Dynamics of Large Proteins through Hierarchical Levels of Coarse-Grained Structures

PEMRA DORUKER,[1] ROBERT L. JERNIGAN,[2] IVET BAHAR[3]

[1]*Chemical Engineering Department and Polymer Research Center, Bogazici University, Bebek 80815, Istanbul, Turkey*

[2]*Molecular Structure Section, Laboratory of Experimental and Computational Biology, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-5677*

[3]*Center for Computational Biology & Bioinformatics, and Department of Molecular Genetics & Biochemistry, School of Medicine, University of Pittsburgh, Pennsylvania 15213*

Received 26 February 2001; Accepted 15 May 2001

Abstract: Elastic network models have been successful in elucidating the largest scale collective motions of proteins. These models are based on a set of highly coupled springs, where only the close neighboring amino acids interact, without any residue specificity. Our objective here is to determine whether the equivalent cooperative motions can be obtained upon further coarse-graining of the protein structure along the backbone. The influenza virus hemagglutinin A (HA), composed of $N = 1509$ residues, is utilized for this analysis. Elastic network model calculations are performed for coarse-grained HA structures containing only $N/2$, $N/10$, $N/20$, and $N/40$ residues along the backbone. High correlations (>0.95) between residue fluctuations are obtained for the first dominant (slowest) mode of motion between the original model and the coarse-grained models. In the case of coarse-graining by a factor of $1/40$, the slowest mode shape for HA is reconstructed for all residues by successively selecting different subsets of residues, shifting one residue at a time. The correlation for this reconstructed first mode shape with the original all-residue case is 0.73, while the computational time is reduced by about three orders of magnitude. The reduction in computational time will be much more significant for larger targeted structures. Thus, the dominant motions of protein structures are robust enough to be captured at extremely high levels of coarse-graining. And more importantly, the dynamics of extremely large complexes are now accessible with this new methodology.

 J Comput Chem 23: 119–127, 2002

Key words: Gaussian network model; anisotropic fluctuations; vibration dynamics; collective motions; influenza virus hemagglutinin

Introduction

It is possible to explore the molecular motions of proteins around the folded state by fully atomistic molecular dynamics (MD) simulations and normal mode analysis (NMA).[1–3] However, these techniques become computationally ineffective for the investigation of large quaternary structures and biomolecular complexes due to the increase in the system size. In this respect, coarse-grained approaches have proven to be useful in elucidating the functionally important collective motions of large proteins.[4–8]

The representation of protein structures as elastic networks is a coarse-grained approach that has been effective in predicting the fluctuation dynamics of proteins around their native conformation.[9–11] In this model, named the Gaussian network model (GNM), the α-carbons of the folded protein are chosen as the nodes of the elastic network. And the springs connect each node to all other neighboring nodes that are located within a cutoff distance encompassing at least the first coordinate shell around each residue.[12] Following Tirion's original work,[13] all springs are assigned the same harmonic force constant γ, whose absolute value can be adjusted by comparison with experiment. GNM efficiently determines the magnitude of residue fluctuations in the slow collective modes that are relevant to protein function. The fluctuations are assumed to be isotropic, and give rise to $N - 1$ independent modes of motion for a protein composed of N residues. Unlike simulations, the GNM yields an analytical solution without sampling inaccuracies. The major utility of the GNM lies in its application to large biopolymeric systems, whose MD simulations are simply not feasible. Previous studies have shown that the GNM reproduces very closely the X-ray crystallographic Debye-Waller factors,[4, 9–11, 14–17] the H/D exchange free energies,[18] and the or-

Correspondence to: R. L. Jernigan; e-mail: Robert_Jernigan@nih.gov

der parameters from NMR-relaxation measurements.[19] Also, Tama et al.[20] have shown that normal modes of proteins can be well reproduced by considering blocks of six amino acids at a time.

Coarse graining of structures has its origin in polymer computations[21] where a chain segment of multiple bonds is chosen to be equivalent to one element of an approximate chain model having simpler properties. The most common of these equivalent chain models is the freely jointed chain in which a single bond having no correlations with its neighbors is utilized to represent a larger number of actual bonds having a similar level of flexibility. This freely jointed bond has no resistance from either bond angle or torsion angle restraints. Such models have also been employed to represent crosslinked elastic networks. The motivations for these equivalences differ from that employed in the present work. In these previous coarse-grained equivalent chain models, the aim has been to match the behavior of the model chains. In the present case, for protein dynamics, the purpose is more pragmatic—simply to make accessible calculations on larger structures than would otherwise be impossible.

A recent enhancement of the GNM is the anisotropic network model (ANM), which takes into account the anisotropy of positional fluctuations.[22] Here, the fluctuation vectors, in addition to the magnitudes, are determined for each residue. Thus, $3N - 6$ modes of motion are obtained for a 3D structure of N residues. As a result, with ANM calculations the mechanism of collective motions can be elucidated for quaternary structures such as influenza virus A hemagglutinin.[23] A direct comparison of MD simulation results with GNM and ANM has validated that the collective modes can be effectively determined with these elastic network models.[24]

In this study, we will apply ANM to 3D structures that have been further coarse-grained along the protein backbone. We will show that the dominant motions can still be effectively extracted from highly (extremely) coarse-grained structures. This result points out the possible application of this method to even larger quaternary structures. In the next section, we will introduce the methodology of ANM and the different levels of coarse-graining applied to the 3D structure of the specific protein (HA). Then, the comparison of results from highly coarse-grained and original all-residue calculations will be illustrated. The last section discusses the implications and utility of coarse-grained ANM calculations.

