JCTC Journal of Chemical Theory and Computation

Internal Dynamics of an Analytically Coarse-Grained Protein

Michael J. M. Mazack, Alessandro Cembran, and Jiali Gao*

Department of Chemistry, Digital Technology Center, and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455, United States

Received July 31, 2010

Abstract: An analytically coarse-grained model (ACG) is introduced to represent individual macromolecules for the simulation of dynamic processes in cells. In the ACG model, a macromolecular structure is treated as a fully coarse-grained entity with a uniform mass density without the explicit atomic details. The excluded volume and surface of the ACG macromolecular species are explicitly treated by a spherical harmonic representation in the present study (although ellipsoidal, solid, and radial augmented functions can be used), which can provide any desired accuracy and detail depending on the problem of interest. The present paper focuses on the description of the internal fluctuations of a single ACG macromolecule, modeled by the superposition of low frequency quasiharmonic modes from explicit molecular dynamics simulation. A procedure for estimating the amplitudes, time scales of the quasiharmonic motions, and the corresponding phases is presented and used to synthesize the complex motion. The analytical description and numerical algorithm can provide an adequate representation of the internal protein fluctuations revealed from the corresponding atomistic simulations, although the internal motions of ACG macromolecules do not explore motions not exhibited in the dynamic simulations.

1. Introduction

Molecular dynamics simulations of biological macromolecules in explicit aqueous solution offer the most detailed information at the atomic level, which is essential for the understanding of dynamics, binding, and activity of these systems.[1] Tremendous progress has been made both in the advance of computer architecture and in the development of computational algorithms, enabling atomistic dynamics simulations to treat systems containing millions of atoms[2] and the dynamics lasting up to milliseconds.[3,4] However, the spatial and temporal scales needed to address questions relevant to cellular processes such as protein−protein and protein−nucleic acid interactions and macromolecular assembly dwarfs the most sophisticated atomistic approaches available today and perhaps in the distant future.[5] In such a mesoscopic system, it is necessary to use a coarse-grained approach to describe the individual macromolecular components.[6,7] In this series of papers, we present an analytically coarse-grained (ACG) model to represent macromolecular entities such as proteins and nucleic acids as single building blocks that can be used to study macromolecular interactions and assembly in biological cells.

In the past decade, significant efforts have been devoted to the development of coarse-grained models to circumvent the need for describing molecular systems with increasing demands in size and time.[8] Nevertheless, the concepts of atoms and molecules are deeply rooted in our perception of intermolecular interactions; not surprisingly, coarse-grained models typically involve interaction sites in terms of reduced representation of the detailed atomic features of the target system. The early united-atom force field is an example of this type of coarse graining,[9,10] and recent advances have enabled a much larger number of atoms to be grouped into a united site along with continuum elastic network models.[8] However, to model macromolecular interactions and assembly such as the mechanism of a virus capsid formation, the detailed sequence and the "united-atom" constituents are no longer critical, and it becomes unnecessary to enumerate the specific pairwise interactions between coarse-grained groups among proteins. On the other hand, the use of regular

* Corresponding author e-mail: gao@jialigao.org.

10.1021/ct100426m
Published on Web 10/01/2010

geometrical shapes seems too crude. Indeed, the key structural components are the specific shape and the excluded volume of each capsid protein along with their intrinsic dynamic fluctuations and the accompanying surface electrostatic potential and surface tension. The detailed atomistic interactions, of course, are essential for recognition and binding when two macromolecular particles are in close contact, but these are not the types of details needed for transport processes in the cell. It appears to be desirable to develop a theoretical method that is not restricted to the detailed features of atoms and molecules or coarse-grained interacting groups in a large system; yet, the individual macromolecular species still retain the information of, and are constructed based on the detailed atomistic coordinates determined experimentally, which also provide all physical and biological properties needed to model the dynamic system. Furthermore, the intrinsic fluctuations of the coarse-grained macromolecules relevant to the time scale of the dynamic model for the mesoscopic system need to be taken into account.[8] To limit the scope of discussion, the present article is only concerned with the representation of the internal dynamic fluctuations of an ACG macromolecule itself, which are derived from explicit molecular dynamics simulations.

To this end, we make use of the mathematical tools of spherical harmonic analysis to represent the macromolecular particles of interest; spherical harmonic analysis has been extensively applied in a wide range of areas such as geopotential,[11] topography,[12] and physics as well as motion picture and gaming animation. Since analytical harmonic basis functions are used, the method that we design for modeling cellular processes is called the analytical coarse-graining (ACG) of macromolecules. The harmonic representation can also be used directly to describe the physical interactions and to model the dynamics of the system, which will be detailed in the next papers. Our strategy provides a single, unifying theory and computational algorithm to study macromolecular systems consisting of thousands of macromolecular particles and entities. In such an approach, each macromolecular unit, such as a protein, is "coarse-grained" as a single moiety of uniform mass density whose excluded volume is encompassed by its solvent-accessible surface that is represented by a set of analytical harmonic basis functions. Furthermore, its physical and biological properties can be treated by exactly the same mathematical procedure as the representation of the macromolecule. Here, we describe the treatment of the intrinsic dynamic fluctuations of a single ACG protein using spherical harmonic functions.

Although the mathematical tools of spherical harmonic computation have been established since the 1780s and modern numerical techniques have greatly enhanced the computational speed,[13,14] Max and Getzoff,[15] and Olson and co-workers, were among the first to apply spherical harmonic functions to the visualization of molecular surfaces as a graphics rendering tool.[16−20] Recent years have seen the increased usage of this technique to model protein−ligand interactions and protein docking.[21−25] Buchete et al. used spherical harmonics to analyze coarse-grained potentials for folding calculations.[26] In a subsequent publication, Duncan and Olson described the possibility of animating the dynamic motion of a protein to render real-time graphics visualization;[27] however, it does not appear that specific investigations have been reported. The approach described by Olson and co-workers was aimed to follow the time-dependent Cartesian coordinates of a protein surface as modeled by normal mode dynamics; the method provides an efficient procedure to generate graphics rendering, but it does not guarantee single-valued properties on the surface, nor is it suitable for modeling protein dynamic motions.[19,27] In contrast, the method described in this article focuses on radial fluctuations of an ACG protein surface that concerns no atomic details, but it is designed to model the most significant dynamic motions revealed from an explicit molecular dynamics trajectory, in which the radial fluctuations are decomposed based on quasiharmonic dynamic analysis. The latter has been extensively explored to characterize large amplitude motions and quasiharmonic vibrational modes of proteins and nucleic acids;[28−32] it provides an adequate analytical procedure to describe the global large amplitude motions of a fully coarse-grained macromolecule with spherical harmonic representation. In addition, for a set of well-chosen numerical quadrature points in the spherical harmonic analysis, our approach provides an efficient procedure for evaluation of molecular properties.

In the following, we first present the analytically coarse-grained (ACG) model and computational details, focusing on the use of spherical harmonic basis functions. Then, we describe a procedure for constructing a mathematical approac to describe the intrinsic quasiharmonic dynamic fluctuations of a protein based on the information from molecular dynamics simulations in explicit solvent. This is followed by an illustrative example to show the feasibility of modeling protein fluctuations without explicit atomic details using the ACG model. Finally, we summarize the key findings of the present study.

2. Method

Throughout this article, we mainly use the homodimeric enzyme, orotidine 5′-monophosphate decarboxylase (OMP-DC), as an illustrative example, which consists of two β-barrels of eight strands of β-sheet and eight α-helices.[33] In this article, for convenience of discussion, we focus on the use of a spherical harmonic basis, with which the ACG method is applicable to any macromolecular systems that have star-like topology.[15] We note here that this is not a restriction because any geometrical shapes including non-star-like macromolecules can be represented in the ACG model by augmenting radial functions such as the Zernike function or Slater-type radial functions,[23−25,34,35] but we shall not discuss these approaches here. In the rest of the paper, we interchangeably use the terms of "protein" and "macromolecule", which include proteins, nucleic acids, lipids, and other components of a macromolecular assembly, without specific distinction. In this section, we first outline a qualitative description of the coarse-grained macromolecular model. Then, we provide a brief summary of spherical harmonic representation of the surface of a macromolecular structure. This is followed by the description of incorporating

internal dynamic fluctuations of the protein based on the information obtained from principal component analysis (PCA)[29,30] of the atomistic molecular dynamic trajectory of OMPDC in water.[33] The purpose in this article is not aimed at studying the dynamics of proteins using quasiharmonic dynamics; the latter has been thoroughly investigated and its applications and limitations have been characterized.[29−32] The goal here is to illustrate the capability and procedure of incorporating low-frequency, large-amplitude dynamic fluctuations, as revealed by an explicit dynamics simulation, into a fully coarse-grained protein model.

