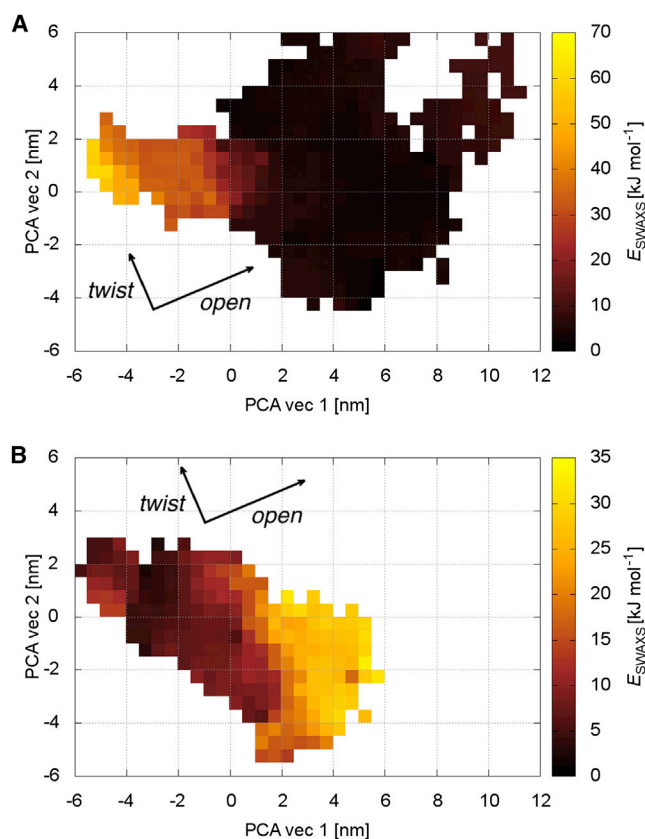
Apo- and holo-LBP structures from PDB:1USG and PDB:1USI, respectively. Simulation data reported from 10 × 50 ns replicates with SD. Calculated using CRY SOL default parameters (20). LIVBP data from Olah et al. (40).

the SWAXS curves and pair distribution functions  $P(r)$  computed from the final frames of the SWAXS-driven MD are in close agreement to the target curves (Fig. 2, D and E). A moderate  $E_{\text{SWAXS}}$  bias of 15–30 kJ/mol (Fig. 1, E and G) was sufficient to trigger rapid transitions, which was followed by an equilibration toward the target state within 10 ns (Fig. 2 A). Decomposing  $E_{\text{SWAXS}}$  into contributions from different scattering vectors  $q$  demonstrate that transitions were primarily driven by the difference signal at  $2 \text{ nm}^{-1}$  (Fig. S4). We note that during the opening transition (Fig. 1 F), some overcompensation occurred due to memory effects (Eq. 6), but such effects can be lessened by reducing the constant  $k_c$  at the cost of slower transitions (Fig. S5).

The driving forces for the conformational transitions are further analyzed Figs. 2 A and 3, which visualize transition pathways and  $E_{\text{SWAXS}}$ , respectively, along the eigenvectors from a principal component analysis (PCA). As shown in Fig. 2 A, we observed a further relaxation of LBP along a twisting motion after the equilibration of  $E_{\text{SWAXS}}$ , both after opening and closing transitions (orange and blue curves, respectively). This twisting motion is almost perpendicular to SWAXS-derived forces (arrows in Fig. 2), suggesting that the MD force field is primarily responsible for these twisting relaxations, and not the SWAXS data. The  $E_{\text{SWAXS}}$  landscape of LBP is analyzed in detail in Fig. 3, showing a large gradient of  $E_{\text{SWAXS}}$  along the opening/closing motion. In contrast,  $E_{\text{SWAXS}}$  is nearly invariant along the twisting motion, suggesting that only the MD force field can discriminate between states of different twist. Taken together, we find that the MD force field is required to complement the SWAXS signals in order to characterize the LBP solution states, while the SWAXS signals trigger the conformal transitions along the opening motion within the accessible simulation time.

#### Comparison to rigid-body modeling

The role of the applied force field was further investigated by comparing the SWAXS-driven MD to rigid-body modeling, as, for instance, implemented in the SASREF tool of the ATSAS software package (42). In contrast to atomistic MD simulations, SASREF employs rigid-body motions to minimize the intensity difference with the experiment, and further ranks complexes using a knowledge-based scoring function over residue-residue contacts. Details of our ATSAS modeling are given in the Supporting



**FIGURE 3** Energy landscape of  $E_{\text{SWAXS}}$  during SWAXS-driven simulations of LBP, binned along the first two PCA eigenvectors (see Fig. 2). The analysis demonstrates that the SWAXS data mainly encodes structural transitions along the opening motion. Transitions along the twisting motion are not coupled by SWAXS restraints, and hence require sampling under the influence of the MD force field. (A) Simulations targeting open LBP. (B) Simulations targeting closed LBP. Data was collected from the equilibration period (4–10 ns,  $t > 4\tau$ ). In order to sample the  $E_{\text{SWAXS}}$  landscape exhaustively, a total of 30 trajectories were simulated for each figure using coupling strengths of  $k_c = 100, 300,$  and  $1000$ .  $E_{\text{SWAXS}}$  was scaled to match  $E_{\text{SWAXS}}$  simulations using  $k_c = 1000$ . To see this figure in color, go online.

**Material.** To compare SWAXS-driven MD and ATSAS approaches, the final structures from rigid-body modeling are shown as plus-symbols in the PCA projection in Fig. 2 A, and their RMSDs to cluster-center structures are shown in Table 2.

We found consistently smaller RMSD of LBP structures derived by SWAXS-driven MD, as compared to rigid-body modeling. In particular, a large spread of ATSAS apo-LBP models along the second PCA vector is seen, loosely corresponding to the twisting motion found in SWAXS-driven MD (green plus symbols in Fig. 2 A). This suggests that ATSAS modeling lacks information necessary to refine a correct twist between the N- and C-terminal domains, despite additional distance constraints. In comparison, holo-LBP models found by ATSAS are of comparable accuracy to those found by SWAXS-driven MD, implying

**TABLE 2** LBP RMSD versus cluster-center structures, for conformations determined by ATSAS rigid-body modeling or SWAXS-driven MD

Description	RMSD <sub>apo</sub>	RMSD <sub>holo</sub>
ATSAS fitting		
Apo → apo	4.40 ± 1.09	6.29 ± 0.77
Holo → apo	3.13 ± 0.24	5.08 ± 0.16
Holo → holo	3.69 ± 0.22	2.34 ± 0.30
Apo → holo	3.93 ± 0.85	2.77 ± 1.58
SWAXS-driven MD		
Apo → apo	1.47 ± 0.34	4.43 ± 0.44
Holo → apo	2.23 ± 0.59	4.33 ± 0.65
Holo → holo	4.41 ± 0.62	1.38 ± 0.34
Apo → holo	3.35 ± 0.37	2.08 ± 0.61

ATSAS models were built using N- and C-terminal domains of the apo- and holo-LBP crystal structures as rigid bodies, and targeting the ensemble curves in Fig. 1 C. SWAXS-driven MD values obtained by averaging the last nanosecond of five independent simulations, conducted at force constant  $k_c = 1000$ . See the Supporting Material for details.

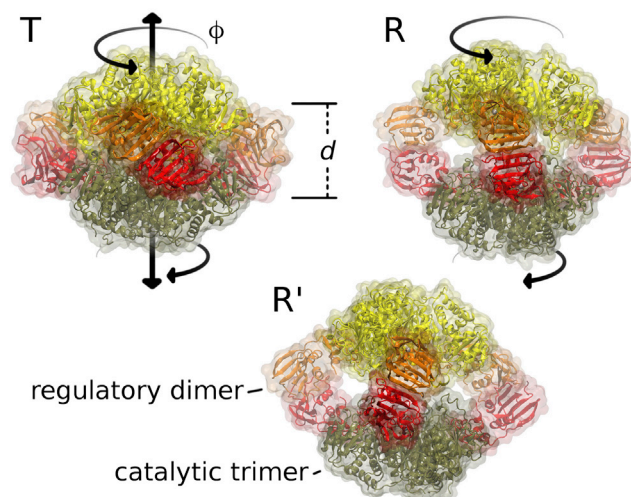
that both ATSAS- and SWAXS-driven MD possess sufficient information here. We further rationalize this difference from the  $E_{\text{SWAXS}}$  landscape of apo- and holo-LBP (Fig. 3). While the holo-LBP state is relatively well defined in  $E_{\text{SWAXS}}$ , the apo-LBP landscape is almost flat along the twisting motion. This explains the large potential distribution in ATSAS modeling, while SWAXS-driven MD additionally benefits from the information provided by the MD force field.

#### SWAXS-driven MD using an uncertainty-weighted SWAXS potential

In order to explicitly account for uncertainties of  $I(q)$ , we carried out SWAXS-driven MD of LBP using the weighted SWAXS potential of Eq. 3. Because we here coupled the simulations to calculated SWAXS curves, we assumed that the relative experimental error linearly increases with  $q$ , resembling the errors of the experimental data we used in a previous study (26). The uncertainties  $\sigma_e(q)$ ,  $\sigma_{\text{bur}}(q)$ , and  $\sigma_c(q)$  are shown in Fig. S6, demonstrating that, for the LBP model, the uncertainty of the buffer density dominates at low  $q$ , while the experimental errors dominates at high  $q$ . Hence, disregarding any of these error contributions would lead to a significant underestimate of the overall uncertainty for certain  $q$  ranges. The  $d_{\text{sep}}$ ,  $E_{\text{SWAXS}}$ , and the PCA analysis of multiple transitions are visualized in Fig. S7. These transitions resemble those from the nonweighted SWAXS-driven MD, suggesting that the functional form for  $E_{\text{SWAXS}}$  plays only a minor role for LBP. For simplicity, we therefore applied the nonweighted coupling scheme for the remainder of the article.

#### ATCase

ATCase is the first member of the pyrimidine synthesis pathway. It is composed of 12 subunits with three regulatory dimers surrounding a pair of catalytic trimers (Fig. 4).



**FIGURE 4** Snapshots of  $T$  (top left),  $R$  (top right), and  $R'$ -ATCase, viewed from one of the three regulatory dimers (orange and red; left, center and right domains) connecting the opposing catalytic trimers (yellow and tan; top and bottom domains). (Black arrows) Concerted motion connecting  $T \leftrightarrow R \leftrightarrow R'$ , involving an expansion along the trimer separation  $d$  and corotation of trimers in  $\phi$ . To see this figure in color, go online.

