Quasiequivalence of multiscale coevolution and ensemble MD simulations: A demonstration with lactoferrin

J. Yang, A. Singharoy, Yu.V. Sereda, P.J. Ortoleva

Center for Theoretical and Computational Nanoscience, Department of Chemistry, Indiana University, Bloomington IN 47405, United States

ARTICLE INFO

Article history:
Received 12 August 2014
In final form 8 October 2014
Available online 16 October 2014

ABSTRACT

Ensemble molecular dynamics computations are performed because a single MD simulation may not represent all the possible scenarios. Ensemble MD consumes a great amount of resources. Here, the similarity of ensemble MD and coupled all-atom, coarse-grained simulation (the multiscale coevolution method) is assessed. Quasiequivalence of two simulations is defined and shown to provide a similarity measure. Quasiequivalence compares trajectories on coarse and relatively finer scales. Good agreement between multiscale coevolution and ensemble MD simulations is demonstrated for lactoferrin. Quasiequivalence along with greater CPU efficiencies of multiscale coevolution relative to ensemble MD, underscore the advantages of multiscale coevolution for nanosystem modeling.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Conventional MD simulations starting with essentially the same initial data can lead to qualitatively different scenarios [1]. Ensemble MD (EMD) is carried out to ensure that the results are representative of the possible scenarios that the system may express for specified conditions. EMD is defined as the synthesis of a set of MD runs; each of the MD runs has the same initial state but with different seeds for the randomization of atomic velocities imposed to simulate isothermal conditions. However, EMD consumes a great amount of computational resources. To address this, a multiscale coevolution approach (implemented as the Deductive Multiscale Simulator, DMS [2,3], see Section 2) was developed to model the dynamics of macromolecular systems via coevolving a set of coarse-grained (CG) variables together with an ensemble of all-atom configurations, thereby, simultaneously enumerating multiple discrete atomistic scenarios [4–8]. Consequently, under practical computational limits, DMS has been suggested to be more statistically significant than a single MD run over similar (long) time period [4,9]. It has been suggested that a DMS simulation is similar to EMD. Here, the nature of this DMS–EMD similarity is qualitatively and quantitatively explored.

Assessing the accuracy of a multiscale simulation relative to conventional MD requires consideration of several factors. Firstly, two traditional MD simulations started with same initial states only differing by a small degree will tend to depart in their trajectories in 6N-dimensional position-momentum space due to the orbital instability of the N-particle Newtonian dynamical systems [10]. Indication of this orbital instability is that the distance between the 6N-dimensional microstates diverges exponentially with time. Furthermore, if the simulations are isothermal, the system is continuously disturbed at random via the Langevin or Nose–Hoover thermostat [11]. In this way, even two traditional MD simulations cannot be readily compared if the criterion for agreement is in terms of the 6N-dimensional trajectory. In addition, the multiscale Langevin simulation in DMS involves statistical errors arising from the size of the ensemble used to construct the thermal-average forces, as well as from the random force in the Langevin equations for the CG variables [8]. Furthermore, there are uncertainties due to the establishment of the CG states, as well as the microstates consistent with it after each of the discrete CG timestep updates (see Section 2.B). With the above, measures of the DMS–EMD similarity must be clearly defined.

Given the above uncertainties in simulations, a concept of ‘quasiequivalence’ of two simulations is introduced for assessing the DMS–EMD similarity. Two simulations are quasiequivalent if the timecourses of the associated CG variables are closely lying, and if the ensembles of relatively finer scale properties (notably here CG velocities) over a characteristic time of system dynamics are similar. The CG velocities fluctuate on a timescale between that of atomic collisions/vibrations and the characteristic time of CG dynamics. The quasiequivalence of an ensemble of all-atom trajectories with multiscale coevolution implies that their evolving CG
probability distributions and free energy landscapes are similar. Thus, quasiequivalence between trajectories suggests similarity on both the CG scale and a scale intermediate between the CG and microscopic timescale, and provides a computationally tractable mean to demonstrate the DMS-EMD equivalence, and to validate the coevolution concepts.

