

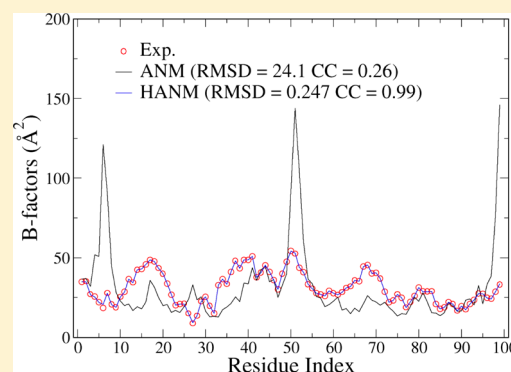
Robust Heterogeneous Anisotropic Elastic Network Model Precisely Reproduces the Experimental B-factors of Biomolecules

Fei Xia,[†] Dudu Tong,[†] and Lanyuan Lu*

School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore, 637551

S Supporting Information

ABSTRACT: A computational method called the progressive fluctuation matching (PFM) is developed for constructing robust heterogeneous anisotropic network models (HANMs) for biomolecular systems. An HANM derived through the PFM approach consists of harmonic springs with realistic positive force constants, and yields the calculated B-factors that are basically identical to the experimental ones. For the four tested protein systems including crambin, trypsin inhibitor, HIV-1 protease, and lysozyme, the root-mean-square deviations between the experimental and the computed B-factors are only 0.060, 0.095, 0.247, and 0.049 Å², respectively, and the correlation coefficients are 0.99 for all. By comparing the HANM/ANM normal modes to their counterparts derived from both an atomistic force field and an NMR structure ensemble, it is found that HANM may provide more accurate results on protein dynamics.



1. INTRODUCTION

Protein function is closely related to its dynamics, which plays the key role in various biological processes such as ligand binding and allosteric structural change. Apart from the widely used molecular dynamics (MD) simulation, a conventional method for the investigation of protein dynamics is the normal-mode analysis (NMA) on the basis of all-atom force fields,^{1–5} which approximates the potential energy surface to be harmonic and describes the motions of macromolecules in collective ways. Numerous successful applications of NMA to the studies of protein dynamics^{6–15} have shown that several lowest-frequency vibrational normal modes are responsible for the dominant conformational changes of proteins near the equilibrium state. However, the computational cost of NMA based on all-atom force fields is still considerable for large biological systems, and the required energy minimization for NMA might lead to structure deformation from what observed in experiments.

To overcome the difficulties encountered in the all-atom NMA, the coarse-grained (CG) elastic network models (ENMs), typically at the C α level, are often implemented for the studies of protein dynamics, generally referring to the Gaussian network model (GNM)^{16–18} and the anisotropic network model (ANM),¹⁹ as well as many other types of similar models developed on the basis of the original ideas of GNM/ANM. In both GNM and ANM, the residues in a protein are usually represented by the C α atoms along the backbone. Each pair of C α atoms is connected by a harmonic bond with a uniform force constant for the entire protein, while GNM and ANM employ slightly different potential energy functions. The theory of GNM assumes that the probability distribution of the fluctuation vectors of C α atoms in the network is isotropic and

follows Gaussian distributions so that the isotropic fluctuations of C α atoms can be directly estimated from the Kirchhoff matrix that is related to the network connectivity.¹⁷ The ANM takes the anisotropic fluctuations of C α atoms into account, and it has been implemented to model the high resolution X-ray measurements with the anisotropic Debye–Waller factors, or B-factors.²⁰ GNM and ANM have been shown to yield reasonably good results compared to experimental B-factors. The average correlation coefficient (CC) for 113 proteins is 0.59 with GNM and 0.55 with ANM.^{21,22} Note that in GNM/ANM, the directions of the normal modes are solely determined by the topology information, so is the CC. As a result, GNM/ANM produces the experimental B-factors instead of fitting them by tuning the model's parameters. The advantages of ENMs lie in the fact that for NMA not only the requirement of energy minimization is eliminated but also the computational cost is substantially reduced. Additionally, the potential energy function is constructed on the basis of experimental structure, and no force field inaccuracy is involved. Moreover, the lowest-frequency vibrational modes calculated from ENMs accord with the slow conformational changes of proteins observed experimentally very well.^{23–28}

Due to the great amount of successes in the applications^{29–35} of ENMs to protein conformational dynamics, many researchers put their efforts to improve the validity of ENMs. It is expected that an accurate ENM should reproduce the experimentally observed B-factors, which are, in fact, usually the only physical quantities that link the ENM parameters and the experiments in structural biology. To achieve this goal, more

Received: January 17, 2013

Published: July 3, 2013



sophisticated ENMs have been developed based on various strategies. One category of the methods tends to mimic the experimental crystallization environment in computer modeling, such as taking into account the rigid body motions and the crystal packing effect.^{36–43} For example, Kundu et al.⁴¹ considered the crystal packing effect by explicitly including the neighboring molecules in the crystal lattice in GNM, with a CC improved from 0.59 to 0.66. Song and Jernigan⁴² took both the rigid body motions and the crystalline packing effect into account and developed a new version of GNM called vGNM, in which they postulated that the amplitudes of the internal protein motions were not necessarily proportional to the inverses of the eigenvalues. In this approach, the least-squares fit to experimental B-factors with three parameters corresponding to the translational, rotational, and internal motions leads to an improved average CC of 0.80 for the proteins. The translation libration screw (TLS) model originally introduced by Schomaker and Trueblood³⁶ emphasizes the significant contributions of rigid body motions to the crystallographic B-factors in the crystalline environment, rather than those of the internal motions of proteins. The extended TLS model proposed by Soheilifard et al.⁴³ with a modification on the flexible chain ends generates an average CC up to 0.88 for a set of 300 proteins. Although the B-factor results can be improved in these approaches, the interactions are much more complex than those in the original GNM or ANM, which usually causes additional computational difficulties and complicates the physical picture of ENM. Alternatively, the B-factor results can be enhanced by applying some moderate modifications within the ENM framework on the potential energy functions or by considering the spring heterogeneity in ENM.^{44–55} However, the computed B-factors from this type of approaches still have substantial room for improvement, except those from Erman.⁴⁷ In Erman's approach, an iterative algorithm was developed to update the Kirchhoff matrix of GNM, in which the connections of neighboring $C\alpha$ atoms on the backbone are fixed, and the strength of the interactions between pairs of residues are varied until it gives a good B-factor fit to the experimental values. While this approach generates computed B-factors that are very close to their experimental counterparts, a considerable portion of the spring force constants have the opposite sign, indicating some effective forces that pull the residue pairs away from their native pairwise distances. Recently, a study upon the formalism of ENMs reported by Leioatts and Grossfield⁵⁶ claims that the formalism complexity of interaction potentials has little bearing on the accuracy of ENMs, while the fitting of spring parameters can make a substantial improvement. Finally, it should be noted that ENM is actually anharmonic in terms of Cartesian coordinates, and the computed B-factors are somehow different from the NMA results if the higher order terms are considered.⁵⁴

Therefore, following the simplicity of the conventional ENMs and considering the heterogeneity of force constants, in this article, a new method called the progressive fluctuation matching (PFM) is devised to construct a robust heterogeneous anisotropic elastic network model (HANM) with individual force constants, which relies on the normal-mode analysis based fluctuation matching (NMA-FM) method that was proposed before.⁵⁷ The central idea of PFM is to use restraint potentials to iteratively update the calculated B-factors, progressively approaching the experimental values. The NMA-FM method⁵⁷ can “convert” an arbitrary fine-grained model to a coarse-grained ENM by means of NMA and fluctuation

matching.^{58,59} Although NMA-FM was originally developed for CG modeling, coarse-graining is actually not required in NMA-FM, as the method can also transfer two potential energy functions at the similar resolution from one to another. In the PFM approach, apart from an ANM potential, a set of potential restraints are applied to the $C\alpha$ atoms. These restraints are harmonic in terms of the Cartesian coordinates, and they modify the total potential to generate B-factors closer to the experimental targets. The NMA-FM method is subsequently implemented to convert the amended potential to an HANM. The HANM can be iteratively refined in the same manner until the final HANM yields satisfactory results relative to experiments. The PFM method completely includes the merits of the previous NMA-FM method, as the method is computationally cheap and no MD simulation is required, in comparison with some other methods.^{60,61} More importantly, the HANM derived from the PFM method reproduces the experimental B-factors precisely with very high CCs, dramatically outperforming most of the ENMs mentioned. To the best of the authors' knowledge, it is the first time that the experimental B-factors are accurately reproduced by an ENM model, yet with physically meaningful force constants. In HANM, the interaction potentials are still harmonic and the crystalline effect is implicitly incorporated in the heterogeneous force constants.

