Protein elastic network models and the ranges of cooperativity

Lei Yang[a,b,c], Guang Song[a,c,d], and Robert L. Jernigan[a,b,c,1]

[a]Bioinformatics and Computational Biology Program, [b]Department of Biochemistry, Biophysics and Molecular Biology, [c]L. H. Baker Center for Bioinformatics and Biological Statistics, and [d]Department of Computer Science, Iowa State University, Ames, IA 50011

Edited by Ken A. Dill, University of California, San Francisco, CA, and approved June 3, 2009 (received for review February 26, 2009)

Elastic network models (ENMs) are entropic models that have demonstrated in many previous studies their abilities to capture overall the important internal motions, with comparisons having been made against crystallographic B-factors and NMR conformational variabilities. ENMs have become an increasingly important tool and have been widely used to comprehend protein dynamics, function, and even conformational changes. However, reliance upon an arbitrary cutoff distance to delimit the range of interactions has presented a drawback for these models, because the optimal cutoff values can differ somewhat from protein to protein and can lead to quirks such as some shuffling in the order of the normal modes when applied to structures that differ only slightly. Here, we have replaced the requirement for a cutoff distance and introduced the more physical concept of inverse power dependence for the interactions, with a set of elastic network models that are parameter-free, with the distance cutoff removed. For small fluctuations about the native forms, the power dependence is the inverse square, but for larger deformations, the power dependence may become inverse 6th or 7th power. These models maintain and enhance the simplicity and generality of the original ENMs, and at the same time yield better predictions of crystallographic B-factors (both isotropic and anisotropic) and of the directions of conformational transitions. Thus, these parameter-free ENMs can be models of choice whenever elastic network models are used.

B factors | conformational entropy

Protein dynamics can provide important insights into protein function. Most proteins carry out their functions through conformational changes of the structures. In X-ray structures, the information about thermal motions is provided by the Debye-Waller temperature factors or B-factors, which are proportional to the mean square fluctuations of atom positions in a crystal. Thus, accurate predictions of crystalline B-factors offer a good starting point for understanding the functional dynamics of proteins.

A number of computational and statistical approaches has been proposed to predict protein B-factors from protein sequence (1–7), atomic coordinates (8–13), and electron density maps (14). The atomic coordinate-based methods such as molecular dynamics (MD) (15–18) and normal mode analysis (NMA) (19–22) are computationally expensive, and thus, in the past decade, the elastic network model (ENM), with NMA, using a single parameter harmonic potential, usually coarse-grained, has been widely used for studying protein dynamics including B-factor predictions. The ENM for isotropic fluctuations is called the Gaussian network model (GNM) (23, 24), where only the magnitudes of the fluctuations are considered. Its anisotropic counterpart, where the directions of the collective motions are also examined, is called the anisotropic network model (ANM) (25), and these can be compared with the experimental anisotropic temperature factors (26–29), but generally these comparisons are worse than when atomic representations are used (30).

Previous studies have shown that in many cases, GNM-predicted B-factors are in quite good agreement with the experimental B-factors determined by crystallographers (8–11). Kundu et al. (10) studied 113 X-ray protein structures and found that GNM is able to predict the experimental B-factors well, yielding an average correlation between prediction and experiment of ≈0.59. The results of Sen et al. showed that the correlations between experimental B-factors and the GNM-predicted values are quite similar at either coarse-grained or atomic levels (12), which is consistent with its being overall an entropy model. The ANM can also be used for B-factor predictions, although, in reproducing the isotropic B-factors, it was noted that ANM generally performs slightly worse than GNM (10).