This study focuses on a case that does not involve major barrier crossing events, so that associated complexities in constructing isothermal, isovolumetric trajectories are avoided. The choice of system to demonstrate quasiequivalence in DMS and EMD is made according to the following criteria: (a) the system must be large enough so that the timescale separation between individual atomicistic and overall structural dynamics warrants a multiscale approach; (b) the system must be small enough so that ensemble MD is feasible. However, even if EMD could be run for a large system, the \( N^{-1/2} \) scaling of deviations from average behavior would mask the fluctuation effects which are important for quantifying DMS-EMD quasiequivalence. To respect these criteria, the open-to-close transition in iron binding protein lactoferrin is chosen for demonstration.

2. Method

The multiscale coevolution approach was implemented as DMS. DMS simulations are achieved by coevolving the coarse-grained (CG) and microscopic states. This accounts for the exchange of information across multiple spatiotemporal scales. Given an all-atom structure, a set of CG variables is constructed (Section 2.1). Next, several NAMD runs (1–10 ps) are performed [12] to compute factors needed to update CG variables. Then, the CG variables are updated (Section 2.2) via Langevin dynamics (Eq. (14)). Since factors for the CG dynamics are computed from the inter-atomic force-field (CHARMM [13]) via MD, the dependence on CG states of factors in the Langevin equation is avoided. The atom-resolved structure for the next microstate is constructed from the set of updated CG variables via Eq. (1). Thus, CG variables control the equilibrium probability of atomistic configurations, while the latter provide information that underlies the dynamics of CG variables. In this way, the ensemble of atomic configurations coevolves with the CG variables. Here, the degree of similarity between DMS and EMD simulations is quantified. The EMD simulations are performed on NAMD [12] directly. Since the version of DMS studied here uses NAMD to generate the ensemble of all-atom states, the comparison is readily made due to the identity of I/O formats for generating EMD and DMS trajectories.

2.1. CG variables

Consider a macromolecular assembly described via the positions of its \( N \) constituent atoms labeled \( i = 1, \ldots, N \). Let the \( i \)-th atom be moved from its reference position \( \bar{r}_i \) to

\[
\bar{r}_i = \sum_k U_{ik} \bar{r}_k + \bar{\sigma}_i ;
\]

\( \bar{\sigma}_i \) is a residual displacement (at time \( t \)) in addition to the coherent deformation generated by the \( k \) sum.

An explicit expression for \( \Phi_k \) is obtained by minimizing the mass-weighted square residual (i.e., \( \sum_i m_i \bar{\sigma}_i^2 \)) with respect to \( \Phi_k \) [8]:

\[
\Phi_k = \frac{\sum_{i=1}^N m_i U_{ik} \bar{r}_i}{\mu_k} ; \quad \mu_k = \sum_{i=1}^N m_i U_{ik}^2 ; \quad \mu_k \]

\( m_i \) is the atom \( i \) mass. As previously, CG variables labeled by indices \( k = \{000, 100, 010, 001\} \) are denoted lower-order and the rest are higher-order [7]. Inclusion of \( m_i \) in developing Eq. (2) gives \( \Phi_k \) the character of a generalized center-of-mass (CM). For example, if \( U_{2i} \) is independent of \( i \) (as for the Legendre polynomial of zeroth order) then \( \Phi_k \) is proportional to the CM. A subset of lower-order CG variables defined in this way constitutes a strain tensor accounting for compression-extension-rotation; the higher-order ones describe more complex deformations [6,14,15]. The \( \mu_k \) serve as effective masses for each CG variable, implying the temporal scales. The masses primarily decrease with increase in complexity of \( U_{ki} \) [14,15]. Thus, CG variables with higher \( k \) probe smaller regions in space. A model based on such CG variables probes structures over a diverse range of spatial scales (see earlier [14] and Section SI of supporting information), and evolves on timescales much greater than that of individual atomic collisions/vibrations [3,14,16].

2.2. CG dynamics

The multiscalar analysis starts with a transformation of the N-atom probability density from \( \rho(\Gamma, t) \) formulation to one that makes the multiple ways on which \( \rho \) depends on \( \Gamma, t \) more explicit. \( \Gamma \) is a point in the \( N \)-particle phase space, and \( t \) is time.

\[
\rho(\Gamma, t) \rightarrow \rho(\Gamma, \Phi(\Gamma), t_0(t), \xi(t); \varepsilon) ;
\]