Two conformational states  $T$  and  $R$  corresponding to low and high activities are known, related by a concerted transition that is triggered by binding of a cosubstrate and further regulated by nucleotide binding (37,43). Whereas the crystallographic  $T$  state is compatible with SAXS data of  $T$ , Fetler and Vachette (37) and Svergun et al. (44) noticed that the solution state of  $R$  is more expanded as compared to the crystal structure of  $R$  (here termed  $R_{\text{cryst}}$ ). Hence, we used SWAXS-driven MD to characterize the solution state of  $R$ , based on published experimental SAXS data (37).

To begin, we conducted conventional simulations starting at  $T$  and  $R_{\text{cryst}}$ . SAXS curves computed from these simulations are compared to experimental data in Fig. 5 A, and conformational transitions are quantified in Fig. 5 B using the center-of-mass distance  $d$  as well as the rotational angle  $\phi$  between the two trimers (Fig. 4). Simulations starting at  $T$  were highly stable, and SAXS curves computed from the simulations favorably agree with available SAXS data (Fig. 5 A, red solid line versus red/orange crosses). In contrast, simulations that are restrained to  $R_{\text{cryst}}$  are incompatible with experimental SAXS data, suggesting that the crystal lattice stabilized a compact conformation (Fig. 5 A, purple solid line versus green/blue crosses). Indeed, ATCase in free simulations starting from  $R_{\text{cryst}}$  spontaneously expand, coupled to relative rotation of the two trimers (Fig. 5 B, green). That transition leads to an improved agreement with SAXS data of the  $R$  state at  $q \approx 0.8 \text{ nm}^{-1}$  (Fig. 5 A, green curve). However, none of our equilibrium simulations could explain the difference between the  $R$  and  $T$  curves at  $q \approx 2 \text{ nm}^{-1}$  (Fig. 5 A). These findings suggest that our free MD

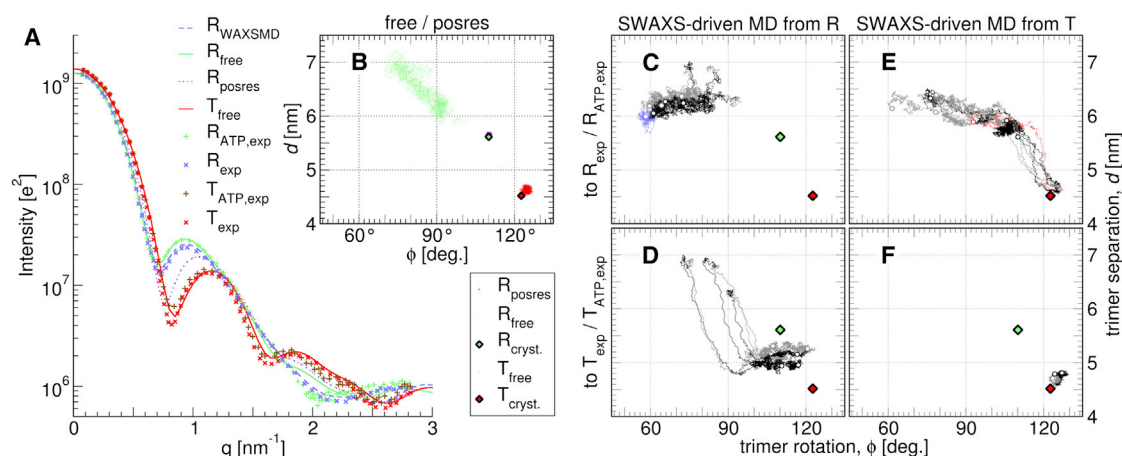


FIGURE 5 (A) SAXS curves of free ATCase simulations starting from crystallographic  $T_{\text{cryst}}$  (solid dark-red) or crystallographic  $R_{\text{cryst}}$  (solid light-green), and position-restrained to  $R_{\text{cryst}}$  (dotted purple). (Blue dashed line) SWAXS-driven MD targeting the experimental curve  $R_{\text{exp}}$ . The ensembles used to compute solid and dashed curves in (A) are shown as dots of the same colors in subplots (B) and (C). Experimental curves (37), shown as pluses/crosses, were fitted with function  $I_{\text{fit}} = fle + c$ , as described previously (26). Experimental curves are shown for both ATP-bound (green/tan pluses) and ATP-free states (blue/red crosses). (B) Unbiased and position-restrained ensembles of (A) plotted along relative catalytic trimer rotation  $\phi$  and separation  $d$ . Legend shown below the plot. (Diamonds) Starting crystallographic positions ( $T_{\text{cryst}}$  and  $R_{\text{cryst}}$ ); (dots) simulation frames. Note that the position-restrained  $R$  ensemble ( $R_{\text{posres}}$ , purple dots) is hardly visible behind the (green) diamond. (C–F) SWAXS-driven MD simulations; (C)  $6 \times R \rightarrow R$ , (D)  $6 \times R \rightarrow T$ , (E)  $6 \times T \rightarrow R$ , and (F)  $5 \times T \rightarrow T$ . (Gray lines) Simulations coupled to ATP-bound SAXS curves  $R_{\text{ATP,exp}}/T_{\text{ATP,exp}}$ . (Black lines) Simulations coupled to ATP-unbound SAXS curves  $R_{\text{exp}}/T_{\text{exp}}$ . (Blue dashed line) Ensemble from SWAXS-driven MD targeting  $R_{\text{exp}}$  used to compute the (blue dashed) SAXS curve in (A). (Red dashed line) Control simulation targeting a computed unbiased  $R$ -state curves. (Circles) Last frames. Trajectories span 20–23 ns, using parameters of  $\tau = 1$  ns and  $k_c = 500$ . To see this figure in color, go online.

simulations did not fully describe the solution  $R$  state, further emphasized by variations observed between individual simulation trajectories.

Results from SWAXS-driven MD simulations are presented in Fig. 5, C–F, starting from final frames of free  $R$  or free  $T$  simulations. Associated  $I(q)$  and  $P(r)$  curves are shown in Figs. S9 and S10. As controls,  $T \rightarrow T$  simulations are stable, and  $R \rightarrow T$  simulations reach a closed position near  $T$  (Fig. 5, D and F). Here we find that further closure is prevented by flexible loops of the catalytic trimer, whose rearrangement is necessary for reforming the  $T$ -interface. Because such details are not encoded in the SAXS signal, longer simulations or enhanced sampling techniques would be required to fully reach the  $T$  state.

Simulations coupled to the  $R$  curve rapidly open and consistently suggest that the trimers are on average separated by  $\sim 6.3$  nm in solution,  $\sim 0.7$  nm more as compared to  $R_{\text{cryst}}$  (Fig. 5, C and E). Compared to free simulations and previous modeling efforts, however, the SWAXS-driven simulations suggest ATCase conformations characterized by a further trimer rotation, corresponding to smaller  $\phi$  (Fig. 4,  $R'$ ). We note that after releasing the SWAXS-derived potential, those simulations partly returned to larger  $\phi$  angles, suggesting that the simulations were not yet able to accommodate a stable rotated structure within the simulation time of  $\sim 20$  ns (Fig. S9).

To identify the subset of experimental scattering data that encourages rotated  $R'$ -ATCase, we conducted additional

SWAXS-driven MD with restricted coupling ranges. The two main features of  $R$ -SAXS were coupled to individual simulations: 1)  $\sim 0.8$   $\text{nm}^{-1}$  or 2) at  $2$   $\text{nm}^{-1}$  (Fig. S8). Coupling only to the  $0.8$   $\text{nm}^{-1}$ -range fixed the trimer separation near  $6.3$  nm but induced no additional rotation, suggesting that this feature mainly encodes the trimer separation. In contrast, coupling to the feature at  $2.0$   $\text{nm}^{-1}$  was found to influence trimer rotations  $\phi$ , yielding  $R'$ -conformers similar to those found by the coupling to the entire  $q$  range. These findings further suggest that highly rotated conformers are significantly populated in the solution state of ATCase. More importantly, the analysis demonstrates how SWAXS-driven MD interprets features of a SWAXS curve in terms of specific collective modes of the biomolecule.

## Nuclear exportin 1

CRM1 is a nuclear exportin composed of 21 HEAT repeats arranged in a ringlike structure. Its structure-function relationships are complex: crystallographic studies yield a ring-open conformation for the apo form (45), but closed when in complex with cofactors or cargo (46,47). However, cryo-electron microscopy identifies different degrees of openness apo-CRM1 (45). Here, we revisited CRM1 by means of SWAXS-driven MD coupled to experimental SAXS patterns (48), to understand the solution state of apo CRM1.



To test whether MD simulations alone would provide a consistent solution ensemble of CRM1, we first simulated the open apo-CRM1 using two different force fields and salt concentrations (Fig. 6, A and C, Fig. S11). Surprisingly, we observe spontaneous ring-closure under CHARMM22\* (49) but ring-opening under the AMBER99sb force field (50). Simulations with 100 mM NaCl instead of only counterions exhibited slightly higher ring-open populations, but a significant force-field bias remains. These discrepancies suggest that the energy surface underlying CRM1 conformations is very shallow and, hence, highly sensitive to force-field differences. Furthermore, the SAXS patterns of these distributions suggest that neither distribution correctly reproduced the experimental data (Fig. 6 B).

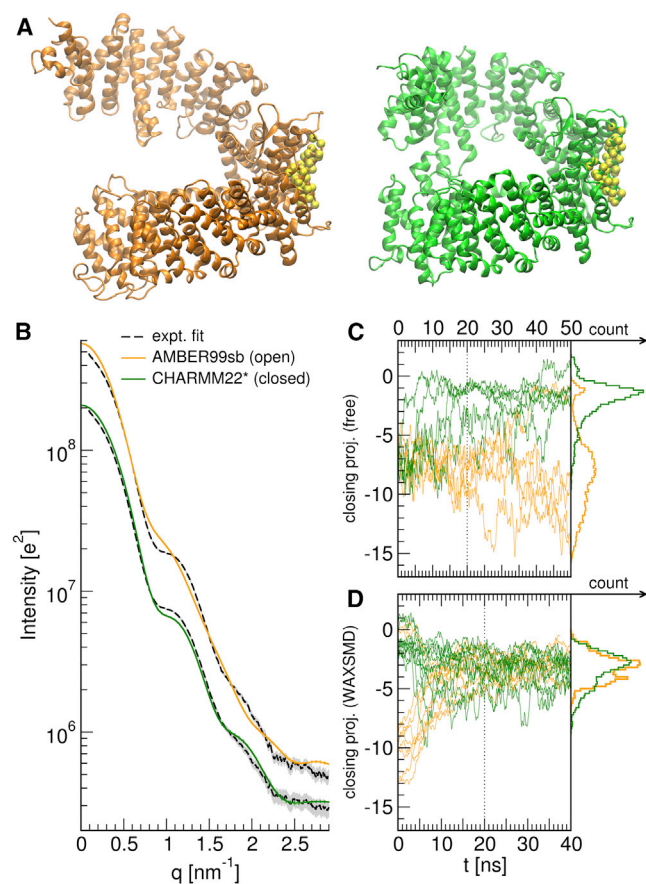


FIGURE 6 (A) Snapshots of CRM1 simulated with the AMBER99sb (orange, left) and CHARMM22\* (green, right) force fields, respectively, in a ring-open and ring-closed conformation. (B) Predicted SAXS curves of open (AMBER99sb, lighter orange) and closed (CHARMM22\*, darker green) CRM1 computed from the last 30 ns of the simulations (D); (dashed black line) fitted, smoothed experimental curve; (solid gray) error bars. (C) Display of 50-ns trajectories and last 30-ns histograms of free CRM1 in AMBER99sb (lighter orange) and CHARMM22\* (darker green), plotted according to the atomic difference vector between open (PDB:4FGV) and closed (PDB:3GJX) CRM1, the same metric as used in Monecke et al. (45) Significant force-field dependence is found. (D) A display of the 40-ns SWAXS-driven MD of CRM1, exhibiting no dependence on the force field. To see this figure in color, go online.

