JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Hierarchical Order Parameters for Macromolecular Assembly Simulations. 1. Construction and Dynamical Properties of Order Parameters

Abhishek Singharoy, Yuriy Sereda, and Peter J. Ortoleva*

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

Supporting Information

ABSTRACT: Macromolecular assemblies often display a hierarchical organization of macromolecules or their subassemblies. To model this, we have formulated a space warping method that enables capturing overall macromolecular structure and dynamics via a set of coarse-grained order parameters (OPs). This article is the first of two describing the construction and computational implementation of an additional class of OPs that has built into them the hierarchical architecture of macromolecular assemblies. To accomplish this, first, the system is divided into subsystems, each of which is described via a representative set of OPs. Then, a global set of variables is constructed from these subsystem-centered OPs to capture overall system organization. Dynamical properties of the resulting OPs are compared to those of our previous nonhierarchical ones, and implied conceptual and computational advantages are discussed for a 100 ns, 2 million atom solvated human Papillomavirus-like particle simulation. In the second article, the hierarchical OPs are shown to enable a multiscale analysis that starts with the *N*-atom Liouville equation and yields rigorous Langevin equations of stochastic OP dynamics. The latter is demonstrated via a force-field-based simulation algorithm that probes key structural transition pathways, simultaneously accounting for all-atom details and overall structure.

I. INTRODUCTION

Macromolecular assemblies exhibit a complex hierarchy of structural organization.[1,2] For example, viruses display a hierarchical organization of atoms forming protomers, pentamers, or hexamers that ultimately assemble into capsids via different types of bonded and nonbonded interactions. Such architecture provides several advantages to biomolecular structures. For example, the existence of hierarchical organization in macromolecular assemblies optimizes the efficiency with which they are assembled and surface area available for interaction with the environment, while simultaneously minimizing the amount of building materials, energy expended, and metabolic cost.[3] Hierarchical organization also leads to novel forms of control: the higher level structure constrains the motion and, hence, position or reactivity of the smaller subcomponents.[3] Thus, hierarchy results in the multiple space and time scale character of macromolecular structure and associated pathways of assembly. An all-atom model, capturing the cross-talk among various scales in space and time is often necessary to completely elucidate the structure and function of these assemblies.[4−6] However, direct computational implementation of such models is limited by the system size, as reflected in hardware requirements.

Interestingly, the very hierarchical nature of macromolecular assemblies can be mapped onto computational algorithms for designing reduced dimensionality frameworks. For example, collective motions of multiple groups of atoms facilitate the hierarchical organization of assemblies as they promote orderly binding of intermediates consisting of several monomers.[7] The number of modes capturing such motion is much less than the total atomic degrees of freedom. This implies that macromolecular models based only on coherent modes involve tracking a much smaller number of dynamical variables relative to those based on all-atom description. Thus, the computational cost of implementing these reduced dimensional models is moderate. This idea has resulted in the development of coarse-grained models for macromolecules and their assemblies. An incomplete list would include bead,[8,9] shape-based,[10] rigid region decomposition,[11] symmetry constrained,[12] and curvilinear coordinate[13] models, as well as Principal Component Analysis (PCA) and Normal Mode Analysis (NMA) guided approaches.[14−17] These models have been successful in investigating structural transitions in a very rich set of biomolecules including alanine polypeptides,[18] ligand binding proteins,[19] transmembrane proteins,[20,21] RNA segments,[22,23] and virus capsids of different symmetries.[10,24] However, they suffer from one or more of the following difficulties: (1) Characteristic variables are not slowly varying in time. (2) Nonlinear motions like macromolecular twist are not readily accounted for. (3) Internal dynamics, and hence inelasticity of collisions, is neglected. (4) Symmetry breaking processes cannot be accounted for. (5) Forces involved must be calibrated for most new applications. (6) Generating intermediate all-atom trajectories for "on-the-fly" dimensionality reduction becomes very expensive for large systems.

Significant effort has been devoted in furthering the accuracy and sensitivity of the coarse-grained models. Notable progress would include techniques like Multiscale Coarse Graining where the atomic and coarse-grained representations are coupled via boundary conditions,[20] variational coarse-graining where coarse-grained sites on centers of masses of various subgroups of C_α atoms are identified via PCA or NMA,[25] and Reconstruction Algorithm for Coarse-Grained Structures where

Received: August 16, 2011
Published: March 13, 2012

low-energy all-atom protein structures can be constructed from configurations with only C_α atoms.[26] Structural fluctuations and internal dynamics are the central features of several biological processes. For example, in the presence of energy barriers, the fluctuations are utilized to allow for the self-organization of lipids in membranes.[27] Fluctuations are also important in expressing the conformational diversity of macromolecules that allows for large deformations upon drug binding. One of the suggested ways to incorporate internal dynamics in coarse-grained models is via superposition of low frequency quasiharmonic modes. This idea has been demonstrated on the dynamics of a dimeric enzyme orotidine 5-monophosphate decarboxylase.[28] However, this model, as well as several others (reviewed in ref 18), does not provide all-atom details.

In a series of studies, we have discovered novel multiscale techniques that probe the cross-talk among multiple scales in space and time and preserve the key all-atom details of the macromolecular assemblies.[4−6,29−34] This is achieved via the introduction of a set of space warping order parameters (OPs) that describe coherent, overall structural changes, while our mathematical reformulation of the underlying molecular physics simultaneously captures high frequency atomic fluctuations.[30,33,35,36] In effect, our OPs filter out high frequency atomistic fluctuations to track low frequency coherent modes that enable coarse-grained description of the system.[30,32] These slow variables capture polyalanine folding from a linear to a globular state,[35] Ostwalds ripening in nanocomposites,[29,35] nucleation and front propagation pathways in a virus capsid structural transition,[31] counterion induced collapse in viral RNA, and stability of RNA−protein complexes.[32] The OPs are generalized center of mass variables that include a strain tensor accounting for visco-elastic deformations and hence probe complex motions like macromolecular twist or bend.[32,35] They serve as the basis of a multiscale analysis that starts with the N-atom Liouville equation and yields rigorous Smoluchowski/Langevin equations[4,29] and the Fokker−Planck equation[34] of stochastic OP dynamics. The Langevin equations have been numerically implemented via a force-field-based algorithm in our ***SimNanoWorld*** software.[32,33] However, as mentioned earlier,[32,35] interpretation of all motions associated with these OPs is nontrivial. This is in part due to the very definition of our OPs, and more generally due to the many-to-one mapping between all-atom and reduced representations involved with any coarse-graining technique.[20] Henceforth, these OPs are referred to as the nonhierarchical ones.

Here, we define a new set of OPs that takes more explicit account of the hierarchical nature of assembly architecture. First, the system is divided into subsystems, each of which is described via a representative set of OPs. Then, a global set of variables is constructed from these subsystem-centered OPs to capture overall organization of the composite system. The structural subsystems (e.g., protomers, pentamers, or hexamers within a virus capsid) are considered as separate beads with associated subsystem-centered OPs probing their position, orientation, and internal structure. Overall bead deformations like extension, compression, rotation, and translation, as well as resulting interbead motions that probe temporal dynamics of the assembly are captured via the global set of OPs. In addition, high frequency fluctuations within each subsystem and across the assembly are captured via a quasi-equilibrium ensemble of all-atom configurations. Thus, assembly structure and dynamics is understood in terms of atomistic fluctuations, intrasubsystem and global OPs, and the interplay between them. Consequently, the hierarchical OP approach ensures that higher level structure mediates subsystem and atomistic dynamics. The reverse transfer of information (from shorter to longer scales) is shown in the second part of this presentation[37] to emerge from the effect of short-scale processes on factors in the dynamical equations for the global variables. In this way, the present hierarchical framework is a natural way to understand biological phenomena involving hierarchical organization. In the rest of this article and the following one, the terms "subsystem" and "bead" are used interchangeably.

In the following, we construct the hierarchical OPs and show that they satisfy key criteria for being the basis of a multiscale methodology and simulation approach (section II). We numerically demonstrate the properties of hierarchical OPs by comparing them to those of our nonhierarchical ones using all-atom configurations derived from Molecular Dynamics (MD) and ***SimNanoWorld*** simulations of T = 1 L1 *Human Papilloma Virus* 16 (HPV16) Virus-Like Particle (VLP) disassembly/collapse (section III), and discuss conceptual and computational advantages of the hierarchical approach (section IV). Conclusions are drawn in section V. In the second article,[37] we formulate a multiscale methodology that captures assembly dynamics through the coupled evolution of hierarchical OPs and all-atom ensembles. Applicability of Langevin equations to hierarchical OP dynamics is tested. Advantages of implementing these OPs within a practical simulation framework are demonstrated in the context of computational efficiency and tracking emergence of new collective modes. Finally, relative accuracy and efficiencies of the OP enabled multiscale and MD simulations are evaluated.

II. CONSTRUCTION OF HIERARCHICAL ORDER PARAMETERS

A central element of our multiscale analysis is the construction of OPs that describe the coarse-grained features of a macromolecular assembly. The method for constructing these OPs is built on techniques presented in earlier studies.[30,31,35] A description of the latter and detailed extensions that constitute the current work are provided below.

A. Nonhierarchical Order Parameters. Consider a macromolecular assembly described via the positions of its N constituent atoms $\vec{r}_i$ labeled i = 1, ..., N. In our approach, $\vec{r}_i$ is related to a reference position $\vec{r}_i^{\,0}$. Deformation of space taking $\vec{r}_i^{\,0}$ to $\vec{r}_i$ is continuous and is used to introduce OP $\vec{\Phi}_{\underline{k}}$ via the transformation

$$\vec{r}_i = \sum_{\underline{k}} \vec{\Phi}_{\underline{k}} U_{\underline{k}i} \tag{1}$$

where $\vec{\Phi}_{\underline{k}}$ is termed the $\underline{k}$th nonhierarchical OP. As the $\vec{\Phi}_{\underline{k}}$ changes, space is deformed, and so are the macromolecules embedded in it. The factor $U_{\underline{k}i}$ is defined in terms of a basis function $U_{\underline{k}}(\vec{r}_i^{\,0})$, i.e., $U_{\underline{k}i} \equiv U_{\underline{k}}(\vec{r}_i^{\,0})$ for reference position $\vec{r}_i^{\,0}$ of atom i. Index $\underline{k}$ labeling the $\vec{\Phi}$ is a set of three integers $\{k_1k_2k_3\}$ such that $U_{\underline{k}i} \equiv U_{k_1k_2k_3}(\vec{r}_i^{\,0})$ is a product of Legendre polynomials of orders k_1, k_2, and k_3 for the X, Y, and Z components of $\vec{r}_i^{\,0}$, respectively, i.e., $U_{k_1k_2k_3}(\vec{r}_i^{\,0}) = U_{k_1}(X_i^0)\, U_{k_2}(Y_i^0) U_{k_3}(Z_i^0)$.[33] The curly bracket {} is used to denote a set of variables. Since we seek a dimensionality reduction, the

number of $\vec{\Phi}_{\underline{k}}$ is much less than the number N of atoms. Thus, we take a finite truncation of the $\underline{k}$ sum in eq 1; this necessitates the introduction of a residual (denoted $\vec{\sigma}_i$) to correct the coherent deformation generated by the $\vec{\Phi}_{\underline{k}}$. With this,

$$\vec{r}_i = \sum_{\underline{k}} \vec{\Phi}_{\underline{k}} U_{\underline{k}i} + \vec{\sigma}_i \tag{2}$$

To maximize information content of the OPs, the magnitude of $\vec{\sigma}_i$ is minimized by the choice of basis functions and the number of terms in the $\underline{k}$ sum. Conversely, imposing a permissible size threshold for the residuals allows one to determine the number of terms to include in the $\underline{k}$ sum.

