JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

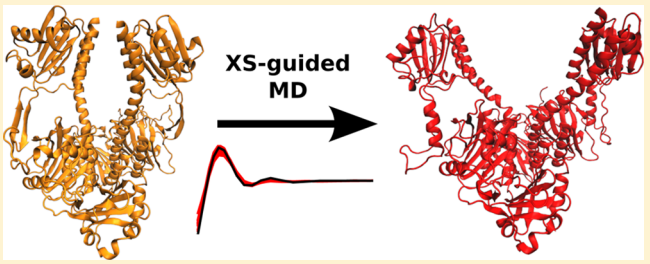
Deciphering Solution Scattering Data with Experimentally Guided Molecular Dynamics Simulations

Alexander Björling,[†] Stephan Niebling,[†] Moreno Marcellini,[‡] David van der Spoel,*[,‡] and Sebastian Westenhoff*[,†]

[†]Department of Chemistry and Molecular Biology, University of Gothenburg, SE-405 30 Gothenburg, Sweden

[‡]Uppsala Center for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, SE-751 05 Uppsala, Sweden

ABSTRACT: Time-resolved X-ray solution scattering is an increasingly popular method to measure conformational changes in proteins. Extracting structural information from the resulting difference X-ray scattering data is a daunting task. We present a method in which the limited but precious information encoded in such scattering curves is combined with the chemical knowledge of molecular force fields. The molecule of interest is then refined toward experimental data using molecular dynamics simulation. Therefore, the energy landscape is biased toward conformations that agree with experimental data. We describe and verify the method, and we provide an implementation in GROMACS.



1. INTRODUCTION

Solution X-ray scattering techniques are important for studying the conformations of molecules in solution.[1] Time-resolved X-ray scattering has been developed to study the structural dynamics of molecular reactions. In the experiment, a reaction is triggered and scattering is recorded as a function of reaction time. Time-resolved data are displayed and interpreted as difference scattering patterns, where the scattered intensity of a reference state is subtracted from that of a certain time point during the reaction.

Initially, time-resolved scattering studies focused on photoreactions of small molecules.[2−5] A few years later, the technique was also applied to proteins. First, time-resolved scattering during carboxy-hemoglobin photolysis[6] and the photoreaction of bacteriorhodopsin and proteorhodopsin[7] were reported. Several other investigations were conducted on the structural dynamics of, for example, myoglobin,[8] photoactive yellow protein,[9] and a phytochrome.[10] These studies were carried out at synchrotron sources and covered time scales from 100 ps to seconds. The availability of ultrashort X-ray bursts at free electron lasers has enabled time-resolved scattering studies with femtosecond time resolution.[11]

Today, time-resolved X-ray scattering experiments on proteins can be carried out reliably, but they are usually limited by data interpretation. Difference scattering reflects a difference in pair distribution functions and thus encodes the structural change of the molecules.[5] However, the information content in the one-dimensional spherically averaged (isotropic) scattering curves is inherently low. As in conventional small-angle X-ray scattering (SAXS), a curve typically contains only tens of independent data points.[12,13] Unlike SAXS, difference X-ray scattering has the advantage that the protein hydration shell typically only gives a minor contribution to the scattering signal.[14] A drawback is that a resting state structure, the starting point of structural change, is required for structural interpretation.

Structural rearrangements can be uniquely identified from difference scattering data, when only a few atoms participate in the reaction.[2−5] For proteins with many hundreds, often thousands, of atoms, fitting atomistic models without additional structural information is impossible. Earlier attempts to surmount this challenge used rigid body refinement,[7,15] dynamic annealing of pseudo atoms,[9] and selection of suitable frames from molecular simulations.[10,11] In all cases, many candidate structures are generated, the scattering of each is computed, and the best fits to the experimental scattering data are identified. Thus, sampling and comparison with experimental data are done in sequence. Since the sampling in conformational space is usually limited, uniqueness of the solutions is not guaranteed. The ideal method to identify structural change from difference scattering data would be to harvest its limited information content while simultaneously taking chemical information, such as bonded and nonbonded interactions, into account.

Biomolecular force fields encode such chemical information, and are used in molecular dynamics (MD) or Monte Carlo (MC) simulations that sample thermal ensembles of proteins in solution. In structural biology, many experimental techniques are increasingly aided by molecular simulation for data interpretation. Refinement by simulation of protein structure based on X-ray diffraction (XRD) data is now common-

Received: October 31, 2014
Published: January 13, 2015

DOI: 10.1021/ct5009735
J. Chem. Theory Comput. 2015, 11, 780−787

place.[16−21] In NMR spectroscopy, the use of MD with experimental restraints has been even more instrumental,[18,20,22−24] since such experiments have historically not supplied a sufficient amount of data to determine a protein structure from scratch. The use of data from small-molecule crystal structures,[25] restraints found from sequence and chemical shift homology,[26] and molecular mechanics force fields[27] supplied the missing information. More recently, orientational restraints from residual dipolar couplings have been used to refine protein structures from NMR data[28,29] and to evaluate the performance of unrestrained simulations.[30]

Conventional SAXS data are usually interpreted by *ab initio* reconstruction of low-resolution envelopes or by assembly of rigid high-resolution elements.[31,32] These strategies do not benefit from the sampling power of molecular simulation. They also do not use chemical knowledge beyond rudimentary considerations to identify likely conformations. A notable exception is the attempt by Grishaev et al. to improve NMR refinement by also taking SAXS intensities into account.[33] As with NMR data, absolute SAXS intensities have also been used for after-the-fact validation of MD trajectories.[34] To our knowledge, no systematic attempts to interpret difference X-ray scattering data with atomistic force fields have been published to date.

Here, we describe and implement X-ray scattering-guided molecular dynamics simulations (XS-guided MD), which is a technique that refines structures against difference scattering data. The method can be used to guide proteins in MD simulations from one structure to another, and therefore is well-suited for elucidating the structural dynamics of biomolecules. In what follows, we shall only be concerned with difference X-ray scattering data, and as explained and justified below, neglect the scattering contribution from the protein's solvation layer.

