Exploring the Suitability of Coarse-Grained Techniques for the Representation of Protein Dynamics

Agustí Emperador,* Oliver Carrillo,*† Manuel Rueda,* and Modesto Orozco*†‡

*Molecular Modeling and Bioinformatics Unit, Joint Research Program in Computational Biology, Institute for Research in Biomedicine, Barcelona 08028, Spain, and Barcelona Supercomputing Center, Barcelona 08034, Spain; †Department de Bioquímica i Biologia Molecular, Facultat de Biologia, University of Barcelona, Barcelona 08028, Spain; and ‡National Institute of Bioinformatics, Parc Científic de Barcelona, Barcelona 08028, Spain

ABSTRACT A systematic study of two coarse-grained techniques for the description of protein dynamics is presented. The two techniques exploit either Brownian or discrete molecular dynamics algorithms applied in the context of simple C_α-C_α potentials, like those used in coarse-grained normal mode analysis. Coarse-grained simulations of the flexibility of protein metafolds are compared to those computed with fully atomistic molecular dynamics simulations using state-of-the-art physical potentials and explicit solvent. Both coarse-grained models efficiently capture critical features of the protein dynamics.

INTRODUCTION

A relationship between protein structure and function was first envisaged in the 1960s. However, only recently has a link between flexibility and function been uncovered, and several lines of evidence indicate that proteins have evolved not only to have certain structures but also to display dynamical properties that favor the specific conformational transitions required for their biological action (1–9).

The experimental description of protein flexibility is a challenging task (10) and has motivated the simultaneous exploitation of theoretical models (11,12). Atomistic molecular dynamics (MD) simulation is a very powerful approach insofar as it represents protein dynamics in physiological environments using physical potentials (13–21), the development of which proceeds from the rigorous formalism of molecular physics. Unfortunately, despite ongoing methodological advances, fully atomistic MD still demands very significant computational resources as well as substantial user expertise. As a result, for many proteins, trajectories of longer than 10 ns remain impractical. This significantly limits the applicability of MD to the study of protein flexibility in organelle or cellular environments, as not only intramolecular protein dynamics but also intermolecular dynamics need to be considered. Although the former take place on a timescale of nano- to microseconds, the latter may have timescales ranging from milliseconds to seconds, which vastly exceeds the limits of current atomistic MD. As a result, only coarse-grained models can currently provide a practical framework for such cellular or organelle simulations.

In a previous study (12), we showed that normal model analysis based on a $C_\alpha - C_\alpha$ quasiharmonic potential (22) provides a reasonable description of the equilibrium deformability pattern predicted by atomistic MD simulations. Here we examine the suitability of this same coarse-grained potential in the context of two inexpensive dynamic algorithms, namely, Brownian molecular dynamics (BD (23)) and discrete molecular dynamics (DMD (24)). The former relies directly on the basic algorithms of MD but also introduces strong simplifications in the protein and its environment, thereby reducing the cost of each integration step (at the possible expense of accuracy). DMD, by contrast, derives from the use of ballistic equations of motion, with the dynamic behavior of residues simplified to an interaction of elastic beads following square-well potentials. The simplicity of the potential function removes the need for numerical integration of the equations of motions. These protocols permit the study of large conformational movements that occur over long timescales and originate from equilibrium geometries.

The gain in speed obtained by using coarse-grained DMD or BD is clear; however, the level of accuracy achieved for the study of typical proteins is not as evident. Here we present, for the first time to our knowledge, a broad analysis of these two coarse-grained simulation techniques using the largest database of fully atomistic protein trajectories available ((11); http://mmb.pcb.ub.es/MODEL). Our analysis provides a thorough benchmark of BD and discrete MD and illustrates their respective strengths and weaknesses. The potential utility of these two techniques for the study of protein dynamics in crowded environments like cellular organelles is discussed.

METHODS

We compared three algorithms: i), reference MD with a fully atomistic force field and explicit solvent; ii), BD with a pseudoharmonic potential linking C_α atoms; and iii), DMD with C_α-C_α interactions represented by discontinuous square wells.

Submitted September 17, 2007, and accepted for publication February 29, 2008.

Agustí Emperador, Oliver Carrillo, and Manuel Rueda contributed equally to this work.

Address reprint requests to Modesto Orozco, E-mail: modesto@mmb.pcb.ub.es.

Editor: Klaus Schulten.

0006-3495/08/09/2127/12 $2.00

doi: 10.1529/biophysj.107.119115

Molecular dynamics

Trajectories of all protein metafolds (see Table 1 in the Supplementary Material, Data S1) taken from our $\mu-$MODEL database (http://mmb.pcb.ub.es/MODEL) were considered. All protein structures were titrated, neutralized by ions, minimized, hydrated, heated, and equilibrated (for at least 0.5 ns) using an established protocol (11). Trajectories were collected using three all-atom force fields (AMBER (25), CHARMM (26,27), and OPLS/AA (28–31)), which have potential functions such as

$$E = E_{\mathrm{bonded}} + E_{\mathrm{nonbonded}}, \tag{1}$$

where for bonded interactions,

$$E_{\mathrm{bonded}} = \sum_{\mathrm{bonds}} K_{\mathrm{s}}(l - l_0)^2 + \sum_{\mathrm{angles}} K_{\mathrm{b}}(\theta - \theta_0)^2 + \sum_{\mathrm{torsions}} \sum_{\mathrm{i}=1}^{3} \frac{V_{\mathrm{i}}}{2}(1 + \cos(i\phi - \xi)), \tag{2}$$

where l and θ refer to bond lengths and angles, respectively (with subscript 0 for parametric equilibrium values), K_{s} and K_{b} are the associated force constants, ϕ is a torsion angle, the set of V_{i} are the amplitudes associated with the Fourier terms used to represent torsional potentials, and ξ is a phase angle. For nonbonded interactions

$$E_{\mathrm{nonbonded}} = \sum_{\mathrm{a,b}} \frac{Q_{\mathrm{a}} Q_{\mathrm{b}}}{r_{\mathrm{ab}}} + \sum_{\mathrm{a,b}} \left[\left(\frac{C_{\mathrm{ab}}}{r_{\mathrm{ab}}} \right)^{12} - \left(\frac{D_{\mathrm{ab}}}{r_{\mathrm{ab}}} \right)^{6} \right], \tag{3}$$

where Q is a partial atomic charge, C and D denote diatomic van der Waals parameters, and r_{ab} is the interatomic distance.

The particle mesh Ewald approach was used to address long-range nonbonded interactions (32). Integration of the equations of motion proceeded with a time step of 1 fs; vibrations of bonds involving hydrogen atoms were removed by the SHAKE/RATTLE algorithm (33,34). Production runs were obtained using AMBER8 (35) and NAMD2.6 (36,37) and were extended for at least 10 ns with each force field. Jorgensen's TIP3P model (38,39) was used to represent aqueous solvent. As described elsewhere (11,12), trajectories obtained with different force fields give similar results for flexibility and as such can be combined in a metatrajectory, which should provide an improved description of the equilibrium dynamics of the protein. All comparisons were subsequently made to this ''metatrajectory'' as reference.

Brownian dynamics

In BD, the protein is in a stochastic bath that maintains constant temperature and modulates the otherwise extreme oscillations of the residues (40,41). The bath is simulated with two terms accounting for velocity-dependent friction and stochastic forces caused by the solvent environment such that

$$m \dot{\vec{v}}_{\mathrm{i}} = -\gamma \vec{v}_{\mathrm{i}} + \vec{F}_{\mathrm{i}} + \eta_{\mathrm{i}}, \tag{4}$$

where m is the effective mass of C_{α} (see below), $\vec{v}$ and $\dot{\vec{v}}$ are velocity and acceleration, respectively, $\vec{F}$ represents the force, γ is the inverse of a characteristic time, over which the particle loses its energy in a given solvent, and the stochastic term $\eta(t)$ is considered Gaussian white noise with autocorrelation given by

$$\langle \eta_{\mathrm{l}}(t)\, \eta_{\mathrm{n}}(t') \rangle = 2\, m\, k_{\mathrm{B}} T\, \gamma \delta_{\mathrm{ln}} \delta(t - t'), \tag{5}$$

where k_{B} is Boltzmann's constant and T is the temperature of the stochastic bath. The Dirac functions δ_{ln} and $\delta(t - t')$ ensure the independence of the components of the noise vector.

