Thermal nanostructure: An order parameter multiscale ensemble approach

S. Cheluvaraja and P. Ortoleva

Additional information on J. Chem. Phys.

Thermal nanostructure: An order parameter multiscale ensemble approach

S. Cheluvaraja and P. Ortoleva[a]

Department of Chemistry, Center for Cell and Virus Theory, Indiana University, Bloomington, Indiana 47405, USA

Deductive all-atom multiscale techniques imply that many nanosystems can be understood in terms of the slow dynamics of order parameters that coevolve with the quasiequilibrium probability density for rapidly fluctuating atomic configurations. The result of this multiscale analysis is a set of stochastic equations for the order parameters whose dynamics is driven by thermal-average forces. We present an efficient algorithm for sampling atomistic configurations in viruses and other supramillion atom nanosystems. This algorithm allows for sampling of a wide range of configurations without creating an excess of high-energy, improbable ones. It is implemented and used to calculate thermal-average forces. These forces are then used to search the free-energy landscape of a nanosystem for deep minima. The methodology is applied to thermal structures of Cowpea chlorotic mottle virus capsid. The method has wide applicability to other nanosystems whose properties are described by the CHARMM or other interatomic force field. Our implementation, denoted SIMNANOWORLD™, achieves calibration-free nanosystem modeling. Essential atomic-scale detail is preserved via a quasiequilibrium probability density while overall character is provided via predicted values of order parameters. Applications from virology to the computer-aided design of nanocapsules for delivery of therapeutic agents and of vaccines for nonenveloped viruses are envisioned. © *2010 American Institute of Physics*. [doi:10.1063/1.3316793]

I. INTRODUCTION

The behavior of nanosystems like viruses and nanocapsules has recently been described in terms of order parameters (OPs), which coevolve with a quasiequilibrium distribution of atomistic configurations.[1–4] The OPs describe nanoscale features and evolve via stochastic equations. In the present context, OPs describe overall features of a nanosystem such as shape, global orientation, or large scale structure. A necessary requirement for an OP is that it be obtained by averaging over a large number of atomic variables, i.e., is a collective variable. This use of the term OP is distinct from that of some other authors (see papers in Ref. 5), notably which are not collective variables (e.g., torsional angle or reaction coordinate that are highly fluctuating and incoherent). The driving forces for OP dynamics are a coherent thermal-average contribution and a random one. The key to our approach is an automatable formalism for constructing OPs and a rigorous framework for deriving equations for their stochastic dynamics. The overall logic of the physical picture is suggested in Fig. 1. Here, we focus on the computation of the thermal-average forces which must be executed with great efficiency considering that the nanosystems of interest are supramillion atoms in scale.

OPs describe larger-scale features involving many atoms and hence they evolve on timescales much longer than 10^{-14} s characteristic of atomic collisions/vibrations. Hence, a novel multiscale analysis was used to derive rigorous equations for the stochastic dynamics of OPs.[2–4,6] These equations are driven by thermal-average forces that are the gradient of the free energy constructed from an ensemble constrained to fixed OP values. Here, we present an algorithm for constructing these thermal-average forces and demonstrate it for determining the thermal structure of a virus.

Many nanosystems like viruses undergo structural transitions that take place on timescales much greater than that of atomic collisions/vibrations.[7–12] Multiscale analysis is a natural way to understand these phenomena.[13–15] In the present approach, OPs are used to capture the slow overall transformation of a system from one state to another. These OPs are introduced and our deductive multiscale analysis indicates that they determine the characteristics of the ensemble of atomic configurations. In turn, the latter drives the evolution of the OPs via thermal-average forces (Fig. 1). In a deductive multiscale approach, the Liouville equation is solved via a perturbation expansion in a ratio of characteristic masses, lengths, or times. The solution to the Liouville equation for the N-atom probability density is undetermined to lowest order unless, as with classic ensemble theory, an entropy maximization hypothesis is adopted. This theme has been explored in detail in the context of nanosystems.[3] In the next order in the multiscale theory, an equation of stochastic OP evolution is obtained wherein the coherent dynamics is driven by a diffusivity matrix and thermal-average forces, and by a random force whose correlation properties are determined by the same diffusivities. To apply this multiscale methodology, the thermal-average forces and diffusion factors for a set of OPs must be calculated. As these factors change with the values of the OPs, they must be computed dynamically as the system evolves.

[a]Electronic mail: ortoleva@indiana.edu. Tel.: (812) 8566000.

132, 075102-1
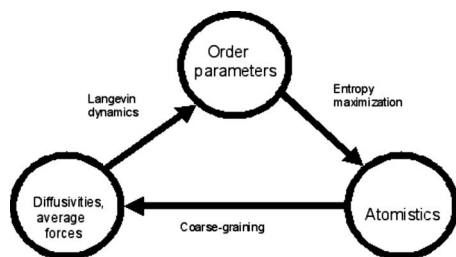© 2010 American Institute of Physics

FIG. 1. OPs characterizing nanoscale features affect the probability of atomistic configurations which, in turn, determine forces driving OP dynamics. This is an essential feature of the dynamics of a nanosystem and our multiscale approach as implemented in SIMNANOWORLD™.

Our long-range goal is to predict bionanostructure via a quantitative, calibration-free approach. To do so, we use an underlying all-atom description which enables us to use an interatomic force field (e.g., CHARMM, AMBER96) to avoid recalibration with experimental data. To construct the thermal-average forces within this framework requires a Monte Carlo approach and an efficient algorithm for generating a representative sample of atomistic configurations constrained to fixed OP values. Difficulties in ensemble generation and free energy computation are discussed extensively in Refs. 16–18. Here, we present a novel approach based on a set of structural OPs and calculate the forces driving their dynamics by efficient sampling of atomistic configurations. We briefly mention other multiscale approaches and compare our method with them.