2.1. Description of a Macromolecular Particle. We consider a macromolecular structure, which can be a protein or a domain of a protein complex, a segment of nucleic acids, or a protein−nucleic acid complex, as a single entity of uniform mass density. The excluded volume of a given macromolecular structure is defined as the cavity enclosed by the solvent-accessible surface (or the van der Waals surface depending on needs), originating from the detailed three-dimensional atomic structure determined experimentally by X-ray crystallography or NMR, or generated computationally by homology modeling and protein-folding prediction in the absence of experimental data. Note that the solvent-accessible surface encloses a molecular volume which may be significantly greater than that defined by its van der Waals surface, the latter of which is more appropriate for evaluating the macromolecular density. All biochemical functions and physical properties of the macromolecule are fully encoded in the three-dimensional structure, necessary for microscopic and mesoscopic modeling of intermolecular interactions, including electrostatics and hydrophobic surface tension. The characteristic features of a macromolecular structure are considered to have distinguishing features both in size and property from small molecules, peptide fragments, ligands and cofactors, ions and solvent molecules, although the specific criteria depends on a particular application.

From the onset, we do not consider the detailed atomic coordinates or interaction sites; the entire macromolecular unit is a single coarse-grained entity. This is justified as a result of statistical averaging over the spatial and temporal scales to be used to model the dynamic system, which is the cell. However, the size, as defined by the excluded volume occupied by the individual atoms, ligands, and perhaps a small number of buried or surface solvent molecules, and the shape, as represented by the solvent-accessible surface, of a given macromolecule are necessary and critical to a physics-based approach; they are well-defined in our ACG model (vide infra) by a single mathematical approach for all types of macromolecular particles of different sizes and shapes, which can be used to systematically provide any desired accuracy and detail of the coarse-grained macromolecule. The definition of a uniform mass density within its excluded volume is akin to the use of a continuum solvent model and a single interior low dielectric constant for a protein in Poisson−Boltzmann calculations and is consistent with our goal of modeling the dynamics of the entire system, which involves the integration of equations of motion at a time step in the order of tens of picoseconds to nanoseconds per iteration. Thus, the representation of the macromolecular species is an average of the system over the time series of the coarse-grained model,[8,36,37] involving the internal atomic fluctuation and spatial orientation when the center of mass of the macromolecule is chosen as a reference point.

Throughout this article, the method of Lee and Richards is used to define the macromolecular surface and the excluded volume,[38] although other approaches are available.[39]

2.2. Spherical Harmonic Representation of a Macromolecular Particle. The method of using spherical harmonics functions to represent the surface of globular proteins was described by Max and Getzoff[15] and later by Duncan and Olson and by others.[16,17,19,20] Here, we briefly summarize the key elements of this approach and highlight the numerical details employed in our implementation.

Arising from the solution of Laplace's equation in spherical coordinates, the spherical harmonic expansion is the spherical coordinate analog of the widely used Fourier series expansion. Spherical harmonic representations of macromolecules provide not only a smooth and aesthetically pleasing surface, but also the ability for evaluating surface properties such as normal vectors and principal curvature. For any star-like surface, which is single valued in the radial direction of (θ,ϕ) with respect to an origin, there exists a spherical harmonic expansion given as follows:

$$S(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_l^m(\theta,\phi) \tag{1}$$

where θ denotes the latitudinal or zenith angle ($0 \leq \theta \leq \pi$), ϕ specifies the longitudinal or azimuth angle ($0 \leq \phi < 2\pi$), $S(\theta,\phi)$ is the radial distance of the surface at angular coordinate (θ,ϕ), a_{lm} are the expansion coefficients and $Y_l^m(\theta,\phi)$ are the real spherical harmonic basis functions, which are orthonormal under the L^2 inner product. We have used the center of mass as the origin in all calculations, and the local axis is generally chosen to coincide with the principal moments of inertia. In general, $S(\theta,\phi)$ can be any scalar physical or chemical property mapped on to the surface of a unit sphere.

The real spherical harmonic basis functions are defined by

$$Y_l^m(\theta,\phi) = \begin{cases} \frac{1}{\sqrt{2\pi}} \bar{P}_l^0(\cos\theta) & m = 0 \\ \frac{1}{\sqrt{\pi}} \bar{P}_l^m(\cos\theta)\cos m\phi & m > 0 \\ \frac{1}{\sqrt{\pi}} \bar{P}_l^{-m}(\cos\theta)\sin m\phi & m < 0 \end{cases} \tag{2}$$

where $\bar{P}_l^m(\cos\theta)$ denotes the normalized associated Legendre polynomial of the first kind, of order m and degree l. Methods exist for evaluating the Legendre polynomials of eq 2, one of which is discussed in ref 40, and ref 22 lists a technique yielding more numerically stable results.

Making use of the orthonormal property of the real spherical harmonic basis functions under the L^2 inner product, the expansion coefficients in eq 1 are given by

$$a_{lm} = \int_0^{2\pi} \int_0^{\pi} S(\theta,\phi) Y_l^m(\theta,\phi) \sin\theta \, d\theta \, d\phi \tag{3}$$

In practice, the infinite sum of eq 1 is truncated to a relatively small value L, which produces an approximate surface **S**. Thus, it is only necessary to determine $(L + 1)^2$ coefficients and function values for the evaluation of points on the surface:

$$S(\theta,\phi) \approx \sum_{l=0}^{L} \sum_{m=-1}^{l} a_{lm} Y_l^m(\theta,\phi) \tag{4}$$

The Cartesian coordinates of vertices used to form a triangulated mesh for graphics display are obtained from the corresponding values in the polar spherical coordinates. This differs from the approach of Duncan and Olson,[16,17] who used spherical harmonics expansions to directly approximate the Cartesian coordinates of surface points. Although the direct representation of the surface Cartesian coordinates is convenient for graphics display, it is not suitable for computation of molecular properties of the system, including intermolecular interactions.

2.3. Dynamic Motion. *2.3.1. Quasiharmonic Dynamics.* We use the lowest frequency modes from principal component analysis (PCA) of a molecular dynamics trajectory of a solvated protein to represent its dynamic fluctuations; the dimeric enzyme OMPDC is employed as an illustrative example. The PCA results show the directionality and frequency of protein dynamic motions, in which the lowest frequency modes are typically correlated with protein conformational changes and have been used to interpret conformational variations observed experimentally.[41] Although other approaches such as the continuum elastic network model can be used,[42] for an analytically represented coarse-grained protein without the explicit details of atomic structure, the PCA modes provide the most direct connection to the dynamic motions sampled during an explicit molecular dynamics simulation. The animation of atomic motions following a given normal mode and quasiharmonic dynamics simulation of the complex motions of a macromolecule have been widely used in structure and dynamics analyses at the atomic details. Voth and co-workers described a method to map coarse-grained sites on the basis of PCA modes,[43] and the model has been extended to using the low-frequency normal modes of an elastic network model for the protein.[44] Our method follows a different route of representation than that of Zhang et al.;[43,44] in ACG, the model is used to represent and animate the low-frequency PCA modes, rather than being derived from PCA. We apply the approach of quasiharmonic dynamics to model the internal fluctuations of coarse-grained macromolecules.

The overall protein fluctuation is obtained by the superposition of individual quasiharmonic modes:

$$\mathbf{R}_j(t) = \mathbf{R}_j(0) + \sum_{k}^{K} \mathbf{Q}_{jk} \sigma_k \cos(\omega_k t + \lambda_k) \tag{5}$$

where $\mathbf{R}_j(0)$ and $\mathbf{R}_j(t)$ are the coordinates of atom j at time 0 and t, respectively, and K is the number of quasiharmonic modes used to animate the total dynamic fluctuation of the system. In eq 5, the parameters associated with mode k, ω_k, $\mathbf{Q}_k$, λ_k, and σ_k are the frequency, mode direction eigenvector, phase, and amplitude, respectively. The phase λ_k is associated with the initial atomic positions, and the thermal average of the second moment of the amplitude distribution is given by $\sigma_k^2 = k_B T/\omega_k^2$, where k_B is Boltzman's constant and T is temperature. The value of K in eq 5 is restricted by the integration time increment, τ, used to propagate the dynamic equations of the coarse-grained system such that $\tau > 2\pi/\omega_K$. Typically, the inclusion of the lowest 10 to 20 modes is more than sufficient to represent the most significant large-amplitude motions.