Thus an ansatz is made that the reformulated probability density \( \rho \) depends on the \( N \)-atom state \( \Gamma \) both directly and, via a set of CG variables \( \Phi(\Gamma) \), indirectly. Similarly, \( \rho \) depends on the sequence of times \( t_0 \), \( \xi(t) = \varepsilon \times t \) and \( \varepsilon \) is a small parameter reflecting the ratio of the characteristic time of individual atomic collisions/vibrations to that of large-scale system structural change. The times \( t_0 \) for \( n > 0 \) are introduced to account for the slower behaviors in \( \rho \). \( t_0 \) accounts for fast processes (i.e., when \( t_0 \) changes by one unit, \( t \) changes by \( 10^{-14} s \)). Dependence of \( \rho \) on all the times tracks the processes evolving on the various characteristic times which are assumed to scale as \( \varepsilon^{-n} \). Introduction of \( \rho \) as in Eq. (3) to the Liouville equation (LE) [17] and the chain rule imply the multiscale LE [18]:

\[
\sum_{n=0}^{\infty} \varepsilon^n \frac{\partial \rho}{\partial t_n} = (L_0 + \varepsilon L_1) \rho ;
\]

where \( L_0 \) is the zeroth order Liouville operator that involves partial derivatives with respect to \( \Gamma \) at constant \( \Phi \) (when operating on \( \rho \) in the form Eq. (3)); conversely for \( L_1 \) (see Section SI):

\[
L_0 = -\sum_{i=1}^N \left( \frac{\partial}{\partial \bar{r}_i} + \bar{F}_i \cdot \frac{\partial}{\partial \bar{F}_i} \right) ;
\]

\[
L_1 = -\sum_k \bar{P}_k \cdot \frac{\partial}{\partial \Phi_k} ;
\]

\( \bar{F}_i \) is the momentum conjugated to \( \Phi_k \); \( \bar{F}_i \) and \( \bar{P}_k \) are the momentum and net force on atom \( i \). With this, the LE can be solved for \( \rho \) as an expansion in \( \varepsilon \): \( \rho = \sum_{n=0}^{\infty} \rho_n \). Solving order by
order [19] as shown in Section III, the lowest-order solution ρ₀ is found to be
\[ ρ₀ = \tilde{ρ}(Γ, Φ)W(Φ, t), \]
\[ \tilde{ρ} = e^{-βH} Q(Φ, β), \]
\[ Q(Φ, β) = \int dΓΔ(Φ - \tilde{Φ})e^{-βH}, \] (7)
the factor \( \tilde{ρ} \) is the equilibrium probability density for constraining state \( Γ \) such that \( \tilde{Φ}(Γ) = Φ \), where \( \tilde{Φ} \) expresses the relationship between CG variables and the atomistic state \( Γ \). \( β = 1/k_B T \) and \( Q \) is the \( Φ \)-constrained partition function for the isothermal conditions considered. \( Δ \) is a product of Gaussian functions that restricts \( Γ \) such that \( \tilde{Φ}(Γ) \) is near a fixed \( Φ \), \( W(Φ, t) \) found to be the probability density for CG variables as follows.