In the ENM, a parameter—the cutoff distance is used to define the connections and placement of springs between residue pairs. Only the residue pairs within a given cutoff distance are considered to be connected to one another. The cutoff in GNM is generally set at 7–8 Å. Kundu et al. (10) showed that 7.3 Å gave the best results for B-factor prediction for most of the proteins in the dataset. In ANM, the optimal cutoff is usually larger, ≈13–15 Å (25). Besides the parameter for cutoff distance, the spring constant is another parameter. However, in current practice, most researchers use a uniform spring constant for all connected residue pairs, and this spring constant is used simply to scale overall the range of B-factors so the spring constant is not actually a fundamental parameter in the model in the same sense that the cutoff distance is. In some studies (31, 32), stronger spring constants have been assigned for rigid elements in the structure. Erman empirically fitted the experimental B-factors by iteratively changing the spring constants of the Kirchoff matrix for GNM (33). One concern about this approach is that the results of such a fitting (or overfitting) may not have a real physical basis, even though the correlations achieved between experiment and prediction can be excellent. The Hinsen (34–37) and the Phillips (29) groups have used a distance-dependent force constant in their studies of protein dynamics. They proposed a force constant exhibiting an exponential decay over the distance between a pair of interacting atoms. The Hinsen model has some physical basis, but, it still does not fully avoid using an arbitrary distance parameter. In his model, the force constant k is defined as an exponential decay of distance r: $k(r) = c\ \exp(-r^2/r_0^2)$, where r_0 is an arbitrary parameter (in Hinsen's studies, r_0 was taken as 3.0 Å). There are several problems with using a cutoff distance. First, the values taken are arbitrary to a significant extent. Second, the optimal values vary somewhat for different proteins. Last, there is a discontinuity in the residue-residue interaction potential at the cutoff distance.

Author contributions: L.Y., G.S., and R.L.J. designed research; L.Y. and G.S. performed research; L.Y., G.S., and R.L.J. analyzed data; and L.Y., G.S., and R.L.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: jernigan@iastate.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0902159106/DCSupplemental.

There can be model artifacts such as the same mode for 2 only slightly different structures being significantly different with the corresponding mode indices shuffled.

Recently Lin et al. proposed a weighted contact number (WCN) model (38) for B-factor predictions. Using the WCN model, they found a good correlation between the calculated WCN and the experimental B-factors. The main difference between the WCN and previous contact number-based models (39, 40) is that the WCN model does not use any cutoff distance, but instead considers the contacts between all pairs of residues. The effect of a contact between a pair of residues is weighted by the inverse of their square distance. One advantage of the WCN model is that it avoids choosing an arbitrary cutoff distance. Inspired by the simple idea of the WCN model, here we propose a parameter-free elastic network model (pfENM). The idea is to assume that there is an interaction between all residue pairs, and to make their interaction strengths inversely proportional to the square distance of their separation. In this way, we can avoid using an arbitrary cutoff distance in the models and therefore avoid the problems outlined above. And, we learn that this refined ENM actually shows significant improvements over the original ENMs. In our parameter-free models, we use an inverse 2nd power (square distance) to model the spring constants. Past work had used an inverse 6th power spring (29, 36) or a spring of exponential form (34–37). However, our results show that our pfENM with inverse 2nd power springs clearly outperforms them all in B-factor predictions (see Table S1 and Table S2).

While a cutoff value was used in the original atomic elastic network model of Tirion (41), the use of a cutoff distance may also have been supported conceptually by the typical reliance upon a cutoff distance for defining coarse-grained contact energies derived from a set of known structures (42, 43). In the latter case the use of a cutoff distance is an attempt to avoid some of the details of interactions and to capture average effects. However, when computing molecular motions, the averaging represented in the use of a cutoff distance is not likely appropriate because it will inevitably eliminate some of the structural details, and ignore the characteristic physical nature of interactions between atoms that should diminish with increasing separation, since the springs should weaken to represent less cohesion between more distant parts. In fact, we find that the use of a cutoff distance, in addition to the use of the distance power of -2, always yields worse fits for the B-factor data (See Table S3 and Table S4).

We test these parameter-free models (pfGNM and pfANM) using a large nonredundant dataset containing 1,220 X-ray structures and a dataset that contains 341 high-resolution structures having anisotropic B-factor entries. Our results show that for B-factor predictions, the pfENM models have an even better performance than the original ENM. Since these parameter-free models also maintain the simplicity of ENMs, we thus propose that in the future, these parameter-free ENMs ought to be the models of choice whenever elastic network models are required.

Results

Isotropic B-Factor Predictions – GNM. In this section, the GNM and pfGNM are compared for scalar isotropic B-factor predictions. The correlation between the computed B-factors and the experimental values are calculated for the 1,220 protein structures in our dataset. We have performed investigations to determine the best inverse power to use, and the detailed results are given in Tables S1–S4.

In the work of Kundu et al. (10), the mean of the correlations between experimental B-factors and the GNM predicted values was 0.59. For our dataset, we obtain a mean correlation of 0.55 using the same GNM, and 0.60 when using pfGNM, both of which are similar to their results. Here, we point out that we must be cautious in interpreting these "mean values" since they are dataset dependent. With our dataset, the mean correlation values from pfGNM (0.60) and GNM (0.55) clearly indicate that pfGNM performs significantly better than GNM.