Enhancing the information content in $\vec{\Phi}_{\underline{k}}$ is achieved by minimizing the sum of mass-weighted square residuals ($m_1\sigma_1^2$ + ...$m_N\sigma_N^2$) with respect to the $\vec{\Phi}_{\underline{k}}$, where m_i is the mass of atom i. This implies

$$\sum_{\underline{k}'} B_{\underline{k}\underline{k}'} \vec{\Phi}_{\underline{k}'} = \sum_{i=1}^{N} m_i U_{\underline{k}i} \vec{r}_i,\ B_{\underline{k}\underline{k}'} = \sum_{i=1}^{N} m_i U_{\underline{k}i} U_{\underline{k}'i} \tag{3}$$

For convenience, we reorganize the factors $U_{\underline{k}i}$ to be mass-weighted orthogonal. In that case, the B matrix in eq 3 is diagonal; i.e., $B_{\underline{k}\underline{k}'}$ is 0 for $\underline{k} \neq \underline{k}'$. With this,

$$\vec{\Phi}_{\underline{k}} = \frac{\sum_{i=1}^{N} m_i U_{\underline{k}i} \vec{r}_i}{\tilde{\mu}_{\underline{k}}};\ \tilde{\mu}_{\underline{k}} = \sum_{i=1}^{N} m_i U_{\underline{k}i}^2 \tag{4}$$

Here, $\tilde{\mu}_{\underline{k}}$ serves as an effective mass associated with $\vec{\Phi}_{\underline{k}}$, giving the latter a generalized center of mass character. In particular, $\vec{\Phi}_{\underline{k}}$ for $\underline{k}$ = {000} is the center of mass (CM), and those for $\underline{k}$ = {100, 010, 001} constitute a strain tensor describing overall deformations like extension−compression−rotation.[32]

Stochastic evolution of the nonhierarchical OPs is captured via Langevin-type equations.[4−6] These OPs are uniquely defined for a set of atomic positions (eq 4). However, the converse is not true; i.e., there exist multiple all-atom configurations consistent with a given set of $\vec{\Phi}_{\underline{k}}$. Thus, the OP construction scheme constitutes a many-to-one mapping from the all-atom to the coarse-grained description. As a consequence, the OP description and associated Langevin equations retain overall structural information, losing all-atom details. To restore all-atom resolution, an equilibrium ensemble approach is introduced that allows generation of atomic configurations consistent with the overall state of the system provided by a set of OPs, $\underline{\Phi} \equiv \{\vec{\Phi}_{\underline{k}}\}$. These multiscale ideas, implemented in our Φ-based ***SimNanoWorld*** simulation software, are summarized in section SI1 of Supporting Information.

B. Subsystem-Centered Order Parameters. In the previous subsection, one set of OPs is introduced for the whole system. However, macromolecular assemblies like viruses and ribosomes are composed of several subunits organized in a structure that can spontaneously change symmetry (e.g., during ribosomal assembly or capsid maturation). This organization is naturally captured in terms of OPs that characterize the deformations of individual subsystems labeled S = 1, ..., N^{sys}. Thus, in this framework, a system is divided into N^{sys} subsystems, each one of which is described by different sets of subsystem-centered OPs denoted $\vec{\Phi}^S_{\underline{k}}$. In analogy to eqs 1 and 2, $\vec{\Phi}^S_{\underline{k}}$ is introduced for subsystem S via

$$\vec{r}^S_j = \sum_{\underline{k}} \vec{\varphi}^S_{\underline{k}} U^S_{\underline{k}j} + \vec{\sigma}^S_j \tag{5}$$

where $\vec{\sigma}^S_j$ denotes the residual displacement of atom j in subsystem S resulting from a finite truncation of the $\underline{k}$ sum. Using the same series of transformations as for the nonhierarchical OPs (but restricted to individual subsystems) yields

$$\vec{\varphi}^S_{\underline{k}} = \frac{\sum_{j=1}^{n^S} m_j^S U^S_{\underline{k}j} \vec{r}^S_j}{\tilde{\mu}^S_{\underline{k}}};\ \tilde{\mu}^S_{\underline{k}} = \sum_{j=1}^{n^S} m_j^S (U^S_{\underline{k}j})^2 \tag{6}$$

where n^S is the total number of atoms in the subsystem S. m_j^S and $\vec{r}^S_j$ are the mass and position of the jth atom in S. $U_{\underline{k}i} \equiv U_{\underline{k}}(\vec{r}^{S,0}_j)$, where $\vec{r}^{S,0}_j$ is the reference position of atom j. Details of derivations yielding eq 6 are provided in section SI2 of the Supporting Information.

C. Hierarchical Order Parameters. As system complexity (e.g., the number of subsystems and their internal structure) increases, one may increase the number of OPs, i.e., the range of the $\underline{k}$ sum in eq 2 or 5, to capture smaller spatial features that tend to evolve on shorter time scales.[35] With this, the characteristic rate of change of nonhierarchical or subsystem-centered OPs increases. However, a separation in time scale between the coherent (slow) and noncoherent (fast) degrees of freedom is necessary for probing the multiscale behavior of macromolecular assemblies.[4] To address this, hierarchical OPs $\vec{\Psi}_{\underline{K}}$ are introduced, which maintain the necessary separation of characteristic time scales between the slow and fast degrees of freedom and simultaneously capture the overall assembly deformations as well as those internal to the subsystem architecture.

The $\vec{\Psi}_{\underline{K}}$'s are constructed using a three-level description of the assembly. First, the N-atom assembly is divided into N^{sys} subsystems, the Sth of which contains n^S atoms. The configuration within subsystem S is provided by the set of the n^S positions denoted $\underline{r}^S \equiv \{\vec{r}^S_j\}$; j = 1, ..., n^S. The set of positions of all atoms in the system is denoted $\underline{r} \equiv \{\underline{r}^S\}$; S = 1, ..., N^{sys}. At an intermediate level, each subsystem is described via a finite set of the subsystem-centered OPs $\underline{\varphi}^S \equiv \{\vec{\varphi}^S_{\underline{k}}\}$. At the highest level, a global set of variables, $\underline{\Psi} \equiv \{\vec{\Psi}_{\underline{K}}\}$, is constructed from these subsystem-centered OPs to capture the overall organization of the system.

In analogy with eq 2, the hierarchical OPs $\vec{\Psi}_{\underline{K}}$ are introduced as

$$\vec{\varphi}^S_{\underline{k}} = \sum_{\underline{K}} \vec{\Psi}_{\underline{K}} U^S_{\underline{K}\underline{k}} + \vec{\varphi}^S_{\underline{k}} \tag{7}$$

where $U_{\underline{K}\underline{k}}{}^S$ depends on the subsystem-centered OPs $\vec{\varphi}^S_{\underline{k}}$ in a manner similar to which the $U_{\underline{k}i}$ depend on the all-atom configuration $\vec{r}_i^{\,0}$ (eq 1), i.e., $U_{\underline{K}}(\vec{\varphi}^S_{\underline{k}}) \equiv U_{K_1K_2K_3}(\vec{\varphi}^S_{\underline{k}}) = U_{K_1}(\varphi^S_{\underline{k}X}) U_{K_2}(\varphi^S_{\underline{k}Y}) U_{K_3}(\varphi^S_{\underline{k}Z})$. Here, $\varphi_{\underline{k}\alpha}{}^S$ is αth Cartesian

component of $\vec{\varphi}^S_{\underline{k}}$. However, unlike our nonhierarchical formulation, here $U_{\underline{K}\underline{k}}{}^S$ depends on $\underline{\varphi}^S$ and not a fixed reference configuration of subsystem-centered OPs $\underline{\varphi}^{0,S}$.

Since $\tilde{\mu}_{\underline{k}}{}^S$ serves as an effective mass for $\vec{\varphi}^S_{\underline{k}}$ (eq 6), minimization of mass-weighted square residual $\sum_{S,\underline{k}}\tilde{\mu}_{\underline{k}}{}^S(\vec{\sigma}^S_{\underline{k}})^2$ with respect to $\vec{\varphi}^S_{\underline{k}}$ yields

$$\sum_{\underline{K}'} B_{\underline{K}\underline{K}'}\vec{\Psi}_{\underline{K}'} = \sum_{S,\underline{k}} \tilde{\mu}^S_{\underline{k}} U^S_{\underline{K}\,\underline{k}}\vec{\varphi}^S_{\underline{k}};\quad B_{\underline{K}\underline{K}'} = \sum_{S,\underline{k}} \tilde{\mu}^S_{\underline{k}} U^S_{\underline{K}\,\underline{k}} U^S_{\underline{K}'\,\underline{k}} \tag{8}$$

Again, for convenience, we choose the basis functions $U_{\underline{K}}$ to be mass-weighted orthogonal. In that case, the B matrix (eq 8) is diagonal, i.e., $B_{\underline{K}\underline{K}'}$ is 0 for $\underline{K} \neq \underline{K}'$. With this,

$$\vec{\Psi}_{\underline{K}} = \frac{\sum_{S,\underline{k}} \tilde{\mu}^S_{\underline{k}} U^S_{\underline{K}\,\underline{k}}\vec{\varphi}^S_{\underline{k}}}{\tilde{\mu}_{\underline{K}}};\ \tilde{\mu}_{\underline{K}} = \sum_{S,\underline{k}} \tilde{\mu}^S_{\underline{k}}(U^S_{\underline{K}\,\underline{k}})^2 \tag{9}$$

The above summation runs through all subsystems, S = 1, ...N^{sys}, and a selected range of $\underline{k}$ that minimizes the subsystem-centered residuals $\vec{\sigma}^S_j$ (eq 5) for each one of these subsystems. Equation 9 expresses overall structural characteristics of the assembly ($\vec{\Psi}_{\underline{K}}$) as a function of position, orientation, and deformation of individual subsystems ($\vec{\varphi}^S_{\underline{k}}$). Thus, $\underline{r}$ explicitly defines $\underline{\varphi}^S$ (eq 6), which in turn yields $\underline{\Psi}$ (eq 9), thereby reflecting the structural hierarchy of an assembly.

For simplicity, the $U_{\underline{K}\underline{k}}{}^S$ can be constructed using only the subsystem CMs, i.e., $\vec{\varphi}^S_{\underline{0}}$.[32] Letting $\vec{R}^S \equiv \vec{\varphi}^S_{\underline{0}}$ be the CM of subsystem S, eq 9 becomes

$$\vec{\Psi}_{\underline{K}} = \frac{\sum_{S=1}^{N^{\mathrm{sys}}} M^S U^S_{\underline{K}}\vec{R}^S}{\tilde{\mu}_{\underline{K}}};\ \tilde{\mu}_{\underline{K}} = \sum_{S=1}^{N^{\mathrm{sys}}} M^S\{U^S_{\underline{K}}\}^2 \tag{10}$$

where $M^S = \tilde{\mu}_{\underline{0}}$ is the total mass of subsystem S. As with eq 7, $U_{\underline{K}}{}^S \equiv U_{\underline{K}}(\vec{R}^S)$ is constructed from the collection of CMs $\underline{R}^S$ and not a fixed reference configuration of subsystem CMs $\underline{R}^{S,0}$. A detailed numerical procedure for the construction of basis factors and OPs is provided in section SI3 of the Supporting Information. Briefly, Legendre polynomials are arranged in one-dimensional matrices which are then Gram−Schmidt orthogonalized with a mass-weighted metric to obtain the basis factors U that diagonalize the B matrix (eqs 3 and 8). The product of the position matrix with these orthogonalized basis-set polynomial matrices yields the OPs.

Now, we make use of the Liouville equation to elucidate the rate of OP dynamics. The Liouville operator is defined via $L = -\sum_{i=1}^{N}[(\vec{p}_i/m_i)\cdot(\partial/\partial\vec{r}_i) + \vec{F}_i\cdot(\partial/\partial\vec{p}_i)]$, where $\vec{p}_i$ and $\vec{F}_i$ are the momentum of and the net force on atom i, respectively. Given eq 10, one may compute $d\underline{\Psi}/dt$ as $-L\underline{\Psi}$. With this, the rate of change of $\vec{\Psi}_{\underline{K}}$ is given by

$$\frac{\partial\vec{\Psi}_{\underline{K}}}{\partial t} = \frac{1}{\tilde{\mu}_{\underline{K}}}\sum_S \vec{P}^S\left\{U^S_{\underline{K}} + \vec{R}^S\cdot\frac{\partial U^S_{\underline{K}}}{\partial\vec{R}^S} - 2U^S_{\underline{K}}\left(\vec{\Psi}_{\underline{K}}\cdot\frac{\partial U^S_{\underline{K}}}{\partial\vec{R}^S}\right)\right\} \tag{11}$$

where $\vec{P}^S = \sum_{j=1}^{n^S}\vec{p}^S_j$ is the total momentum of subsystem S. Details of derivation yielding eq 11 are provided in Appendix A.