Method

Anisotropic Network Model (ANM)

This is an analytical model developed as an extension of the Gaussian Network Model (GNM) for treating the fluctuations of biomolecular structures. It takes account of the anisotropy of fluctuations; whereas GNM assumes all fluctuations to be isotropic. The major advantage of ANM over GNM is that the three components of the fluctuation vectors ($\Delta\mathbf{R}_i$), are computed with the ANM, while GNM predicts only mean-square fluctuation amplitudes $\langle(\Delta\mathbf{R}_i)^2\rangle$. Knowledge of fluctuation vectors permits us to construct—and explicitly view—pairs of alternative conformations sampled by the action of individual modes, simply by adding the fluctuation vectors $\pm\Delta\mathbf{R}_i$ to the equilibrium (native state) position vectors. The total potential of a structure of N interaction sites is expressed in the ANM as

$$V = (\gamma/2)\Delta\mathbf{R}^{\mathrm{T}} H \Delta\mathbf{R} \tag{1}$$

where $\Delta\mathbf{R}$ is a $3N$-dimensional vector of the fluctuations $\Delta\mathbf{R}_i$ in the position vectors $\mathbf{R}_i$ of the individual sites ($1 \le i \le N$), $\Delta\mathbf{R}^{\mathrm{T}}$ is its transpose, and H is the Hessian matrix composed of the second derivatives of the potential

$$V = (\gamma/2)\sum_i \sum_j h(r_c - R_{ij})(\Delta\mathbf{R}_j - \Delta\mathbf{R}_i)^2 \tag{2}$$

Here the summations are performed over all interaction sites, $h(x)$ is the Heavyside step function [$h(x) = 1$ if $x \ge 0$, and zero otherwise), R_{ij} is the distance between sites i and j, and r_c is the cutoff distance defining the range of direct interactions. H is conveniently expressed as a function of N^2 submatrices (or super elements) $\mathbf{H}_{ij}$ of the form

$$\mathbf{H}_{ij} = \begin{bmatrix} \partial^2V/\partial X_i\partial X_j & \partial^2V/\partial X_i\partial Y_j & \partial^2V/\partial X_i\partial Z_j \\ \partial^2V/\partial Y_i\partial X_j & \partial^2V/\partial Y_i\partial Y_j & \partial^2V/\partial Y_i\partial Z_j \\ \partial^2V/\partial Z_i\partial X_j & \partial^2V/\partial Z_i\partial Y_j & \partial^2V/\partial Z_i\partial Z_j \end{bmatrix} \tag{3}$$

Here X_i, Y_i, and Z_i are the components of $\mathbf{R}_i$. Note that $\partial^2V/\partial X_i\partial Y_j = -\partial^2V/\partial X_j\partial Y_i = -\gamma(X_j - X_i)(Y_j - Y_i)/R_{ij}^2$ for $i \ne j$, and $\partial^2V/\partial X_i\partial Y_i = \gamma\sum_j(X_j - X_i) \times (Y_j - Y_i)/R_{ij}^2$.

In conformity with the GNM approach, the force constant, γ, is taken to be identical for all bonded and nonbonded pairs, including even disulfide bridges. γ represents the curvature of the interaction potential near the minimum point; it does not depend on the radial separation, at the energy minimum. The adoption of a fixed γ value for all interactions, is equivalent to assuming that the same curvature holds for all pairs near the energy minimum, which has been confirmed to be a valid approximation in previous GNM studies.

The cross-correlations between the fluctuations of sites i and j are found from

$$\begin{aligned} &\langle\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j\rangle \\ &\quad = (1/Z_N)\int (\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j)\exp\{-V/k_BT\}\, d\{\Delta\mathbf{R}\} \\ &\quad = (3k_BT/\gamma)\,\mathrm{tr}\left[H^{-1}\right]_{ij} \end{aligned} \tag{4}$$

where k_B is the Boltzmann constant, T is the absolute temperature in degrees Kelvin), Z_N is the conformational partition function, and $\mathrm{tr}[H^{-1}]_{ij}$ designates the trace of the ijth submatrix $[H^{-1}]_{ij}$ of $\mathbf{H}^{-1}$. $\langle\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j\rangle$ can be expressed as a sum over the contributions $[\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j]_k$ of the $3N - 6$ individual modes, as $\langle\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j\rangle = \sum_k[\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j]_k$. The contribution of the kth mode is

$$[\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j]_k = (3k_BT/\gamma)\,\mathrm{tr}\left[\lambda_k^{-1}\mathbf{u}_k\mathbf{u}_k^{\mathrm{T}}\right]_{ij} \tag{5}$$

where λ_k is the kth nonzero eigenvalue and $\mathbf{u}_k$ is the corresponding eigenvector. The eigenvalues are representative of the frequencies of the individual modes, and the eigenvectors describe their shape, i.e., their effect on the positions of the N interaction sites. The eigenvalues are usually organized in ascending order (omitting the six zero eigenvalues), such that λ_1 refers to the lowest frequency,

