REACH Coarse-Grained Biomolecular Simulation: Transferability between Different Protein Structural Classes

Kei Moritsugu*†‡ and Jeremy C. Smith*

*Center for Molecular Biophysics, University of Tennessee/Oak Ridge National Laboratory, Oak Ridge, Tennessee; †Computational Molecular Biophysics, Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg, Heidelberg, Germany; and ‡Computational Science Research Program, RIKEN, Wako, Japan

ABSTRACT Coarse graining of protein interactions provides a means of simulating large biological systems. The REACH (Realistic Extension Algorithm via Covariance Hessian) coarse-graining method, in which the force constants of a residue-scale elastic network model are calculated from the variance-covariance matrix obtained from atomistic molecular dynamics (MD) simulation, involves direct mapping between scales without the need for iterative optimization. Here, the transferability of the REACH force field is examined between protein molecules of different structural classes. As test cases, myoglobin (all α), plastocyanin (all β), and dihydrofolate reductase (α/β) are taken. The force constants derived are found to be closely similar in all three proteins. An MD version of REACH is presented, and low-temperature coarse-grained (CG) REACH MD simulations of the three proteins are compared with atomistic MD results. The mean-square fluctuations of the atomistic MD are well reproduced by the CGMD. Model functions for the CG interactions, derived by averaging over the three proteins, are also shown to produce fluctuations in good agreement with the atomistic MD. The results indicate that, similarly to the use of atomistic force fields, it is now possible to use a single, generic REACH force field for all protein studies, without having first to derive parameters from atomistic MD simulation for each individual system studied. The REACH method is thus likely to be a reliable way of determining spatiotemporal motion of a variety of proteins without the need for expensive computation of long atomistic MD simulations.

INTRODUCTION

Molecular dynamics (MD) computer simulation is a useful tool for characterizing protein dynamics. The accuracy of an MD simulation depends on that of the protein three-dimensional structure providing the initial coordinates, on the force field defined by the potential energy function, and on the simulation methodology.

Atomistic force fields, such as CHARMM (1), AMBER (2), and GROMOS (3), have been widely used for protein MD studies. However, simulating biological systems on long timescales (e.g., microseconds) is difficult because of limitations of computational power. To reduce computational costs for large systems, coarse-graining methods, in which the number of degrees of freedom is reduced, have been derived. For example, a protein molecule can be represented by point residues centered at the C_α atoms.

An early simplified residue-based coarse-grained (CG) force field, the elastic network model (ENM), uses a nearest-neighbor harmonic approximation and enables collective vibrational normal modes to be rapidly calculated (4–8). Although extremely simple, ENM has been usefully applied in various studies of protein function (9–13). More sophisticated force fields, including virtual local (bond, angle, torsion, etc.) and nonlocal interactions, have been proposed for CGMD simulations (14–18).

Parameters for CG models have been calculated using the Boltzmann inversion of the MD-derived probability distribution (14,15) or by iteratively matching the internal coordinate fluctuations (16) or the forces on the CG sites (17) to all-atom MD. In another approach, the implicit degrees of freedom are separated dynamically from the CG degrees of freedom using a united-residue energy function (18). Although the above approaches all have merits and hold promise for determining dynamic aspects of macromolecular function using CGMD (14–22), the parameters of the CG model are determined a posteriori so as to reproduce, e.g., vibrational amplitudes. As a result, the derived parameters do not have clearly attributable physical origins, in contrast to all-atom force fields.

In previous work (23), we introduced a methodology, REACH (Realistic Extension Algorithm via Covariance Hessian), for deriving residue-scale ENM force constants from the variance-covariance matrix calculated from atomic-detail MD simulation. The REACH method is a direct mapping, requiring no iterative fitting and no input of experimental data. The REACH normal modes were shown to reproduce the amplitudes and frequencies of MD-derived fluctuations in myoglobin (23).

The aim of this study is to examine the transferability of the REACH CG force constants among protein molecules. To do this, the REACH method was applied to three model proteins of different structural classes (24): myoglobin (an α-fold of eight α-helices), plastocyanin (a β-fold of eight β-strands), and dihydrofolate reductase (DHFR) (an α/β-fold, comprising 4 α-helices and 10 β-strands). The structures of the three proteins are shown in Fig. 1.

Submitted February 18, 2008, and accepted for publication April 18, 2008.

Address reprint requests to Jeremy C. Smith, Center for Molecular Biophysics, University of Tennessee/Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831. Tel.: 865-574-9635; Fax: 865-576-7651; E-mail: smithjc@ornl.gov.

Editor: Gregory A. Voth.

0006-3495/08/08/1639/10 $2.00

doi: 10.1529/biophysj.108.131714

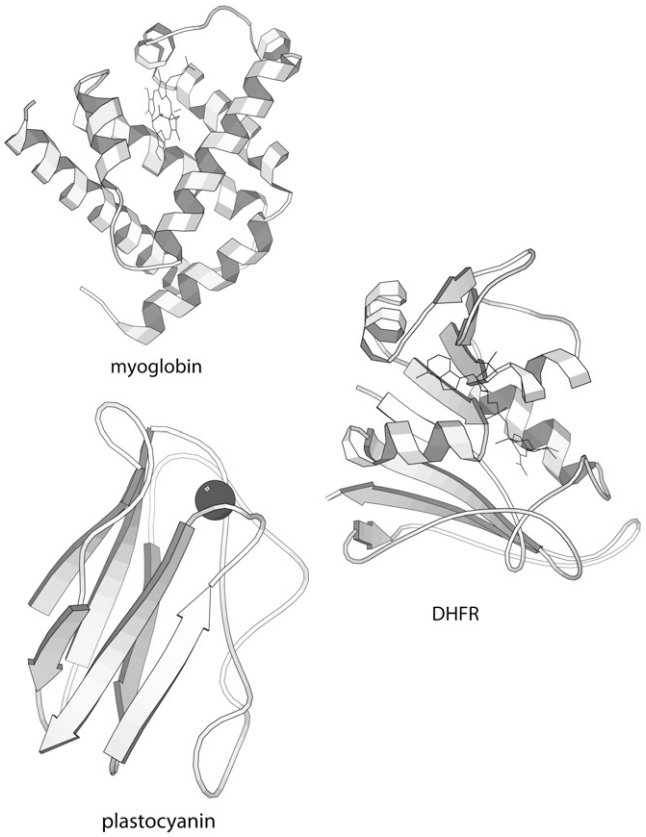

FIGURE 1 Three-dimensional structures of myoglobin, DHFR, and plastocyanin. Drawn with the program MOLSCRIPT (38).

The REACH force constants for the three proteins are calculated from 10-ns atomistic MD trajectories. The interactions within α-helices and β-strands, and the interactions between β-strands (forming β-sheets) are focused on. Analytical model functions are derived representing the distance dependence of the force constants. The REACH mean-square fluctuations are calculated and compared with those from the atomistic MD. In a further development, the method is extended from normal mode analysis to MD. CGMD simulations are performed with the REACH force field. The mean-square fluctuations, the vibrational densities of states, and the dynamic cross-correlation matrices are calculated from the atomistic and coarse-grained MD and compared. A generic, transferable analytical REACH force field is derived that reproduces well the fluctuations from proteins of different structural classes.

