JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

SAXS-Guided Metadynamics

Dari Kimanius,[†] Ingrid Pettersson,[‡] Gerd Schluckebier,[‡] Erik Lindahl,[†,§] and Magnus Andersson*[,§]

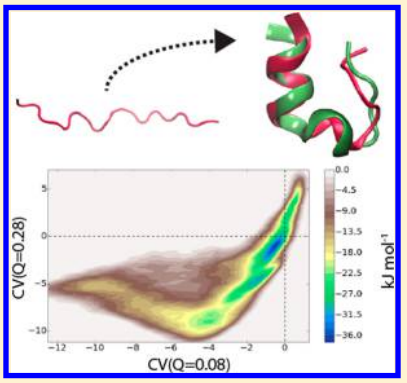
[†]Department of Biochemistry and Biophysics, Center for Biomembrane Research, Stockholm University, SE-106 91 Stockholm, Sweden

[‡]Modeling and Structural Biology, Protein Engineering, Global Research, Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark

[§]Department of Theoretical Physics and Swedish e-Science Research Center, Science for Life Laboratory, KTH Royal Institute of Technology, SE-171 21 Solna, Sweden

Supporting Information

ABSTRACT: The small-angle X-ray scattering (SAXS) methodology enables structural characterization of biological macromolecules in solution. However, because SAXS provides low-dimensional information, several potential structural configurations can reproduce the experimental scattering profile, which severely complicates the structural refinement process. Here, we present a bias-exchange metadynamics refinement protocol that incorporates SAXS data as collective variables and therefore tags all possible configurations with their corresponding free energies, which allows identification of a unique structural solution. The method has been implemented in PLUMED and combined with the GROMACS simulation package, and as a proof of principle, we explore the Trp-cage protein folding landscape.



1. INTRODUCTION

Small-angle X-ray scattering (SAXS) is routinely used to gather low-resolution structural information of biomolecules in solution.[1,2] Combining ingenious experimental and modeling approaches has allowed the structural basis of protein−protein interactions,[3] fibrillation processes,[4] and protein−ligand interactions to be visualized.[5,6] Because aggregation states of peptide-based drugs can be resolved using SAXS,[7,8] the formulation step in drug discovery particularly benefits from this technology. Recent technical advances at synchrotrons have enabled the extension of X-ray scattering into the wide-angle regime (WAXS),[9] enabling visualization of finer structural details such as protein secondary structures and folds.[10,11] In addition, visualizing membrane protein targets seems to be conceivable considering the development of nanodisc technology[12] and modeling of surrounding lipids/detergents.[13] Moreover, time-resolved WAXS (TR-WAXS) of soluble proteins[14,15] and membrane proteins[16,17] shows great promise for adaptation to X-ray free electron laser (XFEL) technology.[18,19]

The molecules in the sample are not oriented in space, and the diffraction pattern obtained will therefore represent a rotational average in reciprocal space corresponding to the multitude of displayed molecular orientations. The resulting low-resolution data are inherently difficult to model because several hypothetical molecular configurations can reproduce the one-dimensonal scattering profile. The task of any structural refinement method, therefore, becomes to single out the molecular configuration that corresponds to the native state of the protein. The need to account for displaced solvent and increased density in the solvation shell add to the complexity of modeling SAXS/WAXS data. While existing refinement techniques show little variation in how the scattering from the target molecule is evaluated, there are drastic differences in accounting for the surrounding solvent molecules. Refinement protocols that model the solvent implicitly typically fit the predicted scattering pattern to the experimental data using free parameters such as solvation shell density, excluded volume, and atomic group radii.[20−23] Because these free parameters are difficult to determine experimentally, refinement protocols that rely on implicit solvent models come with the caveat of overfitting. Therefore, treating solvent molecules explicitly minimizes the risk of overfitting and was also shown to produce better fits to experimental data compared to implicit model refinement methods.[24] Therefore, atomistic molecular dynamics (MD) simulations provide the most accurate descriptions of both the protein−water boundary and the excluded volume.[25] Recently, unrestrained MD simulations have shown excellent reproduction of target scattering profiles.[26] Such atomistic representations come at a significant additional computational cost, which can be alleviated by coarse graining the protein structure.[11]

Rather than performing simulation-based refinement processes independently from the experimental data, it would be preferable to integrate a comparison between the simulated configurations and the target scattering profile. Because several

Received: March 31, 2015
Published: June 10, 2015

DOI: 10.1021/acs.jctc.5b00299
J. Chem. Theory Comput. 2015, 11, 3491−3498

molecular configurations can reproduce the low-resolution scattering data, simply biasing the potential to drive a single simulation toward the experimental data should be avoided. Bias-exchange metadynamics is a simulation method that allows exploration of the free energy landscape with respect to several so-called collective variables (CVs).[27] The CVs summarize and average the microscopic description of the system into a few coordinates more relevant for the degrees of freedom in macromolecular sampling problems, such as the distance between two domains. Recently, nuclear magnetic resonance (NMR) chemical shifts were used as CVs to characterize the folding landscape of protein G from streptococcal bacteria (GB3).[28] The bias-exchange metadynamics approach is particularly well-suited for refinement of SAXS/WAXS data because it inherently avoids being trapped in local minima and therefore minimizes the risk of identifying solutions that reproduce the experimental data while they do not necessarily represent the target native configuration with the lowest free energy.

In this work, we present a novel structural refinement method aimed at SAXS/WAXS data by implementing scattering intensities at different angles as CVs in a bias-exchange metadynamics setting. The folding free energy landscape of the Trp-cage protein is used as proof of principle, and we observe reproduction of the target NMR structure to within 2.4 Å root-mean-square deviation (RMSD). The Gromacs package,[29] version 5.0.4, was chosen as the underlying MD engine, but the presented refinement protocol is generalizable to other MD engines such as NAMD,[30] LAMMPS,[31] and Amber.[32] While our method development focused on X-ray solution scattering, the refinement protocol is equally suitable for neutron scattering data with minor adjustments.