The BD equation of motion (Eq. 4) was integrated using Verlet's algorithm (15), which gives for the velocities (Eq. 6) and positions (Eq. 7) after time Δt

$$\vec{v}_{\mathrm{i}} = e^{-\Delta \mathrm{t}/\tau} \vec{v}_{\mathrm{i}}^{\,0} + \frac{1}{\gamma}(1 - e^{-\Delta \mathrm{t}/\tau}) \vec{F}_{\mathrm{i}}^{\,0} + \Delta \vec{v}_{\mathrm{i}}^{\;\mathrm{G}} \tag{6}$$

and

$$\vec{r}_{\mathrm{i}} = \vec{r}_{\mathrm{i}}^{\,0} + \tau(1 - e^{-\Delta \mathrm{t}/\tau})\, \vec{v}_{\mathrm{i}}^{\,0} + \frac{\Delta t}{\gamma}\left(1 - \frac{\tau}{\Delta t}(1 - e^{-\Delta \mathrm{t}/\tau})\right) \vec{F}_{\mathrm{i}} + \Delta \vec{\mathrm{r}}_{\mathrm{i}}^{\,\mathrm{G}}, \tag{7}$$

where $\tau = \mathrm{m}\gamma^{-1}$ is the characteristic time and $\Delta \vec{r}_{\mathrm{i}}^{\,\mathrm{G}}$ and $\Delta \vec{v}_{\mathrm{i}}^{\,\mathrm{G}}$ are the changes in position and velocity, respectively, induced by the stochastic term (see the Appendix for a full derivation of these equations).

The potential used for the computation of forces in Eq. 4 uses a coarse-grained representation of the protein (C_{α} only) and a quasiharmonic representation of intersite interactions

$$U_{\mathrm{ij}} = \frac{1}{2} C \left(\frac{r^{*}}{|\mathbf{r}_{\mathrm{ij}}^{0}|} \right)^{6} (\vec{r}_{\mathrm{ij}} - \vec{r}_{\mathrm{ij}}^{\,0})^{2}, \tag{8}$$

where $\vec{r}_{\mathrm{ij}} = \vec{r}_{\mathrm{i}} - \vec{r}_{\mathrm{j}}$ is the vector connecting C_{α} atoms i and j and C and $r*$ are constants to be chosen. This is similar to the potential proposed by Kovacs et al. (43).

The initial condition is a native structure (in this case the fully atomistic MD-averaged conformation), which is assumed to be in the minimal energy state and from which the relative vectors $\vec{r}_{\mathrm{ij}}^{\,0}$ are computed. After several trials, the factor C was taken to be 40 kcal/mol-Å^2 and $r*$, which is the mean distance between two consecutive C_{α} atoms, was set to 3.8 Å (43). The mass of all C_{α} atoms was set to 100 D (i.e., that of an average residue). The velocity-dependent friction γ was considered to have the same value as that for pure water (i.e., 0.4 ps^{-1}). BD simulation timescales were equivalent to those considered in MD.

Discrete molecular dynamics

In this approach (24,44–51) the proteins were modeled as a system of beads (C_{α} atoms) interacting through a discontinuous potential (square wells in our study). Outside the discontinuities, potentials were considered constant, thereby implying a ballistic regime for the particles (constant potential, constant velocity) in all conditions, except at such time as when the particles reach a potential discontinuity (this is called ''an event'' or ''a collision''). At this time, the velocities of the colliding particles are modified by imposing conservation of the linear momentum, angular momentum, and total energy. In our study, since the particles were constrained to move within a configurational space where the potential energy is constant (infinite square wells), the kinetic energy remained unchanged and therefore all collisions were assumed to be elastic.

DMD has a major advantage over techniques like MD or BD because, as it does not require the integration of the equations of motion at fixed time steps, the calculation progresses from event to event. In practice, the time between events decreases with temperature and density and depends on the number of particles $N \sim$ as $N^{-1/2}$. The equations of motion, corresponding to constant velocity, are solved analytically

$$\vec{r}_{\mathrm{i}}(t + t_{\mathrm{c}}) = \vec{r}_{\mathrm{i}}(t) + \vec{v}_{\mathrm{i}}(t) t_{\mathrm{c}}, \tag{9}$$

where t_{c} is the minimum among the collision times t_{ij} between each pair of particles i and j, given by

$$t_{\mathrm{ij}} = \frac{-b_{\mathrm{ij}} \pm \sqrt{b_{\mathrm{ij}}^{2} - v_{\mathrm{ij}}^{2}(r_{\mathrm{ij}}^{2} - d^{2})}}{v_{\mathrm{ij}}^{2}}, \tag{10}$$

where r_{ij} is the square modulus of $\vec{r}_{\mathrm{ij}} = \vec{r}_{\mathrm{j}} - \vec{r}_{\mathrm{i}}$, v_{ij} is the square modulus of $\vec{v}_{\mathrm{ij}} = \vec{v}_{\mathrm{j}} - \vec{v}_{\mathrm{i}}$, $b_{\mathrm{ij}} = \vec{r}_{\mathrm{ij}} \cdot \vec{v}_{\mathrm{ij}}$, and d is the distance corresponding to a discontinuity in the potential (the signs + and − before the radical are used for particles approaching one another and moving apart, respectively).

As the integration of Newton's equations is no longer the rate-limiting step, the use of efficient algorithms for predicting collisions (52) allows the extension of calculations for very long simulation periods and large systems (44,53,54).

The collision between particles i and j is associated with a transfer of linear momentum in the direction of the vector $\vec{r}_{\mathrm{ij}}$. Thus,

$$m_{\mathrm{i}}\vec{v}_{\mathrm{i}} = m_{\mathrm{i}}\vec{v}_{\mathrm{i}}' + \Delta\vec{p} \tag{11}$$

$$m_{\mathrm{j}}\vec{v}_{\mathrm{j}} + \Delta\vec{p} = m_{\mathrm{j}}\vec{v}_{\mathrm{j}}', \tag{12}$$

where the prime indexes variables after the event.

To calculate the change in velocities, the velocity of each particle is projected in the direction of the vector $\vec{r}_{\mathrm{ij}}$ so that the conservation equations become one-dimensional along the interatomic coordinate

$$m_{\mathrm{i}}v_{\mathrm{i}} = m_{\mathrm{i}}v_{\mathrm{i}}' + \Delta p \tag{13}$$

$$m_{\mathrm{j}}v_{\mathrm{j}} + \Delta p = m_{\mathrm{j}}v_{\mathrm{j}}', \tag{14}$$

which implies

$$m_{\mathrm{i}}v_{\mathrm{i}} + m_{\mathrm{j}}v_{\mathrm{j}} = m_{\mathrm{i}}v_{\mathrm{i}}' + m_{\mathrm{j}}v_{\mathrm{j}}' \tag{15}$$

$$\frac{1}{2}m_{\mathrm{i}}v_{\mathrm{i}}^2 + \frac{1}{2}m_{\mathrm{j}}v_{\mathrm{j}}^2 = \frac{1}{2}m_{\mathrm{i}}v_{\mathrm{i}}'^2 + \frac{1}{2}m_{\mathrm{j}}v_{\mathrm{j}}'^2. \tag{16}$$

From Eqs.13–16 the transferred momentum is readily determined as

$$\Delta p = \frac{2m_{\mathrm{i}}m_{\mathrm{j}}}{m_{\mathrm{i}} + m_{\mathrm{j}}}(v_{\mathrm{i}} - v_{\mathrm{j}}), \tag{17}$$

and the final velocities of particles i and j are determined through Eqs. 11 and 12.

The interaction potentials were defined as infinite square wells, such that the particle-particle distances varied between $d_1 = (1-\sigma)R_0$ and $d_2 = (1+\sigma)R_0$, R_0 being the distance in the native conformation and 2σ the width of the square well. As in BD, the MD-averaged conformation was taken as the native conformation. Residue-residue interaction potentials were defined only for the particles at a distance smaller than a cutoff radius R_{c} in the native conformation. For nonconsecutive C_α particles, $R_{\mathrm{c}} = 8$ Å and $\sigma = 0.1$ were used, whereas for consecutive pairs of residues a smaller well width ($\sigma = 0.05$) was chosen to keep the C_α-C_α distances closer to the expected value: 3.8 Å. This definition of the potential gave a good reproduction of the shape of the wells obtained by the distance-dependent pseudoharmonic model of Kovacs et al. (43) and in a preevaluation of the method was found to reproduce the essential dynamics of six test proteins well (representative of small, medium, and large macromolecules). As in BD, we used an average mass of 100 D for the beads. Our DMD simulations were performed for a simulation period equivalent to those considered in MD.