The rest of this paper is organized as follows: section 2 gives the theoretical background and the method. In section 3, the applications of HANM to four proteins crambin, trypsin inhibitor, HIV-1 protease, and lysozyme are discussed. Section 4 is for conclusion and future work.

2. Theory and Computational Details. 2.1. Elastic Network Model. The total potential V of an ANM is expressed as the sum of the harmonic interactions as

$$V_{\text{ANM}} = \sum_i^N \sum_{j>i}^N \frac{1}{2} k_u (r_{ij} - r_{ij}^0)^2 H(r_c - r_{ij}^0) \quad (1)$$

where k_u is a uniform harmonic force constant, N is the total number of particles, r_{ij} is the distance between particles i and j , r_{ij}^0 is the equilibrium distance between particles i and j in the crystal structure, r_c is the cutoff distance, and $H(x)$ is the Heaviside unit step function. $H(x) = 0$ if $x < 0$; otherwise $H(x) = 1$. In general, one can adjust the parameters of ANM to obtain the best fitting to the experimental B-factors. In this work, an ANM is generated from a web-server²² as the starting point to derive an HANM. The final HANM has the potential energy formulizm similar to eq 1 as

$$V_{\text{HANM}} = \sum_i^N \sum_{j>i}^N \frac{1}{2} k_{ij} (r_{ij} - r_{ij}^0)^2 H(r_c - r_{ij}^0) \quad (2)$$

where the spring constant k_{ij} denotes the heterogeneous force constant of the pair of particles i and j . Obviously, ANM can be considered as a special case of HANM with uniform force constants.

2.2. NMA-FM Method. The details of NMA-FM have been presented elsewhere,⁵⁷ and here, a brief review is given. The NMA-FM method aims at matching the fluctuations from the fine-grained input model and the CG output model. Both sets of fluctuations are calculated by NMA, in which the standard eigenvalue equations³² of a system composed of N particles are solved:

$$\mathbf{H}\mathbf{U} = \mathbf{U}\mathbf{\Lambda} \quad (3)$$

where \mathbf{H} is the mass-weighted Hessian matrix, $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues, and \mathbf{U} is the orthonormal matrix formed by the eigenvectors u_k ($1 \leq k \leq 3N$). The elements H_{ij} of the Hessian matrix are calculated as the second derivatives of the total potential energy V in eq 1 or 2, namely, $\partial^2 V / \partial x_i \partial x_j$ with respect to the atomic Cartesian displacements x_i and x_j ($1 \leq i, j \leq 3N$). The mean square fluctuations of the particles are readily calculated according to classical mechanics:⁶²

$$\langle \Delta R_i^2 \rangle = \frac{k_B T}{m_i} \sum_j^{\text{modes}} \frac{u_{ij}^2}{\omega_j^2} \quad (4)$$

where ΔR_i and m_i are the displacement and the mass of the i th particle, k_B is the Boltzmann's constant, T is the temperature, and u_{ij} is the Cartesian subvector of the projection of the j th eigenvector with the frequency ω_j on particle i . The summation in eq 4 runs over all $3N - 6$ vibrational normal modes for the calculation of the atomic fluctuations ΔR_i . In the process of fluctuation matching, the bond fluctuations $\Delta r_{ij,FG}^2$ and $\Delta r_{ij,CG}^2$ corresponding to the input model and the output one are evaluated by projecting the normal modes on the pairwise bonds connecting particles i and j .⁶² The spring constants between all pairs of beads i and j are simultaneously estimated and updated according to

$$\frac{1}{k_{ij}^{n+1}} = \frac{1}{k_{ij}^n} - \varepsilon (\Delta r_{ij,FG}^2 - \Delta r_{ij,CG}^2) \quad (5)$$

where n labels the iteration number and the parameter ε controls the size of the steps in the optimization of the spring constants. The procedure repeats until the difference between the bond fluctuations is less than a predefined threshold value σ . In practice, the threshold is defined as the relative deviation of the bond fluctuations, calculated as $\sigma = (|\Delta r_{ij,FG}| - |\Delta r_{ij,CG}|) / |\Delta r_{ij,FG}|$. If all of the calculated σ values are lower than 10^{-3} , the convergence of the spring constants is reached and the iteration process stops.

In PFM, the input model is an ANM/HANM with additional harmonic restraints, while the output model is an HANM defined in eq 2. Although no coarse-graining occurs during the calculation, the procedure described is still valid.

2.3. PFM Method. For an HANM, the theoretical B-factors of the $C\alpha$ atoms are calculated as $B_i^{\text{cal}} = (8\pi^2/3) \langle \Delta R_i^2 \rangle$ compared to the experimental B-factors B_i^{exp} , where i denotes the residue index of the $C\alpha$ atom and $\langle \Delta R_i^2 \rangle$ is calculated according to eq 4. To find out a set of heterogeneous force constants that can produce the correct B-factor values, the iterative approach called PFM is adopted, and the initial guess ENM (i.e., ANM in this article) or an intermediate HANM is refined in each round of the iterations. More specifically, the refinement includes two steps: (a) to add some potential restraints to the input ENM and obtain improved B-factors using the new potential energy function; (b) to convert the output from step a to a HANM without restraints on the basis of the NMA-FM method.⁵⁷

In the first step of the refinement, a restraint potential V_i^{res} is added to an input ENM to produce the computed B-factors closer to the experimental ones. The simplest harmonic restraint potential is used here, where the restraint potential $V_i^{\text{res}} = 1/2 k_i^{\text{res}} \Delta R_i^2$ is imposed on the i th atom. The amplitude of the restraint force constant k_i^{res} imposed on the i th $C\alpha$ atom is determined according to the equation

$$k_i^{\text{res}} = \beta k_B T \cdot 8\pi (B_i^{\text{cal}} - B_i^{\text{exp}}) / (B_i^{\text{cal}} B_i^{\text{exp}}) \quad (6)$$

where β is a numerical scale factor to adjust the restraint strength, k_B is the Boltzmann's constant, and T is the temperature. The rationale of eq 6 can be briefly described as follows: Near the equilibrium position, the motion of particle i can be regarded as harmonic oscillations around its equilibrium position with an effective force constant k_i . According to the equipartition theorem, the ensemble average of atom i 's potential energy is $3/2 k_B T$, thus the effective force constants under experimental and calculated fluctuations can be derived from the relation $1/2 k_i^{\text{exp}} \langle \Delta R_i^2 \rangle = 1/2 k_i^{\text{cal}} \langle \Delta R_i^2 \rangle = 3/2 k_B T$. Here, a restraint potential should be added to fill the gap between the experimental and computational positional fluctuations, neglecting the correlation between different $C\alpha$ atoms. So, the restraint force constant $k_i^{\text{res}} = k_i^{\text{exp}} - k_i^{\text{cal}}$, and the expression of k_i^{res} can be obtained as eq 6. In practice, the parameter β , which is chosen to be 0.3 in the PFM algorithm, is added in eq 6 to scale the calculated k_i^{res} ; that is, only a small correction is added in each PFM iteration to ensure the convergence. This is due to the following reasons: (a) The above estimation may be inaccurate because of the correlation between the atoms. The restraint imposed on atom i also alters the fluctuations of the other atoms because of the ENM potential. Thus, too large restraints may result in unphysical spring constants (e.g., negative ones). (b) There might be multiple solutions for the given experimental B-factors. Therefore, gradually adding small perturbations starting from the conventional ANM should give rise to a set of physically meaningful force constants. Actually, reasonable force constants were obtained for all the molecular systems tested so far, including many cases not shown in this article. As discussed later, the solution is even not very sensitive to the initial ANM spring constant. In principle, a restraint force constant can be further decomposed into its three Cartesian components to reproduce the high-resolution anisotropic B-factors,²⁰ whereas all restraints are isotropic in this work.

The next step is to convert the potential with the restraints to the more physical HANM using NMA-FM. In NMA, for a system with the total potential energy $V_R = V + V^{\text{res}}$, the normal modes can be computed by solving the following equation

$$\mathbf{H}_R \mathbf{U}_R = \mathbf{U}_R \mathbf{\Lambda}_R \quad (7)$$

where \mathbf{H}_R , \mathbf{U}_R , and $\mathbf{\Lambda}_R$ are the counterparts of the items in eq 3 for a restrained system. The pairwise bond fluctuations $\Delta r_{ij,Res}^2$ are then calculated by projecting the modes on the bonds. These bond fluctuations are taken as the target to derive an HANM by fluctuation matching, as presented in section 2.2. In summary, given an initial ANM as the start point, the overall numerical procedure of the PFM method is outlined briefly:

- (1) For a given ANM/HANM, calculate the restraint force constant k_i^{res} for every atom i according to eq 6.
- (2) Add the total restraint potential V^{res} to the original ANM/HANM potential, and solve eq 7 to obtain the pairwise bond fluctuations $\Delta r_{ij,Res}^2$.
- (3) Set the bond fluctuations $\Delta r_{ij,Res}^2$ as the target fluctuations (i.e., $\Delta r_{ij,FG}^2$ in section 2.2), and derive a new HANM through fluctuation matching.

