A minimalist network model for coarse-grained normal mode analysis and its application to biomolecular x-ray crystallography

Mingyang Lu* and Jianpeng Ma*†§

*Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, BCM-125, Houston, TX 77030; and †Department of Bioengineering, Rice University, Houston, TX 77005

Communicated by William N. Lipscomb, Harvard University, Cambridge, MA, July 22, 2008 (received for review June 15, 2008)

In this article, we report a method for coarse-grained normal mode analysis called the minimalist network model. The main features of the method are that it can deliver accurate low-frequency modes on structures without undergoing initial energy minimization and that it also retains the details of molecular interactions. The method does not require any additional adjustable parameters after coarse graining and is computationally very fast. Tests on modeling the experimentally measured anisotropic displacement parameters in biomolecular x-ray crystallography demonstrate that the method can consistently perform better than other commonly used methods including our own one. We expect this method to be effective for applications such as structural refinement and conformational sampling.

all-atom normal modes | elastic network model | energy minimization | normal mode analysis | x-ray refinement

Normal mode analysis is a powerful tool for describing the global, collective, and functional motions of protein complexes (1–5). In this approach, the potential energy function of a protein is assumed to be harmonic so that protein motions can be described as a linear combination of a set of independent harmonic modes. Typically, only a few lowest-frequency modes with the largest amplitudes are sufficient to account for a majority of experimentally observed conformational fluctuations.

Conventional normal mode analysis requires the calculation of an all-atom second-derivative matrix, the Hessian matrix, by using typical molecular mechanics force fields such as CHARMM (6–8) or AMBER (9–12). Thus, it requires substantial computer memory and processing power to perform the matrix diagonalization, which becomes a severe bottleneck in studies of supramolecular complexes. Moreover, to satisfy the harmonic approximation, the conventional method requires a lengthy initial energy-minimization step.

To reduce the computational cost, many types of coarse-grained normal mode analyses have been developed (5, 13, 14). The most notable types include the rotations-translations of blocks (RTB) method (15) [also called block normal mode analysis (16)], all-atom-derived coarse-grained methods (17, 18), and different variations of the elastic network model (19–29). The latter constructs the Hessian matrix from a highly simplistic Hamiltonian rather than a realistic molecular force field. Although elastic network models have been very successful, the quality of their modes is not optimal according to some recent studies, probably because of the usage of oversimplified Hamiltonians. In several reports that evaluated the use of low-frequency modes for interpreting crystallographic thermal parameters (28, 30, 31), the RTB approach with the CHARMM force field predicts anisotropic displacement parameters (ADPs) better than various elastic network models (28).

Another limitation of the elastic network model is that it ignores molecular interaction details, which can be very important for certain cases (32). Also, in many applications, the elastic network model performs better on more uniformly packed systems (33); however, for systems containing components with substantially different stiffness, such as proteins, nucleic acids, and lipid molecules, the construction of a network may become tricky. Although it is possible to adjust the parameters, such as density of network nodes and cutoff distance (50), to better reflect the difference in stiffness, it is difficult to find a universal scheme for all systems.

On the other hand, methods that retain molecular interaction information normally require a lengthy initial energy-minimization step. The validity of these methods hinges on an assumption that a nearby local energy minimum is still a reasonable representation of the native structure. Although such an assumption seems reasonable for many applications, it falls short for other more demanding applications such as x-ray crystallographic refinement (34–37) and harmonic-mode-based Monte Carlo (38), for which the structural deviation resulting from initial energy minimization may introduce substantial errors in sampling. Fig. 1 shows the typical distribution of the RMSDs of heavy atoms from their original positions after energy minimization. In most cases, the minimization results in $\approx$1.5 Å RMSD. In some ultraflexible systems, the RMSD can be as large as 4 to 5 Å.

Hence, it is desirable to design a scheme that (*i*) places the native structures at the potential energy local minimum such that initial energy minimization is not required yet also (*ii*) retains detailed molecular interaction information. The former feature is indeed a distinct advantage of many elastic network models (14, 19, 21, 39), including an enhanced version (29) that was extremely successful in refining ADPs for lower-resolution x-ray crystallographic data of flexible complexes (36, 37). None of these elastic network models, however, are capable of including detailed molecular interaction information.

Because the commonly used molecular mechanics force fields are not parameterized to place a given crystal structure at an energy minimum, an alternative to using energy minimization on the force fields is to modify the molecular interactions so that minimization is no longer needed. Because the lowest-frequency normal modes are insensitive to small changes in the local stiffness (40), it is thus possible to perturb the Hessian matrix to avoid the initial energy minimization without affecting the overall nature of the lowest-frequency modes.

Inspired by this idea, we developed the minimalist network model (MNM), in which the Hessian matrix is slightly modified to be positive semidefinite. It is called "minimalist" because it uses a minimal representation of molecular interactions. It differs from traditional elastic network models in that it directly derives the Hessian matrix from an all-atom force field, which ensures the inclusion of important molecular interaction details. It also differs

Author contributions: M.L. and J.M. designed research; M.L. performed research; M.L. and J.M. analyzed data; and M.L. and J.M. wrote the paper.

The authors declare no conflict of interest.

§To whom correspondence should be addressed. E-mail: jpma@bcm.tmc.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0806072105/DCSupplemental.