The second and third terms in eq 11 arise from the spatial derivative of $U_{\underline{K}}{}^S$; such terms are negligible if the $U_{\underline{K}}{}^S$ is constructed using polynomials that vary slowly across the system. The latter is realized by constructing polynomials using a fixed reference configuration of CMs. In this case, $\underline{\Psi}$ characterizes deformation from a fixed structure, in analogy with our earlier formulation.[30] This implies

$$\frac{\partial\vec{\Psi}_{\underline{K}}}{\partial t} = \frac{1}{\tilde{\mu}_{\underline{K}}}\sum_S U^S_{\underline{K}}\vec{P}^S \tag{12}$$

However, effects of the omitted terms in eq 11 on the applicability of a fixed reference structure in constructing the basis-set polynomials are discussed in section III.

To reveal the time scale on which $\vec{\Psi}_{\underline{K}}$ evolves, it is convenient to define a smallness parameter $\varepsilon = m/M$, where m is a typical atomic mass and M is the mass of the whole assembly. For any of $\vec{\Psi}_{\underline{K}}$, letting $\vec{v}^S_j$ be the velocity of particle j in subsystem S, the definition of ε and eq 12 yields

$$\frac{d\vec{\Psi}_{\underline{K}}}{dt} = \varepsilon\frac{\sum_{S=1}^{N^{\mathrm{sys}}}\sum_{j=1}^{n^S}\hat{m}^S_j U^S_{\underline{K}}\vec{v}^S_j}{\mu_{\underline{K}}} = \varepsilon\frac{\vec{\Pi}_{\underline{K}}}{\mu_{\underline{K}}} \tag{13}$$

where $\mu_{\underline{K}} = \tilde{\mu}_{\underline{K}}/M$, $\hat{m}^S_j = m^S_j/m$ and $\vec{\Pi}_{\underline{K}}$ is the hierarchical OP momentum. Equation 13 suggests that the hierarchical OPs $\vec{\Psi}_{\underline{K}}$ change slowly, at a rate of $O(\varepsilon)$. This is under the assumption that the $\vec{\Pi}_{\underline{K}}$ are linear combinations of individual atomic momenta, which tend to cancel near equilibrium, thereby reducing the rate of OP evolution. Furthermore, there is an ensemble of rapidly fluctuating all-atom configurations consistent with a given set of $\vec{\Psi}_{\underline{K}}$. When the OPs vary on long time scales the Gibbs hypothesis suggests that this ensemble is equilibrium in character. Thus, the set of atomistic configurations visited during a period that is long relative to the characteristic time of interatomic fluctuations yet short relative to those of the slow $\vec{\Psi}_{\underline{K}}$ evolution can be represented by an equilibrium ensemble at fixed $\vec{\Psi}_{\underline{K}}$.[4,29] A numerical procedure for generation of these ensembles is discussed in Appendix B. Atomic configurations from this ensemble capture key internal dynamics within subsystems for which the overall assembly structure is characterized by $\underline{\Psi}$ and subsystem level structure is described by $\underline{\varphi}^S$.

Note that all functions of the atomistic configuration are not slowly varying in time. Even the OPs presented here evolve on a range of time scales related to the spatial scale of motions they

characterize, e.g., assembly-wide versus single monomer motion. From among them, there is a subset which evolves on a time scale much greater than that of individual atomistic fluctuations. The subset manifests the time scale separation between slow and fast variables needed to develop a multiscale simulation algorithm (as discussed in part 2).[37] This separation is made explicit via (a) qualitative arguments used above to justify eq 13 and (b) numerical demonstration of slow OP dynamics, as provided below and in the second part. In the next section, we numerically construct the hierarchical OPs and analyze their pros and cons with respect to those of our previous nonhierarchical ones.

III. NUMERICAL VALIDATION

A study is undertaken to assess the viability of the hierarchical OPs of section II as collective variables that capture the structure and dynamics of macromolecular assemblies. We compare properties of the hierarchical OPs to those of the nonhierarchical ones using all-atom structures from NAMD[38] and ***SimNanoWorld*** simulations with the CHARMM27 force field.[39,40] Details of conditions used for these simulations are provided in Table 1. An evaluation of other coarse-grained variables in this context has been presented previously.[32]

Table 1. Input Parameters for the NAMD and *SimNanoWorld* Simulations

parameter	values
Temperature	300 K
Langevin damping	5
Timestep	1 fs
fullElectFrequency	2 fs
nonbondedFreq	1 fs
Box size	330 Å × 330 Å × 330 Å
Force-field parameter	par_all27_prot_na.prm
1−4scaling	1.0
Switchdist	10.0 Å
Cutoff	12.0 Å
Pairlistdist	20.0 Å
Stepspercycle	2
Rigid bond	Water
OP time step	125 ps

The $T = 1$ L1 HPV16 VLP is the demonstration system. This assembly contains 12 pentamers joined by "attacking arms" that stabilize it via strong hydrophobic interactions.[41,42] Each pentamer is composed of five L1 protein monomers. The C-terminal of the L1 protein consists of four helical regions h2, h3, h4, and h5 that are responsible for intra- and inter-pentameric stabilization. While h2, h3, and h5 are responsible for L1 protein folding and pentameric stability, h4 maintains interpentamer connectivity and, thereby, overall $T = 1$ structure.[41,42] As suggested by experiments,[42] we truncated h2, h3, and h4 from the L1 protein and simulated the disassembly/collapse of the resulting VLP via ***SimNanoWorld***. The simulation elucidates all-atom dynamics of this 2×10^6 atom system (including water and ions, 0.25 M NaCl) for 100 ns (Figure 1), revealing mechanisms of the transition pathway. Since this transition involves the sequential break-up of an assembly into finer scale components, it provides an ideal example for demonstrating the viability of $\vec{\Psi}_{\underline{K}}$ as variables that capture hierarchical organization. Thus, all-atom configurations from this trajectory are used to reconstruct the evolution of hierarchical and nonhierarchical OPs. With this, advantages of the hierarchical OPs over the nonhierarchical ones for simulating macromolecular assembly dynamics are evaluated.

It has been previously shown that a $\underline{\Phi}$-mediated ***SimNanoWorld*** trajectory is equivalent to an ensemble of MD generated ones, but the former is generated much more efficiently.[32] Several simulation controls are implemented in ***SimNanoWorld*** to ensure this consistency.[43] Thus, for the numerical comparison between $\underline{\Psi}$ and $\underline{\Phi}$ presented below, atomic structures are chosen from the $\underline{\Phi}$-mediated trajectory. These structures are equivalent to those from a 100 ns MD, but generating them for the million atom system under consideration is much more efficient via our multiscale simulation.[43] To further validate this point, in the second article,[37] structures from multiscale trajectories of the HPV VLP are compared to those from all-atom MD ones and experimental observations.

In the following, we numerically construct hierarchical OPs ($\vec{\Psi}_{\underline{K}}$) and compare their properties with those of non-hierarchical ones ($\vec{\Phi}_{\underline{k}}$).[32,35] Observed differences between these two classes of OPs are rationalized, and implied advantages for conceptualizing a hierarchical OP based simulation algorithm are discussed. The rest of section III illustrates various facets of $\vec{\Psi}_{\underline{K}}$ evolution.

A. Criteria for Defining Subsystems. A cornerstone in defining hierarchical OPs is the specification of subsystem boundaries. The latter explicitly imply characteristics of the hierarchical OPs. Let each of the 12 pentamers that constitute the VLP be a subsystem (Figure 1c). Subsystem S ($S = 1, \ldots 12$) is described by four subsystem-centered vector OPs $\vec{\varphi}^{S}_{\underline{k}}$ (i.e., the CM and three corresponding to overall subsystem extension−compression−rotation, indices $\underline{k} = \{000, 100, 010, 001\}$). The choice of these $\vec{\varphi}^{S}_{\underline{k}}$ is justified by their dynamical coherence, as observed earlier,[32,33] and in the next subsection. OP specification using the triplet indices $\underline{k}$ is discussed in section II.

The 100 ns $\underline{\Phi}$-mediated ***SimNanoWorld*** simulation provides an ensemble of atomic configurations at every Langevin time step. OP evolution trajectories are obtained by computing the $\vec{\Psi}_{\underline{K}}$ and $\vec{\Phi}_{\underline{k}}$ from configurations of this evolving ensemble. Typical $\vec{\Psi}_{\underline{K}}$ and $\vec{\Phi}_{\underline{k}}$ trajectories are plotted in Figure 2a−d. To track coherence in the variation of OPs, fluctuations (λ) relative to a running average are computed. Details on λ computations are provided in Appendix C and ref 32. Even though the trajectories of lower-order $\vec{\Psi}_{\underline{K}}$ and $\vec{\Phi}_{\underline{k}}$ (i.e., for indices $\underline{K} = \underline{k} = \{000, 100, 010, 001\}$) are similar, the fluctuations for the higher-order $\vec{\Psi}_{\underline{K}}$ (i.e., for indices $\underline{K} = \underline{k} \neq \{000, 100, 010, 001\}$) are about 2 orders of magnitude less than those of $\vec{\Phi}_{\underline{k}}$ (Figure 2e,f). Thus, the higher-order hierarchical OPs are much more coherent than the nonhierarchical ones. In stark contrast, if the VLP is divided into 60 L1 protein monomer subsystems (Figure 1d) where each monomer is described via the same four $\vec{\varphi}^{S}_{\underline{k}}$ as above (i.e., for the L1 pentamer subsystem), the resulting higher-order $\vec{\Psi}_{\underline{K}}$ lose coherence and behave more like the higher-order $\vec{\Phi}_{\underline{k}}$ (Figure 2b,d). Therefore, subsystem size plays a key role in determining the coherence of hierarchical