Steps 1~3 are repeated until the largest correction on the calculated B-factors in an iteration is smaller than 0.5%. The input potential in the first iteration is the conventional ANM, while in the rest of the iterations the HANM as the output of

the previous step is further refined. Steps 2 and 3 are based on the NMA-FM method.

All Hessian matrices were solved analytically¹⁹ with and without restraint potentials. The mathematical operations in the normal-mode analyses were performed using the related routines in the GNU Scientific Library (<http://www.gnu.org/software/gsl/>). The PFM method was implemented in a C language program attached in the Supporting Information.

The current approach adopts an ANM uniform force constant as the initial guess of the model, and adjusts the individual force constants according to the potential restraints suggested by the atomistic B-factor values. During the optimization process, both eigenvalues and eigenvectors from the model are modified to ensure that the projections of the normal modes on the atoms are consistent with the experimental atomistic fluctuations (B-factors). In this sense, HANM can be considered as a perturbative improvement of the conventional ANM. Nevertheless, the PFM method described is indeed general, and besides ANM, many ENM-type models can be selected as the initial input of the protocol. For instance, the PFM method can in principle be implemented to refine an ENM with distance-dependent force constants.⁵³

To quantify the performance of the derived HANM, the correlation coefficients (CC) of the studied systems were calculated according to the definition

$$CC = \frac{\sum_{i=1}^N (B_i^{\text{exp}} - B_{\text{ave}}^{\text{exp}})(B_i^{\text{cal}} - B_{\text{ave}}^{\text{cal}})}{\sqrt{\sum_{i=1}^N (B_i^{\text{exp}} - B_{\text{ave}}^{\text{exp}})^2 \sum_{i=1}^N (B_i^{\text{cal}} - B_{\text{ave}}^{\text{cal}})^2}} \quad (8)$$

where B_i^{exp} and B_i^{cal} are the experimental and calculated B-factors for atom i , while $B_{\text{ave}}^{\text{exp}}$ and $B_{\text{ave}}^{\text{cal}}$ are the corresponding averaged values of the N atoms.

Another quantity to directly measure the difference between the calculated B-factors and the experimental ones is the root mean squared deviation (RMSD) that is defined as

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N (B_i^{\text{cal}} - B_i^{\text{exp}})^2}{N}} \quad (9)$$

where N is the total number of $C\alpha$ atoms, i denotes the i th atom in a model, and the sum counts the deviations of B-factors between calculated and experimental values for all atoms.

To measure the similarity of two normal modes, the cosine of the angle between the two eigenvectors are calculated, according to the formula

$$|\cos \theta| = \frac{|\sum_i^n a_i b_i|}{|\mathbf{a}| \cdot |\mathbf{b}|} \quad (10)$$

where \mathbf{a} and \mathbf{b} are two eigenvectors, a_i and b_i represent the components on the i th atom's three $x/y/z$ directions, and θ is the angle between \mathbf{a} and \mathbf{b} . The absolute value of the cosine ranges from 0 to 1. The larger this value is, the more similar the two modes are.

Due to the possible mode shuffling, the cumulative overlap⁶³ C_i^n is computed to indicate that to what extent the first n modes in the second model reflect the motion of the i th mode in the first model. The definition of the overlap is

$$C_i^n = \sqrt{\sum_j^n (u_i v_j)^2} \quad (11)$$

where u_i is the i th mode from the first model and v_j is the j th mode from the second model. The number n denotes the first n modes from the second model, and the result determines the percentage for the description of the i th mode from the first model with the set of n modes from the second model. A cumulative overlap of 1 means that the subspace spanned by the i th mode from the first model can be completely described by the first n modes from the second model. In this article, the cumulative overlaps are calculated to study the similarity between the modes from atomistic force field and from HANM/ANM, as well as between the principal components from NMR ensemble and from HANM/ANM. In literature, the spanning coefficient,⁶⁴ which equals to the square of C_i^n , was also used to illustrate the mode similarity of two models.

3. RESULTS AND DISCUSSION

3.1. Convergence and Cutoff of HANM. To estimate the performance of our method, the previously studied^{18,45} protein crambin (PDB ID: 1CRN)⁶⁵ with 46 residues was selected as a test system. Figure 1a shows the calculated B-factors based on HANM after certain iterations using the PFM method, compared with the results of ANM and the experiment. In the plot, the black curve denotes the initial B-factors calculated from ANM, in which the optimal force constant was directly taken from the online ANM web-server²² and the cutoff was set as the default value 15 Å. By comparing the results of ANM with the experimental ones, the most obvious difference is reflected by the sharp peak around the residue index 19. It is obvious that ANM substantially overestimates the B-factors for the residue indices 18–20 and 36–39. In addition, the B-factors in the regions with the residue indices 1–3 and 25–34 are slightly underestimated. By applying the PFM method, the calculated B-factors with the indices 18–20 and 36–39 are quickly reduced after 5 iteration cycles, as shown by the magenta curve in Figure 1a. The B-factors for other residue indices are also improved a little over the black curve. After only 10 cycles, the B-factors corresponding to the green curve exhibit a considerable enhancement in the regions with the residue indices 1–3, 18–20, 25–35, and 36–43. Finally after 22 cycles, the calculated B-factors from the HANM of crambin have already converged below the threshold, and the results are quite close to the experimental values. The test for crambin indicates that the PFM method is able to efficiently improve the calculated B-factors at the cost of a small number of iterations. It has to be mentioned that the parameter β in eq 6 for updating the restraint force constants has been optimized so that it can efficiently speed up the convergence of the calculated B-factors to the experimental ones. The PFM method is further validated by constructing the HANMs with different cutoff distances. It is known that ANM often adopts a large cutoff distance in the range 15–24 Å.²² In order to demonstrate the robustness of the PFM approach on different cutoff distances, the cutoff values of 13–16 Å were selected to construct the HANMs of crambin and estimate the respective B-factors. Figure 1b displays the calculated RMSDs between the experimental and calculated B-factors using HANMs with the cutoff values in the range 13–16 Å. The comparison indicates that the evaluated RMSDs based on different cutoffs are similar with each other, with the close values 0.063, 0.060, 0.060, and 0.044 Å² for 13, 14, 15, and 16 Å, respectively. The larger cutoff distances lead to the slightly smaller RMSDs, which agrees with the observed tendency of ANMs in the previous study.²² Due to the robustness of PFM and the default value 15 Å used for

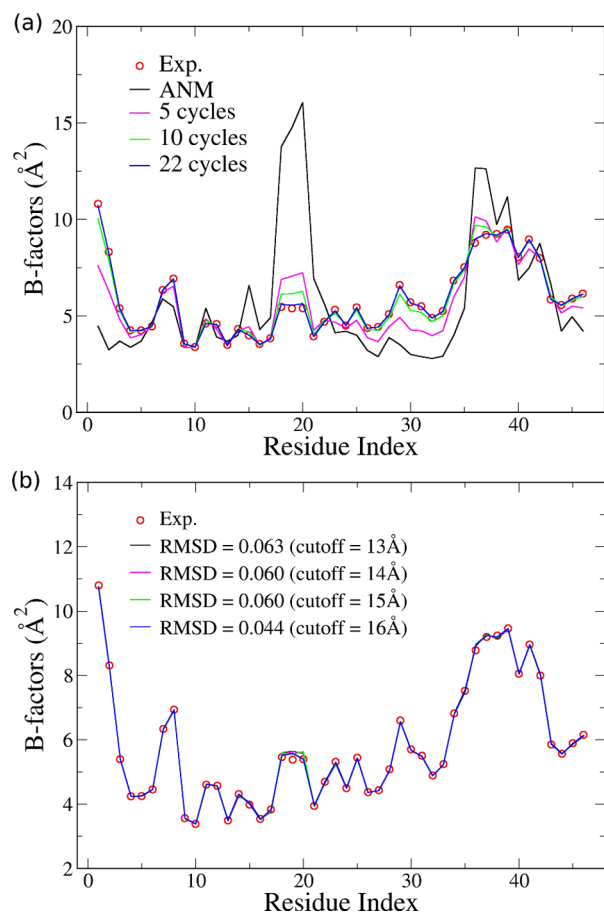


Figure 1. (a) Comparison of the experimental and calculated B-factors of crambin. The experimental B-factors are denoted by the red line, while the black curve is calculated from ANM with the cutoff 15 Å. The magenta, green, and blue curves denote the calculated B-factors based on HANM with the cutoff 15 Å after 5, 10, and 22 iterations, respectively. (b) Comparing the experimental B-factors with the results calculated from the HANMs of crambin with different cutoff distances 13, 14, 15, and 16 Å, denoted by the black, magenta, green, and blue curves. The RMSDs between the experimental B-factors and the calculated converged results are 0.063, 0.060, 0.060, and 0.044 Å² for 13, 14, 15, and 16 Å, respectively.