The method uses an additional pseudo-energy term, based on the Debye scattering equation, which guides the simulation toward states that reproduce target data. The energy landscape of the simulation is biased toward the experimentally observed conformation and distinctions can be made between alternative conformations with similar free energies. As a result, conformers, which are not normally seen in simulations—either because they are transient species observed in dynamic experiments, because they are kinetically inaccessible or because they are incorrectly disfavored by the force field—can potentially be observed and characterized in MD simulations. The algorithm reduces the computational time to determine the structures that agree with experiment, and retains the chemical knowledge contained in MD force fields. The method is implemented in the popular and freely available GROMACS simulations package and is open to further development.

2. THEORY AND METHODS

2.1. X-ray Scattering. Identical but randomly oriented collections of spherical scatterers give rise to a scattered intensity described by the Debye formula.[35]

$$S(q) = \sum_i \sum_j f_i(q) f_j(q) \frac{\sin qr_{ij}}{qr_{ij}} \tag{1}$$

Here, f_i is the scattering factor for scatterer i,[36] q is the magnitude of the scattering vector ($q = 4\pi \sin\theta/\lambda$, with 2θ the scattering angle), and $r_{ij} = |\mathbf{r}_{ij}| = |\mathbf{r}_i - \mathbf{r}_j|$, with $\mathbf{r}_i$ the coordinate vector for particle i. For molecules, the sums typically run over all atoms.

For an actual protein solution, one typically measures the excess scattering of the solution compared to the solvent. In such cases, eq 1 must be corrected for the solvent displaced by the protein. This can be approximately done by using appropriately modified scattering factors $f_i(q)$.[37]

Equation 1 ignores the fact that the solvent and electrolyte densities around a macromolecule are generally different from their bulk values, which, in principle, leads to additional scattering terms. To our knowledge, taking this into account involves either (*i*) making extensive and explicit simulations for each scattering evaluation[34,38,39] or (*ii*) making *ad hoc* assumptions about excess surface densities.[40,41] We ignore this effect, which is an approximation that appears to be reasonable when the data are recorded as differences between two states, $\Delta S(q) = S(q) - S_{\mathrm{ref}}(q)$, as is the case with time-resolved data. Then, solvation contributions largely cancel.[14]

2.2. Force Calculation. For biasing MD simulations toward configurations that agree with experimental scattering curves, following Ahn *et al.*,[15] an additional term (U_{XS}) is added to the usual MD potential energy (U_{MD}):

$$U = U_{\mathrm{MD}} + U_{\mathrm{XS}} \tag{2}$$

$$= U_{\mathrm{MD}} + \left(\frac{k_\chi}{2}\right)\chi^2 \tag{3}$$

$$\chi^2 = \sum_q \frac{1}{\sigma_q^2}[\Delta S_{\mathrm{exp}}(q) - \alpha\{S_{\mathrm{calc}}(q) - S_{\mathrm{ref}}(q)\}]^2$$

$$= \sum_q \frac{1}{\sigma_q^2}[\Delta S_{\mathrm{exp}}(q) - \alpha S_{\mathrm{calc}}(q) + \alpha S_{\mathrm{ref}}(q)]^2 \tag{4}$$

Here, k_χ and σ_q are weighting factors for the scattering bias and for every q-point, respectively. $S_{\mathrm{calc}}(q)$ is the scattering computed from the current state of the simulation. For refinement against *difference* data, $\Delta S_{\mathrm{exp}}(q)$ is the experimental difference curve, $S_{\mathrm{ref}}(q)$ is the scattering from the starting structure and α is the fraction of the observed sample that undergoes conformational change. For refinement against *absolute* data, $\Delta S_{\mathrm{exp}}(q) = 0$ and $\alpha = 1$ while $S_{\mathrm{ref}}(q)$ is the experimental target scattering.

The total force on a particle k is

$$\mathbf{F}_k = -\nabla_k U = -\nabla_k U_{\mathrm{MD}} - \left(\frac{k_\chi}{2}\right)\nabla_k \chi^2 \tag{5}$$

The first term on the right is given by the force field, so the present problem reduces to finding $\nabla_k \chi^2$. The initial structure and experimental data are the same throughout a simulation, so $S_{\mathrm{ref}}(q)$ and $\Delta S_{\mathrm{exp}}(q)$ are constant and $\nabla_k S_{\mathrm{ref}}(q) = \nabla_k \Delta S_{\mathrm{exp}}(q) = 0$. Thus,

$$\nabla_k \chi^2 = \sum_q \frac{1}{\sigma_q^2}\nabla_k[\Delta S_{\mathrm{exp}}(q) - \alpha S_{\mathrm{calc}}(q) + \alpha S_{\mathrm{ref}}(q)]^2$$

$$= -2\alpha \sum_q \frac{1}{\sigma_q^2}[\Delta S_{\mathrm{exp}}(q) - \alpha\{S_{\mathrm{calc}}(q) - S_{\mathrm{ref}}(q)\}]\nabla_k S_{\mathrm{calc}}(q) \tag{6}$$

The last step is to determine the vector $\nabla_k S_{\mathrm{calc}}(q)$, which, by taking the appropriate derivatives of eq 1, comes out as

DOI: 10.1021/ct5009735
J. Chem. Theory Comput. 2015, 11, 780−787

$$\nabla_k S_{\mathrm{calc}}(q) = 2\sum_j f_k(q) f_j(q)\left[\cos qr_{kj} - \frac{\sin qr_{kj}}{qr_{kj}}\right]\frac{\mathbf{r}_{kj}}{r_{kj}^2} \quad (7)$$

Finally, the force contribution from the scattering energy term acting on particle k is obtained by combining eqs 5, 6, and 7:

$$\mathbf{F}_k^{\mathrm{scat}} = 2k_\chi\alpha\left\{\sum_q \frac{1}{\sigma_q^2}[\Delta S_{\mathrm{exp}}(q) - \alpha\{S_{\mathrm{calc}}(q) - S_{\mathrm{ref}}(q)\}] \times \sum_j f_k(q) f_j(q)\left[\cos qr_{kj} - \frac{\sin qr_{kj}}{qr_{kj}}\right]\frac{\mathbf{r}_{kj}}{r_{kj}^2}\right\} \quad (8)$$

Calculating these forces amounts to evaluating two double sums over all scatterers, first for $\{S_{\mathrm{calc}}(q) - S_{\mathrm{ref}}(q)\}$, as in eq 1, then for the forces, as in eq 8.