Here, we use the lowest frequency quasiharmonic motions to represent the internal dynamic fluctuations of analytically coarse-grained macromolecule particles. The limitation of this approach is that it will not produce information for even larger amplitude motions that have not been uncovered in the explicit molecular dynamics simulation. Thus, if the protein undergoes folding and unfolding exchange, it is not appropriate to use the present model; however, it is possible to incorporate into the present treatment conformational transitions for which structures in different conformation substates have been determined experimentally (e.g., by X-ray crystallography or NMR). Nevertheless, our approach is not a simple reproduction of the fluctuation of the molecular dynamics trajectory itself because collision between different coarse-grained macromolecular species can cause random changes in the amplitude and phase of each quasiharmonic mode, resulting in different combinations of modes and trajectories.

Previously, Duncan and Olson proposed a method for shape analysis of protein dynamic surfaces.[27] In that approach, the Cartesian coordinate displacements of surface points corresponding to triangulation vertices were obtained from the static surface and the expansion coefficients for the normal mode eigenvectors projected to these points. Although the method is extremely useful for fast visualization of surface motion, it is not designed to model real-time dynamics. Furthermore, the displacements of triangulation vertices in such a shape analysis algorithm are not suited for property evaluation because the quadrature points and weights will have to be recomputed, which is impractical for real time dynamics simulations. In our approach, the surface deformation is restricted to the direction along the radial vector, consequently preserving the angular coordinates (θ_i,ϕ_j) and the precomputed numerical weights.

2.3.2. Definition of Surface Displacement Vector. We begin with a molecular dynamics trajectory $\{\mathbf{R}(t_n); n = 0, 1, ..., N\}$ that was saved at time slice t_n, where $\mathbf{R}(t_n)$ is a vector of all atomic coordinates. Principal component analysis of this trajectory yields a set of quasiharmonic vibrations with frequencies $\{\omega_k\}$ and eigenvectors $\{\mathbf{Q}_k\}$. The K lowest frequency modes will be used to model the total dynamic motions that have been sampled by the original molecular dynamics simulation.

For each mode k, we use two distorted configurations generated by following the eigenvector direction stretched to $-2\sigma_k$ and $+2\sigma_k$ from its mean, denoted by $\mathbf{R}_k^{-2\sigma}$ and $\mathbf{R}_k^{+2\sigma}$, to represent approximately the "lower" and "upper" bound of an amplitude, respectively. Let $\{\mathbf{S}_k^{-2\sigma}\}$ and $\{\mathbf{S}_k^{+2\sigma}\}$ be the

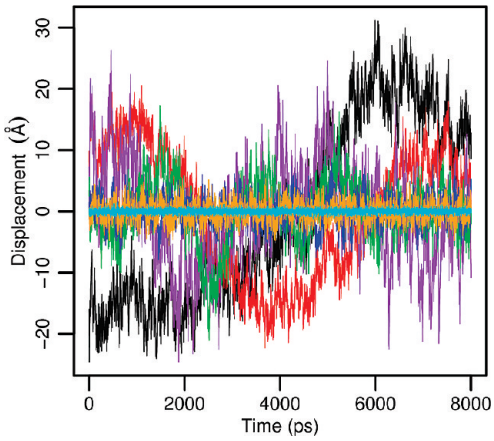

Figure 1. Histogram of the computed projection of instantaneous molecular structure of OMPDC in water onto normalized eigenvector directions of the lowest quansiharmonic mode (black), the second (red), the fourth (purple), the tenth (green), the fiftieth (blue), the one hundredth (orange), and the one thousandth (cyan) modes.

solvent-accessible surfaces (SAS) for the two extreme configurations associated with mode k. Given a set of surface points, $\{u_{ij} = (\theta_i,\phi_j);\ i = 1, ..., M_\theta;\ j = 1, ..., M_\phi\}$, where M_θ and M_ϕ are the number of quadrature points, the radial displacement, due to quasiharmonic vibration of mode k, at the surface point $u_{ij} = (\theta_i,\phi_j)$ is defined as the $2\sigma_k$ variance:

$$q_{ij}^k = \frac{1}{2}[S_k^{2\sigma}(\theta_i,\phi_j) - S_k^{-2\sigma}(\theta_i,\phi_j)] \tag{6}$$

The vector $\mathbf{q}^k$, which can be considered as a property of a unit sphere, represents the approximate amplitude (see below) and direction of surface deformation associated with PCA quasiharmonic mode k:

$$\mathbf{q}^k = \begin{pmatrix} q_{11}^k \\ q_{21}^k \\ \cdots \\ q_{P_\theta P_\phi}^k \end{pmatrix} \tag{7}$$

Consequently, the radial displacement vector can also be expressed by a spherical harmonic expansion whose coefficients are determined using the same procedure as for the molecular surface itself (eq 3):

$$q_{ij}^k = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}^k Y_l^m(\theta_i,\phi_j) \tag{8}$$

2.3.3. Frequency and Phase. Although the rank of the low frequency modes from principal component analysis is very reasonable, the quantitative values of the lowest quasiharmonic vibrational frequencies and the associated time scales are not expected to be accurate to represent the real time dynamic motion.[29−32] Thus, one needs to seek a different way to obtain the desired oscillatory frequencies. We examined the time dependence of the projections of the instantaneous coordinate vector onto the normalized quasiharmonic eigenvectors, seven of which are shown in Figure 1, corresponding to PCA mode numbers 1, 2, 4, 10, 50, 100, and 1000 over a total of 8 ns MD trajectory. Although not unexpected, we are pleased to see the oscillatoroy behavior of each mode, and the amplitudes and frequencies of these oscillations roughly coincide with the order of the PCA modes. For modes above number 50, the fluctuations illustrated in Figure 1 can be considered as noise (friction) with respect to the motions of the lowest frequency modes. Importantly, it appears that the PCA mode-projection results can be used to estimate the quantitative frequencies as well as the phase with respect to the structure at time $t_0 = 0$ needed to animate the complex motion of the superimposed fluctuations.

To this end, we used a sinusoidal fitting procedure to optimize the amplitude A_k (not used for mode animation, see below), frequency ω_k, and phase λ_k in eq 9 for each mode to best reproduce the time-dependent quasiharmonic mode projection data.

$$M_k(t) = A_k \cos(\omega_k t + \lambda_k) \tag{9}$$

In Figure 2, we depict the fitted curves against the raw data for modes 1, 2, 4, and 10, whereas the results for the first 20 modes are given as Supporting Information. Table 1 lists the optimized amplitudes, frequencies, and phases for the first 20 modes, the first 10 of which are used to represent the overall protein dynamics fluctuations as an illustrative example in this article. An alternative approach is to use the spectral transform of the autocorrelation function of the quasiharmonic mode fluctuations.