2. THEORY AND METHODS

2.1. X-ray Scattering Profiles. The calculated average scattering intensity I_C of a large set of identical particles in random orientations is given by the Debye formula[33]

$$I_C(Q) = \sum_i \sum_j f_i f_j \frac{\sin Qr_{ij}}{Qr_{ij}} \tag{1}$$

where f_i is the form factor of the ith atom, $r_{ij} = |\boldsymbol{r}_j - \boldsymbol{r}_i|$ is the distance between atoms i and j, and Q is the amplitude of the scattering vector defined as $Q = 4\pi \sin\theta/\lambda$, where λ is the wavelength of the X-ray beam and θ represents half of the scattering angle. Because solvent effects are inherent to atomistic MD simulations, the form factors were evaluated according to Fraser et al.[34] without correcting for excluded volume and boundary solvent molecules.

2.2. Bias-Exchange Metadynamics. Metadynamics calculations enhance sampling by introducing additional bias forces to the simulated thermodynamic forces and thereby prevent previously explored regions in the configurational space from being visited.[35,36] The bias forces are derived from a history-dependent potential defined in a space spanned by an arbitrary set of collective variables (CVs). This potential is built up by repulsive Gaussian potential contributions deposited at regions previously visited by the simulation. After convergence, the sum of all of the deposited contributions to the potential will be a negative approximation of the underlying free energy surface in the particular CV space. Denoting the ith CV as ξ_i, the applied metadynamics potential V_G for the current set of microscopic coordinates, $\boldsymbol{r}$, at time t is given as

$$V_G(\boldsymbol{r},t) = \omega \sum_{t' \le t} \exp\left(-\sum_i \frac{[\xi_i(\boldsymbol{r},t) - \xi_i^{t'}]^2}{2\sigma_i^2} \right) \tag{2}$$

where the first sum iterates over the set of Gaussian potentials previously deposited and the second sum is over all of the CVs in the simulation; ω and σ_i are the height and width of the deposited Gaussian potential in the ith CV, respectively. The biasing force F_m then becomes

$$F_m(\boldsymbol{r},t) = -\nabla_m V_G(\boldsymbol{r},t) = -\omega \sum_{t' \le t} \sum_i \frac{\xi_i(\boldsymbol{r},t) - \xi_i^{t'}}{\sigma_i^2} \exp\left(-\sum_i \frac{[\xi_i(\boldsymbol{r},t) - \xi_i^{t'}]^2}{2\sigma_i^2} \right) \nabla_m \xi_i(\boldsymbol{r},t) \tag{3}$$

where ∇_m denotes the gradient with respect to the mth atom. Hence, in addition to descriptions of the CVs $\xi_i(\boldsymbol{r},t)$, any metadynamics method development requires an explicit expression for the gradient ∇_m with respect to the microscopic coordinates. To further enhance the exploration of the free energy landscape, a bias-exchange protocol was used that allows periodic configurational exchanges between the biasing potentials.[27]

2.3. Collective Variables. In this work, we defined a novel CV based on X-ray scattering intensity by using an experimental intensity at a single Q-value reference point

$$\xi_Q = I_C(Q) - I_E(Q) = C \sum_i \sum_j f_i f_j \frac{\sin Qr_{ij}}{Qr_{ij}} - I_E(Q) \tag{4}$$

where $I_C(Q)$ is calculated according to eq 1 and $I_E(Q)$ is the experimental intensity at Q. The scale factor C functions as a fitting parameter and is evaluated using a weighted least-squares fit, given as

$$C = \frac{\sum_k I(Q_k) \times I_E(Q_k) \times w^2(Q_k)}{\sum_k I^2(Q_k) \times w^2(Q_k)} \tag{5}$$

where k is the number of Q-values defined by the user to describe the scattering profile and the weighting function w is based on the experimental error at Q to embed error data into the calculations and will be defined here as the reciprocal experimental error.[22] The set of Q-values should be equally distributed over the relevant scattering range. Each $Q(k)$ will be assigned a scattering intensity corresponding to a running average over the kth part of the scattering profile. The number and position of the Q-values should be selected to represent the different features of the profile and were set here to 20 over the interval $Q = [0.1,0.5]$. We set these parameters equally for all of the CVs, hence rendering equal C-values across the entire trajectory.

By letting the CV at a specific angle differentiate between configurations that result in scattering intensities that are both above and below the target experimental scattering, we allow for a better resolution of the free energy landscape compared to less versatile solutions such as, e.g., χ^2 fitting procedures. A key feature of the SAXS methodology is that the full range of distances (5−60 Å) is measured simultaneously. To explore the free energy landscape with respect to all distances and thus to

DOI: 10.1021/acs.jctc.5b00299
J. Chem. Theory Comput. 2015, 11, 3491−3498

mimic the experimental results, several CVs, each defined at specific scattering angles, should be used. In our protocol, this is accomplished using the bias-exchange protocol with four CVs at Q = 0.08, 0.14, 0.20, 0.28 Å^{-1}, respectively (Figure 1). In addition, a two-dimensional ensemble combining Q = 0.08 and 0.28 Å^{-1} was also added to the bias-exchange protocol.

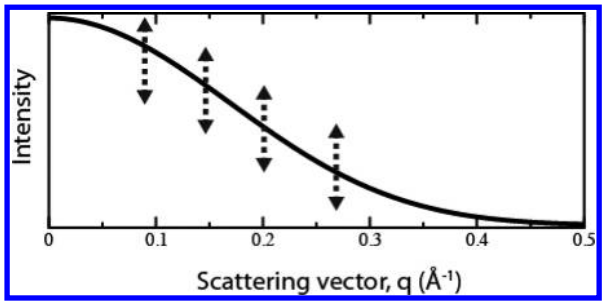


Figure 1. Definition of SAXS CVs. CV(Q = 0.08), CV(Q = 0.14), CV(Q = 0.20), and CV(Q = 0.28) are shown superimposed on the theoretical scattering profile for the Trp-cage NMR structure.