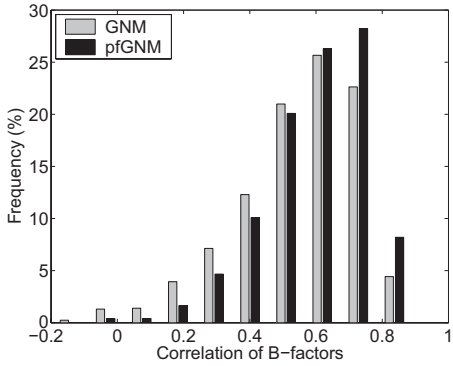

Fig. 1. Distribution of correlations between experimental and the computed B-factors from pfGNM and GNM. It can be seen clearly that pfGNM yields higher correlations than GNM. Approximately 73% of proteins show better agreement with pfGNM than with GNM.

A more detailed comparison of the correlations is found in the histograms of the correlation distributions. See Fig. 1, which shows histograms of the distribution of correlations between experimental B-factors and the computed values for GNM and pfGNM. The percentage of proteins having correlation values >0.5 is $\approx$75% using pfGNM, which is better than the 67% found by using the original GNM with a cutoff distance. For $\approx$73% of proteins, pfGNM yields better results than GNM.

To further compare the results from GNM and pfGNM, we calculate the percentage of B-factor correlations improved by using pfGNM over GNM. Fig. 2 displays the improvement with pfGNM over the original GNM for B-factor prediction. For $>$57% of proteins, the pfGNM correlation is at least 5% better than the GNM correlation.

Further Isotropic B-Factor Predictions. Similar comparisons are performed between pfANM and ANM. Fig. 3 shows the histograms of the distribution of correlations between experimental B-factors and ones computed using ANM and pfANM. Similar to the comparison between pfGNM and GNM, the mean correlation values from pfANM (0.55) and ANM (0.46) indicate that pfANM peforms significantly better than ANM. The percentage of proteins having correlation values >0.5 is $\approx$65% using pfANM, which is better than that (50%) using the original ANM with a cutoff distance. For $\approx$83% of proteins, pfANM yields better results than ANM.

The improvement of pfANM over ANM is further shown by the increased percentage of B-factor correlations in Fig. 4. For

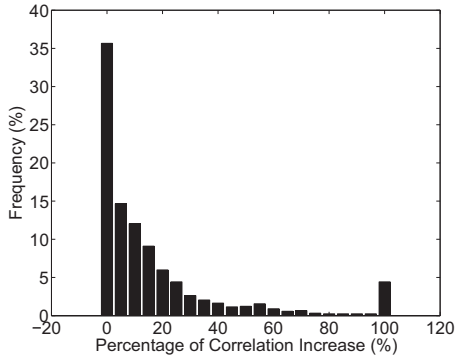

Fig. 2. Percentage (%) of correlation improvement in B-factors by using pfGNM in comparison with GNM. (The last bar represents all of the cases where improvements are 100% or higher).

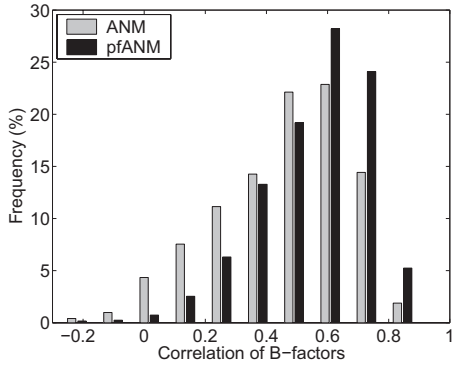

Fig. 3. Correlations between experimental and calculated B-factors from pfANM and ANM. It is clearly seen that pfANM does significantly better than ANM. Approximately 83% of proteins show improvements with pfANM over ANM.

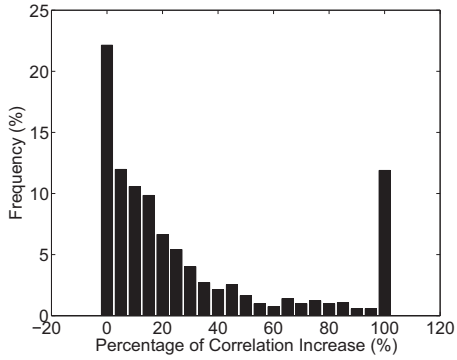

Fig. 4. Percentage (%) of correlation increase in B-factors by using pfANM over ANM. (The last bar represents all cases where the increase is greater than or equal to 100%).