First, none of the other multiscale methods use our OPs. Our approach in this regard provides self-consistency/completeness criteria in terms of the relative magnitude of the correlation function for the OP velocities. If this correlation time is too long, it signals that key OPs have been omitted. Since our OPs are automatically generated, this enables an algorithm which can then be triggered to introduce additional OPs and restart the simulation. Model-free methods[19] have the advantage that the diffusion factors and thermal-average forces occurring in the Langevin equation of OP dynamics need not be computed (unlike our multiscale approach). However, fluctuations are not treated rigorously and the full coupling across scales in space and time is not maintained. Projection operator methods[20,21] are often used as the basis of a multiscale scheme. This approach involves memory functions. If the latter have no long-time tails, then the formalism is equivalent to our fully coupled multiscale perturbation scheme. If the memory functions have long-time tails, then they are not practical to compute via molecular dynamics (MD). This raises doubts as to the merit of projection operator techniques versus our approach. If the memory functions have long-time tails, then it is known that more OPs should be added. This suggests another advantage of our approach, i.e., automated generation of OPs. Force matching[22,23] represents a system by lumped elements and uses MD data to calibrate the effective force field between them, thus needing recalibration with each new system or condition (e.g., pH). Our approach is calibration free because it explicitly calculates thermal-average forces for each OP introduced, only requiring the interatomic force field such as CHARMM or AMBER. More fundamentally, force matching (1)

ignores the dynamical updating of the relation between the thermal-average forces and the OPs as the system evolves, (2) ignores the highly frictional character of the internal processes within each element and the stochastic forces on them. Much of these difficulties stem from the dilemma that if the lumped elements are too small, there is little computational advantage, while if they are too large the internal degrees of freedom/dissipative effects are ignored despite their likely importance. Similar difficulties occur for simulating a macromolecule embedded in a membrane, where the role of dissipative and nonlinear effects in the continuum is ignored. In metadynamics,[24] the free energy surface is explored by approximating the local free energy as a running sum over Gaussians and then traveling in the direction of decreasing free energy. The free energy derivatives are calculated by imposing these Gaussian constraints. However, this method has only been demonstrated in small systems like alanine dipeptide and not larger ones like viral capsids. Milestoning[25] uses local ensembles at specified points in a given reaction pathway to calculate first-passage time distributions that are then integrated to determine the overall time scale of the system. Both milestoning and metadynamics require prior knowledge of the reaction coordinate. In our formalism, no such prior information is needed and thermal-average forces for each OP value are calculated by sampling the fixed OP ensembles, and the system evolves in the direction of decreasing free energy.

In Sec. II, we present OPs relevant for the analysis of structural transitions in a nanosystem. It was shown that they provide a natural starting point for a multiscale analysis.[1,3] The end point of this analysis is a Smoluchowski equation for stochastic OP dynamics driven by thermal-average forces (Sec. III). In Sec. IV, we present a sampling algorithm for computing the thermal-average forces. Our formalism provides the probability density for atomic configurations and a stochastic (Langevin) dynamic for the OPs. We also provide an algorithm for finding deep minima of the free-energy landscape, and in Sec. V we present results using an implementation of our OP thermal-average forces as applied to viral capsid structural transitions. Conclusions are drawn in Sec. VI.

II. ORDER PARAMETERS

In the present context, OPs are variables that characterize nanometer-scale features of a nanosystem.[26] In our approach, a mapping is introduced that transforms a starting point $\vec{r}^o$ into a new one $\vec{r}$ via

$$\vec{r} = \sum_{\underline{k}} \vec{\Phi}_{\underline{k}} U_{\underline{k}}(\vec{r}^o). \tag{1}$$

In our implementation, the basis functions $U_{\underline{k}}$ are products of Legendre functions for the x, y, and z directions, i.e., $U_{\underline{k}}(\vec{r}^o) = u_{k_1}(x^o) u_{k_2}(y^o) u_{k_3}(z^o)$ with indices $\{k_1, k_2, k_3\}$ and $\vec{r}^o = (x^o, y^o, z^o)$. However, other basis functions can be used. The OPs $\vec{\Phi}_{\underline{k}}$ generate the deformation of space and, as shown below, structural changes in a nanosystem. As one may implement the construction of the OPs in a general way, only the limits must be specified on the $\underline{k}$ sum of Eq. (1) to fix the

detail of the coarse-grained description.[1] The transformation in Eq. (1) can be formulated such that it takes a reference state (e.g., X-ray structure) to that for conditions of interest. The $\vec{\Phi}_k$ determine the mix of the deformations that the nanostructure can undergo. In contrast to coarse-grained bead models (where one conjectures which structural subunits are rigid and can be represented by structureless beads), our method keeps track of the deformation of all space in the system, including alterations within all subunits.

In our method, the transformation Eq. (1) accounts for the coherent, large-scale deformation of the system. A residual displacement $\vec{\sigma}_i$ for each atom $i = 1, \ldots, N$ is used to track the more random movement of atoms over and beyond the coherent motion induced by the $\vec{\Phi}_k$. With this, the position $\vec{r}_i$ of atom i is written

$$\vec{r}_i = \sum_k \vec{\Phi}_k U_k(\vec{r}_i^o) + \vec{\sigma}_i. \tag{2}$$

This mixed coherent-residual representation takes atom i from an initial position $\vec{r}_i^o$ (e.g., an input structure) to an evolved one. To define the $\vec{\Phi}_k$ more concretely, we construct them to have the maximum information by minimizing the mass-weighted total square residual.[1–4,6] One finds $\vec{\Phi}_k$ is the solution of

$$\sum_{k'} B_{kk'} \vec{\Phi}_{k'} = \sum_{i=1}^N m_i U_k(\vec{r}_i^o)\vec{r}_i, \tag{3}$$

$$B_{kk'} = \sum_{i=1}^N m_i U_k(\vec{r}_i^o) U_{k'}(\vec{r}_i^o). \tag{4}$$

This expresses the $\vec{\Phi}_k$ in terms of the positions $\vec{r}_i$ of the N atoms. With this and the basic physical (N-atom) description, deductive multiscale techniques imply rigorous equations for the stochastic dynamics of the $\vec{\Phi}_k$.

For convenience, one may choose the basis U_k to be orthogonal. In that case, the B-matrix (Eq. (4)) is diagonal. The orthogonalization can be carried out using a standard Gram–Schmidt procedure starting from any nonorthogonal basis set. Multiplying each element of U_k by $\sqrt{m_i}$, inputting the resulting modified basis into a GS package, and dividing the output component-by-component by $\sqrt{m_i}$ yields orthogonal base functions. Thus, the resulting vectors are mass-weighted orthonormalized. In this basis, the B-matrix is the identity; thus, Eq. (3) simplifies. Unlike the original basis, the orthonormal one has the advantage that, when the basis is enlarged to include more OPs, the pre-extension basis-derived OPs are not affected as the GS method creates the orthogonal basis in an incremental fashion. The GS orthogonalization preserves the physical nature of the three largest OPs (100X, 010Y, and 001Z) because they are always chosen to be the first three members of the basis. These OPs continue to directly track the overall radius and shape of the nanosystem.