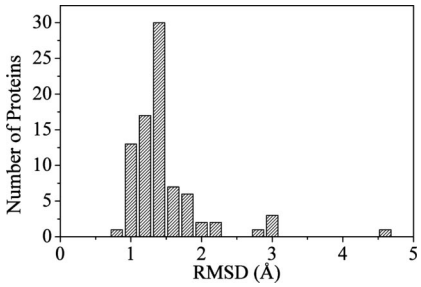

Fig. 1. The distribution of heavy-atom RMSD between protein structures before and after initial energy minimization. The histogram was generated from 83 ultra-high-resolution protein crystal structures, which have at least 50 residues, are at least 1 Å in resolution, and share <50% sequence identity. For more details on the protein test set and minimization protocol, see *Methods*.

from standard normal mode analysis in that it does not require initial energy minimization. Moreover, the MNM contains no additional parameters after coarse graining, and it is computationally very fast.

In this article, we first describe a general pairwise decomposition (PD) scheme starting from the RTB method. Next, we show how to modify the Hessian matrix to avoid initial energy minimization. The MNM was compared with various other normal mode analysis methods in terms of similarity of modes and ability to predict the ADPs measured by x-ray crystallography experiments. Our results show that the MNM not only delivers reliable eigenvectors for low-frequency modes but also models the experimental ADPs consistently better than any of the other available methods. Given that the MNM directly calculates modes from the unminimized structures without sacrificing important molecular interaction details, we expect this method to be a useful tool for applications such as structural refinement and conformational sampling.

Theory

The MNM is based on RTB and a PD scheme. In this section, conventional normal mode analysis and RTB are reviewed first, followed by an introduction of the PD scheme. Finally, the MNM Hessian is derived from the PD method.

Conventional Normal Mode Analysis and the RTB Method. For a molecule of N atoms whose structure is at a local energy minimum, the normal modes can be obtained from a $3N \times 3N$ mass-weighted second-derivative matrix, or Hessian matrix, $\mathbf{H}$, defined in a molecular force field. The eigenvalue λ of a single mode and its associated $3N \times 1$ eigenvector $\mathbf{r}$ can be obtained by solving the eigenvalue equation, $\mathbf{Hr} = \lambda\mathbf{r}$.

For large systems such as supramolecular complexes, a coarse-grained method called the RTB method (15) has been devised to reduce the computational cost. In RTB, atoms in one or more residues are grouped into a rigid-body block, the motion of which is described by six translational and rotational degrees of freedom. If the molecule is divided into n blocks, the corresponding Hessian matrix in RTB, $\mathbf{H}_{RTB}$, is a $6n \times 6n$ matrix. It is related to the all-atom Hessian by $\mathbf{H}_{RTB} = \mathbf{P}^T\mathbf{HP}$, where $\mathbf{P}$ is a $3N \times 6n$ orthogonal projection matrix. The mapping between the $6n \times 1$ rigid-body displacement vector $\mathbf{x}$ of RTB and atomic displacement vector $\mathbf{r}$ is given by

$$\mathbf{r} = \mathbf{Px}. \qquad [1]$$

The PD Scheme. The purpose of the PD scheme is to decompose the interactions of the whole molecule into pairwise interactions of small subsystems (blocks). For any isolated molecule of n blocks at a local energy minimum, the external motions of which produce no net forces, the RTB Hessian $\mathbf{H}_{RTB}$ obeys

$$\mathbf{H}_{RTB}\Omega = \mathbf{0}, \qquad [2]$$

where Ω is the $6n \times 6$ eigenvector matrix for the six external translational-rotational modes. The matrix Ω can be calculated from Eq. **1** by $\Omega = \mathbf{P}^T\mathbf{P}_1$, where $\mathbf{P}_1$ is the $3N \times 6$ projection matrix for the RTB that regards the whole molecule as a single block (i.e., $n = 1$).

The PD Hessian $\mathbf{H}_{PD}$ can be obtained from $\mathbf{H}_{RTB}$ by

$$\frac{1}{2}\mathbf{x}^T\mathbf{H}_{PD}\mathbf{x} = \frac{1}{2}\sum_{i<j}\begin{pmatrix}\mathbf{x}_i\\ \mathbf{x}_j\end{pmatrix}^T\mathbf{H}_{ij}\begin{pmatrix}\mathbf{x}_i\\ \mathbf{x}_j\end{pmatrix}, \qquad [3]$$

where $\mathbf{x}_i$ is the 6×1 rigid-body component of the displacement vector $\mathbf{x}$ for block i, and $\mathbf{H}_{ij}$ is the 12×12 decomposed Hessian matrix for the ijth block pair:

$$\mathbf{H}_{ij} = \begin{pmatrix} -\frac{1}{2}(\mathbf{K}_{ij}\Gamma_{ij}^{-1} + (\Gamma_{ij}^{-1})^T\mathbf{K}_{ij}^T) & \frac{1}{2}(\mathbf{K}_{ij} + (\Gamma_{ij}^{-1})^T\mathbf{K}_{ij}^T\Gamma_{ij}) \\ \frac{1}{2}(\mathbf{K}_{ij}^T + \Gamma_{ij}^T\mathbf{K}_{ij}\Gamma_{ij}^{-1}) & -\frac{1}{2}(\mathbf{K}_{ij}^T\Gamma_{ij} + \Gamma_{ij}^T\mathbf{K}_{ij}) \end{pmatrix}. \qquad [4]$$

Here $\Gamma_{ij} = \Omega_i\Omega_j^{-1}$, Ω_i is the 6×6 nonsingular submatrix of Ω for block i, $\mathbf{K}_{ij} = \partial^2E/\partial\mathbf{x}_i\partial\mathbf{x}_j$ is the $_{ij}$ submatrix of $\mathbf{H}_{RTB}$, and E is the total energy of the whole molecule.

