

Equally Weighted Multiscale Elastic Network Model and Its Comparison with Traditional and Parameter-Free Models

Weikang Gong, Yang Liu, Yanpeng Zhao, Shihao Wang, Zhongjie Han, and Chunhua Li*



Cite This: *J. Chem. Inf. Model.* 2021, 61, 921–937



Read Online

ACCESS |



Metrics & More

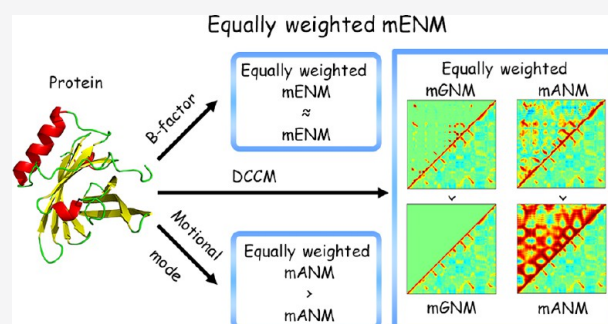


Article Recommendations



Supporting Information

ABSTRACT: Dynamical properties of proteins play an essential role in their function exertion. The elastic network model (ENM) is an effective and efficient tool in characterizing the intrinsic dynamical properties encoded in biomacromolecule structures. The Gaussian network model (GNM) and anisotropic network model (ANM) are the two often-used ENM models. Here, we introduce an equally weighted multiscale ENM (equally weighted mENM) based on the original mENM (denoted as mENM), in which fitting weights of Kirchhoff/Hessian matrixes in mENM are removed since they neglect the details of pairwise interactions. Then, we perform its comparison with the mENM, traditional ENM, and parameter-free ENM (pfENM) in reproducing dynamical properties for the six representative proteins whose molecular dynamics (MD) trajectories are available in <http://mmb.pcb.ub.es/MoDEL/>. In the results, for B-factor prediction, mENM performs best, while the equally weighted mENM performs also well, better than the traditional ENM and pfENM models. As to the dynamical cross-correlation map calculation, mENM performs worst, while the results produced from the equally weighted mENM and pfENM models are close to those from MD trajectories with the latter a little better than the former. Furthermore, encouragingly, the equally weighted mANM displays the best performance in capturing the functional motional modes, followed by pfANM and traditional ANM models, while the mANM fails in all the cases. This work is helpful for strengthening the understanding of the elastic network model and provides a valuable guide for researchers to utilize the model to explore protein dynamics.



INTRODUCTION

Protein structural dynamics is intimately related to their functions, which is reflected in many biological processes such as protein–ligand interactions,¹ signal transduction,² and assembly of macromolecular machines and allosteric regulation.^{3–5} Thus, obtaining accurate protein dynamical characteristics is critical for understanding and deducing their functions.

Protein dynamics and conformational changes can be calculated from their structure ensembles which can be obtained from experimental methods including X-ray crystallography,⁶ NMR spectroscopy,⁷ and cryo-electron microscopy (Cryo-EM)⁸ under different conditions. However, these experimental methods are both expensive and time consuming. Theoretically, molecular dynamics (MD) simulation provides a useful tool at the atomic level to analyze the mechanical, structural, and thermodynamic properties of biomolecules.^{9,10} However, its application requires enormous computer resources and does not always fully sample the entire conformational space accessible to a protein. Therefore, some coarse-grained methods have been developed,^{11–13} and among them, the elastic network model (ENM) is a harmonic potential-based and cost-effective computational method.¹¹ The ENM has achieved great success in predicting the large-

amplitude collective motion for proteins^{14–18} and even for RNAs.¹⁹ The Gaussian network model (GNM)^{15,16} and anisotropic network model (ANM)²⁰ are the two often-used ENM models.

In the traditional ENM, a protein structure is modeled as an elastic network of C_α atoms in which the residue pairs within a given cutoff distance are considered to have interactions and are connected by a set of Hookean springs with a uniform force constant. Generally, the low-frequency motion modes obtained by ENM represent the large-scale collective motions relevant to molecular functions. In the past few years, variant methods of ENM have been proposed. Considering residue specificity, Hamacher and McCammon developed the extended ANM (eANM) which uses the Miyazawa–Jernigan (MJ) contact potential to weight the spring force constant and achieves a good performance in reproducing the crystallo-