For convenience, we define

$$A_q = \frac{k_\chi}{2\sigma_q^2}[\Delta S_{\mathrm{exp}}(q) - \alpha\{S_{\mathrm{calc}}(q) - S_{\mathrm{ref}}(q)\}] \quad (9)$$

which gives

$$\mathbf{F}_k^{\mathrm{scat}} = 4\alpha\sum_q\left\{A_q\sum_j f_k(q) f_j(q)\left[\cos qr_{kj} - \frac{\sin qr_{kj}}{qr_{kj}}\right]\frac{\mathbf{r}_{kj}}{r_{kj}^2}\right\} \quad (10)$$

2.3. Scaling of Experimental Data. When evaluating the Debye equation, the resulting intensity carries units given by the atomic scattering factors. These are often given in "electron units" (e.u.), that is, as relative scattering amplitudes referenced to a classical unit point charge.[35] If concentrations, path lengths, incoming X-ray intensities, and detector efficiencies are precisely known, experimentally measured intensities can be converted to such units. However, this is often not the case in practice, and experimental data are instead scaled to predicted curves. For absolute data, this is straightforward. Either the experimental curve is scaled to each calculated profile, or it is rescaled such that the experimental extrapolation of $S(q = 0)$ matches the calculated value. The latter is independent of conformation (see eq 1).

For time-resolved difference measurements scaling is more problematic and involves a parameter α (eq 4). If the experimental system has pure initial and final states with $I(q) = S_i$ and S_f, and is initially distributed so that the relative populations over these states are α_0 and $1 - \alpha_0$, then with a relative yield α,

$$\Delta S_{exp}(q) = \underbrace{(\alpha + \alpha_0)S_f + (1 - \alpha - \alpha_0)S_i}_{\text{final scattering}} - \underbrace{[\alpha_0 S_f + (1 - \alpha_0)S_i]}_{\text{initial scattering}} = \alpha(S_f - S_i) \quad . \quad (11)$$

provided that the input $\Delta S_{\mathrm{exp}}(q)$ has meaningful units (this can be approximately achieved by scaling the detector reading to $S_{\mathrm{calc}}(q)$ and then scaling $\Delta S_{\mathrm{exp}}(q)$ by the same number), the parameter α in eq 4 can be identified with the relative yield of the difference experiment.

2.4. Implementation. XS-guided MD is implemented in GROMACS 5 and, therefore, profits from the efficient core of GROMACS. Moreover, it can, in principle, be used in combination with any other featured algorithm, for example, with the Simulated Annealing Method or Replica Exchange Method, to improve sampling. The current implementation supports parallelization using OpenMP.

The Debye scattering terms are represented as bonded interactions of a new type (pairs type 3). Each such interaction is explicitly specified in the topology. Any atom or virtual interaction site can act as a scatterer, leaving it up to the user to design an appropriate simulation. The Debye summation is an N^2 problem that becomes computationally demanding for large biomolecules. In some cases, coarse-graining the scattering calculation[14,42] or even the molecular representation[43] is appropriate. The code allows for combinations of these choices, and we provide a tool to expedite topology construction. To speed up simulations further, the slowly varying scattering forces can be calculated less often than the other forces. We have implemented a dual time-step algorithm, as described in detail in the GROMACS 5.0 manual.

We provide implementations of atomic scattering factors with and without a displaced-solvent correction,[36,37] commonly used in SAXS intensity prediction tools.[40,41] For flexibility, we also implement the library-averaged and coarse-grained scattering factors calculated by Niebling et al.[14] for amino acid[42] and MARTINI-bead[43] scatterers.

In addition to structural refinement, the user can calculate Debye sums from single frames or of entire trajectories, using a tool shipped with the code.

2.5. Computational Details. All simulations were performed using GROMACS 5.0, with the XS-guided MD implementation added. The Verlet cutoff scheme was used throughout. A constant temperature of 300 K was achieved in all simulations using a modified Berendsen thermostat[44] with $\tau = 0.5$. All bonds were constrained using the LINCS algorithm.

Dibutyl Ether. Dibutyl ether was simulated in vacuum, using the OPLS/AA force field.[45] A periodic box was used, with electrostatic cutoffs and a time step of 2 fs. The 10 ns equilibrium simulation was run directly from the initial model with velocities drawn from a Maxwell distribution corresponding to 300 K. The first 500 ps were considered equilibration and removed from the analysis. XS-guided MD simulations were run with atomistic vacuum form factors[36] against scattering data predicted using the Debye equation (eq 1).

LAO Lysine Ligation. Simulations started from crystal structures of the *apo* and *holo* forms of LAO, with PDB codes 2LAO and 1LST,[46] respectively. Residues 1−238 were included. Missing atoms were filled in using the WHAT IF web server (http://swift.cmbi.ru.nl/whatif/). The CHARMM27 force field[47,48] was used for all LAO simulations. A periodic box was used, and electrostatics were treated with PME, using fourth-order interpolation and a grid spacing of 0.12 nm.

The initial models were converted and protonated with the GROMACS tool pdb2gmx. Using cubic simulation boxes initially 2 nm larger than the protein in each direction, the starting structures were first energy-minimized to a maximum force of 2000 kJ/mol/nm. The systems were energy-minimized again after the addition of TIP3P water and ions corresponding to ∼50 mM NaCl (in addition to two or three cations to neutralize the protein). The simulation boxes were equilibrated under NVT conditions for 50 ps, then subsequently under NPT conditions for 500 ps using the Parrinello−Rahman barostat ($\tau = 2$ ps, $P = 1$ bar). All non-hydrogen atoms were position-restrained to their initial positions during equilibration (force constants = 1000 kJ/mol/nm^2).