III. DEDUCTIVE MULTISCALE ANALYSIS

The above description of the OPs does not provide a complete theory since no closed equation for their dynamics is provided. However, deductive multiscale analysis does provide such a theory wherein the OPs are coevolved with a quasiequilibrium probability for the atomic state.[2,3,27] According to Liouville's picture, the N-atom probability density ρ characterizes the evolving statistical state of the system. As the OPs are slowly varying and the atomic degrees of freedom fluctuate rapidly, the system arrives at a quasiequilibrium state consistent with the OPs. A key parameter is ε, the ratio of the mass of a typical atom to that of the nanoscale feature of interest. As $\varepsilon \to 0$, ρ has the form

$$\rho = \hat{\rho} W, \tag{5}$$

with

$$\hat{\rho} = \exp(-\beta H)/Q(\Phi), \tag{6}$$

where Q is the reduced partition function for fixed values of the OPs Φ, and H is the energy expressed in terms of the set of positions and momenta (Γ) of all atoms, and

$$Q(\Phi) = \int d\Gamma^* \Delta(\Phi - \Phi^*)e^{-\beta H^*}, \tag{7}$$

Δ, a product of functions shaped like Π in each direction in Φ space, is centered about zero and has a width dictated by the small range of Φ included in the Γ^* integration. The factor W is the reduced probability density for Φ, i.e.,

$$W \equiv \int d\Gamma^* \Delta(\Phi - \Phi^*)\rho(\Gamma^*, t) \tag{8}$$

for time t.

Since the OPs evolve slowly, it is convenient to introduce time $\tau = \varepsilon^2 t$. Multiscale techniques[3] (notably solving the Liouville equation for small ε) imply that as $\varepsilon \to 0$ W depends on τ and satisfies

$$\frac{\partial W}{\partial \tau} = \sum_{kk'} \frac{\partial}{\partial \vec{\Phi}_k} \left[\vec{\vec{D}}_{kk'} \left[\frac{\partial}{\partial \vec{\Phi}_{k'}} - \beta \vec{f}_{k'} \right] W \right]. \tag{9}$$

The diffusivity factors $\vec{\vec{D}}_{kk'}$ are related to the correlation function of the time derivatives of the OPs. The thermal-average force $\vec{f}_k$ is given by

$$\vec{f}_k = -\frac{\partial F}{\partial \vec{\Phi}_k}, \tag{10}$$

$$Z = \int d^M \Phi \exp(-\beta F(\Phi, \beta)). \tag{11}$$

The equation for the evolution of W is of the generalized Smoluchowski form. Equivalent to it is an ensemble of time courses of Langevin equations

$$\frac{\partial \vec{\Phi}_{\underline{k}}}{\partial \tau} = \beta \sum_{\underline{k}'} [\vec{\vec{D}}_{\underline{k}\underline{k}'}, \vec{f}_{\underline{k}'}] + \vec{\xi}_{\underline{k}}. \tag{12}$$

The coherent part of the evolution is determined by the product of the diffusion factors and the thermal-average forces; the stochastic evolution is determined by the random noise $\vec{\xi}_{\underline{k}}$. The random noise is constrained by requiring the integral of its autocorrelation function to be equal to the diffusion coefficient.

From the above equation, evolving the OPs requires calculation of the thermal-average forces and the diffusion factors. The most probable equilibrium state of the OP is when the $\vec{f}_{\underline{k}}$ vanish. The equilibrium probability distribution for the OPs is W^{eq} from which one can compute RMS fluctuations, and is calculated from

$$W^{\mathrm{eq}} = \exp(-\beta F)/Z, \tag{13}$$

$$Z = \int d^M \Gamma^* \exp(-\beta H^*). \tag{14}$$

The atomistic fluctuations, i.e., of Γ, is given by $\hat{\rho}$. This conditional probability $\hat{\rho}$ (i.e., for Γ at given Φ) is $\Delta(\Phi - \Phi^*)\exp(-\beta H^*)/Q(\Phi)$.

IV. AN EFFICIENT MONTE CARLO SCHEME FOR CONSTRUCTION OF THERMAL-AVERAGE FORCES

The thermal-average forces are expressed as high-dimensional integrals. Consider a quantity A; its thermal average $\langle A \rangle$ in the ensemble represented by $\hat{\rho}$ is given by

$$\langle A \rangle = \int d^{6N} \Gamma^* \Delta(\underline{\Phi} - \underline{\Phi}^*) A \frac{\exp(-\beta H^*)}{Q}, \tag{15}$$

where Φ is the set of OP values and A is a function of Γ. The $*$ on a quantity indicates evaluation at Γ^* over which integration is carried out. High-dimensional integrals are often calculated via a Monte Carlo sampling technique.[28–31] To make the computations feasible requires constructing an ensemble of atomistic states for which the Δ factor is appreciable, i.e., a representative set of Γ^* for which Φ^* is near Φ. Thermal ensembles of states can be generated using many well-known methods,[16–18,32,33] but simultaneous imposition of the Δ factor makes ensemble generation more difficult. Unlike in other methods, we need to generate ensembles with given OP values and sample the OP space efficiently, requiring us to develop a new method for ensemble generation. Methods for generating other types of constrained ensembles have been presented in Refs. [5], [16], and [34–36]. However, the present method is tailored to the type of constraints and OPs used here.