Received: October 9, 2020

Published: January 26, 2021



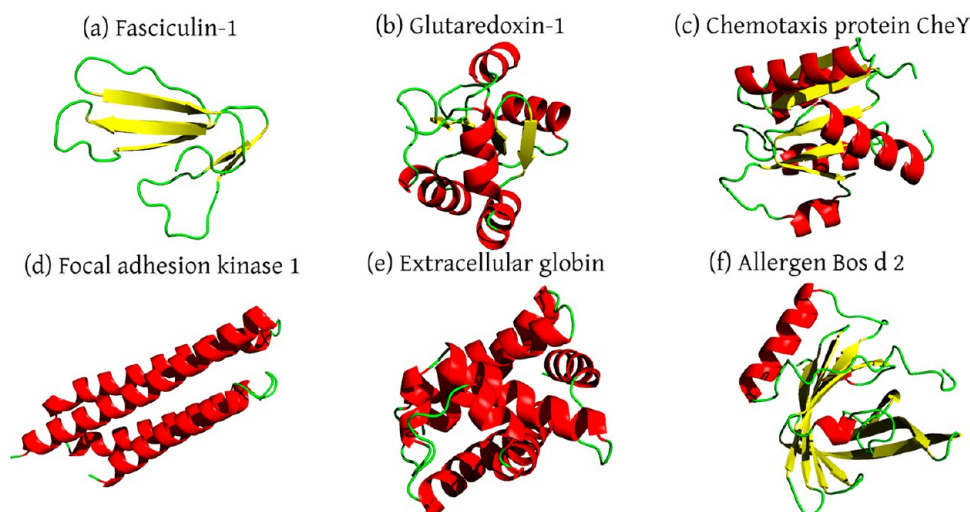


Figure 1. Six test proteins including Fasciculin-1 (a), Glutaredoxin-1 (b), Chemotaxis protein CheY (c), focal adhesion kinase 1 (d), extracellular globin (e), and allergen Bos d 2 (f) with PDB IDs being 1FAS, 1KTE, 1CHN, 1K40, 1ASH, and 1BJ7, respectively.

graphic temperature factors (B-factor).²¹ Later, with the long-range effect of interactions taken into account, Yang et al. proposed a parameter-free ENM (pfENM), in which the spring force constant is proportional to different inverse powers of the inter-residue distance.²² On the basis of pfENM method, Zhang et al. introduced the relative solvent accessibility and developed RpfGNM method.²³ As we know, in biomolecules, different kinds of interactions, such as covalent, van der Waals, hydrophobic, and electrostatic interactions, have their own characteristic action ranges. Thus, recently, considering the multiscale features of the interactions, Xia et al. introduced a multiscale ENM (mENM) which incorporates different scales of interactions into the spring force constant. The mENM can successfully capture the multiscale properties of proteins and significantly improves the accuracy of B-factor calculation.²⁴ A previous study has shown that the strong power of some models in predicting B-factors does not mean that they also have a tough power in reproducing other dynamical properties.²⁵ A systematic comparison needs to be done for variant ENM models in multiple dynamical property calculations.

Generally, B-factors, dynamical cross-correlation maps (DCCMs), and motional modes are the widely used parameters to describe protein dynamics. The B-factors of atoms contain important information about their thermal motions in a macromolecule, reflecting the local structural flexibility. The DCCM shows the strength of the collective motions between residues, which has been verified to be useful in understanding protein large domain movements.^{26,27} As for motional modes, generally low frequency modes represent the global collective motions related to protein functions, which provide a substantial understanding of the protein-folding process²⁸ and allosteric transition.²⁹

In this paper, we introduce an equally weighted mENM model based on the original one and perform its comparison with the widely used traditional ENM and pfENM models in capturing protein dynamical properties including B-factors, DCCMs, and functional motional modes. For the latter two properties, the results from ENM models are compared with those from MD ensembles.

MATERIALS AND METHODS

Test Cases. In order to evaluate the performance of ENM models, we selected six representative proteins (Figure 1) as test cases whose MD trajectories are available in the Molecular Dynamics Extended Library (MoDEL).³⁰ The six proteins have different compositions of secondary structures and varied residue numbers ranging from 61 to 150 (Table S1). These proteins are Fasciculin-1 (PDB ID: 1FAS),³¹ mainly containing beta-sheet secondary structures, focal adhesion kinase 1 (1K40),³² and extracellular globin (1ASH)³³ primarily composed of α helix secondary structures and the remaining Glutaredoxin-1 (1KTE),³⁴ Chemotaxis protein CheY (1CHN),³⁵ and Allergen Bos d 2 (1BJ7)³⁶ made up of α helix, beta-sheet, and loop secondary structures.

Traditional ENM. In traditional ENM, a protein structure is represented as an elastic network of some atoms (here C_α) where the node pairs within a cutoff distance (r_c) are connected by springs with a uniform force constant.³⁷

In traditional GNM, the total internal potential energy of the network of N nodes can be written as

$$V_{\text{GNM}} = \frac{1}{2} \gamma [\Delta \mathbf{R}^T (\mathbf{\Gamma} \otimes \mathbf{E}) \Delta \mathbf{R}] \quad (1)$$

where γ is the spring force constant, the column vector $\Delta \mathbf{R}$ represents the fluctuation of the N nodes, $\Delta \mathbf{R}^T$ is the transpose matrix of $\Delta \mathbf{R}$, \mathbf{E} is the unitary matrix, $\mathbf{\Gamma}$ is the matrix direct product, and $\mathbf{\Gamma}$ is the $N \times N$ symmetric Kirchhoff matrix, the elements of which are described as

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0, & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{j,j \neq i}^N \Gamma_{ij}, & \text{if } i = j \end{cases} \quad (2)$$

where r_c is the cutoff distance, and R_{ij} is the distance between the i th and j th nodes. The B-factors, dynamical cross-correlation map (DCCM), and motional modes can be calculated from the pseudoinverse of the Kirchhoff matrix, as shown in the following section of the calculation of B-factors and dynamical cross-correlations based on ENMs.