It is easy to verify that $\mathbf{H}_{ij}$ satisfies Eq. **2** when the block number is set to two, i.e.

$$\mathbf{H}_{ij}\begin{pmatrix}\Omega_i\\ \Omega_j\end{pmatrix} = 0, \qquad [5]$$

so $\mathbf{H}_{ij}$ is sufficient to represent the Hessian matrix of an isolated system of two blocks. Similarly, $\mathbf{H}_{PD}$ can represent the Hessian of the whole molecule [the derivations of the PD scheme can be found in supporting information (SI)]. Although the PD scheme is designed for minimized structures, $\mathbf{H}_{ij}$ can still be calculated from Eq. **4** for unminimized structures but with great care, because $\mathbf{H}_{PD}$ derived on unminimized structures is not guaranteed to be positive semidefinite.

Perturbation theory is applied to further assess the difference between the PD and the RTB schemes. For the normal modes with eigenvalues $\lambda_{(k)}$ and eigenvectors $\mathbf{x}_{(k)}$ (k is the index of the modes), and $\Delta\mathbf{H} = \mathbf{H}_{PD} - \mathbf{H}_{RTB}$, perturbation theory gives

$$\Delta\lambda_{(k)} = \mathbf{x}_{(k)}^T\Delta\mathbf{H}\mathbf{x}_{(k)} \qquad [6a]$$

$$\Delta\mathbf{x}_{(k)} = \sum_{k=} c_{(kl)}\mathbf{x}_{(l)} \qquad [6b]$$

$$c_{(kl)} = \frac{\mathbf{x}_{(k)}^T\Delta\mathbf{H}\mathbf{x}_{(l)}}{\lambda_{(k)} - \lambda_{(l)}}. \qquad [6c]$$

It can be shown that

$$\left\langle\frac{\Delta\lambda_{(k)}}{\lambda_{(k)}}\right\rangle \approx 0 \qquad [7a]$$

$$\sigma\left(\frac{\Delta\lambda_{(k)}}{\lambda_{(k)}}\right) \propto \frac{1}{\sqrt{n}} \qquad [7b]$$

$$|\langle c_{(kl)}\rangle| \le \gamma\frac{\lambda_{(k)} + \lambda_{(l)}}{|\lambda_{(k)} - \lambda_{(l)}|}, \qquad [7c]$$

where n is the number of blocks if the blocks are uniform in size, and γ is a small scaling factor (see SI for more details, and see *Results* for the numerical tests). This implies that the PD scheme can produce almost the same eigenvalues as RTB. In addition,

Eq. **7c** implies that each of the low-frequency eigenvectors in PD can be approximated as a linear combination of the RTB eigenvectors with similar frequencies.

The MNM Method. The MNM method is developed by modifying the PD scheme, and it guarantees that the Hessian matrix is positive semidefinite. This process is essentially equivalent to modifying the molecular interactions expressed in the original force field. In the MNM, all PD $\mathbf{H}_{ij}$ values are replaced by their nearest (in terms of the Frobenius norm) symmetric positive semidefinite matrices $\mathbf{H}_{ij}^{+}$ (see SI for an alternative approach).

In linear algebra, any symmetric matrix $\mathbf{M}$ can be expressed as $\mathbf{M}^{+} + \mathbf{M}^{-}$, where $\mathbf{M}^{+} = \mathbf{U}\Lambda^{+}\mathbf{U}^{T}$ is a nearest symmetric positive semidefinite matrix (41). Here, $\mathbf{U}$ is the eigenvector matrix of $\mathbf{M}$, and Λ^{+} is the nonnegative part of the diagonal matrix of $\mathbf{M}$ (the negative eigenvalues are reset to zero).

In this way, the interaction energy of each pair of blocks is at its minimum so the interaction energy of the whole system (i.e., the sum of all pairwise interactions) is also at the energy minimum. Note that this condition is also satisfied by the elastic network model (19, 21). Thus, unlike the PD scheme, the MNM is applicable to unminimized structures by modifying $\mathbf{H}_{ij}$ values obtained from the RTB $\mathbf{H}_{RTB}$ of the unminimized structures.

As in Eq. **3**, the MNM Hessian matrix can be calculated from

$$\frac{1}{2}\mathbf{x}^{T}\mathbf{H}_{MNM}\mathbf{x} = \frac{1}{2}\sum_{i<j}\begin{pmatrix}\mathbf{x}_i\\ \mathbf{x}_j\end{pmatrix}^{T}\mathbf{H}_{ij}^{+}\begin{pmatrix}\mathbf{x}_i\\ \mathbf{x}_j\end{pmatrix}. \qquad [8]$$

All $\mathbf{H}_{ij}^{+}$ values satisfy Eq. **5** and $\mathbf{H}_{MNM}$ satisfies Eq. **2**. The eigenvectors of the MNM may be transformed to the all-atom representation by Eq. **1**.

Results

Tests on Protein Structures with Energy Minimization. It is worth comparing various methods quantitatively. We first performed the comparison on 83 protein structures after initial energy minimization (see *Methods*).

For PD on minimized structures, the eigenvalues are almost identical to those from the original RTB when comparing modes sorted by eigenvalue magnitude. The relative difference is ≈1% for the lowest-frequency normal modes (eigenvalue magnitudes <0.2 kcal·g^{-1}·Å^{-2}), and it approaches zero for the higher-frequency modes. This property is universal to proteins of various sizes, as shown in Fig. 2 *Upper*. To verify the scaling behavior of the standard deviation (SD) of $\Delta\lambda/\lambda$ in Eq. **7b**, we analyzed the eigenvalues of the first 1,000 lowest-frequency modes whose magnitudes are in the range of 0.2 to 0.8 kcal·g^{-1}·Å^{-2} for each protein. As shown in Fig. 2 *Lower*, $\sigma(\Delta\lambda/\lambda) \approx 0.04/\sqrt{n}$ with a 0.97 linear correlation coefficient.