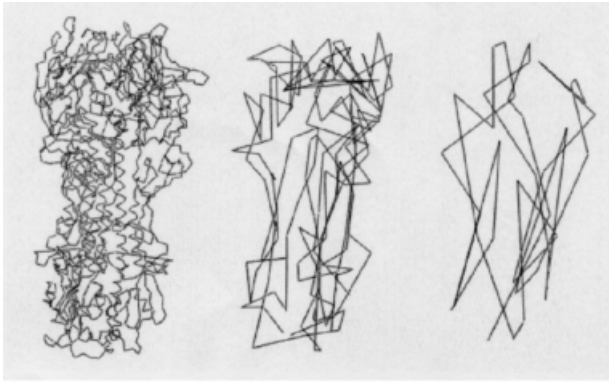


Figure 1. X-ray structure of hemagglutinin (HA) showing α-carbon atoms of (a) all residues ($N = 1509$), and coarse grained with (b) $N/10$ and (c) $N/40$ residues.

also called the global, mode of motion, and $[\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j]_1$ is the correlation driven by this mode. Likewise, $[(\Delta\mathbf{R}_i)^2]_1$ is the mean-square (ms) fluctuation in the position of site i induced by mode 1. The slowest modes usually play a dominating role in the collective dynamics of the structure. They are the only surviving or operating modes at long times, hence their implications for biological function.

Protein

Influenza virus hemagglutinin A (HA) is an integral membrane glycoprotein, which is involved in the binding of the virus to target cells and in the fusion of viral and endosomal membranes at the low pH environment. The X-ray structure of HA has been determined[25] and refined[26] by Wiley and coworkers, as shown in Figure 1a. HA is a cylindrically shaped homotrimer of 1509 residues. The monomers are composed of two polypeptides chains each, HA1 (residues 1–328) and HA2 (residues 329–503) that are linked by two disulfide bridges. The three monomers are assembled into a central coiled coil that forms the stem-like domain, and three globular heads containing the receptor binding sites. The fibrous stem-like domain consists of all HA2 and some HA1 residues. Each monomer is anchored in the viral membrane by a single trans-membrane peptide near the C-terminus of its HA2 chain. The globular heads are formed by the HA1 residues 116–261 folded into a jelly roll motif of eight antiparallel β-strands. The distal tips of the globular heads contain the receptor binding sites.

Coarse-Graining Procedures

A detailed study on the functional motions of hemagglutinin[23] has been performed based on the two analytical methods—GNM and ANM—discussed above. In that study, all residues' interactions were considered, the interaction sites being identified with the positions of α-carbon atoms ($N = 1509$) taken from the X-ray structure of HA. Our present study will focus on the consequences of adopting increasingly lower resolution models on the collective motions extracted by ANM. The coarse-graining is performed at a series of hierarchical levels by retaining $(N/2)$, $(N/10)$, $(N/20)$, and $(N/40)$ α-carbon atoms of the original structure. Figure 1b and c display results for individual cases of coarse-grained 3D structures for the $(N/10)$ and $(N/40)$ cases, respectively.

Here, we coarse-grain HA's structure uniformly on the basis of its primary sequence, i.e., along the chain backbone. If we sequentially number the residues of HA in the original PDB file (2HMG, reference), the residue indices of the three identical monomers would be (1–503), (504–1006), and (1007–1509). Taking $(N/10)$ structure as a specific example, nonsymmetrical coarse-graining would lead to retention of residues (1, 11, 21, . . . , 1501) for further ANM calculations. In this case, the first (N-terminal) residues of the three identical monomers that are retained are the first, eighth, and fifth residues on their primary structures sequentially. Thus, the resulting coarse-grained monomers are not identical any more, in terms of sequence or 3D structure. In symmetrical coarse-graining, we would keep the first residue of each monomer so that the resultant coarse-grained structure is still a homotrimer with identical monomers (first monomer: 1, 11, . . . , 501, second monomer: 504, 514, . . . , 1004, and third monomer: 1007, 1017, . . . , 1507). Here we indicated only one of the 10 possible subsets of residues that form $N/10$ structure. There are nine other frames of coarse-graining for $N/10$, which would be utilized by shifting the index of first retained residue from 1 through 10.

Results and Discussion

Comparison of Equilibrium Fluctuations with Experiments

Figure 2 compares the Debye–Waller factors measured by X-ray crystallography (solid curve) with those predicted by the ANM (dotted curves) for the HA trimer.[26] The three monomers exhibit practically the same behavior both in experiments and calculations. Therefore, the fluctuations of residues are presented as averages over the three monomers, although the calculations have been performed for the trimeric structure containing 1,509 residues.

The experimental Debye–Waller factors, the so-called B-factors, are related to the ms fluctuations of individual residues $(1 \leq i \leq N)$ as $B_i = (8\pi^2/3)\langle(\Delta\mathbf{R}_i)^2\rangle$. There is a single parameter in the present analytical calculations, the force constant γ, which is adjusted to obtain the best match between the average values of ms fluctuations predicted by ANM and the experimental B-factors. The force constant is a measure of the strength of intramolecular potentials that stabilize the native fold. A typical force constant of 0.95 kcal/(mol Å^2) is obtained for HA using a cutoff of 13 Å for interresidue interactions. This specific cutoff value and the corresponding force constant fall within the range suggested by other recent ANM studies.[22] It may eventually become possible to evaluate these constants from neutron scattering experiments.[27]

The agreement between theory and experiments is noteworthy in that no residue specificity or nonlinear effects are included in the theory. Moreover, the resolution of the examined crystal structure is relatively good (3 Å), and the ms fluctuations are usually subject to even larger experimental uncertainties than the mean positions. The correlation coefficients between theory and experiment are found to be 0.79 and 0.70 for the HA1 and HA2 chains, respectively.