The GNM model can provide the amplitudes of residue fluctuations but no information about their directions, and this information is considered in ANM. In traditional ANM, the total potential energy of the network can be written as

$$V_{\text{ANM}} = \frac{1}{2} \gamma \sum_{i,j}^N (R_{ij} - R_{ij}^0)^2 \quad (3)$$

where R_{ij} and R_{ij}^0 refer to the instantaneous and equilibrium distances between node i and j , respectively. The protein dynamical properties are determined by a Hessian matrix \mathbf{H} whose element is a submatrix with size of 3×3 . The submatrix \mathbf{h}_{ij} is calculated as the matrix of second-order derivatives of the potential with respect to the Cartesian coordinates of the nodes.²⁰ When $i \neq j$, the corresponding \mathbf{h}_{ij} is

$$\mathbf{h}_{ij} = \begin{pmatrix} \frac{\partial V}{\partial X_i \partial X_j} & \frac{\partial V}{\partial X_i \partial Y_j} & \frac{\partial V}{\partial X_i \partial Z_j} \\ \frac{\partial V}{\partial Y_i \partial X_j} & \frac{\partial V}{\partial Y_i \partial Y_j} & \frac{\partial V}{\partial Y_i \partial Z_j} \\ \frac{\partial V}{\partial Z_i \partial X_j} & \frac{\partial V}{\partial Z_i \partial Y_j} & \frac{\partial V}{\partial Z_i \partial Z_j} \end{pmatrix} \quad (4)$$

When $i = j$, the submatrix is

$$\mathbf{h}_{ii} = - \sum_{i \neq j} \mathbf{h}_{ij} \quad (5)$$

The dynamical properties can be calculated from the pseudoinverse of the Hessian matrix, as shown in the following section of calculation of B-factors and dynamical cross-correlations based on ENMs.

Parameter-Free ENM (pfENM). Different from the traditional cutoff-based ENM, the parameter-free ENM (pfENM) adopts a distance-dependent spring constant set. In pfENM, some atoms (here also C_α) are selected as nodes. All the node pairs are considered to be interacting with each other with the strength being inversely proportional to their square distance. In pfGNM, the elements of the Kirchhoff matrix Γ^{pfGNM} are calculated as²²

$$\Gamma_{ij}^{\text{pfGNM}} = \begin{cases} -R_{ij}^{-2}, & \text{if } i \neq j \\ -\sum_{j,j \neq i}^N \Gamma_{ij}^{\text{pfGNM}}, & \text{if } i = j \end{cases} \quad (6)$$

where R_{ij} is the distance between the i th and j th nodes. In pfANM, each submatrix of the Hessian matrix is weighted by the inverse of the square distance between two nodes.²² When $i \neq j$, $\mathbf{h}_{ij}^{\text{pfANM}}$ is

$$\mathbf{h}_{ij}^{\text{pfANM}} = \mathbf{h}_{ij} R_{ij}^{-2} \quad (7)$$

where \mathbf{h}_{ij} is of the form of eq 4. When $i = j$, the submatrix is

$$\mathbf{h}_{ii}^{\text{pfANM}} = - \sum_{i \neq j} \mathbf{h}_{ij}^{\text{pfANM}} \quad (8)$$

Multiscale ENM (mENM). In multiscale GNM (mGNM),²⁴ the multiscale interactions are considered. Different exponential decay kernel functions are used to represent interactions, which can be written as

$$\Phi(R_{ij}; \eta, \kappa) = e^{-\left(\frac{R_{ij}}{\eta}\right)^\kappa}, \quad \kappa > 0 \quad (9)$$

where R_{ij} is the distance between the i th and j th nodes, and the parameters η and κ control the decay extent. Multiscale interactions can be represented as the sum of different kernel functions, and the corresponding Kirchhoff matrix of the n th kernel function can be described as

$$[\Gamma_n^{\text{mGNM}}]_{ij} = \begin{cases} -\Phi_n(R_{ij}; \eta_n, \kappa_n), & \text{if } i \neq j \\ -\sum_{j,j \neq i}^N [\Gamma_n^{\text{mGNM}}]_{ij}, & \text{if } i = j \end{cases} \quad (10)$$

When n (here $n = 2$) kernel functions are considered, the multiscale Kirchhoff matrix can be given by

$$\Gamma^{\text{mGNM}} = \sum_n a_n \Gamma_n^{\text{mGNM}} \quad (11)$$

where Γ_n^{mGNM} is the Kirchhoff matrix corresponding to the n th kernel function, and the parameters a_n can be obtained by the least-square method (LSM)

$$\text{Min}_{a_n} \left\{ \sum_i \left| \sum_n a_n [\Gamma_n^{\text{mGNM}}]_{ii} - \frac{1}{B_i^e} \right|^2 \right\} \quad (12)$$

where B_i^e is the experimental B-factor of the i th node.

In mANM,²⁴ each submatrix of the Hessian matrix is also of the form of eq 4. When $i \neq j$, the submatrix corresponding to the n th kernel function can be described as

$$[\mathbf{h}_n^{\text{mANM}}]_{ij} = \Phi_n(R_{ij}; \eta_n, \kappa_n) \mathbf{h}_{ij} \quad (13)$$

When $i = j$, the submatrix is

$$[\mathbf{h}_n^{\text{mANM}}]_{ii} = - \sum_{i \neq j} [\mathbf{h}_n^{\text{mANM}}]_{ij} \quad (14)$$

When n (here $n = 2$) kernel functions are considered, the multiscale Hessian matrix can be given by

$$\mathbf{H}^{\text{mANM}} = \sum_n a_n \mathbf{H}_n^{\text{mANM}} \quad (15)$$

The weights a_n are determined by

$$\text{Min}_{a_n} \left\{ \sum_i \left| \sum_n a_n \sum_{j=1}^3 [\mathbf{H}_n^{\text{mANM}}]_{3(i-1)+j, 3(i-1)+j} - \frac{1}{B_i^e} \right|^2 \right\} \quad (16)$$

Equally Weighted mENM. We noticed that in mENM the least-square method (LSM) used to determine the weights of the Kirchhoff/Hessian matrixes only considers their diagonal elements which are equal to the negative sums of the off-diagonal elements. Thus, it is possible to have different values of off-diagonal elements and yet obtain the same values of diagonal elements, which implies that the details of the pairwise interactions (off-diagonal elements) are not taken into account in determination of the weights, thereby abolishing the “network” contribution in the ENM model. To avoid it, we try a simple way to let all the weights be equal to 1, and we call it the equally weighted mENM.