Fig. 6 D presents trajectories of SWAXS-driven MD started from the final frames of free MD simulations. Associated  $I(q)$  and  $P(r)$  curves are shown in Fig. S12. We find that SWAXS-driven MD suggests an intermediate average solution state of CRM1, which is more open than predicted by CHARMM22\* but more closed than predicted by AMBER99sb. The SAXS curve of the refined ensemble (Fig. S13 B) resembles that of the target curve. Only small deviations at very small angles remain, which we did not couple to the simulations, in order to exclude concentration-dependent experimental artifacts. We further verified the refinement by performing the  $\chi^2_{\text{free}}$  analysis suggested by Rambo and Tainer (12) over individual SWAXS-driven trajectories (Fig. S14). Although the SAXS curve for the total ensemble of 10 replica is in good agreement to experiment, we find significant variations in individual  $\chi^2_{\text{free}}$ , suggesting that the MD force field imposed tight restraints on possible conformations, and that the refinement is far from overfitting.

The alignment of ring-geometries to an intermediate conformation, however, does not exclude the possibility of a more heterogeneous open/closed solution ensemble. We therefore modeled the CRM1 ensemble as a linear combination of open and closed states from CHARMM22\* simulations, and evaluated the fits with our  $\chi^2$  metric (Eq. S1 in the Supporting Material) and the  $\chi^2_{\text{free}}$  metric (Fig. S1). We found that our  $\chi$ -metric reports an optimum ratio of 60% open/40% closed CRM1, while the  $\chi^2_{\text{free}}$ -metric reports ~75% open populations. Although the exact ratios differ, both are comparable to 2:1, suggested by cryo-electron microscopy experiments (45).

Thus, the ensemble-averaged SAXS data encodes the average solution structure, but it is in this case insufficient to determine the variance of the open/closed distribution. In order to interpret such heterogeneous ensembles, it will be highly interesting to couple multiple replicas of a simulation to one ensemble-averaged SWAXS curve, thereby also refining the relative weights of the replicas. Because that approach will require additional care to avoid overfitting, we leave that extension to a future study.

## CONCLUSIONS

Many proteins carry out their biological function via conformational transitions. High-resolution crystallographic structures are frequently only available for one state, such as a ground state or a holo state, whereas SWAXS data may be available for the complete conformational cycle. However, due to the large number of degrees of freedom, the interpretation of the SWAXS data in terms of conformational transitions would be ill defined without prior physical knowledge. Starting from related crystallographic structures, we here employed explicit-solvent MD simulations to refine structures against SWAXS curves. The simulations provide an accurate prior distribution of accessible

conformational states, and they are not subject to the restriction associated with rigid-body MD or normal mode fitting, which were previously used to refine structures against SAXS data (51,52). The commitment to explicit solvent and atomistic models of scattering also allows the accurate calculation of curves at wide angles (25,26,33). The explicit-solvent calculations require, apart from an overall scaling factor for the intensities, only one additional fitting parameter that accounts for uncertainties in the buffer subtraction. In particular, and in contrast to implicit solvent calculations, they do not require fitting of the solvation layer or excluded solvent, thus retaining the ability to differentiate small alterations in protein structure (26).

The three biomolecular systems presented here highlight the complementarity between SWAXS and MD techniques. As frequently found for biomolecular systems, we observed in all three examples that conventional MD simulations were insufficient to fully characterize the solution states of the proteins, due to limited sampling (LBP and ATCase) or force-field inaccuracies (CRM1). Incorporating SWAXS curves into the simulation reduced a possible force-field bias and accelerated conformational transitions without prior knowledge of possible reaction paths. In particular, the algorithm does not push the biomolecule along preselected collective modes such as PCA or normal modes, as would be common for many accelerated MD algorithms. Instead, the information of the SWAXS curve is merely added to the MD force field, and the simulation decides which collective motions are most appropriate to meet both experimental and physical restraints. In turn, during structure refinement, the MD simulation added physical knowledge to the SWAXS curves, thereby 1) restricting the refinement to physically realistic motions and 2) allowing further relaxation along collective motions that are not encoded in the SWAXS curve.

We validated SWAXS-driven MD using two different metrics of the coupling potential. First, a nonweighted metric on a logarithmic scale was applied (Eq. 2), which is conceptually simple, numerically stable and, hence, suitable from a practical perspective. In particular, this metric is rather insensitive to poor buffer matching, which is a common problem in SWAXS experiments, and it imposes reasonable statistical weights on small versus wide-angle data. Second, a more rigorous coupling metric was implemented by imposing the overall uncertainties as inverse weights, given by experimental  $\sigma_e$  and calculated  $\sigma_c$  statistical errors as well as a systematic error  $\sigma_{\text{buf}}$  from the buffer subtraction (Eqs. 3 and 4). We found incorporating  $\sigma_{\text{buf}}$  to be crucial for SWAXS-driven MD, in order to account for cases of poor buffer matching. Ignoring  $\sigma_{\text{buf}}$  and instead applying  $\sigma_e$  alone would impose spuriously high weights to small  $q$ , which frequently lead to numerical instabilities during SWAXS-driven simulations (not shown). The error  $\sigma_{\text{buf}}$ , on the contrary, dominates the overall error at small  $q$ , thereby reducing those weights to physically justified

values. As a consequence, the uncertainty-weighted SWAXS-driven MD simulations were numerically stable, capable of interpreting intensity variations over the entire  $q$  range. However, finding a rigorous procedure to distribute the weights along  $q$  may require more research.

Our implementation of SWAXS-driven MD uses certain heuristic parameters to achieve efficient sampling: coupling strength  $k_c$ , and memory time  $\tau$ . Their optimal choices depend upon the energy landscape of the system, and we noted above that adopting a constant  $k_c$  does risk temporary overcompensation, as observed for LBP (Fig. 1 F). We are investigating possibilities for adaptive coupling schemes that would avoid such artifacts arising from an ad hoc selection of these parameters. For instance, it will be interesting to test the effect of  $\tau$  on the heterogeneity of the refined ensemble, and to compare it with the protein flexibility as quantified from Kratky and Porod-Debye plots (53). In addition, the calculations shown here required us to fit experimental curves against calculated curves before starting the SWAXS-driven MD (Eq. S1 in the Supporting Material), in order to remove some of the uncertainty from poor buffer matching. In a future study, we instead aim to include the fitting step explicitly into the coupling scheme, if possible by using an inferential (Bayesian) approach (54). In addition, future efforts aim to render the methods shown here easily accessible to researchers that have little computational background. We note that SWAXS-driven MD is applicable to neutron scattering data, and it can be straightforwardly generalized to more heterogeneous ensembles by coupling multiple replicas of the biomolecule simultaneously to a target SWAXS curve. We expect SWAXS-driven MD presented here, complemented by such future developments, to become a useful protocol for biomolecular research.

## SUPPORTING MATERIAL

Supporting Materials and Methods, 14 figures, and two tables are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(15\)00403-8](http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00403-8).

## AUTHOR CONTRIBUTIONS

P.-c.C. and J.S.H. designed and performed research and wrote the article; P.-c.C. analyzed data.

## ACKNOWLEDGMENTS

We thank T. Monecke, A. Dickmanns, and R. Ficner for the SAXS curve of CRM1; B. Voss and H. Grubmüller for the AMBER99sb simulation system of CRM1; and A. Dickmanns and K. Atkovska for critically reading the article. Computing time by the North-German Supercomputing Alliance is gratefully acknowledged.

This study was supported by the Deutsche Forschungsgemeinschaft (grant No. HU 1971/1-1).

## SUPPORTING CITATIONS

References (56–70) appear in the Supporting Material.