2.3.4. Time Evolution of the Dynamic Fluctuation. The SAS surface $\mathbf{S}(t_0)$ corresponding to the structure $\mathbf{R}(t_0)$ at time $t_0 = 0$ in the dynamic trajectory is chosen as the starting configuration and is expressed in terms of spherical harmonics basis as follows:

$$S_{ij}(0) \equiv S(t = 0, \theta_i, \phi_j) = \sum_{l=0}^{L}\sum_{m=-l}^{l} \mathbf{a}_{lm}^o Y_l^m(\theta_i,\phi_j) \tag{10}$$

where $S_{ij}(0)$ is the radial distance at an angular coordinate $u_{ij} = (\theta_i,\phi_j)$, and the coefficients $\{\mathbf{a}_{lm}^o\}$ are determined according to eq 3. We assume that the dynamic modulation of the protein surface associated with mode k also has the same frequency. Thus, the atomic coordinates in eq 5 are replaced by protein surface points, and we write the total surface radial displacement at point $u_{ij} = (\theta_i,\phi_j)$ as follows:

$$\begin{aligned} S_{ij}(t) &= S_{ij}(0) + \sum_{k=1}^{K} W_k q_{ij}^k \cos(\omega_k t + \lambda_k) \\ &= \sum_{l=0}^{L}\sum_{m=-l}^{l} [\mathbf{a}_{lm}^o + \sum_{k=1}^{K} W_k c_{lm}^k \cos(\omega_k t + \lambda_k)] Y_l^m(\theta_i,\phi_j) \end{aligned} \tag{11}$$

where W_k is a mode weighting factor to be determined by least-squares fit to the instantaneous surfaces $\tilde{\mathbf{S}}(t_n)$ of the structures sampled by the explicit molecular dynamics simulations in the entire trajectory, $\{\mathbf{R}(t_n) \rightarrow \tilde{\mathbf{S}}(t_n);\ n = 0, 1, ..., N\}$. Equation 11 preserves the angular coordinates, consequently all precomputed values of the spherical harmonic functions and quadrature weights needed for

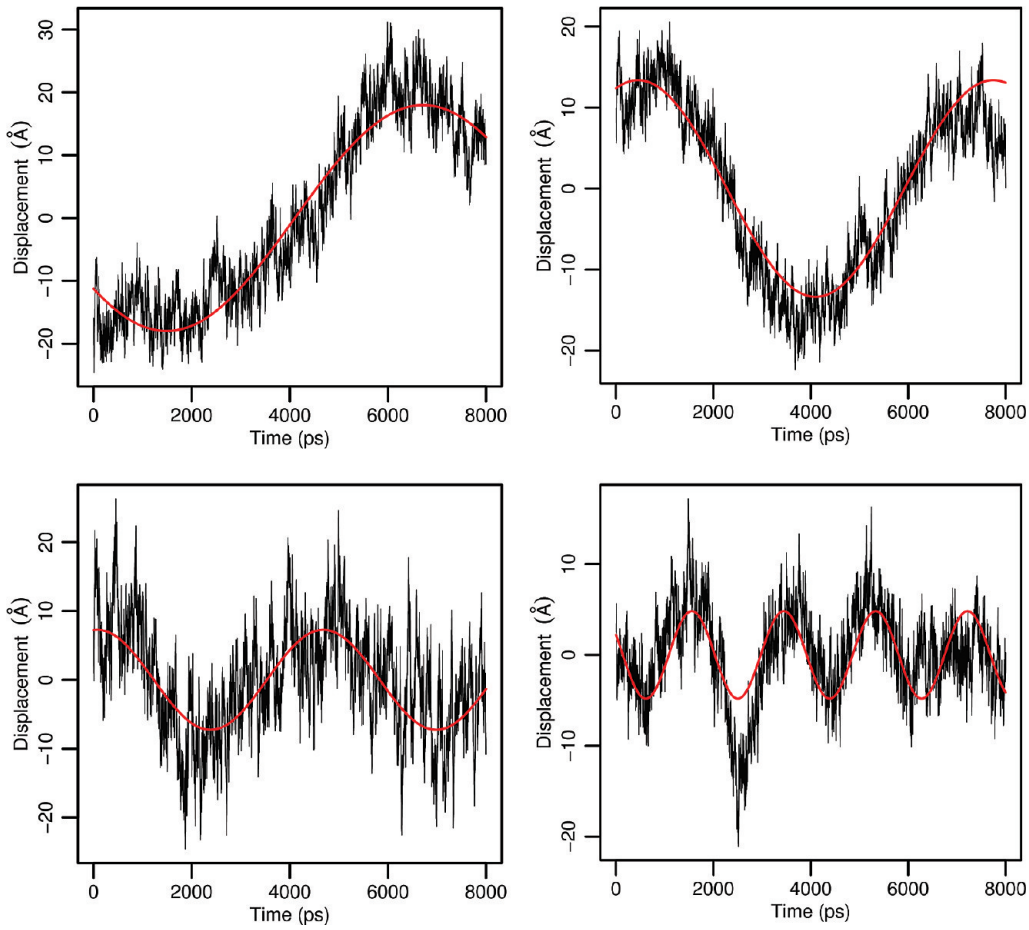

Figure 2. Sinusoidal fit of harmonic frequencies and phases (with respect to the structure used at the start of the molecular dynamic simulation of OMPDC) to the oscillatory structural projections illustrated in Figure 1 for modes number 1, 2, 4, and 10.

Table 1. Optimized Amplitudes (Å), Frequencies (rad/ns), and Phase (rad) for the Time-Dependent Quasiharmonic Mode Projection along the Molecular Dynamics Trajectory of the Protein Orotidine Monophosphate Decarboxylase As Represented by Eq 16

mode	A_i	ω_i	λ_i
1	18.0	0. 60	2.25
2	13.4	0. 86	5.89
3	9.8	1.55	1.05
4	7.3	1.37	6.17
5	3.0	2.71	2.50
6	1.3	7.37	4.02
7	6.5	2.12	2.97
8	5.0	2.72	1.62
9	3.9	2.66	3.32
10	4.8	3.33	1.10
11	3.2	3.37	6.00
12	3.3	2.65	1.22
13	0.4	3.79	0.69
14	2.7	4.10	5.47
15	2.6	4.17	2.66
16	0.9	5.01	6.21
17	2.1	4.72	1.33
18	2.5	4.22	5.78
19	1.7	7.59	1.71
20	1.2	7.59	1.60

property evaluations (as well as for real-time graphics animation).[45]

The mode weighting factors in eq 11 are determined by minimizing the following error function:

$$\varepsilon = \frac{1}{N}\sum_{n=1}^{N}\sum_{ij}[S_{ij}(t_n) - \tilde{S}_{ij}(t_n)]^2 \tag{12}$$

It is straightforward to show that the minimization yields a linear equation that can be conveniently solved.

$$\sum_{k=1}^{K} B_{qk}W_k = D_q; \quad q = 1, \ldots, K \tag{13}$$

where the matrix elements are defined as follows:

$$D_q = \frac{1}{N}\sum_{n=1}^{N}\sum_{ij}[\tilde{S}_{ij}(t_n) - \tilde{S}_{ij}(0)]U_{ij}^{q}(t_n) \tag{14}$$

$$B_{qk} = \frac{1}{N}\sum_{n=1}^{N}\sum_{ij}U_{ij}^{q}(t_n)U_{ij}^{k}(t_n) \tag{15}$$

$$U_{ij}^{q}(t_n) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}^{q}\cos(\omega_q t_n + \lambda_q)Y_l^m(\theta_i,\phi_j) \tag{16}$$

3. Numerical Considerations

The spherical harmonic expansion coefficients are determined by sampling surface values (coordinates) to approximate the integral of eq 3. A number of methods for optimizing surface point distribution are available including the use of a geodesic unit sphere.[46] In the present application, realizing that the

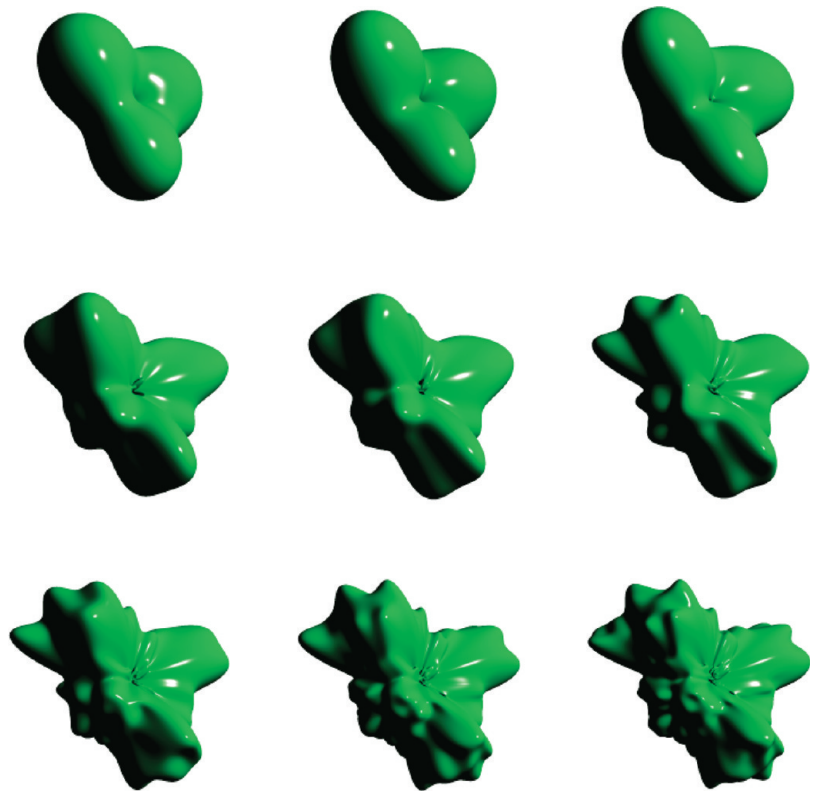

Figure 3. Spherical harmonic reconstruction of the Lee and Richards surface for the trimer complex of the capsid protein of turnip yellow mosaic virus using representation degrees of $L = 3$, 4, 5, 10, 12, 15, 20, 25, and 30 numbered from top left to bottom right.