In addition to the developed SAXS CVs, we also used two established CVs: radius of gyration and backbone interactions.[37] The radius of gyration CV was defined as

$$\xi_{\mathrm{Gyr}} = \left(\frac{\sum_i^N m_i |\boldsymbol{r}_i - \boldsymbol{r}_{\mathrm{COM}}|^2}{\sum_i^N m_i} \right)^{1/2} \tag{6}$$

Here, N is the number of atoms; m_i and $\boldsymbol{r}_i$ correspond to the mass and position of the ith atom, respectively; and $\boldsymbol{r}_{\mathrm{COM}}$ is the center of mass, defined as $\boldsymbol{r}_{\mathrm{COM}} = \sum_i^N \boldsymbol{r}_i m_i / \sum_i^N m_i$. The backbone interaction CV was defined as the double sum of contacts between all H-bond acceptors and donors along the backbone. A single contact was described using the continuous switching function

$$s_{ij} = \frac{1 - (r_{ij} - d_0/r_0)^n}{1 - (r_{ij} - d_0/r_0)^m} \tag{7}$$

r_{ij} is the distance between H-bond donors and acceptors and d_0, r_0, m, and n are constants defining the shape of the switching function.

2.4. Implementation. The code describing the new SAXS CV was implemented within the PLUMED framework.[37] The atomic coordinates generated by the MD engine were imported into PLUMED at each time step. However, because temporal fluctuations are significantly slower in the CVs compared to those in the atomic coordinates, a multi-time-step protocol was introduced where the CVs were evaluated only at certain preset time steps.[38] At each metadynamic step, a representative number of intensities for a set of Q-values spanning the relevant region in Q-space were calculated using an efficient algorithm for SAXS profiling.[22] The resulting global scattering profile determined the scale factor in eq 5, which was used to calculate the intensity at the Q-value of choice for the specific CV (eq 4). In addition to the calculated scalar value of the CV, a gradient with respect to the atom coordinates included in eq 4 was determined and used to calculate the bias forces according to

$$\nabla_m \xi_Q = C \sum_{i \neq m} f_m f_i \frac{Q r_{mi} \cos Q r_{mi} - \sin Q r_{mi}}{Q r_{mi}^3} \boldsymbol{r}_{mi} \tag{8}$$

where ∇m is the gradient with respect to atom coordinates of atom index m and $\boldsymbol{r}_{mi} = \boldsymbol{r}_i - \boldsymbol{r}_m$ is the vector difference between atoms m and i. The atomic form factors f_m and f_i were treated as constants in the evaluation of the gradient. The gradient was scaled to properly account for the metadynamics multitime step. The time step iteration was finalized by adding the bias forces to the thermodynamic forces in the MD engine. In our simulations, intensities at low-Q regions of the calculated scattering profiles showed larger fluctuations compared to those at higher scattering angles. The width of the deposited potential was therefore chosen as 0.20, 0.18, 0.12, 0.10 au for CV(Q = 0.08), CV(Q = 0.14), CV(Q = 0.20), and CV(Q = 0.28), respectively. The potential height ω was set to 0.1 kJ mol^{-1} throughout the simulations, since this resulted in good sampling in earlier metadynamics simulations of the Trp-cage system.[39]

To benchmark our developed SAXS-CVs, a control simulation was performed using a bias-exchange metadynamics approach with already established CVs chosen to describe general properties of a protein structure in solution: the radius of gyration and backbone interactions, described in Section 2.3, as well as helicity and end-to-end distance. The helicity CV describes the helical content of the protein by comparing the RMSD distance between collections of six continuous amino acid residues in the simulated structure at each frame with an idealized α-helix. The CV was calculated as the sum of switching functions, eq 7, where r is the mentioned RMSD distance.[40] The end-to-end distance CV was defined as the distance between Cα's of the first and last residue.

2.5. Computational Details. All simulations were performed using the GROMACS molecular dynamics package[29] with a time step of 2 fs. The AMBER03 force field[41] was used because it has been shown to give reliable results of the Trp-cage NMR structure.[39,42] The temperature was kept at 282 K using a V-rescale thermostat[43] with a time constant of 0.1 ps and a Parrinello−Rahman pressure coupling[44] with a time constant of 2.0 ps. The particle mesh Ewald was used for calculations of long-range interactions based on cubic interpolation. Short-range neighborlist, electrostatic, and van der Waals cutoffs were set to 1.0 nm. All bonds were constrained to equilibrium length using the LINKS (LINear Costraint Solver) algorithm.[29] The protein was solvated in 9704 water molecules, using the tip3p water model[45] in a periodic dodecahedron-shaped simulation box. Starting from a random coil, each replica was run for 81 ns, resulting in a total of 648 ns of bias-exchange metadynamics simulation time distributed over 8 replicas based on 6 CVs:

- Four SAXS CVs were defined at Q-values of 0.08, 0.14, 0.20, and 0.28 Å^{-1} with decreasing σ-values of 0.20, 0.18, 0.12 and 0.10, respectively, and were evaluated at every tenth MD step.
- Radius of gyration as described by eq 6 with σ-values set to 0.01.
- Backbone interaction count as described by eq 7 with parameters m = 12, n = 6, r_0 = 0.2, d_0 = 0.0, and σ = 1.0. The calculations were restricted to donors and acceptors in the backbone.

Replicas 1−6 were biased with each of the 6 CVs, whereas the remaining two replicas were run in a two-dimensional CV space with SAXS-CV at Q = 0.08 paired with Q = 0.28 and radius of gyration paired with backbone interaction. Gaussians

DOI: 10.1021/acs.jctc.5b00299
J. Chem. Theory Comput. 2015, 11, 3491−3498

were deposited every 1 ps, and exchange attempts were performed every 4 ps.

The radius of gyration and backbone interaction CVs were identically defined in the control simulation. The two additional CVs, helicity and end-to-end distance, were described with σ-values of 0.2 and 0.1, respectively. The parameters of the switching function used to describe the helicity CV were set to $m = 0$, $n = 6$, $r_0 = 0.08$, and $d_0 = 0.0$. A total of six replicas were used, where replicas 1−4 were biased using each of the four CVs separately, replica 5, using the radius of gyration CV combined with the backbone interaction CV, and replica 6, using the helicity CV in combination with the end-to-end distance CV. Each replica in the control simulation was run for 87 ns, which sums to a total of 520 ns.