Similar to the comparison of PD to RTB, the relative differences between the MNM and RTB eigenvalues fall into a curve that is independent of protein size (Fig. 2 *Upper*). Compared with PD, the MNM has larger deviations for the lower-frequency modes. Because the negative definite components of $\mathbf{H}_{ij}$ values are eliminated in the MNM, it produces modes with relatively higher frequencies ($\Delta\lambda > 0$).

The eigenvectors produced by both PD and the MNM were compared with those from the original RTB as well. The results are presented for two proteins, an all-helical protein, myoglobin (PDB ID code 1a6m), and an all-sheet protein, Con A (PDB ID code 1nls). As shown in Fig. 3, the subspace of the first 50 lowest-frequency modes of PD is almost the same as that of RTB, whereas the overlap between MNM modes and RTB subspaces becomes weaker only for the highest-frequency modes in the range (i.e., the 50-mode subspace). As a control, the eigenvector overlap was also evaluated between RTB and an all-atom-based elastic network model (modified elNémo, see *Methods*), both on minimized structures. The eigenvector subspace of elNémo is much less similar to those of RTB and the MNM.

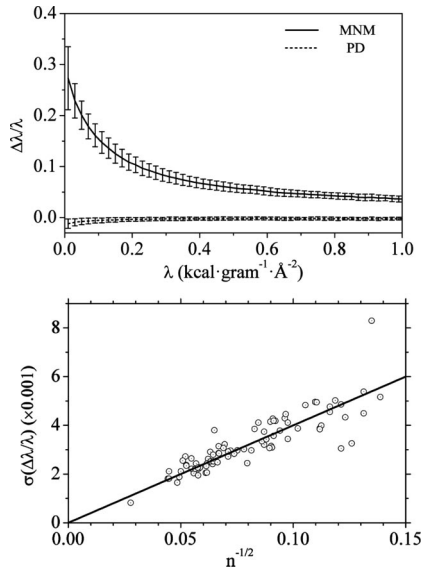

Fig. 2. Relative differences of eigenvalues. (*Upper*) Average relative difference of eigenvalues as a function of RTB eigenvalues for PD (dashed line) and the MNM (solid line). The statistics were derived from the first 1,000 modes with eigenvalues of <1.0 kcal·g^{-1}·Å^{-2} on all 83 ultra-high-resolution protein structures. The SDs are shown as vertical lines. (*Lower*) SDs for relative differences of eigenvalues (circles) as a function of inverse square root of the number of residues (*n*). For each of the 83 test proteins, the SD was calculated for the first 1,000 normal modes with eigenvalues ranging from 0.2 to 0.8 kcal·g^{-1}·Å^{-2} (as a representative range). The linear correlation coefficient is 0.97, and the slope is ≈0.04.

The close similarity of the low-frequency eigenvector subspaces of PD and the MNM to those of RTB on minimized structures indicates that PD and the MNM produce reliable modes on minimized structures.

Tests on Protein Structures Without Energy Minimization. The primary purpose of the MNM is to calculate modes based on unmini-

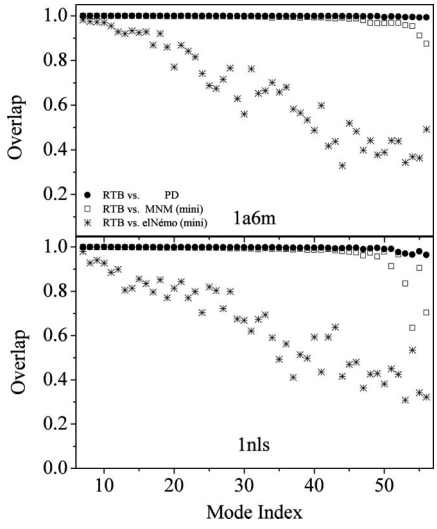

Fig. 3. Overlap of the lowest-frequency mode subspaces between various normal mode analyses. Tests were performed on two proteins: PDB ID codes 1a6m (*Upper*) and 1nls (*Lower*). The methods RTB, PD, MNM, and elNémo were applied to the structures after energy minimization, and overlap was calculated by projecting the modes of each method onto the 50-lowest-frequency-normal-mode subspace of RTB.

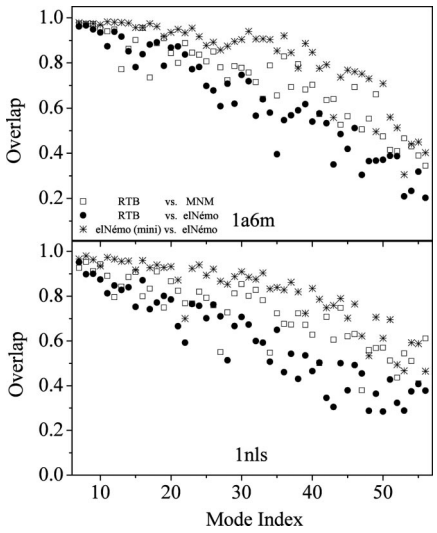

Fig. 4. Overlap of the lowest-frequency mode subspaces between various normal mode analyses. Tests were performed on two proteins: PDB ID codes 1a6m (*Upper*) and 1nls (*Lower*). The MNM was applied to relaxed structures, RTB to native structure, and elNémo to both native and minimized structures. For all comparisons, overlap was calculated by projecting the modes of each method onto the 50-lowest-frequency-normal-mode subspace of RTB, with the exception of the overlap between the two elNémo methods (stars), for which the modes were projected onto the 50-lowest-frequency-normal-mode subspace of elNémo on minimized structures.