DOI: 10.1021/ct5009735
J. Chem. Theory Comput. 2015, 11, 780−787

Equilibrium simulations covering 100 ns were carried out for the *apo* and *holo* forms, the latter both with and without bound lysine.

The XS-guided MD simulations were carried out in a a larger cubic box, with dimensions of 13.5 nm, separately minimized and equilibrated as above. The starting structure was the same as for the lysine-free *holo* simulation. Amino acid–based scattering factors[14,42] were used, centered on virtual interaction sites representing the center of mass of each residue. The target scattering was calculated from the crystal structures using the Debye equation (eq 1).

Phytochrome Photoconversion. XS-guided MD simulations of phytochrome photoconversion started from the crystal structure with PDB code 4O01.[10] The CHARMM27 force field[47,48] was used, with the chromophore co-factor added by manually adapting the parameters published by Kaminski et al.[49] to the GROMACS format. A time step of 2 fs was used for these simulations.

The initial structure was energy-minimized, solvated, and equilibrated as described for LAO above, in a 25 nm cubic box, with ions corresponding to 100 mM NaCl in addition to the 48 cations necessary to neutralize the protein. XS-guided MD simulations used amino-acid scattering factors centered on the C-beta atoms, and were run against previously published experimental difference scattering data.[10]

3. RESULTS

The implementation of XS-guided MD described above was validated against three test systems. The first is an illustrative theoretical experiment which explores how the conformational distribution of a small organic molecule can be biased toward reproducing theoretical scattering curves. The second is more complex, as a real protein movement is reproduced by providing theoretical scattering data corresponding to two observed crystal structures. The third case involves actual difference scattering data, and shows that XS-guided MD reproduces earlier results that were based on a knowledge-based *ad hoc* approach. In the following, we review the results and discuss the limitations and possibilities of the method.

3.1. Dibutyl Ether. The linear dibutyl ether molecule was first simulated without experimental constraints under vacuum for 10 ns, and the degree of openness of the carbon–oxygen chain considered. Figure 1A shows examples of the conformations encountered, and the distribution of end-to-end distances (R_{C1-C8}) is shown in Figures 1B and 1C (black curve).

After calculating the theoretical scattering curves of an "open" dibutyl ether conformation (R_{C1-C8} = 9.8 Å) and of a "closed" (R_{C1-C8} = 5.0 Å) dibutyl ether conformation, these curves were used as input for XS-guided MD. Figures 1B and 1C show that the artificial scattering energy (U_{XS}) has a dramatic effect on the end-to-end distance distribution. For both simulations, increasing the coupling coefficient k_χ skews the distance distribution toward the target structure. At high coupling, all conformations that do not display the degree of openness of the target conformation are avoided. Meanwhile, the calculated scattering curves also approach their targets, as shown in the insets of Figures 1B and 1C.

This establishes that the method and implementation work for this simple test system, where one degree of freedom essentially determines the scattering profile.

3.2. LAO Lysine Ligation. The lysine/arginine/ornithine-binding protein (LAO) from *Salmonella typhimurium* undergoes large-scale domain movements when binding ligands.[46,50] Specifically, lysine binding causes one of its domains to rotate 52° about an axis formed by two hinge points located on adjacent beta strand termini. The movement is shown in Figure 2D. Reproducing this structural change constitutes a more complex, albeit hypothetical, test case for the XS-guided MD method.

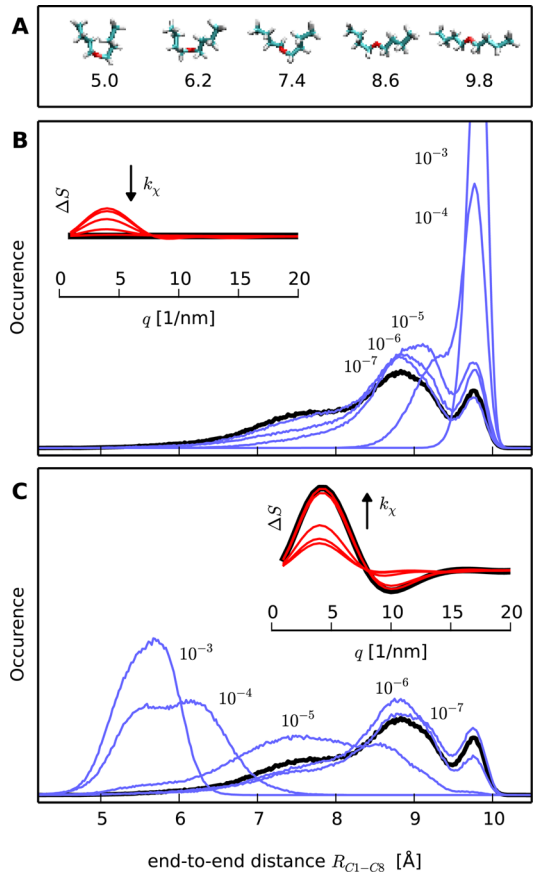


Figure 1. Dibutyl ether guided toward an open or a closed conformation based on calculated X-ray scattering. (A) Example conformations from an unrestrained simulation, with various end-to-end distances in Å, chosen at random for each distance. (B, C) Distance distributions of an unrestrained simulation (black) and of XS-guided MD simulations aiming at an open conformation (panel (B), 9.8 Å) and a closed conformation (panel (C), 5.0 Å) with k_χ values as indicated (blue). Insets show average calculated scattering profiles for each k_χ, as a difference relative to the open conformation. There, red curves correspond to different k_χ values, while target curves are shown in black.

Unrestrained MD simulations starting from equilibrated crystal structures show that both the lysine-bound *holo* and the unbound *apo* forms are stable on the 100 ns time scale (Figure 2A). In contrast, with the lysine ligand removed, the *holo* form becomes ill-defined and deviates from its initial conformation without reaching the *apo* state (Figure 2A, bottom panel).