In equally weighted mGNM, when n (here $n = 2$) kernel functions are considered, the multiscale Kirchhoff matrix can be given by

$$\Gamma^{\text{equally weighted mGNM}} = \sum_n \Gamma_n^{\text{mGNM}} \quad (17)$$

where Γ_n^{mGNM} is the same as eq 10. In equally weighted mANM, the Hessian matrix can be given by

$$H^{\text{equally weighted mANM}} = \sum_n H_n^{\text{mANM}} \quad (18)$$

where H_n^{mANM} can be obtained by eqs 13 and 14.

Here, although the weights are equal to 1, for a residue pair, its long-range and short-range interactions represented by two kernel functions are added at some proportion. For the pair of residues that are far apart (close to each other), a relative small (or nearly zero (large) short-range interaction and a not very small (still large) long-range interaction are added together.

Calculation of B-Factors and Dynamical Cross-Correlations Based on ENMs. In GNMs, the mean-square fluctuation of each residue and dynamical cross-correlation between two residues are in proportion to the diagonal and off-diagonal elements of the pseudoinverse of Kirchhoff matrix, respectively.

$$\langle \Delta R_i \cdot \Delta R_i \rangle = \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ii} \quad (19)$$

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ij} \quad (20)$$

where k_B is the Boltzmann constant, and T is the thermodynamic temperature. The dynamical cross-correlations of all the residue pairs can construct a map denoted as dynamical cross-correlation map (DCCM). The pseudoinverse of Kirchhoff matrix can be decomposed as

$$\Gamma^{-1} = \sum_{k=2}^N \lambda_k^{-1} \mu_k \cdot \mu_k^T \quad (21)$$

where λ_k and μ_k are the k th eigenvalue and eigenvector of Kirchhoff matrix, respectively.

In ANMs, the two corresponding dynamical properties can be written as

$$\langle \Delta R_i \cdot \Delta R_i \rangle = \frac{k_B T}{\gamma} (H_{3i-2,3i-2}^{-1} + H_{3i-1,3i-1}^{-1} + H_{3i,3i}^{-1}) \quad (22)$$

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \frac{k_B T}{\gamma} (H_{3i-2,3j-2}^{-1} + H_{3i-1,3j-1}^{-1} + H_{3i,3j}^{-1}) \quad (23)$$

The pseudoinverse of the Hessian matrix can be decomposed as

$$H^{-1} = \sum_{k=7}^{3N} \lambda_k^{-1} \mu_k \cdot \mu_k^T \quad (24)$$

where λ_k and μ_k are the k th eigenvalue and eigenvector of Hessian matrix, respectively. According to Debye–Waller theory, the theoretical B-factor of the i th node can be calculated with the expression

$$B_i^t = \frac{8\pi^2}{3} \langle \Delta R_i \cdot \Delta R_i \rangle \quad (25)$$

The dynamical cross-correlation is usually normalized as

$$C_{ij} = \frac{\langle \Delta R_i \cdot \Delta R_j \rangle}{[\langle (\Delta R_i)^2 \rangle \times \langle (\Delta R_j)^2 \rangle]^{1/2}} \quad (26)$$

The cross-correlation value varies from -1 to 1 . The positive values indicate that the residues move in the same direction, and the negative ones indicate that they move in the opposite direction. The higher the absolute value is, the more the two residues are correlated. The zero value means that the motions of residues are completely uncorrelated.

Calculation of DCCMs and Motional Principle Components Based on MD Ensembles. For the six representative test proteins whose MD trajectories are available in MoDEL, Table S1 gives the information on MD simulations. MoDEL is a database of protein trajectories obtained by atomistic MD simulations at near-physiological conditions. In order to get the DCCMs and motional principle components from MD trajectories, we collect the C_α atoms' conformations every 1 ps from the trajectories after a 5 ns run when all the systems reach an equilibrium.

The cross-correlation between two residues can be calculated with the formula

$$C_{ij} = \frac{\langle \Delta R_i(t) \cdot \Delta R_j(t) \rangle}{[\langle \Delta R_i(t)^2 \rangle \times \langle \Delta R_j(t)^2 \rangle]^{1/2}} \quad (27)$$

where $\Delta R_i(t) = R_i(t) - \langle R_i(t) \rangle$, $R_i(t)$ is the position vector of C_α atom of the i th residue at time t , and $\langle \rangle$ refers to a time average of the quantity within the brackets.

The motional principle components (PCs) are obtained by performing the principal component analysis (PCA) method on the collected C_α atoms' conformations. The principal components are sorted in a decreasing order of eigenvalues and referred to as PC1, PC2, PC3, and so on. The first few PCs capture a significant part of the ensemble variance, with PC1 giving the largest contribution, followed by PC2 and PC3 and so forth.