## REFERENCES

- Putnam, C. D., M. Hammel, ..., J. A. Tainer. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40:191–285.
- Graewert, M. A., and D. I. Svergun. 2013. Impact and progress in small and wide angle x-ray scattering (SAXS and WAXS). *Curr. Opin. Struct. Biol.* 23:748–754.
- Rodi, D. J., S. Mandava, ..., R. F. Fischetti. 2007. Detection of functional ligand-binding events using synchrotron x-ray scattering. *J. Biomol. Screen.* 12:994–998.
- He, L., S. André, ..., H.-J. Gabius. 2003. Detection of ligand- and solvent-induced shape alterations of cell-growth-regulatory human lectin galectin-1 in solution by small angle neutron and x-ray scattering. *Biophys. J.* 85:511–524.
- Pietri, R., S. Zerbs, ..., L. M. Miller. 2012. Biophysical and structural characterization of a sequence-diverse set of solute-binding proteins for aromatic compounds. *J. Biol. Chem.* 287:23748–23756.
- Mould, A. P., E. J. H. Symonds, ..., M. J. Humphries. 2003. Structure of an integrin-ligand complex deduced from solution x-ray scattering and site-directed mutagenesis. *J. Biol. Chem.* 278:39993–39999.
- Cammarata, M., M. Levantino, ..., H. Ihee. 2008. Tracking the structural dynamics of proteins in solution using time-resolved wide-angle x-ray scattering. *Nat. Methods.* 5:881–886.
- Andersson, M., E. Malmerberg, ..., R. Neutze. 2009. Structural dynamics of light-driven proton pumps. *Structure.* 17:1265–1275.
- Ramachandran, P. L., J. E. Lovett, ..., J. J. van Thor. 2011. The short-lived signaling state of the photoactive yellow protein photoreceptor revealed by combined structural probes. *J. Am. Chem. Soc.* 133:9395–9404.
- Grishaev, A., J. Wu, ..., A. Bax. 2005. Refinement of multidomain protein structures by combination of solution small-angle x-ray scattering and NMR data. *J. Am. Chem. Soc.* 127:16621–16628.
- Schwieters, C. D., J.-Y. Suh, ..., G. M. Clore. 2010. 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.* 132:13026–13045.
- Rambo, R. P., and J. A. Tainer. 2013. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature.* 496:477–481.
- Schneidman-Duhovny, D., M. Hammel, and A. Sali. 2011. Macromolecular docking restrained by a small angle x-ray scattering profile. *J. Struct. Biol.* 173:461–471.
- Karaca, E., and A. M. J. J. Bonvin. 2013. On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys. *Acta Crystallogr. D Biol. Crystallogr.* 69:683–694.
- Neutze, R., and K. Moffat. 2012. Time-resolved structural studies at synchrotrons and x-ray free electron lasers: opportunities and challenges. *Curr. Opin. Struct. Biol.* 22:651–659.
- Levantino, M., A. Spilotros, ..., A. Cupane. 2012. The Monod-Wyman-Changeux allosteric model accounts for the quaternary transition dynamics in wild type and a recombinant mutant human hemoglobin. *Proc. Natl. Acad. Sci. USA.* 109:14894–14899.
- Bernadó, P., and D. I. Svergun. 2012. Structural analysis of intrinsically disordered proteins by small-angle x-ray scattering. *Mol. Biosyst.* 8:151–167.
- Yang, S., L. Blachowicz, ..., B. Roux. 2010. Multidomain assembled states of Hck tyrosine kinase in solution. *Proc. Natl. Acad. Sci. USA.* 107:15757–15762.
- Rózycki, B., Y. C. Kim, and G. Hummer. 2011. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure.* 19:109–116.
- Svergun, D., C. Barberato, and M. H. J. Koch. 1995. CRY SOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* 28:768–773.
- Merzel, F., and J. C. Smith. 2002. 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 Biol. Crystallogr.* 58:242–249.
- Bardhan, J., S. Park, and L. Makowski. 2009. SOFTWAXS: a computational tool for modeling wide-angle x-ray solution scattering from biomolecules. *J. Appl. Cryst.* 42:932–943.
- Grishaev, A., L. Guo, ..., A. Bax. 2010. Improved fitting of solution x-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. *J. Am. Chem. Soc.* 132:15484–15486.
- Poitevin, F., H. Orland, ..., M. Delarue. 2011. AQUASAXS: a web server for computation and fitting of SAXS profiles with non-uniformly hydrated atomic models. *Nucleic Acids Res.* 39:W184–W189.
- Köfinger, J., and G. Hummer. 2013. Atomic-resolution structural information from scattering experiments on macromolecules in solution. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 87:052712.
- Chen, P. C., and J. S. Hub. 2014. Validating solution ensembles from molecular dynamics simulation by wide-angle x-ray scattering data. *Biophys. J.* 107:435–447.
- Rambo, R. P., and J. A. Tainer. 2013. Super-resolution in solution x-ray scattering and its applications to structural systems biology. *Annu. Rev. Biophys.* 42:415–441.
- Brünger, A. T., J. Kuriyan, and M. Karplus. 1987. Crystallographic R factor refinement by molecular dynamics. *Science.* 235:458–460.
- Lange, O. F., N.-A. Lakomek, ..., B. L. de Groot. 2008. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science.* 320:1471–1475.
- Trabuco, L. G., E. Villa, ..., K. Schulten. 2008. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure.* 16:673–683.
- Moore, P. B. 2014. The effects of thermal disorder on the solution-scattering profiles of macromolecules. *Biophys. J.* 106:1489–1496.
- Oroguchi, T., and M. Ikeguchi. 2011. Effects of ionic strength on SAXS data for proteins revealed by molecular dynamics simulations. *J. Chem. Phys.* 134:025102.
- Park, S., J. P. Bardhan, ..., L. Makowski. 2009. Simulated x-ray scattering of protein solutions using explicit-solvent models. *J. Chem. Phys.* 130:134114.
- Doyle, P. A., and P. S. Turner. 1968. Relativistic Hartree-Fock x-ray and electron scattering factors. *Acta Crystallogr. A.* 24:390–397.
- Sorenson, J. M., G. Hura, ..., T. Head-Gordon. 2000. What can x-ray scattering tell us about the radial distribution functions of water? *J. Chem. Phys.* 113:9149–9161.
- Gumerov, N. A., K. Berlin, ..., R. Duraiswami. 2012. A hierarchical algorithm for fast Debye summation with applications to small angle scattering. *J. Comput. Chem.* 33:1981–1996.
- Fetler, L., and P. Vachette. 2001. The allosteric activator Mg-ATP modifies the quaternary structure of the R-state of *Escherichia coli* aspartate transcarbamylase without altering the T↔R equilibrium. *J. Mol. Biol.* 309:817–832.
- Vestergaard, B., and S. Hansen. 2006. Application of Bayesian analysis to indirect Fourier transformation in small-angle scattering. *J. Appl. Cryst.* 39:797–804.
- Hess, B., C. Kutzner, ..., E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.
- Olah, G. A., S. Trakhanov, ..., F. A. Quiocho. 1993. Leucine/isoleucine/valine-binding protein contracts upon binding of ligand. *J. Biol. Chem.* 268:16241–16247.

41. Magnusson, U., B. Salopek-Sondi, ..., S. L. Mowbray. 2004. X-ray structures of the leucine-binding protein illustrate conformational changes and the basis of ligand specificity. *J. Biol. Chem.* 279:8747–8752.
42. Petoukhov, M. V., and D. I. Svergun. 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* 89:1237–1250.
43. Macol, C. P., H. Tsuruta, ..., E. R. Kantrowitz. 2001. Direct structural evidence for a concerted allosteric transition in *Escherichia coli* aspartate transcarbamylase. *Nat. Struct. Biol.* 8:423–426.
44. Svergun, D. I., C. Barberato, ..., P. Vachette. 1997. Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state. *Proteins.* 27:110–117.
45. Monecke, T., D. Haselbach, ..., R. Ficner. 2013. Structural basis for cooperativity of CRM1 export complex formation. *Proc. Natl. Acad. Sci. USA.* 110:960–965.
46. Monecke, T., T. Güttler, ..., R. Ficner. 2009. Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science.* 324:1087–1091.
47. Dong, X., A. Biswas, ..., Y. M. Chook. 2009. Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature.* 458:1136–1141.
48. Dölker, N., C. E. Blanchet, ..., A. Dickmanns. 2013. Structural determinants and mechanism of mammalian CRM1 allostery. *Structure.* 21:1350–1360.
49. Piana, S., K. Lindorff-Larsen, and D. E. Shaw. 2011. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* 100:L47–L49.
50. Hornak, V., R. Abel, ..., C. Simmerling. 2006. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins.* 65:712–725.
51. Gorba, C., O. Miyashita, and F. Tama. 2008. Normal-mode flexible fitting of high-resolution structure of biological molecules toward one-dimensional low-resolution data. *Biophys. J.* 94:1589–1599.
52. Zheng, W., and M. Tekpinar. 2011. 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.* 101:2981–2991.
53. Rambo, R. P., and J. A. Tainer. 2011. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers.* 95:559–571.
54. Rieping, W., M. Habeck, and M. Nilges. 2005. Inferential structure determination. *Science.* 309:303–306.
55. Svergun, D. I. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* 25:495–503.
56. Wang, J., K. A. Stieglitz, ..., E. R. Kantrowitz. 2005. Structural basis for ordered substrate binding and cooperativity in aspartate transcarbamylase. *Proc. Natl. Acad. Sci. USA.* 102:8881–8886.
57. Gouaux, J. E., R. C. Stevens, and W. N. Lipscomb. 1990. Crystal structures of aspartate carbamoyltransferase ligated with phosphonoacetamide, malonate, and CTP or ATP at 2.8-Å resolution and neutral pH. *Biochemistry.* 29:7702–7715.
58. Mandell, D. J., E. A. Coutsias, and T. Kortemme. 2009. Sub-Ångstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods.* 6:551–552.
59. Brooks, B. R., C. L. Brooks, 3rd, ..., M. Karplus. 2009. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30:1545–1614.
60. Bjelkmar, P., P. Larsson, ..., E. Lindahl. 2010. Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* 6:459–466.
61. Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
62. Essmann, U., L. Perera, ..., L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.
63. Miyamoto, S., and P. A. Kollman. 1992. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13:952–962.
64. Hess, B. 2008. P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 4:116–122.
65. Feenstra, K. A., B. Hess, and H. J. C. Berendsen. 1999. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* 20:786–798.
66. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.
67. Berendsen, H. J. C., J. P. M. Postma, ..., J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
68. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
69. Hoover, W. G. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A.* 31:1695–1697.
70. van Gunsteren, W. F., and H. J. C. Berendsen. 1988. A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* 1:173–185.

# Supporting Information for: Interpretation of solution X-ray scattering by explicit-solvent molecular dynamics

Po-chia Chen<sup>1</sup> and Jochen S. Hub<sup>1\*</sup>

<sup>1</sup> Institute for Microbiology and Genetics  
Georg-August-University Göttingen  
Justus-von-Liebig-Weg 11  
37077 Göttingen, Germany

\* E-mail: jhub@gwdg.de; phone: +49 551 39-14189

## Simulation Details

### Choice and preparation of source coordinates

Apo and holo structure of LBP were taken from protein data bank (PDB codes 1USG and 1USI, respectively) (1).

The CTP-bound ATCase from [PDB code 1ZA1 (2)] and ATP-bound ATCase from [PDB code 7AT1 (3)] are commonly accepted to represent the crystallographic *T* and *R* states, respectively. Bound ATP and CTP were retained during all simulations, but the substrate mimic PALA was removed from the *R* coordinates. Although it is necessary *in-vivo* to form the *R* state, we observed no collapse back to *T* during 50 ns-simulations (see main text).

CRM1 simulations were based on the ring-open structure from *chaetomium thermophilum* CRM1 [ctCRM1, PDB code 4FGV (4)], as an appropriate model for mammalian CRM1 in the absence of binding partners (40% sequence identity). ctCRM1 represents the most complete structure and also retains a C-terminal helix stretching across the ring. In comparison, the structure of mouse CRM1 (PDB code 3GJX) (5) was not suitable due to non-trivial rearrangement upon ranGTP binding, while the structure of human CRM1 (PDB code 3GB8) (6) lacks several long loops and the N-terminal HEAT repeat and would require speculative modeling. Missing loops near the N-terminus of ctCRM1 were modelled using the kinematic loop modeling module from the Rosetta3.4 suite (7).