First, define the reduced probability density \( \tilde{W} \) via
\[ \tilde{W} = \int dΓωΔ(Φ - \tilde{Φ}(Γ))ρ; \] (8)
\( ω \) is a constant arising from the conversion of a quantum state distribution to a classical state integration. As \( ε → 0, ρ → ρ₀ \), thus the above implies \( \tilde{W} → W(Φ, t) \), where \( Δ \tilde{ρ} \) is normalized to 1. \( \tilde{W} \) and the LE imply
\[ \frac{∂\tilde{W}}{∂t} = \int dΓωΔ(\tilde{Φ}(Γ))ρ, \] (9)
using \( L = L₀ + ΕL₁ \) and \( ρ = ρ₀ + ερ₁, L₀ \) and \( ρ₀ \) do not contribute to Eq. (9). The first order correction \( ρ₁ \) has the form
\[ ρ₁ = e^{L₀A₁} - L₀\tilde{ρ} \frac{∂W}{∂t₂} - \sum_k \int_{-t₀}^{t₀} dt'e^{-L₀t₀} \tilde{Π}_k \cdot \frac{∂}{∂Φ_k} \left( \frac{W(\tilde{Q}(Φ, β))}{Q(Φ, β)} \right), \] (10)
as derived in Section III [3,8]. Here, \( A₁ \) is the value of \( ρ₁ \) at \( t₀ = 0 \). Inserting \( ρ₁ \) into Eq. (9), one obtains the Smoluchowski equation for \( W(Φ, t₂) \) [8] (Section III):
\[ \frac{∂W}{∂t₂} = \sum_{kk'} \frac{∂}{∂Φ_k} \left[ D_{kk'} \left( \frac{∂}{∂Φ_k} - ΕF_k \right) W \right]. \] (11)
where \( t₂ = ε²t \), and the diffusivity factors \( D_{kk'} \) and thermal-average forces \( F_k \) are given by
\[ D_{kk'} = \frac{1}{μ_k μ_k'} \int_{-∞}^{0} dt (\tilde{Π}_k \tilde{Π}_k'), \] (12)
\[ \tilde{F}_k = -\frac{∂E}{∂Φ_k} = \langle F_k \rangle. \] (13)
As above, \( \tilde{Π}_k \) is the momentum associated with \( Φ_k \) and \( μ_k \) serve as effective masses. \( ε \) implies configurational average. \( F = -(1/β)lnQ(Φ, β) \) is the \( Φ \)-constrained Helmholtz free energy with \( Q(Φ, β) \) being the partition function (Eq. (7)), and \( \tilde{F}_k = \frac{1}{N} \sum_i U_i(\tilde{R}_i) \tilde{F}_i \) is the \( k \)-th CG force in terms of the net force \( \tilde{F}_i \) on atom \( i \) with \( ρᵢ \) being its original position. The Smoluchowski equation is equivalent to an ensemble of solutions to the Langevin equation [20]:
\[ \frac{∂Φ_k}{∂t} = \sum_k D_{kk'} \tilde{F}_k + ξ_k; \] (14)
the random force \( ξ_k \) is constrained by \( D \) via the classic fluctuation-dissipation relation [5]. Eq. (14) is the basis of CG evolution within DMS. Eq. (14) is used to update the CG state; this updated state is used via a constrained method to generate an ensemble of atom-resolved state consistent with it. These microstates are then used with short MD simulations to construct factors needed to advance the CG state one timestep. Thus, coupled evolution of the system is simultaneously achieved across multiple scales [3,4,7].

Next, the equilibrium properties of the probability \( ρ \) are analyzed and compared to those from a long isothermal MD run. Consider \( \rho₀ \) (Eq. (7)) in the limit \( t → ∞ \), when the system is closed to mass transfer and in good contact with a thermal bath. The longtime form of \( ρ₀ \) depends on that of \( W \) as \( ρ \) is time independent. In turn, temporal evolution of \( W \) is given by Eq. (11). As \( t → ∞, W → W^{eq} \), where \( W^{eq} = e^{-βH}/Z \). Here, \( Z = \int dΓωΔ(Φ - \tilde{Φ})e^{-βH} \) is the partition function. Note, \( Z \) is different from the \( Φ \)-constrained partition function \( Q(\tilde{Φ}) \) (Eq. (7)) since \( Z \) accounts for all thermally allowed states, not only those for which the values of CG variables are constrained to be near \( Φ \).

The asymptotic analysis is continued for \( ρ₁ \) (Eq. (10)). Since the eigenvalue spectrum of \( L₀ \) is purely imaginary, the initial value (\( A₁ \)) term remains bounded for all time, and notably oscillates around zero or may have a constant contribution when \( A₁ \) lies in the null space of \( L₀ \). Because physically \( ρ₁ \) must be bounded for all \( t₀, dW/dt₁ \) must be 0, i.e., \( W \) is independent of \( t₁ \) [3]. The third term in Eq. (10) vanishes for \( W = W^{eq} \) as can be verified algebraically. Thus, if \( A₁ \) lies in the null space of \( L₀ \), as \( t → ∞, ρ → ρ^{eq} \) term of \( O(ε) \) which oscillates around 0 plus a small term that lies in the null space of \( L₀ \). With this, as \( t → ∞, ρ = ρ^{eq} = e^{-βH}/Z \), indicating the ensemble reached after many Langevin timesteps should be consistent with the Boltzmann distribution after thermal-average forces have essentially vanished. The same holds for EMD. With this, the DMS and EMD generated equilibrium ensembles are seen to be equivalent. However, constructing an equilibrium ensemble that is adequately representative of the very high dimensional configuration space in large macromolecular systems is computationally unfeasible. As an alternative, similarity of the DMS equilibrium CG probability \( W^{eq} \) to that constructed from MD simulations is demonstrated through the comparison of DMS and EMD in terms of CG variables, CG velocities, and the free-energy profile along the CG variables trajectory. This similarity is denoted quasiequivalence (Section I) here to distinguish from equilibrium at the atomistic level. In the next section, this DMS-EMD quasiequivalence is demonstrated computationally.