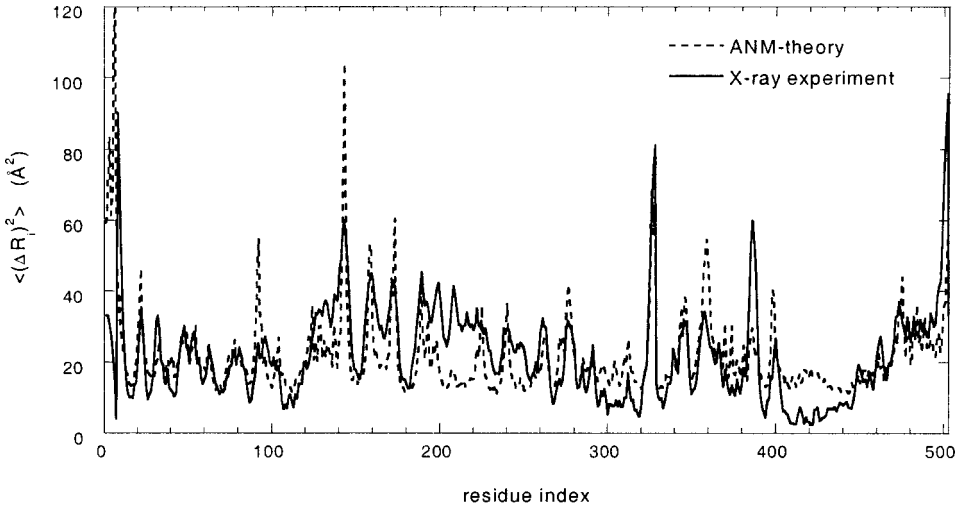


Figure 2. Comparison of experimental and theoretical (ANM) ms fluctuations as a function of residue number for HA.

Comparison of the Results from Hierarchical Coarse-Graining Levels with the Original ANM Results

ANM calculations are performed for HA structures that have been nonsymmetrically coarse-grained at higher levels. Table 1 summarizes these models containing $N = 1509$ (original), $N/2$, $N/10$, $N/20$, and $N/40$ of the total number of residues. The original cutoff radius of 13 Å necessarily must be increased for higher levels of coarse-graining so that force constants comparable to the original value of 0.95 kcal/(mol Å^2) are obtained. The slight variations in the force constants, γ, for different values of cutoff distance (Table 1) follow from the normalization of the theoretical results to match crystallographic fluctuation data. If we were to keep a constant cutoff of 13 Å for more highly coarse-grained structures, then the ANM analysis would yield more than six zero eigenvalues, and certain residues would exhibit extremely large amplitude unrealistic fluctuations.[22] The implementation of larger cutoff radii removes such physically unrealistic behavior by smoothing over the missing parts of the structure. Next, we will discuss the correlations among the ANM results obtained for the entities in Table 1.

Mean Square Fluctuations

ANM calculations for $N/2$ residues lead to almost indistinguishable ms fluctuations from the original results ($N = 1509$ residues), as shown in Figure 3a. The results are averaged over the three monomers because they exhibit almost the same behavior. The results of further coarse-graining the structure at the level of $N/10$ and $N/40$ residues are displayed in Figure 3b and c, respectively. Here the results for all three monomers are displayed, because the ms fluctuations of the three monomers are not identical due to nonsymmetric method of coarse-graining. In each panel, the ms fluctuations obtained from the original all-residue calculations are plotted for only those specific residues that are retained in the coarse-grained structures for clarity. Moreover, the correlations of coarse-grained ms fluctuations with those from original calculations are given in Table 2. These results are for a specific frame of nonsymmetrical coarse-graining starting at the first residue of HA. Other frames would yield similar correlation coefficients. As can be observed, the correlation coefficients are quite high even in the extreme case of coarse-graining, $N/40$.

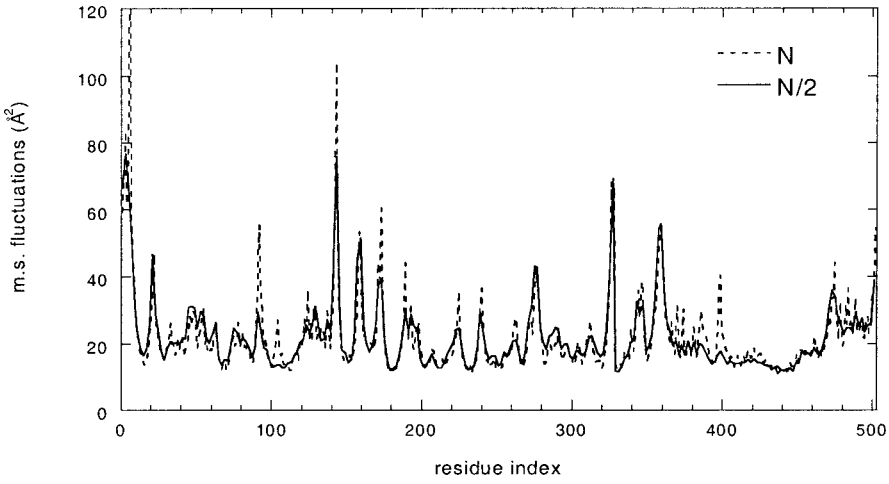
Table 1. Details of Coarse-Grained ANM Calculations.