### Simulation setup and parameters

Equilibrium simulations were conducted using the GROMACS 4.6 simulation software (8). SWAXS-driven MD simulations were conducted using our in-house modification of GROMACS 4.6. Hydrogen atoms were added with the pdb2gmx software (8), keeping co-crystallised water molecules and the ligands noted above. The proteins were solvated in a rhombic-dodecahedron box with 15~10 Å buffer regions, and ionised to 100 mM NaCl, if not stated otherwise. CRM1 simulations without salt were neutralized by Na<sup>+</sup> counterions. Parameters for LBP and ATCase simulations were taken from the CHARMM27

and CHARMM22\* force fields, respectively (9–11). CRM1 was simulated either with the Amber99sb or with the CHARMM22\* force field, as noted in the main text (12). Water was described by the TIP3P water model (13).

Electrostatic interactions were computed using the particle-mesh Ewald algorithm (14). Short-range dispersive interactions were modeled by a Lennard-Jones potential using dual-range cut-off between 10 and 12 Å using force-switching. Water molecules were kept rigid using the SETTLE algorithm (15), and all other bonds were constrained using P-LINCS (16). Hydrogen bonds were described by virtual interaction sites, allowing a time step of 4 fs (17). The pressure was kept at 1 bar using the Berendsen barostat during initial equilibrations and the Parrinello-Rahman barostat during equilibrium simulations (time constant 1 ps) (18, 19). Likewise, the temperature was controlled at 298.15 K using velocity-rescaling during equilibration and using Nosé-Hoover thermostat during equilibrium simulations (time constant 5 ps) (20, 21). Protein and non-protein atoms have been coupled to separate thermostats. Stochastic dynamics integrators (22) were used during SWAXS-driven MD simulations because SWAXS-driven MD does not conserve energy, which might otherwise results in gradual heating of the system.

## Free (unbiased) simulations

Each system was equilibrated over 2 ns using stepwise decreasing restraints to relax the solute, first releasing the sidechains and then the backbones. From each equilibrated system, a number of free (*i.e.*, not SWAXS-driven) simulation were conducted, as listed in Table S1. SWAXS curves from equilibrium simulations were computed as explained previously (23). The principal component analysis (PCA) for LPB was conducted on the C<sub>α</sub> atoms, after combining free 100-ns simulations of apo and phenylalanine-bound LBP.

## SWAXS-driven simulations

Target experimental patterns for ATCases have been transcribed from Fetler *et al.* (24), while the target data for ctCRM1 were provided by the Ralf Ficner group. In the latter case, sample and buffer intensities were obtained to conduct and verify background matching in-house. To reduce the influence of noise upon WAXSMD simulations, the buffer-subtracted curve was passed through a 20-point running average. This smoothed curve was used in SWAXS-driven MD. For modeling the solution ensemble of CRM1 as a combination of open and closed states, the unsmoothed curve was adopted, allowing for direct comparison between our *chi*-metric and  $\chi_{free}^2$  validations according to Rambo and Tainer (25). In this work, we adopt convention  $I_{\text{sample}} - I_{\text{buffer}}$  for buffer subtraction.

Before coupling the simulations to an experimental SWAXS curve, the experimental curves were fitted to a SWAXS curve calculated from a short equilibrium simulation. Accordingly, the experimental curve was fitted by minimizing

$$\chi^2 = \sum_i [\log I_c(q_i) - (f \log I_e(q_i) + c)]^2, \quad (\text{S1})$$

where  $f$  adjusts the arbitrary overall scale of scattering curves, and  $c$  is an offset that approximately absorbs experimental uncertainties due to the buffer subtraction. Fitting only these two parameters, we

recently found excellent agreement between calculated and experimental WAXS curves (23). Here, the fit was conducted on a log scale, compatible to the coupling potential  $E_{\text{SWAXS}}$  in eq. 2.

Parameters applied for SWAXS-driven MD are listed in Table S1. A second simulation for ATCase  $R \rightarrow R$  was conducted at a higher  $k_c$  to resolve the discrepancy at  $2 \text{ nm}^{-1}$ , producing the blue ensemble in the main text and Fig. S9. As noted the main text, the simulation frames were superimposed onto a reference structure in reference orientation ( $\omega$ ) before evaluating the averages in eqs. 7, 8, and 13 (in main text), using a least-square fit on the  $C_\alpha$  atoms. For each protein, the same reference structure was used for all SWAXS-driven simulations.

The spatial envelope for LBP was constructed around simulation frames after combining 100-ns trajectories of apo and phenylalanine-bound LBP, keeping a distance of 0.7 nm between protein and envelope. The envelope was constructed as explained previously (23). Because the maximum extent of CRM1 and ATCase were not known prior to SWAXS-driven simulation, we assigned spherical envelopes with radii 5.2 and 8.8 nm, respectively, based on free simulations and solvation layer thickness of 0.7 nm. A snapshot of each system with its envelope are shown in Figure S1. The same envelope was shared across all SWAXS calculations of the same system.

## Rigid-body modeling

Rigid body modeling was conducted with SASREF(26) starting either from apo- or holo- LBP crystal structures, and targeting the computed SAXS curve of either open or closed LBP ensembles in Fig. 1. This resulted in four sets of models, colored as plus symbols in Fig. 2). The N- and C-terminal domains of LBP were defined as follows: the N-terminal domain spanning D1-L120 and P249-K327, and the C-terminal domain spanning D121-M248 and G332-K346. Four linker residues were removed between K327 and G332. To constrain the possible quaternary arrangements, constraints at 3, 4, 5, 6 and 7 Å (in steps of 1 Å) were imposed between L120-D121 and P249-M248. Five replicates were conducted for each constraint value, resulting in a total of 25 modeling attempts per structure-SAXS combination.

## P(r) calculations with GNOM

The program GNOM (27) was used to back-calculate the pair distribution function  $P(r)$  from simulated and experimental curves. For each protein and curve, a range of maximum extent  $R_{max}$  values is chosen to compute candidate  $P(r)$  curves. The curve with the highest fitness score calculated by GNOM was chosen, subject to the condition that it did not exhibit a large region of  $\sim 0$  density. These resultant  $R_{max}$  values are also reported in Table S2. In order to maintain comparability with experimental data, the maximum- $q$  of the simulated SWAXS curves considered for  $P(r)$  calculations for ATCase and ctCRM1 were set a  $5 \text{ nm}^{-1}$ . The cut-off for LBP was set at  $10 \text{ nm}^{-1}$ .

## Computational resource usage

All simulations were conducted on the GWDG cluster at the Georg-August University Göttingen, and at the North-German Supercomputing Alliance (HLRN) cluster in Hanover using between 16 to 192 CPU cores depending on architecture and system size. SWAXS-driven MD simulations take approximately

10% longer than conventional MD simulations. According to GROMACS internal cycle-accounting, the SWAXS-related calculations took (relative to total computing time)  $7.6 \pm 1.4\%$  during LBP,  $11.7 \pm 2.2\%$  during ATCase, and  $9.9 \pm 2.0\%$  during ctCRM1 simulations.

## Supporting References

1. Magnusson, U., B. Salopek-Sondi, L. A. Luck, and S. L. Mowbray, 2004. X-ray Structures of the Leucine-binding Protein Illustrate Conformational Changes and the Basis of Ligand Specificity. *J. Biol. Chem.* 279:8747–8752.
2. Wang, J., K. A. Stieglitz, J. P. Cardia, and E. R. Kantrowitz, 2005. Structural basis for ordered substrate binding and cooperativity in aspartate transcarbamoylase. *PNAS* 102:8881–8886.
3. Gouaux, J. E., R. C. Stevens, and W. N. Lipscomb, 1990. Crystal structures of aspartate carbamoyltransferase ligated with phosphonoacetamide, malonate, and CTP or ATP at 2.8-Å resolution and neutral pH. *Biochemistry* 29:7702–7715.
4. Monecke, T., D. Haselbach, B. Voß, A. Russek, P. Neumann, E. Thomson, E. Hurt, U. Zachariae, H. Stark, H. Grubmüller, A. Dickmanns, and R. Ficner, 2013. Structural basis for cooperativity of CRM1 export complex formation. *Proc. Natl. Acad. Sci. USA* 110:960–965.
5. Monecke, T., T. Güttler, P. Neumann, A. Dickmanns, D. Görlich, and R. Ficner, 2009. Crystal Structure of the Nuclear Export Receptor CRM1 in Complex with Snurportin1 and RanGTP. *Science* 324:1087–1091.
6. Dong, X., A. Biswas, K. E. Süel, L. K. Jackson, R. Martinez, H. Gu, and Y. M. Chook, 2009. Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* 458:1136–1141.
7. Mandell, D. J., E. A. Coutsias, and T. Kortemme, 2009. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* 6:551–552.
8. Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl, 2008. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 4:435–447.
9. Brooks, B. R., C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, 2009. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* 30:1545–1614.
10. Piana, S., K. Lindorff-Larsen, and D. E. Shaw, 2011. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* 100:L47–L49.
11. Bjelkmar, P., P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl, 2010. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* 6:459–466.
12. Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65:712–725.



13. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
14. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.
15. Miyamoto, S., and P. A. Kollman, 1992. SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13:952–962.
16. Hess, B., 2008. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* 4:116–122.
17. Feenstra, K. A., B. Hess, and H. J. C. Berendsen, 1999. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* 20:786–798.
18. Parrinello, M., and A. Rahman, 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.
19. Berendsen, H. J. C., J. P. M. Postma, W. F. v. Gunsteren, A. DiNola, and J. R. Haak, 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
20. Bussi, G., D. Donadio, and M. Parrinello, 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
21. Hoover, W. G., 1985. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* 31:1695–1697.
22. Van Gunsteren, W. F., and H. J. C. Berendsen, 1988. A Leap-frog Algorithm for Stochastic Dynamics. *Mol. Simul.* 1:173–185.
23. Chen, P.-c., and J. S. Hub, 2014. Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data. *Biophys. J.* 107:435–447.
24. Fetler, L., and P. Vachette, 2001. The Allosteric activator Mg-ATP Modifies the Quaternary Structure of the R-state of Escherichia coli Aspartate Transcarbamylase Without Altering the T $\leftrightarrow$ R Equilibrium. *J. Mol. Biol.* 309:817–832.
25. Rambo, R. P., and J. A. Tainer, 2013. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496:477–481.
26. Petoukhov, M. V., and D. I. Svergun, 2005. Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data. *Biophys. J.* 89:1237–1250.
27. Svergun, D. I., 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 25:495–503.