integral in θ is formally a Fourier transform, the numerical integration can be evaluated using M equispaced points to take advantage of fast Fourier transform (FFT) at a computing scaling of O(Mlog M). There are two ways of selecting points in ϕ; the first is to use Gauss−Legendre quadrature nodes and weights, which needs only $M/2$ points, whereas a set of equally spaced points can be selected, which is equivalent to Chebychev nodes in cos ϕ.[45,47] The latter is convenient to use but less efficient computationally and requires a total of M points for the same accuracy. The numerical scaling for integrating in ϕ is O(M^3).

If the sampling points used in the evaluation of the integral in eq 3 are chosen to coincide with the numerical quadrature values, $\{(\theta_p,\phi_q); p = 1, ..., M_\theta; q = 1, ..., M_\phi\}$, the numerical procedure for property calculation can be greatly simplified. In all cases, the associated Legendre function values are precomputed along with the measure, sin θ_i, and quadrature weights (see Appendix) for a given structure and stored. The use of $(\sin \theta_i)^{1/2}P_l^m(\theta_i,\phi_j)$ preconditioning in property calculations can greatly increase numerical stability by keeping the product roughly constant.[47]

All computations and illustrations are performed using a software package written in our laboratory.

4. Discussion

Figure 3 illustrates the spherical harmonic rendering of the trimeric structure of the capsid protein (2FZ2) of turnip yellow mosaic virus at various degrees of representation up to $L = 30$. By simple inspection, the domed triangular shape is not directly associated with a unit sphere, but the 3-fold symmetry of the complex is already represented at $L = 3$, and the domed structural feature is clearly visible with $L =$ 4 and 5. As the degree of spherical harmonic basis increases, the molecular shape and detail is well described with an L above 10, while greater local features can be found using higher degrees. In principle, the spherical harmonic representation can yield any desired accuracy by increasing the value of L. However, it should be kept in mind that the protein or macromolecular structure that we model is a coarse-grained representation of a distribution over the time scale of the integration step used in Brownian dynamics simulations. Thus, there is no reason to use a very high degree of L to generate an "accurate" surface that is in fact beyond the variance of the surface amplitude fluctuation over the time interval in Brownian dynamics simulations. In fact, a certain degree of fuzziness is especially desired for these computations, a subject to be addressed in the future. We have found that a value of $L = 10-15$ is adequate to provide a compromise of quantitative shape and volume description and sufficient distinguishing details of different proteins. At this level of representation, a total of 121 to 256 terms is needed in the spherical harmonic expansion in eq 1.

To animate the dynamic fluctuation of spherical harmonics coarse-grained OMPDC, we have determined the surface radial displacement amplitudes of the ten lowest quasiharmonic modes of vibration from principal component analysis, which contribute to the overall fluctuation throughout the 8 ns molecular dynamics simulation. The optimized mode weighting factors in eq 11 for the first ten modes are listed in Table 2. The weighting factors are about 0.5 for these low frequency modes, which is a reflection that the approximate "amplitudes" used to define the surface radial

Table 2. Computed Mode Weighting Factors to Represent the Overall Complex Protein Fluctuations Using the First Ten Lowest Quasiharmonic Modes

mode	W_k
1	0.53892
2	0.48497
3	0.48692
4	0.42098
5	0.09645
6	0.05428
7	0.42884
8	0.29127
9	0.33406
10	0.37536

displacement vectors by stretching the quasiharmonic deformation to $\pm 2\sigma$ limits is used so as to obtain a large contrast in the analysis. The minimization procedure reduces the initial large variations to about $\pm 1\sigma$, further suggesting that the procedure employed in the present study is a reasonable approximation to represent the overall protein fluctuation. However, as the frequency (mode number) increases, one standard deviation is not a good measure of the dynamic contributions due to stochastic collision and coupling with fast motions. The small weighting factors for modes 5 and 6 indicate that the associated fluctuations from the principal component analysis may not be well described by quasiharmonic vibrational motions (Figure 4), perhaps due to conformational jumps, or a longer equilibration that is needed in the original MD simulation, or the fact that the explicit molecular dynamics simulation is rather short.

Using the frequencies, phases, and amplitudes optimized using the procedure outlined in section 2 by means of principal component analysis of a molecular dynamics trajectory to train the large amplitude dynamic behavior of the ACG model for OMPDC, we carried out quasiharmonic dynamics animation of the compounded motion of the ten lowest frequency quasiharmonic modes for 0.25 μs at an integration increment of 25 ps per step, which took about 1 min on a desktop workstation. Figure 5 shows three structures from the trajectory using the initial conditions listed in Tables 1 and 2. Although it is difficult to distinguish the relatively small surface variations in the static pictures, a movie that is given as Supporting Information (SMovie 1) does provide a more vivid depiction of the dynamic fluctuations of the trajectory. Figure 6 shows the computed volume histogram of the ACG protein (see below). Not surprisingly, the primary periodicity is dictated by the lowest frequency mode, which has the largest amplitude contributions to the complex motion (Tables 1 and 2), and the spectral transform of Figure 6 shows frequencies that coincide with the input listed in Table 1.

Note that although the surface radial displacement vectors were obtained by considering the corresponding quasiharmonic modes of atomic vibrations, the radial vectors do not possess an orthorgonality relationship, and the least-squares fitting procedure used to optimize the displacement amplitudes also introduces contributions from other modes not specifically characterized purely by each quasiharmonic mode. Further, the amplitude for each quasiharmonic mode represents an average fluctuation sampled in the original molecular dynamics simulation; however, the maximum

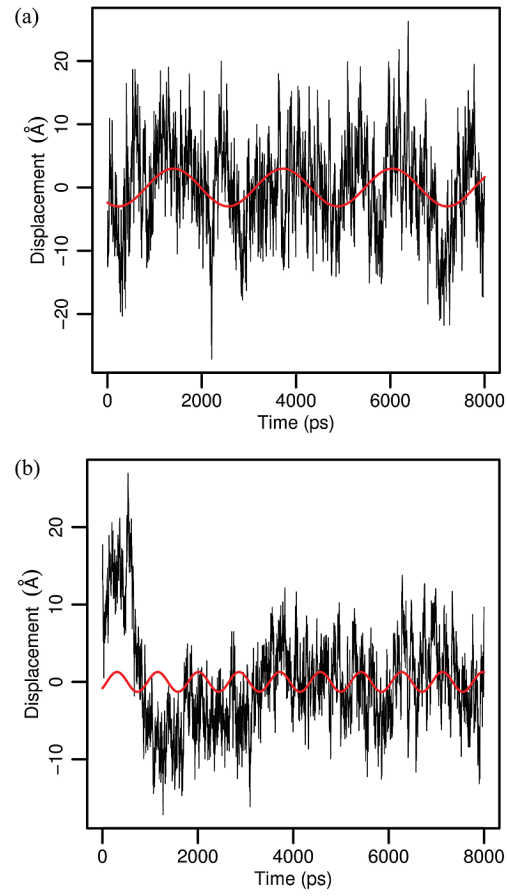

Figure 4. Sinusoidal fit of harmonic frequencies and phase (with respect to the structure used at the start of the molecular dynamic simulation of OMPDC) to the oscillatory structural projections illustrated in Figure 1 for modes number 5 (a) and 6 (b). Note that if the trajectory of the first 1 ns is discarded in mode 6 evaluation, the frequency and amplitude are both reasonable, suggesting there is either a conformational jump or change in the first 1 ns.

fluctuations can be significantly greater than the individual averages due to stochastic collisions with solvent molecules as well as mode coupling. Consequently, stochastic effects may be included in mode synthesis by randomly increasing and decreasing the amplitudes that yield the correct means over a long trajectory and satisfy the condition of the second dissipation theorem.