3. RESULTS

The SAXS-driven bias-exchange metadynamics approach presented here was developed using the NMR structure (PDB ID 1L2Y) of the designed 20-residue Trp-cage protein as validation.[46]

3.1. Numerical Integration. To verify that the developed SAXS CV was implemented correctly, we evaluated the numerical integration of the bias forces and compared it to the biased potential. In a correctly implemented CV, the potential calculated by averaging over the integrated bias forces should approximately capture the tendencies of the actual bias potential, up to a constant offset. The integration with respect to a single Cartesian component x was performed according to

$$V^{t+1} = V^t - \Delta x^t d_x^{t+1} F^{t+1} = V^t - (x_{t+1} - x_t) d_x^{t+1} F^{t+1} \quad (9)$$

where V^t is the bias potential at time step t, x is the atomic position component, d_x^t is the corresponding component of the gradient at time step t, and F is the scalar force applied along the direction of the gradient. Averaging of the three-component bias force vector allows it to be directly compared to the scalar bias potential.

The comparison of the average numerical integration of the bias forces and the biased potential was performed using a single CV set to $Q = 0.2$ Å^{-1}. The height of the applied Gaussian potential was set to $\omega = 10$ kJ mol^{-1}, the spread was $\sigma = 2$ au, and the potentials were observed for 1000 time steps (Figure 2). During this short simulation, the structure did not change significantly. However, we were able to monitor changes in the potential, and it was clear that both the integrated bias forces and the bias potential followed similar patterns and hence the CV was correctly implemented. The evaluation of the bias forces involves the Δx components and will, therefore, incorporate a contribution from the thermodynamic potential in the simulation. This force field dependent discrepancy can be observed as the differences between the potentials in Figure 2.

3.2. Folding Landscape of the Trp-Cage Protein. To test the implemented SAXS-CV, we explored the folding landscape of the Trp-cage protein using bias-exchange metadynamics simulations with seven independent ensembles. The SAXS-CV was used as CV($Q = 0.08$), CV($Q = 0.14$), CV($Q = 0.20$), and CV($Q = 28$). In addition, one ensemble was run using two SAXS CVs simultaneously: CV($Q = 0.08$) and CV($Q = 0.28$). We also included two established CVs: radius of gyration (RG) and backbone interaction (BB). Each ensemble was simulated for 81 ns with a stride for the SAXS-CV evaluation of 10 time steps.

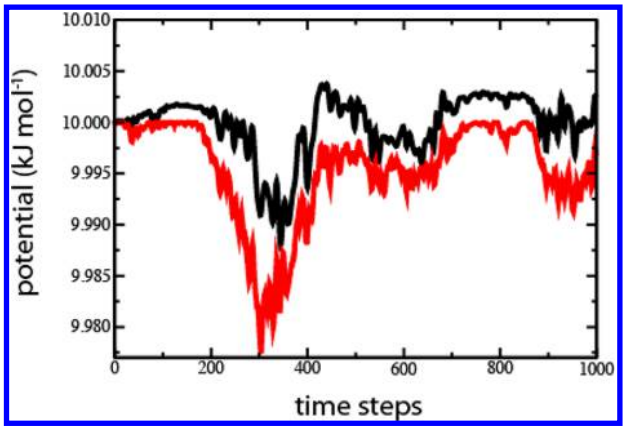


Figure 2. Verification of the CV implementation. The average of the integrated biased forces is shown in black, and the biased potential is shown in red. The starting height and spread of the applied Gaussian were $\omega = 10$ kJ mol^{-1} and $\sigma = 2$ au, respectively.

The resulting free energy landscape with respect to the simultaneously evaluated CV($Q = 0.08$) and CV($Q = 0.28$) is presented in Figure 3A. The metadynamics simulation started

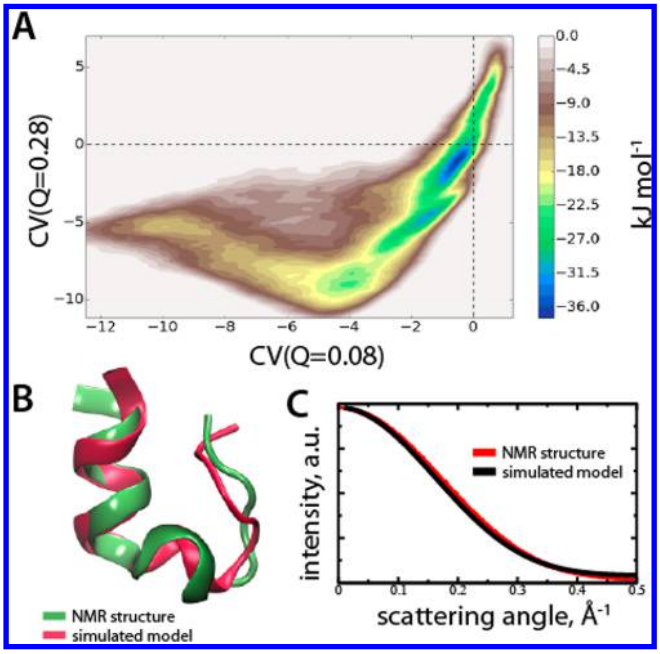


Figure 3. Identification of the Trp-cage native state by combining bias-exchange metadynamics and SAXS data. (A) Free energy landscape represented by the ensemble using two-dimensional CVs: CV($Q = 0.08$) and CV($Q = 0.28$). (B) Average structure from the cluster closest to the global free energy minimum with >3.7 backbone interactions (magenta) superimposed on the target NMR structure (green). (C) Predicted and target X-ray scattering profiles shown in black and red, respectively.

with the Trp-cage protein in an unfolded configuration generated in PyMOL[47] with CV($Q = 0.08$) = −13 au and CV($Q = 0.28$) = −6 au. The main features of the free energy landscape were produced within the first tens-of-nanoseconds of the simulation. Two free energy minima were observed that deepened continuously and were clearly separated during the final course of the simulation, where the minimum closer to the origin showed significantly lower free energies. The wide range of possible rotational averages from the overall unfolded-to-

DOI: 10.1021/acs.jctc.5b00299
J. Chem. Theory Comput. 2015, 11, 3491−3498

folded configurations manifests itself in the shape of the two-dimensional free energy landscape, i.e., the low-angle CV(Q = 0.08) showed a significantly higher degree of variation compared to that of the higher-angle CV(Q = 0.28).