Essential dynamics

To facilitate comparison between MD and the two coarse-grained methods, we used essential dynamics (55) to extract the essential deformation patterns of the proteins from MD, BD, and DMD simulations. Accordingly, the Cartesian covariance matrix collected from the trajectories (only C_α atoms were considered) was diagonalized, thereby yielding a set of eigenvalues (λ_{i}) and corresponding eigenvectors ($\boldsymbol{\nu}_{\mathrm{i}}$). Note that the eigenvalues appear in units of length squared, but they can easily be transformed into energy units according to

$$k_{\mathrm{l}} = \frac{k_{\mathrm{b}}T}{\lambda_{\mathrm{l}}}. \tag{18}$$

The diagonalization of the mass-weighted covariance matrix yields frequencies which can be manipulated using pseudoharmonic models to derive entropies (56,57). Here we report only results obtained from the Andricioaei-Karplus method; however, very similar values can be obtained using Schlitter's approach. In all cases, entropies were calculated only over C_α atoms, which all have a mass of 100 D.

Statistical descriptors for comparison

Several complementary features have been considered to quantify the similarity between the samplings obtained with MD, BD, or DMD simulations:

1. Global deformability was measured by analyzing the i), total variance, ii), entropy, iii), strength of the softer deformation modes (Eq. 18), iv), dimensionality (58), and v), number of modes required to account for 90% of the trajectory variance.
2. Deformation space overlap was determined using Hess's metrics (59–61):

$$\gamma_{\mathrm{XY}} = \frac{1}{m}\sum_{\mathrm{i=1}}^{\mathrm{m}}\sum_{\mathrm{j=1}}^{\mathrm{m}}(\boldsymbol{\nu}_{\mathrm{i}}^{\mathrm{X}} \cdot \boldsymbol{\nu}_{\mathrm{j}}^{\mathrm{Y}})^2, \tag{19}$$

where X and Y index the two methods, i and j index the eigenvectors (ranked on the basis of their contribution to structural variance), and m is the number of eigenvectors in the ''important space''. The significance of a given γ value can be quantified by an associated Z-score:

$$Z_{\mathrm{score}} = \frac{(\gamma_{\mathrm{XY}}(observed)) - (\gamma_{\mathrm{XY}}(random))}{std(\gamma_{\mathrm{XY}}(random))}. \tag{20}$$

Fully random models were obtained by diagonalization of a pseudo-covariance matrix obtained from random permutation of the C_αs for each snapshot; the standard deviation in this quantity was obtained by considering 500 pseudocovariance matrices. Additional random models were built in such a way that the chemical connectivity was maintained and steric collapses were avoided. For this purpose, we performed several 10-ns DMD simulations for the diverse proteins using a simplified force field defined by covalent C_α-C_α contacts plus a hard sphere

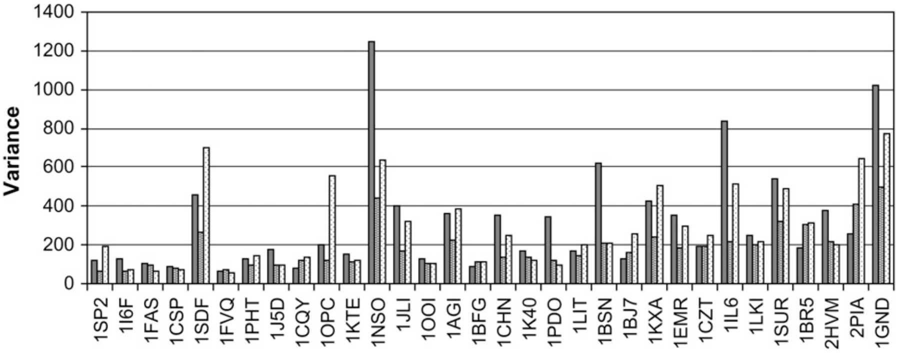

FIGURE 1 Total variances (in Å^2) computed for the set of proteins using MD (*dark gray*), BD (*gray*), and DMD (*light gray*).

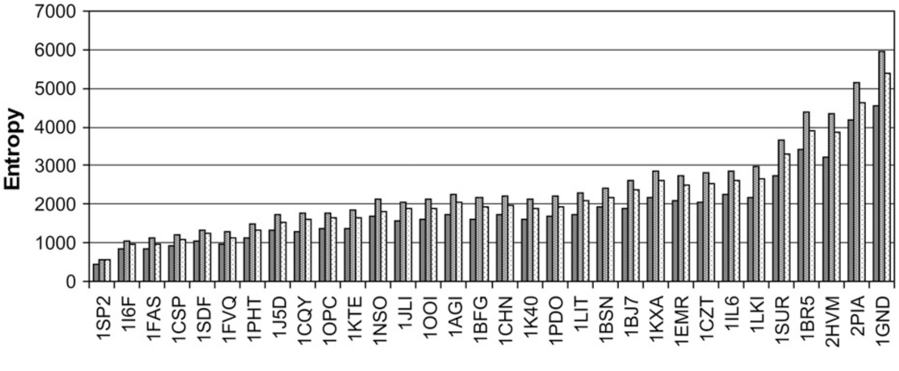

FIGURE 2 Entropies (in cal/molK) associated with the samplings obtained using MD (*dark gray*), BD (*gray*), and DMD (*light gray*).

potential for each residue. Essential dynamics from these trajectories provided sets of ''pseudorandom'' eigenvectors, which, although random, were consistent with the physical structure of the proteins. The set of *Z*-scores obtained using these ''pseudorandom'' eigenvectors were labeled with ''*'' to avoid confusion with standard *Z*-score measures.

To determine the pair correspondence between eigenvectors obtained from MD, BD, and DMD, we computed the difference in rank between the eigenvectors showing the largest overlap and also the eigenvector ''spread function'' (62):

$$s_i = \sqrt{\sum_{j=1}^{m} j^2 \eta_{ij}^2 - \left(\sum_{j=1}^{m} j \eta_{ij}^2\right)^2}, \qquad (21)$$

which indicates the number of MD modes in which BD/DMD eigenvectors are distributed. In Eq. 21, $\eta_{ij} = \boldsymbol{v}_i^X \cdot \boldsymbol{v}_j^Y$ and m is the number of degrees of freedom. Overlaps are scaled to ensure that $\sum_j \eta_{ij}^2 = 1$. Note that for two identical sets of modes, η_{ij}^2 is nonzero only for $i = j$ and the spread becomes 0.

3. Relative distribution of deformational pattern. A direct comparison of trajectories was made using the α and Ω indices defined by

$$\alpha_{AB} = \frac{1}{M_A M_B} \sum_{k=1}^{M_A} \sum_{k=1}^{M_B} \left(\frac{1}{N} \sum_{l=1}^{3N} (x_{Al} - x_{Bl})^2 \right)^{1/2}, \qquad (22)$$

where N is the number of atoms and M is the number of frames, and

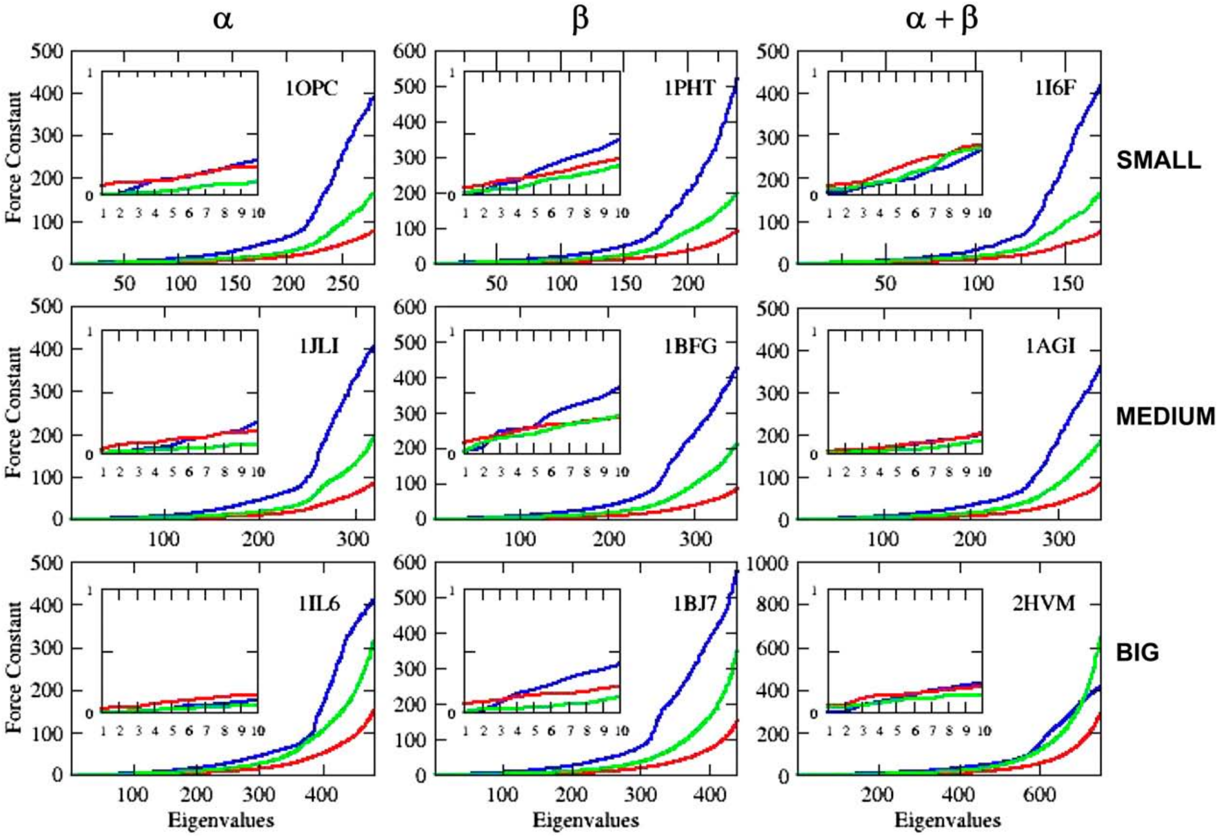

FIGURE 3 Harmonic force constants (kcal/mol Å^2) associated with the deformation along eigenvectors derived from MD (*blue*), BD (*red*), and DMD (*green*) for representative proteins. The inset corresponds to values obtained for the first 10 eigenvectors.