Comparison of Theoretical and Experimental B-Factors. The Pearson's correlation coefficient (PCC)²² is used to evaluate the correlation extent between theoretical and experimental B-factors, which is given by

$$B_{\text{PCC}} = \frac{\sum_{i=1}^N (B_i^t - \overline{B^t})(B_i^e - \overline{B^e})}{[\sum_{i=1}^N (B_i^t - \overline{B^t})^2 \sum_{i=1}^N (B_i^e - \overline{B^e})^2]^{1/2}} \quad (28)$$

where B_i^t and B_i^e are the theoretical and experimental B-factors for the i th node, respectively, and $\overline{B^t}$ and $\overline{B^e}$ are their mean values. A perfect correlation between them gives a value of 1 , whereas perfect anticorrelation gives -1 .

It should be pointed out that for traditional ENM and mENM models (except for pfENM due to no parameters in it), the optimized parameters are obtained to construct the corresponding ENM models through maximizing the PCC (B_{PCC}) value between the theoretical and experimental B-factors for each protein. During the process, for traditional GNM/ANM, the cutoff distance value systematically varies in the range of $[1 \text{ \AA}, 16 \text{ \AA}]/[1 \text{ \AA}, 30 \text{ \AA}]$ with a step size of 1 \AA , and for mENM, the parameter η/κ systematically varies in the range of $[1 \text{ \AA}, 50 \text{ \AA}]/[1, 10]$ with step size $1 \text{ \AA}/1$. Here, it is noted that in the optimization process, for a given set of η and κ parameters, there are still parameters a_n in eqs 11 and 15 to

Table 1. B_PCC Values between Experimental and Theoretical B-Factors Calculated by Four Kinds of ENM Models on Six Proteins^a

| PDB ID | Model | | | | | | | |
|--------|-----------------|-----------------|-------|-------|-------------|-------------|-----------------------|-----------------------|
| | traditional ENM | | pfENM | | mENM | | equally weighted mENM | |
| | traditional GNM | traditional ANM | pfGNM | pfANM | mGNM | mANM | equally weighted mGNM | equally weighted mANM |
| 1FAS | 0.74 | 0.69 | 0.63 | 0.39 | 0.74 | 0.88 | 0.73 | 0.78 |
| 1KTE | 0.65 | 0.64 | 0.66 | 0.63 | 0.67 | 0.72 | 0.70 | 0.64 |
| 1CHN | 0.66 | 0.69 | 0.72 | 0.71 | 0.75 | 0.81 | 0.74 | 0.73 |
| 1K40 | 0.79 | 0.70 | 0.72 | 0.58 | 0.82 | 0.80 | 0.81 | 0.78 |
| 1ASH | 0.72 | 0.56 | 0.65 | 0.56 | 0.76 | 0.79 | 0.75 | 0.63 |
| 1BJ7 | 0.69 | 0.73 | 0.63 | 0.66 | 0.79 | 0.78 | 0.73 | 0.73 |

^aTwo highest B_PCC values from GNMs and ANMs, respectively, for each protein are shown in bold.

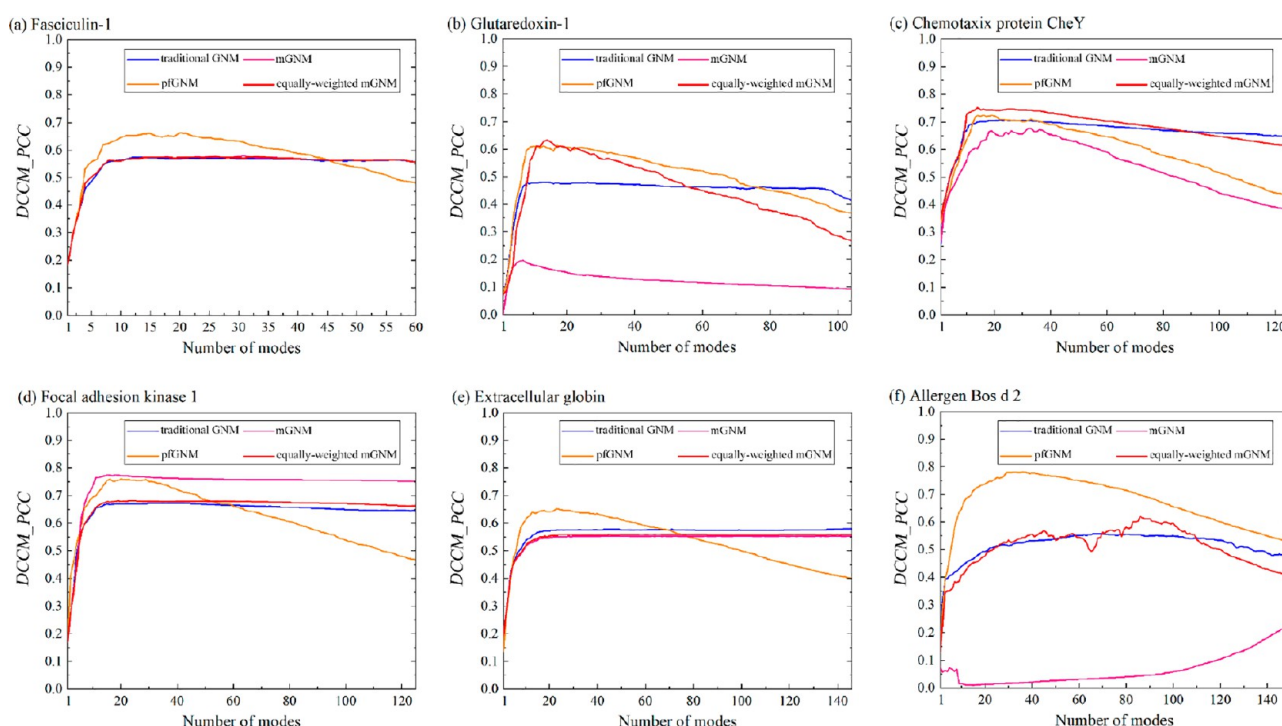


Figure 2. Changes of pearson correlation coefficients between DCCMs ($DCCM_PCC$) obtained from MD ensembles and four kinds of GNMs with the number of motional modes considered in GNMs for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e), and 1BJ7 (f).

be determined, for which the least-square method is used as shown in eqs 12 and 16. For clarity, the flowchart of the optimization is shown in Figure S1.