>72% of proteins, the pfANM correlation is at least 5% better than the ANM correlation.

Anisotropic B-Factor Predictions. To test the models more thoroughly, in this section we will look at the performance of these parameter-free ENMs (pfENM) for anisotropic B-factor prediction. In comparison with isotropic B-factors, anisotropic B-factors provide not only the magnitudes of mean square fluctuations, but also information about the directions of the fluctuations. In PDB files, the anisotropic B-factors are given as 6 numbers that represent the anisotropic tensor of the fluctuations. Hence the anisotropic B-factors provide more data about the fluctuations. These are only available for some of the higher resolution crystal structures. Fortunately, because of improving experimental structure quality, the number of such proteins with anisotropic B-factors has been increasing rapidly.

In recent work (26), we applied the ANM to study the anisotropic fluctuations of 341 proteins. Here, we carry out a similar study using pfANM. Again, we will see that pfANM performs significantly better than ANM.

The results of anisotropic B-factor predictions from both pfANM and ANM are listed in Table 1. The experimental and calculated anisotropic B-factor tensors, which can be represented by ellipsoids, are compared [see (26) for details]. Specifically, the first and second anisotropies (κ and χ) of the experimental anisotropic B-factors are compared with those calculated from the models (pfANM or ANM). The first/second anisotropy is defined as the ratio of the length of the shortest (or second shortest) axis to the longest axis. The first column in Table 1 shows the mean difference between the experimental first anisotropy and the calculation. The means are taken over all proteins, and for each protein, over all its residues (each residue is represented by its alpha carbon). Similarly for the second column, the mean difference is for the second anisotropy. It is seen in Table 1 that pfANM clearly does better than ANM in predicting anisotropies even though the gain is small. The mean θ and ϕ values in the third and fourth column describe how well the directions of the ellipsoids match between experiment and calculation. Specifically, θ is the angle between the longest axes of the ellipsoids (which represent the anisotropic B-factor tensors of experiment and calculation), and ϕ is the rotation angle required to align the second longest axes of the 2 ellipsoids once their longest axes are aligned (26). Therefore, a perfect prediction would render both θ and ϕ as 0. From the Table 1, we see that neither pfANM nor ANM does well in this regard, for possible reasons that were discussed in (26). However, the point of interest here is that pfANM does slightly better than ANM.

A more rigorous measure for comparing anisotropic tensors are the correlation coefficient of tensors proposed by Merritt (44). If U and V are 2 anisotropic B-factor tensors, then the correlation coefficient between them is given by:

$$cc(U,V) = \frac{(detU^{-1}detV^{-1})^{1/4}}{[(1/8)det(U^{-1}+V^{-1})]^{1/2}},$$

The normalized correlation coefficient is given by:

$$ncc(U,V) = \frac{cc[U,(U_{eq}/V_{eq})\mathrm{V}]}{cc(U,U_{iso})cc(V,V_{iso})},$$

where U_{iso} and V_{iso} describe a pair of isotropic atoms, with $U_{iso}{}^{11} = U_{iso}{}^{22} = U_{iso}{}^{33} = U_{eq} = \mathrm{trace}(U)/3$ and similarly for V_{iso}. This normalized correlation coefficient *ncc* in Eq. **2** will be >1 if the 2 atoms described by U and V are more similar to each other than to an isotropic case, and will not be >1 otherwise. Thus, *ncc* provides an excellent measure to compare the size, orientation, and direction of the 2 tensors. In practice, a simple count of how many atoms in a structure have their normalized correlation coefficients larger than 1 provides a good measure of the quality of anisotropic B-factor prediction.

The percentages of residues with *ncc* larger than 1, as computed by pfANM and ANM, are listed in the last column of Table 1. Using this *ncc* measure, it can clearly be seen that pfANM yields significant improvements over ANM in anisotropic B-factors predictions.

Table 1. Predictions of anisotropic B-factors using pfANM and the original ANM

	$<\kappa_{exp} - \kappa_{calc}>$	$<\chi_{exp} - \chi_{calc}>$	$<\theta>$	$<\phi>$	ncc > 1
ANM	0.04 ± 0.17	0.07 ± 0.17	50.6 ± 22.6	55.4 ± 22.2	68%
pfANM	0.02 ± 0.16	0.02 ± 0.16	45.0 ± 23.0	54.2 ± 22.4	82%

All symbols in the column headers are defined in the text (κ and χ are the first and second anisotropies, θ and ϕ angles are direction preferences, and *ncc* is the normalized correlation coefficient).