Because solution scattering involves dynamical ensembles of structures rather than a single static native state, we used a clustering procedure to find the representative lowest free energy structure, corresponding to the native state. First, a QT-clustering algorithm[48] using a 1 Å RMSD cutoff was used to bin similar structural configurations generated in the ensemble. Each cluster was then assigned with an average free energy corresponding to the free energies of all included configurations. To remove random coil configurations that display low free energies, we added a cutoff based on the average backbone interactions, which was determined to be 3.7 in the simulations. Taking this cutoff into consideration (i.e., >3.7 backbone interactions), the cluster average configuration representing the lowest free energy showed a backbone RMSD of 2.6 Å (Figure 3B) and reproduced the scattering profile of the Trp-cage NMR structure with a χ^2 value of 21.2 (Figure 3C).

By plotting the free energies of all clusters against the RMSD to the target structure, it becomes clear that there are many configurations that reproduce the main features of the target structure (Figure 4A). This provides an excellent illustration of the difficulties inherent to structural refinement of solution scattering experiments. However, because all configurations are tagged with individual free energies in our simulation-based refinement protocol, it becomes possible to single out the lowest free energy native state configuration. In addition, by introducing the backbone-interaction cutoff, random coil configurations with low free energies can be discarded (red boxes, Figure 4A) to favor the identification of native-like configurations (black circles, Figure 4A). We note that, while the predicted scattering profile of a cluster belonging to the configurational population with a backbone interaction count <3.7 reproduces the target scattering quite well (χ^2 = 5.6) (Figure 4B), the structural fit of the cluster average to the NMR target structure is far from satisfactory (RMSD = 7.1 Å) (Figure 4C).

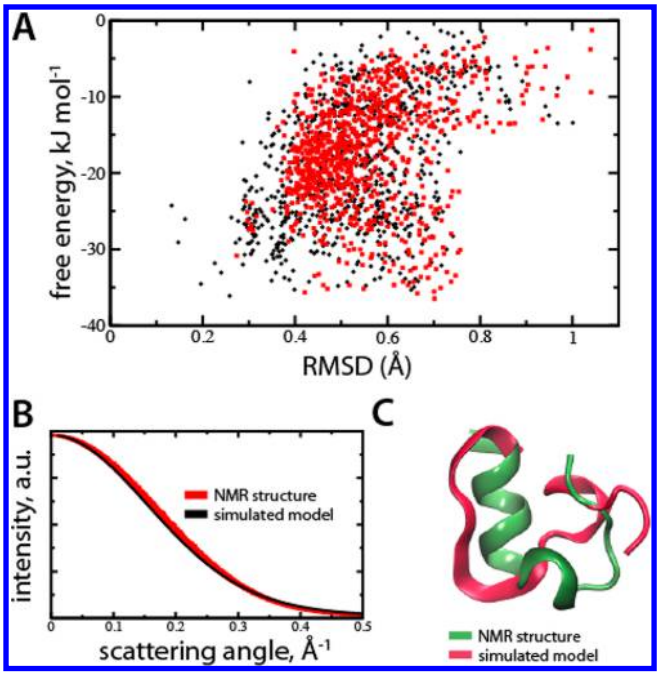


Figure 4. Configurational space in the global free energy minimum. (A) Free energies and RMSD values to the target NMR structure of all average clusters in the simulation. Configurational clusters with backbone interactions >3.7 are displayed as black circles, and clusters with backbone interactions <3.7 are shown as red boxes. (B) Scattering profile of a simulated low-energy configurational cluster with <3.7 backbone interactions (black) relative to the target NMR structure (red). (C) Structure of a low-energy configurational cluster with <3.7 backbone interactions (magenta) superimposed on the target NMR structure (green).

To benchmark our developed SAXS-CVs, a control bias-exchange metadynamics simulation was performed on an identical configurational starting point of the Trp-cage system using only already established CVs that were chosen to describe the general properties of a protein structure in solution. Following similar analyses of the results as those for the SAXS-guided metadynamics simulation, we did not observe any of the free energy minima to contain structures with low RMSD to the native Trp-cage NMR structure (Supporting Information Figure S1A). Instead, the free energy minima were populated by random coil states with low helical content, which, in turn, prevented identification of the native state by applying the backbone interaction criterion (Supporting Information Figure S1B).

3.3. Comparison to Single-Bias Simulations. In our implementation, the point at the origin represents a perfect fit to the target experimental scattering profile. The fact that the lowest free energy native state of the Trp-cage protein is slightly offset with respect to the origin (Figure 3A) illustrates the strength of our implemented SAXS-guided metadynamics approach. By sampling configurations around the target scattering profile, the developed method allows for configurations along a free energy spectrum to be discriminated. In this way, the risk of non-native solutions is avoided. To test this explicitly, we benchmarked our approach against methods that rely on direct biasing of the potential to drive the simulation toward the experimental target. We performed a 5 ns MD simulation, applying a harmonic potential of 100 kcal mol nm^{-1} centered around the target scattering data. Although a perfect fit (χ^2 = 0.15) to the target scattering was obtained (Figure 5A), the corresponding configuration showed poor agreement with the target NMR structure (RMSD = 7.6 Å).

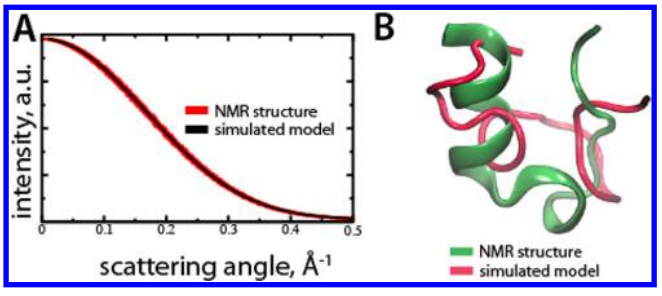


Figure 5. Direct potential biasing to reproduce target scattering profile. (A) Predicted and target X-ray scattering profiles shown in black and red, respectively. (B) Configuration corresponding to the perfect fit to target data (magenta) superimposed on the target NMR structure (green).