Coarse Graining	Number of Residues	Cutoff Radius (Å)	Force Constant (kcal/mol Å^2)	Mean Coordination Number
N	1509	13	0.95	39.2
$N/2$	755	18	0.75	44.2
$N/10$	151	30	1.21	28.4
$N/20$	76	40	1.28	25.8
$N/40$	38	60	1.38	22.6

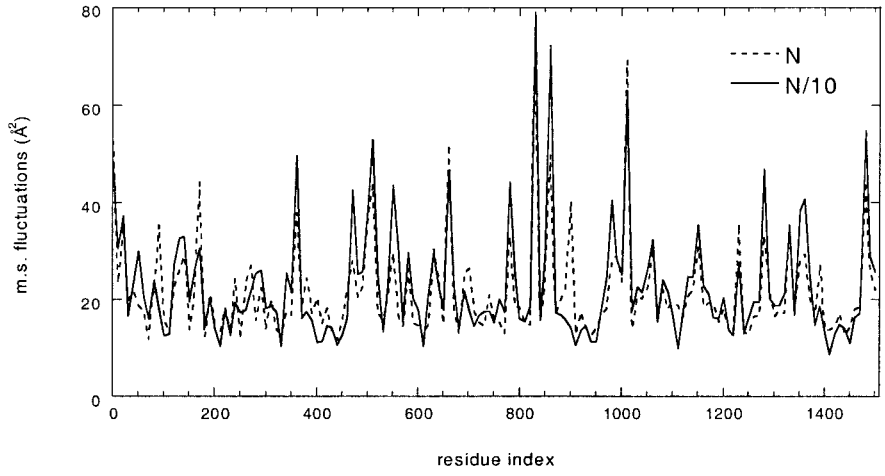
Collective Motions

The low-frequency modes, also called global modes, provide insights about the mechanisms of the cooperative conformational motions of native structures. Such motions are expected to be relevant to biological function. Our aim here is to see whether we can still extract reliable information about the global motions of HA even if its structure is extremely coarse-grained.

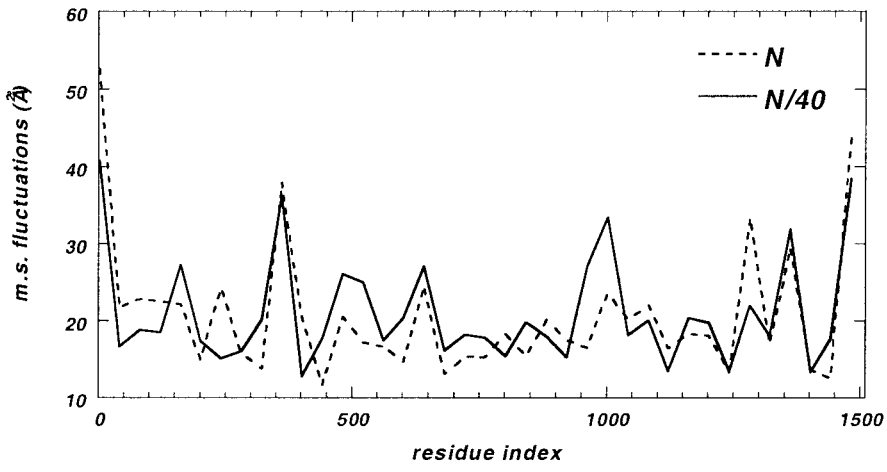
In Figure 4a, a cubic-spline fit is performed on the first (slowest) mode shape of $N/40$ and all-residue calculations. Here, all $N = 1509$ residues are considered in the fit to all-residue calculations (solid curve), and the resulting mode shape is symmetric, as expected. The curve for the model including only $N/40$ of the residues, on the other hand, is the best fitting cubic spline through the results (solid curve) for the sparse interaction sites (1509/40 of them) included in the coarse-grained model. The important point is that the nonsymmetric mode shape of the $N/40$ residues considered can still capture the important details of the first mode shape. Figure 4b displays the slowest mode shape for only those specific $N/40$ residues preserved in a single $N/40$ calculation. The fit is excellent with a correlation coefficient of 0.98. The correlation co-



(a)



(b)



(c)

Figure 3. Comparison of ms fluctuations from original all-residue ANM calculations with (a) $N/2$ (average of three monomers), (b) $N/10$, and (c) $N/40$ cases.

Table 2. Correlations of Coarse-Grained Results with the Original All-Residue Case.

	All modes	First Mode	Second Mode
$N/2$	0.94	1.00	0.96
$N/10$	0.86	0.95	0.92
$N/20$	0.82	0.98	0.96
$N/40$	0.79	0.98	0.81

efficients for the first two modes of motion are listed in Table 2 for all levels of coarse-graining. As can be observed, the slowest mode shapes show excellent fit to the original all-residue calculations, much better than the correlation for all modes.

Distribution of Eigenvalues

Figure 5 shows the fractional contribution of each mode as a function of mode number. The ordinate represents the weighted contribution of each mode to the ms fluctuations. The first three slowest modes with the smallest eigenvalues make the greatest contribution to the overall motion, and this is true regardless of the level of coarse graining. Notably these "slowest" motions usually reflect the overall shape of the structure and not any of the details. The cumulative contribution of the first three modes sum up to 0.22 in the $N/40$ case, whereas this value decreases to 0.15 for the original case, due to the longer tail of the $(3N-6)$ eigenvalue distributions. The present distributions unambiguously demonstrate that the dispersion of the most dominant (slowest) modes is almost unchanged in the models with different levels of coarse-graining. Adoption