The fluctuation of the excluded volume defined by the protein surface for OMPDC, which can be conveniently determined by

$$\begin{aligned} V &= \int_0^{2\pi}\int_0^{\pi}\int_0^{S(\theta,\phi)} r^2 \sin\theta \, d\theta \, d\phi \, dr \\ &= \frac{1}{3}\int_0^{2\pi}\int_0^{\pi} S^3(\theta,\phi)\sin\theta \, d\theta \, d\phi \\ &\approx \frac{1}{3}\sum_{ij} S^3(\theta_i,\phi_j)\sin\theta_i \, w_i w_j \end{aligned} \tag{17}$$

is shown in Figure 6 at different degrees of approximation from $L = 5$ to $L = 20$. In eq 17, w_i and w_j are the quadrature weights and the values $\{S(\theta_i,\phi_j)\}$ are already determined during the dynamics animation. There is no

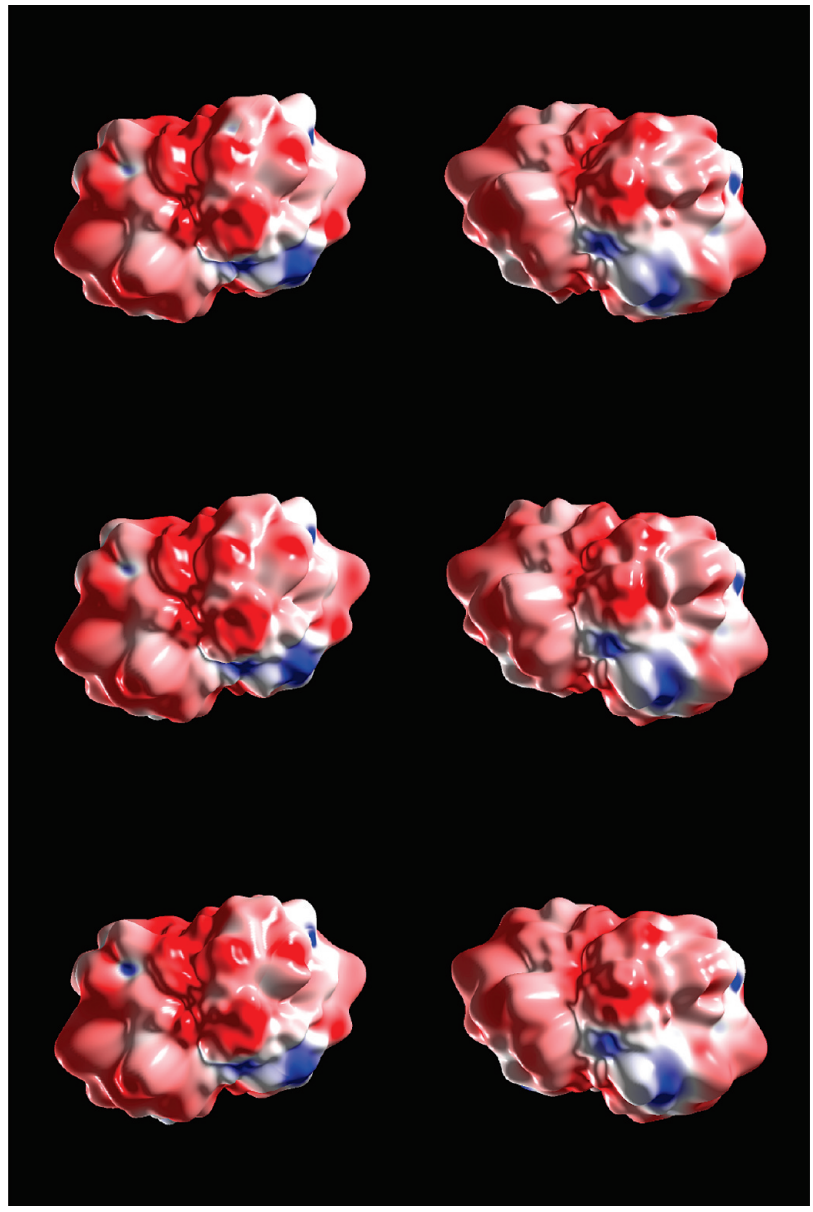

Figure 5. Snapshots of three structures of the analytically coarse-grained (ACG) protein OMPDC using spherical harmonic basis at a representation degree of $L = 15$ from the 0.25 μs composite fluctuation trajectory using the amplitudes, frequencies, and phases listed in Tables 1 and 2. The three structures on the right-hand side are the same as the corresponding ones on the left, rotated by 180°. The ACG protein surfaces are colored by the surface charge density for the illustration with red representing negative and blue positive charge densities, respectively.

major difference for the results obtained using $L = 15$ and $L = 20$, suggesting that the use of a spherical harmonic representation of degree 15 is sufficiently accurate to model the molecular volume. At $L = 10$, the average volume is about 84.3 ± 0.2 nm^3, which is less than 0.4% smaller than an average value of about 84.6 ± 0.2 nm^3 at higher orders. The lower order at $L = 5$ introduces an error of about 1% in volume. Note that the excluded volume determined by solvent-accessible surface is significantly larger than that encompassed by the van der Waals surface of the macromolecule. Using the initial structure in the molecular dynamics simulation, the ratio between the volumes defined by the solvent-accessible surface and the Bondi van der Waals surface (scaled by 1.20 as proposed by Luque and Orozco[48] in the calculation of solvation free energies treating the solvent as a polarizable dielectric continuum) is 1.46. If this factor is taken into account, the average molecular density of dry (without solvent molecules) OMPDC is estimated to be 1.315 ± 0.002 g/cm^3, in excellent agreement with the typical protein density (1.35−1.40 g/cm^3) estimated experimentally.[49,50] If the extra volume enclosed by the solvent-accessible surface is filled with water molecules (about 850 water molecules) at the bulk density, the average density of cavity included in the spherical harmonics coarse-grained protein is estimated to be about 1.20 g/cm^3, which may be considered as the macromolecular structure solvated by one solvent shell.

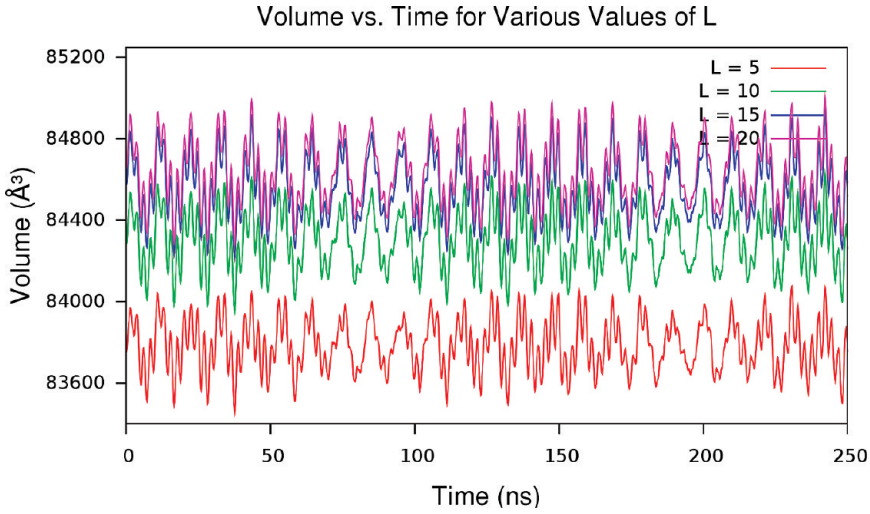

Figure 6. Histogram of the fluctuation of the excluded volume of the ACG OMPDC protein at various resolutions in *L*, ranging from 5 to 20 along the internal quasiharmonic fluctuation trajectory. The excluded volume illustrated in this figure is defined as the cavity enclosed by the Lee−Richards surface, which is about one solvent layer larger than the van der Waals surface.