ANM, only the results of HANM with the cutoff distance 15 Å are presented in the rest part of the article.

Additionally, the HANM force constants derived from two different initial ANMs are compared, as shown in Figure 2. The comparison in Figure 2a indicates that the different initial guesses lead to very similar HANM force constants in PFM. A linear regression of the data points in Figure 2a generates a slope of 0.93 for the red straight line and a correlation coefficient 0.97, demonstrating the high similarity between the two sets of force constants. The two sets of force constants k_x and k_y generate nearly the same curves of B-factors, as shown in Figure 2b. It is natural to believe that a given set of B-factors for a protein corresponds to multiple solutions of HANM. For instance, it is trivial to calculate the force constants of the atomic restraints in eq 6 according to the equipartition theorem, if the ENM part of potential is zero. In this case the calculated B-factors can be perfectly consistent with the experimental values, whereas the model is of little interest because it consists of unphysical positional restraints. The solution from Erman⁴⁷ also reproduces the experimental B-

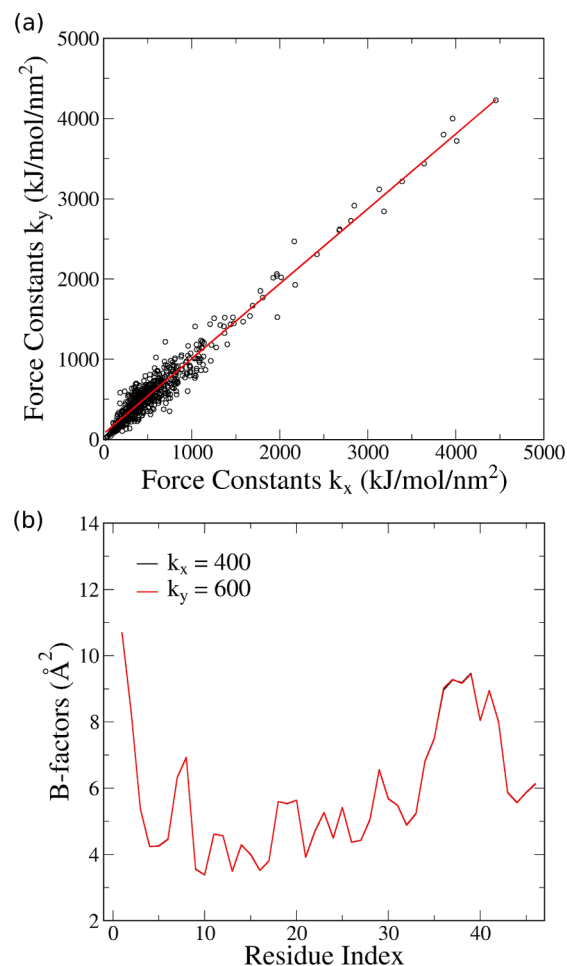


Figure 2. (a) Comparison of two sets of force constants k_x and k_y derived from the ANMs with different initial force constants 400 and 600 kJ/(mol·nm²) after convergence using the PFM method. The linear regression generates a slope 0.93 as the red line. (b) Comparison of the B-factors calculated from the two sets of force constants k_x and k_y .

factors at reasonable precision, while many force constants have the opposite sign unlike the PFM case. A possible reason for the unphysical force constants from Erman's approach is that the algorithm has an oversimplified physical picture, in which *all* force constants connecting to an atom are increased, if the calculated fluctuation of the atom is larger than the experimental value. The iterative algorithm in PFM is based on fluctuation matching, and it starts from the widely accepted ANM model. Thus, PFM is able to improve the B-factor results when the original ANM is not accurate enough, and the method generates physically reasonable force constants as shown in the molecular examples. A more extensive study on the influence from the initial ANM parameter is necessary.

The fact that HANM in general more accurately reproduces the experimental atomic positional fluctuations can be understood as the nonuniform force constants provide more degrees of freedom for the fitting procedure. More importantly, the results illustrate that the effective interactions between the residues are heterogeneous in nature, which is discussed later. The importance of heterogeneity is not only reflected in the B-factor fitting, but also revealed by the more sophisticated analysis below on protein motions. In the next section, a comprehensive comparison between ANM and HANM is

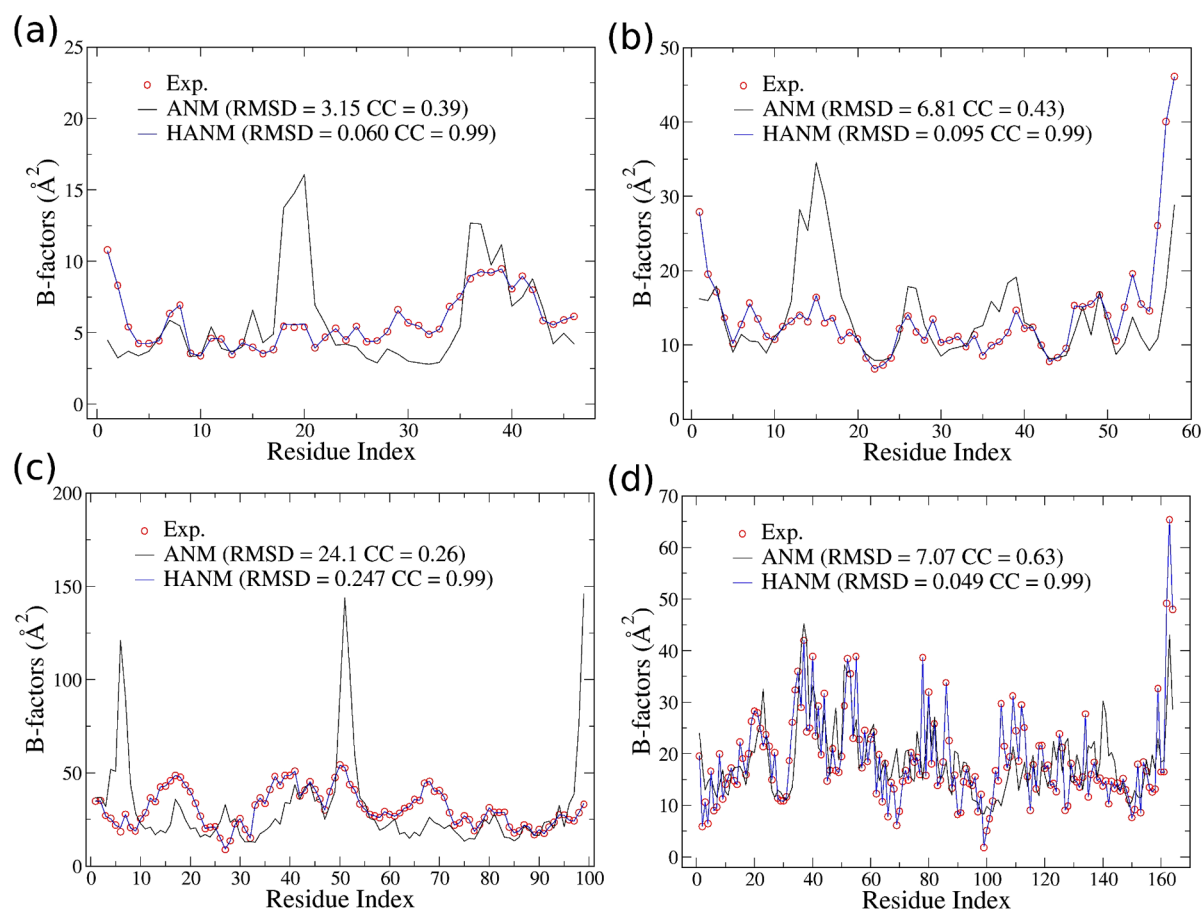


Figure 3. Comprehensive comparisons of the experimental and calculated B-factors for (a) crambin, (b) trypsin inhibitor, (c) HIV-1 protease, and (d) lysozyme. The experimental values are denoted by the hollow red cycles, and the results of ANM and HANM with the cutoff distance 15 Å are denoted by the black and blue curves. The calculated RMSDs for ANM and HANM are 3.15 vs 0.060 Å² (a), 6.81 vs 0.095 Å² (b), 24.1 vs 0.247 Å² (c), and 7.07 vs 0.049 Å² (d). The calculated CCs from ANM are 0.39, 0.43, 0.26 and 0.63, respectively, while all CCs from HANM are 0.99.

provided for four proteins: crambin, HIV-1 protease, trypsin inhibitor, and lysozyme.