mized structures. Because no conventional normal mode analyses can be properly performed on structures without energy minimization, the MNM on unminimized structures can only be compared with RTB on minimized structures. The results for the two test proteins used earlier (PDB ID codes 1a6m and 1nls) are presented in Fig. 4. Eigenvector overlap between the MNM (on relaxed structures, see *Methods* for the protocol) and RTB (on minimized structures) is better than that between elNémo (on native structures) and RTB (on minimized structures), which indicates that the MNM modes are closer to those of RTB than of elNémo. Even so, a relatively large difference in subspace overlap is found between the MNM and RTB, which may be a result of the structural deviation after minimization in RTB and/or the Hessian matrix modification in the MNM. As a control, eigenvector overlap between elNémo on minimized structures and on native structures is also shown in Fig. 4; the large difference in subspace overlap suggests that the impact of minimization on the modes is substantial.

Tests on Fitting of Experimental ADPs. Because no standard answers exist for normal modes on native structures without minimization, the consequence of the difference between the MNM and RTB are unclear. Thus, a more objective way to judge the quality of modes in the MNM is to model the ADPs by fitting against the experimental data.

In the left-most column of Table 1, the average Kullack–Leibler (KL) distances (the fitting score, after optimization) between the ADPs of various modal analysis methods and experimental data are shown. The smallest is found to be by the MNM. When applied to native structures, elNémo has an average KL distance similar to that of RTB, but worse when applied to minimized structures.

To examine the prediction and fitting quality in another way, we evaluated the correlation coefficient between theoretical (after KL-distance minimization) and experimental ADPs. In total, four types of correlation coefficients were calculated and averaged over proteins in the test sets. The results are listed in Table 1. It is clear that all correlation coefficients for the MNM are consistently better than those for RTB. For two elNémo methods, the coefficients are similar but worse than the MNM and RTB.

To evaluate the quality of the calculated ADP correlation coefficients, we compared them with the correlations of experimental ADPs for the same protein in different crystal environments. On the basis of the results in the first row of Table 1 and in tables 1 and 2 of ref. 30, we find that the MNM gives worse ADP correlation coefficients than the coefficients between the same protein in identical crystal form (experimental upper limit), but better than those between the same protein in different crystal forms. Thus, our results indicate that the fitting procedure for the MNM is reasonable for capturing the basic features of experimental ADPs, because the procedure is sensitive enough to detect differences in experimental ADPs resulting from crystal packing.

The fitting procedure in this study includes the contributions from overall rigid-body motions by using the six zero-frequency normal modes (see *Methods*). However, for RTB on minimized structures, the zero modes of minimized structures do not necessarily represent rigid-body motions of native structures. We also performed fitting for both the MNM and RTB with zero modes replaced by the zero modes obtained from native structures. These modes are slightly nonorthogonal to the other normal modes calculated on minimized structures, but the results are likely not too sensitive to the orthogonality. As shown in the last two rows in Table 1, no apparent improvements were found for the MNM, because the structures are almost unchanged after relaxation; whereas only slight improvements were found for RTB.

Fitting tests were also performed with all-atom CHARMM modes on several proteins after energy minimization. The fitting results are often worse than for RTB (data not shown), presumably because RTB restrains surface side-chain motions during initial minimization.

Finally, we also analyzed the improvements in KL distances from RTB to the MNM for each atom as a function of structural deviation after initial minimization. As shown in Fig. 5, the average

Table 1. Comparisons between calculated ADPs and experimental ADPs of various normal mode analyses

Normal mode analysis method	KL distance	Pearson's correlation coefficient with experimental data			
		All ADPs	Diagonal ADPs	Off-diagonal ADPs	Cα isotropic B
MNM	0.117	0.850	0.687	0.496	0.874
RTB	0.126	0.836	0.647	0.456	0.845
elNémo	0.127	0.792	0.608	0.415	0.835
elNémo (mini)	0.132	0.801	0.604	0.415	0.824
MNM (zero)	0.117	0.850	0.687	0.497	0.874
RTB (zero)	0.125	0.838	0.653	0.461	0.848

The test was performed by using 83 ultra-high-resolution protein crystal structures (see *Methods*), and the average results are presented here. After minimizing the KL distance between calculated and experimental ADPs, the Pearson's correlation coefficients of various groups of ADPs were calculated (right-most four columns). ElNémo was applied to unminimized ("elNémo") and minimized structures ["elNémo (mini)"]. The last two rows [with "(zero)"] show the results of modified versions of the MNM and RTB, in which the calculated zero modes were replaced by the zero modes of the native structures.

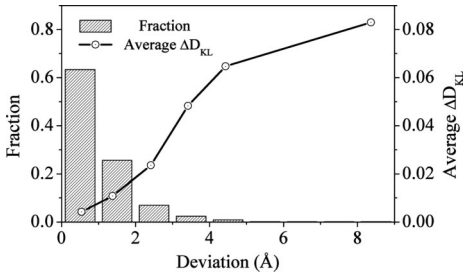

Fig. 5. Average KL distance improvement from RTB (on minimized structures) to the MNM (on unminimized structures) as a function of atomic positional deviation after energy minimization. The changes in KL distance from RTB to the MNM ($\Delta D_{KL} = D_{KL}^{RTB} - D_{KL}^{MNM}$) were calculated and averaged over atoms with similar positional deviations. This figure presents both the histogram of deviations (*y* axis to the left) and the average improvements in KL distance (*y* axis to the right). The average KL distance from RTB to the MNM improves for all deviation ranges (i.e., $\Delta D_{KL} > 0$), and the improvements tend to be larger for larger structure deviations.