5. Conclusion

An analytical coarse-graining (ACG) model has been introduced to represent biological macromolecules, making use of a spherical harmonic basis in the present study. In our approach, a macromolecular structure is treated as a fully coarse-grained entity with a uniform mass density without the explicit description of atomic details or "coarse-grained" interaction sites. The use of a uniform density of the ACG macromolecule is justified because the model represents an ensemble average relevant to the time series used in the dynamics simulation of cellular processes. However, the excluded volume and specific shape of the ACG macromolecule species are critical, which are explicitly treated by a spherical harmonic representation. In principle, spherical harmonic analysis can provide any desired accuracy and detail of the macromolecular surface. The present paper focuses on the first issue in a fully coarse-grained protein model, that is, the description of the internal fluctuation of the ACG macromolecule. Here, we make use of the dimeric enzyme OMPDC, consisting of 416 amino acids and 2 substrate molecules in the active site, as an illustrative example.

The internal fluctuation of the ACG protein is modeled by the superposition of a selected number of lowest frequency quasiharmonic modes of vibration, which are derived from an explicit molecular dynamics simulation of the fully solvated protein in water. A procedure for estimating the amplitudes, time scales (frequencies) of the quasiharmonic motions, and the corresponding phase is presented and used to synthesize the complex motion (note that the eigenvalues of the lowest quasiharmonic modes are close to zero and they are not quantitative for description of the time scales of the corresponding motions). In principle, all modes up to a frequency, limited by the time interval of the coarse-grained dynamics, can be included, but as numerous studies have shown, when employing principal component analysis and quasiharmonic essential dynamics, only a fraction of the lowest frequency modes are important in such a representation. The analytical description and numerical algorithm presented here can in principle provide a representation of the internal protein fluctuations as closely as needed in comparison with the atomistic molecular dynamics simulation; however, the internal motion is restricted by the short-time nature of molecular dynamic trajectories, and the present method is not designed for the description of unfolding events unless such transitions occur during the molecular dynamics simulation.

Appendix: Algorithm for Spherical Harmonic Expansion and Evaluation

In this Appendix, we summarize the numerical procedure for spherical harmonic expansion and evaluation used in our implementation; additional details may be found in ref 13 (see also refs 47 and 51). The numerical methods are widely used in geopotential modeling and an expansion of degree, and order up to 3800 has been reported corresponding to a ground resolution of about 5 km[52] (for the protein OMPDC considered here, it would correspond to an astonishing surface resolution of about 0.005 Å).

It is useful to recast eq 1 for the surface function $S(\theta,\phi)$ in terms of harmonic coefficients as follows:

$$S(\theta,\phi) = \sum_{m=0}^{L}\sum_{l=m}^{L} a_{lm}^{c}\bar{P}_{l}^{m}(\cos\theta)\cos m\phi + a_{lm}^{s}\bar{P}_{l}^{m}(\cos\theta)\sin m\phi \tag{A1}$$

Notice also that the order of the summations over degree and order has been switched. This is especially important in modern spherical harmonic analysis because the latitude and longitude data can now be treated independently, resulting in a two-step computational algorithm.[13,14,47] The expansion coefficients are determined based on a set of sampling data $\{S(\theta_i,\phi_j)\}$ on a grid of equispaced $2M$ points in ϕ (for Fourier transform) and M Gauss−Legendre quadrature nodes in $\cos\theta$ (for integration):

$$a_{lm}^{c} = \int_{0}^{\pi}\left\{\int_{0}^{2\pi}\frac{1}{\sqrt{(1+\delta_{m0})\pi}}S(\theta,\phi)\cos(m\phi)\mathrm{d}\phi\right\}\bar{P}_{l}^{m}(\cos\theta)\sin\theta\,\mathrm{d}\theta \tag{A2}$$

$$a_{lm}^{s} = \int_0^{\pi}\left\{\int_0^{2\pi}\frac{1}{\sqrt{(1+\delta_{m0})\pi}}S(\theta,\phi)\sin(m\phi)\mathrm{d}\phi\right\}\bar{P}_l^m(\cos\theta)\sin\theta\,\mathrm{d}\theta \tag{A3}$$

The grid points are defined as follows:

$$\phi_j = j\Delta\phi = j\frac{\pi}{M};\quad j = 0, 1, \ldots, 2M-1 \tag{A4}$$

$$P_{M+1}(\cos\theta_i) = 0;\quad i = 1, 2, \ldots, M+1 \tag{A5}$$

where $P_{M+1}(\cos\theta_i)$ is the Legendre polynomial. The Gauss−Legendre quadrature weights can be determined using the expression[40,45]

$$w_i = 2\left[\frac{\sin(\theta_i)}{(M+1)P_M(\cos\theta_i)}\right]^2 \tag{A6}$$

From this set of data, the maximum degree of the expansion coefficients that can be determined is $L = M$ because the maximum number of terms in the summation over l in eq A1, which is $L+1$ when $m = 0$, must be less than or equal to the number of sampling points in the longitudinal direction θ. Here, we do not address the issue of aliasing,[53] and we typically use a larger number of sampled grid points than the maximum degree used in the expansion.

Alternatively, if equal spaced points are used in θ, which is equivalent to Chebychev nodes in $\cos\theta$, at least $2M+1$ sampling points are needed for degree L in the expansion coefficients as opposed to a minimum of $M = L+1$ points with the Gauss−Legendre quadrature. In this case, the Chebychev weights are obtained using the formula below:[47,54,55]

$$w_i = \frac{\sqrt{2}}{M}\sum_{l=0}^{M-1}\frac{1}{2l+1}\sin([2l+1]\theta_i) \tag{A7}$$

For convenience in the rest of the discussion, we use the degree of the expansion L to define the grid divisions throughout.

A.1. Spherical Harmonic Expansion. For the spherical harmonics expansion, a two-step computation algorithm is used. The first step corresponds to a Fourier transform in the inner integrals in eqs A2 and A3. For a given value m, the discretized Fourier series in the inner integrals of eqs A2 and A3 are expressed as follows:

$$U(\theta_i,\hat{m}) = \frac{1}{\sqrt{(1+\delta_{m0})\pi}}\sum_{j=0}^{2L-1}S(\theta_i,\phi_j)\cos m\phi_j;\quad i = 1, \ldots, L+1 \tag{A8}$$

$$V(\theta_i,\hat{m}) = \frac{1}{\sqrt{(1+\delta_{m0})\pi}}\sum_{j=0}^{2L-1}S(\theta_i,\phi_j)\sin m\phi_j;\quad i = 1, \ldots, L+1 \tag{A9}$$

where δ_{m0} is the Kronecker delta, and the notation $\hat{m}$ is used to emphasize that the Fourier series can be efficiently performed by Fast Fourier Transform (FFT).

The second step involves integration by Gauss−Legendre quadrature (or equally spaced Chebychev quadrature which requires twice as many sampling points) to yield the $2\times(L-m+1)$ expansion coefficients:

$$a_{lm}^{c} = \sum_{i=1}^{L+1}w_i\bar{P}_m^l(\cos\theta_i)U(\theta_i,\hat{m});\quad l = m, \ldots, L \tag{A10}$$

$$a_{lm}^{s} = \sum_{i=1}^{L+1}w_i\bar{P}_m^l(\cos\theta_i)V(\theta_i,\hat{m});\quad l = m, \ldots, L \tag{A11}$$

If the values $U(\theta_i,\hat{m})$ and $V(\theta_i,\hat{m})$ are arranged as column vectors $\mathbf{u}(\hat{m})$ and $\mathbf{v}(\hat{m})$, respectively, the above equations can be conveniently written in matrix form:

$$\mathbf{a}_m^c = \mathbf{P}(m)^T\mathbf{W}\mathbf{u}(\hat{m}) \tag{A12}$$

$$\mathbf{a}_m^s = \mathbf{P}(m)^T\mathbf{W}\mathbf{v}(\hat{m}) \tag{A13}$$

where $\mathbf{a}_m^c$ and $\mathbf{a}_m^s$ are the expansion coefficients of eqs A10 and A11 arranged as column vectors, $\mathbf{W} = diag\{w_i\}$ is an $(L+1)\times(L+1)$ diagonal matrix consisting of the quadrature weights and the matrix for the precomputed values of the normalized associated Legendre polynomial is arranged as follows:

$$\mathbf{P}(m) = \begin{pmatrix}\bar{P}_m^m(\cos\theta_1) & \cdots & \bar{P}_L^m(\cos\theta_1)\\ \cdots & \cdots & \cdots\\ \bar{P}_m^m(\cos\theta_{L+1}) & \cdots & \bar{P}_L^m(\cos\theta_{L+1})\end{pmatrix} \tag{A14}$$