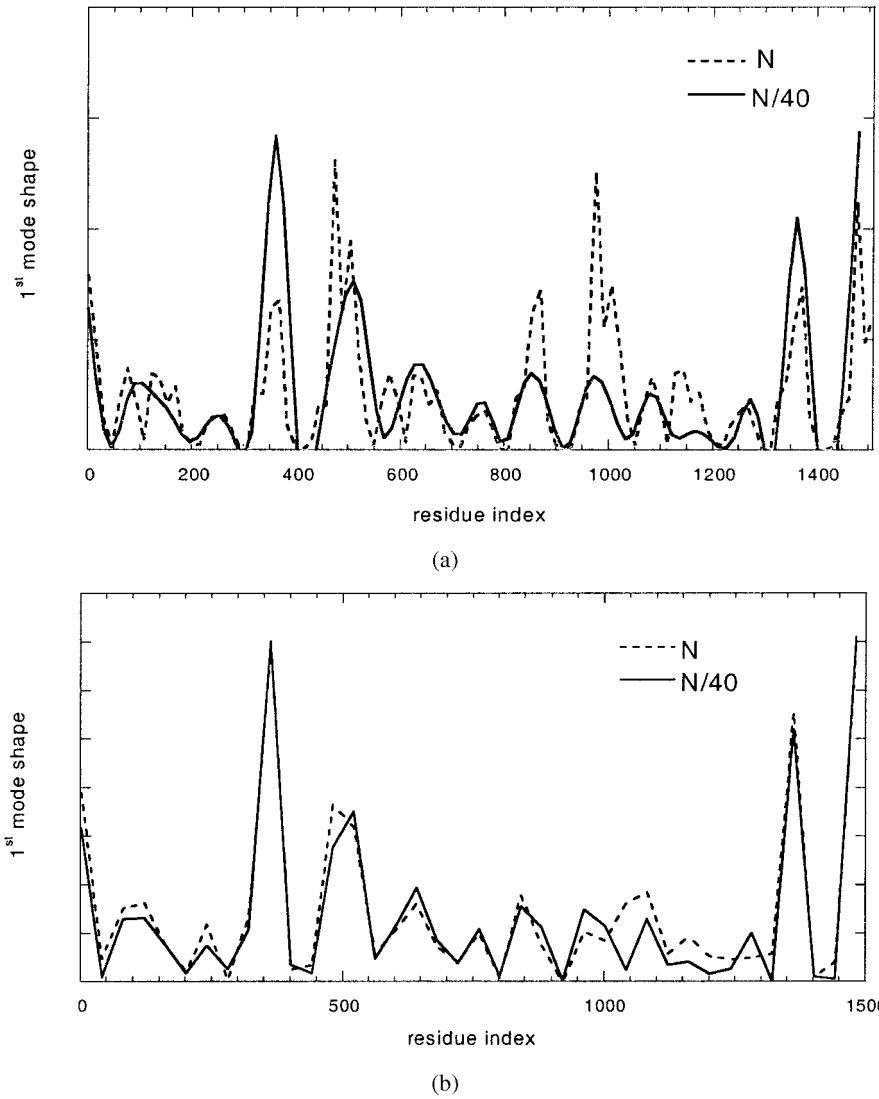


Figure 4. Comparison of distribution of fluctuations in the first mode of original ANM calculations with that for $N/40$. (a) The distributions are plotted after performing cubic-spline fits for both curves. (b) Fluctuations are shown only for the retained $N/40$ residues for both cases.

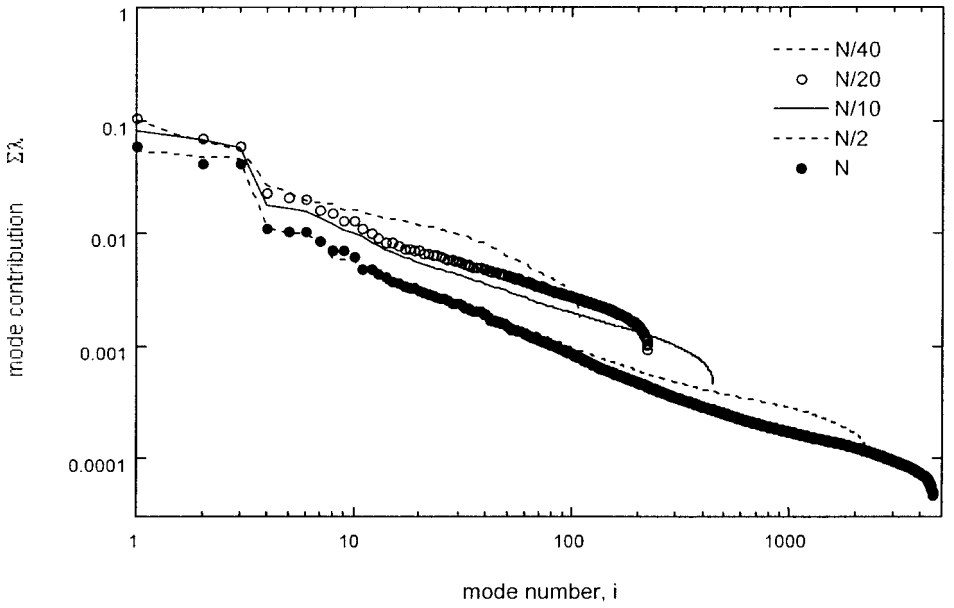


Figure 5. Fractional contribution of each mode to the total motion against the mode number for all-residue and coarse-grained ANM calculations.

of higher resolution models simply introduces more modes at the high-frequency end of the spectrum, while having a minimal effect on the characteristics of the lowest frequency modes.