KL distance improves for all RMSD regions, with larger improvements found in regions with larger structural deviations. The magnitude of the absolute value of ΔD_{KL} is substantial compared with the value of KL distances in the second column of Table 1. Once again, this finding highlights the importance of eliminating the structure shift attributable to energy minimization, which is perhaps particularly true for functionally relevant regions, such as protein active site loops, that are often located in the most flexible regions.

Discussion

In this study, we developed a method for coarse-grained normal mode analysis called the MNM. Starting from the RTB blocking scheme of the Hessian matrix computed from any molecular force field, the MNM uses a PD scheme to approximate the interactions between blocks. Next, the decomposed Hessian is modified such that the quadratic portion of the local interaction energy between blocks is retained. Such a scheme avoids the lengthy and sometimes problematic initial energy minimization. Another distinct advantage of the MNM is that it retains detailed molecular interaction information during coarse graining.

According to our results, the MNM preserves the lowest-frequency normal modes. More importantly, MNM modes predict experimental ADPs measured by x-ray crystallography better than the modes from any other method. Presumably, this result is largely attributable to the elimination of structural shifts caused by initial energy minimization. Thus, we expect that modes calculated by the MNM will perform in normal-mode-based x-ray refinement better than those in our previous studies (29, 36, 37).

Similar to RTB, the MNM is applicable to proteins at different coarse-graining levels. In this study, we applied the MNM only at the residue level, but for very large protein complexes, PD and the MNM can be applied to rigid multiresidue blocks such as secondary structure elements. From the test on all-α protein SLT70 (PDB ID code 1qsa), for which each helix element or loop residue is chosen as a block, the MNM produced similar results.

Although the interaction between a pair of blocks in the MNM is modeled as quadratic, much like the interactions between two nodes in the elastic network model (21), the nature of block–block interactions in the MNM is different from that between two nodes in the elastic network model. In the MNM, the two blocks are not two mass points; thus, relative motions between them are much more complex. For example, torsional motions between two blocks can occur, which is a reflection of more detailed molecular interactions. Thus, the MNM is fundamentally different even from elastic network models that have a nonuniform distribution of force constants between nodes.

The MNM is also applicable to other biomolecules such as nucleic acids and lipids. It is expected to be particularly advantageous for mixed systems such as protein–nucleic acid or protein–lipid complexes. Blocking schemes for nonprotein components can be chosen on the basis of chemical intuition. For the example of DNA, each nucleotide can be grouped into a block.

Finally, the low-frequency normal modes are excellent choices for generating conformational move sets in sampling algorithms, because they allow for large-scale structural changes without substantially increasing potential energy (38, 42). This feature is useful, for example, in refining computationally predicted structural models (43). In general, a modal analysis that can generate modes directly from a given structure offers a major advantage in conformational sampling.

Methods

Protein Test Set. The protein test set used in this study is derived from ref. 28, which consists of 83 ultra-high-resolution proteins ($\geq$1-Å resolution) with the number of residues ranging from 55 to 1,300. Before calculating normal modes, all proteins were preprocessed: (*i*) the residues of unrecognizable types (only found near the termini) were deleted; and (*ii*) the coordinates of missing atoms in some proteins were built by using CHARMM (8), including rebuilding all hydrogen atoms.

Conventional All-Atom Normal Modes. Conventional all-atom normal modes were computed by the CHARMM package. Similar to other studies in the literature (16), the united atom CHARMM19 force field (7) was used together with the EEF1 solvation model (44). Disulfide bonds were modeled according to information provided in the PDB files. Before the modes were calculated, multiple cycles of energy minimization by the adapted-basis Newton–Raphson method were performed with decreasing harmonic constraints. The minimizations were terminated at RMS energy gradients of 1×10^{-8} kcal·mol^{-1}·Å^{-1}. Conventional normal modes were then coarse-grained by using the RTB scheme (see *Theory*), by which all of the atoms in a residue were grouped into a block.

Structure Relaxation. In our MNM analysis on unminimized structures, to eliminate bad contacts in the crystal structures, we applied 100 steps of steepest-descent energy minimization: the first 50 steps with very weak harmonic constraints and the remaining 50 steps without. Such structure relaxation resulted in 0.13 Å RMSD from the crystal structure on average, which is substantially smaller than that from deep energy minimization in conventional normal mode calculations.

Elastic Network Model. Because the MNM and RTB are applied to all-atom systems, an all-atom-based elastic network model, elNémo (39), was compared to the MNM and RTB. elNémo was found to produce low-frequency modes similar to those of Cα-based elastic network models (19, 21, 39). In this study, a slight modification was made to elNémo to include hydrogen atoms. More specifically, all neighbor atom pairs within 8 Å are connected with springs of various strengths, depending on atom types: the strength of those between heavy atom pairs is unity; between heavy atoms and hydrogen atoms is 0.3; and between hydrogen pairs is 0.1. As in the original elNémo method (39), the RTB scheme was applied after obtaining the all-atom mass-weighted Hessian matrix.