## Supplementary Tables

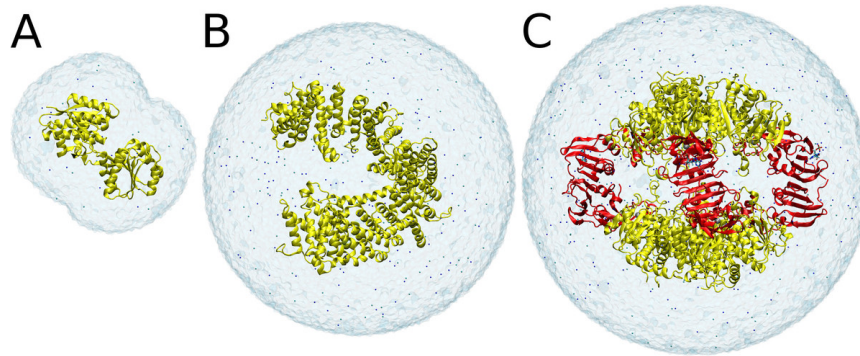
**Table S1.** Simulation parameters for free ensembles. Replicas are reported per functional state. Initial  $t_{eq}$  ns discarded as equilibration.

Variables	LBP	ATCase	bs-RNase	ctCRM1
replicas	10 (ensemble),1 (WAXS)	3	5	10
$t_{sim}$ [ns]	50,100	$T: 20, R: 40$	100	50
$t_{eq}$ [ns]	10	$T: 0, R: 20$	20	20

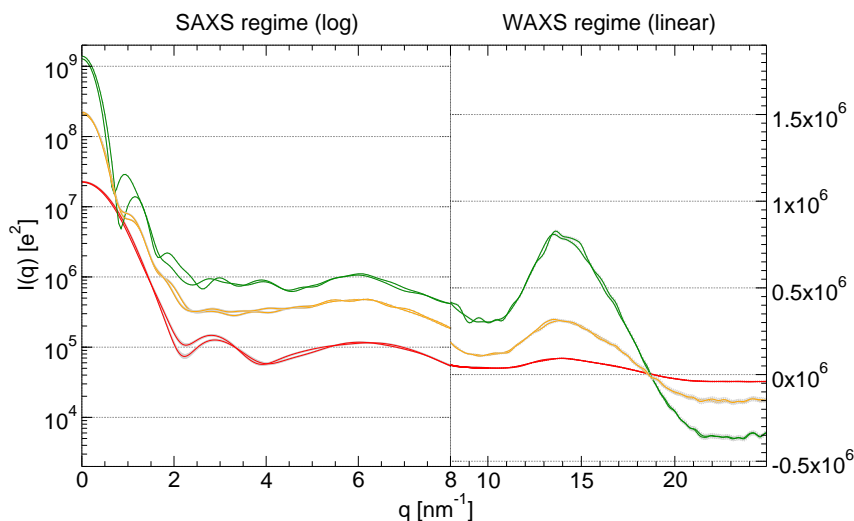
**Table S2.** Simulation Parameters for SWAXS-driven MD, and structural information used to assist in selection of scattering vectors. Maximum extent,  $R_{max}$ , taken from GNOM calculations, and converted to width of a Shannon channel  $q_{Shannon} = \pi/R_{max}$ , and total number of Shannon channels within coupling range  $n_{Shannon}$ . Scattering vectors coupled in SWAXS-driven MD are distributed linearly between  $q_{min}$  and  $q_{max}$ , resulting in a corresponding separation  $q_{sep}$  that is directly comparable to  $q_{Shannon}$ . \*: Additional ATCase simulations with single- $q$ -point restraints are conducted with increased  $k_c$ .

System	LBP	ATCase	ctCRM1
$R_{max}$ [nm]	6.2	13	11
$q_{Shannon}$ [ $\text{nm}^{-1}$ ]	0.51	0.24	0.29
$n_{Shannon}$	15.8	9.1	8.4
$q_{sep}$ [ $\text{nm}^{-1}$ ]	0.5	0.2	0.2
$n_q$	16	11	12
$q_{min}$ [ $\text{nm}^{-1}$ ]	0.5	0.2	0.4
$q_{max}$ [ $\text{nm}^{-1}$ ]	8.0	2.2	2.6
$k_c$	100~3000	500 (2500)*	1000
$\tau$ [ns]	1.0	1.0	2.5
$\delta t$ [ps]	2.0	2.0	2.0
Total simulation time [ns]	10	20~23	40

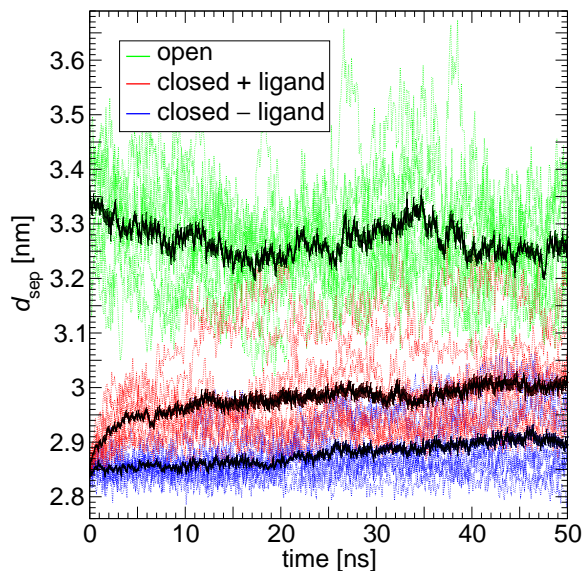
## Supplementary Figures



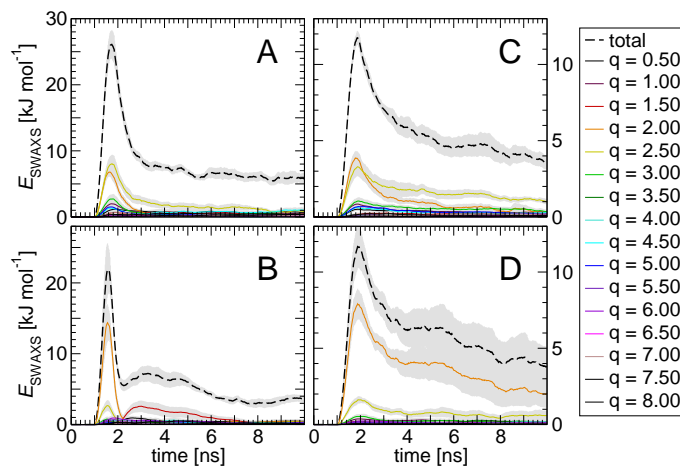
**Figure S1.** Biological systems presented in this paper, displayed to the same scale. The solute and bound ligands (if any) are shown in cartoon and stick formats, respectively. Solvating ions are shown as small spheres, while water molecules are abstracted for clarity. Each solute is enclosed in an envelope, in which all atoms are considered for SWAXS scattering calculations (see Methods of the main text). (A) LBP in the apo-state. (B) ctCRM1 in its ring-open state. (C) ATCase in the *R*-state, with bound ATP. Catalytic trimers and regulatory dimers are colored yellow and red, respectively.



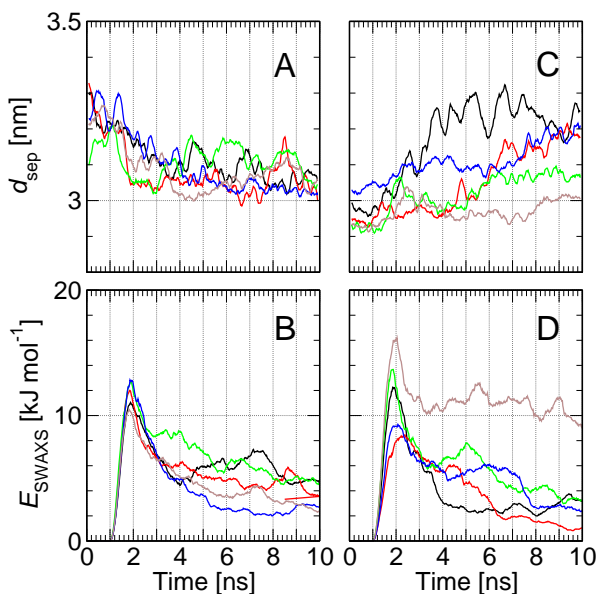
**Figure S2.** SAXS and WAXS patterns of the biological systems considered in this study, superimposed without rescaling to indicate relative intensities. Left panel shows  $\log I(q)$  over  $q < 8 \text{ nm}^{-1}$ , in the SAXS/near-WAXS regime, while the right panel shows  $I(q)$  over  $8 \text{ nm}^{-1} < q < 25 \text{ nm}^{-1}$  in the WAXS regime. Error bars (where visible,) are shown in grey. Colours are as follows: Red – LBP in the apo- (open) and holo- (closed) states. Orange – ctCRM1 simulated under CHARMM22\* forcefield, in position restrained (ring-open) and free (ring-closed) simulations. Green – ATCase in the R- and T- states, in free simulations. The curves show that the intensities mainly change in the SAXS regime upon conformational transition, but hardly in the WAXS regime, justifying why we here coupled SWAXS-driven MD only to SAXS and near-WAXS data.



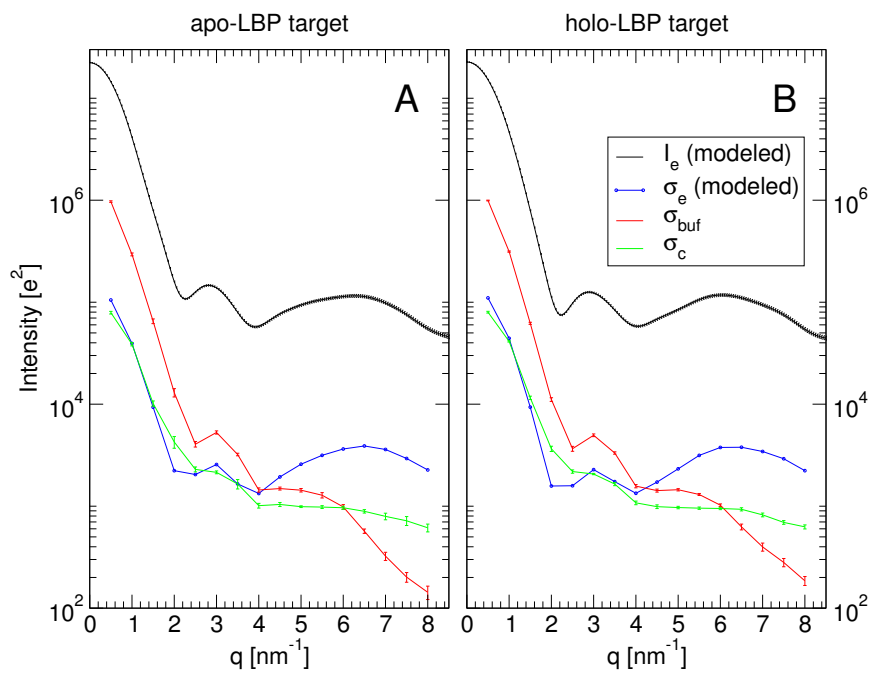
**Figure S3.** Domain separation  $d_{\text{sep}}$  during free 50ns-simulations of LBP, starting from open (green), closed (red), and closed with ligand removed (blue) configurations. Colored, dotted lines represent individual trajectories, and thick black lines represent average over ten replicates.