Comparison of DCCMs from ENMs and MD Ensembles. To detect the effectiveness of different ENM models in obtaining DCCMs, we compare the similarity between the DCCMs from ENMs and MD ensembles based on their Pearson's correlation coefficient,³⁸ which can be given by

$$DCCM_PCC = \frac{\sum_{i=1}^{N \times N} (X_i - \bar{X})(Y_i - \bar{Y})}{[\sum_{i=1}^{N \times N} (X_i - \bar{X})^2 \sum_{i=1}^{N \times N} (Y_i - \bar{Y})^2]^{1/2}} \quad (29)$$

where X and Y represent the DCCM's elements obtained from ENMs and MD ensembles, respectively, and \bar{X} and \bar{Y} denote the average values of X and Y , respectively.

Comparison of Motional Modes from ANMs and PCs from MD Ensembles. In order to explore the effectiveness of different ENMs in obtaining motional modes, we compare the motional modes from ANMs and PCs from MD ensembles. The 10 different performance metrics are used, which include

the maximum and cumulative overlaps between the first 20 modes from ANMs and each of PC1, PC2, and PC3, respectively, (denoted as O_1^{\max} , O_2^{\max} , O_3^{\max} and CO_1^{20} , CO_2^{20} , CO_3^{20}), and the root mean square inner products ($RMSIPs$) between the first 20 ANM modes and sets of the first 3, 6, 10, and 20 PCs (denoted as $RMSIP_3^{20}$, $RMSIP_6^{20}$, $RMSIP_{10}^{20}$, and $RMSIP_{20}^{20}$). The overlap between PC eigenvector p_i from MD ensembles and a given motional mode u_j from ANMs is calculated as

$$O_{ij} = \frac{|p_i \cdot u_j|}{\|p_i\| \|u_j\|} \quad (30)$$

where $\|p_i\|$ and $\|u_j\|$ refer to the lengths of p_i and u_j vectors, respectively. The maximum overlap between any of the first 20 modes and the PC eigenvector p_i is calculated as

$$O_i^{\max} = \max_{j=1 \text{ to } 20} O_{ij} \quad (31)$$

The cumulative overlap between the first 20 modes and PC eigenvector p_i is calculated as

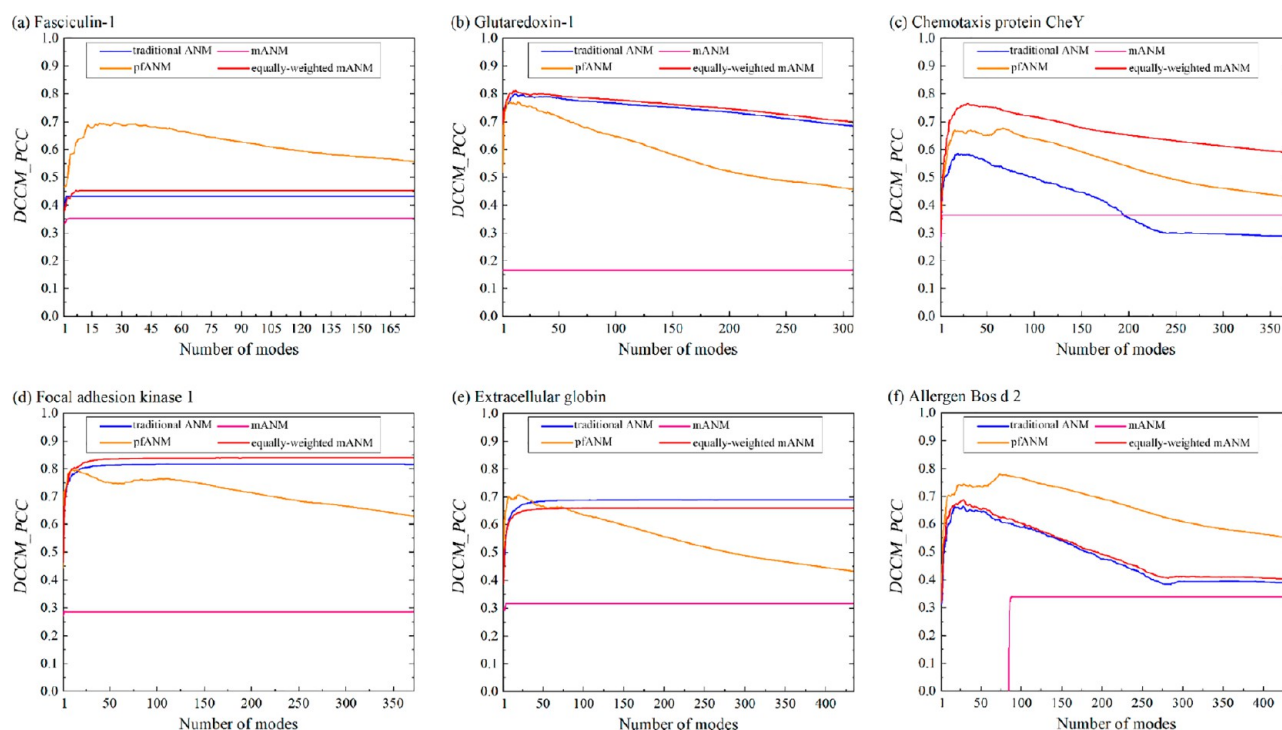


Figure 3. Changes of pearson correlation coefficients between DCCMs (*DCCM_PCC*) obtained from MD ensembles and four kinds of ANMs with the number of motional modes considered in ANMs for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e), and 1BJ7 (f).