We ran simulations guided by theoretical scattering profiles toward the *apo* state, starting from the lysine-free *holo* structure. The target scattering data was computed as the scattering of the *apo* minus the *holo* state. Figure 2B shows how the structural similarity to both states, as well as the energy term U_{XS}, develop over time at several coupling strengths k_χ. The figure shows that guiding the simulation toward the theoretical target scattering data causes the *holo*-to-*apo* transition to readily occur. As

DOI: 10.1021/ct5009735
J. Chem. Theory Comput. 2015, 11, 780–787

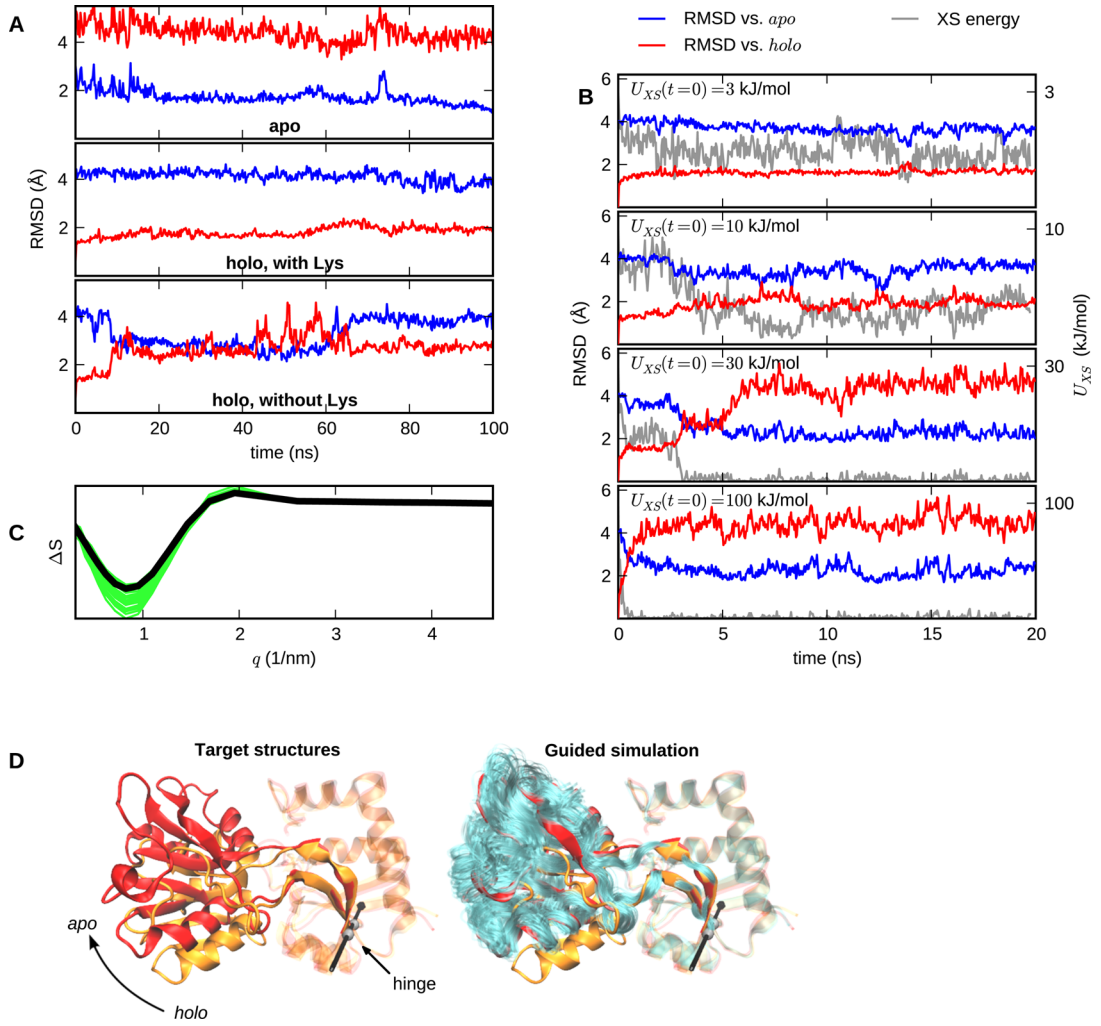


Figure 2. Method validation against the lysine/arginine/ornithine-binding protein (LAO). (A) Unrestrained simulations of the *apo*, *holo*, and lysine-free *holo* states. (B) XS-guided MD trajectories aiming at the *apo* state, starting from the lysine-free *holo* state, with various coupling strengths k_χ. The plot shows RMSD:s compared to initial and target structures, as well as the evolution of the scattering energy. (C) 25 scattering curves, extracted from the second half of the 30 kJ/mol run, together with the target curve (thick line). (D) Graphical representation of the *apo* (red) and *holo* (orange) conformations. The right-hand model also shows a trajectory view over the second half of the 30 kJ/mol run. The hinge axis is indicated.

expected, increasing the coupling parameter k_χ causes the transition to happen earlier in the simulation. The final root-mean-square deviation (RMSD), with respect to the *apo* state, is consistent with the equilibrium *apo* simulation (Figure 2A), at ∼2 Å. Figure 2C confirms that typical difference X-ray scattering patterns from the simulations agree well with the target data. Figure 2D gives a structural view of the XS-guided MD simulation and, again, confirms the *holo*-to-*apo* transition suggested by the RMSD traces.

Interestingly, if even stronger coupling to scattering data is imposed, the simulation fails to find the *apo* state (data not shown). This indicates that the scattering energy term contains barriers that become insurmountable on the time scale simulated here, if k_χ is set too high.

The ability of XS-guided MD to quickly reproduce the *apo* state of LAO, based solely on a difference scattering curve, shows that the curve contains enough information to bias the macromolecular force field toward the correct conformation. It also shows that XS-guided MD provides sufficient sampling to find the correct structure in a relatively short simulation, at least for this particular system.