Eigenvector Comparison. The eigenvectors from different normal mode analyses can be compared by means of a vector dot product. To evaluate the similarity of the subspaces of the lowest-frequency normal modes generated by two different methods, we used an overlap parameter defined as

$$\rho_i = \sum_j (\mathbf{y}_{(i)}^T \mathbf{x}_{(j)})^2, \qquad [9]$$

which projects an eigenvector $\mathbf{y}_{(i)}$ computed by one method onto the subspace of lowest-frequency-mode eigenvectors $\{\mathbf{x}_{(j)}\}$ calculated by another method. In this study, the first 50 lowest-frequency modes for each method were used for comparison.

Comparison Criteria for ADPs. Using normal modes to model experimental ADPs is an effective way of judging the quality of modes calculated by various methods (28, 30). This type of test is particularly important in this study, given the lack of a theoretical correct answer for unminimized native structures. ADPs of each atom consist of three diagonal parameters (U_{xx}, U_{yy}, and U_{zz}) and three off-diagonal parameters (U_{xy}, U_{xz}, and U_{yz}). Two measures were evaluated to com-

pare experimental ADPs with theoretical values modeled by normal mode analysis (see *Modeling of ADPs by Normal Modes* for the procedure). The first measure is Pearson's correlation coefficient (28, 30). The other measure is the KL distance (45), representing the difference between the Gaussian probability distributions defined by the theoretical and experimental ADPs (17, 30, 46). The KL distance for each atom can be expressed in terms of the eigenvalues ($\omega_{(p)}^{mode}$, $\omega_{(p)}^{data}$, $p \in 1, 2, 3$) and eigenvectors ($\mathbf{a}_{(p)}^{mode}$, $\mathbf{a}_{(p)}^{data}$, $p \in 1, 2, 3$) of the theoretical (mode) and experimental (data) ADP matrices as

$$D_{KL} = -\frac{3}{2} + \frac{1}{2}\sum_{p=1}^{3} \ln \frac{\omega_{(p)}^{mode}}{\omega_{(p)}^{data}} + \frac{1}{2}\sum_{p=1}^{3}\sum_{q=1}^{3} \frac{\omega_{(p)}^{data}}{\omega_{(q)}^{mode}} |(\mathbf{a}_{(p)}^{data})^T \mathbf{a}_{(q)}^{mode}|^2. \quad [10]$$

This equation is derived from ref. 30.

Modeling of ADPs by Normal Modes. ADPs can be modeled by normal modes in two ways. One is by a direct calculation, in which ADPs are obtained from the atomic thermal fluctuation predicted by normal modes (28, 30). The other way is by a fitting procedure, in which ADPs are obtained by optimizing normal mode parameters against the experimental data (e.g., diffraction data or experimental ADPs). This latter approach can produce more realistic ADPs as demonstrated in the normal-mode-based x-ray crystallographic refinements (34–37) and other theoretical studies (31). In this study, we used the fitting procedure to evaluate the ADPs.

The 3×3 ADP matrix for each atom in terms of normal modes is expressed as

$$\mathbf{U}^{mode} = \sum_k \sum_l \sigma_{kl} \mathbf{v}_{(k)} \mathbf{v}_{(l)}^T, \quad [11]$$

where σ_{kl} is the klth element of matrix $\Sigma = \mathbf{S}\mathbf{S}^T$, $\mathbf{S}$ is an $m \times m$ lower triangular matrix introduced to ensure positive semidefiniteness of Σ, m is the number of modes, and $\mathbf{v}_{(k)}$ represents the three components of the atom in the kth normal mode eigenvector. The summations are both over the chosen subset of low-frequency normal modes.

In the fitting procedure, 20 lowest-frequency normal modes, including six zero modes representing the contributions of the external motions (47, 48), were used. The elements S_{kl} were set to be zero for $k > 6$, $l \leq 6$ to ignore the correlations between external and internal motions. Consequently, a total of $6 \times 7/2 + 14 \times 15/2 = 126$ parameters were involved. The average KL distance was calculated for all heavy atoms, except for those with defects in experimental data (e.g., those with no ADPs) with only U_{xx} components or with non-positive-definite covariance matrices of ADPs. Our study used a much larger experimental data set than any of the analyses in recent studies, in which only Cα data were considered (28, 30, 31). As a result, there were $\approx$6,000 ADP data points for a protein of 100 residues, providing a reasonable data-to-parameter ratio for appropriate fitting.

In the optimization process, $\mathbf{S}$ was initially set as $S_{kk} \propto \lambda_{(k)}^{-1/2}$ for the diagonal elements [$\lambda_{(k)}$ is the eigenvalue of the kth mode] and small random numbers for the rest. Then, $\mathbf{S}$ was further scaled so that the average magnitude of ADPs matched that of the experimental data. The independent elements of $\mathbf{S}$ were then optimized by minimizing the average KL distance by using Brent's principal axis method (49).

ACKNOWLEDGMENTS. We thank Athanasios D. Dousis for critical reading of the manuscript. This work was supported by National Institutes of Health Grant GM067801-06A1, National Science Foundation Grant MCB-0818353, and Welch Grant Q-1512.