Reconstruction of All-Residue Slowest Mode Shape from N/40 Residue Calculations

The excellent correlation among the actual $N/40$ residues of the extremely coarse-grained and the corresponding elements of the original structure, shown in Figure 4b, indicates that even the $N/40$ residue representation is sufficient to obtain the amplitudes of fluctuations in the global modes of HA. It would be interesting if we could reconstruct the original symmetric mode shape for all $N = 1509$ (or $N/2$) residues by shifting the frame of the retained $N/40$ residues to calculate the fluctuation amplitudes for the missing residues. For the $N/40$ case, we need to perform ANM calculations with the addition of different frames to restore the complete mode shapes. Figure 6 displays such a reconstructed first mode shape for $N/2$ residues, giving a comparison of original vs. reconstructed $N/2$ results, based on the repetitive use of the $N/40$ model. Figure 6 is obtained from symmetric coarse-graining of the monomers, and the reconstructed $N/2$ mode shape. The correlation coefficients are 0.84 and 0.73 for the smoothed and non-smoothed data points, respectively. Here, smoothing is performed

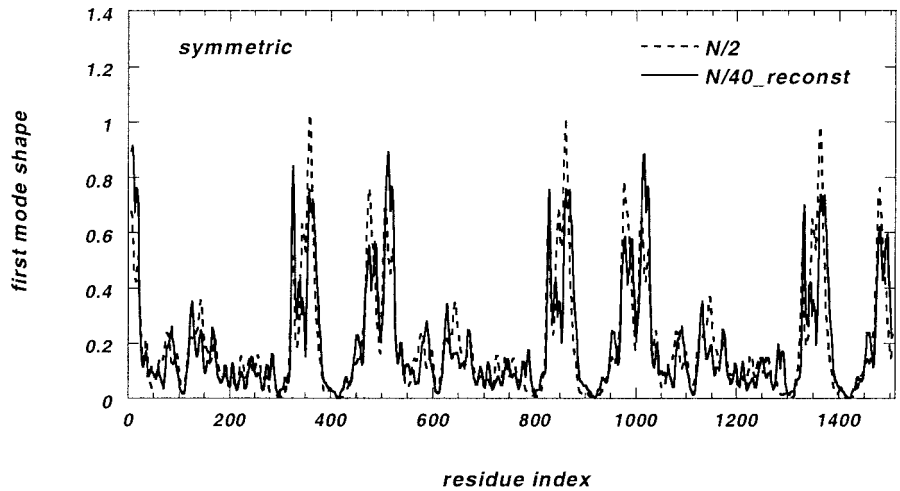


Figure 6. Reconstruction of the $N/2$ first mode shape by using 20 different frames of $N/40$ calculations, using symmetric coarse-graining.

by averaging over the fluctuations in windows of five successive points.

The high correlation between the two sets of results presented in Figure 6 shows that even adopting an extremely high level of coarse-graining very closely reproduces the global mode shape. Calculations indicate that this mode leads to a global twisting of the trimeric molecule about its axis of cylindrical symmetry (not shown). In particular, the base portion of the stem-like domain is highly mobile in this mode. Such a conformational mobility has been ascribed to the translocation and aggregation of HA molecules preceding the opening of a viral membrane pore via a biaxial deformation mechanism.[23] Experimental studies suggest that the collection of three or more HA molecules is required for membrane pore opening.[28]

Figure 7 illustrates the mechanism of motion of the second slowest mode for (a) the all-residue and (b) $N/10$ cases. This mode of motion indicates bending of the whole molecule around the same hinge site, at both levels of coarse-graining. Such a global bending has been proposed by several groups as a major mechanism underlying the association of the viral and endosomal membranes prior to their fusion.[29, 30]

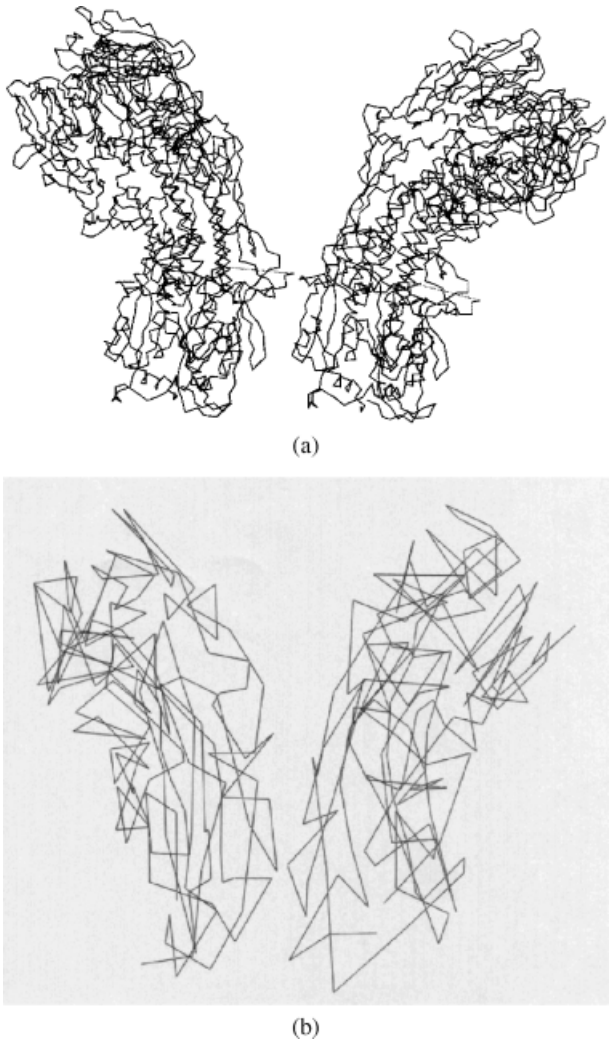


Figure 7. Mechanism of second mode of motion from (a) all-residue, and (b) $N/10$ calculations. The two structures shown represent the conformations that HA visits in the opposite directions according to the motion of the second mode.