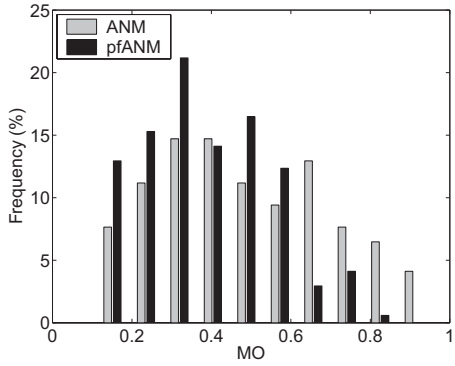

Fig. 5. Distribution of Maximum overlap (MO) values between the conformational change and a single mode of ANM or pfANM for all protein structure pairs studied.

Discussion

Besides testing the performance of pfANM in isotropic and anisotropic B-factor predictions, we have also looked at how well the modes of pfANM are related to experimentally observed conformational changes, especially the conformational differences between pairs of experimental structures of the same protein that has both an "open" and a "closed" form (31, 32, 45–49). We conduct this study using the same approach and dataset as in our previous work (32), except that here we now recalculate using pfANM. There are a total of 170 structure pairs in the dataset.

The conformational changes between the "open" and "closed" forms of the 170 proteins are interpreted with the modes calculated from both pfANM and ANM. We employ 2 measures to quantify how well the conformational changes can be explained by pfANM or ANM modes. One of them is the maximum overlap (MO, cosine of angle between 2 vectors), which is the maximum of the overlap values between the observed conformational change and the direction of a mode. If there is a single mode that contributes significantly to the observed conformational change, it has a high overlap value and it might be a mode that is functionally important. Fig. 5 shows the maximum overlap (MO) values for both pfANM and ANM. The 2 models are quite comparable even though the modes of ANM match slightly better with the observed conformational changes than do those of the pfANM. It is, however, likely that distortions of different magnitudes may exhibit different distance dependences. First, both ANM and pfANM assume harmonic interactions among the residues and consequently, both models are most valid for interpreting local fluctuations (B-factors). When applying these models to interpret larger conformational changes, caution needs to be taken. The larger deformations observed in these conformational transitions often correspond to the proteins moving as if they were comprised of rigid domains or clusters, that is, the local cohesion within a domain/cluster appears to be much stronger relative to the interdomain cohesiveness. Compared with the ANM model with a cutoff distance, the pfANM with inverse square distance dependence has stronger long-range cohesion that does not permit discrete domains to move sufficiently. Indeed, when we increase the power of the inverse square function built into the pfANM to a higher power to strengthen the short range interactions over the longer range interactions, to permit domains to move, then the pfANM model outperforms the ANM in its interpretation of conformational changes. As shown in Table S5, the best power dependence of the pfANM for the large conformational transitions is in the range of r^{-6} to r^{-8}.

In this work, we developed a parameter-free ENM (pfENM) and compared its performance with the conventional ENM that uses a cutoff distance to define binary interactions—either on or off. We have found that our pfENMs are not only simpler (without any cutoff distances), but they also perform better than usual ENMs, for both isotropic and anisotropic B-factor predictions. The removal of the cutoff distance may also remedy the problem of the scrambling of the modes that sometimes occurs between structures that differ only slightly (13), and other problems caused by the discontinuity of the interactions at the boundary of the cutoff distance. Thus, these parameter-free ENMs should be the models of choice wherever elastic network models are used. Recall that the ENMs are conformational entropy models, so these distance-dependent springs are better able to capture the conformational entropic fluctuations than can the use of a cutoff distance. This means that the cohesiveness of protein structures is better represented by interactions decaying with the inverse square of the separation distances. In some ways this corresponds to others' findings where they used inverse 6th power springs (29, 36)

Improvements to the ENMs are critical for their use in developing mechanisms (50–53). Previously springs of different strengths for different classes of interactions were not been found to improve the motions computed with ENMs (54). The improvements found here do suggest that the introduction of different power dependences for different extents of deformation leads to gains. This introduces an approach, different in formulation from distance-dependent interaction energies, for distance-dependent entropy-related cohesion.