THEORY AND METHODS

Atomistic MD simulations

MD simulations of myoglobin, DHFR, and plastocyanin were performed. The starting structures were obtained from the Protein Data Bank (25): 1A6G (26), 1RX1 (27), and 1PLC (28), respectively.

The model systems were constructed as follows: A rectangular primary simulation box was built of dimensions, 60 Å × 60 Å × 60 Å for myoglobin, 62 Å × 52 Å × 48 Å for DHFR, and 62 Å × 54 Å × 52 Å for plastocyanin. Then 7181 TIP3P water molecules (29) were placed in the box for myoglobin, 3173 for DHFR, and 4413 for plastocyanin, together with one chloride ion for myoglobin, six potassium ions for DHFR and seven potassium ions for plastocyanin. This procedure constructed electrically neutral systems of 21,110 atoms for myoglobin, 12,094 for DHFR, and 14,712 for plastocyanin. Periodic boundary conditions were imposed on the boxes.

The simulations were performed using the program NAMD2 (30). The CHARMM all-atom parameter set 22 (31) was used for the potential function. Electrostatic interactions were calculated with a dielectric constant of 1 using the particle mesh Ewald method (32), i.e., without truncation. The systems were energy minimized with 1000 steps of the conjugate gradient method. Then, each simulation system was uniformly heated to the target temperature of 120 K over 30 ps and equilibrated for 100 ps with velocity scaling in the NVE ensemble. Subsequent simulations were carried out at constant temperature (120 K) and pressure (1 atm) conditions (the NPT ensemble) for a 500-ps equilibrium run and 10-ns production runs. The temperature of 120 K was selected so as to reduce the contribution of anharmonic protein motion: the previous study showed a clear decrease of the force constants with temperature corresponding to softening of protein dynamics (23). Subsequent work will focus on construction of a force field capable of accurately simulating anharmonic dynamics at physiological temperatures. Langevin dynamics was used to control the temperature and pressure (33). All production runs were performed for 10 ns. The atomic coordinates were saved every 50 fs for analysis.

REACH force constant calculation: secondary-structural elements

The REACH CG method uses only C_α atom coordinates. Heteroatoms such as ions and ligands are not included. The mass of each CG residue is defined as the sum of the atomic masses comprising the corresponding residue.

The REACH method is described in detail in the literature (23). Here we give a brief summary. The potential energy in the ENM is written as follows (4):

$$V = \frac{1}{2}\sum_{i<j} k_{ij}(r_{ij} - r_{0,ij})^2, \tag{1}$$

where r_{ij} ($r_{0,ij}$) is the distance between the dynamic (equilibrium) positions of the C_α atoms in residues i and j, and k_{ij} is the force constant for the harmonic spring between i and j. Calculation of the Hessian (second-derivative) matrix from Eq. 1 leads to the relation:

$$k_{ij} = -tr(\mathbf{K}_{ij}), \tag{2}$$

where $\mathbf{K}_{ij}$ is the off-diagonal component of the Hessian associated with i and j. Making the harmonic approximation under the equilibrium condition at constant temperature, T, allows the Hessian matrix to be calculated from the variance-covariance matrix, $\mathbf{C} = (C_{ij}) = \left(\langle(r_i - \langle r_i\rangle)(r_j - \langle r_j\rangle)\rangle\right)$ as

$$\mathbf{K} = k_B T \mathbf{C}^{-1}, \tag{3}$$

where k_B is the Boltzmann constant. k_{ij} is then derived by combining Eqs. 2 and 3 as follows:

$$k_{ij} = -k_B T tr(\mathbf{C}_{ij}^{-1}) \tag{4}$$

In this study, the 10-ns MD trajectories were each separated into ten 1-ns trajectories, allowing the calculation of $\mathbf{C}$ from each 1-ns trajectory. Subsequently, the ten $\mathbf{C}$ matrices were averaged, and the associated force constants derived using Eqs. 3 and 4. To calculate the force constants for local interactions, i.e., the virtual 1−2 (between residues i and $i + 1$), 1−3 (between residues i and $i + 2$) and 1−4 (between residues i and $i + 3$) interactions, segments of 20 residues were fitted individually to calculate submatrices of $\mathbf{C}$. This avoids problems associated with the best fit to the overall structure arising from the incorporation of external motions of the segments, resulting in errors in the pairwise covariances (23).

An analysis was performed of the dependence of the REACH force constants on the structural fold, i.e., the secondary-structural elements. For this, the DSSP program (34) was used to define the residues comprising secondary structures. The interactions within α-helices and β-strands, and between β-strands, were calculated using the corresponding secondary structural segments: each secondary structure was fitted individually to calculate a submatrix of **C,** and then the force constants were derived. For α-helices, a virtual 1−5 (between residues i and $i + 4$) interaction was included in the force field arising from the helical pitch of ~3.6 residues.

Mathematical model functions describing the distance dependence of the force constants are useful for obtaining a simplified understanding of protein dynamics and for convenient application in CG simulations. Terms were constructed for each interaction type and include a dependence on the secondary-structural elements. The local (1,2, 1–3, 1–4, and 1–5) interactions and the inter-β-strand interaction were modeled with a single force constant. In contrast, the nonbonded interaction force constants were modeled as distance dependent with a double-exponential function. The resulting force constant parameters are as follows:

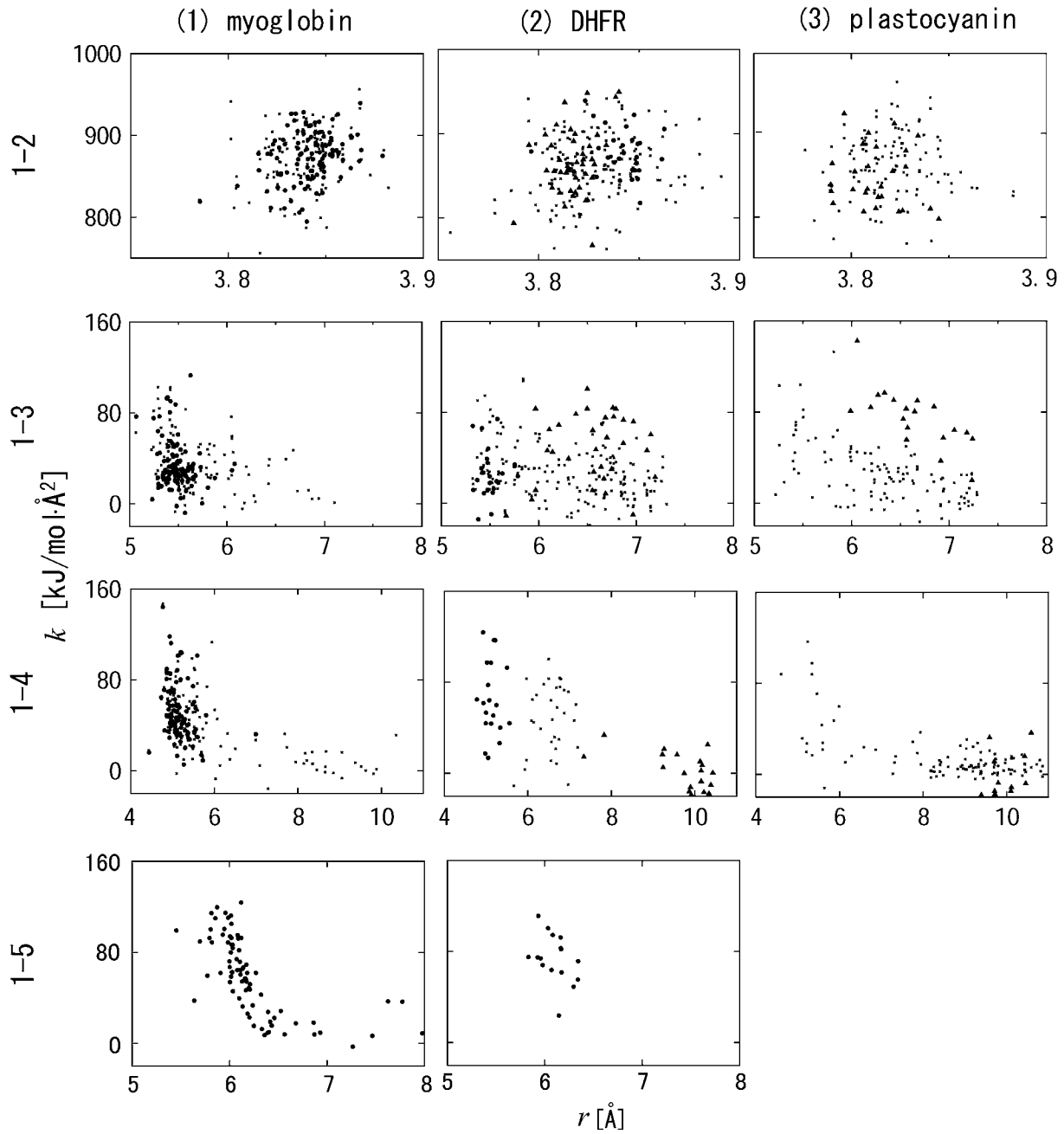

FIGURE 2 REACH force constants, k, of virtual 1−2, 1−3, 1−4, and 1−5 interactions in (*1*) myoglobin, (*2*) DHFR, and (*3*) plastocyanin are plotted as a function of pairwise distance, r. Interactions within α-helices and β-strands are plotted with circles and triangles, respectively. The virtual 1−2 interaction via *cis* peptide bonds (residues 95–96 in DHFR and residues 15–16 and 35–36 in plastocyanin) is excluded from the figure.

$$k(r) = \begin{cases} k_{12}(k_{12,\alpha}, k_{12,\beta}) & 1-2 \\ k_{13}(k_{13,\alpha}, k_{13,\beta}) & 1-3 \\ k_{14}(k_{14,\alpha}, k_{14,\beta}) & 1-4 \\ k_{15,\alpha} & 1-5 \\ k_{\text{inter-}\beta} & \text{inter-}\beta \\ a_f \exp(-b_f r) + a_s \exp(-b_s r) & \text{nonlocal} \end{cases}, \quad (5)$$

where α and β mean the interactions within α- and β-secondary-structural elements, respectively. The model parameters were calculated by fitting the associated model functions to the force-constant distributions.

REACH CG normal-mode calculation and MD simulation

Using the potential energy (Eq. 1) together with the equilibrium structure and the force constant models, residue-scale CG normal mode calculations and MD simulations were performed.

Normal-mode eigenvalues, $\{\lambda_i\}$, and eigenvectors, $\{\mathbf{v}_i\}$, were calculated by diagonalizing the Hessian matrices derived from the potential energy. The mean-square fluctuation of residue n, x_n^2, at temperature T is given by

$$x_n^2 = \frac{k_B T}{m_n} \sum_i v_{n,i}^2 / \lambda_i, \quad (6)$$

where $v_{n,i}$ is the displacement of residue n in mode i.

CGMD simulations were performed using the program DL_POLY (35). For this, a Morse potential, $v(r) = A[1 - \exp\{-(r - r_0)/\sqrt{2}\}]^2$, was applied for the nonlocal interactions, replacing the harmonic potential in Eq. 1, i.e.,

$$V = \sum_{i<j,\text{local}} \frac{1}{2} k(r_{0,ij})(r_{ij} - r_{0,ij})^2 + \sum_{i<j,\text{nonlocal}} A(r_{0,ij}) \left[1 - \exp\left\{ -\frac{(r_{ij} - r_{0,ij})}{\sqrt{2}} \right\}\right]^2. \quad (7)$$

The harmonic force constant k was replaced with A, and the same equilibrium distance, r_0, was used to map the harmonic potential energy onto the Morse potential around r_0. Simulations were performed at 120 K so that the dynamics remains mostly harmonic.

Initial coordinates were taken as the C_α coordinates of the average structure over the 10-ns atomistic MD simulation. Both 1-ns equilibration and 1-ns production runs were performed at constant volume and temperature conditions. The time step applied was 50 fs. Langevin dynamics (33) was used to maintain the constant temperature with a friction constant of $\gamma = 3\ \text{ps}^{-1}$, a value chosen to be similar to the average friction exerted on the low-frequency motions $< 100\ \text{cm}^{-1}$ in atomistic solvated myoglobin at 120 K (36). Nonlocal interactions were truncated at $r_c < 40$ Å. The mean-square fluctuation of each residue was calculated from the trajectories derived and compared with the atomistic MD and CG normal-mode results.

Calculations of vibrational density of states and dynamic cross-correlation matrix from MD trajectory

The vibrational density of states, $g(\omega)$, was calculated from the MD trajectory as the Fourier transform of the velocity autocorrelation function; i.e.,

$$g(\omega) = \frac{1}{2\pi} \sum_{i=1}^{N} \int_{-\infty}^{\infty} \langle \mathbf{v}_i(0) \cdot \mathbf{v}_i(t) \rangle / \langle \mathbf{v}_i^2 \rangle e^{-i\omega t} dt. \quad (8)$$

To compare the spectrum from CGMD with that from atomistic MD, only the C_α atoms were used for the calculation; i.e., N in Eq. 8 is the number of residues. A total of 2^{14} MD trajectory frames, separated by 50 fs (total length of ~0.82 ns), were used for calculating the velocity autocorrelation function and $g(\omega)$.

The dynamic cross-correlation matrix of atomic fluctuations consists of elements $c_{i,j}$ defined as

$$C_{i,j} = \frac{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle) \cdot (\mathbf{r}_j - \langle \mathbf{r}_j \rangle) \rangle}{\sqrt{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)^2 \rangle \langle (\mathbf{r}_j - \langle \mathbf{r}_j \rangle)^2 \rangle}} \quad (9)$$

where $\langle \cdot \rangle$ denotes the average over the overall MD trajectory. Again, only the C_α atom pairs, i and j, were used for the calculation from the atomistic MD. A completely correlated or anticorrelated motion has $c_{ij} = 1$ or -1, respectively, whereas c_{ij} is 0 if a pairwise motion is perfectly uncorrelated.