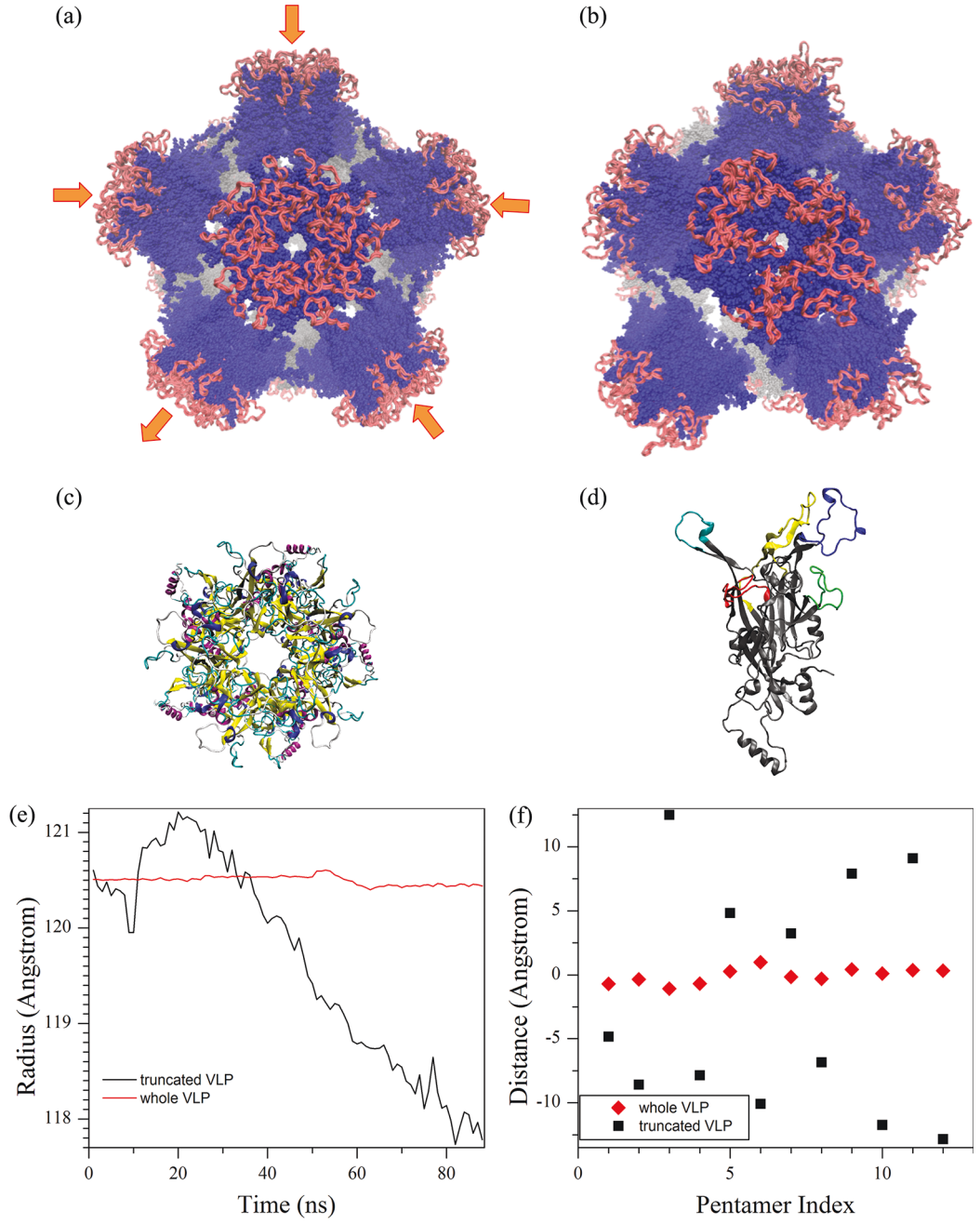

Figure 1. Snapshots at (a) 0 ns and (b) 100 ns illustrating symmetry-breaking deformations of the helix-truncated HPV16 VLP starting from an icosahedral organization of pentamers; arrows indicate inward and outward motions of the pentamer. Structure of subsystems belonging to the VLP: (c) pentamer and (d) L1 protein monomer. Plots of the (e) radius of gyration versus time and (f) displacement of the pentamer CMs after 100 ns from their initial positions showing that the truncated VLP is less stable than the nontruncated one. A positive displacement implies pentamer movement away from the VLP center and vice versa. A majority of the pentamers move inward, yielding VLP collapse.

OPs and hence should be considered when specifying the boundaries defining subsystems. In summary, the $\vec{\Psi}_{\underline{K}}$ constructed for the larger subsystems are more coherent than those from smaller ones. An explanation for this observation and resulting computational advantages is provided in the next subsection.

Let $\Psi_{\underline{K}\alpha}$ be the αth Cartesian component of $\vec{\Psi}_{\underline{K}}$. These OPs have the following interpretation. $\Psi_{\underline{K}X}$ mediates compression−extension in the X direction, whose spatial variation across the system is indicated by the basis function $U_{\underline{K}}^{S}$ and similarly for $\alpha = Y$ and Z.[32] With this, Ψ_{100X}, Ψ_{010Y}, and Ψ_{001Z} track overall compression−extension of the VLP in the three Cartesian directions. In contrast, overall rotation of the VLP is tracked by the set of nine $\Psi_{100\alpha}$, $\Psi_{010\alpha}$, and $\Psi_{001\alpha}$ constituting a 3×3 matrix that is a pure rotation when it has a unit determinant.[32] Therefore, OPs track the overall shape, size, and orientation of the collection of subsystems in the VLP. For example, during disassembly/collapse, the simple compression−extension OPs (Ψ_{100X}, Ψ_{010Y}, and Ψ_{001Z}) initially increase and then either remain constant or decrease (Figure 2a). Consequently, these variables capture the evolution of the VLP radius of gyration (Figure 1e). Furthermore, differences in the evolution profiles

dx.doi.org/10.1021/ct200574x | J. Chem. Theory Comput. 2012, 8, 1379−1392

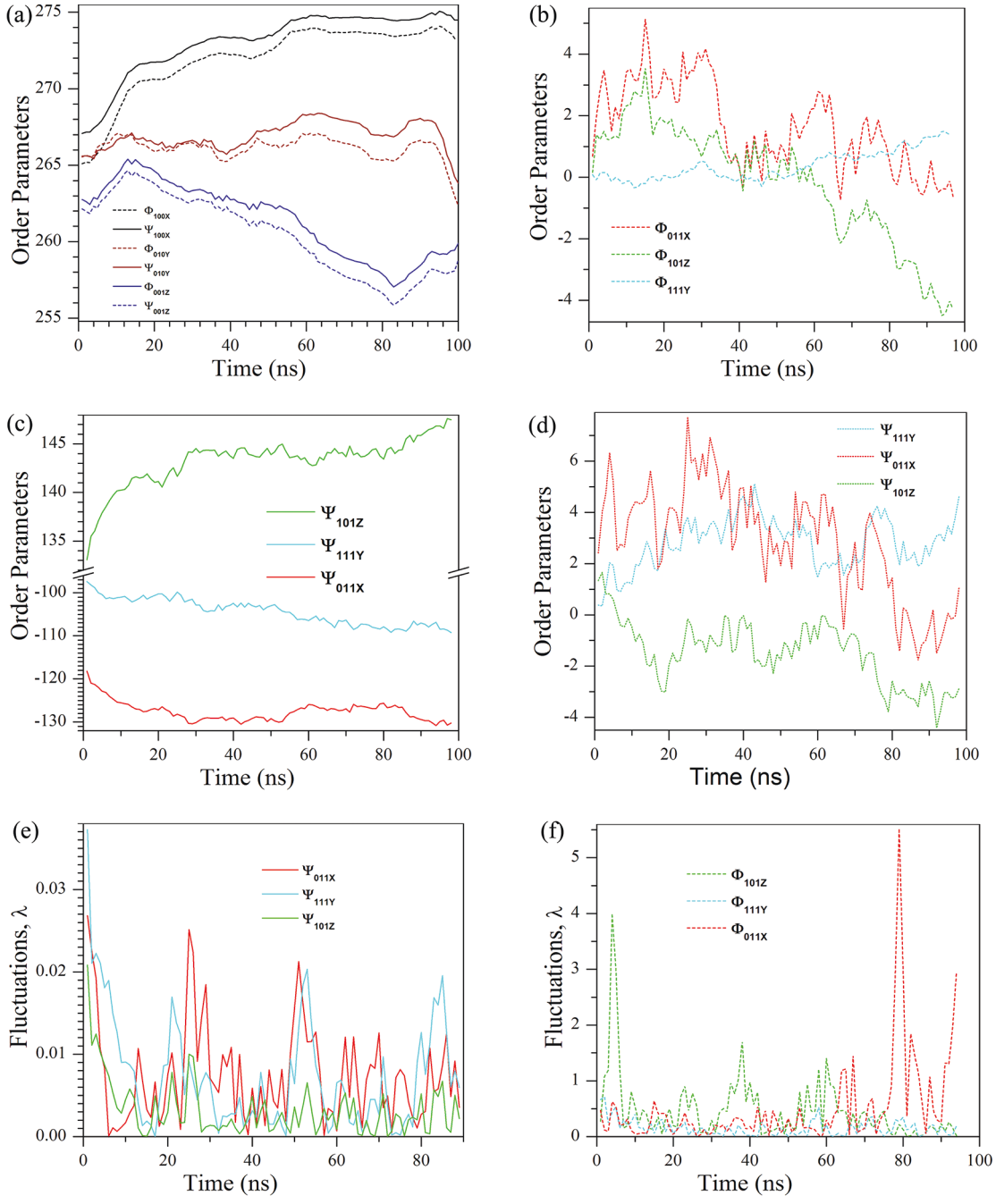

Figure 2. Evolution of (a) typical lower-order $\vec{\Psi}_{\underline{K}}$ and $\vec{\Phi}_{\underline{k}}$, higher-order (b) $\vec{\Phi}_{\underline{k}}$ and (c) $\vec{\Psi}_{\underline{K}}$ with a pentamer as the subsystem, and (d) higher-order $\vec{\Psi}_{\underline{K}}$ with an L1 monomer as the subsystem. (e, f) Fluctuations λ of OPs presented in b and c showing the higher-order $\vec{\Psi}_{\underline{K}}$ are more coherent than $\vec{\Phi}_{\underline{k}}$.

of the three OPs indicate different extents of deformation along the three axes. This is explained as follows.

Closed structures like that of the nontruncated VLP sustain strong radial, lateral, and tangential stress via the interpenetrating helices that compose the attacking arms.[44] On helix truncation, these attacking arms are disrupted, and the pentamers become free to move. With this, the overall stress on the system is released, and most of the pentamers diffuse away from their original location. Consequently, the VLP expands, and hence the OPs increase for the first 10 ns. However, this expansion is not uniform and involves simultaneous translation and rotation of the pentamers. As a result, some pentamers move outward, while most others interact with their neighbors via strong interloop hydrophobic interactions and move inward (Figure 1f; Movie 1, Supporting Information). Thus, the VLP gradually shrinks, indicating collapse. At the level of OPs, Ψ_{100X} probes the expansion along the X direction that results from outward motion of the pentamers, while Ψ_{010Y} and Ψ_{001Z} track contraction along the Y and Z directions due to the interloop hydrophobic interactions between neighboring pentamers. This anisotropy in deformations along the three Cartesian axes illustrates the symmetry-breaking nature of VLP disassembly/

dx.doi.org/10.1021/ct200574x | J. Chem. Theory Comput. 2012, 8, 1379–1392

collapse starting from an icosahedral arrangement. The results also illustrate that VLP instability involves three processes: (1) short time scale atomistic vibrations/collisions underlying free energy driving forces, (2) intermediate time scale interactions between closely lying subsystems, and (3) long time scale subsystem migration. Thus, proper definition of subsystems assists in clarifying the hierarchy of processes across scales in space and time.

Another criterion for defining subsystems is the influence of bonded and nonbonded interactions on the choice of boundaries. If the subsystem boundary traverses a covalent bond, then the ensemble generation procedure via eq B1 would treat constituent atoms in different subsystems independently. This would unphysically stretch the covalent bonds, resulting in the generation of a significant population of very high energy configurations. Generation of such improbable configurations would increase the cost of hybrid sampling required for a convergent Boltzmann averaging of thermal forces.[33] Alternatively, if subsystems natural to an assembly like protomers or pentamers are chosen, the covalent bonds are not stretched by relative motions of the subsystems. In contrast, nonbonded interactions bridging subsystems can still be perturbed, but they do not produce very high-energy configurations. The latter is because (1) such perturbations are achieved by moving atoms in the ensemble with a measure of coherence (Appendix B) and (2) energy changes generated by perturbing nonbonded interactions are smaller than those for bonded ones. Short MD simulations starting from these structures are run to enrich the ensemble, as well as to anneal the nonbonded interactions that are affected by the choice of subsystem boundary. Hence, subsystems are specified such that their boundaries can cross nonbonded interactions if required but never traverse bonded ones. The hybrid ensemble generation procedure together with a proper choice of boundary facilitates sampling modest energy configurations. Thus, the size and boundary criteria are used to guide our choice of subsystems.

A similar partitioning of bonded and nonbonded interactions has been used to obtain a coarse-grained description of virus capsids (as noted in section I). In this representation, covalent interactions are between lumped elements; e.g., peptides are conserved within "trapezoidal tiles" of capsid protein that interact via weaker noncovalent interactions.[24] However, this approach does not address the shortest scale in the hierarchy, i.e., the all-atom state.

B. Rationale for the Behavior of Hierarchical Order Parameters. Now, the evolution characteristics of hierarchical OPs are investigated and compared with those of the nonhierarchical ones. As suggested by eq 11, and more generally by eqs A12 and A13, the rate of change of OPs depends on the associated masses, spatial distribution of the polynomial basis functions that allows for cancellation of atomic momenta, and the evolution characteristics of subsystem-centered OPs. Contributions from these effects are compared to validate the applicability of a fixed reference structure (eqs 12 and 13) and hence explain differences between the $\vec{\Psi}_{\underline{K}}$ and $\vec{\Phi}_{\underline{k}}$ behaviors of Figure 2.