3.3. Phytochrome Photoconversion.

As a final test system, we applied the XS-guided MD algorithm to the bacterial phytochrome from *Deinococcus radiodurans*. The dimeric photosensory module of this light-sensing protein was recently shown to undergo dramatic structural change during photoconversion.[10] Upon illumination with red light, the dark-adapted form, labeled the Pr state, undergoes a series of Ångström-scale structural transformations that amplify and ultimately cause the homodimer to partially open in a nanometer-scale movement, forming the so-called Pfr state. As a result, the distance between the globular PHY (phytochrome-specific) domains on opposing monomers increases. In ref 10, we reported dark and illuminated crystal and solution structures, which showed that the opening motion is larger in solution than in the crystals.[10]

Simulations guided by X-ray scattering were run starting from the illuminated crystal structure of the photosensory core from the *D. radiodurans* phytochrome.[10] The simulations were biased toward an experimental difference scattering curve corresponding to the Pr-to-Pfr transition, by comparing the calculated scattering profiles of the simulation to that of the

DOI: 10.1021/ct5009735
J. Chem. Theory Comput. 2015, 11, 780−787

proposed Pr solution structure. This data was previously used to propose the solution structure for the illuminated Pfr state.[10]

Figure 3A shows that guiding the simulation toward experimental scattering data causes the distance between the

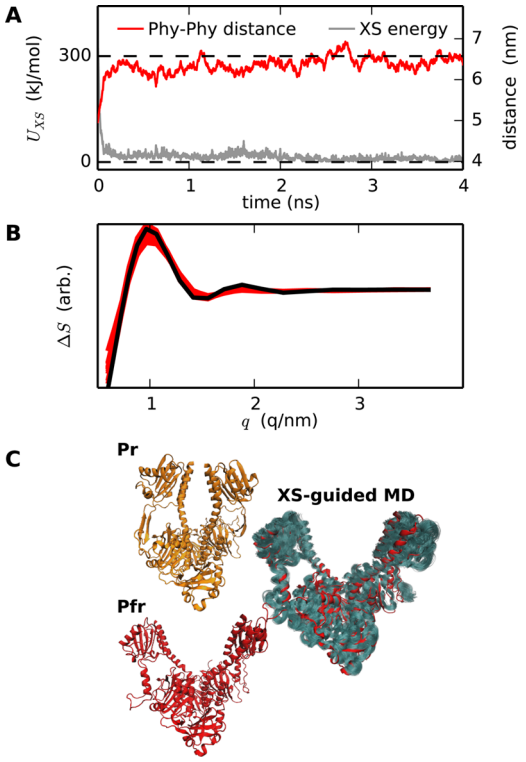


Figure 3. Method validation against the *D. radiodurans* bacterial phytochrome. (A) The center-of-mass distances between the globular PHY domains (residues 330−445 and 480−503) on opposing monomers, as well as the evolution of the scattering energy term; the initial scattering energy was 300 kJ/mol. (B) Theoretical difference scattering during the second half of the 4 ns run. (C) Structural views of the Pr and Pfr solution structures,[10] and a trajectory view over the second half of the run, superimposed on the Pfr target structure.

phytochrome dimer's opposing PHY domains to grow.[51] As the PHY−PHY distance approaches the value previously found,[10] the global conformation of the dimer reproduces the Pfr solution structure. Indeed, the structural overlap shown in Figure 3C illustrates that these structures are identical at the low-resolution level. Thus, XS-guided MD simulations are capable of refining solution structures based on an initial model and experimental X-ray difference scattering data.

We note that, while the application to phytochrome photoconversion validates the XS-guided MD method as such, and it lends confidence to the nature of the previously proposed structural change,[10] the magnitude of the opening between the PHY domains cannot be rigorously defined at very large openings. This was already recognized in ref 10 and does not affect the conlusions drawn in that paper; however, it does illustrate a potential problem in refinement against difference X-ray solution scattering data. The parameter α, which describes the experimental conversion efficiency and is needed for the correct scaling of calculated to experimental data, was estimated based on the original analysis, where α was arbitrarily scaled for best fit to data.[10] In fact, the absolute size of the difference signal, and therefore the precise value of α, affects the degree to which the phytochrome dimer opens up in XS-guided MD refinement. This is illustrated in Figure 4, which shows predicted peak positions for difference scattering curves based

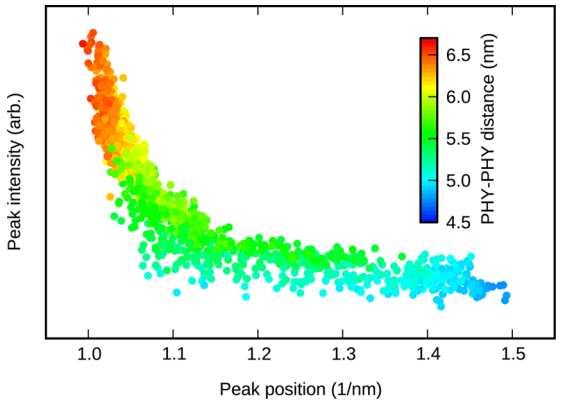


Figure 4. Magnitude and position of the ΔS peak at $q \cong 1$/nm for various PHY−PHY distances. The data are taken from the trajectories presented by Takala *et al.*[10]

on the previously published trajectories.[10] As the dimer opening increases, the peak first shifts along the q-axis, and then, after an initial opening to ∼5.5 nm, instead grows in magnitude. Thus, experimental determination of the yield parameter α is important for successful structural refinement.

4. DISCUSSION

Using three test cases, ranging from small and theoretical to large and experimental, we have shown that XS-guided MD simulations may serve as a tool to structurally interpret difference solution X-ray scattering data. We have focused on the use of *difference* scattering data, as encountered in time-resolved X-ray scattering experiments of proteins in solution. However, the method could, in principle, be equally well-applied to absolute data, provided that attention is paid to the limitations of the Debye equation (eq 1) in predicting absolute X-ray scattering. In contrast to macromolecular crystallography and SAXS, which have seen the development of rigorous methodologies in the last decades,[31,52−55] there is no established way of structurally interpreting difference scattering data. The algorithm presented here may provide a starting point for such development.