From the definition of the OPs, the associated forces $\vec{f}_{\underline{k}}^*$ at Γ^* are obtained from atomic forces $\vec{F}_i^*$ via

$$\sum_{\underline{k}'} B_{\underline{k}\underline{k}'} \vec{f}_{\underline{k}'} = \sum_{i=1}^N U_{\underline{k}}(\vec{r}_i^o) \vec{F}_i, \tag{16}$$

where $U_{\underline{k}}$, $\vec{r}_i^o$, and $B_{\underline{k}\underline{k}'}$ are as in Eqs. (2)–(4). With the above equation, the Monte Carlo approximation to $\vec{f}_{\underline{k}}$ becomes

$$\vec{f}_{\underline{k}} = \frac{\sum_\lambda \Delta(\underline{\Phi} - \Phi^{(\lambda)\underline{k}}) \vec{f}_{\underline{k}}^\lambda \exp(-\beta H^{(\lambda)})}{\sum_\lambda \Delta(\underline{\Phi} - \Phi^{(\lambda)\underline{k}}) \exp(-\beta H^{(\lambda)})}, \tag{17}$$

where λ labels the configurations. A configuration Γ^* chosen at random will not usually contribute either (1) because the Δ- or (2) the Boltzmann-factor is small. Thus, a method for choosing the $\Gamma^{(\lambda)}$ that avoids these improbable states is required if the thermal-average forces for the supramillion atom systems of interest are to be constructed on a practical computational time frame. This is even more critical since solving the Langevin equations requires that the $\vec{f}_{\underline{k}}$ be calculated at every time step.

To generate an ensemble of atomistic configurations for given Φ, one might randomly vary the residuals in Eq. (2). However, this usually yields high energies as a result of overlapping configurations or bond length/angle distortions. Thus, an alternative approach to ensemble generation is used. In our approach, some coherent character in the moves from one sample to the next is used and the resulting all-atom configurations do not alter bond lengths/angles appreciably, while a significant movement of local clusters is achieved. This sampling strategy is based on the use of co-ordinated atomic moves generated by a sum over OP-like terms for $\underline{k}$ not in the set labeling the chosen OPs Φ. These residuals $\vec{Y}_{\underline{k}}$ are chosen at random, and thereby the residuals $\vec{\sigma}_i$ are sampled in a partially coherent, yet random manner. Thus, we adopt the prescription

$$\vec{\sigma}_i = \sum_{(\mathrm{Re}\,s)} \vec{Y}_{\underline{k}} U_{\underline{k}}(\vec{r}_i^o), \tag{18}$$

where (Re s) indicates the set of k taken and does not include those labeling the OPs Φ. In this manner, the $\vec{\Phi}_{\underline{k}}$ that are true OPs are kept constant over the ensemble. For the $\vec{Y}_{\underline{k}}$ (which generate the residuals), one can take for example,

$$\vec{Y}_{\underline{k}} = \frac{c\vec{\xi}(|\underline{k}| - k_{\max})^2}{a^2 + (|\underline{k}| - k_{\max})^2} \frac{b^2}{b^2 + (k_c - |\underline{k}|)^2},$$
$$|\underline{k}|^2 = k_1^2 + k_2^2 + k_3^2, \tag{19}$$

where $\vec{\xi}$ is a random vector whose components are in $(-1,1)$. Parameters a and b determine the range over which the random variation in $\vec{Y}_{\underline{k}}$ is applied and both are roughly $(k_c + k_{\max})/2$. The amplitude of the randomness is fixed by c and is several angstroms to a nanometer; a, b, and c are positive. The goal is to enrich the ensemble by making c as large as possible without creating many large energy configurations. To minimize CPU time, k_c should not be too large to avoid evaluating many $U_{\underline{k}}$. Several values of $k_{\max}$, k_c, a, b, and c, were explored by numerical experimentation to determine ranges of values that lead to a rich ensemble of configurations, yet minimizing the number of high energy ones. The important point to note in Eq. (19) is that the residuals for larger modes must decrease with wave number, just as the higher order Fourier components of a smooth function must fall to zero. High frequency modes correspond to local motions that could otherwise result in atom overlaps or excessive bond stretching, and these motions must be avoided by
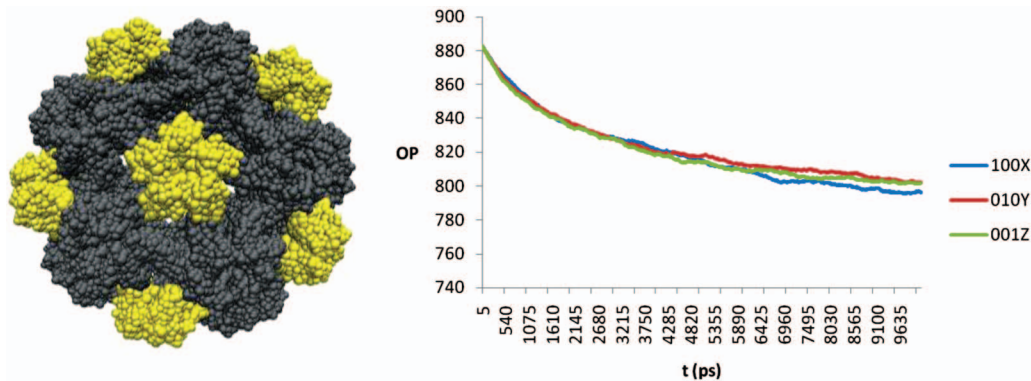
FIG. 2. (a) An all-atom bead representation of the CCMV capsid. The pentamers are shown in a lighter shade, whereas the hexamers filling the spaces between pentamers are shown in a darker shade. (b) The OPs 100X,010Y, and 001Z as a function of time for the first 10ns in an MD simulation. The OPs decrease slowly with time and track the overall shrinking of the capsid.

choosing suitable forms. Also, a simpler prescription using

$$\vec{Y}_{\underline{k}} = \frac{c\vec{\xi}}{b^2 + |\underline{k}|^2}, \quad |\underline{k}|^2 = k_1^2 + k_2^2 + k_3^2, \tag{20}$$

with $b=1$ and $c=0.005$ worked very well to generate OP configurations (for 2^3 OPs) having low energies. The value of c depends on the number of OPs varied; as more higher order OPs are changed, c must be made smaller to counter the effect of the highly oscillating basis functions ($U_{\underline{k}}$). Additional minimization of high energy configurations by annealing the bond lengths/angles before calculating their contribution to the thermal-average forces was used to improve sampling. The thermal-average forces obtained for different values of these parameters are then compared to see if they remain unchanged, i.e., if the sample is sufficiently rich.