First, the inertial effects are considered. Effective OP masses for hierarchical and nonhierarchical OPs of the L1 VLP are shown in Figure 3. The masses primarily decrease with an increase in the order of polynomial basis functions. Considering systems with spatially uniform mass distribution,[28] this suggests that the higher-order OPs with smaller effective masses probe

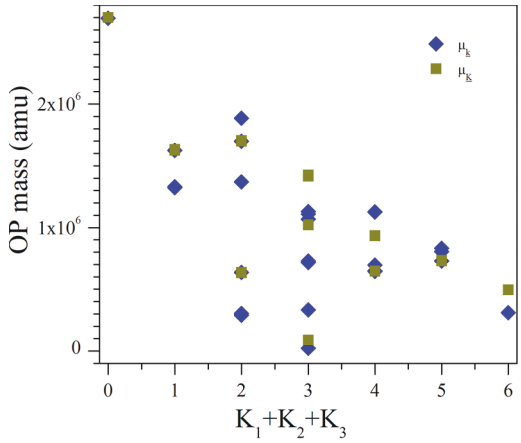

Figure 3. Effective OP masses $\tilde{\mu}_{\underline{K}}$ and $\tilde{\mu}_{\underline{k}}$ for the hierarchical and nonhierarchical OPs versus $K_1 + K_2 + K_3 \equiv k_1 + k_2 + k_3$ show similar ranges of magnitudes, which demonstrates that inertial effects do not account for the difference in coherence between $\vec{\Psi}_{\underline{K}}$ and $\vec{\Phi}_{\underline{k}}$.

relatively smaller regions in space. However, these OP masses vary over similar ranges of magnitude for both the hierarchical and nonhierarchical cases. This implies, for the same order of polynomial, i.e., $\underline{k} = \underline{K}$, that $\vec{\Psi}_{\underline{K}}$ and $\vec{\Phi}_{\underline{k}}$ characterize deformations involving spatial regions of comparable size. Consequently, inertial effects are not responsible for the difference in behavior between the $\vec{\Psi}_{\underline{K}}$ and $\vec{\Phi}_{\underline{k}}$. Therefore, differences in OP behavior can be attributed to the spatial distribution of the basis functions and evolution characteristics of the $\underline{\varphi}^S$.

Now consider the evolution characteristics of the subsystem-centered OPs $\underline{\varphi}^S$. The time evolution of a typical subsystem-centered OP φ^S_{001Z} (Z component of $\vec{\varphi}^S_{001}$) is plotted for 12 pentamer subsystems, i.e., $S = 1, \ldots 12$ (Figure 4a). The fluctuation λ of φ^S_{001Z} is plotted in Figure 4c; we present λ only for the pentamer having the maximum fluctuation. Results suggest that the lower-order $\vec{\varphi}^S_{\underline{k}}$ (e.g., $\vec{\varphi}^S_{\underline{k}}$ with $\underline{k} = \{000, 100, 010, 001\}$, which track pentamer position and orientation) have significantly less fluctuation than do the higher-order nonhierarchical OPs defined for the entire VLP, e.g., Φ_{101Z} and Φ_{011X} (Figure 2b–f). However, with the choice of the L1 monomer as a subsystem, the φ^S_{001Z} (and other members of $\underline{\varphi}^S$ not shown here) are more rapidly fluctuating than when a pentamer is the subsystem (Figure 4b,c).

The $\vec{\Psi}_{\underline{K}}$ are linear combinations of $\underline{\varphi}^S$ (eq 9), and the atomic configurations consistent with fixed values of $\vec{\Psi}_{\underline{K}}$ belong to an equilibrium ensemble (eq 13). This suggests that the rate of $\vec{\Psi}_{\underline{K}}$ evolution is either slower than or comparable to that of $\vec{\varphi}^S_{\underline{k}}$. As a result, the rate of $\underline{\varphi}^S$ evolution provides an upper bound to that of $\underline{\Psi}$. Figure 4 indicates that larger subsystems enable construction of slower $\underline{\varphi}^S$. Consequently, they yield more coherent $\vec{\Psi}_{\underline{K}}$. Therefore, given a set of subsystems, coherence of the associated lower-order $\underline{\varphi}^S$ filters out the higher frequency fluctuations from $\vec{\Psi}_{\underline{K}}$ dynamics. Physically, this signifies that overall assembly behavior is slower than intrasubsystem processes. Through the mathematical construction of our OPs, this property of hierarchically organizing systems is

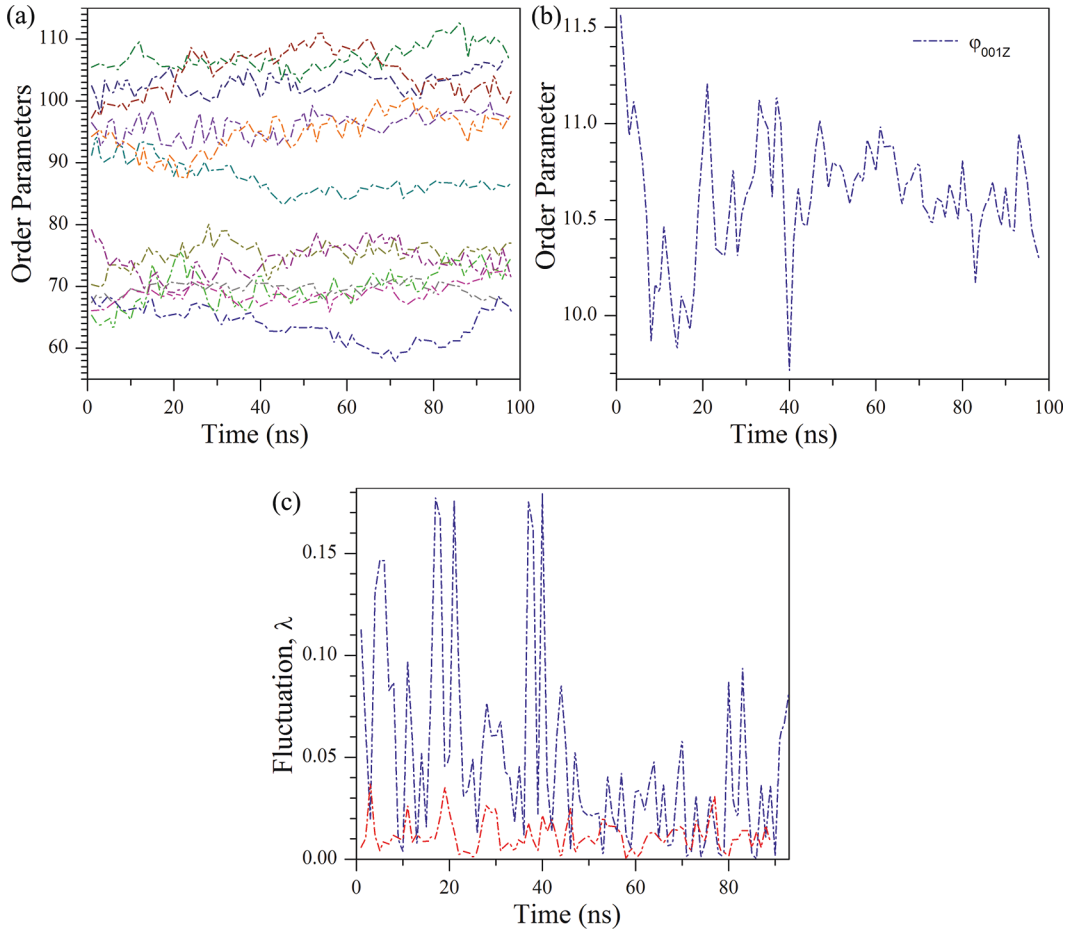

Figure 4. Evolution of φ^{S}_{001Z} for the choice of the (a) pentamer and (b) L1 protein as a subsystem and (c) associated fluctuation λ (pentamer, red; monomer, blue), illustrating that smaller subsystems have less coherent $\vec{\varphi}^{S}_{\underline{k}}$.

naturally captured. In contrast, when $\underline{k}$ is the same as $\underline{K}$, the $\vec{\Phi}_{\underline{k}}$, e.g., Φ_{101Z} and Φ_{011X}, are constructed directly from the all-atom configuration and hence express more high frequency fluctuations than $\vec{\Psi}_{\underline{K}}$. This, in part, is the reason for obtaining a slower and more coherent $\vec{\Psi}_{\underline{K}}$ relative to $\vec{\Phi}_{\underline{k}}$ at $\underline{K} = \underline{k} \neq$ {000, 100, 010, 001}.

The subsystem-centered OP φ^{S}_{001Z}, by construction, probes deformations along the Z Cartesian axis for Sth subsystem.[32] Truncation of the helices h2 and h3 disrupts inter-L1 monomer interactions, triggering local deformations within the intra- and interpentameric organization. Observed differences in the evolution characteristics of φ^{S}_{001Z} (Figure 4a) for the 12 different pentamers elucidate the anisotropic nature of these deformations. While VLP shape and size follow simple trends (Figure 1e), subsystem-centered OPs reflect the anisotropy of local deformations that ultimately manifests in the symmetry-breaking nature of the VLP disassembly. Therefore, accounting for intermediate-scale deformations is necessary to probe the local events that underlie VLP wide processes, i.e., symmetry breaking here. For large $\underline{K}$, the $\vec{\Psi}_{\underline{K}}$ evolution probes the dynamics of spatial regions with size similar to those of the constituting subsystems (Figure 3). Thus, such intermediate-scale deformations are captured. In the companion article,[37] this feature of the higher-order hierarchical OPs is demonstrated for the specific case of a deforming pentamer. Next, the evolution of polynomial basis functions is considered.

Terms omitted from eq 11 for obtaining 12 express the dependence of OP velocity on the rate of change of the basis-set polynomials A1. To investigate the effect of such changes on $\underline{\Psi}$ dynamics, first define the quantity $\Theta_{\underline{K}}(t)$ via

$$\Theta_{\underline{K}}(t) = \frac{\sum_{S=1}^{N^{sys}} M^{S}\, U^{S}_{\underline{K}}(t)\, U^{S}_{\underline{K}}(0)}{\sum_{S=1}^{N^{sys}} M^{S}\{U^{S}_{\underline{K}}(0)\}^{2}} \tag{14}$$

where $U_{\underline{K}}{}^{S}(t) \equiv U_{\underline{K}}[\vec{R}^{S}(t)]$, $\vec{R}^{S}(t)$ is the position of the S-th subsystem CM at time t, and $U_{\underline{K}}{}^{S}(0) \equiv U_{\underline{K}}[\vec{R}^{S}(0)]$. The polynomials $U_{\underline{K}}{}^{S}(t)$ are computed at each Langevin time step from an evolving set of subsystem CMs as generated here using ***SimNanoWorld***. For the nonhierarchical OPs $\vec{\Phi}_{\underline{k}}$ an expression analogous to eq 14, $\Theta_{\underline{k}}(t)$, is used with M^{S} replaced by m_i and $U_{\underline{K}}{}^{S}$ replaced by $U_{\underline{k}i}$ for $i = 1, \ldots N$. This implies

$$\Theta_{\underline{k}}(t) = \frac{\sum_{i=1}^{N} m_i U_{\underline{k}i}(t)\, U_{\underline{k}i}(0)}{\sum_{i=1}^{N} m_i\{U_{\underline{k}i}(0)\}^{2}} \tag{15}$$