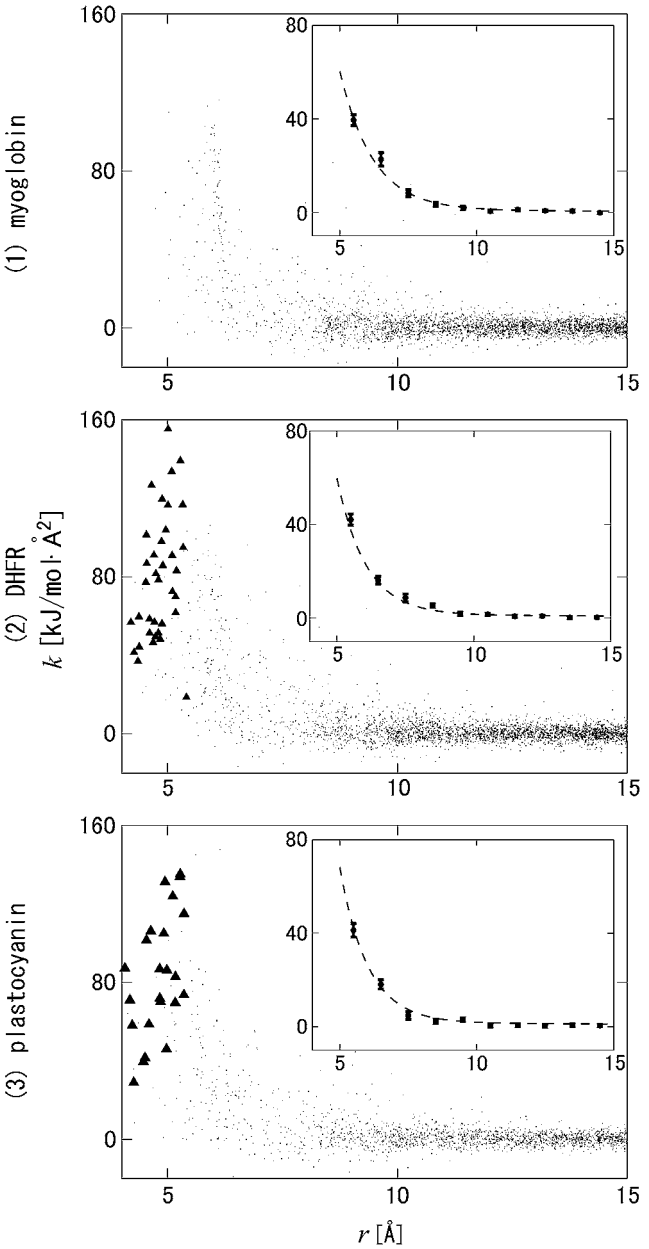

FIGURE 3 REACH force constants, k, of nonbonded interactions in (*1*) myoglobin, (*2*) DHFR, and (*3*) plastocyanin are plotted with dots as a function of pairwise distance, r. Interactions between β-strands are plotted with triangles. In the insets the averages of k within bins of 1 Å width are shown together with the corresponding fitted model function (*dashed curve*).

TABLE 1 REACH force constants derived from atomistic MD trajectories

Protein	k_{12}	k_{13}	k_{14}	a_f	b_f	a_s	b_s
Myoglobin	873 ± 2.5	32.8 ± 1.7	40.7 ± 2.1	8330 ± 1000	0.902 ± 0.022	0.755 ± 1.3	0.0489 ± 0.048
DHFR	860 ± 2.7	26.7 ± 1.7	17.0 ± 1.5	6570 ± 1100	0.869 ± 0.038	1.15 ± 1.0	0.0456 ± 0.11
Plastocyanin	865 ± 3.6	26.6 ± 2.5	14.8 ± 2.0	10200 ± 1400	0.960 ± 0.024	0.912 ± 0.55	0.0235 ± 0.040

$k_{\text{nonlocal}}(r) = a_X \exp(-b_X r)$, where the subscripts X = *f*, *s* denote fast and slow components, respectively. Units are kJ/mol·Å^2, except for b_X, which is in Å^{-1}.

RESULTS

REACH force constants: transferability among proteins of different structural classes

The REACH force constants for the three proteins, myoglobin, DHFR, and plastocyanin, were calculated from the corresponding atomistic MD trajectories and are compared. The interactions within and between secondary-structural elements are also examined.

In Fig. 2, the REACH force constants, k, for interactions between the local residue pairs are plotted as a function of the pairwise distance, r. The distributions, $k(r)$, of the virtual 1−2, 1−3, 1−4, and 1−5 interactions are very similar among the three proteins. The distributions of the 1−2 interactions are quite different from the others, with a larger k, peaked at ~3.83 Å, arising from virtual bonds between neighboring amino acids. The 1−2 interaction via *cis* peptide bonds has a smaller k and r than that via *trans* peptide bonds: *cis* peptide bonds occur at residues 95−96 in DHFR ($k = 491$ kJ/mol·Å^2 and $r = 2.97$ Å), and residues 15−16 ($k = 612$ kJ/mol·Å^2 and $r = 3.05$ Å) and 35−36 ($k = 587$ kJ/mol·Å^2 and $r = 3.11$ Å) in plastocyanin.

The local interactions within secondary structures are also shown in Fig. 2. Small differences in $k(r)$ among the three proteins are caused by differences in interactions within the secondary-structural elements. The distribution of interactions within α-helices in DHFR is narrower than that in myoglobin because there is more variety in the structure and dynamics of the eight helices (comprising 113 residues) in myoglobin than in the four helices (comprising 32 residues) in DHFR. The 1−2 interactions within the α-helices are slightly larger than those within the β-strands. The average equilibrium distance, $\langle r \rangle$, is also slightly larger for α-helices (3.84 Å for both myoglobin and DHFR) than for β-strands (3.82 Å for DHFR and 3.81 Å for plastocyanin). For the 1−3, 1−4, and 1−5 interactions, the distributions of force constants within α-helices are narrower than those within β-strands, with $\langle r \rangle$ being 5.49, 5.16, 6.22 Å for myoglobin and 5.51, 5.13, 6.10 Å for DHFR, respectively. For α-helices, the 1−4 interaction has a smaller $\langle r \rangle$ than the 1−3 interaction because of the helical pitch of ~3.6 residues. In contrast, the 1−3 and 1−4 interactions within β-strands have larger $\langle r \rangle$ and also more scattered distributions because of the extended structure of β-strands.

Fig. 3 shows k for the nonlocal interaction residue pairs as a function of r. The distributions of the nonlocal interactions, including the inter-β-interactions, are again very similar among the three proteins. Because of the hydrogen bonds between β-strands forming β-sheets, the distributions of the inter-β-interactions are found, on average, at shorter pairwise distances than those within β-strands, with $\langle r \rangle$ for the former being 4.83 Å for DHFR and 4.81 Å for plastocyanin.