We tested our method for calculating thermal-average forces associated with various OPs to better understand OP dynamics they would generate. Large forces change the overall structure of the nanosystem, whereas small ones are present when the system is near a free energy minimum. The relation between OP dynamics and thermal-average forces is not direct since values of the diffusion matrix and the random force both drive the OPs.

The thermal average forces can also be calculated directly by doing short MD runs as the OPs only vary appreciably over longer time scales [Fig. 2(b)]. The instantaneous OP forces for the set of configurations generated during an MD run can be ensemble averaged to obtain the thermal-average forces. Both approaches were used, but a hybrid method was found to be more efficient in which multiple short MD runs are performed starting from configurations generated by the enhanced sampling method. This retains the advantages of both approaches; higher order OPs are well sampled in the first method, while a short MD trajectory minimizes the number of very high energy configurations.

Having calculated the thermal-average forces for a given OP configuration, one may determine free energy minimizing structures (i.e., structures for which the thermal-average forces on all OPs are negligible). Thus, one may map out conditions under which a given thermal structure is stable.

OP free energy minimization (OPFM) starts from an initial structure and performs dynamics with the Langevin noise

set to zero in Eq. (12). Thus, the only factor driving the OPs is the product of the thermal-average forces and the diffusion matrix. A steady state is reached when the thermal-average forces vanish. The thermal-average force and the diffusion matrix define this dynamics by

$$\frac{\partial \vec{\Phi}_{\underline{k}}}{\partial \tau} = \beta \sum_{\underline{k}'} [\vec{\vec{D}}_{\underline{k}\underline{k}'} \vec{f}_{\underline{k}'}]. \tag{21}$$

Multiplying by $\vec{f}_{\underline{k}}$ and summing over k we get

$$\frac{\partial F}{\partial \tau} = -\beta \sum_{\underline{k}\underline{k}'} [\vec{\vec{D}}_{\underline{k}\underline{k}'} \vec{f}_{\underline{k}} \vec{f}_{\underline{k}'}]. \tag{22}$$

Since the diffusion matrix is positive definite, the free energy decreases with time and reaches its minimum value at the end of the minimization. Free energy minimization is based on the observation that exploring the OP space using thermal average forces is more efficient than exploring the complete phase space, as will be shown in the next section.

V. RESULTS

The system studied was the swollen capsid of Cowpea chlorotic mottle virus (CCMV) in vacuum at 300 K. The solvent was not included since the purpose of the present work is to test the new methodology and to get a coherent physical picture of the results and advantages using our method. However, the capsid in vacuum is a good starting point for developing new ideas and methodologies for viral systems. Furthermore, the CCMV capsid already exhibits interesting behavior seen in other well-studied viral structural transitions[7–9,12,37–39] that can be analyzed using our multiscale framework.

The CCMV capsid consists of 180 identical protomers organized as 12 pentamers and 20 hexamers arranged in an icosahedron, having in all 432, 120 atoms. An all-atom bead representation of the virus (created using vmd software) is shown in Fig. 2(a). In vacuum, the swollen structure (radius 140 Å) is unstable and shrinks to reach a native configuration (radius 118 Å). A 10 ns MD trajectory beginning from the swollen X-ray structure[37] was initiated and the configurations generated were stored for later analysis and benchmark-
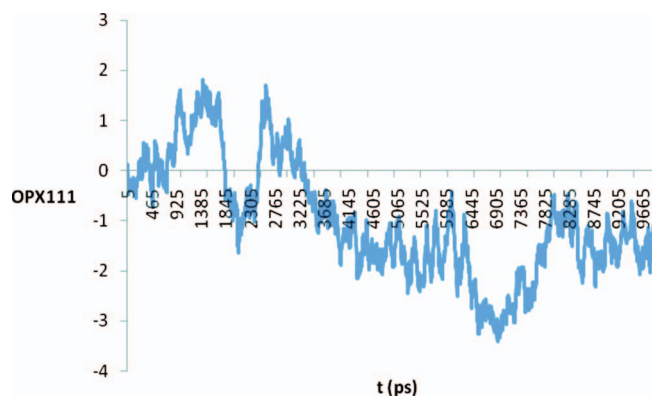
FIG. 3. The higher order OP 111X for the same period as in Fig. 2. Note the much smaller values and the absence of any coherent trend as in the previous graph. The large short-time scale fluctuations suggest that this OP is not suitable for long-time scale Langevin evolution.

FIG. 5. Correlation plot for the thermal-average OP forces and OP velocities for the 001Z component. The lack of a good correlation is due to the fact that the MD effectively includes the random force not included in the present calculation.

ing. Thermal-average forces were calculated for some of these structures using Eq. (16) and our sampling method discussed earlier. Configurations were stored every 5 ps for calculation of OPs, OP velocities, and thermal-average forces. OPs and OP velocities were calculated using Eq. (3). The largest OPs are the X, Y, and Z components of the 100, 010, and 001 modes, respectively. The other OPs like 111 are much smaller in magnitude and correspond to local motions in the capsid. Cases with 2^3 and 3^3 OPs were both studied.

After obtaining the thermal-average forces, we explored the approach to the free energy minimum along a trajectory in the OP space of the capsid, starting from the swollen structure using Eq. (21). The solution of this equation we refer to as OP free energy minimization (OPFM). Each step of OPFM requires calculation of thermal-average forces that are calculated by short atomic scale simulations, as described in Sec. IV. At each step, the forces were calculated by sampling the fixed OP atomic configuration space and the Langevin equation was used to evolve the OPs to their new values. Multiple short MD runs 0.5 ps (50 samples) long after a 1 ps step equilibration were used to enhance the ensemble for calculating OP forces. The new atomic coordinates were calculated from Eq. (2) and these were then used to calculate new forces, and the process was repeated. The time step in the Langevin evolution was chosen optimally so that modest
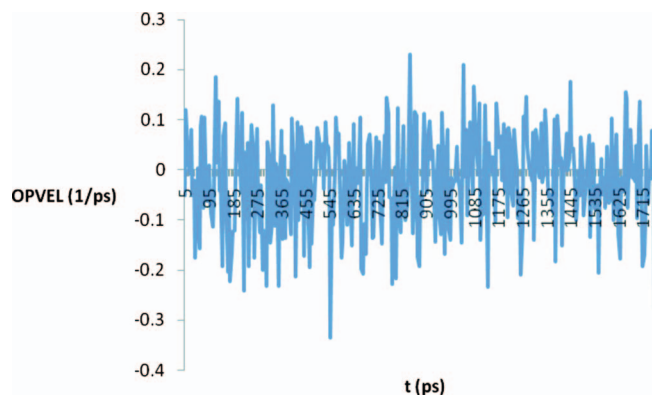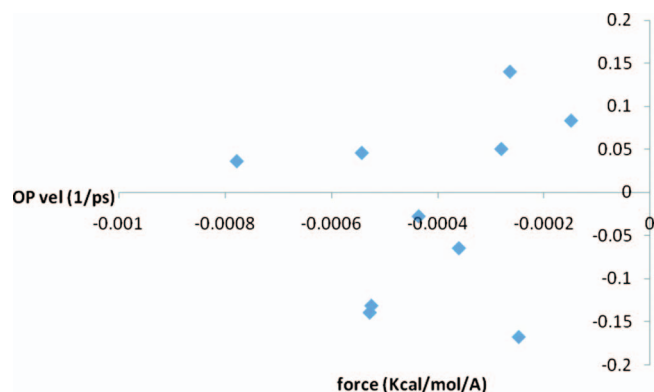