The basis-set polynomials associated with nonhierarchical and hierarchical OPs are organized into one-dimensional matrices (i.e., vectors) of dimension $(N \times 1)$ and $(N^{sys} \times 1)$, respectively (section SI3 in the Supporting Information). Here,

dx.doi.org/10.1021/ct200574x | J. Chem. Theory Comput. 2012, 8, 1379–1392

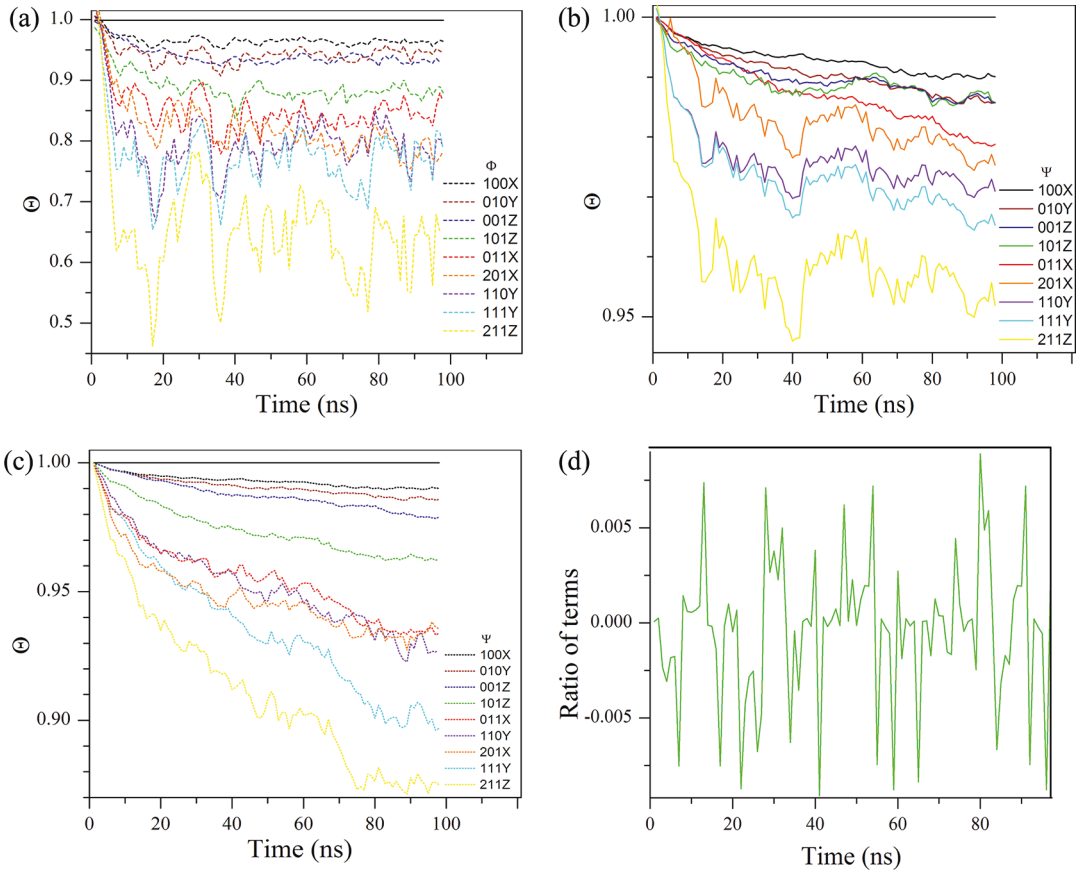

Figure 5. Evolution of $\Theta(t)$, the cosine of the angle of rotation from the original basis vector orientation to that at time t, for typical (a) $\vec{\Phi}_{\underline{k}}$ and $\vec{\Psi}_{\underline{K}}$ with the (b) L1 pentamer and (c) monomer as a subsystem. This shows that the hierarchical basis vectors rotate slower than the nonhierarchical ones. (d) Ratio of the last two terms of eq 10 to the first term, showing that the rate of change of hierarchical OP basis-set polynomials is negligible relative to that of the OPs themselves.

N^{sys} = 12 or 60 for the choice of 12 pentamers or 60 L1 monomers as subsystems. For the nonhierarchical OPs, $N = 4 \times 10^5$, as the VLP contains that many atoms. With this, $\Theta(t)$ tracks the cosine of rotation angles of these vectors from their initial orientation. $\Theta_{\underline{k}}$ changes significantly with time for the $\vec{\Phi}_{\underline{k}}$ with higher $\underline{k}$ (i.e., $\underline{k} \neq \{000, 100, 010, 001\}$; Figure 5a). However, for the hierarchical case with pentamers as subsystems, this change diminishes, implying that the $U_{\underline{K}}{}^S$ are virtually insensitive to the evolving configuration of CMs (Figure 5b).

Higher-order basis-set polynomials express significant spatial variation and thus tend to probe local motions. However, for the hierarchical formalism, these are expressed as a function of the subsystem CMs $\underline{R}^S$ (eq 10) or OPs φ^S (eq 9), which by definition are mass-weighted averages of atomic positions (eq 6). Since the arguments $\underline{R}^S$ of $U_{\underline{K}}$ vary slowly in time, they imply negligible rotations of the associated N^{sys}-dimensional basis vectors (Figure 5b). For the choice of a smaller subsystem, i.e., an L1 monomer versus a pentamer, these arguments express more high-frequency local motions. This is reflected in the greater variations of corresponding $\Theta_{\underline{K}}$ (Figure 5c).

Significant changes in Θ for the L1 monomer defined $\vec{\Psi}_{\underline{K}}$ and higher-order $\vec{\Phi}_{\underline{k}}$ suggest that the time dependence of the basis-set polynomials is not negligible. This provides a sufficient criterion for indicating the incompleteness of eq 12, which neglects the rate of change of the basis-set polynomials. In contrast, negligible changes in $\Theta_{\underline{K}}$ for the pentamer subsystem defined as $\vec{\Psi}_{\underline{K}}$ do not guarantee the correctness of eq 12. This is because $\Theta_{\underline{K}}$ tracks the time dependence of only one component of the N^{sys}-dimensional vector of basis-set polynomials. Evolution characteristics of the other $(N^{sys} - 1)$ components should be taken into account for validating eq 12 and hence the $O(\varepsilon)$ scaling of our OPs (eq 13).

Evaluating $\partial\underline{\Psi}/\partial t$ as $-L\underline{\Psi}$, the rate of change of $U_{\underline{K}}{}^S$ enters through spatial derivatives of these polynomials (Appendix A). In Figure 5d, the time evolution of the ensemble-averaged ratio of the last two terms in eq 11 to the first term, $\langle \sum_S \vec{P}^S \cdot \{ \vec{R}^S \cdot (\partial U_{\underline{K}}{}^S / \partial \vec{R}^S) - 2 U_{\underline{K}}{}^S (\vec{\Psi}_{\underline{K}} \cdot (\partial U_{\underline{K}}{}^S / \partial \vec{R}^S)) \} / \sum_S \vec{P}^S U_{\underline{K}}{}^S \rangle$ is plotted. Results suggest that contributions from the omitted terms in eq 12 are several orders of magnitude lower than the considered one. Numerically, these are as small as the fluctuations in the hierarchical OPs (Figure 2e) and are hence negligible. Thus, eqs 12 and 13 are restored, and the slow temporal evolution of the hierarchical OP basis-set polynomials relative to the nonhierarchical ones is confirmed. This suggests that a fixed reference configuration of subsystem CMs for $\vec{\Psi}_{\underline{K}}$ construction holds for long periods of time.

In summary, the higher-order $\vec{\Psi}_{\underline{K}}$ are slower than $\vec{\Phi}_{\underline{k}}$ for a given $\underline{K} = \underline{k}$. This is because the scaled variables $\underline{\varphi}^S$ (e.g., subsystem CMs) vary more slowly in time than do the atomic

dx.doi.org/10.1021/ct200574x | J. Chem. Theory Comput. 2012, 8, 1379−1392

coordinates. The choice of subsystem size determines the dynamical properties of $\underline{\varphi}^S$. This in conjunction with the slow variation of the basis-set polynomials used to define $U_{\underline{K}}^S$ makes the $\vec{\Psi}_{\underline{K}}$ extremely coherent. For the demonstration system, choosing the subsystem to be a pentamer implies slowly varying $\underline{\varphi}^S$ and hence even slower $\underline{\Psi}$. In contrast, if the L1 monomer is the subsystem, the $\underline{\varphi}^S$ are less coherent and so are the resulting $\vec{\Psi}_{\underline{K}}$. This rationalizes the results presented in Figure 2.

The hierarchical OP formulation provides a "double filter" of noise. Choosing $\underline{\varphi}^S$ (or $\underline{R}^S$) at the subsystem level itself filters some local fluctuations. In turn, the noise in $\underline{\varphi}^S$ is refiltered over the entire system via $U_{\underline{K}}^S$ for $S = 1, ..., N^{sys}$ to yield more coherent $\underline{\Psi}$ dynamics. However, the lower-order OPs exhibit similar evolution for both hierarchical and nonhierarchical formulations (Figure 2a). This is explained in section SI4 of the Supporting Information.

IV. DISCUSSION

Results presented imply that hierarchical OPs ($\vec{\Psi}_{\underline{K}}$) express much less fluctuation than nonhierarchical ones ($\vec{\Phi}_{\underline{k}}$) (Figure 2), even if they capture deformations involving spatial regions of similar size (Figure 3). This follows from the enhanced filtering of fluctuations embedded in the construction of the hierarchical OPs. The number and character of the hierarchical OPs depend explicitly on the definition of subsystems. For example, choosing L1 monomers as subsystems for the $T = 1$ HPV assembly results in $4 \times 60 = 240$ subsystem-centered OPs. Thus, the number of global OPs is limited to 240, i.e., those for which $K_1 + K_2 + K_3 \leq 11$. However, the OPs in this formulation are about as coherent as the nonhierarchical ones. In contrast, if the subsystem is taken to be an L1 pentamer, the number of subsystem OPs decreases to $4 \times 12 = 48$. With this, the number of global OPs decreases to those with $K_1 + K_2 + K_3 \leq 6$. As shown above, these hierarchical OPs are more coherent than the nonhierarchical ones. The associated basis functions change slowly in time, thereby enabling the application of a fixed reference configuration over long time periods (Figure 5). Furthermore, the $\vec{\Psi}_{\underline{K}}$ enable reconstruction of all-atom structures consistent with global and subsystem assembly architecture. This is achieved through the generation of quasi-equilibrium ensembles that account for small-scale displacement of each atom in addition to the coherent deformation generated by the OPs (Appendix B). Thus, the hierarchical OPs yield conceptual and computational advantages for multiscaling macromolecular systems. These are summarized as follows:

- The time scale of $\vec{\Psi}_{\underline{K}}$ evolution is well separated from that of atomic fluctuations (eq 13 and Figure 2a and c). With this, the set of $\vec{\Psi}_{\underline{K}}$ enables one to uncloak multiple spatial and temporal dependencies of the N-atom probability density and thereby serve as the basis of a multiscale analysis. As discussed in a companion study,[37] this methodology starts with the N-atom Liouville equation and yields Langevin equations for $\vec{\Psi}_{\underline{K}}$ evolution, while mathematical reformulation of the underlying molecular physics simultaneously captures high frequency atomic fluctuations.
- A judicious choice of subsystems enables the hierarchical OP construction scheme to generate a greater number of slow variables than for the nonhierarchical approach. Therefore, one can perform a multiscale simulation which still provides all-atom detail as with the nonhierarchical OPs but now with a richer set of slow descriptive variables. Accounting for the evolution of a greater number of OPs through associated Langevin equations facilitates the capturing of multiple pathways to structural transition.[30]
- The effective OP masses, $\tilde{\mu}_{\underline{K}}$ (eq 9 or 10), decrease for larger values of $\underline{K}$. This suggests that OPs with a higher $\underline{K}$ probe deformation of smaller regions in space. Thus, a model based on the hierarchical OPs captures a spatially diverse range of coherent deformations through the $\vec{\Psi}_{\underline{K}}$ for a set of $\underline{K}$.
- Enhanced coherence of the hierarchical OPs relative to the nonhierarchical ones enables the application of larger Langevin timesteps for numerically simulating the $\underline{\Psi}$ evolution. Furthermore, the hierarchical strategy tracks deformation with respect to a fixed reference structure. This avoids the need for frequent rereferencing and recalculation of the basis functions while carrying out a long-time simulation. Thus, interplay between the characteristic time of OP evolution and the frequency of rereferencing reflects in the overall simulation efficiency.
- The $\underline{\Psi}$ facilitates generation of an ensemble of all-atom configurations. This captures the exchange of information across multiple scales in space and time and provides criteria for judging the completeness of an OP description (as shown explicitly in the companion study[37]). For example, the appearance of long time tails in the OP velocity autocorrelation functions or a systematic growth in the size of residuals in these ensembles indicates the necessity of adding more OPs to the coarse-grained description. Consequently, one can initialize multiscale simulations with a finite number of OPs, e.g., 48 or less when pentamers are chosen as subsystems, but change the number on the fly to account for emergent motions.