Model functions of the distance dependence of the force constants were constructed. The model functions of the local interactions and the inter-β-interactions, which have narrow $k(r)$ distributions and no significant distance dependence, were assumed to be constant with distance, as in the previous REACH implementation (23) and another CGMD study (14). In contrast, the model functions for the nonbonded interactions were assumed to be a double-exponential in form. The slower decaying exponential term, which is the smaller in magnitude, was found to be essential for reproducing the atomistic MD fluctuations (results not shown). The resulting parameters and standard-deviation errors are listed in Tables 1 and 2. Again, the force constant parameters are similar among the three proteins. The force constant values obtained for interactions within secondary-structural elements for the three proteins are very close to each other, and indeed, the values obtained from the three proteins by averaging the interactions within each α- or β-element, i.e., $k_{12\alpha}$ or $k_{12\beta}$ for, e.g., the 1−2 interaction, are closer to each other than the values averaged over all the 1−2 interactions (i.e., k_{12}).

Mean-square fluctuation from REACH normal modes and MD simulations

With the REACH force-constant model functions listed in Tables 1 and 2, residue-scale CG normal-mode analyses and MD simulations were performed for the three proteins, myoglobin, DHFR, and plastocyanin, and the mean-square fluctuations, x^2, were calculated.

TABLE 2 Secondary-structural REACH force constants derived from atomistic MD trajectories

Protein	$k_{12\alpha}$	$k_{12\beta}$	$k_{13\alpha}$	$k_{13\beta}$	$k_{14\alpha}$	$k_{14\beta}$	$k_{15\alpha}$	$k_{\text{inter-}\beta}$	a_f	b_f	a_s	b_s
Myoglobin	874 ± 3.0		34.0 ± 2.2		51.3 ± 3.0		58.9 ± 3.7		3690 ± 570	0.826 ± 0.030	1.50 ± 2.7	0.0666 ± 0.065
DHFR	878 ± 5.2	866 ± 6.0	26.4 ± 4.3	48.4 ± 5.3	65.4 ± 7.2	−3.86 ± 4.5	73.8 ± 5.2	79.5 ± 5.3	6770 ± 1800	0.951 ± 0.052	2.08 ± 1.4	0.0589 ± 0.043
Plastocyanin		859 ± 6.9		74.8 ± 5.8		−3.45 ± 5.5		83.5 ± 6.4	9010 ± 1800	0.982 ± 0.042	2.27 ± 1.2	0.0458 ± 0.037

$k_{\text{nonlocal}}(r) = a_X \exp(-b_X r)$, where the subscripts X = *f*, *s* denote fast and slow components, respectively. Units are kJ/mol·Å^2, except for b_X, which is in Å^{-1}.

Fig. 4 shows x^2 from the normal modes (CGNM,noss) calculated using Eq. 5. As summarized in Table 3, the magnitude of x^2 from the REACH normal modes is slightly larger than from the atomistic MD but is in satisfactory agreement given the simplicity of the REACH method. The value of x^2 for residue 95 in DHFR is much larger from the atomistic MD than from the REACH calculation because of the presence of the *cis* peptide bond between residues 95 and 96.

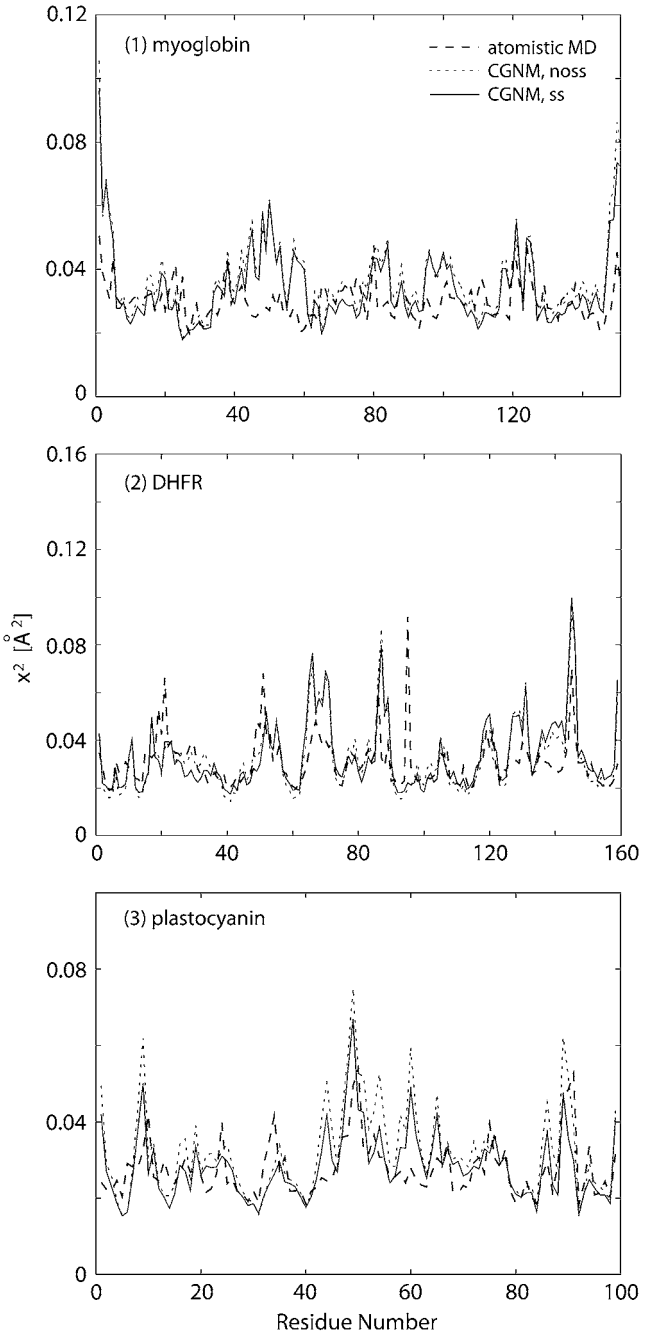

FIGURE 4 Mean-square fluctuation, x^2, of each residue, derived from atomistic MD (*dashed curve*) and REACH normal modes with (CGNM,ss, *solid curve*) and without (CGNM,noss, *dotted curve*) secondary-structure force constants: (*1*) myoglobin, (*2*) DHFR, and (*3*) plastocyanin.

Normal-mode analyses were also performed using a modified potential function in which the force constant parameters were derived separately for the secondary-structural elements (CGNM,ss). The residue dependences of x^2 with and without the secondary-structural force constants are slightly different (see Fig. 4), as are the magnitudes of x^2 and the correlation coefficients with the atomistic MD results (see Table 3). The agreement in x^2 magnitude with the atomistic MD is slightly better when the force constant model parameters defined separately for secondary structural elements. However, this may only be because the additional parameters enable better fitting to the $k(r)$ distributions.