3.2. B-factor Comparison of ANM and HANM. Figure 3 presents a comprehensive comparison of experimental and theoretical B-factors of four proteins. Apart from the crambin system mentioned above, other tested proteins are trypsin inhibitor with 58 residues (PDB ID: SPTI),⁶⁶ HIV-1 protease with 99 residues (PDB ID: 1HHP),⁶⁷ and lysozyme with 164 residues (PDB ID: 2LZM).⁶⁸ All the results were derived from ANM and HANM with the cutoff 15 Å and the PFM final converged data. For crambin in Figure 3a, it is evident that the ANM prediction is considerably different from the experimental data. Two artificial large peaks on the computed curve exist in the region with the residue indices 18–20 and 36–39. The calculated RMSD and CC from ANM are 3.15 Å² and 0.39, while the HANM has a much smaller RMSD 0.060 Å² and a dramatically higher CC 0.99, indicating that the calculated B-factors matches the experimental curve at high precision.

Figure 3b shows the results for another small protein trypsin inhibitor. The B-factors are somewhat overestimated for the atoms with the residues indices 13–17, 26–27, and 36–39, and the results are underestimated in the region with the indices 52–58. The results of ANM lead to a larger deviation value with the RMSD 6.81 Å², while the calculated RMSD from the HANM for trypsin inhibitor is only 0.095 Å². For all Cα atoms in trypsin inhibitor, the calculated HANM B-factors with the high CC 0.99 accord very well with the experimental data.

The third case is the biologically important HIV-1 protease whose function has been widely studied before.²² However, ANM fails to produce a satisfactory B-factor curve for HIV-1 protease. In Figure 3c, the B-factors predicted by the ANM for the atoms with the residue indices 5–6, 50–51, and 98–99 are extremely high, up to 150 Å², which conflicts with the physical scenario of atomic fluctuations indicated by the crystal B-factors. The RMSD calculated for HIV-1 protease reaches 24.1 Å², and the CC is as low as 0.26. Nevertheless, the B-factors calculated from HANM are in good agreement with the experimental information. The calculated RMSD and CC are 0.247 Å² and 0.99 respectively, dramatically higher than the ANM values. The results for HIV-1 protease strongly demonstrate the advantage of the PFM method and encourage us to apply it to more proteins.

The last example involves the larger protein lysozyme, whose structures have been widely investigated.^{69–72} Figure 3d shows the B-factor curves calculated from ANM and HANM. The computed RMSD and CC are 7.07 Å² and 0.63 for ANM, and 0.049 Å² and 0.99 for HANM, respectively. Using the PFM method, the obtained CCs for the four systems are around 0.99, and all RMSDs are below 1.0 Å². More applications of PFM for the construction of HANMs of proteins are shown in Figure S1 of the Supporting Information.

Apart from the applications on the basis of X-ray B-factors, the PFM method can also be used to construct HANM according to the effective B-factors from an NMR structure

ensemble. Figure 4 shows the B-factor results of HANM/ANM and the effective B-factors calculated from the NMR measure-

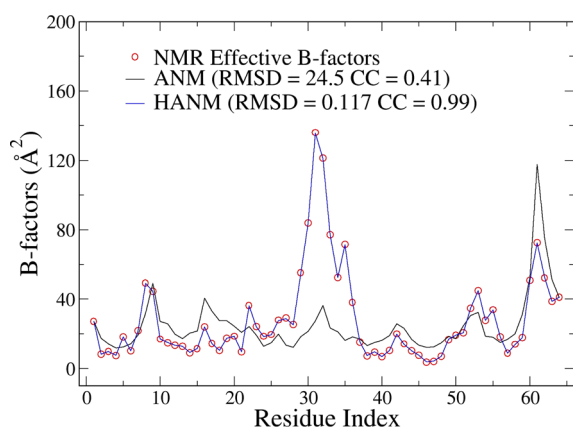


Figure 4. Comparison of the effective B-factors of the β -neurotoxin (PDB ID: 1B3C) with the calculated results of HANM and ANM. The effective B-factors denoted by the red circles are derived from the mean square fluctuations estimated from an ensemble of NMR structures of the beta-neurotoxin in solution. The calculated RMSDs/CCs for HANM (blue line) and ANM (black line) are 0.117 Å²/0.99 and 24.5 Å²/0.41, respectively.

ment of the beta-neurotoxin (PDB ID: 1B3C).⁷³ The average structure of the ensemble of the beta-neurotoxin in solution was taken as the reference structure for the construction of HANM and ANM. The effective B-factors were derived from the mean squared fluctuations of the NMR ensemble of structures, considering the average structure as the reference. As shown in Figure 4, the experimentally derived effective B-factor curve has several peaks at the indices 31, 32, 53, and 61. The blue fitting curve using the PFM method generates a higher CC 0.99 and a lower RMSD 0.117 Å², compared with the 0.41 and 24.5 Å² of ANM. This case validates the application of the PFM method to the construction of HANM on the basis of experimental NMR data.

In ANM/GNM, the CC between experimental and computed B-factors is an intrinsic property of the protein topology, which is not related to the single spring constant. By contrast, the per-protein spring constant in ANM/GNM is typically adjusted to minimize the B-factor RMSD.²² While the overall scales of the normal modes are optimized during the optimization process, the directions of the modes are intact. Compared to the single-parameter models, both the CC and the RMSD from HANM depend on the model parametrization on the basis of individual protein B-factor data. The increased complexity of HANM allows heterogeneous interactions and incorporates the experimental information better, resulting in drastically enhanced CC and RMSD.

3.3. Normal Mode Comparison of ANM and HANM.

Since HANM is able to correctly reflect the experimental fluctuations and those fluctuations can be considered as the projections of the normal modes on the atoms, it is very interesting to compare the modes generated by HANM and the conventional ANM and to investigate whether HANM could generate more reasonable protein dynamics. To validate the derived modes of HANM in the study of protein dynamics, NMA for the proteins crambin, trypsin inhibitor, HIV-1 protease, and lysozyme were performed using ANM and HANM, respectively, and the normal modes from the two

models were compared. The calculated maximal overlaps of the first five eigenvectors between ANM and HANM are listed in Table 1. It shows that the modes of proteins from ANM and

Table 1. Maximal Overlaps of the First Five Lowest-Frequency Modes of HANM and ANM Calculated According to Equation 10

| crambin | trypsin inhibitor | HIV-1 protease | lysozyme |
|---------|-------------------|----------------|----------|
| 0.812 | 0.601 | 0.736 | 0.979 |
| 0.951 | 0.591 | 0.725 | 0.988 |
| 0.221 | 0.751 | 0.457 | 0.962 |
| 0.482 | 0.068 | 0.160 | 0.883 |
| 0.492 | 0.011 | 0.439 | 0.604 |

HANM have reasonable overlaps with each other, especially for the first two low-frequency modes with the values greater than 0.6.¹² For lysozyme, the first two modes of ANM and HANM have the high overlaps larger than 0.95 and completely reproduce the typical bending and twisting modes,⁷¹ which are shown in Figure 5, parts a and b, respectively. Considering

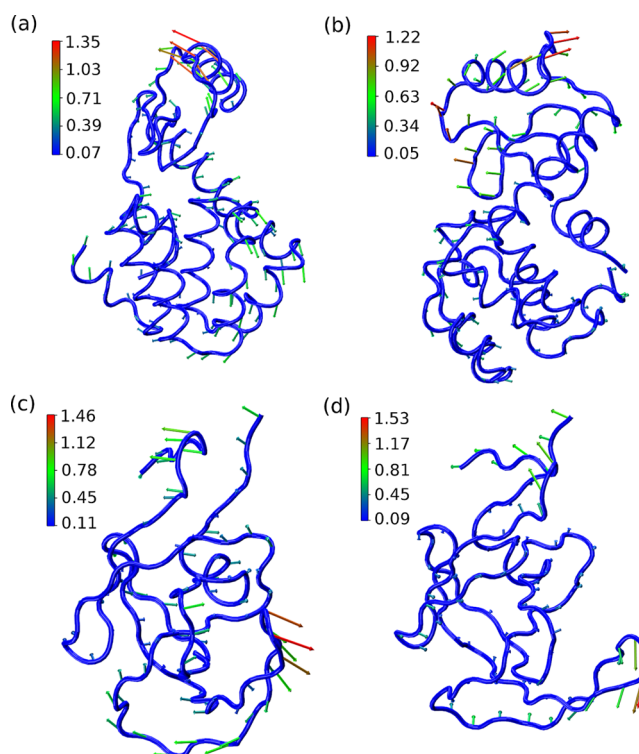


Figure 5. Porcupine plots indicating the motion directions and amplitudes of the normal modes, which are calculated by HANM. Length and color of the arrows show the motion amplitude, while the direction of the arrows shows the motion direction. (a and b) The seventh and eighth normal modes of Lysozyme. The two modes can be clearly identified as bending (seventh) and twisting (eighth) modes. (c and d) The seventh and eighth normal modes of HIV-1 protease.

the possibility of the mode shuffling, it is also necessary to compare the mode from one model to the nearby modes from a different model. More calculated overlaps between the modes 7–11 of ANM and HANM are provided in Table S1 of the Supporting Information, and the conclusion from these data is similar to that from Table 1. It is not surprising that the modes of lysozyme of HANM and ANM accord very well with each

other since ANM reproduces the experimental B-factors with the CC 0.63, as shown in Figure 3d.