Methods

Protein Dataset. We use PDB-PEPRDB (55) to select protein structures from the Protein Data Bank (PDB) (56). We choose protein structures determined by X-ray crystallography at resolutions better than 2.0 Å and with R-factors better than 0.2. We exclude protein fragments or membrane proteins. All protein sizes are at least 50 residues with sequence identities not $>25\%$ and structure similarities differing by >10 Å (parameter in PDB-PEPRDB). We exclude structures that only have backbone atoms or alpha carbons. We also remove any structure that does not provide experimental B-factors. Finally, we obtain and use 1,220 protein structures in our dataset.

For anisotropic B-factors, we choose to include in our dataset all protein crystal structures in the Protein Data Bank (PDB) that have resolution equal or better than 1.2 Å and have anisotropic B-factor entries. There were 341 such structures in our dataset. The dataset is the same one that we used in (26).

Coarse-Grained Gaussian Network Model (GNM). In the GNM (8), a protein structure is represented by its alpha carbons only. The relative fluctuations between a pair of contacting residues are assumed to follow a Gaussian distribution in its dependence on the square of the separation (hence the name of the model). Alpha carbon pairs are considered to be in contact when their separation distance is smaller than a preset cutoff distance ($\approx$7–8 Å, here we use 7.3 Å). All springs are taken to be at equilibrium for the input structure, and the input structure is assumed to be the lowest in energy, for the computed fluctuations around it. One advantage of this approach is that the fluctuations of each atom around its equilibrium position, and their cross-correlations can be expressed in simple analytical form. To determine the atomic fluctuations, we first form the Kirchhoff matrix based on the contact information:

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } r_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } r_{ij} > r_c \\ -\Sigma_{i,j\neq i}\Gamma_{ij} & \text{if } i = j \end{cases},$$

where r_{ij} is the distance between atoms i and j, and r_c is the cutoff distance. The mean square fluctuations of each atom are given by:

$$\langle \Delta R_i^2 \rangle = (3k_B T / \gamma)[\Gamma^{-1}]_{ii} ,$$

where γ is the spring constant. Then the theoretical B-factors can be conveniently expressed as:

$$B_i = 8\pi^2 \langle \Delta R_i^2 \rangle / 3 .$$

Anisotropic Network Model (ANM). In the ANM (25), the Hessian matrix contains the second derivatives of the energy function. For a structure with n residues, the Hessian matrix H contains n by n superelements each of size 3 by 3. The ijth superelement of H is given as:

$$H_{ij} = \begin{bmatrix} \frac{\partial^2 V}{\partial X_i \partial X_j} & \frac{\partial^2 V}{\partial X_i \partial Y_j} & \frac{\partial^2 V}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Y_i \partial X_j} & \frac{\partial^2 V}{\partial Y_i \partial Y_j} & \frac{\partial^2 V}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Z_i \partial X_j} & \frac{\partial^2 V}{\partial Z_i \partial Y_j} & \frac{\partial^2 V}{\partial Z_i \partial Z_j} \end{bmatrix}$$

where X_i, Y_i, and Z_i are the Cartesian components of residue i, and V represents the potential energy of the system. The potential between residues i and j is harmonic, modeled by a Hookean spring, for residues i and j within the cutoff distance. In the ANM (25), a larger cutoff distance than for GNM, say 13 Å, is normally used to establish the spring connections between residues. The Hessian matrix H is the counterpart of the Kirchhoff matrix Γ in the GNM. Similarly, the inverse of H contains n by n superelements, with the superelements of H^{-1}, being 3 by 3 matrices, describing the correlations between the components of ΔR_i. and ΔR_j. Specifically,

$$\langle \Delta R_i^2 \rangle = (3k_B T / \gamma) * trace([H^{-1}]_{ii}) .$$

Parameter-Free ENM (pfENM). In the pfENM models, there is no cutoff distance. All pairs of residues are considered to be interacting with one another, although their interaction strengths are weighted by the inverses of their square distances. (A higher power of distances could have been used, however, from our preliminary results, the differences would be small.). Thus, residue pairs that are far apart have weaker interactions than those that are close. In pfGNM, the elements of the Kirchhoff matrix Γ are calculated as:

$$\Gamma_{ij}^{pf} = \begin{cases} r_{ij}^{-2} & if\ i \neq j \\ -\Sigma_{i,j \neq i} \Gamma_{ij} & if\ i = j \end{cases} ,$$

where r_{ij} is the distance between residues i and j. In the pfANM, each superelement of the Hessian matrix is weighted by the inverse of the square distance between that residue pair, that is,

$$H_{ij}^{pf} = H_{ij} r_{ij}^{-2} .$$

Comparing Computed B-Factors with Experimental Values. The correlation between experimental and calculated B-factors is given by:

$$corr(B^{exp}, B^{calc}) = \frac{B^{exp} - \langle B^{exp} \rangle}{||B^{exp} - \langle B^{exp} \rangle||} \frac{B^{calc} - \langle B^{calc} \rangle}{||B^{calc} - \langle B^{calc} \rangle||} .$$

A perfect correlation between the 2 vectors gives a value of 1 whereas perfect anti-correlation gives −1. Here, the computed B-factors can be from either the GNM, ANM, or pfENM models.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grants R01GM073095, R01GM072014, and R01GM081680.