In Fig. 5, x^2 from the CGMD is shown for the three proteins. The averages and correlation coefficients with the atomistic MD are listed in Table 3. The average x^2 from the CGMD is smaller and in better agreement with the atomistic MD than that obtained from the normal modes. The finite timescale of 1 ns may decrease protein fluctuations compared with the normal-mode results, which cover an infinite timescale, given the harmonic approximation to the potential energy. Again, the large x^2 from the atomistic MD at residue 95 in DHFR, probably because of the *cis* peptide bond between residues 95 and 96, is not reproduced by the CGMD even though in this case a smaller $k = 491$ [kJ/mol·Å^2] was used for the residue pair instead of the $1-2$ interaction force constant. Additional parameters for the $1-3$ and $1-4$ interactions may be necessary for future modeling of *cis* peptide bonds.

Finally, we investigated whether a single set of REACH force constants can be used in CGMD of the three proteins, in an analogous way that atomistic MD simulations of the three proteins are performed with a single atomistic force field (in our study, CHARMM 22 (31)). To do this, a new set of force constant parameters was constructed by averaging over the three proteins. For the local interactions, the derived ''constant'' model parameters were averaged. For the nonlocal interactions, the three distributions of $k_{\mathrm{nonlocal}}(r)$ were superimposed and fitted with a new model function: the resulting function derived is, $k_{\mathrm{nonlocal}}(r) = 4810\exp(-0.872r) + 1.7\exp(-0.068r)$[kJ/mol·Å^2]. The new, common force-constant parameters were then applied to perform CGMD (CGMDave) of the three proteins, and x^2 was calculated.

The results, shown in Fig. 5 and Table 3, indicate that the adoption of the average set of force constant parameters increases the average fluctuations for the three proteins. However, the magnitude of the average x^2 remains similar to the atomistic MD, and the residual x^2 is still highly correlated with the atomistic MD.

Internal dynamics from atomistic and CGMD

Protein internal dynamics simulated by the residue-scale REACH CGMD is now compared with that obtained from the atomistic MD. To do this, the vibrational density of states, $g(\omega)$, and the dynamic cross-correlation matrix elements, c_{ij}, were calculated from the MD trajectories of the three proteins.

TABLE 3 Average mean-square fluctuation from atomistic MD, normal modes, and CGMD

Protein	Atomistic MD	CGNM, noss	CGNM,ss	CGMD	CGMD,ave
Myoglobin	0.0301	0.0366 (0.41)	0.0342 (0.42)	0.0316 (0.44)	0.0333 (0.36)
DHFR	0.0310	0.0336 (0.54)	0.0332 (0.52)	0.0328 (0.47)	0.0370 (0.43)
Plastocyanin	0.0276	0.0321 (0.61)	0.0284 (0.58)	0.0281 (0.59)	0.0360 (0.63)

Units are Å^2. Correlation coefficients with atomistic MD are shown in parentheses. See text for details.

Fig. 6 shows $g(\omega)$ for the three proteins. In the low-frequency region of interest ($\omega < 200$ cm^{-1}) $g(\omega)$ calculated from the atomistic MD by averaging over all the atoms is found to be similar to that averaging over only the C_α atoms (results not shown), a result following from the global, collective nature of low-frequency protein motion. The $g(\omega)$ from the CGMD lacks intensity at $\omega = 100-150$ cm^{-1} relative to the all-atom model; i.e., there is an absence of vibrational modes at high frequencies, as was also found in previous work demonstrating that the residue-scale CG force field cannot reproduce well > 100 cm^{-1} all-atom protein dynamics (23). In contrast, the $g(\omega)$ intensity at $\omega < 100$ cm^{-1} is larger in the CGMD than in the atomistic MD spectrum, indicating that additional lower-frequency motions are added in the CGMD to satisfy the amplitudes present in the atomistic MD.

Fig. 7 shows the cross correlations c_{ij} for the three proteins. The agreement in c_{ij} between the atomistic and CGMD is good for the three proteins. Particularly good agreement is seen for the correlations of C_α atom pairs with large positive values. Thus, the correlations of C_α atom pairs in atomistic MD simulations are well reproduced by the REACH CGMD. However, the magnitude of cross correlation from the CGMD is in general smaller than that from the atomistic MD. One reason for this may be that the cross-correlation matrix is related to the contact map, defined by the pairwise distance or the interaction strength. Atom pairs with larger magnitudes of c_{ij} are likely to be close in distance and to interact strongly. Coarse-graining eliminates explicit covalent interactions between C_α atom pairs and thus decreases the dynamic correlations.

CONCLUDING REMARKS

This work examines the transferability of the REACH CG simulation potential among proteins of different structural classes: myoglobin (α-fold), DHFR (α/β-fold), and plastocyanin (β-fold).

The REACH force constants for the three proteins were obtained from the variance-covariance matrices calculated from the atomistic MD trajectories. The force constants, when plotted as a function of the residue-residue pairwise distance, exhibit distinct distributions for the local (the virtual $1-2$, $1-3$, $1-4$, and $1-5$) and nonlocal interactions. These distributions were found to be similar among the three proteins. Model functions of distance dependence of the force constants, constructed for each type of interaction, were again found to be similar for the three proteins. Both the REACH CG normal modes and MD simulations reproduce reasonably well the mean-square fluctuations from the atomistic MD. The introduction of separate model function parameters for describing interaction within α- and β-secondary-structural elements improves slightly the agreement in x^2 values. A further introduction of interaction model function parameters, e.g., including residue-type dependence, may decrease the width of the distributions in the distance-dependent force constants, increase similarity of the distributions among the three proteins, and also improve the x^2 agreement.

Force-constant model functions were constructed by averaging the results over the three proteins, with the goal of examining whether a single analytical residue-scale force field, transferable between structural classes, could be generally applied, analogously to atomistic force fields such as CHARMM (1), AMBER (2), and GROMOS (3). This force field averaged over the three proteins is also found to reproduce reasonably well the mean-square fluctuations for all the three proteins, indicating that a single force field for the residue-scale CG model can be applied to different protein classes. Therefore, the REACH force constants from the three proteins can be, for practical purposes, considered identical. One implication of this finding is that the REACH method indeed maps directly the atomistic MD force field onto a generic residue-scale CG potential energy function.

To examine whether not only fluctuations but also time-scales and residue-pair correlations found in atomistic MD protein internal dynamics are also reproduced by the CGMD, the vibrational densities of states and the dynamic cross-correlation matrices were calculated and compared. Coarse graining decreases the spectral amplitude of higher-frequency motions ($>\sim 100$ cm^{-1}) as a result of the reduction in the number of degrees of freedom, and the magnitude of the C_α atom correlations is also reduced somewhat. Nevertheless, the correspondence between the atomistic and CGMD spectral and correlation results is satisfactory given the simplicity of the REACH method.

The REACH method is self-consistent, involving a direct mapping of atomistic MD results onto the CG model. This self-consistency is clearly evident in the finding of this study that the distributions of distance-dependent force constants and the associated model functions are similar for the three proteins and that the CG force field constants averaged over those of the three proteins lead to a generalized set of force fields that reproduce the mean-square fluctuations and vibrational frequencies from the atomistic MD for all three proteins.