However, the results obtained for HIV-1 protease reveal a large gap between the modes of ANM and HANM. To validate the modes derived from HANM, we compared the similarity of the modes from an all-atom model and HANM as well as ANM for HIV-1 protease. We used the Gromos96 vacuum parameter set⁷⁴ as the all-atom model, and NMA were performed to generate the corresponding all-atom modes for the C α atoms. The calculated maximum overlap values for the modes 7–11 of the all-atom model with the first five lowest-frequency modes of HANM and ANM are listed in Table 2. All modes of HANM

Table 2. Maximal Overlaps between the Modes 7–11 of the All-Atom Model of HIV-1 Protease and the First Five Lowest-Frequency Modes of HANM and ANM Calculated According to Equation 10

| all-atom modes | HANM | ANM |
|----------------|-------|-------|
| 7 | 0.739 | 0.606 |
| 8 | 0.709 | 0.664 |
| 9 | 0.421 | 0.168 |
| 10 | 0.243 | 0.248 |
| 11 | 0.435 | 0.256 |

except mode 10 have overlap values larger than the corresponding ANM ones, and the HANM overlap values for the first two modes are larger than 0.7. Figure 5c,d display the modes 7 and 8 of HIV-1 protease calculated from HANM. The mode 7 of HANM in Figure 5c indicates a relatively twisting motion between the top and bottom regions of HIV-1 protease, while the mode 8 exhibits a collective bending motion between the two regions. More calculated overlaps of the modes 7–11 between the all-atom model and HANM as well as ANM are provided in Table S2 of the Supporting Information, including the overlaps between adjacent modes. These data also suggest that the mode similarity between the all-atom force field and HANM is higher. In addition, we calculated the cumulative overlaps for the modes of the all-atom model with the modes of HANM and ANM. The cumulative overlaps reflect how the subspaces spanned by the all-atom modes are described by another set of modes of HANM and ANM. The calculated cumulative overlaps for the first five modes 7–11 of the all-atom model are listed in Table 3. The results show that the overall cumulative overlaps of HANM are larger than those of ANM. Only one value 0.542 of HANM is smaller than 0.7, which indicates that the significant lowest-frequency modes

Table 3. Calculated Cumulative Overlap C_i^n of the Modes 7–11 of HIV-1 Protease^a

| C_i^n ($n = 5, i = 7\sim 11$) | HANM | ANM |
|-----------------------------------|-------|-------|
| 7 | 0.770 | 0.721 |
| 8 | 0.799 | 0.769 |
| 9 | 0.542 | 0.384 |
| 10 | 0.806 | 0.630 |
| 11 | 0.728 | 0.522 |

^aThe overlaps are between the all-atom model modes and the HANM/ANM ones. The subscript i denotes the i th mode of the all-atom model, while the superscript n denotes the number of modes of HANM and ANM for the calculation of the sum in eq 11. The first five lowest-frequency modes of HANM and ANM are used for the calculation of the cumulative overlaps.

from the all-atom model are better expressed by the first 5 modes of HANM. Since the atomistic force field is not parametrized on the basis of B-factors, certain discrepancies between the results from the force field and from HANM/ANM are expected. Nevertheless, here, the comparison with the force field results is considered as a validation of our model because the force field is widely used for describing protein dynamics.

The discrepancy of the modes observed from ANM and HANM can be readily explained from the fitted results to the B-factors. As shown in Figure 3c, the uniform ANM extremely overestimates the B-factors at the indices 6, 50, and 99 apparently with three sharp peaks, even though the rest part of the curve is a reasonable approximation to the experimental one. It is well known that the lowest-frequency modes make the most significant contributions to the atomic fluctuations so that it can be easily understood that the incorrect fitting of B-factors for these residues might lead to the discrepancy of the collective modes.

Finally, we compared the HANM/ANM modes with the principal components that were calculated from the NMR structure ensemble of the β -neurotoxin, following the work of Yang et al.⁶³ The experimentally inferred protein motions were calculated by applying the principal component analysis (PCA)⁷⁵ to the NMR structures. Table 4 shows the cumulative overlaps, in which the first three NMR modes are expressed as 3, 6, and 20 HANM/ANM modes. The overlap values from the HANM model are obvious higher than their ANM counterparts, indicating that the similarity between the NMR and HANM modes is higher. The comparison on the basis of the individual modes shows the same conclusion, and these data are included in Table S3 and S4 of the Supporting Information. In this case HANM is directly validated by the NMR experimental data, which demonstrates the model's potential in the study of protein dynamics.

3.4. Heterogeneity of Force Constants. The heterogeneity of the derived force constants in HANM needs to be further studied. Compared with the uniform force constant of ANM, the heterogeneous force constants are more physically meaningful for accounting for the heterogeneous interactions of residues in proteins. Actually, the heterogeneity in proteins has been explicitly revealed in the previous studies. For instance, Ming and Wall⁷⁶ used a backbone-enhanced ENM to produce the correct bimodal distribution of the density-of-states from an all-atom model. Moritsugu et al.,⁵¹ Lyman et al.,⁵³ and Orellana et al.⁷⁷ parametrized and refined heterogeneous ENMs by extracting the information from MD trajectories. Yang et al.⁶³ scaled the interaction elements of the Hessian matrix by the inverse square of the pairwise distances. In HANM, the heterogeneous force constants are naturally generated by fitting to experimental data and the magnitudes are basically inversely proportional to the pairwise fluctuations. Therefore, the force constants are intrinsically correlated with the B-factors of C α atoms. To demonstrate this view, the correlation coefficients of the sum of force constants of C α atoms and their corresponding B-factors in the proteins examples were calculated and listed in Table 5. It is apparent that all the proteins have high negative correlation coefficients with the absolute values larger than 0.5, and the average correlation coefficient is -0.74 . This indicates a high anticorrelation between force constants and B-factors. In biology, B-factors are usually regarded as the indicator of structure flexibility,⁷⁸ thus the force constants obtained here also show an anticorrelation

Table 4. Mode Overlaps between the NMR Principal Components (PCs) and the Accumulative ANM/HANM Modes for the β -Neurotoxin^a

| | NMR PCs described by ANM modes | | | NMR PCs described by HANM modes | | |
|----------|--------------------------------|----------|----------|---------------------------------|----------|----------|
| | PC1 | PC2 | PC3 | PC1 | PC2 | PC3 |
| 3 modes | 0.61255 | 0.420617 | 0.248221 | 0.82626 | 0.460111 | 0.449248 |
| 6 modes | 0.636389 | 0.757205 | 0.560815 | 0.849601 | 0.782422 | 0.545887 |
| 20 modes | 0.788109 | 0.855851 | 0.681719 | 0.897546 | 0.869179 | 0.813346 |

^aA larger value corresponds to a greater overlap.**Table 5.** Calculated Correlation Coefficients between the Sum of the Force Constants of the Bonds Connected to a $C\alpha$ Atom and the $C\alpha$ Atom's B-factor^a

| PDB | CC |
|------|--------|
| 1A6M | −0.907 |
| 1CC8 | −0.819 |
| 1CRN | −0.608 |
| 1GNU | −0.796 |
| 1HG7 | −0.732 |
| 1HHP | −0.658 |
| 1IQZ | −0.783 |
| 1TU9 | −0.780 |
| 2LZM | −0.539 |
| SPTI | −0.781 |

^aThe averaged correlation coefficient over ten proteins is −0.740.

with structure flexibility. During the fitting process, the force constants of the springs involving $C\alpha$ atoms from well-ordered structures are generally tuned to be larger, while those involving $C\alpha$ atoms from somewhat flexible loop regions are typically tuned to be smaller. Lezon et al.⁷⁹ proposed that the magnitudes of force constants depend on the factors such as contact order, secondary structure, and packing density. Here, we use crambin as an example to characterize the diversity of the generated forces constants. Figure 6a shows the secondary structure of crambin with two short β -sheets at the residue indices 1–4 and 32–35, two α -helices at the residues 6–19 and 23–30, and a long flexible loop for 36–46, which are colored according to the sum of all force constants of the bonds connected to a $C\alpha$ atom. Figure 6b shows the distribution of the amplitudes of the force constants of HANM for crambin with respect to the two-dimensional coordinates of residue index. Based on GNM,⁸⁰ it is known that the fluctuations of atoms are closely associated with the local packing density, namely, the number of contacts of harmonic bonds. For small crambin, the local packing density is highly heterogeneous inside, depending on the environment of the residues surrounded. For instance, the first β sheet with the residues 1–4 forms many contacts with the surrounded residues within the cutoff distance 15 Å, except for the residues 18–20 that are far away from them, as shown in Figure 6b. This situation leads to the weaker heterogeneous force constants between residues 1–4 and the surrounded residues such as the residues 9–14 and 25–35, indicated by the light blue color at the bottom of Figure 6b. By contrast, there is a yellow region with the stronger force constants in the middle of the plot, roughly in the range of the indices 13–28 as seen from the abscissa, whose stiffness can be explained from the two aspects below. First, the sequence of 13–28 corresponds to the two segments of α -helices 6–19 and 23–30 in the secondary structure, which have a compact structure indicated by the B-factors. Second, the two helices have a low packing density since they do not form