DOI: 10.1021/acs.jctc.5b00299
J. Chem. Theory Comput. 2015, 11, 3491−3498

4. DISCUSSION

Metadynamics provides a remarkably efficient approach to refine solution scattering data using collective variables (CVs) that describe intensities at different angles gathered from SAXS experiments. The proof-of-principle system chosen in our studies was the designed Trp-cage protein, which is a popular model system in simulation-based method development using, e.g., replica exchange,[49−51] transition path sampling (TPS),[52] transition interface sampling,[53] and metadynamics[27,39] approaches. By incorporating a set of CVs, each defining a scattering angle in a bias-exchange protocol, accounting for the full scope of the distances describing the experimental data can be achieved. We note that ensembles exploring the free energy landscape with more than one SAXS CV simultaneously provide not only efficient exploration of the free energy landscape but also a meaningful way of representing it. In fact, the resolution of the free energy landscape will depend upon the scattering angles at which the CVs are defined. For instance, while the low-angle CV in our two-dimensional ensemble did not satisfactorily discriminate between low free energy configurations, the high-angle CV resolved these configurations into two separate major free energy minima (Figure 3A). During the simulation, the free energy minimum closest to the origin became significantly more favorable relative to the more distant free energy minimum. In addition, because this global free energy minimum was found closest to the origin, it represented a better fit to the target X-ray scattering profile. The slight offset between the origin and the global free energy minimum can likely be attributed to experimental errors, force field effects, and limited computational sampling. However, given the continual improvements of force fields and ever-increasing computational resources, this offset can, in principle, be used to compensate for inherent experimental errors, such as large error bars in the wide-angle region of scattering data, by finding the solution at the free energy minimum rather than at the origin.

We found that by introducing a cutoff set at the average number of backbone interactions observed during the simulation it was possible to discard random coil solutions (Figure 4A). Although this is not necessary to find the unique solution, similar filtering approaches can be used to assist the structural refinement process. Finally, we note that the free energy maps were constructed with respect to the calculated scattering profile of the Trp-cage NMR structure. It is certainly possible that using the target scattering from a simulated average of the Trp-cage would provide even better discrimination between low-energy configurations. The failure of the control bias-exchange metadynamics simulation to identify the native Trp-cage state implies that the chosen CV set was not ideal for discriminating between random coils and the native state. With a different set of CVs, but with similar computational sampling, Piana et al. successfully explored the folding free energy landscape of the Trp-cage.[27] This highlights the inherent difficulty in selecting CVs that will be able to resolve the free energy landscape of the system or process of interest. In this case, it appears that the SAXS-based CVs were well-suited for discriminating between the native Trp-cage state and random coils.

The proposed atomistic metadynamics approach to refinement of scattering data is valuable in several respects. First, while refinement methods relying on implicit solvent descriptions are computationally less demanding compared to that with atomistic simulations, they come with the caveat of overfitting because parameters describing the solvent effects, such as excluded solvent volume and boundary effects, are not easily determined and hence are used as fitting parameters.[24] In addition, as noted by Zagrovic et al., native-like scattering profiles will also be obtained from structural configurations that are noncompact and not close to the native state.[54] Therefore, simply biasing the potential to drive the simulation toward the experimental target scattering may not be a desirable approach. To test this, we implemented a direct biasing potential designed to force the simulation toward the target scattering. A perfect fit was obtained (χ^2 = 0.15), but the representative structural configuration was not native-like (RMSD = 7.6 Å) (Figure 5). Even so, bias potential refinement methods were recently used to successfully model time-resolved scattering data[55] and conformational transitions.[56] Because both methods start from defined structures that are allowed to undergo transitional dynamics, the possible configurational space to search is significantly more limited compared to finding the native state starting from an unfolded random coil. Therefore, while direct biasing methods are indeed suitable to model time-resolved and transition scattering data, our results suggest that such methods are not appropriate for modeling static solution scattering data.

In an alternative approach to reduce the risk of overfitting, simulation-generated configurations were used to refine statistical weights of ensembles in a structural transition.[57] In this coarse-grained simulation approach, it was possible to discriminate between open and closed conformations of the CHMP3 protein. Our metadynamics approach builds on similar principles to avoid overfitting, but it does not require full sampling of the free energy landscape. Instead, each generated structural configuration will be tagged with a specific free energy with respect to the collective variables defined by scattering intensities at an arbitrary number of scattering angles.

5. CONCLUSIONS

We have developed a metadynamics-based approach for structural refinement of solution scattering data using the Trp-cage as a proof-of-principle system. We expect that our approach can be extended to more complex proteins and used to verify homology models. The protocol was designed to be general, using freely available software (PLUMED and GROMACS), and is easily adjusted to also refine neutron scattering data. The implementation is freely available from the authors upon request.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Bias-exchange metadynamics simulation with non-SAXS CVs. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00299.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: magnus.andersson@scilifelab.se.

Funding

This work was funded by grants from Marie Curie Career Integration Grant (FP7-MC-CIG-618558), Magnus Bergvalls Stiftelse (2014-00170), and Åke Wibergs Stiftelse (M14-0245) to M.A. and from the Swedish e-Science Research Center (SeRC) and Swedish Research Council (2013-5901) to E.L.

DOI: 10.1021/acs.jctc.5b00299
J. Chem. Theory Comput. 2015, 11, 3491−3498

Computational resources were provided by the Swedish National Infrastructure for Computing (2014/11-33).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Koch, M. H.; Vachette, P.; Svergun, D. I. Small-Angle Scattering: A View on the Properties, Structures and Structural Changes of Biological Macromolecules in Solution. *Q. Rev. Biophys.* **2003**, *36*, 147−227.