$$CO_i^{20} = \sqrt{\sum_{j=1}^{20} O_{ij}^2} \quad (32)$$

The RMSIP between the first l PC eigenvectors and the set of first 20 modes is calculated as

$$RMSIP_l^{20} = \sqrt{\frac{1}{l} \sum_{i=1}^l \sum_{j=1}^{20} (p_i \cdot u_j)^2} \quad (33)$$

RESULTS

Theoretical B-Factor Calculation. B-factors reflect the local flexibility of molecules. We used the PCC (B_PCC) values between the theoretical and experimental B-factors to evaluate the performances of ENM models, with the results shown in Table 1. From Table 1, for GNMs, it is clear that mGNM performs best with only one case (1KTE) having a little worse B_PCC value than that from the equally weighted mGNM. The previous work has reported that mGNM is significantly better than traditional GNM in B-factor prediction.²⁴ For ANMs, the best method is still the mANM model which attains the highest values in all the cases. Thus, we can see that with the multiscale interactions considered, mENM achieves the best performance among the four ENM models in B-factor calculation. Additionally, it is noted that with the weights of multiscale Kirchhoff/Hessian matrixes being equal, interestingly, the equally weighted mGNM/mANM only has a little worse performance than mGNM/mANM with an average B_PCC value drop of 0.01/0.08 but still has a much better performance than the corresponding traditional ENM and pfENM models.

As for pfENM, besides the inverse second power model, we also performed the inverse 12th power model (Table S2). Generally, the latter performs worse than the former,

consistent with the previous result that the inverse second power model is relatively good in B-factor prediction.²² Thus, the inverse second power is adopted in pfENM model for the following calculations.

Comparing DCCMs from ENMs and MD Ensembles.

The DCCM can give the collective motion information and help understand the functional movements between protein domains. In order to get the similarity between the DCCMs obtained by ENMs and MD ensembles, we calculated them and then evaluated the similarity using their *DCCM_PCC* value (see Materials and Methods). Figures 2 and 3 show the changes of *DCCM_PCC* values with the number of modes considered in GNMs and ANMs, respectively. From Figure 2, for traditional GNM, it can be seen that the curves tend to be steady after a rapid rise for all the proteins, which is consistent with the previous observation by Tekpinar and Yildirim on six proteins.³⁸ That means for traditional GNM the motional modes beyond a certain number of low frequency modes have almost no contributions to the cross-correlation calculation. In contrast, an approximately inverse parabolic shape is observed for pfGNM on all the proteins, which means that the motional modes beyond a certain number of low frequency modes can cause different levels of decrease in cross-correlation calculations. Interestingly, when setting a cutoff distance 7.0 Å to truncate interactions in pfGNM, the corresponding curves become steady after a rapid rise for all the proteins (Figure S2), which implies that the long-range interactions need to be considered specially and carefully in ENM models.³⁹ As for mGNM, its performance is unstable, like traditional GNMs on three proteins (1FAS, 1K40, and 1ASH), like pfGNM's on one protein (1CHN), and with failed results obtained on two proteins (1KTE and 1BJ7). Excitedly, the equally weighted mGNM displays a significant improvement on the three proteins 1KTE, 1BJ7, and 1CHN, especially on the former two where mGNM fails.

Table 2. Best DCCM_PCC Values between DCCMs from MD Ensembles and Four Kinds of ENMs for Six Proteins^a

| PDB ID | Model | | | | | | | |
|--------|-----------------|-----------------|-------------|-------------|-------------|------|-----------------------|-----------------------|
| | traditional ENM | | pfENM | | mENM | | equally weighted mENM | |
| | traditional GNM | traditional ANM | pfGNM | pfANM | mGNM | mANM | equally weighted mGNM | equally weighted mANM |
| 1FAS | 0.57 | 0.43 | 0.66 | 0.70 | 0.58 | 0.35 | 0.58 | 0.45 |
| 1KTE | 0.48 | 0.80 | 0.61 | 0.77 | 0.20 | 0.17 | 0.63 | 0.81 |
| 1CHN | 0.71 | 0.59 | 0.72 | 0.68 | 0.67 | 0.36 | 0.75 | 0.77 |
| 1K40 | 0.67 | 0.82 | 0.76 | 0.80 | 0.78 | 0.29 | 0.68 | 0.84 |
| 1ASH | 0.58 | 0.69 | 0.65 | 0.71 | 0.55 | 0.32 | 0.56 | 0.66 |
| 1BJ7 | 0.56 | 0.66 | 0.78 | 0.78 | 0.23 | 0.34 | 0.62 | 0.69 |

^aTwo highest values obtained from GNMs and ANMs, respectively, for each protein are shown in bold.