$$\Omega_{AB} = \frac{\alpha_{AA} + \alpha_{BB}}{2\alpha_{AB}}. \tag{23}$$

More regional or atom-based measures were obtained by analyzing the (C_α) B-factors and Lindemann's index (11,63):

$$\Delta_L = \frac{\left(\sum_i \langle \Delta r_i^2 \rangle / N\right)^{1/2}}{a'}, \tag{24}$$

where a' is the most probable nonbonded near-neighbor distance, N is the number of atoms, and $\langle \Delta r^2 \rangle$ is the mean-square displacement of an atom from its equilibrium position.

The benchmark

Thirty-two proteins representative of all protein metafolds were selected, as described elsewhere (see Table S1, Data S1) (11,64). This database contained highly representative proteins with distinct folds, amino acid compositions, secondary structure, topology, and stability. Movies for trajectories can be found at http://mmb.pcb.ub.es/CG/.

RESULTS AND DISCUSSION

Global stiffness

In general, MD samplings explored the largest conformational space, as reflected in the total variance (see Fig. 1). BD led to the most rigid representation, whereas DMD displayed a variance, which was, on average, ~15% lower than that predicted by MD. Most of the difference between DMD and MD was attributable to a few proteins that displayed large deformations (such as 1NSO, 1IL6, and 1BSN), which are not well reproduced by quasiharmonic techniques. Interestingly and perhaps surprisingly, the largest variances predicted by MD did not necessarily correlate with larger entropies, and indeed coarse-grained models generally produced slightly larger entropies than MD (Fig. 2). This apparent paradox is resolved by considering the different dependence of variance (linear) and entropies (logarithmic) with the eigenvectors of the mass-weighted covariance matrix (see next paragraph and Fig. 3). Overall, the results strongly suggest i), that the distribution of eigenvalues in MD and the two coarse-grained methods differs (see below); and ii), that general concepts like ''flexibility'' must be defined in a precise way, as they depend on the physical property used for measurement purposes.

The stiffness profile (i.e., the force-constant associated with each eigenvector) shows that MD simulations were dominated by a few very soft modes, whereas DMD, and especially BD, distributed the flexibility in a larger number of deformation modes (see Fig. 3), thereby depicting a more complex scenario where the partition between high and low relevance modes is not as clear as in MD. This was confirmed by the analysis of dimensionality (i.e., the number of vectors along which the protein is displaced by at least 1 Å at room temperature) and by the number of vectors required to explain 90% of the variance (see Fig. 4). It is worth noting that the intrinsic difference in the distribution of protein variance along modes between coarse-grained methods and MD did not depend on the force constant (C in Eq. 8) or the well width used in BD or DMD but depended mainly on the intrinsic nature of the three techniques and on the use of universal stiffness parameters (force constants or well dimensions) to describe all C_α-C_α interactions in these calculations. Finally, a detailed analysis of the behavior of the proteins indicated

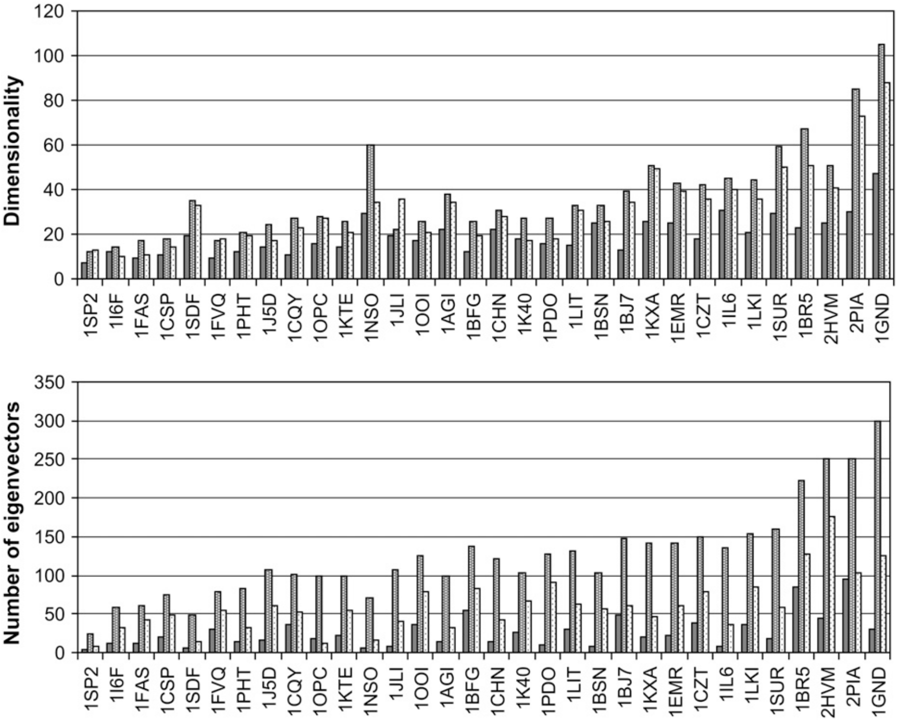

FIGURE 4 Dimensionality (*top*) and number (*bottom*) of essential modes required to explain 90% of the variance for the set of proteins using MD (*dark gray*), BD (*gray*), and DMD (*light gray*).

that the capacity of DMD and BD to reproduce MD trajectories does not depend on the CATH family, on the size, or on the presence/absence of saline bridges or disulfide bridges.

Analysis of the deformation pattern

No general rank-pair correspondence between the eigenvectors obtained from MD, BD, and DMD simulations was observed (Fig. 5). Furthermore, each DMD/BD eigenvector was distributed in a variety of MD modes, as shown by the spread plots (Fig. 6). However, when the analysis focused only on the first eigenvectors, a greater agreement was found between MD and the two coarse-grained techniques: i), the distance (in rank) between the best overlapped eigenvectors reached very small values, and ii), the spread took values close to zero (see Figs. 5 and 6). These observations suggest that although individual DMD- or BD-derived essential movements might not be accurate individually, when considered together (defining an essential deformation space), they provide information similar to that obtained by MD.

To further study the capacity of BD and DMD to reproduce (at equal simulation times) the dynamics of proteins compared to predictions from atomistic MD, we computed the similarity indices among BD, DMD, and MD (γ; Eq. 19) for the essential space required to reproduce 90% of (MD) variance (Fig. 7) and the associated *Z*-scores (Fig. 8). For the sake of completeness, this analysis was repeated using a constant number (50) of eigenvectors (12), which on average reproduced most of the (MD) protein variance (see also Fig. 4). Remarkable similarity was observed between MD and the simplified DMD ($\gamma = 0.51$) and BD ($\gamma = 0.55$) techniques (Fig. 7). This similarity increased (especially for small proteins) to an average value of 0.61 (DMD) and 0.66 (BD) when the analysis was limited to 50 eigenvectors.

The statistical significance of this similarity is supported by values of the *Z*-scores (typically in the range 50–250) for both definitions of the important space (see Fig. 8). When *Z*-scores* (see Methods) were considered, the statistical significance of the similarities was maintained (see Fig. S1 in Data S1), supporting the utility of the coarse-grained models. Finally, the level of similarity found between the MD meta-trajectories and the coarse-grained models was similar to that obtained when the three MD trajectories (AMBER, OPLS, and CHARMM) were compared to one another (see Figs. S2 and S3, Data S1) and to that observed when various 10-ns sections of a long trajectory were compared (see Table S2, Data S1).

The global similarity between MD and coarse-grained methods was also examined by comparing collected snapshots using similarity indices α and Ω (Eqs. 22 and 23).