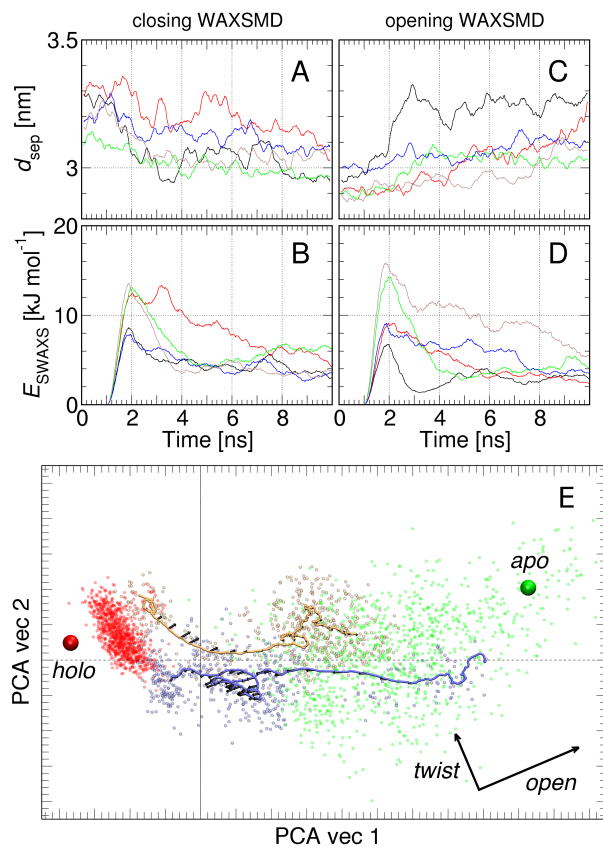
**Figure S4.** Decomposition of  $E_{\text{SWAXS}}$  during LBP SWAXS-driven MD simulations, demonstrating that the conformational transitions are mainly driven by variations of the intensity at  $q = 2 \text{ nm}^{-1}$ . The total  $E_{\text{SWAXS}}$  (thick black, dotted) is plotted along with contributions from individual scattering angles (colored as shown in legends). All values represent averages over 5 replicate trajectories, with standard errors shown as a solid light-grey background. (A/C) open LBP targeting SWAXS curves of closed LBP, using (A)  $k_c = 1000$  or (C)  $k_c = 300$ . (B/D) closed LBP targeting SWAXS curves of open LBP, using (B)  $k_c = 1000$  or (D)  $k_c = 300$ .



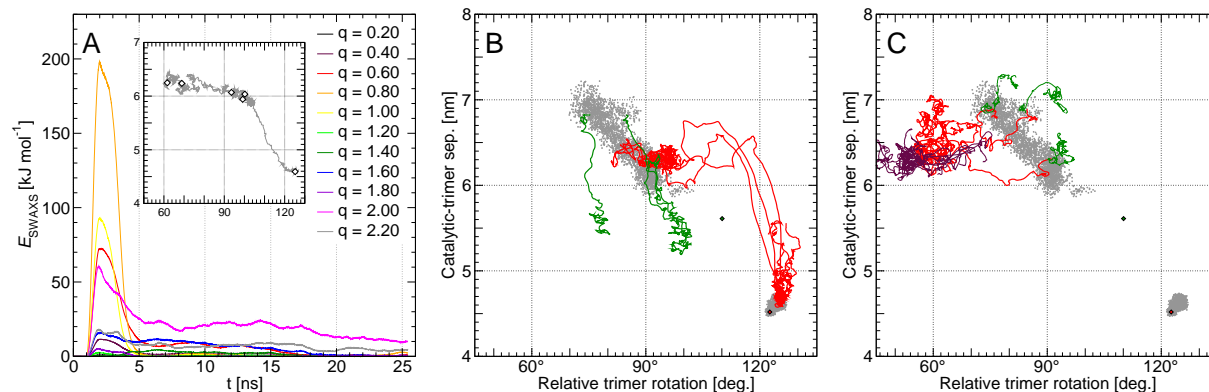
**Figure S5.** Opening and closing trajectories of SWAXS-driven MD simulations of LBP at  $k_c = 300$  and  $\tau = 1 \text{ ns}$ . (A/C) smoothed domain separation  $d_{\text{sep}}$  and (B/D) SWAXS potential  $E_{\text{SWAXS}}$  versus simulation time. (A/B)  $d_{\text{sep}}$  and  $E_{\text{SWAXS}}$  for closing transforms; (C/D)  $d_{\text{sep}}$  and  $E_{\text{SWAXS}}$  for opening transforms.



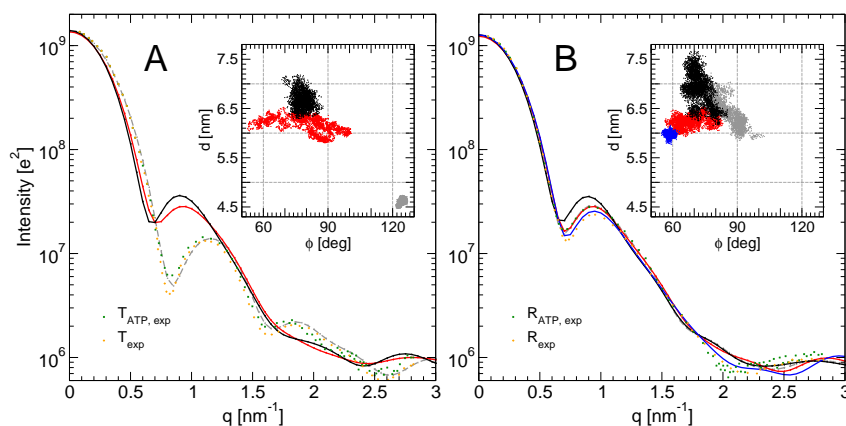
**Figure S6.** Statistical and systematic uncertainties applied as inverse weights for SWAXS-driven MD of LBP (eq. 3 of the main text). (A) Targeting the apo state, and (B) the holo state of LBP; black: Target SWAXS curves; blue: experimental error, modeled as a linear increase from 0.5% at  $q = 0$  to 5% at  $q = 8 \text{ nm}^{-1}$  of  $I_e$ , resembling the statistical error in the experimental data used in a previous study(23); red: systematic error in  $I_c(q)$  due to an uncertainty of 1% in the buffer density; green: statistical error of  $I_c(q)$ . The overall uncertainty is dominated by the buffer at low  $q$ , and by the experimental error at high  $q$ .



**Figure S7.** Opening and closing trajectories of SWAXS-driven MD simulations of LBP at  $k_c = 1$  and  $\tau = 1$  ns, using the errors as inverse weights instead of uniform weighting (eq. 3 of the main text, and Fig. S6). (A/C) smoothed domain separation  $d_{\text{sep}}$  and (B/D) SWAXS potential  $E_{\text{SWAXS}}$  versus simulation. (A/B) for closing transitions; and (C/D) for opening transitions. (E) PCA analysis of the black trajectory in (A-D), using the same color scheme as Fig. 2. The green and red spheres indicated the apo and holo crystal structures, respectively.

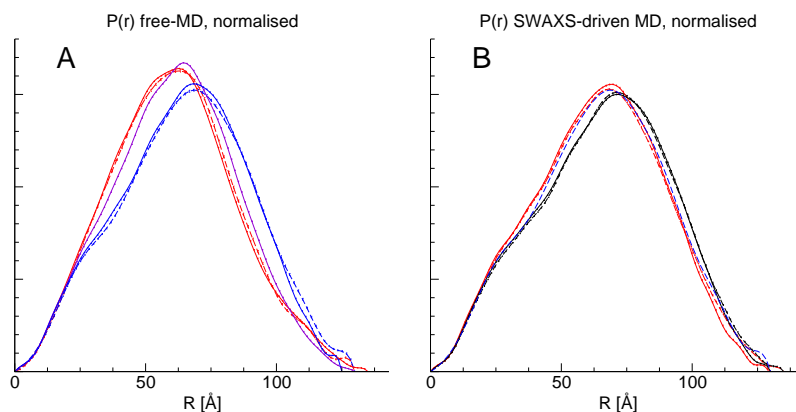


**Figure S8.** (A) Decomposition of  $E_{\text{SWAXS}}$  into contributions from different scattering vectors  $q$ , shown for one typical  $T \rightarrow R$  trajectory of ATCase. Main figure:  $E_{\text{SWAXS}}$  versus time; inset: trimer separation  $d_{\text{sep}}$  and relative rotation  $\phi$  during that simulations, with positions every 5 ns labelled with diamonds. The decomposition shows that, while the rapid expansion during the first transition is triggered by restraints near  $q = 1 \text{ nm}^{-1}$ , the subsequent rotation along  $\phi$  was triggered by restraints at  $q = 2 \text{ nm}^{-1}$ . (B) Partial SWAXS-driven MD trajectories of  $T$  and  $R$  ATCase, coupled to  $0.2 \sim q \sim 1.6 \text{ nm}^{-1}$  in order to fit only the first SAXS peak. Green lines indicate smoothed  $R \rightarrow T$  trajectories, and red lines indicate smoothed  $T \rightarrow R$  trajectories. Exclusion of the second peak during SWAXS-driven MD results in transitions primarily along  $d_{\text{sep}}$  with only diffusive rotation along  $\phi$ . (C) Partial  $R \rightarrow R$  and  $R \rightarrow T$  SWAXS-driven trajectories coupled only to the second peak signal at  $q = 2 \text{ nm}^{-1}$ . Red lines indicate  $R \rightarrow R$  trajectories using  $k_c = 500$ , maroon lines indicate  $R \rightarrow R$  trajectories using  $k_c = 2500$ , and green lines indicate  $R \rightarrow T$  trajectories using  $k_c = 500$ . Those trajectories demonstrate that coupling purely to the SAXS signal at  $q = 2 \text{ nm}^{-1}$  results in a rotation along  $\phi$  without any significant change in  $d_{\text{sep}}$ .