FIG. 4. The OP velocity 100X during the first 2ns of the MD simulation. The negative average velocities correlate with the decrease in the OPs observed in Fig. 2 and the overall shrinking of the capsid.

energy configurations are generated after each evolution step. A very small time step will cause very slow exploration of the OP space while a very large one will lead to unphysical configurations with higher energies when we revert back to atomic coordinates in Eq. (2). 2000–5000 steps of energy minimization were used to reduce high energy Van der Waal's contacts that arise after going from OPs to atomic coordinates. In OPFM, the diffusion constants are taken to be typical values to give an appropriate time scale to the dynamics, although temporal evolution was not the objective here (i.e., we seek only free energy minima). However, as a consistency check the diffusion constant was also calculated to ensure that the time scale matched that of MD dynamics.

The time course of the larger OPs (100X, 010Y, and 001Z) over a 10 ns MD trajectory starting from the swollen capsid is shown in Fig. 2(b). At first the OPs change rapidly and continue changing even after 10ns. Some higher order OPs(111X) are shown in Fig. 3 and they fluctuate more rapidly than the lower order ones and do not show any coherent trend. These OPs correspond to more localized motions of the capsid during its MD evolution and are small in magnitude compared to the major ones in Fig. 2(b). OP velocities for the 100, 010, and 001 modes are negative on the average during this period (Fig. 4). Although the OP velocities are fluctuating, a closer analysis reveals that they are negative on the average and this behavior is consistent with the decrease in OPs in Fig. 2(b). OPs and OP velocities do not decrease monotonically because of the Langevin noise term present implicitly in the MD simulation. Also, at the initial rapid stage of capsid evolution inertial effects may be important and thus a Fokker–Planck, and not a Smoluchowski, description of the dynamics might be more appropriate. The sign of the thermal-average forces was negative when the lower order OPs were decreasing and the capsid shrinking. The thermal-average forces fluctuate just as the OP velocities do, but the forces and velocities shown are for OP configurations separated by 5 ps and their fluctuation is much smaller than that of atomic forces and velocities. Figure 5 shows the correlation between OP forces and OP velocities for the 001Z component. This shows that the OP velocities are not always in the direction of the OP forces because of the random noise term implicit in the MD. However, the OP velocities are
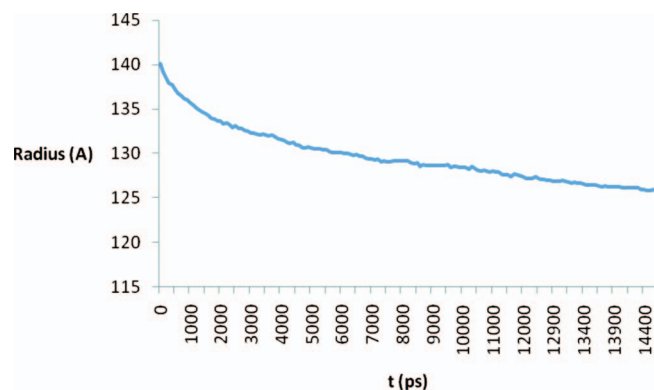
FIG. 6. OPFM trajectory after 200 cycles. The capsid radius has dropped by almost 14 Å. Note the decrease in the rate of shrinking after the rapid initial decrease. The time axis shows that the effective time of the simulation is over 14ns.

FIG. 8. Progress of capsid RMSD from its initial state over a 4ns period obtained with MD and OPFM. The RMSD increases as the capsid shrinks and both trajectories almost overlap.

negative on the average as is seen in the downward trend of the OPs. These results are similar to those of simple Brownian motion where there is no instantaneous correlation between the velocity of the Brownian particle and the random forces in the system even if the Brownian particle shows an overall drift in the direction of a persistent externally applied force.

In Fig. 6, we show a long OPFM trajectory (200 cycles) that starts from the swollen capsid, with resulting shrinkage of more than 14 Å during minimization. The time step used in the Langevin evolution was 10000 in units of (KT/D) and corresponds to a physical time of about 80 ps. The rate of shrinking is rapid at first and then decreases. A closer examination shows that electrostatic energy decreases with radius, lowering the value of the associated thermal-average forces causing it to shrink at a lower rate as time progresses. This is an indication that the structure is slowly approaching its new shrunken state. We have also compared the initial part of the trajectory obtained by OPFM to that obtained by conventional MD. If we match the radius of the structures obtained by OPFM to those from MD, we can reproduce our MD trajectory by assuming that each cycle of OPFM is approximately 80 ps. This is the same number one gets by calculating the diffusion factor from the OP velocity autocorrelation function.[40] By choosing this unit of advancement for OPFM,

we made more detailed comparisons of our trajectory with that obtained by MD. Figure 7 shows the radius of the capsid obtained with the two methods and Fig. 8 shows the progress of the RMSD as a function of time. Agreement of the radius and the RMSD between MD and OPFM is excellent up to 4ns and this further confirms that OP minimization is generating configurations consistent with the same value of the OPs that arise in MD. By extending the OPFM even further, we can reach much longer time scales than we can in MD. This is shown in Fig. 6 where the capsid shrinks by almost 14 Å, and using the unit of time determined earlier corresponds to a MD simulation of almost 14 ns. We note that even after 14 ns the capsid continues shrinking. The complete shrinking of the capsid is expected to take hundreds of nanoseconds. In our OPFM study, smaller and larger time steps of 2500–20000 were used to check for consistency of the results and an optimum value was chosen as a compromise between speed and numerical stability. The thermal-average force for the 001Z component is shown as a function of time in Fig. 9 for the first 4ns. The OP force, large initially, makes the capsid shrink rapidly at first and then decreases in magnitude causing a more gradual reduction in size. The potential energy of the system decreases with de-