3. Results and discussion
The system used for probing DMS-EMD quasiequivalence is lactoferrin. This protein is composed of a distal and two proximal lobes (Figure 1a). Two free energy minimizing conformations have been demonstrated experimentally: dierfer with closed proximal lobes (PDB code 1LF), and apo with open ones (PDB code 1LFH) [21]. Simulations here start with open lactoferrin structure without iron ions and we simulate its closing in vacuum. Simulations are done in vacuum because (a) strong electrostatic interactions between oppositely charged surface peptides enable the open proximal lobes to close in reasonable compute time and (b) it is a simple example wherein only protein-protein interactions affect the overall system dynamics through the thermal-average forces. Understanding of solute-solvent interactions and their effect on the Langevin CG dynamics is more complicated as shown earlier [3,14,22], and is thus avoided. The open-to-close transition in lactoferrin is analyzed via an ensemble of eight 300 ns MD trajectories. Ensemble MD simulations are performed with NAMD (simulation parameters are shown in Section IV). During the simulation most of the solution phase tertiary structure is conserved, as can be expected for large proteins such as lactoferrin [23,24]. However, as salt bridge interactions grow stronger in vacuum than in solution, the molecule undergoes an open-to-close transition.
First, it is tested whether the above MD simulations adequately sample the reduced dimensional CG space necessary to probe the open-to-close transition. As introduced in Section 2, and several of our previous publications [3,6–9,15,17,19], the CG variables employed for the present demonstration are generalized center of masses that capture overall system translation, extension-compression, rotation, twisting, bending, tapering, and other more complicated higher-order motions. Figure 2 shows time evolution of the velocity autocorrelation functions (VACFs) for the CG variables. The VACFs decay on a timescale of 1 ps, implying a simulation of 300 ns yields $3 \times 10^3$ statistically independent CG configurations. Since an estimate of an observable, e.g., CG variables, based on greater than 20 statistically independent configurations is considered reliable [25], length of the current trajectory is sufficient to sample the CG states. Thus, provided correct number of CG variables is chosen, the structural transition can be accurately captured in the CG space. Sampling of the low dimensional CG space is further probed via analysis of all-to-all RMSDs [13,25], root-mean-square inner product (RMSIP) of essential covariance eigenvectors [26].

Presented in Section 5.1, such measures quantitatively show that the EMD simulations successfully sample multiple CG states of the system, which are further shown below to be possibly the most biologically relevant. Therefore, the above EMD simulations can be used for benchmarking the accuracy of our DMS simulations.

DMS simulations of lactoferrin are performed using $3^2$ CG variables with a Langevin timestep of 25 ps; these CG variables are particularly suited for capturing extension-compression type motions. The ps-scale timestep is achieved provided CG-constrained quasiequilibrium ensembles of at least 200 all-atom configurations are constructed for computing thermal-average forces and diffusions at every Langevin timestep employing swarms of twenty 10 ps trajectories (See Section 4.4). With this, eight 300 ns ensemble MD with configurations saved every ps, and a 300 ns DMS simulation with 200 configurations saved every 25 ps provides isothermal ensembles of identical size. Reliability of the DMS simulations relative to those from EMD is tested as follows. First, consider the large-scale system dynamics. Besides the visual comparison between DMS and EMD both of which generated closed structures at equilibrium (Figure 1b), in Figure 3 large-scale variables from DMS and EMD are compared. These trajectories are within numerical error, i.e., CG evolution predicted by DMS lies within the standard deviation of 5–10% observed for the approximately convergent EMD results; the observed difference is due to finiteness of the EMD trajectories (i.e., limited number of trajectories and length of simulation). Similarly, due to the finiteness of the swarm of MD trajectories used with DMS for all-atom ensemble generation at every Langevin timestep, there exists a numerical uncertainty in the DMS results. As shown in Figure 3, the error bars on DMS predictions are small and strongly overlap with those from EMD; this implies the statistical equivalence of the CG variables from EMD and DMS. For DMS, the error bar at a given point is obtained by measuring standard deviations within the swarm of short MD trajectories at approximately constrained CG variable. For EMD, the error is measured as deviations of individual MD trajectories from the mean quantity at that time.