These properties of the hierarchical OPs make them ideal variables for developing a multiscale computational approach that captures slow processes in macromolecular assemblies. In the second part of this work, we formulate a $\underline{\Psi}$-mediated multiscale framework and provide further numerical validation of the computational gain obtained by exploiting this algorithm.

V. CONCLUSION

This study is a part of a broader quest to discover OPs for classical and quantum nanosystems.[13,30,35,36,45] A subset of OPs presented in this paper (e.g., ones with $K_1 + K_2 + K_3 \leq 6$) is shown to evolve coherently and operate on time scales that are several orders of magnitude greater than those of atomistic fluctuations. These OPs capture the slowly varying modes of a macromolecular assembly. The atomic scale fluctuations are accounted for via the quasi-equilibrium ensemble of all-atom configurations generated using OP-like variables (B1) and enriched via short MD simulations. With this, the overall as well as internal dynamics of macromolecules or their assemblies is captured. Furthermore, hierarchical OPs enable the use of a fixed reference structure for defining the OPs, an assumption that is violated by the higher-order $\underline{\Phi}$. In summary, we demonstrate the conceptual and technical advantages of the hierarchical OPs over our nonhierarchical ones for multiscaling

macromolecular assemblies. Putting the metaphysical question of the exact "meaning" of an OP aside, even though as shown above some physical meaning can usually be imparted, we suggest that the purpose of our OPs is to serve as the basis of a multiscale formalism designed to facilitate the all-atom simulations of multimillion atom systems. In the follow-up paper,[37] such an algorithm based on the Langevin evolution of hierarchical OPs and associated ensemble of atomic structures is formulated and demonstrated.

APPENDIX A

Hierarchical Order Parameter Kinetics

Here, we determine the rate of change of hierarchical OPs via the Liouville equation. Using eq 10, the rate of change of $\vec{\Psi}_{\underline{K}}$ is given by

$$\frac{\partial \vec{\Psi}_{\underline{K}}}{\partial t} = \frac{1}{\tilde{\mu}_{\underline{K}}}\left(\sum_S M^S U^S_{\underline{K}} \frac{\partial \vec{R}^S}{\partial t} + \sum_S M^S \frac{\partial U^S_{\underline{K}}}{\partial t}\vec{R}^S - \vec{\Psi}_{\underline{K}}\frac{\partial \tilde{\mu}_{\underline{K}}}{\partial t}\right) \tag{A1}$$

Equation A1 and the Liouville equation yields

$$\frac{\partial \vec{\Psi}_{\underline{K}}}{\partial t} = -\frac{1}{\tilde{\mu}_{\underline{K}}}\{\sum_S M^S U^S_{\underline{K}}(L\vec{R}^S) + \sum_S M^S (LU^S_{\underline{K}})\vec{R}^S - \vec{\Psi}_{\underline{K}}(L\tilde{\mu}_{\underline{K}})\} \tag{A2}$$

We proceed via analysis of the three terms appearing on the RHS of eq A2. The first term includes

$$L\vec{R}^S = -\sum_{i=1}^{N} \frac{\vec{p}_i}{m_i}\cdot\frac{\partial}{\partial \vec{r}_i}\vec{R}^S \tag{A3}$$

With eq A3, the definition of subsystem CM $\vec{R}^S = \sum_{j=1}^{n^S} m_j^S \vec{r}_j^S / M^S$ and $(\partial/\partial r_{i\alpha}) r_{i\alpha'} = \delta_{ij}\delta_{\alpha\alpha'}$

$$\begin{aligned} L\vec{R}^S &= -\frac{\sum_{\alpha=1}^{3}\hat{e}_\alpha \sum_{j=1}^{n^S}\frac{\vec{p}_{j\alpha}^S}{m_i} m_j^S \delta_{ij}}{M^S} \\ &= -\frac{\sum_{j=1}^{n^S}\vec{p}_j^S}{M^S} \\ &= -\frac{\vec{P}^S}{M^S} \end{aligned} \tag{A4}$$

where $\hat{e}_\alpha$ is the unit vector along the α Cartesian direction.

The second term includes

$$LU^S_{\underline{K}} = -\sum_{i=1}^{N}\frac{\vec{p}_i}{m_i}\cdot\frac{\partial}{\partial \vec{r}_i}U^S_{\underline{K}} \tag{A5}$$

Furthermore,

$$\begin{aligned} \frac{\partial}{\partial \vec{r}_i}U^S_{\underline{K}} &= \sum_{\alpha'=1}^{3}\sum_{\alpha=1}^{3}\delta_{\alpha\alpha'}\frac{\partial R^S_{\alpha'}}{\partial r_{i\alpha}}\frac{\partial U^S_{\underline{K}}}{\partial R^S_{\alpha'}}\hat{e}_{\alpha'} \\ &= \sum_{\alpha=1}^{3}\frac{\partial R^S_\alpha}{\partial r_{i\alpha}}\frac{\partial U^S_{\underline{K}}}{\partial R^S_\alpha}\hat{e}_\alpha \end{aligned} \tag{A6}$$

With eq A6, $\vec{R}^S = \sum_{j=1}^{n^S} m_j^S \vec{r}_j^S / M^S$, and $(\partial/\partial r_{i\alpha})\partial r_{i\alpha'} = \partial_{ij}\partial r_{\alpha\alpha'}$

$$\frac{\partial}{\partial \vec{r}_i}U^S_{\underline{K}} = \delta_{ij}\frac{m_j^S}{M^S}\frac{\partial U^S_{\underline{K}}}{\partial \vec{R}^S} \tag{A7}$$

Putting eq A7 into eq A5 yields

$$LU^S_{\underline{K}} = -\frac{\sum_{j=1}^{n^S}\vec{p}_j^S}{M^S}\cdot\frac{\partial U^S_{\underline{K}}}{\partial \vec{R}^S} = -\frac{\vec{P}^S}{M^S}\cdot\frac{\partial U^S_{\underline{K}}}{\partial \vec{R}^S} \tag{A8}$$

Finally, the third term includes

$$L\tilde{\mu}_{\underline{K}} = -\sum_{i=1}^{N}\frac{\vec{p}_i}{m_i}\cdot\frac{\partial}{\partial \vec{r}_i}\sum_S M^S (U^S_{\underline{K}})^2 \tag{A9}$$

Using eq A7 in eq A9 yields

$$L\tilde{\mu}_{\underline{K}} = -2\sum_S U^S_{\underline{K}}\vec{P}^S\cdot\frac{\partial U^S_{\underline{K}}}{\partial \vec{R}^S} \tag{A10}$$

Putting eqs A4, A8, and A10 into eq A2,

$$\begin{aligned} \frac{\partial \vec{\Psi}_{\underline{K}}}{\partial t} &= \frac{1}{\tilde{\mu}_{\underline{K}}}\left\{\sum_S U^S_{\underline{K}}\vec{P}^S + \sum_S\left(\vec{R}^S\cdot\frac{\partial U^S_{\underline{K}}}{\partial \vec{R}^S}\right)\vec{P}^S - 2\vec{\Psi}_{\underline{K}}\sum_S U^S_{\underline{K}}\left(\vec{P}^S\cdot\frac{\partial U^S_{\underline{K}}}{\partial \vec{R}^S}\right)\right\} \\ &= \frac{1}{\tilde{\mu}_{\underline{K}}}\sum_S \vec{P}^S\left\{U^S_{\underline{K}} + \vec{R}^S\cdot\frac{\partial U^S_{\underline{K}}}{\partial \vec{R}^S} - 2U^S_{\underline{K}}\left(\vec{\Psi}_{\underline{K}}\cdot\frac{\partial U^S_{\underline{K}}}{\partial \vec{R}^S}\right)\right\} \end{aligned} \tag{A11}$$

Thus, eq 11 is obtained.

Defining $\vec{\Psi}_{\underline{K}}$ via eq 9 instead of eq 10 modifies eq A11 to

$$\frac{\partial \vec{\Psi}_{\underline{K}}}{\partial t} = \frac{1}{\tilde{\mu}_{\underline{K}}}\sum_{\underline{k},S}\vec{\Pi}^S_{\underline{k}}\left\{U^S_{\underline{K}\underline{k}} + \vec{\varphi}^S_{\underline{k}}\cdot\frac{\partial U^S_{\underline{K}\underline{k}}}{\partial \vec{\varphi}^S_{\underline{k}}} - 2U^S_{\underline{K}\underline{k}}\left(\vec{\Psi}_{\underline{K}}\cdot\frac{\partial U^S_{\underline{K}\underline{k}}}{\partial \vec{\varphi}^S_{\underline{k}}}\right)\right\} \tag{A12}$$

where $\vec{\Pi}^S_{\underline{k}} = \sum_{j=1}^{n^S} U^S_{\underline{k}j}\vec{p}_j^S$.

dx.doi.org/10.1021/ct200574x | J. Chem. Theory Comput. 2012, 8, 1379−1392

Repeating steps similar to A1 through A11 for $\vec{\Phi}_{\underline{k}}$ yields

$$\frac{\partial \vec{\Phi}_{\underline{k}}}{\partial t} = \frac{1}{\tilde{\mu}_{\underline{k}}} \sum_{i=1}^{N} \vec{p}_i \left\{ U_{\underline{k}i} + \vec{r}_i \cdot \frac{\partial U_{\underline{k}i}}{\partial \vec{r}_i} - 2U_{\underline{k}i} \left(\vec{\Phi}_{\underline{k}} \cdot \frac{\partial U_{\underline{k}i}}{\partial \vec{r}_i} \right) \right\} \tag{A13}$$

APPENDIX B

All-Atom Reconstruction from Hierarchical Order Parameters

Consider an extended set $\underline{\varphi}^S_{ex}$ of subsystem-centered variables that include the $\vec{\varphi}^S_{\underline{k}}$ for $\underline{k}$ in the list of OPs that compose $\underline{\Psi}$ (via eq 9 or 10), plus additional variables $\vec{\varphi}^S_{\underline{k}\mathrm{res}}$ for $\underline{k}$res not in the list of OP indices. Thus, we write the position of atom j in subsystem S as

$$\vec{r}^S_j = \sum_{\underline{k}}^{\mathrm{OP}} \vec{\varphi}^S_{\underline{k}} U^S_{\underline{k}j} + \sum_{\underline{k}res}^{\mathrm{res}} \vec{\varphi}^S_{\underline{k}\mathrm{res}} U^S_{\underline{k}j,\mathrm{res}} \tag{B1}$$

where $U^S_{\underline{k}j,\mathrm{res}} = U_{\underline{k}\mathrm{res}}(\vec{r}^{S,0}_j)$. This equation maps $\underline{\varphi}^S_{ex}$ onto the all-atom configuration variables $\underline{r}^S$. The mapping is 1:1 when the total number of $\vec{\varphi}^S_{\underline{k}}$ and $\vec{\varphi}^S_{\underline{k}\mathrm{res}}$ equals the number of atoms n^S in the subsystem S. An expression for the residuals $\vec{\sigma}^S_j$ in terms of the $\vec{\varphi}^S_{\underline{k}}$ is obtained by comparing eq 5 and eq B1, i.e., $\vec{\sigma}^S_j = \sum^{\mathrm{res}}_{\underline{k}\mathrm{res}} \vec{\varphi}^S_{\underline{k}\mathrm{res}} U^S_{\underline{k}j,\mathrm{res}}$S. With this, eq B1 provides a way to generate an ensemble of atomic configurations consistent with a given value of $\vec{\varphi}^S_{\underline{k}}$ (and hence $\vec{\Psi}_{\underline{K}}$). For a set of fixed $\vec{\Psi}_{\underline{K}}$, these configurations are achieved by accounting for the residuals $\vec{\sigma}^S_j$ via randomly varying $\vec{\varphi}^S_{\underline{k}\mathrm{res}}$. However, this procedure typically leads to many very high energy, low Boltzmann probability configurations. The practical difficulty is readily avoided as long as $\vec{\sigma}^S_j$ is chosen by constraining the magnitude of $\vec{\varphi}^S_{\underline{k}\mathrm{res}}$ to small values for higher $\underline{k}$res.[32,33] This procedure provides major structural variations by moving atoms in the ensemble with a measure of coherence, avoiding near atom overlap, and is simultaneously performed for all of the subsystems. Thus, rich ensembles at fixed $\underline{\Psi}$ and with modest energies (and hence Boltzmann relevance) are generated. In practice, short MD runs are performed starting with configurations from the $\vec{\varphi}^S_{\underline{k}\mathrm{res}}$-generated ensemble to enrich fluctuations about the constant set of OPs $\underline{\Psi}$. This is self-consistent when the MD runs are much shorter than the characteristic time of OP evolution. This procedure for generating ensembles is henceforth termed hybrid sampling.[33]