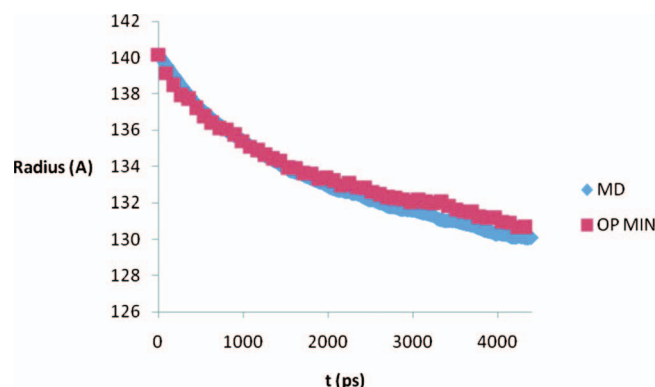FIG. 7. Comparison of capsid radius over a 4ns period obtained with MD and OPFM with both trajectories showing the same rate of contraction. This shows that OPFM captures the nanoscale dynamics.
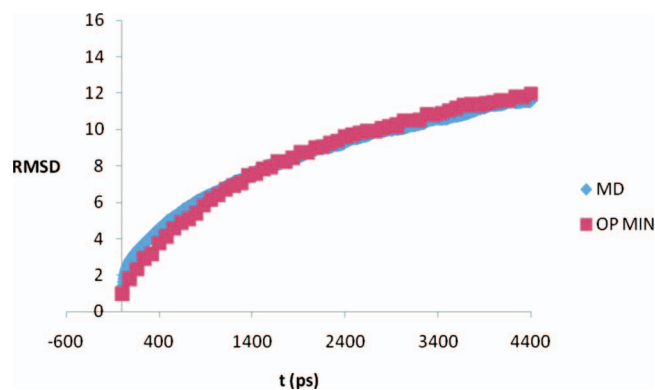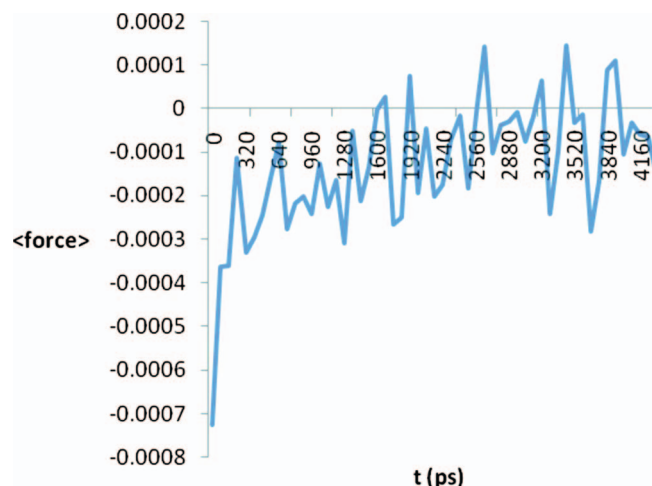

FIG. 9. Thermal-average (001Z) force as a function of time over a 4ns period. The forces are large and negative at first when the capsid is large and then become smaller in magnitude as it shrinks. The units are as in Fig. 5.
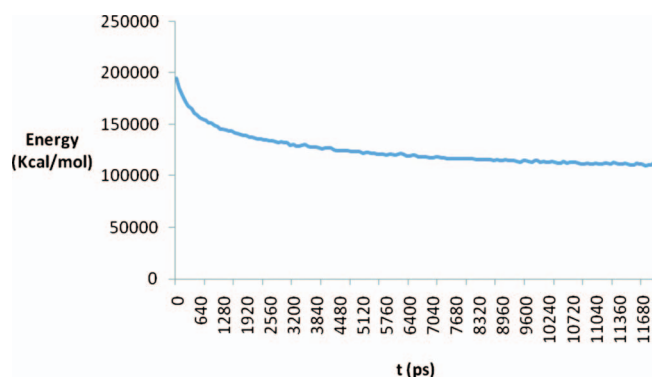
FIG. 10. Potential energy as a function of time during OPFM. The system energy decreases as expected showing that OPFM shrinks the capsid and results in lower energy structures.

creasing radius and is shown in Fig. 10. This shows that the structures generated by OPFM have not only the physically correct geometry but are also energetically more probable in the ensemble sense. As expected, the energies and forces decrease rapidly at first and then decrease more gradually as the capsid keeps shrinking.

In summary, OPFM reproduces the crucial features of a 4ns MD trajectory like, shrinking of the capsid, progress of the RMSD, and decrease in energy that are all observed and expected features of CCMV capsid behavior. By extending our simulations we see further contraction of the capsid by a total of almost 14 Å, a feature observable only in long CPU intensive MD runs. Note that in OPFM each cycle is about 80 ps, a factor of 80, 000 over the step size of 1fs used in MD. We have achieved a considerable acceleration over conventional MD by following Langevin moves in OP space. However, OPFM requires calculating thermal-average forces and solving the Langevin equation, but these operations can be considerably optimized to yield a significant speed-up in overall wall-clock computational time. Our calculations also provide us with thermal-average forces, yielding insights into the complex free energy landscape of the bionanosystem and this information is useful in understanding the overall structural changes in the system. Furthermore, for our OP generated trajectory the appropriate comparison is with an ensemble of replica MD trajectories since we use multiple samples (50) to calculate thermal-average forces and this makes our method even more efficient.