1. Brooks B, Karplus M (1983) Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA* 80:6571–6575.
2. Go N, Noguti T, Nishikawa T (1983) Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc Natl Acad Sci USA* 80:3696–3700.
3. Levitt M, Sander C, Stern PS (1985) Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *J Mol Biol* 181:423–447.
4. Brooks BR, Janezic D, Karplus M (1995) Harmonic analysis of large systems. I. Methodology. *J Comput Chem* 16:1522–1542.
5. Ma J (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure (London)* 13:373–380.
6. MacKerell AD, *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616.
7. Neria E, Fischer S, Karplus M (1996) Simulation of activation free energies in molecular systems. *J Chem Phys* 105:1902–1921.
8. Brooks BR, *et al*. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217.
9. Wang JM, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* 21:1049–1074.
10. Ponder JW, Case DA (2003) Force fields for protein simulations. *Adv Protein Chem* 66:27–85.
11. Case DA, *et al*. (2005) The amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688.
12. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174.
13. Ma J (2004) New advances in normal mode analysis of supermolecular complexes and applications to structural refinement. *Curr Protein Pept Sci* 5:119–123.
14. Bahar I, Rader AJ (2005) Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol* 15:586–592.
15. Tama F, Gadea FX, Marques O, Sanejouand YH (2000) Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* 41:1–7.
16. Li G, Cui Q (2002) A coarse-grained normal mode approach for macromolecules: An efficient implementation and application to Ca(2+)-ATPase. *Biophys J* 83:2457–2474.
17. Ming D, Wall ME (2005) Allostery in a coarse-grained model of protein dynamics. *Phys Rev Lett* 95:198103, and erratum (2006) 96:159902.
18. Zhou L, Siegelbaum SA (2008) Effects of surface water on protein dynamics studied by a novel coarse-grained normal mode approach. *Biophys J* 94:3461–3474.
19. Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* 77:1905–1908.
20. Haliloglu T, Bahar I, Erman B (1997) Gaussian dynamics of folded proteins. *Phys Rev Lett* 79:3090–3093.
21. Atilgan AR, *et al*. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 80:505–515.
22. Doruker P, Jernigan RL, Bahar I (2002) Dynamic of large proteins through hierarchical levels of coarse-grained structures. *J Comput Chem* 23:119–127.
23. Kurkcuoglu O, Jernigan RL, Doruker P (2005) Collective dynamics of large proteins from mixed coarse-grained elastic network model. *QSAR Comb Sci* 24:443–448.
24. Eom K, Baek SC, Ahn JH, Na S (2007) Coarse-graining of protein structures for the normal mode studies. *J Comput Chem* 28:1400–1410.
25. Ming D, Kong Y, Lambert M, 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.
26. Tama F, Wriggers W, Brooks CL (2002) Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. *J Mol Biol* 321:297–305.
27. Hinsen K (2000) The molecular modeling toolkit: A new approach to molecular simulations. *J Comput Chem* 21:79–85.
28. Kondrashov DA, Van Wynsberghe AW, Bannen RM, Cui Q, Phillips GN Jr. (2007) Protein structural variation in computational models and crystallographic data. *Structure (London)* 15:169–177.
29. Lu M, Poon B, Ma J (2006) A new method for coarse-grained elastic normal-mode analysis. *J Chem Theor Comput* 2:464–471.
30. 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–I184.
31. Song G, Jernigan RL (2007) vGNM: A better model for understanding the dynamics of proteins in crystals. *J Mol Biol* 369:880–893.
32. Van Wynsberghe A, Li G, Cui Q (2004) Normal-mode analysis suggests protein flexibility modulation throughout RNA polymerase's functional cycle. *Biochemistry* 43:13083–13096.
33. Bagci Z, Kloczkowski A, Jernigan RL, Bahar I (2003) The origin and extent of coarse-grained regularities in protein internal packing. *Proteins* 53:56–67.
34. Kidera A, Gō N (1990) Refinement of protein dynamic structure: normal mode refinement. *Proc Natl Acad Sci USA* 87:3718–3722.
35. Diamond R (1990) On the use of normal modes in thermal parameters refinement: Theory and application to the bovine pancreatic trypsin inhibitor. *Acta Crystallogr A* 46:425–435.
36. Poon BK, *et al*. (2007) Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-A crystallographic resolution. *Proc Natl Acad Sci USA* 104:7869–7874.
37. Chen X, Poon BK, Dousis A, Wang Q, Ma J (2007) Normal-mode refinement of anisotropic thermal parameters for potassium channel KcsA at 3.2 Å crystallographic resolution. *Structure (London)* 15:955–962.
38. Wu Y, *et al*. (2005) Folding of small helical proteins assisted by small-angle X-ray scattering profiles. *Structure (London)* 13:1587–1597.
39. Suhre K, Sanejouand YH (2004) elNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res* 32:W610–W614.
40. Lu M, Ma J (2005) The role of shape in determining molecular motions. *Biophys J* 89:2395–2401.
41. Higham NJ (1988) Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra Appl* 103:103–118.
42. Noguti T, Go N (1995) Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins. *Biopolymers* 24:527–546.
43. Stumpff-Kane AW, Maksimiak K, Lee MS, Feig M (2008) Sampling of near-native protein conformations during protein structure refinement using a coarse-grained model, normal modes, and molecular dynamics simulations. *Proteins* 70:1345–1356.
44. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. *Proteins* 35:133–152.
45. Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22:79–86.
46. Ming DM, Wall ME (2005) Quantifying allosteric effects in proteins. *Proteins* 59:697–707.
47. Kidera A, Go N (1992) Normal mode refinement: Crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. *J Mol Biol* 225:457–475.
48. Schomaker V, Trueblood KN (1968) On the rigid-body motion of molecules in crystals. *Acta Crystallogr B* 24:63–76.
49. Brent RP (2002) *Algorithms for Minimization Without Derivatives* (Dover, Mineola, NY).
50. Yang LW, *et al*. (2006) oGNM: Online computation of structural dynamics using the Gaussian Network Model. *Nucleic Acids Res* 34:W24–W31.