APPENDIX C

λ Computation

Fluctuations λ in OPs are computed via

$$\lambda(f, t_i) = \frac{\sqrt{\sum_{i=0}^{N_f-1} [f(t_i) - \langle f \rangle]^2 / N_f}}{f_{\max}(t_i)}; f \equiv \Psi_{\underline{K}\alpha}, \Phi_{\underline{k}\alpha} \tag{C1}$$

where t_i is the ith Langevin timestep, N_f is the total number of time frames used for the moving averages, $\langle f \rangle$ is the moving average (over 50 frames), and $f_{\max}$ is the maximum absolute value of the OP within the range of OPs sampled for the moving average calculation. Fluctuations are defined about a moving absolute maxima rather than a moving average to avoid singularities from zero moving averages. The normalization makes λ dimensionless. Thus, λ can be compared between different hierarchical and nonhierarchical OPs.

ASSOCIATED CONTENT

Supporting Information

(a) Movie depicting the inward motion of pentamers facilitated by strong hydrophobic interactions between neighboring loops that leads to the HPV VLP collapse. (b) Discussion on the Langevin evolution of $\vec{\Phi}_{\underline{k}}$, numerical construction of subsystem OPs and basis functions, and similarity in the behavior of lower-order hierarchical and nonhierarchical OPs. This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ortoleva@indiana.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project was supported in part by the National Science Foundation (Collaborative Research in Chemistry program), National Institute of Health (NIBIB), Department Of Energy (office of basic science), METAcyt, and the Indiana University College of Arts and Sciences through the Center for Cell and Virus Theory.

REFERENCES

(1) Wegst, U. G.; Ashby, M. F. The mechanical efficiency of natural materials. *Philos. Mag.* **2004**, *84* (21), 2167.

(2) Meyers, M. A.; Chen, P.-Y.; Lin, A. Y.-M.; Seki, Y. Biological materials: Structure and mechanical properties. *Prog. Mater. Sci.* **2008**, *53* (1), 1.

(3) Ingber, D. E. The origin of cellular life. *BioEssays* **2000**, *22* (12), 1160.

(4) Ortoleva, P. J. Nanoparticle dynamics: A multiscale analysis of the Liouville equation. *J. Phys. Chem. B* **2005**, *109* (45), 21258.

(5) Miao, Y.; Ortoleva, P. J. Viral structural transitions: an all-atom multiscale theory. *J. Chem. Phys.* **2006**, *125* (21), 214901.

(6) Miao, Y.; Ortoleva, P. J. All-atom multiscaling and new ensembles for dynamical nanoparticles. *J. Chem. Phys.* **2006**, *125* (4), 44901.

(7) Whitelam, S.; Feng, E. H.; Hagan, M. F.; Geissler, P. L. The role of collective motion in examples of coarsening and self-assembly. *Soft Matter* **2009**, *5* (6), 1251.

(8) Uvarov, A.; Fritzsche, S. Friction of N-bead macromolecules in solution: Effects of the bead-solvent interaction. *Phys. Rev. E* **2006**, *73* (1), 011111.

(9) Uvarov, A.; Fritzsche, S. Effects of the bead-solvent interaction on the dynamics of macromolecules, 1 - The dumbbell molecule. *Macromol. Theor. Simul.* **2004**, *13* (3), 241.
(10) Arkhipov, A.; Freddolino, P. L.; Schulten, K. Stability and Dynamics of Virus Capsids Described by Coarse-Grained Modeling. *Structure* **2006**, *14* (12), 1767.
(11) Gohlke, H.; Thorpe, M. F. A Natural Coarse Graining for Simulating Large Biomolecular Motion. *Biophys. J.* **2006**, *91* (6), 2115.
(12) Backofen, R.; Maher, M.; Puget, J.-F. In *Constraint Techniques for Solving the Protein Structure Prediction Problem, Principles and Practice of Constraint Programming — CP98*; Maher, M., Puget, J.-F., Eds.; Springer: Berlin/Heidelberg: 1998; p 72.
(13) Shreif, Z.; Ortoleva, P. Curvilinear All-Atom Multiscale (CAM) Theory of Macromolecular Dynamics. *J. Stat. Phys.* **2008**, *130* (4), 669.
(14) Hayward, S.; Kitao, A.; Gō, N. Harmonicity and anharmonicity in protein dynamics: A normal mode analysis and principal component analysis. *Proteins: Struct., Funct., Bioinf.* **1995**, *23* (2), 177.
(15) Hayward, S.; Kitao, A.; Berendsen, H. J. C. Model-free methods of analyzing domain motions in proteins from simulation: A comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins: Struct., Funct., Bioinf.* **1997**, *27* (3), 425.
(16) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins: Struct., Funct., Bioinf.* **1993**, *17* (4), 412.
(17) Sweet, C. R.; Petrone, P.; Pande, V. S.; Izaguirre, J. A. Normal mode partitioning of Langevin dynamics for biomolecules. *J. Chem. Phys.* **2008**, *128* (14), 145101.
(18) Zhou, J.; Thorpe, I. F.; Izvekov, S.; Voth, G. A. Coarse-Grained Peptide Modeling Using a Systematic Multiscale Approach. *Biophys. J.* **2007**, *92* (12), 4289.
(19) Chang, C.-E.; Shen, T.; Trylska, J.; Tozzini, V.; McCammon, J. A. Gated Binding of Ligands to HIV-1 Protease: Brownian Dynamics Simulations in a Coarse-Grained Model. *Biophys. J.* **2006**, *90* (11), 3880.
(20) Ayton, G. S.; Voth, G. A. Multiscale simulation of transmembrane proteins. *J. Struct. Biol.* **2007**, *157* (3), 570.
(21) West, B.; Brown, F. L. H.; Schmid, F. Membrane-Protein Interactions in a Generic Coarse-Grained Model for Lipid Bilayers. *Biophys. J.* **2009**, *96* (1), 101.
(22) Riccardi, L.; Nguyen, P. H.; Stock, G. Free-Energy Landscape of RNA Hairpins Constructed via Dihedral Angle Principal Component Analysis. *J. Phys. Chem. B* **2009**, *113* (52), 16660.
(23) Pasquali, S.; Derreumaux, P. HiRE-RNA: A High Resolution Coarse-Grained Energy Model for RNA. *J. Phys. Chem. B* **2010**, *114* (37), 11957.
(24) Nguyen, H. D.; Reddy, V. S.; Brooks, C. L. III. Invariant Polymorphism in Virus Capsid Assembly. *J. Am. Chem. Soc.* **2009**, *131* (7), 2606.
(25) Zhang, Z.; Pfaendtner, J.; Grafmüller, A.; Voth, G. A. Defining Coarse-Grained Representations of Large Biomolecules and Biomolecular Complexes from Elastic Network Models. *Biophys. J.* **2009**, *97* (8), 2327.
(26) Heath, A. P.; Kavraki, L. E.; Clementi, C. From coarse-grain to all-atom: Toward multiscale analysis of protein landscapes. *Proteins: Struct., Funct., Bioinf.* **2007**, *68* (3), 646.
(27) Sung, W.; Kim, Y. W. How Nature Modulates Inherent Fluctuations for Biological Self-Organization − The Case of Membrane Fusion. *J. Biol. Phys.* **2005**, *31* (3−4), 639.
(28) Mazack, M. J. M.; Cembran, A.; Gao, J. Internal Dynamics of an Analytically Coarse-Grained Protein. *J. Chem. Theory Comput.* **2010**, *6* (11), 3601.
(29) Pankavich, S.; Shreif, Z.; Miao, Y.; Ortoleva, P. J. Self-assembly of nanocomponents into composite structures: Derivation and simulation of Langevin equations. *J. Chem. Phys.* **2009**, *130* (19), 194115.
(30) Miao, Y.; Ortoleva, P. J. Molecular Dynamics/Order Parameter eXtrapolation (MD/OPX) for Bionanosystem Simulations. *J. Comput. Chem.* **2009**, *30* (3), 423.
(31) Miao, Y.; Johnson, J. E.; Ortoleva, P. J. All-Atom Multiscale Simulation of Cowpea Chlorotic Mottle Virus Capsid Swelling. *J. Phys. Chem. B* **2010**, *114* (34), 11181.
(32) Singharoy, A.; Cheluvaraja, S.; Ortoleva, P. J. Order parameters for macromolecules: Application to multiscale simulation. *J. Chem. Phys.* **2011**, *134* (4), 044104.
(33) Cheluvaraja, S.; Ortoleva, P. Thermal nanostructure: An Order Parameter/Multiscale Ensemble Approach. *J. Chem. Phys.* **2010**, *132* (7), 075102.
(34) Pankavich, S.; Shreif, Z.; Ortoleva, P. J. Multiscaling for Classical Nanosystems: Derivation of Smoluchowski and Fokker-Planck Equations. *Physica A* **2008**, *387* (16−17), 4053.
(35) Jaqaman, K.; Ortoleva, P. J. New space warping method for the simulation of large-scale macromolecular conformational changes. *J. Comput. Chem.* **2002**, *23* (4), 484.
(36) Shreif, Z.; Pankavich, S.; Ortoleva, P. J. Liquid-crystal transitions: A first-principles multiscale approach. *Phys. Rev. E* **2009**, *80* (3), 031703.
(37) Singharoy, A. B.; Sereda, Y. V.; Ortoleva, P. J. Hierarchical Order Parameters for Macromolecular Assembly Simulation. 2. Multiscale Dynamics and Simulation Algorithm. *J. Chem. Theory Comput.* **2011**, Submitted.
(38) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781.
(39) Patel, S.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004**, *25* (1), 1.
(40) Patel, S.; Mackerell, A. D.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004**, *25* (12), 1504.
(41) Bishop, B.; Dasgupta, J.; Klein, M.; Garcea, R. L.; Christensen, N. D.; Zhao, R. Crystal Structures of Four Types of Human Papillomavirus L1 Capsid Proteins. *J. Biol. Chem.* **2007**, *282* (43), 31803.
(42) Bishop, B.; Dasgupta, J.; Chen, X. Structure-based engineering of papillomavirus major capsid L1: controlling particle assembly. *Virol. J.* **2007**, *4* (1), 3.
(43) Joshi, H.; Singharoy, A. B.; Sereda, Y. V.; Cheluvaraja, S. C.; Ortoleva, P. J. Multiscale simulation of microbe structure and dynamics. *Prog. Biophys. Mol. Biol.* **2011**, *107* (1), 200.
(44) Zandi, R.; Reguera, D. Mechanical properties of viral capsids. *Phys. Rev. E* **2005**, *72* (2), 021917.
(45) Pankavich, S.; Shreif, Z.; Chen, Y.; Ortoleva, P. J. Multiscale theory of finite size bose systems: Implications for collective and single-particle excitations. *Phys. Rev. A* **2009**, *79* (1), 013628.