Our simulations were performed using our newly designed SIMNANOWORLD™ platform which incorporates NAMD[41] and by running our programs on the BigRed cluster at Indiana University. For comparison, an ordinary NAMD simulation takes about 1.0day/ns on 128 processors, while our SIMNANOWORLD™ runs at roughly 0.5 days/ns under the same environment. Much further reductions in computer time are possible by further optimizing and integrating our code with NAMD and reducing I/O overheads. Further acceleration is expected when the set of OPs is limited to 001, 010, and 100, as OPs such as 111 are still highly fluctuating, limiting the Langevin step used in this study.

VI. CONCLUSIONS

A novel method for constructing thermal-average forces was demonstrated and these forces were used to explore the

free energy landscape of CCMV capsid. The key to the methodology is the generation of an ensemble of atomistically resolved configurations consistent with the instantaneous values of OPs characterizing the nanometer-scale features of a system. The efficiency of the method was demonstrated, opening the possibility for modeling systems as diverse as viruses, ribosomes and nanocapsules for the delivery of therapeutic agents, all with great generality and without the need for recalibration with each new application. The approach was implemented by accounting for the coupling of OPs and the probability of atomic configurations. This preserves the coupling of processes across scales in space time that is the hallmark of nanosystem dynamics. Our methodology is encoded in the SIMNANOWORLD™ software to be made available at sysbio.indiana.edu and other open source repositories. We have shown that calculation of thermal-average forces, an essential facet of multiscale analysis, can be used to understand large-scale structural changes in viruses and other bionanosystems. More specifically, we have followed the structural transition in the CCMV capsid as it contracts from its swollen state. The multiscale simulation method yields results consistent with an ordinary MD simulation but is faster by a factor of over 100, and has great potential for further optimization.

In ongoing work, we are including the aqueous microenvironment solvent degrees of freedom and the random noise term in the Langevin equation. The inclusion of the solvent is key as it may qualitatively change the behavior of the system.

[1] Y. Miao and P. J. Ortoleva, J. Comput. Chem. **30**, 423 (2009).
[2] P. Ortoleva, P. Adhangale, S. Cheluvaraja, M. Fontus, and Z. Shreif, IEEE Eng. Med. Biol. Mag. **28**, 70 (2009).
[3] S. Pankavich, Y. Miao, J. Ortoleva, Z. Shreif, and P. Ortoleva, J. Chem. Phys. **128**, 234908 (2008).
[4] Z. Shreif, P. Adhangale, S. Cheluvaraja, R. Perera, R. Kuhn, and P. Ortoleva, Sci. Model. Simul. **15**, 363 (2008).
[5] *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, edited by C. Chipot and A. Pohorille (Springer-Verlag, Berlin, 2007).
[6] Z. Shreif and P. Ortoleva, Physica A **388**, 593 (2009).
[7] K. W. Adolph, J. Gen. Virol. **28**, 147 (1975).
[8] J. E. Johnson and J. A. Speir, J. Mol. Biol. **269**, 665 (1997).
[9] J. M. Johnson, J. Tang, Y. Nyame, D. Willits, M. J. Young, and A. Zlotnick, Nano Lett. **5**, 765 (2005).
[10] P. Jonathan and G. Butler, Supermolecular Structure and Function **9**, 131 (2007).
[11] Y. G. Kuznetsov, S. B. Larson, J. Day, A. Greenwood, and A. McPherson, Virology **284**, 223 (2001).
[12] L. Lavelle, J. P. Michel, and M. Gingery, J. Virol. Methods **146**, 311 (2007).
[13] J. E. Shea and I. Oppenheim, J. Phys. Chem. **100**, 19035 (1996).
[14] J. E. Shea and I. Oppenheim, Physica A **247**, 417 (1997).
[15] J. E. Shea and I. Oppenheim, Physica A **250**, 265 (1998).
[16] T. Schlick, Biol. Reprod. **1**, 48 (2009).
[17] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[18] D. M. Zuckerman and E. Lyman, J. Chem. Theory Comput. **2**, 12001202

(2006).

[19] I. G. Kevrekidis, C. W. Gear, and G. Hummer, AIChE J. **50**, 1346 (2004).

[20] J. M. Deutch and I. Oppenheim, Faraday Discuss. Chem. Soc. **83**, 1 (1987).

[21] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).

[22] S. Izvekov and G. A. Voth, J. Phys. Chem. B **109**, 2469 (2005).

[23] S. Izvekov and G. A. Voth, J. Chem. Phys. **123**, 134105 (2005).

[24] M. Iannuzzi, A. Laio, and M. Parrinello, Phys. Rev. Lett. **90**, 238302 (2003).

[25] R. Elber, Biophys. J. **92**, L85 (2007).

[26] K. Jaqaman and P. J. Ortoleva, J. Comput. Chem. **23**, 484 (2002).

[27] Y. Miao and P. Ortoleva, J. Chem. Phys. **125**, 044901 (2006).

[28] P. Ahlrichs and B. Dünweg, J. Chem. Phys. **111**, 8225 (1999).

[29] *The Monte Carlo Method in Condensed Matter Physics*, edited by K. Binder (Springer, New York, 1992).

[30] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Methuen, London, 1975).

[31] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[32] U. E. Hansmann, Chem. Phys. Lett. **281**, 445 (1997).

[33] J. Kim, J. E. Straub, and T. Keyes, Phys. Rev. Lett. **97**, 050601 (2006).

[34] H. Meirovitch, S. Cheluvaraja, and R. P. White, Curr. Protein Pept. Sci. **10**, 229 (2009).

[35] R. Radhakrishnan and T. Schlick, J. Chem. Phys. **121**, 2436 (2004).

[36] J. M. J. Swanson, R. H. Henchman, and J. A. McCammon, Biophys. J. **86**, 67 (2004).

[37] J. A. Speir, S. Munshi, G. Wang, T. S. Baker, and J. E. Johnson, Structure **3**, 63 (1995).

[38] A. Arkhipov, P. L. Freddolino, and K. Schulten, Structure **14**, 1767 (2006).

[39] P. L. Freddolino, A. Arkhipov, S. Larson, A. McPherson, and K. Schulten, Structure **14**, 437 (2006).

[40] P. Adhangale, S. Cheluvaraja, and P. Ortoleva, "Roadmap for Virtual Virus™: A multiscale whole-virus simulator," (unpublished).

[41] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).