The operation for the first step has a computation scale of $O(L\log L)$ using FFT, whereas the second step is of $O(L^2)$ for each order m. Thus, the overall procedure scales $O(L^2\log L) + O(L^3)$. Obviously, the $L+1$ parallels can be fully distributed over different processors, each having an overall computational scaling of $O(L^2)$; this is particularly suited for GPUs by choosing the number of parallels equal to that of the processors. Note that a fast spherical harmonic transform algorithm similar to that of FFT has been described.[47,54]

A.2. Spherical Harmonic Evaluation. Evaluation (or spherical harmonic synthesis) of surface function values also involves two computational steps. For a fixed colatitude θ_i, the first step is to compute intermediate vectors $\hat{\mathbf{u}}(\theta_i)$ and $\hat{\mathbf{v}}(\theta_i)$ over l for $0 \le m \le L$:

$$\hat{\mathbf{u}}(\theta_i) = \mathbf{P}_i(m)\mathbf{a}_m^c \tag{A15}$$

$$\hat{\mathbf{v}}(\theta_i) = \mathbf{P}_i(m)\mathbf{a}_m^s \tag{A16}$$

In the second step, the $2L$ longitudinal values are computed by Fast Fourier Transform for the following discrete series:

$$S(\theta_i,\phi_j) = \sum_{m=0}^{L}\hat{U}(\theta_i,m)\cos m\phi_j + \hat{V}(\theta_i,m)\sin m\phi_j;\quad j = 0, \ldots, 2L-1 \tag{A17}$$

The overall computational scaling is also $O(L^3)$, which can be distributed to L processors as the two computational steps are fully independent. The use of parities in the construction of the associated Legendre polynomials reduces computation by a factor of 2.[13]

Acknowledgment. This work is partially supported by the National Institutes of Health, Grant GM46736.

Supporting Information Available: Projected structural fluctuations on to the first 20 lowest frequency modes and a movie animating the trajectory of the dynamic fluctuation of the ACG protein OMPDC represented by using spherical harmonic basis functions. This material is available free of charges via the Internet at http://pubs.acs.org.

References

(1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646.

(2) Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. *Structure* **2006**, *14*, 437.

(3) Dror, R. O.; Jensen, M. O.; Borhani, D. W.; Shaw, D. E. *J. Gen. Physiol.* **2010**, *135*, 555.

(4) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 1526.

(5) Dunning, T. H.; Schulten, K.; Tromp, J.; Ostriker, J. P.; Droegemeier, K.; Xue, M.; Fussell, P. *Comput. Sci. Eng.* **2009**, *11*, 28.

(6) Arkhipov, A.; Roos, W. H.; Wuite, G. J. L.; Schulten, K. *Biophys. J.* **2009**, *97*, 2061.

(7) Lu, L. Y.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 954.

(8) *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Ed.; Taylor & Francis Group: Boca Raton, FL, 2008.

(9) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.

(10) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(11) Jekeli, C.; Lee, J. K.; Kwon, J. H. *J. Geodesy* **2007**, *81*, 603.

(12) Wieczorek, M. A. *Treatise Geophys.* **2007**, *10*, 165.

(13) Colombo, O. Numerical Methods for Harmonic Analysis on the Sphere, Ph.D. Thesis, Ohio State University, 1981.

(14) Sneeuw, N. *Geophys. J. Int.* **1994**, *118*, 707.

(15) Max, N. L.; Getzoff, E. D. *IEEE Comput. Graph. Appl.* **1988**, *8*, 42.

(16) Duncan, B. S.; Olson, A. J. *Biopolymers* **1993**, *33*, 231.

(17) Duncan, B. S.; Olson, A. J. *Biopolymers* **1993**, *33*, 219.

(18) Duncan, B. S.; Olson, A. J. *J. Mol. Graph.* **1995**, *13*, 258.

(19) Sanner, M. F.; Olson, A. J.; Spehner, J.-C. *Biopolymers* **1996**, *38*, 305.

(20) Ritchie, D. W.; Kemp, G. J. L. *J. Comput. Chem.* **1999**, *20*, 383.

(21) Cai, W. S.; Shao, X. G.; Maigret, B. *J. Mol. Graph. Model.* **2002**, *20*, 313.

(22) Morris, R. J.; Najmanovich, R. J.; Kahraman, A.; Thornton, J. M. *Bioinformatics* **2005**, *21*, 2347.

(23) Mavridis, L.; Hudson, B. D.; Ritchie, D. W. *J. Chem. Inf. Model.* **2007**, *47*, 1787.

(24) Mak, L.; Grandison, S.; Morris, R. J. *J. Mol. Graph. Model.* **2008**, *26*, 1035.

(25) Venkatraman, V.; Sael, L.; Kihara, D. *Cell Biochem. Biophys.* **2009**, *54*, 23.

(26) Buchete, N. V.; Straub, J. E.; Thirumalai, D. *J. Mol. Graph. Model.* **2004**, *22*, 441.

(27) Duncan, B. S.; Olson, A. J. *J. Mol. Graph.* **1995**, *13*, 250.

(28) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4995.

(29) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct. Funct. Genet.* **1993**, *17*, 412.

(30) Brooks, B. R.; Janezic, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522.

(31) Janezic, D.; Brooks, B. R. *J. Comput. Chem.* **1995**, *16*, 1543.

(32) Janezic, D.; Venable, R. M.; Brooks, B. R. *J. Comput. Chem.* **1995**, *16*, 1554.

(33) Wu, N.; Mo, Y.; Gao, J.; Pai, E. F. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2017.

(34) Dey, T. K.; Giesen, J.; Goswami, S. *Algorithms Data Struct. Proc.* **2003**, *2748*, 25.

(35) Klees, R.; Tenzer, R.; Prutkin, I.; Wittwer, T. *J. Geodesy* **2008**, *82*, 457.

(36) Arkhipov, A.; Freddolino, P. L.; Imada, K.; Namba, K.; Schulten, K. *Biophys. J.* **2006**, *91*, 4589.

(37) Izvekov, S.; Swanson, J. M. J.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 4711.

(38) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(39) Connolly, M. L. *Science* **1983**, *221*, 709.

(40) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes*; University of Cambridge: New York, 1992.

(41) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450*, 913.

(42) Arkhipov, A.; Yin, Y.; Schulten, K. *Biophys. J.* **2008**, *95*, 2806.

(43) Zhang, Z.; Lu, L. J.; Noid, W. G.; Krishna, V.; Plaendtner, J.; Voth, G. A. *Biophys. J.* **2008**, *95*, 5073.

(44) Zhang, Z.; Plaendtner, J.; Grafmuller, A.; Voth, G. A. *Biophys. J.* **2009**, *97*, 2327.

(45) Stroud, A. H.; Secrest, D. *Gaussian Quadrature Formulas*; Prentice-Hall Inc.: Englewood Cliffs, NJ, 1966.

(46) Sloan, I. H.; Womersley, R. S. *Adv. Comput. Math.* **2004**, *21*, 107.

(47) Mohlenkamp, M. J. *J. Fourier Anal. Appl.* **1999**, *5*, 159.

(48) Luque, F. J.; Bachs, M.; Orozco, M. *J. Comput. Chem.* **1994**, *15*, 847.

(49) Fischer, H.; Polikarpov, I.; Craievich, A. F. *Protein Sci.* **2004**, *13*, 2825.

(50) White, E. T.; Tan, W. H.; Ang, J. M.; Tait, S.; Lister, J. D. *Powder Technol.* **2007**, *179*, 55.

(51) Blais, J. A. R.; Provins, D. A.; Soofi, M. A. *J. Supercomput.* **2006**, *38*, 173.

(52) Balis, J. A. R. In *Computational Science - ICCS 2008 Part II*; Bubak, M., van Albada, G. D., Dongarra, J., Eds.; Springer-Verlag: Berlin, 2008; Vol. LNCS 5102, p 638.

(53) Jekeli, C. *J. Geodesy* **1996**, *70*, 214.

(54) Mohlenkamp, M. J., A Fast Transform for Spherical Harmonics. Ph.D. Thesis, Yale University, 1997.

(55) Driscoll, J. R.; Healy, D. M. *Adv. Appl. Math.* **1994**, *15*, 202.

CT100426M