While the closing of lactoferrin occurs in all three Cartesian directions, it is most pronounced within the XZ plane. Therefore, CG
variable clusters were constructed for $\Phi_{100X}$ and $\Phi_{001Z}$ (Figure 4a); here, $\Phi_{100X}$ captures extension/compression along the X direction, and $\Phi_{001Z}$ for the Z direction (See Section SI for details). Clusters corresponding to five closed structures are sampled during the eight MD, as well as the DMS trajectories; the associated free energy landscape is shown in Figure 4b. Each of the five CG variable clusters in Figure 4a characterizes a minima on the free energy surface. DMS and EMD visit similar clusters, which further suggests presence of qualitatively comparable large-scale motions in both trajectories.

The above results illustrate the similarity between DMS and EMD predictions at the CG level. Next, similarity on an intermediate scale between CG and atomic levels is shown via the comparison of CG velocities. The ensemble of DMS CG velocities shows they become close to zero at free energy minima, and are non-zero away from the minima (Figure 5a). Comparison between DMS and EMD CG velocities shows strong overlap (Figure 5b), thereby suggesting DMS-EMD equivalence also at an intermediate scale.

Even though the aim of the work is to show DMS-EMD quasiequivalence, analysis at the atomic level is shown in Section SV.B to quantitatively assess the accuracy of the atom-resolved description provided by DMS. It is shown that the degree of DMS-EMD similarity depends on the choice of DMS timestep at atomistic scale. Furthermore, it becomes impossible to computationally reproduce the correct Boltzmann distribution (Section SV.B), though the DMS methodology analytically yields the same (Section 2). In view of the numerical demonstration of such computational limitations, we chose to show quasiequivalence in the CG space, rather than equivalence in the $N$-dimensional space.

 Altogether, via demonstrating the similarity in dynamics, states sampled, and free energy landscape at the CG scale, and CG velocities at an intermediate scale, quasiequivalence of DMS and EMD predicted trajectories is computationally established.

There are subtle large-scale structural differences between the five clusters of sampled states as manifested in the different topologies of the associated free energy minima in Figure 4b. A minimalistic model is constructed to physically interpret the CG variable clusters and free energy of Figure 4a [27]. Degrees of freedom captured by the CG variables include the distances of the center of mass (CM) of each lobe from the CM of the entire protein and angles between the lines joining these CMs (Figure 6a). Here, lobe CMs are labeled with L1–L3 for two proximal lobes (L1 and L2) and one distal lobe (L3); distances are labeled with D1–D3, and angles are labeled with A1–A3. The open lactoferrin state is represented by the bottom right CG cluster in Figure 4a with maximum magnitudes of CG variables corresponding to extended states along the X- and Z-directions. As shown in Table 1, changes in D3 of all

![Figure 4](image-url) (a) Clusters of CG variables showing x component of 100 and z component of 001 (See Section 2 for nomenclature of CG variables): blue, red, green, orange, and pink are 5 clusters reviewed by eight MD trajectories. The eight MD simulations start from the same open state (bottom right). DMS (gray) clusters sampled most regions of EMD clusters. b) The corresponding free energy; the gray region is not visited. (For interpretation of the references to color in figure legend, the reader is referred to the web version of the article.)

<table>
<thead>
<tr>
<th>Table 1</th>
<th>1LH</th>
<th>Blue</th>
<th>Red</th>
<th>Green</th>
<th>Pink</th>
<th>Orange</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle 1 (A1)</td>
<td>75.4</td>
<td>58.4</td>
<td>55.3</td>
<td>56.7</td>
<td>60.7</td>
<td>55.4</td>
</tr>
<tr>
<td>Angle 2 (A2)</td>
<td>143.1</td>
<td>150.3</td>
<td>160.6</td>
<td>160.0</td>
<td>154.0</td>
<td>158.0</td>
</tr>
<tr>
<td>Angle 3 (A3)</td>
<td>141.5</td>
<td>151.3</td>
<td>144.1</td>
<td>143.2</td>
<td>145.1</td>
<td>146.6</td>
</tr>
<tr>
<td>Distance 1 (D1)/Å</td>
<td>26.5</td>
<td>25.0</td>
<td>19.2</td>
<td>19.1</td>
<td>21.4</td>
<td>21.6</td>
</tr>
<tr>
<td>Distance 2 (D2)/Å</td>
<td>26.8</td>
<td>24.2</td>
<td>28.7</td>
<td>28.6</td>
<td>25.9</td>
<td>28.3</td>
</tr>
<tr>
<td>Distance 3 (D3)/Å</td>
<td>20.0</td>
<td>20.2</td>
<td>20.1</td>
<td>20.0</td>
<td>19.3</td>
<td>20.9</td>
</tr>
</tbody>
</table>
Figure 5. (a) DMS CG velocity distribution along x component of CG variable 100 and z component of CG variable 001. Zero velocities of 001Z are observed at free energy minima (Figure 4b). (b) Comparison between DMS (gray) and EMD (red) CG velocities. (For interpretation of the references to color in figure legend, the reader is referred to the web version of the article.)