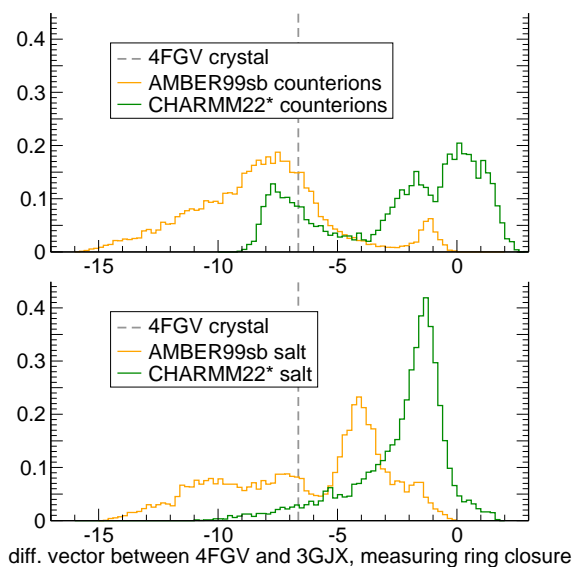


**Figure S9.** SAXS curves and trimer configuration of ATCase ensembles in free simulations before applying SWAXS-derived restraints (grey, dotted), during SWAXS-driven simulations (red and blue), and in free simulations after releasing the SWAXS-derived restraints (black). Good agreement with experimental  $R$  is found at low  $\phi$  (blue line and dots in B). After releasing the SWAXS-derived restraints, the simulations partly returned to the solution  $R$  in free simulations. Insets show corresponding trajectories along the trimer separation  $d_{\text{sep}}$  and relative trimer rotation  $\phi$  in their respective colors. (A) Trajectories starting from the  $T$ -state. (B) Trajectories starting from the  $R$ -state.

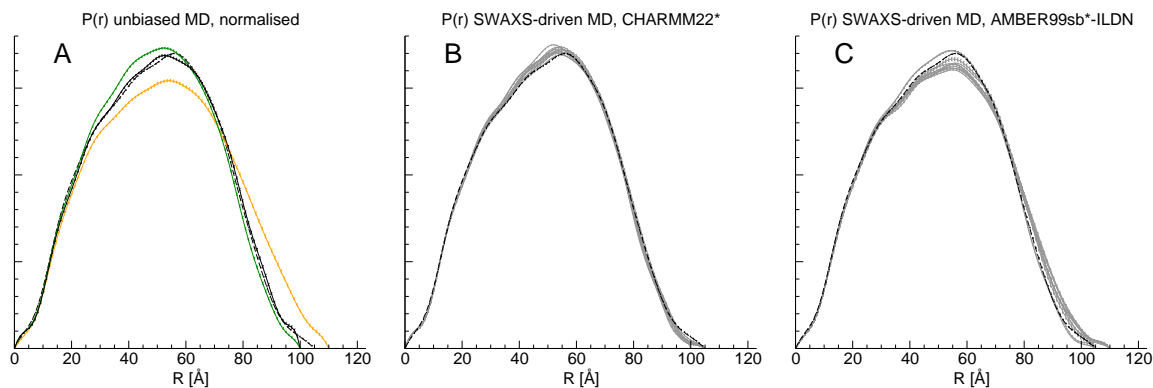




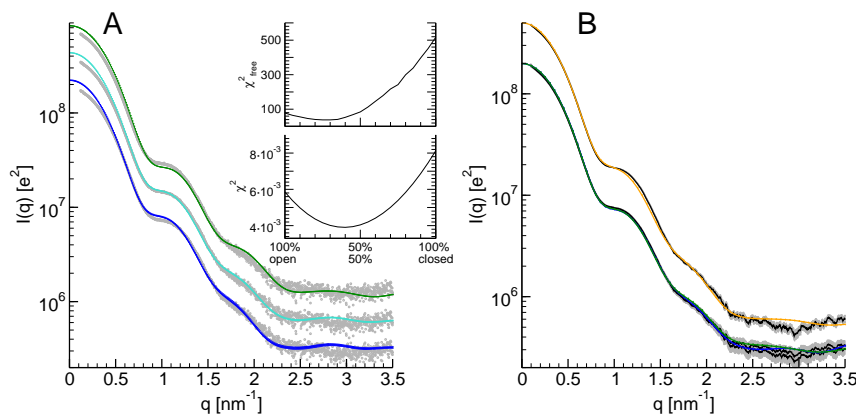
**Figure S10.** Pair distribution functions  $P(r)$  for ATCase, calculated using GNOM based on curves in Figures 5A and S9. (A)  $P(r)$  curves for ATCase. Color scheme as follows – Red-solid: T-ATCase unbiased simulations. Red-dashed: T-ATCase experiment. Purple-solid: R-ATCase unbiased position-restrained simulations. Blue-solid: R-ATCase unbiased simulations. Blue-dashed: R-ATCase experiment. (B)  $P(r)$  curves for ATCase during and after SWAXS-driven MD. Color scheme as follows – Blue-dashed: Experimental R-ATCase, as in A. Red-dashed: T-ATCase driven to R-state spectra. Black-dashed: Relaxed T-ATCase after release of SWAXS constraints. Red-solid: R-ATCase driven to R-state spectra. Black-solid: Relaxed R-ATCase after release of SWAXS constraints.



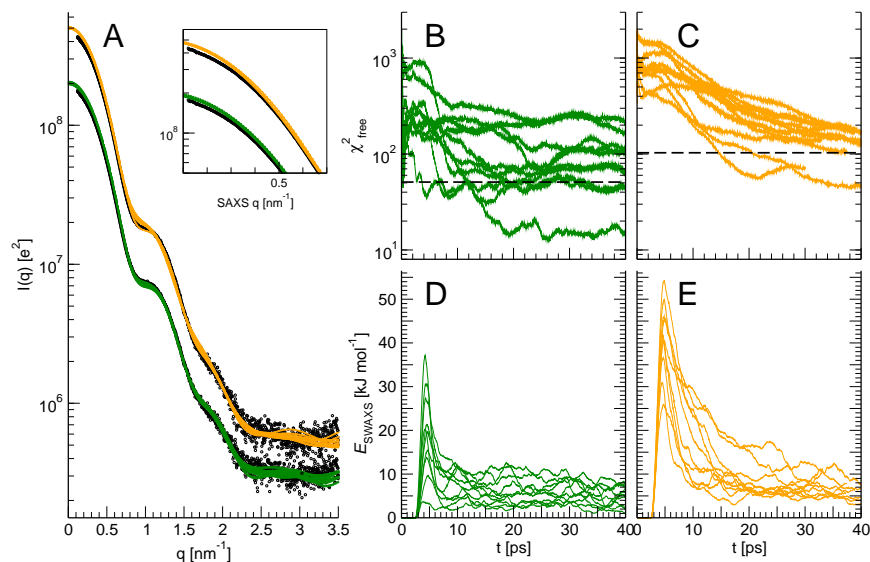
**Figure S11.** Conformations of CRM1 under different conditions, projected onto the difference vector between open (4FGV) and closed (3GJX) CRM1, following Monecke *et al.*(4). Location of 4FGV structure shown as grey lines. Top: Comparison of AMBER99sb and CHARMM22\*, solvated only with counterions. Bottom: Comparison of AMBER99sb and CHARMM22\*, solvated with 100 mM NaCl.



**Figure S12.** Pair distribution functions  $P(r)$  for CRM1, calculated using GNOM based on curves in Figures 6B and S14A. (A)  $P(r)$  curves for ctCRM1. Color scheme as follows – green-solid: ctCRM1 unbiased simulations using CHARMM22\* forcefield. orange-solid: ctCRM1 unbiased simulations using AMBER99sb\*-ILDN forcefield. black-dotted: ctCRM1 experiment. black-solid: Hybrid spectra of 45% free-ctCRM1 (ring-closed) and 55% restrained ctCRM1 (ring-open). (B)  $P(r)$  curves for ctCRM1 during SWAXS-driven MD, under CHARMM22\*. Color scheme as follows – black-dashed: target experimental ctCRM1 curve, as in A. grey-solid: individual SWAXS-driven replica. (C)  $P(r)$  curves for ctCRM1 during SWAXS-driven MD, under AMBER99sb. Color scheme as follows – black-dashed: target experimental ctCRM1 curve, as in A. grey-solid: individual SWAXS-driven replica.



**Figure S13.** (A) SAXS curves of CRM1 ensembles under the CHARMM22\* forcefield, in ring-open (blue) and ring-closed (green) conformations, and a linear combination of 60% open/40% closed spectra (cyan). The unsmoothed experimental spectra (circles) was fit to these curves and displayed for comparison, and curves were vertically shifted for clarity. Ensembles were generated from position-restrained and free simulations. In the insets,  $\chi^2$  fits were conducted on linear combinations of the open and closed curves according to Eq. 1 in the Supporting Material Methods, and  $\chi_{\text{free}}^2$  fits were conducted according to Rambo and Tainer (25). The  $q$ -range for this fit is set between  $0.5 \sim 2.7 \text{ nm}^{-1}$  to avoid the dominating influence of experimental noise at  $q \sim 0.3 \text{ nm}^{-1}$ . (B) SAXS patterns of CRM1 ensembles after SWAXS-driven MD. Blue and green curves: CHARMM22\* ensembles. Orange: AMBER99sb ensemble. Black: smoothed experimental curve with errors (grey).



**Figure S14.**  $\chi^2_{\text{free}}$  analysis of ctCRM1 SWAXS-driven trajectories using the CHARMM22\* (green) or the AMBER99sb force field (orange). (A) Exponentially averaged calculated SAXS curves of CRM1, at the final simulation timepoint of SWAXS-driven MD. All replica are plotted as thin lines over the fitted, unsmoothed experimental data (circles). AMBER99sb curves are vertically offset for clarity. A magnification of the SAXS region is shown on the inset. (B/C)  $\chi^2_{\text{free}}$  values during SWAXS-driven MD. Values are high due to significant underfitting at the Shannon bins between  $0.5 \sim 0.7 \text{ nm}^{-1}$ , and around  $1 \text{ nm}^{-1}$ , suggesting that the MD forcefield imposed tight restraints on possible protein conformations. Thick dotted lines represent  $\chi^2_{\text{free}}$  of the ensemble SAXS curve in Fig. S13B after averaging over all ten replica. (D/E)  $E_{\text{SWAXS}}$  values during SWAXS-driven MD, demonstrating that  $E_{\text{SWAXS}}$  provides similar information as  $\chi^2_{\text{free}}$ .