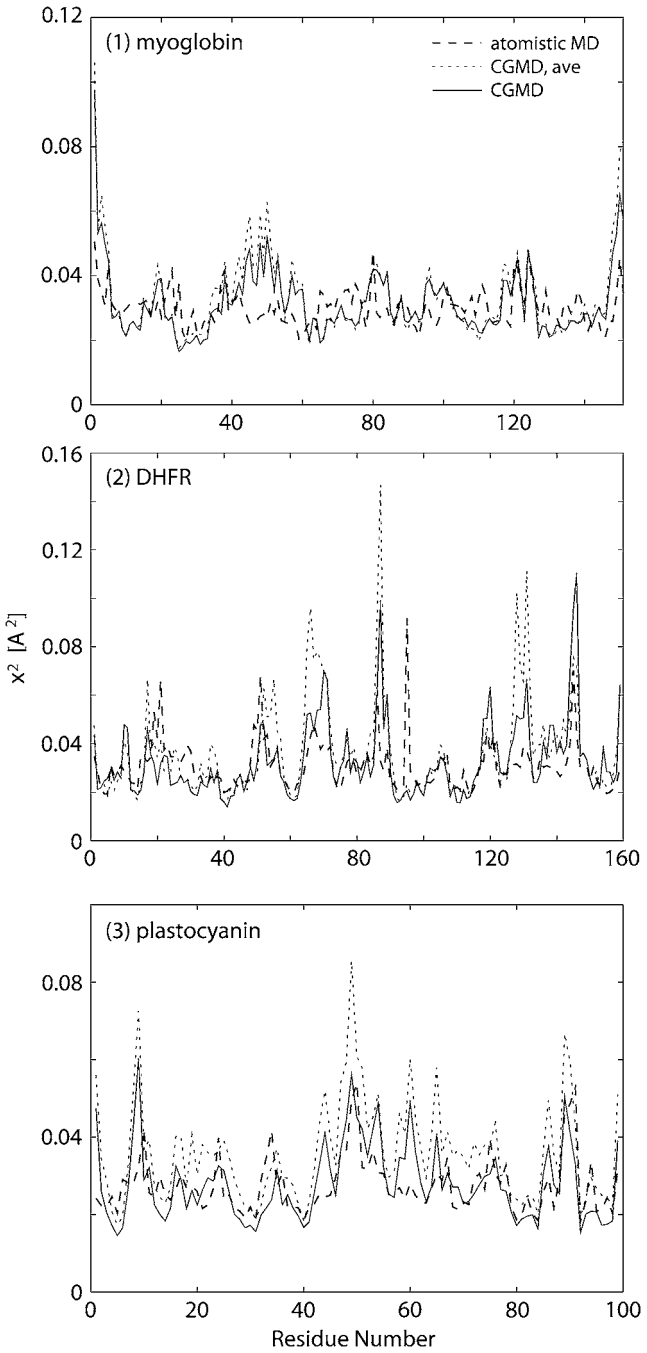

FIGURE 5 Mean-square fluctuation, x^2, of each residue derived from atomistic MD (*dashed curve*) and REACH CGMD with secondary-structure force constants (CGMD, *solid curve*). x^2 from REACH CGMD using REACH force constants averaged over the three proteins (see text in detail) is plotted as a dot profile (CGMD,ave): (*1*) myoglobin, (*2*) DHFR, and (*3*) plastocyanin.

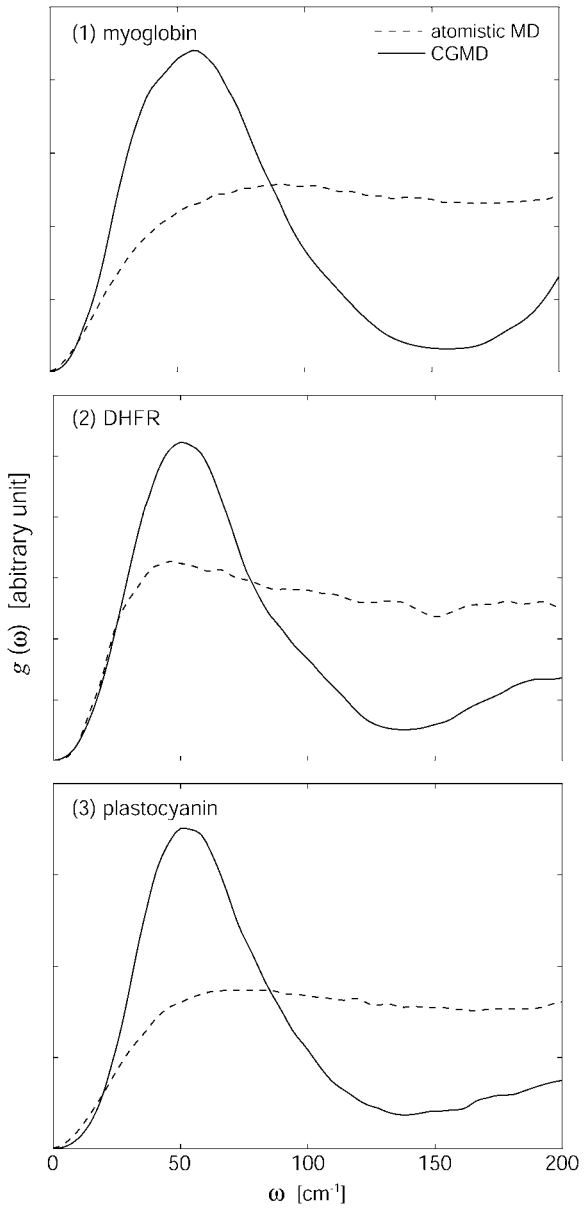

FIGURE 6 Vibrational density of states, $g(\omega)$, derived from atomistic MD (*dashed curve*) and REACH CGMD (*solid curve*): (*1*) myoglobin, (*2*) DHFR, and (*3*) plastocyanin.