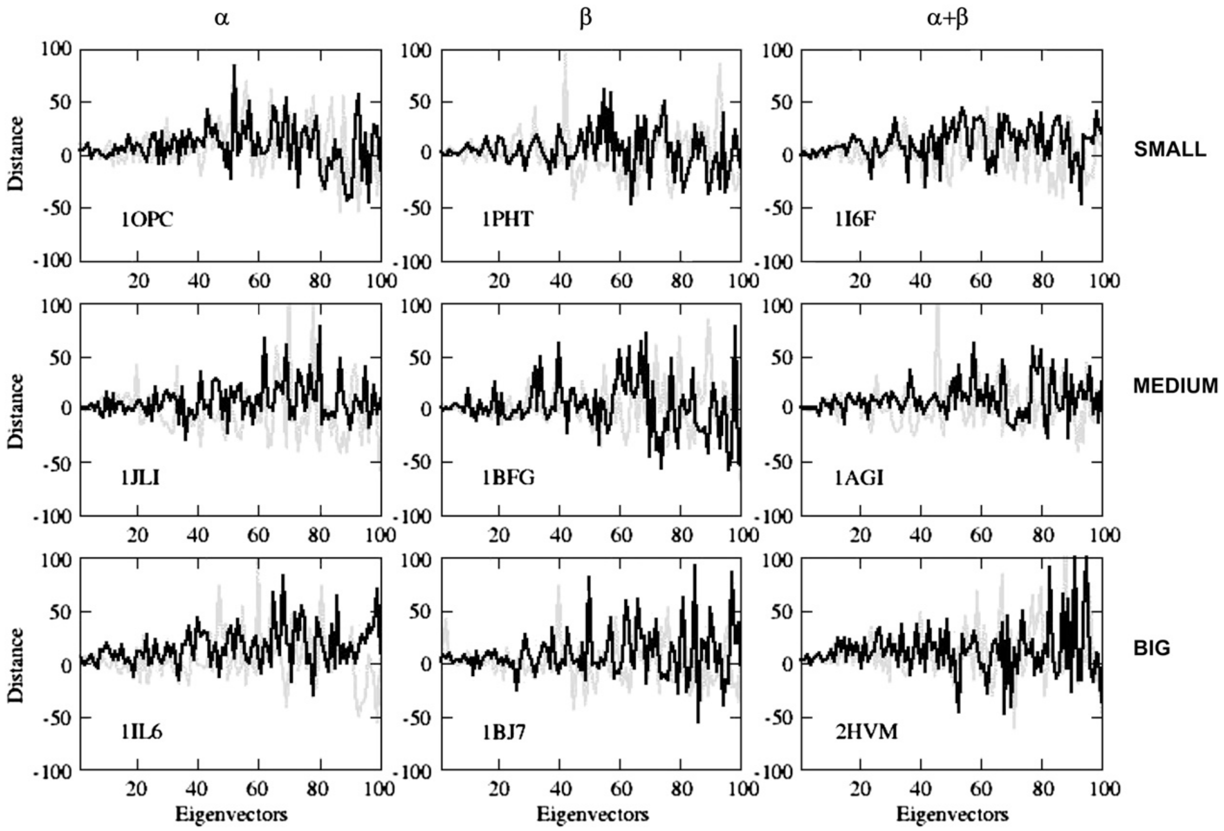

FIGURE 5 Rank distance between the DMD (*light gray*) or BD (*gray*) eigenvectors (*x* axis) and the MD eigenvectors showing the best overlap for representative proteins.

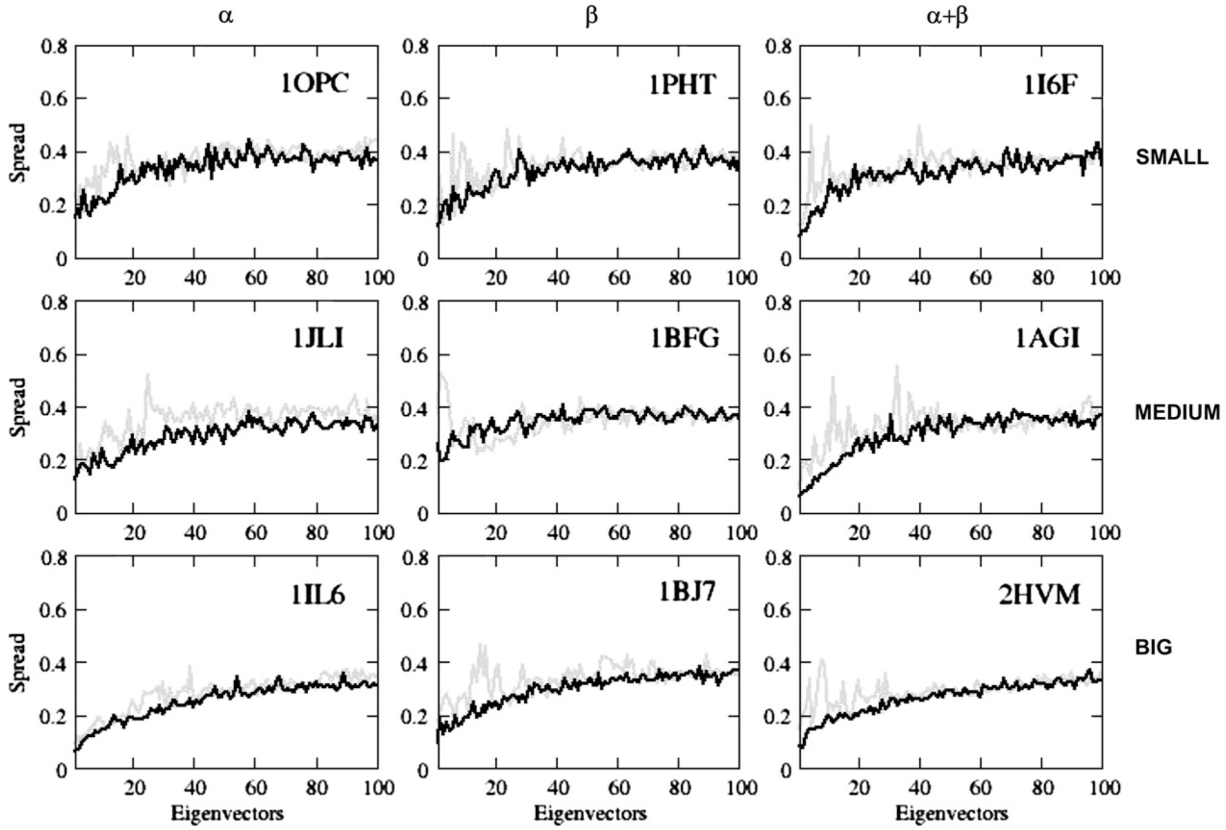

FIGURE 6 Normalized spread (Eq. 21) of DMD (*light gray*) and BD (*gray*) eigenvectors in MD essential space.

Absolute similarity indices α around 2 Å were obtained (Table 1), which are similar to those found when snapshots collected from the same trajectory were compared. This justifies the large relative similarity indices Ω (0.9) obtained between MD and coarse-grained trajectories (see individual data in Table S3, Data S1). Overall, these findings confirm that despite their simplicity, BD and DMD provide reasonable global pictures of the equilibrium dynamics of proteins.

As a final test, we analyzed how protein flexibility can be mapped into residues by computing the C_α B-factors from MD, DMD, and BD trajectories. The B-factor profiles showed good agreement between MD and the two coarse-grained methods (Fig. 9), even though some deviations were detected, typically in residues located at loops or regions without secondary structure, which are predicted to be more mobile in MD than in the two other methods. A similar type of information was obtained from Lindemann's index (see Methods), which measures macromolecular dynamics in terms of gas-like, liquid-like, or solid-like behavior. In agreement with our previous results (11), proteins behaved like liquids in the exterior ($\Delta_L > 0.3$) and like solids in the interior ($\Delta_L \sim 0.2$) by MD simulation (see Table 2; values for proteins are shown in Table S4, Data S1). The two coarse-grained methods were similar in character. However, given that these methods tend to reduce atomic oscillations, a small increase in the ''solid'' character of the proteins was found, this trend being more noticeable for BD. Interestingly, not only was the general Lindemann's distribution well reproduced by BD and DMD, but so were subtle details like the distinct values of Δ_L for different types of secondary structures (see Table 2). These observations suggest that DMD and BD also have the capacity to reproduce the dynamic properties of proteins and their residue anisotropy.

CONCLUSIONS

By performing a broad and systematic comparison of BD and DMD techniques to atomistic MD simulations, we were able to quantify, for the first time to our knowledge, the ability of these two coarse-grained simulation methods to describe the equilibrium dynamics of proteins. On the basis of our findings, we believe that our results will apply to the entire proteome. In particular, we found that methods based on ultrasimplified Hamiltonians (C_α-C_α quasiharmonic potentials or square wells) provide reasonable approximations in many cases to trajectories obtained from atomistic MD simulations with explicit solvent.