(2) Rambo, R. P.; Tainer, J. A. Bridging the Solution Divide: Comprehensive Structural Analyses of Dynamic RNA, DNA, and Protein Assemblies by Small-Angle X-ray Scattering. *Curr. Opin. Struct. Biol.* **2010**, *20*, 128−37.

(3) Hu, S. H.; Whitten, A. E.; King, G. J.; Jones, A.; Rowland, A. F.; James, D. E.; Martin, J. L. The Weak Complex between RhoGAP Protein ARHGAP22 and Signal Regulatory Protein 14-3-3 Has 1:2 Stoichiometry and a Single Peptide Binding Mode. *PLoS One* **2012**, *7*, e41731.

(4) Giehm, L.; Svergun, D. I.; Otzen, D. E.; Vestergaard, B. Low-Resolution Structure of a Vesicle Disrupting α-Synuclein Oligomer That Accumulates During Fibrillation. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3246−51.

(5) Makowski, L.; Gore, D.; Mandava, S.; Minh, D.; Park, S.; Rodi, D. J.; Fischetti, R. F. X-ray Solution Scattering Studies of the Structural Diversity Intrinsic to Protein Ensembles. *Biopolymers* **2011**, *95*, 531−42.

(6) Williams, G. J.; Williams, R. S.; Williams, J. S.; Moncalian, G.; Arvai, A. S.; Limbo, O.; Guenther, G.; SilDas, S.; Hammel, M.; Russell, P.; Tainer, J. A. ABC ATPase Signature Helices in Rad50 Link Nucleotide State to Mre11 Interface for DNA Repair. *Nat. Struct. Mol. Biol.* **2011**, *18*, 423−31.

(7) Mosbaek, C. R.; Konarev, P. V.; Svergun, D. I.; Rischel, C.; Vestergaard, B. High Concentration Formulation Studies of an IgG2 Antibody Using Small Angle X-ray Scattering. *Pharm. Res.* **2012**, *29*, 2225−35.

(8) Nygaard, J.; Munch, H. K.; Thulstrup, P. W.; Christensen, N. J.; Hoeg-Jensen, T.; Jensen, K. J.; Arleth, L. Metal Ion Controlled Self-Assembly of a Chemically Reengineered Protein Drug Studied by Small-Angle X-ray Scattering. *Langmuir* **2012**, *28*, 12159−70.

(9) Graewert, M. A.; Svergun, D. I. Impact and Progress in Small and Wide Angle X-ray Scattering (SAXS and WAXS). *Curr. Opin. Struct. Biol.* **2013**, *23*, 748−54.

(10) Makowski, L.; Rodi, D. J.; Mandava, S.; Devarapalli, S.; Fischetti, R. F. Characterization of Protein Fold by Wide-Angle X-ray Solution Scattering. *J. Mol. Biol.* **2008**, *383*, 731−44.

(11) Yang, S.; Park, S.; Makowski, L.; Roux, B. A Rapid Coarse Residue-Based Computational Method for X-ray Solution Scattering Characterization of Protein Folds and Multiple Conformational States of Large Protein Complexes. *Biophys. J.* **2009**, *96*, 4449−63.

(12) Kynde, S. A.; Skar-Gislinge, N.; Pedersen, M. C.; Midtgaard, S. R.; Simonsen, J. B.; Schweins, R.; Mortensen, K.; Arleth, L. Small-Angle Scattering Gives Direct Structural Information about a Membrane Protein Inside a Lipid Environment. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, *70*, 371−83.

(13) Pérez, J.; Koutsioubas, A. Memprot: A Program To Model the Detergent Corona around a Membrane Protein Based on SEC-SAXS Data. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2015**, *71*, 86−93.

(14) Cammarata, M.; Levantino, M.; Schotte, F.; Anfinrud, P. A.; Ewald, F.; Choi, J.; Cupane, A.; Wulff, M.; Ihee, H. Tracking the Structural Dynamics of Proteins in Solution Using Time-Resolved Wide-Angle X-ray Scattering. *Nat. Methods* **2008**, *5*, 881−6.

(15) Takala, H.; Bjorling, A.; Berntsson, O.; Lehtivuori, H.; Niebling, S.; Hoernke, M.; Kosheleva, I.; Henning, R.; Menzel, A.; Ihalainen, J. A.; Westenhoff, S. Signal Amplification and Transduction in Phytochrome Photosensors. *Nature* **2014**, *509*, 245−8.

(16) Andersson, M.; Malmerberg, E.; Westenhoff, S.; Katona, G.; Cammarata, M.; Wohri, A. B.; Johansson, L. C.; Ewald, F.; Eklund, M.; Wulff, M.; Davidsson, J.; Neutze, R. Structural Dynamics of Light-Driven Proton Pumps. *Structure* **2009**, *17*, 1265−75.

(17) Malmerberg, E.; Omran, Z.; Hub, J. S.; Li, X.; Katona, G.; Westenhoff, S.; Johansson, L. C.; Andersson, M.; Cammarata, M.; Wulff, M.; van der Spoel, D.; Davidsson, J.; Specht, A.; Neutze, R. Time-Resolved WAXS Reveals Accelerated Conformational Changes in Iodoretinal-Substituted Proteorhodopsin. *Biophys. J.* **2011**, *101*, 1345−53.

(18) Arnlund, D.; Johansson, L. C.; Wickstrand, C.; Barty, A.; Williams, G. J.; Malmerberg, E.; Davidsson, J.; Milathianaki, D.; DePonte, D. P.; Shoeman, R. L.; Wang, D.; James, D.; Katona, G.; Westenhoff, S.; White, T. A.; Aquila, A.; Bari, S.; Berntsen, P.; Bogan, M.; van Driel, T. B.; Doak, R. B.; Kjaer, K. S.; Frank, M.; Fromme, R.; Grotjohann, I.; Henning, R.; Hunter, M. S.; Kirian, R. A.; Kosheleva, I.; Kupitz, C.; Liang, M.; Martin, A. V.; Nielsen, M. M.; Messerschmidt, M.; Seibert, M. M.; Sjohamn, J.; Stellato, F.; Weierstall, U.; Zatsepin, N. A.; Spence, J. C.; Fromme, P.; Schlichting, I.; Boutet, S.; Groenhof, G.; Chapman, H. N.; Neutze, R. Visualizing a Protein Quake with Time-Resolved X-ray Scattering at a Free-Electron Laser. *Nat. Methods* **2014**, *11*, 923−6.