In future work, the REACH force constants will be calculated from atomistic MD simulation of a larger number of proteins so as to obtain better statistics for generalized residue-scale force-constant functions. Further validation of the generic force field will be performed using an extended subset of proteins that are not used for derivation of the REACH model parameters. Furthermore, a more flexible CG potential energy function, including more extensive anharmonic effects, will be implemented in the REACH methodology, allowing the dynamics present in MD at physiological temperatures to be reproduced. As a future extension, an anharmonic potential for C_α atom pairs between secondary structural elements may be introduced while keeping the intrasecondary-structure interactions harmonic. Subsequent development may include the incorporation of anharmonicity into intrasecondary-structure interactions, allowing formation and deformation of secondary-structural elements to be simulated. Dynamic solvent effects will be taken into account, and a methodology for

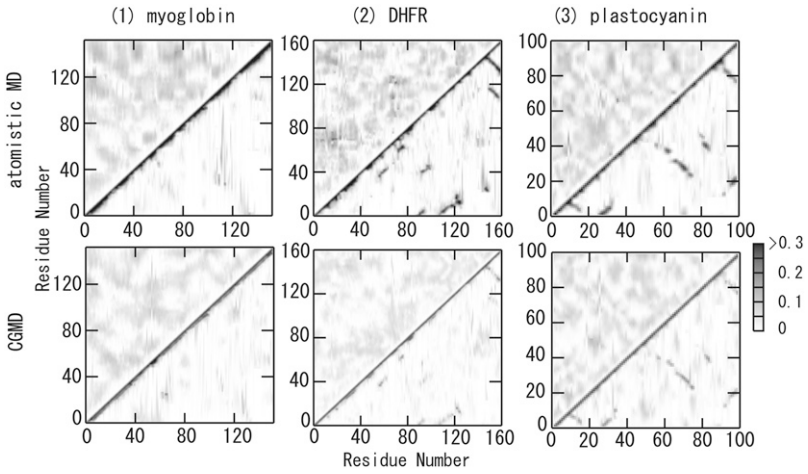

FIGURE 7 Dynamic cross-correlation matrix, c_{ij}, derived from atomistic MD (*upper row*) and REACH CGMD (*lower row*). In lower right (*upper left*), c_{ij} is plotted when c_{ij} is positive (*negative*). Pairs with $c_{ij} > 0.3$ are plotted with dots using the same gray shade as $c_{ij} = 0.3$: (*1*) myoglobin, (*2*) DHFR, and (*3*) plastocyanin.

calculating the Langevin friction on the C_α atoms from atomistic MD trajectories will be implemented (36,37).

We acknowledge funds from the U.S. Department of Energy, the Volkswagen Stiftung (I/80 437), and the European Union (NEST 012835 EMBIO). K.M. acknowledges support by the MEXT grand challenge program using next-generation supercomputing.

REFERENCES

1. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
2. Weiner, S. C., P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106:765–784.
3. Hermans, J., H. J. C. Berendsen, W. F. Van Gunsteren, and J. P. M. Postma. 1984. A consistent empirical potential for water-protein interactions. *Biopolymers.* 23:1513–1518.
4. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.
5. Bahar, I., A. R. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* 2:173–181.
6. Haliloglu, T., I. Bahar, and B. Erman. 1998. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79:3090–3093.
7. Bahar, I., A. R. Atilgan, M. C. Demirel, and B. Erman. 1998. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* 80:2733–2736.
8. Hinsen, K. 1998. Analysis of domain motions by approximate normal mode calculations. *Proteins Struct. Funct. Genet.* 33:417–429.
9. Tama, F., and C. L. Brooks 3rd. 2005. Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *J. Mol. Biol.* 345:299–314.
10. Bahar, I., and A. J. Rader. 2005. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* 15:586–592.
11. Kundu, S., and R. L. Jernigan. 2004. Molecular mechanism of domain swapping in proteins: an analysis of slower motions. *Biophys. J.* 86:3846–3854.
12. Yang, L. W., X. Liu, C. J. Jursa, M. Holliman, A. J. Rader, H. A. Karimi, and I. Bahar. 2005. iGNM: a database of protein functional motions based on Gaussian Network Model. *Bioinformatics.* 21:2978–2987.
13. Tobi, D., and I. Bahar. 2005. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA.* 102:18908–18913.
14. Trylska, J., V. Tozzini, and A. McCammon. 2005. Exploring global motions and correlations in the ribosome. *Biophys. J.* 89:1455–1463.
15. Tozzini, V., and J. A. McCammon. 2005. A coarse grained model for the dynamics of flap opening in HIV-1 protease. *Chem. Phys. Lett.* 413: 123–128.
16. Chu, J. W., and G. A. Voth. 2006. Coarse-grained modeling of actin filament derived from atomistic-scale solutions. *Biophys. J.* 90:1572–1582.
17. Zhou, J., I. F. Thorpe, S. Izvekov, and G. A. Voth. 2007. Coarse-grained peptide modeling using a systematic multiscale approach. *Biophys. J.* 92:4289–4303.
18. Liwo, A., M. Khalili, and H. A. Scheraga. 2005. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proc. Natl. Acad. Sci. USA.* 102:2362–2367.
19. Arkhipov, A., P. L. Freddolino, and K. Schulten. 2006. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure.* 14:1767–1777.
20. Freddolino, P. L., A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. 2006. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure.* 14:437–449.
21. Okazaki, K., N. Koga, S. Takada, J. N. Onuchic, and P. G. Wolynes. 2006. Multiple-basin energy landscapes for large amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA.* 103:11844–11849.
22. Koga, N., and S. Takada. 2006. Folding-based molecular simulations reveal mechanisms of the rotary motor F1-ATPase. *Proc. Natl. Acad. Sci. USA.* 103:5367–5372.
23. Moritsugu, K., and J. C. Smith. 2007. Coarse-grained biomolecular simulation with REACH: realistic extension algorithm via covariance Hessian. *Biophys. J.* 93:3460–3469.
24. 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.
25. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, and H. Wessig. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
26. Vojtechovsky, J., K. Chu, J. Berendzen, R. M. Sweet, and I. Schlichting. 1999. Crystal structures of myoglobin-ligand Complexes at near-atomic resolution. . *Biophys. J.* 77:2153–2174.

27. Sawaya, M. R., and J. Kraut. 1997. Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry*. 36:586–603.

28. Guss, J. M., H. D. Bartunik, and H. C. Freeman. 1992. Accuracy and precision in protein structure analysis: restrained least-squares refinement of the structure of poplar plastocyanin at 1.33 Å resolution. *Acta Crystallogr. B*. 48:790–811.

29. Jorgensen, W. D., J. Chandrasekhar, and J. D. Madura. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*. 79:926–935.

30. 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.

31. MacKerell, A. D., Jr., D. Bashford, R. L. Bellott, R. L. Dunbrack Jr., 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. I. I. I. 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.

32. Pastor, R. W., and M. Karplus. 1988. Parameterization of the friction constant for stochastic simulations of polymers. *J. Phys. Chem*. 92: 2636–2641.

33. Feller, S. E., Y. H. Zhang, R. W. Pastor, and B. R. Brooks. 1995. Constant pressure molecular dynamics simulation – the Langevin piston method. *J. Chem. Phys*. 103:4613–4621.

34. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 22:2577–2637.

35. Smith, W., and T. Forester. 1996. DL_POLY_2.0: a general-purpose parallel molecular dynamics simulation package. *J. Mol. Graph*. 14: 136–141.

36. Moritsugu, K., and J. C. Smith. 2005. Langevin model of the temperature and hydration dependence of protein vibrational dynamics. *J. Phys. Chem. B*. 109:12182–12194.

37. Izvekov, S., and G. A. Voth. 2006. Modeling real dynamics in the coarse-grained representation of condensed phase systems. *J. Chem. Phys*. 125:151101.

38. Kraulis, P. 1991. *MOLSCRIPT*: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst*. 24:946–950.