Care should be taken when applying BD and DMD if large, but local, nonharmonic deformations are accessible, in which case atomic detailed representation of the residues is required. Such representation will also be required when the analysis calls for a correct balance between low- and high-

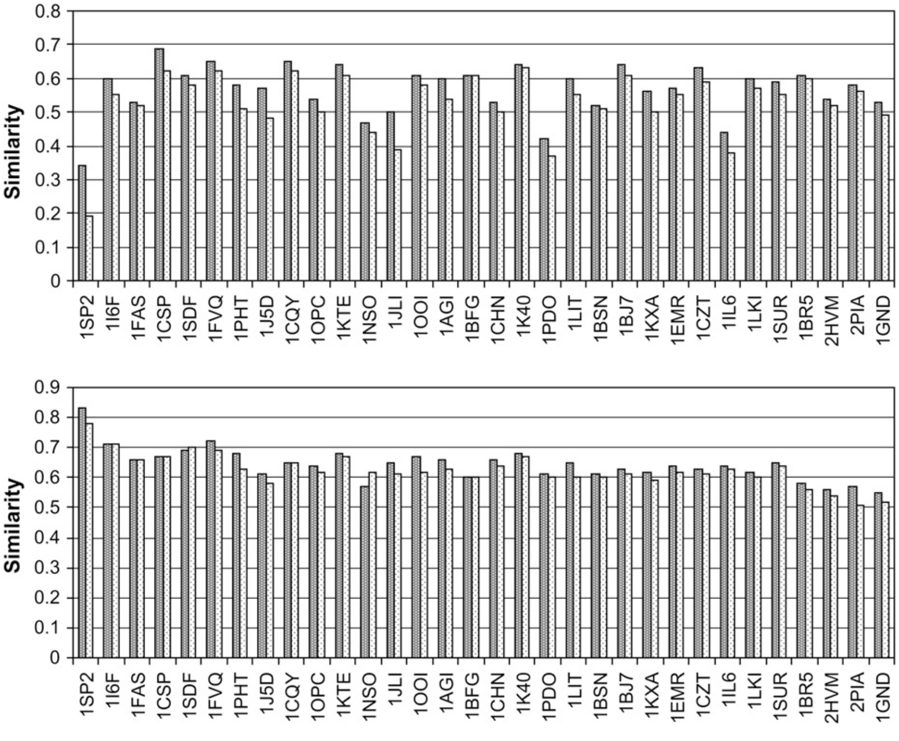

FIGURE 7 Similarity index (γ; Eq. 19) between MD and coarse-grained important spaces in DMD (*light gray*) and BD (*gray*) simulations for the set of proteins. The important space is defined for each protein as (*top*) the minimum number of eigenvectors required to explain 90% of variance, and (*bottom*) the first 50 eigenvectors were selected for all proteins.

frequency movements. In general, good results can be expected from BD and DMD calculations for large proteins, where the key movements are large domain-domain rearrangements or large loop movements. In contrast, local changes in small or medium proteins will require atomistic MD as will cases in which flexibility changes as a result of the presence of ligands, stress, or environmental variations.

Further improvements of the methods are expected to derive from the use of topology or residue-specific force constants, which should correct a tendency of current coarse-

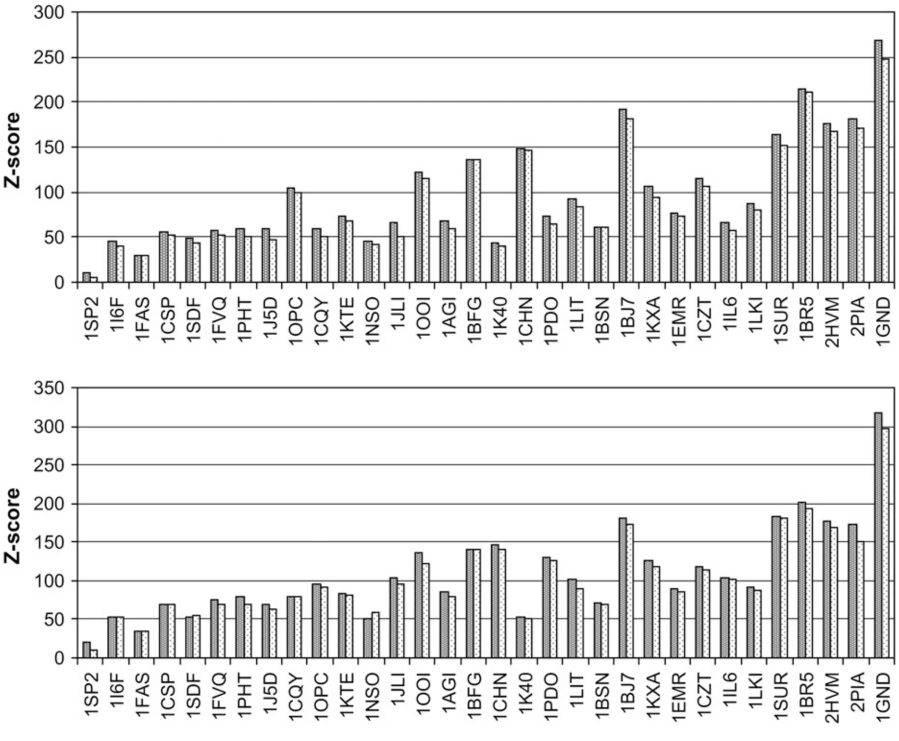

FIGURE 8 *Z*-scores (Eq. 20) associated with similarity indices (Eq. 19 and Fig. 7) between MD and coarse-grained models with DMD (*light gray*) and BD (*gray*). The important space is defined for each protein as (*top*) the minimum number of eigenvectors required to explain 90% of variance and (*bottom*) the first 50 eigenvectors.

TABLE 1 Similarity indices

$\langle\alpha\rangle$ /$\langle\Omega\rangle$	DMD	BD	MD
DMD	1.9/1.0	2.0/0.9	2.2/0.9
BD		1.6/1.0	2.0/0.9
MD			2.0/1.0

Values averaged for all the proteins in the data set.
Absolute (α, in Å; Eq. 22) and relative (Ω; Eq. 23) similarity indices among the three types of dynamics simulations.

grained models to spread the dynamics of proteins in a larger number of modes than those predicted from atomistic MD simulations or from the use of higher resolution protein models that incorporate additional physical interactions.

Overall, our results suggest that BD, and especially DMD, can be profitably employed in at least two major scenarios: i), to represent backbone flexibility in docking experiments where side chains are adjusted to better accommodate ligands; and ii), to represent, in a fast and efficient way, the intramolecular dynamics of proteins in organelle- or cellular-scale simulations, where thousands of proteins are free to move in crowded environments.

APPENDIX

We require a numerical algorithm to solve the following stochastic differential equation (note that for simplicity, the vector properties of positions, velocities, accelerations, and forces in this appendix are not explicitly stated in the equations):

$$m\,\dot{v}_i = -\gamma v_i + F_i + \eta_i. \tag{A1}$$

The first step is to divide both sides of Eq. A1 by the dissipative factor γ and rearrange to

$$\tau\,\dot{v}_i + v_i = \gamma^{-1}F_i + \gamma^{-1}\eta_i, \tag{A2}$$

where $\tau = m\gamma^{-1}$ is the characteristic time (see main text).
The first term of Eq. A2 can be written as

$$\tau\,\dot{v}_i + v_i = \tau\,e^{-t/\tau}\frac{d}{dt}(e^{t/\tau}\,v_i). \tag{A3}$$

Substituting Eq. A3 into Eq. A2 and after some manipulation, we obtain

$$\frac{d}{dt}(e^{t/\tau}\,v_i) = m^{-1}e^{t/\tau}F_i + m^{-1}e^{t/\tau}\eta_i. \tag{A4}$$

Integration of Eq. A4 leads to the following expression for the velocity at time step $t + \Delta t$:

$$v_i = e^{-\Delta t/\tau}\,v_i^0 + m^{-1}e^{-(t+\Delta t)/\tau}\int_t^{t+\Delta t}F_i e^{t'/\tau}dt' + m^{-1}e^{-(t+\Delta t)/\tau}\int_t^{t+\Delta t}\eta_i e^{t'/\tau}dt'. \tag{A5}$$

At this stage, to simplify integration we may assume that the time step is small enough that the force has a constant value F_i^0 during the integration. Note that we cannot take the same approach with the noise function, since it