Figure 6. Overlap of open structure (Figure 1(a), black) and closed structures from 5 clusters (colors are consistent with clusters in Figure 4(a)). (For interpretation of the references to color in figure legend, the reader is referred to the web version of the article.)

five closed structures relative to that for the open state are quite small implying that the collapse does not involve large-scale deformations of the distal lobe. Hence the deformations include only proximal lobes with CMs of L1 and L2. Within a given MD simulation, angle between proximal lobes (A1) in the final state is smaller than that in the initial open state which indicates that in arriving at each of the five structures the lactoferrin experiences open-to-close transition to certain degree. However, symmetric or asymmetric deformations involving two proximal lobes lead the structure to different basins. Symmetric contractions where both D1 and D2 decrease result in the blue (Figure 6b) and pink (Figure 6e) clusters on Figure 4a; the degree of contractions in blue lesser and marginally more symmetric than in the pink cluster of structures. Subsequently, the values of CG variables decrease relative to the open state. Since both proximal lobes collapse to the center creating steric repulsion, A1 in blue and pink structures is relatively greater than that in the other three structures. Any further closing of lobes leads to a decrease in A1 and D1 while D2 increases to avoid steric clashes between lobes. Subsequently, the values of CG variables are relatively smaller in symmetric than asymmetric structures. This
results in the asymmetric structures (for red and orange clusters in Figure 4a, the minimalist models are shown using the same color code in Figures 6c (red) and f (orange)). Green cluster is an intermediate structure between asymmetric and asymmetric structures. As shown in Figure 4, green cluster contains a small population (Figure 4a) and is located in a quite high energy basin (Figure 4b).

Given the open-to-close transition in lactoferrin is mediated by only two lobes, as is also known from experiments [28], and the closing can be either symmetric or asymmetric, the simulated scenarios have sampled all possible biologically relevant mode of closing, again, a testament to the adequacy of the configurations sampled both by DMS and EMD.

4. Conclusion

DMS uses notions from multiscale theory and the construction of quasi-equilibrium ensembles that coevolve with CG variables to capture the large-scale structural changes in macromolecular systems. Since DMS coevolves the CG state with an ensemble of microstates consistent with it, DMS is equivalent to EMD as demonstrated computationally. Since trajectories in 6N-dimensional space could not be readily compared, a coarse-grained criterion for judging two trajectories was needed. The ‘quasi-equivalence’ concept introduced here involved two scales at which similarity was assessed, i.e., the CG variables and CG velocities. This DMS-EMD quasi-equivalence is demonstrated via all-atom simulations of a structural transition in lactoferrin.

The larger-scale structural properties such as radius of gyration, and RMSD show agreement between DMS and EMD. Clusters of CG variables obtained from DMS sampled similar regions as EMD clusters, suggesting presence of qualitatively comparable large-scale motions. At an intermediate level, CG velocities also show strong overlap between DMS and EMD. With this, DMS-EMD quasi-equivalence is demonstrated computationally. Finally, a minimalist model is constructed to interpret the CG variable clusters physically. In conclusion, the DMS-EMD quasi-equivalence validates that the DMS approach yields statistically significant all-atom trajectories along with greater CPU efficiencies.

Acknowledgments

This research was supported in part by the following: NSF INSPIRE program under Grant No. ECCS-1344263; Indiana University College of Arts and Science via the Center for Theoretical and Computational Nanoscience; Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute and the Indiana METACyT initiative.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cplett.2014.10.020.

References