The applicability of the method relies on its ability (i) to calculate X-ray scattering with adequate accuracy, (ii) to distinguish between the true target conformation and other conformations using the combined knowledge of the chemical force field and the experimental data, and (iii) to sample enough conformations to find the target structure within a reasonable simulation time.

We found that requirement (iii) is satisfied for all test systems tried here, usually after simulating ∼10 ns. Regarding requirement (i), it is noted that a shortcoming of the Debye equation (eq 1), where the sums run only over the protein atoms, is that it does not account for the solvation layer scattering. However, when computing difference X-ray scattering data, this effect can often be neglected.[14] This is the case for the phytochrome photoconversion investigated here.

The second requirement, uniqueness of the target structure, is more critical. First, we note that a certain feature in a difference scattering pattern is more likely to reflect a unique structural rearrangement when it occurs at q values that

DOI: 10.1021/ct5009735
J. Chem. Theory Comput. 2015, 11, 780−787

describe the molecular envelope (for proteins, this region is typically described by $q \lesssim 2$ nm^{-1}). Difference scattering in this q-range reflects large-scale protein motions. All the test cases used here can be described by collective motions along a small number of degrees of freedom. For example, the phytochrome rearrangement can be thought of as an increase in distance between the PHY domains, while the structural change of LAO can be seen as rotation around an axis. These structural processes affect the overall shape of the protein and the difference scattering pattern of LAO is dominated by peaks at $q \leq 2$ nm^{-1}, while the phytochrome difference scattering has a distinct feature at $q = 1$ nm^{-1}. Consequently, unique structural fits can be obtained.

At higher q values, multiple candidate structures are more likely to produce overlapping features in difference scattering patterns and they must be discriminated using geometrical or chemical constraints. In the intermediate q range (2 nm$^{-1} < q < 8$ nm^{-1}), typical structural dynamics would involve the rearrangement of secondary or tertiary structural elements. Geometric constraints, such as used in crystallography for dihedrals, angles, and bond lengths, are not likely to be effective in discriminating candidate structures. Instead, the full range of covalent, dispersion, and electrostatic interactions described by molecular force fields must be considered, and the resulting energy landscape sampled in the appropriate ensemble. In previous reports, these considerations were made *ad hoc*. For example, helix movements, hand-picked based on previous knowledge, were refined against difference scattering features at 2 nm$^{-1} < q < 6$ nm^{-1}.[7] In the approach presented here, the required interactions and realistic sampling are inherent in the MD simulations. We consider this to be an important step toward unbiased structural interpretation of time-resolved X-ray scattering data.

Finally, we note that some protein structural changes may be difficult to observe with X-ray scattering. For example, the conformational change upon photolysis of the hemoglobin−carbon monoxide bond corresponds to an RMSD value of 4 Å, but has a relatively small effect on the molecular envelope. Therefore, it affects the X-ray scattering signal at low q only weakly. We have tested a refinement of the hemoglobin structural change against the difference scattering patterns published in ref 6; however, even though the scattering energy decreased and stabilized rapidly, the RMSD observables did not reach the target values. This indicates that structures exist, which describe the difference scattering data of hemoglobin well, but which do not agree with structural changes derived from crystallographic structures. The situation arises because the experimental data contain too little information, and the demands on the force field and sampling become too high. In principle, molecular force fields should be able to distinguish between artificial and true target structures; however, in practice, this would require long and repeated simulations, beyond what is practically possible.

CONCLUSION

We expect that our implementation of XS-guided MD constitutes a founding step in the development of a systematic methodology for the structural interpretation of solution scattering data. The program is integrated in the GROMACS software package, which implies that users will have access to the features from the next release.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: spoel@xray.bmc.uu.se (D. van der Spoel).
*E-mail: westenho@chem.gu.se (S. Westenhoff).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by grants from the Foundation of Strategic Research, Sweden, the Swedish Research Council, and the European Research Council (through Agreement No. 279944).