Conclusion

The anisotropic network model has recently been introduced for the investigation of collective motions of proteins.[22, 24] ANM assumes a simple elastic network structure, formed by springs that connect close α-carbon atoms in the 3D structure of proteins. In this study, we have investigated the consequences of retaining only a fraction of the residues (α-carbons) along the protein's backbone on the collective motions extracted by the ANM. The coarse-grained ANM calculations have been performed on influenza virus hemagglutinin, a homotrimeric enzyme composed of 1509 residues. The results indicate that almost the same shapes for the slowest modes of motion are obtained, even when including only 1/40th of the complete set of residues in HA. Correlation coefficients between the ms fluctuations of residues evaluated from the original all-residue model and those found from coarse-grained models lie in the ranges of 0.95–0.98 and 0.81–0.96 for the first and second slowest modes, respectively, where the upper and lower limits correspond to the adoption of $N/40$ and $N/2$ residues, respectively, in the coarse-grained models. These high correlations unambiguously indicate that the low-frequency collective motions of proteins are robust enough to be almost identically extracted from even extremely coarse-grained descriptions of the 3D structure.

The analysis of the different mode shapes demonstrates that including more detailed descriptions of the 3D structure essentially adds modes with higher frequencies without changing the global mode shapes. Thus, the most cooperative motions of the folded structure, which are also the functionally most important motions, are estimated with high accuracy at the cost of only losing information on the high-frequency mode shapes.

Another interesting finding is that we can reconstruct the original mode shape by shifting successively the frame of the retained $N/40$ residues to calculate the fluctuation amplitudes for all missing residues. The correlation for this reconstructed first mode shape with the original case is 0.73, while the computational time is reduced by more than two orders of magnitude. This result is highly significant because it now permits us to investigate the global motions of extremely large structures by reducing the total number of residues, by repeating smaller calculations by simply shifting the reading frame. Thus, the dynamics of extremely large complexes now become accessible with this new method.

References

1. McCammon, A.; Harvey, S. C. Dynamics of Proteins and Nucleic Acids; Cambridge University Press: Cambridge, 1987.
2. Brooks, C. L.; Karplus, M.; Pettitt, B. M. Adv Chem Phys 1988, 71, 1.
3. Kitao, A.; Go, N. Curr Opin Struct Biol 1999, 9, 164.
4. Bahar, I.; Erman, B.; Jernigan, R. L.; Atilgan, A. R.; Covell, D. G. J Mol Biol 1999, 285, 1023.
5. Bahar, I.; Jernigan, R. L. Biochemistry 1999, 38, 3478.

6. Hinsen, K.; Hinsen, K.; Thomas, A.; Field, M. Proteins 1998, 33, 417; Proteins 1999, 34, 369.
7. Abseher, R.; Nilges, M. Proteins 2000, 39, 82.
8. Hinsen, K.; Kneller, G. R. J Chem Phys 1999, 111, 10766.
9. Bahar, I.; Atilgan, A. R.; Erman, B. Fold Des 1997, 2, 173.
10. Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. Phys Rev Lett 1998, 80, 2733.
11. Haliloglu, T.; Bahar, I.; Erman, B. Phys Rev Lett 1997, 79, 3090.
12. Jernigan, R. L.; Bahar, I. Curr Opin Struct Biol 1996, 6, 195.
13. Tirion, M. M. Phys Rev Lett 1996, 77, 1905.
14. Bahar, I.; Jernigan, R. L. J Mol Biol 1998, 281, 871.
15. Demirel, M. C.; Atilgan, A. R.; Jernigan, R. L.; Erman, B.; Bahar, I. Protein Sci 1998, 7, 2522.
16. Jernigan, R.L.; Demirel, M. C.; Bahar, I. Int J Quantum Chem 1999, 75, 301.
17. Keskin, O.; Jernigan, R. L.; Bahar, I. Biophys J 2000, 78, 2093.
18. Bahar, I.; Wallqvist, A.; Covell, D.; Jernigan, R. L. Biochemistry 1998, 37, 1067.
19. Haliloglu, T.; Bahar, I. Proteins 1999, 37, 654.
20. Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y.-H. Proteins 2000, 41, 107.
21. Flory, P. J. Statistical Mechanics of Chain Molecules; Interscience-Wiley Publishers: New York, 1969.
22. Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Biophys J 2001, 80, 505.
23. Isin, B.; Doruker, P.; Bahar, I. Biophys J, submitted.
24. Doruker, P.; Atilgan, A. R.; Bahar, I. Proteins 2000, 40, 512.
25. Wilson, I. A.; Skehel, J. J.; Wiley, D. C. Nature 1981, 289, 366.
26. Weis, W. I.; Brünger, A. T.; Skehel, J. J.; Wiley, D. C. J Mol Biol 1990, 212, 737.
27. Zaccai, G. Science 2000, 288, 1604.
28. Bentz, J. Biophys J 2000, 78, 227.
29. Tatulian, S. A.; Hinterdorfer, P.; Barber, G.; Tamm, L. K. J Cell Biol 1995, 106, 5514.
30. White, J. M.; Hoffman, L. R.; Arevalo, J. H.; Wilson, I. A. In Structural Biology of Viruses; Chiu, W.; Burnett, R. M.; Garcea, R. L., Eds.; Oxford University Press: New York, 1997, p. 80.