1. Karplus PA, Schulz GE (1985) Prediction of chain flexibility in proteins - A tool for the selection of peptide antigens. *Naturwiss* 72:212–213.
2. Vihinen M, Torkkila E, Riikonen P (1994) Accuracy of protein flexibility predictions. *Proteins* 19:141–149.
3. Smith DK, Radivojac P, Obradovic Z, Dunker AK, Zhu G (2003) Improved amino acid flexibility parameters. *Protein Sci* 12:1060–1072.
4. Radivojac P, et al. (2004) Protein flexibility and intrinsic disorder. *Protein Sci* 13:71–80.
5. Yuan Z, Bailey TL, Teasdale RD (2005) Prediction of protein B-factor profiles. *Proteins* 58:905–912.
6. Schlessinger A, Rost B (2005) Protein flexibility and rigidity predicted from sequence. *Proteins* 61:115–126.
7. Chen P, Wang B, Wong HS, Huang DS (2007) Prediction of protein B-factors using multiclass bounded SVM. *Protein Pept Lett* 14:185–190.
8. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 2:173–181.
9. Haliloglu T, Bahar I (1999) Structure-based analysis of protein dynamics: Comparison of theoretical results for hen lysozyme with X-ray diffraction and NMR relaxation data. *Proteins* 37:654–667.
10. Kundu S, Melton JS, Sorensen DC, Phillips GN, Jr (2002) Dynamics of proteins in crystals: Comparison of experiment with simple models. *Biophys J* 83:723–732.
11. Yang LW, et al. (2006) oGNM: Online computation of structural dynamics using the Gaussian network model. *Nucl Acids Res* 34:W24–W31.
12. Sen TZ, Feng Y, Garcia JV, Kloczkowski A, Jernigan RL (2006) The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. *J Chem Theo Comp* 2:696–704.
13. Song G, Jernigan RL (2007) vGNM: A better model for understanding the dynamics of proteins in crystals. *J Mol Biol* 369:880–893.
14. Ming D, Kong Y, Lambert MA, Huang Z, Ma J (2002) How to describe protein motion without amino acid sequence and atomic coordinates. *Proc Natl Acad Sci USA* 99:8620–8625.
15. Stillinger FH, Rahman A (1974) Improved simulation of liquid water by molecular dynamics. *J Chem Phys* 60:1545–1557.
16. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. *Nature* 267:585–590.
17. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comp Chem* 26:1781–1802.
18. Brooks BR, et al. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* 4:187–217.
19. Nishikawa T, Go N (1987) Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. *Proteins* 2:308–329.
20. Brooks B, Karplus M (1985) Normal modes for specific motions of macromolecules: Application to the hinge-bending mode of lysozyme. *Proc Natl Acad Sci USA* 82:4995–4999.
21. Brooks CL, Karplus M, Pettitt BM (1988) Proteins: A theoretical perspective of dynamics, structure, and thermodynamics. *Adv Chem Phys* 71:1–249.
22. Case DA (1994) Normal mode analysis of protein dynamics. *Curr Opin Struct Biol* 4:285–290.
23. Haliloglu T, Bahar I, Erman B (1997) Gaussian dynamics of folded proteins. *Phys Rev Lett* 79:3090–3093.
24. Bahar I, Jernigan RL (1998) Vibrational dynamics of transfer RNAs: Comparison of the free and synthetase-bound forms. *J Mol Biol* 281:871–884.
25. Atilgan AR, et al. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 80:505–515.
26. Yang, L, Song, G, Jernigan, RL (2009) Comparisons of experimental and computed protein anisotropic temperature factors. *Proteins* 76:164–175.
27. Eyal E, Bahar I (2008) Toward a molecular understanding of the anisotropic response of proteins to external forces: Insights from elastic network models. *Biophys J* 94:3424–3435.
28. Eyal E, Chennubhotla C, Yang LW, Bahar I (2007) Anisotropic fluctuations of amino acids in protein structures: Insights from X-ray crystallography and elastic network models. *Bioinformatics* 23:i175–184.
29. Riccardi D, Cui Q, Phillips GN, Jr (2009) Application of elastic network models to proteins in the crystalline state. *Biophys J* 96:464–475.
30. Kondrashov DA, Van Wynsberghe AW, Bannen RM, Cui Q, Phillips GN, Jr (2007) Protein structural variation in computational models and crystallographic data. *Structure* 15:169–177.
31. Song G, Jernigan RL (2006) An enhanced elastic network model to represent the motions of domain-swapped proteins. *Proteins* 63:197–209.
32. Yang L, Song G, Jernigan RL (2007) How well can we understand large-scale protein motions using normal modes from elastic network models? *Biophys J* 93:920–929.
33. Erman B (2006) The Gaussian network model: Precise prediction of residue fluctuations and application to binding problems. *Biophys J* 91:3589–3599.