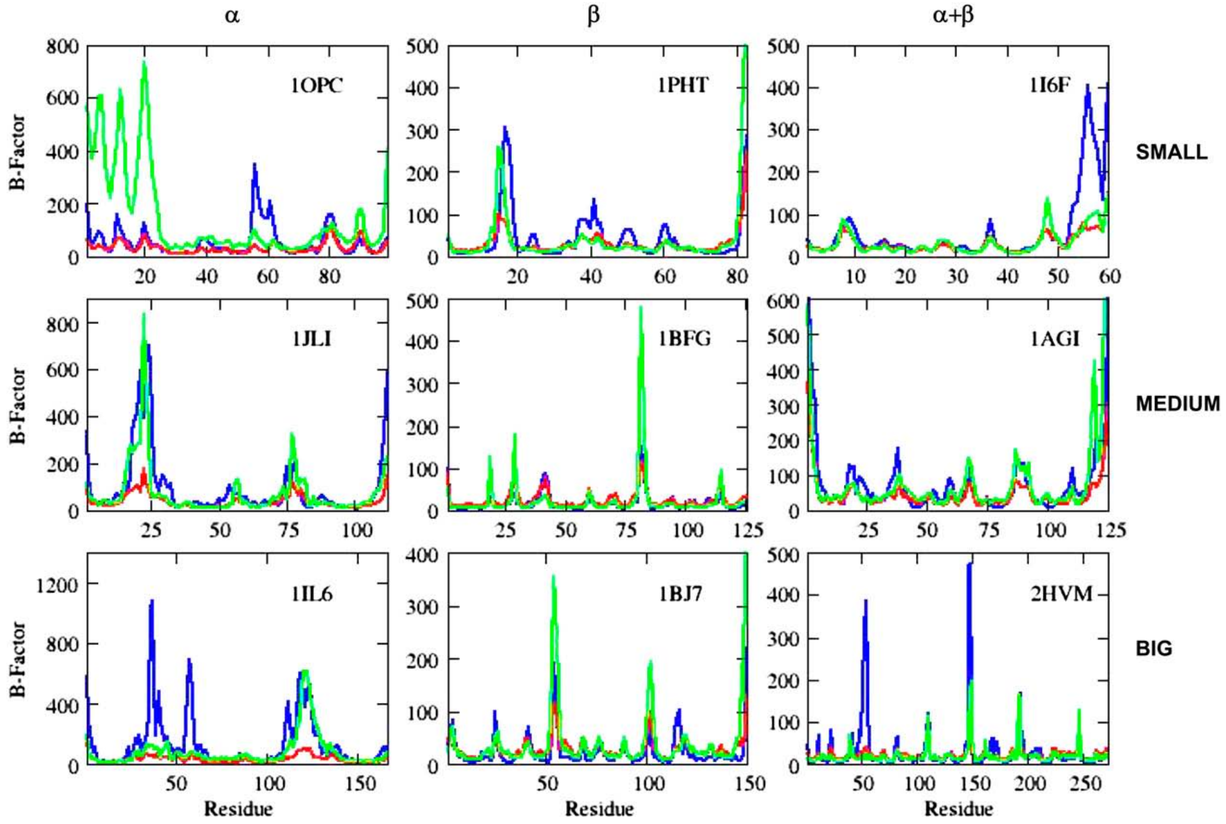

FIGURE 9 α-Carbons B-factors (in Å^2) computed from MD (*blue*), DMD (*green*), and BD (*red*) simulations for representative proteins.

TABLE 2 Lindemann's indices

	DMD	BD	MD
All	0.31	0.26	0.33
Buried	0.22	0.20	0.22
Not buried	0.35	0.28	0.37
α-Helix	0.24	0.23	0.27
β-Sheet	0.23	0.21	0.22
Turn	0.37	0.30	0.40

Values averaged for all the proteins in the data set. Lindemann's indices (considering only C_α; Eq. 24) obtained from DMD, BD, and MD, considering all the residues in the proteins or distinct groups of residues selected on the basis of accessibility or secondary structure.

is not a smooth and deterministic function as the force is. These considerations allow us to rewrite Eq. A5 as

$$v_{\rm i} = e^{-\Delta t/\tau} v_{\rm i}^0 + \frac{1}{\gamma}(1 - e^{-\Delta t/\tau}) F_{\rm i}^0 + \int_{\rm t}^{\rm t+\Delta \mathit{t}} m^{-1} e^{-(\rm t+\Delta t-t')/\tau} \eta_{\rm i} dt', \tag{A6}$$

where we can find the stochastic contribution to the new velocity through the integral

$$\Delta v_{\rm i}^{\rm G} = \int_{\rm t}^{\rm t+\Delta t} m^{-1} e^{-(\rm t+\Delta t-t')/\tau} \eta_{\rm i} dt'. \tag{A7}$$

Note that the updated position $r_{\rm i}$ can be obtained from Eq. A6 by integrating both sides with respect to time,

$$\begin{aligned} r_{\rm i} = r_{\rm i}^0 &+ \int_0^{\Delta \rm t} e^{-s/\tau} v_{\rm i}^0 ds + \frac{1}{\gamma} \int_0^{\Delta \rm t} (1 - e^{-s/\tau}) F_{\rm i}^0 ds \\ &+ \int_0^{\Delta \rm t} ds \int_{\rm t}^{\rm t+s} m^{-1} e^{-(\rm t+s-t')/\tau} \eta_{\rm i} dt', \end{aligned} \tag{A8}$$

where the integration of the stochastic term can be done by parts:

$$\begin{aligned} &\int_0^{\Delta \rm t} ds \int_{\rm t}^{\rm t+\mathit{s}} m^{-1} e^{-(\rm t+s-t')/\tau} \eta_{\rm i} dt' \\ &= \int_{\rm t}^{\rm t+\Delta t} \gamma^{-1} m^{-1} (1 - e^{-(\rm t+\Delta t-t')/\tau}) \eta_{\rm i} dt'. \end{aligned} \tag{A9}$$

Finally, the new position is obtained as

$$\begin{aligned} r_{\rm i} = r_{\rm i}^0 &+ \tau(1 - e^{-\Delta t/\tau}) v_{\rm i}^0 + \frac{\Delta t}{\gamma}\left(1 - \frac{\tau}{\Delta t}(1 - e^{-\Delta t/\tau})\right) F_{\rm i}^0 \\ &+ \int_{\rm t}^{\rm t+\Delta t} \gamma^{-1} m^{-1} (1 - e^{-(\rm t+\Delta t-t')/\tau}) \eta_{\rm i} dt'. \end{aligned} \tag{A10}$$

As before (Eq. A7), there is a stochastic contribution to the new position

$$\Delta r_{\rm i}^{\rm G} = \int_{\rm t}^{\rm t+\Delta t} \gamma^{-1} m^{-1} (1 - e^{-(\rm t+\Delta t-t')/\tau}) \eta_{\rm i} dt'. \tag{A11}$$

Note that these stochastic contributions to updated positions and velocities have the shape of the so-called stochastic integral (40,41):

$$\int_{\rm t}^{\rm t+\Delta \mathit{t}} G(t') dW(t'), \tag{A12}$$

where $dW(t)$ takes the place of $\eta_{\rm i}\, dt$ and corresponds to a differential Wiener process, describing a Brownian motion. This stochastic process has the property (40,41)

$$\langle \Delta W^2 \rangle = 2\, m\, k_{\rm B} T\, \gamma\, \Delta t, \tag{A13}$$

where $\Delta W = W(t+\Delta t) - W(t)$.
Assuming that Δ t is small, we can approximate Eq. A12 up to leading order in Δ W (40,41) as

$$\int_{\rm t}^{\rm t+\Delta t} G(t') dW(t') \approx G(t) \sqrt{2\, m\, k_{\rm B} T\, \gamma\, \Delta t}\, u(t), \tag{A14}$$

where $u(t)$ is a Wiener process of variance equal to 1.
Using the former equations, we can find expressions for the stochastic integrals Eqs. A7 and A11, viz.

$$\begin{aligned} \Delta v_{\rm i}^{\rm G} &\approx \sqrt{\frac{2\, k_{\rm B} T\, \gamma}{m}} e^{-\Delta t/\tau} \Delta t^{1/2} u_{\rm i}(t) \\ \Delta r_{\rm i}^{\rm G} &\approx \sqrt{\frac{2\, k_{\rm B} T}{\gamma\, m}} (1 - e^{-\Delta t/\tau}) \Delta t^{1/2} u_{\rm i}(t). \end{aligned} \tag{A15}$$

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

We thank Prof. F. J. Luque for his helpful suggestions. Thanks also go to Tim Meyer, Alberto Pérez, and Dr. Carles Ferrer-Costa for providing us with molecular dynamics trajectories. Calculations were performed on the *MareNostrum* supercomputer at the Barcelona Supercomputer Center.