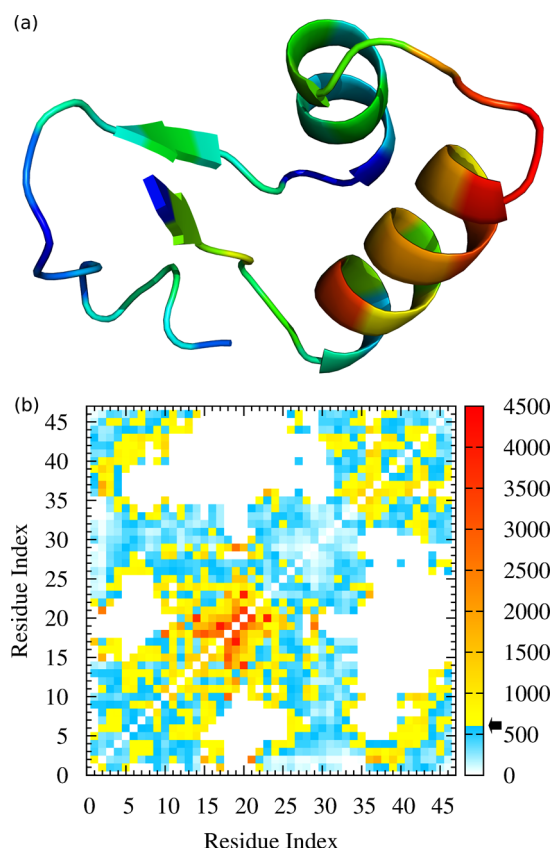


Figure 6. (a) Secondary structure of $C\alpha$ atoms of crambin. Each $C\alpha$ atom in the secondary structure is colored according to the sum of all force constants of the bonds connected to it. The warmer colors of $C\alpha$ atoms indicate that they form stiffer harmonic bonds with the surrounding $C\alpha$ atoms. (b) The distribution of the amplitudes of the heterogeneous force constants (in $\text{kJ}/(\text{mol}\cdot\text{nm}^2)$) calculated from the HANM of crambin with the cutoff distance 15 Å with respect to the two-dimensional coordinates of residue index. The line indicated by the black arrow is set as the value $600 \text{ kJ}/(\text{mol}\cdot\text{nm}^2)$ of the uniform force constant of ANM used for crambin. Above this line, the force constants increase with the color changing from yellow to red. Below this line, the force constants become weaker denoted by the color turning from cyan to white. The white color means a zero force constant, also indicating no existing effective bonded interactions beyond the cutoff value.

contacts with the residues 35–46 within 15 Å, indicated by the larger blank basin in the lower triangular region of Figure 6b. Thus, the segment of the residues 13–28 shows a warmer color than other parts in Figure 6a, especially for the loop at the residues 20–22. Similar to the first β sheet 1–4, the second β sheet with the residues 32–35 also forms a considerable number of contacts with the surrounding residues with the weaker force constants, while the loop with the indices 36–46 only has contacts with the β -sheet 1–4 and themselves. Even

though the loop is more flexible indicated by the B-factors, the less contacts lead to a mixed distribution of weak and strong force constants in the region of the residues 36–46. It can be inferred from the distribution in Figure 6b that for correctly reproducing the experimental B-factors, the resultant force constants not only depend on the heterogeneous packing density determined by the topology but also are associated with the heterogeneous interactions between residues, which is consistent with the conclusions drawn by Lezon et al.⁷⁹ Consequently, an ENM with a uniform constant cannot reproduce the accurate experimental B-factors.

3.5. Benchmark. To study the efficiency of the PFM method, we tested a number of proteins of up to 500 residues, using the very strict criterion that the change of B_i^{cal} compared with previous iteration is smaller than 0.5% for each residue. The CPU time (On a 2.00 GHz Intel Xeon E7-8850 processor running Red Hat Linux) consumed is listed in Table 6. For

Table 6. Benchmark of the PFM Method for the Proteins of Different Residue Numbers

| PDB ID | residue no. | iteration no. | total time | time per iteration |
|--------|-------------|---------------|----------------|--------------------|
| 1CRN | 46 | 22 | 1' 05" | 2.95×10^0 |
| 1HHP | 99 | 35 | 19' 49" | 3.40×10^1 |
| 1A6M | 151 | 16 | 34' 27" | 1.29×10^2 |
| 1IQQ | 200 | 23 | 1h 51' 01" | 2.90×10^2 |
| 1CPO | 298 | 19 | 4h 55' 34" | 9.33×10^2 |
| 1A8H | 500 | 30 | 1d 14h 29' 43" | 4.62×10^3 |

small proteins of less than 100 residues, the algorithm converges within 20 min. Usually, several hours are needed to obtain a converged result for medium size proteins of 100 to 300 residues. For relatively large proteins, such as methionyl-tRNA synthetase (PDB ID: 1A8H)⁸¹ of 500 residues, around one and a half days are needed before our algorithm converges. However, this should be an acceptable time scale, especially when comparing with that of protein MD simulations. It is also notable that the long time needed for large proteins is partially because of the very rigorous convergence criterion. If the criterion is changed to be that CC is larger than 0.90, which is already a better result than most conventional ENMs, only 5 iterations or 6 hours are required (Table S5, Supporting Information) for the 500-residue protein. Thus, our method is reasonably efficient when applying to at least medium size proteins less than 500 residues. Although it is somewhat time-consuming for large systems to achieve a converged result, a feasible solution may be to construct a database storing the spring constants for indexing, similar to iGNM.²¹

4. CONCLUSION

In this paper, we propose a novel PFM method to construct HANM starting from ANM. HANM uses harmonic potentials with heterogeneous force constants to describe the effective bond interactions. The test upon crambin shows that the HANMs are robust to yield accurate B-factor results with different cutoff values. For the four studied proteins crambin, trypsin inhibitor, HIV-1 protease, and lysozyme, all the calculated RMSDs are below 0.1 Å² and the CCs are 0.99, outperforming any ENM proposed before with physically meaningful force constants. For the studied molecular systems, it was found that the HANM modes have a higher similarity to the modes of the all-atom model and those derived from the NMR ensemble structures, compared with the ANM counter-

parts. It is proposed that HANM may be able to provide a more reasonable description of protein dynamics.

The current PFM method is computational efficient even for relatively large systems, though it is more expensive compared with ANM. More studies of protein motions on various biomacromolecules using the PFM method are also in progress. For instance, it is of particular interest to implement the approach on the biological systems containing nucleic acids as the previous ENM results are not satisfactory.⁴⁶

■ ASSOCIATED CONTENT

Supporting Information

Comparison of the modes 7–11 of HANM and ANM for crambin, trypsin inhibitor, HIV-1 protease and lysozyme (Table S1). Another comparison of the modes 7–11 of the all-atom model and HANM as well as ANM for HIV-1 protease (Table S2). Table S3 and S4 show the overlaps between the NMR principal components and the modes from HANM/ANM. The comparison of the experimental B-factors and the calculated results of HANM using the PFM method for six additional proteins with the PDB IDs: 1HG7, 1CC8, 1IQZ, 1GNU, 1TU9, and 1A6M is shown in Figure S1. A file containing all HANM force constants for the ten X-ray crystal protein structures in this paper and the source code of the method is also provided. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: LYL@ntu.edu.sg.

Author Contributions

†These two authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by a start-up grant and a college of science collaborative award, both provided by Nanyang Technological University. The computational resource partially came from an IDA cloud computing grant in Singapore. Xia Fei and Tong Dudu contributed equally in this work.

■ REFERENCES

- (1) Brooks, B. R.; Brucoleri, R. E.; Olafson, D. J.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (2) Jorgensen, W. L.; TiradoRives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- (3) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205–1218.
- (4) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (5) MacKerel Jr., A. D.; Brooks III, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*; John Wiley & Sons: Chichester, 1998; Vol. 1; pp 271.
- (6) Kidera, A.; Go, N. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3718–3722.
- (7) Kidera, A.; Inaka, K.; Matsushima, M. *J. Mol. Biol.* **1992**, *225*, 477–486.
- (8) Hayward, S.; Kitao, A.; Go, N. *Protein Sci.* **1994**, *3*, 936–943.
- (9) Hayward, S.; Kitao, A.; Go, N. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 177–186.
- (10) Ma, J.; Karplus, M. *J. Mol. Biol.* **1997**, *274*, 114–131.