(19) Neutze, R.; Moffat, K. Time-Resolved Structural Studies at Synchrotrons and X-ray Free Electron Lasers: Opportunities and Challenges. *Curr. Opin. Struct. Biol.* **2012**, *22*, 651−9.

(20) Svergun, D.; Barberato, C.; Koch, M. H. J. Crysol—a Program To Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28*, 768−773.

(21) Liu, H.; Morris, R. J.; Hexemer, A.; Grandison, S.; Zwart, P. H. Computation of Small-Angle Scattering Profiles with Three-Dimensional Zernike Polynomials. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2012**, *68*, 278−85.

(22) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. Accurate SAXS Profile Computation and Its Assessment by Contrast Variation Experiments. *Biophys. J.* **2013**, *105*, 962−74.

(23) Azuara, C.; Orland, H.; Bon, M.; Koehl, P.; Delarue, M. Incorporating Dipolar Solvents with Variable Density in Poisson−Boltzmann Electrostatics. *Biophys. J.* **2008**, *95*, 5587−5605.

(24) Grishaev, A.; Guo, L.; Irving, T.; Bax, A. Improved Fitting of Solution X-ray Scattering Data to Macromolecular Structures and Structural Ensembles by Explicit Water Modeling. *J. Am. Chem. Soc.* **2010**, *132*, 15484−6.

(25) Park, S.; Bardhan, J. P.; Roux, B.; Makowski, L. Simulated X-ray Scattering of Protein Solutions Using Explicit-Solvent Models. *J. Chem. Phys.* **2009**, *130*, 134114.

(26) Chen, P. C.; Hub, J. S. Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data. *Biophys. J.* **2014**, *107*, 435−47.

(27) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553−9.

(28) Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. Characterization of the Free-Energy Landscapes of Proteins by NMR-Guided Metadynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6817−22.

(29) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. Gromacs 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(30) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(31) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.

(32) Case, D.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(33) Warren, B. E. *X-ray Diffraction*; Dover Publications: New York, 1990.

DOI: 10.1021/acs.jctc.5b00299
J. Chem. Theory Comput. 2015, 11, 3491−3498

(34) Fraser, R. D. B.; MacRae, T. P.; Suzuki, E. An Improved Method for Calculating the Contribution of Solvent to the X-ray Diffraction Pattern of Biological Molecules. *J. Appl. Crystallogr.* **1978**, *11*, 693−694.
(35) Laio, A.; Gervasio, F. Metadynamics: A Method To Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71*, 126601.
(36) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(37) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R.; Parrinello, M. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(38) Ferrarotti, M.; Bottaro, S.; Pérez-Villa, A.; Bussi, G. Accurate Multiple Time Step in Biased Molecular Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 139−146.
(39) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2009**, *5*, e1000452.
(40) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to Sh3 and Gb1. *J. Chem. Theory Comput.* **2009**, *5*, 2197−2201.
(41) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(42) Xu, W.; Mu, Y. Ab Initio Folding Simulation of Trpcage by Replica Exchange with Hybrid Hamiltonian. *Biophys. Chem.* **2008**, *137*, 116−25.
(43) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(44) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(45) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. Solvent-Induced Forces between Two Hydrophilic Groups. *J. Phys. Chem.* **1994**, *98*, 2198−2202.
(46) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-Residue Protein. *Nat. Struct. Biol.* **2002**, *9*, 425−30.
(47) Delano, W. L. *The Pymol Molecular Graphics System*; Schrödinger: New York. http://www.pymol.org.
(48) Heyer, L. J.; Kruglyak, S.; Yooseph, S. Exploring Expression Data: Identification and Analysis of Coexpressed Genes. *Genome Res.* **1999**, *9*, 1106−1115.
(49) Zhou, R. Trp-Cage: Folding Free Energy Landscape in Explicit Water. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280−13285.
(50) Paschek, D.; Nymeyer, H.; Garcia, A. E. Replica Exchange Simulation of Reversible Folding/Unfolding of the Trp-Cage Miniprotein in Explicit Solvent: On the Structure and Possible Role of Internal Water. *J. Struct. Biol.* **2007**, *157*, 524−33.
(51) Beck, D. A.; White, G. W.; Daggett, V. Exploring the Energy Landscape of Protein Folding Using Replica-Exchange and Conventional Molecular Dynamics Simulations. *J. Struct. Biol.* **2007**, *157*, 514−23.
(52) Juraszek, J.; Bolhuis, P. G. Sampling the Multiple Folding Mechanisms of Trp-Cage in Explicit Solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15859−15864.
(53) Juraszek, J.; Bolhuis, P. G. Rate Constant and Reaction Coordinate of Trp-Cage Folding in Explicit Water. *Biophys. J.* **2008**, *95*, 4246−4257.
(54) Zagrovic, B.; Pande, V. S. Simulated Unfolded-State Ensemble and the Experimental Nmr Structures of Villin Headpiece Yield Similar Wide-Angle Solution X-ray Scattering Profiles. *J. Am. Chem. Soc.* **2006**, *128*, 11742−11743.
(55) Bjorling, A.; Niebling, S.; Marcellini, M.; van der Spoel, D.; Westenhoff, S. Deciphering Solution Scattering Data with Experimentally Guided Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 780−787.
(56) Chen, P. C.; Hub, J. S. Interpretation of Solution X-ray Scattering by Explicit-Solvent Molecular Dynamics. *Biophys. J.* **2015**, *108*, 2573−2584.
(57) Rozycki, B.; Kim, Y. C.; Hummer, G. SAXS Ensemble Refinement of ESCRT-III CHMP3 Conformational Transitions. *Structure* **2011**, *19*, 109−116.

DOI: 10.1021/acs.jctc.5b00299
J. Chem. Theory Comput. 2015, 11, 3491−3498