34. Hinsen K, Kneller GR (1998) Analysis of domain motions by approximate normal mode calculations. *Proteins* 33:417–429.
35. Hinsen K, Kneller GR (1999) A simplified force field for describing vibrational protein dynamics over the whole frequency range. *J Chem Phys* 111:10766–10769.
36. Hinsen K, Petrescu A, Dellerue S, Bellissent-Funel M, Kneller GR (2000) Harmonicity in slow protein dynamics. *Chem Phys* 261:25–37.
37. Hinsen K (2008) Structural flexibility in proteins:Impact of the crystal environment. *Bioinformatics* 24:521–528.
38. Lin CP, et al. (2008) Deriving protein dynamical properties from weighted protein contact number. *Proteins* 72:929–935.
39. Halle B (2002) Flexibility and packing in proteins. *Proc Natl Acad Sci USA* 99:1274–1279.
40. Weiss MS (2007) On the interrelationship between atomic displacement parameters (ADPs) and coordinates in protein structures. *Acta Crystallogr D* 63:1235–1242.
41. Tirion MM (1996) Large amplitude elastic motions from a single-parameter, atomic analysis. *Phys Rev Lett* 77:1905–1908.
42. Tanaka S, Scheraga H (1976) Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules* 9:945–950.
43. Miyazawa S, Jernigan RL (1985) Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules* 18:534–552.
44. Merritt EA (1999) Comparing anisotropic displacement parameters in protein structures. *Acta Crystallogr* D55:Pt 12:1997–2004.
45. Kim MK, Jernigan RL, Chirikjian GS (2002) Efficient generation of feasible pathways for protein conformational transitions. *Biophys J* 83:1620–1630.
46. Kim MK, Jernigan RL, Chirikjian GS (2003) An elastic network model of HK97 capsid maturation. *J Struct Biol* 143:107–117.
47. Kim MK, Jernigan RL, Chirikjian GS (2005) Rigid-cluster models of conformational transitions in macromolecular machines and assemblies, *Biophys J* 89:43–55.
48. Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. *Protein Eng* 14:1–6.
49. Krebs WG, et al. (2002) Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic. *Proteins* 48:682–695.
50. Bahar I, Erman B, Jernigan RL, Covell DG (1999) Collective motions in HIV-1 Reverse Transcriptase: Examination of flexibility and enzyme function. *J Mol Biol* 285:1023–1037.
51. Keskin O, Jernigan RL, Bahar I (2000) Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophys J* 78:2093–2106.
52. Wang Y, Rader AJ, Bahar I, Jernigan RL (2004) Global ribosome motions revealed with elastic network model. *J Struct Biol* 147:302–314.
53. Kurkcuoglu O, Jernigan RL, Doruker P (2006) Loop motions of triosephosphate isomerase observed with elastic networks. *Biochemistry* 45:1173–1182.
54. Sen TZ, Jernigan RL (2006) Optimizing the parameters of the Gaussian network model for ATP-binding proteins. *In Normal Mode Analysis*: *Theory and Applications to Biological and Chemical Systems*, eds Bahar I, Cui Q (Chapman and Hall/CRC, Boca Raton), pp 171–186.
55. Noguchi T, Akiyama Y (2003) PDB-REPRDB: A database of representative protein chains from the Protein Data Bank (PDB) in 2003. *Nucleic Acids Res* 31:492–493.
56. Berman HM, et al. (2000) The protein data bank. *Nucleic Acids Res* 28:235–242.