This work was supported by the Spanish Ministry of Education and Science (BIO2006-01602 and GEN2003-20642-CO9-07), the Fundación Marcelino Botín, and the National Institute of Bioinformatics (Structural Bioinformatics Node).

REFERENCES

1. Ma, J., and M. Karplus. 1998. The allosteric mechanism of the chaperonin GroEL: a dynamic analysis. *Proc. Natl. Acad. Sci. USA*. 95:8502–8507.
2. Daniel, R. M., R. V. Dumm, J. L. Finney, and C. J. Smith. 2003. The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct.* 32:69–92.
3. Eisenmesser, E. Z., D. A. Bosco, M. Akke, and D. Kern. 2002. Enzyme dynamics during catalysis. *Science*. 295:1520–1523.
4. Luo, J., and T. C. Bruice. 2004. Anticorrelated motions as a driving force in enzyme catalysis: the dehydrogenase reaction. *Proc. Natl. Acad. Sci. USA*. 101:13152–13156.
5. Hinsen, K., A. Thomas, and M. J. Field. 1999. Analysis of domain motions in large proteins. *Proteins*. 34:369–382.
6. Waldron, T. T., and K. P. Murphy. 2003. Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics. *Biochemistry*. 42:5058–5064.
7. Yang, L.-W., and I. Bahar. 2005. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure*. 13:893–904.
8. Sacquin-Mora, S., and R. Lavery. 2006. Investigating the local flexibility of functional residues in hemoproteins. *Biophys. J.* 90:2706–2717.

9. Remy, I., I. A. Wilson, and S. W. Michnick. 1999. Erythropoietin receptor activation by a ligand-induced conformation change. *Science*. 283:990–993.
10. Lindorff-Larsen, K., R. B. Best, M. A. Depristo, C. M. Dobson, and M. Vendruscolo. 2005. Simultaneous determination of protein structure and dynamics. *Nature*. 433:128–132.
11. Rueda, M., C. Ferrer-Costa, T. Meyer, A. Perez, J. Camps, A. Hospital, J. L. Gelpi, and M. Orozco. 2007. A consensus view of protein dynamics. *Proc. Natl. Acad. Sci. USA*. 104:796–801.
12. Rueda, M., P. Chacon, and M. Orozco. 2007. Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. *Structure*. 15:565–575.
13. Karplus, M., and J. A. McCammon. 1986. The dynamics of proteins. *Sci. Am.* 254:42–51.
14. McCammon, J. A., B. R. Gelin, and M. Karplus. 1977. Dynamics of folded proteins. *Nature*. 267:585–590.
15. Allen, M. P., and D. J. Tildesley. 1989. Computer Simulation of Liquids. Clarendon Press, Oxford, UK.
16. Brooks III, C. L., M. Karplus, and B. M. Pettitt. 1987. Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics. Cambridge University Press, Cambridge, UK.
17. Warshel, A. 1976. Bicycle-pedal model for the first step in the vision process. *Nature*. 260:679–683.
18. Van Gunsteren, W. F., and M. Karplus. 1982. Protein dynamics in solution and in a crystalline environment: a molecular dynamics study. *Biochemistry*. 21:2259–2274.
19. Berendsen, H. J. C., J. P. M. Postma, W. F. Van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
20. Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9:646–652.
21. Karplus, M., and J. Kuriyan. 2005. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA*. 102:6679–6685.
22. Doruker, P., A. R. Atilgan, and I. Bahar. 2000. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins Struct. Funct. Genet.* 40:512–524.
23. McCammon, J. A., and S. C. Harvey. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, UK.
24. Alder, B. J., and T. E. Wainwright. 1959. Studies in molecular dynamics. I. General method. *J. Chem. Phys.* 31:459–466.
25. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.
26. MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*. 102:3586–3616.
27. MacKerell Jr., A. D., J. Wiorkiewicz-Kuczera, and M. Karplus. 1995. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* 117:11946–11975.
28. Damm, W., A. Frontera, J. Tirado-Rives, and W. L. Jorgensen. 1997. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* 18: 1955–1970.
29. Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118: 11225–11236.
30. Kaminski, G., E. M. Duffy, T. Matsui, and W. L. Jorgensen. 1994. Free energies of hydration and pure liquid properties of hydrocarbons from the OPLS all-atom model. *J. Phys. Chem.* 98:13077–13082.
31. Kaminski, G. A., R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*. 105:6474–6487.
32. Darden, T. L., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
33. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23:327–341.
34. Andersen, H. C. 1983. RATTLE: a velocity version of the SHAKE algorithm for molecular dynamics calculations. *J. Comput. Phys.* 52:24–34.
35. Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, W. S. Ross, C. L. Simmerling, T. L. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. 2004. AMBER8. University of California, San Francisco.
36. Kale, L., R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. 1999. NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* 151:283–312.
37. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
38. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
39. Mahoney, M. W., and W. L. Jorgensen. 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* 112:8910–8922.
40. Gardiner, C. W. 1989. Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences, 2nd ed. Springer, Berlin.
41. Van Kampen, N. G. 1981. Stochastic Processes in Physics and Chemistry. North-Holland, Amsterdam.
42. Reference deleted in proof.
43. Kovacs, J. A., P. Chacon, and R. Abagyan. 2004. Predictions of protein flexibility: first-order measures. *Proteins*. 56:661–668.
44. Ding, F., S. V. Buldyrev, and N. V. Dokholyan. 2005. Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model. *Biophys. J.* 88:147–155.
45. Ding, F., J. M. Borreguero, S. V. Buldyrey, H. E. Stanley, and N. V. Dokholyan. 2003. Mechanism for the alpha-helix to beta-hairpin transition. *Proteins Struct. Funct. Genet.* 53:220–228.
46. Marchut, A. J., and C. K. Hall. 2006. Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations. *Biophys. J.* 90:4574–4584.
47. Zhou, Y. Q., M. Karplus, J. M. Wichett, and C. K. Hall. 1997. Equilibrium thermodynamics of homopolymers and clusters: molecular dynamics and Monte Carlo simulations of systems with square-well interactions. *J. Chem. Phys.* 107:10691–10708.
48. Zhou, Y. Q., and M. Karplus. 1999. Interpreting the folding kinetics of helical proteins. *Nature*. 401:400–403.
49. Smith, A. V., and C. K. Hall. 2001. Alpha helix formation: discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins*. 44:344–360.
50. Nguyen, H. D., and C. K. Hall. 2004. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. USA*. 101:16180–16185.
51. Nguyen, H. D., V. S. Reddy, and C. L. Brooks. 2007. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Lett.* 7:338–344.
52. Smith, W. S., C. K. Hall, and B. D. Freeman. 1997. Molecular dynamics for polymeric fluids using discontinuous potentials. *J. Comput. Phys.* 134:16–30.
53. Sharma, S., F. Ding, and N. V. Dokholyan. 2007. Multiscale modeling of nucleosome dynamics. *Biophys. J.* 92:1457–1470.

54. Peng, S., F. Ding, B. Urbanc, S. V. Buldyrev, L. Cruz, H. E. Stanley, and N. V. Dokholyan. 2004. Discrete molecular dynamics simulations of peptide aggregation. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 69:041908.

55. Amadei, A., A. B. Linssen, and H. J. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.

56. Andricioaei, I., and M. Karplus. 2001. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* 115:6289–6292.

57. Schlitter, J. 1993. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* 215:617–621.

58. Perez, A., J. R. Blas, M. Rueda, J. M. López-Bes, X. de La Cruz, F. J. Luque, and M. Orozco. 2005. Exploring the essential dynamics of B.DNA. *J. Chem. Theory Comput.* 1:790–800.

59. Hess, B. 2000. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 62:8438–8448.

60. Noy, A., T. Meyer, M. Rueda, C. Ferrer, A. Valencia, A. Perez,, X. de La Cruz, J. M. Lopez-Bes, R. Pouplana, J. Fernandez-Recio, F. J. Luque, and M. Orozco. 2006. Data mining of molecular dynamics trajectories of nucleic acids. *J. Biomol. Struct. Dyn.* 23:447–456.

61. Orozco, M., A. Perez, A. Noy, and F. J. Luque. 2003. Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.* 32:350–364.

62. Hinsen, K. 1998. Analysis of domain motions by approximate normal mode calculations. *Proteins.* 33:417–429.

63. Zhou, Y., D. Vitkup, and M. Karplus. 1999. Native proteins are surface-molten solids: application of the Lindemann criterion for the solid versus liquid state. *J. Mol. Biol.* 285:1371–1375.

64. Day, R., D. A. Beck, R. S. Armen, and V. Daggett. 2003. A consensus view of fold space: combining SCOP, CATH, and the Dali domain dictionary. *Protein Sci.* 12:2150–2160.