REFERENCES

(1) Mertens, H. D. T.; Svergun, D. I. *J. Struct. Biol.* **2010**, *172*, 128−141.
(2) Neutze, R.; Wouts, R.; Techert, S.; Davidsson, J.; Kocsis, M.; Kirrander, A.; Schotte, F.; Wulff, M. *Phys. Rev. Lett.* **2001**, *87*, No. 195508.
(3) Davidsson, J.; Poulsen, J.; Cammarata, M.; Georgiou, P.; Wouts, R.; Katona, G.; Jacobson, F.; Plech, A.; Wulff, M.; Nyman, G.; Neutze, R. *Phys. Rev. Lett.* **2005**, *94*, No. 245503.
(4) Ihee, H.; Lorenc, M.; Kim, T. K.; Kong, Q. Y.; Cammarata, M.; Lee, J. H.; Bratos, S.; Wulff, M. *Science* **2005**, *309*, 1223−7.
(5) Plech, A.; Wulff, M.; Bratos, S.; Mirloup, F.; Vuilleumier, R.; Schotte, F.; Anfinrud, P. *Phys. Rev. Lett.* **2004**, *92*, No. 125505.
(6) Cammarata, M.; Levantino, M.; Schotte, F.; Anfinrud, P. a.; Ewald, F.; Choi, J.; Cupane, A.; Wulff, M.; Ihee, H. *Nat. Methods* **2008**, *5*, 881−886.
(7) Andersson, M.; Malmerberg, E.; Westenhoff, S.; Katona, G.; Cammarata, M.; Wöhri, A. B.; Johansson, L. C.; Ewald, F.; Eklund, M.; Wulff, M.; Davidsson, J.; Neutze, R. *Structure* **2009**, *17*, 1265−1275.
(8) Cho, H. S.; Dashdorj, N.; Schotte, F.; Graber, T.; Henning, R.; Anfinrud, P. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 7281−7286.
(9) Kim, T. W.; Lee, J. H.; Choi, J.; Kim, K. H.; van Wilderen, L. J.; Guerin, L.; Kim, Y.; Jung, Y. O.; Yang, C.; Kim, J.; Wulff, M.; van Thor, J. J.; Ihee, H. *J. Am. Chem. Soc.* **2012**, *134*, 3145−3153.
(10) Takala, H.; Björling, A.; Berntsson, O.; Lehtivuori, H.; Niebling, S.; Hoernke, M.; Kosheleva, I.; Henning, R.; Menzel, A.; Ihalainen, J. A.; Westenhoff, S. *Nature* **2014**, *509*, 245−248.
(11) 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.; Kjæ r, 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.; Sjöhamn, J.; Stellato, F.; Weierstall, U.; Zatsepin, N. A.; Spence, J. C. H.; Fromme, P.; Schlichting, I.; Boutet, S.; Groenhof, G.; Chapman, H. N.; Neutze, R. *Nat. Methods* **2014**, *11*, 923−926.
(12) Moore, P. B. *J. Appl. Crystallogr.* **1980**, *13*, 168−175.
(13) Svergun, D. I.; Koch, M. H. J. *Rep. Prog. Phys.* **2003**, *66*, 1735−1782.
(14) Niebling, S.; Björling, A.; Westenhoff, S. *J. Appl. Crystallogr.* **2014**, *47*, 1190−1198.
(15) Ahn, S.; Kim, K. H.; Kim, Y.; Kim, J.; Ihee, H. *J. Phys. Chem. B* **2009**, *113*, 13131−13133.
(16) Brünger, A. T.; Kuriyan, J.; Karplus, M. *Science* **1987**, *235*, 458−460.
(17) Brünger, A. T. *X-PLOR: Version 3.1: A System for X-ray Crystallography and NMR*; Yale University Press: New Haven, CT, 1992.
(18) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 905−921.

DOI: 10.1021/ct5009735
J. Chem. Theory Comput. 2015, 11, 780−787

(19) Brunger, A. T.; Adams, P. D. *Acc. Chem. Res.* **2002**, *35*, 404−412.
(20) Brunger, A. T. *Nat. Protocols* **2007**, *2*, 2728−2733.
(21) Soper, A. *Chem. Phys.* **1996**, *202*, 295−306.
(22) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283−98.
(23) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Marius Clore, G. *J. Magn. Reson.* **2003**, *160*, 65−73.
(24) Choy, W. Y.; Forman-Kay, J. D. *J. Mol. Biol.* **2001**, *308*, 1011−1032.
(25) Engh, R. A.; Huber, R. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1991**, *47*, 392−400.
(26) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289−302.
(27) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615−9620.
(28) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 9279−9283.
(29) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471−1475.
(30) Lange, O. F.; van der Spoel, D.; de Groot, B. L. *Biophys. J.* **2010**, *99*, 647−655.
(31) Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I. *J. Appl. Crystallogr.* **2012**, *45*, 342−350.
(32) Graewert, M. A.; Svergun, D. I. *Curr. Opin. Struct. Biol.* **2013**, *23*, 748−754.
(33) Grishaev, A.; Wu, J.; Trewhella, J.; Bax, A. *J. Am. Chem. Soc.* **2005**, *127*, 16621−16628.
(34) Chen, P.-c.; Hub, J. S. *Biophys. J.* **2014**, *107*, 435−447.
(35) Warren, B. E. *X-ray Diffraction*; Dover Publications: New York, 1990.
(36) Doyle, P.; Turner, P. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theory Gen. Crystallogr.* **1968**, *24*, 390−397.
(37) Fraser, R. D. B.; MacRae, T. P.; Suzuki, E. *J. Appl. Crystallogr.* **1978**, *11*, 693−694.
(38) Park, S.; Bardhan, J. *J. Chem. Phys.* **2009**, *130*, 134114.
(39) Grishaev, A.; Guo, L.; Irving, T.; Bax, A. *J. Am. Chem. Soc.* **2010**, *132*, 15484−15486.
(40) Svergun, D.; Barberato, C.; Koch, M. *J. Appl. Crystallogr.* **1995**, 768−773.
(41) Liu, H.; Hexemer, A.; Zwart, P. H. *J. Appl. Crystallogr.* **2012**, *45*, 587−593.
(42) Yang, S.; Park, S.; Makowski, L.; Roux, B. *Biophys. J.* **2009**, *96*, 4449−4463.
(43) Marrink, S.; Risselada, H. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(44) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(45) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. *J. Chem. Theory Comput.* **2012**, *8*, 61−74.
(46) Oh, B. H.; Pandit, J.; Kang, C. H.; Nikaido, K.; Gokcen, S.; Ames, G. F.; Kim, S. H. *J. Biol. Chem.* **1993**, *268*, 11348−11355.
(47) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−15.
(48) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. *J. Chem. Theory Comput.* **2010**, *6*, 459−466.
(49) Kaminski, S.; Daminelli, G.; Mroginski, M. A. *J. Phys. Chem. B* **2009**, *113*, 945−958.
(50) Hayward, S. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 425−435.
(51) Note that the definitions of the PHY−PHY separation in this paper and in ref 10 differ.
(52) Brünger, A. T. *Nature* **1992**, *355*, 472−475.
(53) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 240−255.
(54) Rambo, R. P.; Tainer, J. A. *Nature* **2013**, *496*, 477−481.
(55) Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole, F. L.; Tsutakawa, S. E.; Jenney, F. E.; Classen, S.; Frankel, K. A.; Hopkins, R. C.; Yang, S.-J.; Scott, J. W.; Dillard, B. D.; Adams, M. W. W.; Tainer, J. A. *Nat. Methods* **2009**, *6*, 606−612.

DOI: 10.1021/ct5009735
J. Chem. Theory Comput. 2015, 11, 780−787