- (11) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 1–7.
- (12) Tama, F.; Sanejouand, Y. H. *Protein Eng.* **2001**, *14*, 1–16.
- (13) Cui, Q.; Li, G.; Ma, J.; Karplus, M. J. *Mol. Biol.* **2004**, *340*, 345–372.
- (14) Petrone, P.; Pande, V. S. *Biophys. J.* **2006**, *90*, 1583–1593.
- (15) Reith, D.; Putz, M.; Muller-Plathe, F. J. *Comput. Chem.* **2003**, *24*, 1624–1636.
- (16) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908.
- (17) Halioglu, T.; Bahar, I.; Erman, B. *Phys. Rev. Lett.* **1997**, *79*, 3090–3094.
- (18) Bahar, I.; Atilgan, A. R.; Erman, B. *Fold. Des.* **1997**, *2*, 173–181.
- (19) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505–515.
- (20) Yang, L.; Song, G.; Jernigan, R. L. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 164–175.
- (21) Yang, L.-W.; Liu, X.; Jursa, C. J.; Holliman, M.; Rader, A. J.; Karimi, H. A.; Bahar, I. *Bioinformatics* **2005**, *21*, 2978–2987.
- (22) Eyal, E.; Yang, L.-W.; Bahar, I. *Bioinformatics* **2006**, *22*, 2619–2627.
- (23) Kim, M. K.; Jernigan, R. L.; Chirikjian, G. S. *Biophys. J.* **2002**, *83*, 1620–1630.
- (24) Maragakis, P.; Karplus, M. J. *Mol. Biol.* **2005**, *352*, 807–822.
- (25) Chu, J.-W.; Voth, G. A. *Biophys. J.* **2007**, *93*, 3860–3871.
- (26) Doruker, P.; Jernigan, R. L.; Bahar, I. J. *Comput. Chem.* **2002**, *23*, 119–127.
- (27) Kurkcuglu, O.; Jernigan, R. L.; Doruker, P. *Polymer* **2004**, *45*, 649–657.
- (28) Kurkcuglu, O.; Turgut, O. T.; Cansu, S.; Jernigan, R. L.; Doruker, P. *Biophys. J.* **2009**, *97*, 1178–1187.
- (29) Case, D. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285–290.
- (30) Ma, J. *Structure* **2005**, *13*, 373–380.
- (31) Bahar, I.; Rader, A. J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.
- (32) Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. *Chem. Rev.* **2010**, *110*, 1463–1497.
- (33) Dykeman, E. C.; Sankey, O. F. J. *Phys.: Condens. Matter* **2010**, *22*, 423202.
- (34) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.
- (35) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.
- (36) Schomaker, V.; Trueblood, K. N. *Acta Crystallogr. B* **1968**, *24*, 63–76.
- (37) Sternberg, M. J. E.; Grace, D. E. P.; Phillips, D. C. J. *Mol. Biol.* **1979**, *130*, 231–253.
- (38) Kuriyan, J.; Weis, W. I. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2733–2737.
- (39) Harata, K.; Abe, Y.; Muraki, M. J. *Mol. Biol.* **1999**, *287*, 347–358.
- (40) Halle, B. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1274–1279.
- (41) Kundu, S.; Melton, J. S.; Sorensen, D. C.; Phillips, G. N. *Biophys. J.* **2002**, *83*, 723–732.
- (42) Song, G.; Jernigan, R. L. J. *Mol. Biol.* **2007**, *369*, 880–893.
- (43) Soheilifard, R.; Makarov, D.; Rodin, G. J. *Phys. Biol.* **2008**, *5*, 026008.
- (44) Hinsen, K. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 417–429.
- (45) Hinsen, K.; Kneller, G. R. J. *Chem. Phys.* **1999**, *111*, 10766–10769.
- (46) Van Wynsberghe, A. W.; Cui, Q. *Biophys. J.* **2005**, *89*, 2939–2949.
- (47) Erman, B. *Biophys. J.* **2006**, *91*, 3589–3599.
- (48) Kondrashov, D. A.; Cui, Q.; Phillips, G. N. *Biophys. J.* **2006**, *91*, 2760–2767.
- (49) Sen, T. Z.; Feng, Y. P.; Garcia, J. V.; Kloczkowski, A.; Jernigan, R. L. J. *Chem. Theory Comput.* **2006**, *2*, 696–704.
- (50) Lu, M.; Ma, J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15358–15363.
- (51) Zheng, W. *Biophys. J.* **2008**, *94*, 3853–3857.
- (52) Riccardi, D.; Cui, Q.; Phillips, G. N. *Biophys. J.* **2009**, *96*, 464–475.
- (53) Yang, L.; Song, G.; Jernigan, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12347–12352.
- (54) Zheng, W. *Biophys. J.* **2010**, *98*, 3025–3034.
- (55) Hafner, J.; Zheng, W. J. *Chem. Phys.* **2010**, *132*, 014111.
- (56) Leioatts, N.; Romo, T. D.; Grossfield, A. J. *Chem. Theory Comput.* **2012**, *8*, 2424–2434.
- (57) Xia, F.; Lu, L. J. *Chem. Theory Comput.* **2012**, *8*, 4797–4806.
- (58) Chu, J.-W.; Izveko, S.; Voth, G. A. *Mol. Simul.* **2006**, *32*, 211–218.
- (59) Chu, J.-W.; Voth, G. A. *Biophys. J.* **2006**, *90*, 1572–1582.
- (60) Moritsugu, K.; Smith, J. C. *Biophys. J.* **2007**, *93*, 3460–3469.
- (61) Lyman, E.; Pfaendtner, J.; Voth, G. A. *Biophys. J.* **2008**, *95*, 4183–4192.
- (62) Brooks, B. R.; Janežič, D.; Karplus, M. J. *Comput. Chem.* **1995**, *16*, 1522–1542.
- (63) Yang, L.; Song, G.; Carriquiry, A.; Jernigan, R. L. *Structure* **2008**, *16*, 321–330.
- (64) Li, G.; Cui, Q. *Biophys. J.* **2002**, *83*, 2457–2474.
- (65) Teeter, M. M. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6014–6018.
- (66) Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. J. *Mol. Biol.* **1984**, *180*, 301–329.
- (67) Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J. *Biochimie* **1991**, *73*, 1391–1396.
- (68) Weaver, L. H.; Matthews, B. W. J. *Mol. Biol.* **1987**, *193*, 189–199.
- (69) McCammon, J. A.; Gelin, B.; Karplus, M.; Wolynes, P. G. *Nature* **1976**, *262*, 325–326.
- (70) Levitt, M.; Sander, C.; Stern, P. S. J. *Mol. Biol.* **1985**, *141*, 423–447.
- (71) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4995–4999.
- (72) Mchaourab, H. S.; Oh, K. J.; Fang, C. J.; Hubbell, W. L. *Biochemistry* **1997**, *36*, 307–316.
- (73) Jablonsky, M. J.; Jackson, P. L.; Trent, J. O.; Watt, D. D.; Krishna, N. R. *Biochem. Biophys. Res. Commun.* **1999**, *254*, 406–412.
- (74) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Humber, T.; Krüger, P.; van Gunsteren, W. F. J. *Phys. Chem. A* **1999**, *103*, 3596–3607.
- (75) Pearson, K. *Philos. Mag.* **1901**, *2*, 572.
- (76) Ming, D.; Wall, M. E. *Phys. Rev. Lett.* **2005**, *95*, 198103.
- (77) Orellana, L.; Rueda, M.; Ferrer-Costa, C.; Lopez-Blanco, J. R.; Chacón, P.; Orozco, M. J. *Chem. Theory Comput.* **2010**, *6*, 2910–2923.
- (78) Drenth, J. In *Principles of Protein X-Ray Crystallography*; Springer: New York, 2007; Chapter 4, pp 18–21.
- (79) Lezon, T. R.; Bahar, I. *PLoS Comput. Biol.* **2010**, *6*, e1000816.
- (80) Sen, T. Z.; Jernigan, R. L. *Optimizing the Parameters of the Gaussian Network Model for ATP-Binding Proteins*, in *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*; Bahar, I., Cui, Q., Eds.; Chapman and Hall/CRC: Boca Raton, 2006; pp 171–186.
- (81) Sugiura, I.; Nureki, O.; Ugaji-Yoshikawa, Y.; Kuwabara, S.; Shimada, A.; Tateno, M.; Lorber, B.; Giege, R.; Moras, D.; Yokoyama, S.; Konno, M. *Structure* **2000**, *8*, 197–208.