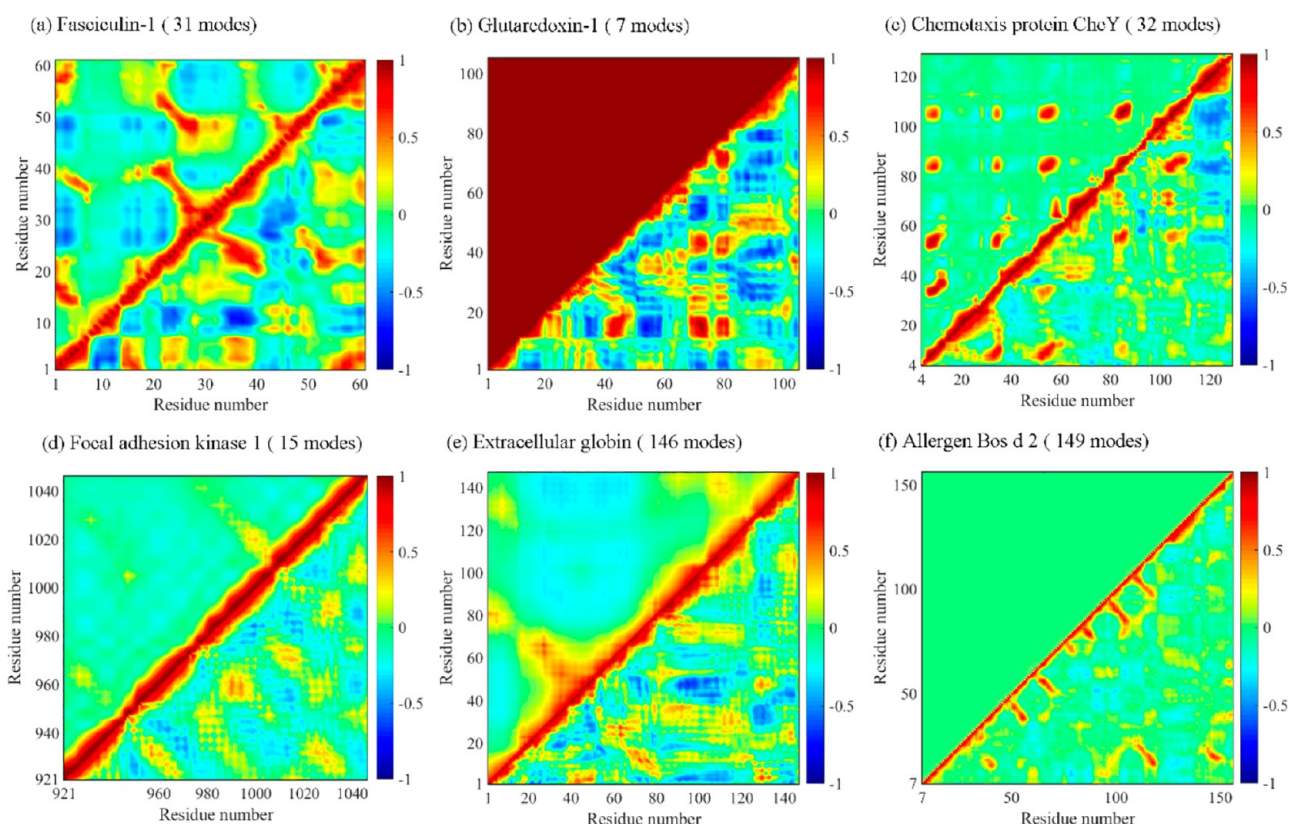


Figure 4. DCCMs obtained from MD ensembles (lower right triangle) and mGNM (upper left triangle) at the best DCCM_PCC value (corresponding number of motional modes given in parentheses) for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e), and 1BJ7 (f).

From Figure 3, the traditional ANM, compared with traditional GNM, gives the curves of different trends on two proteins 1CHN and 1BJ7, dropping first and then keeping steady after a rapid rise. The curve outlines from pfANM are similar to those from pfGNM, and on proteins 1FAS, 1KTE, 1K40 and 1ASH, pfANM performs better than pfGNM. As for mANM, the worst results are observed for all the proteins with the highest values less than 0.4. Interestingly, the equally weighted mANM achieves remarkable improvements for all the proteins. There are four proteins, 1KTE, 1CHN, 1K40, and 1ASH, where the equally weighted mANM performs best among all the GNM and ANM models.

Table 2 reports the best pearson correlation coefficient (DCCM_PCC) values between the DCCMs obtained from ENMs and MD ensembles, respectively, for all the six proteins. The corresponding maps are shown in Figures 4 and 5 for mGNM and equally weighted mGNM and Figure 6 and 7 for mANM and equally weighted mANM respectively (see Figures

S3–S4 for traditional GNM and pfGNM and Figures S5–S6 for traditional ANM and pfANM). From Table 2, when the correlation reaches the best value, for GNMs, the DCCMs from pfGNM are generally closest to those from MD ensembles. There are three cases which have the highest correlation values (0.66 for 1FAS, 0.65 for 1ASH, and 0.78 for 1BJ7), and the lowest value is 0.61 for 1KTE (whose highest value is 0.63 from the equally weighted mGNM). The pfGNM is closely followed by the equally weighted mGNM with two cases having the highest values (0.63 for 1KTE and 0.75 for 1CHN). As shown in Figure S4 and Figure 5, the similarities between the DCCMs from MD ensembles and pfGNM or equally weighted mGNM are more or less evident. As for mGNM, although there is one case having the highest value 0.78 for 1K40, four lowest values (0.20 for 1KTE, 0.67 for 1CHN, 0.55 for 1ASH, and 0.23 for 1BJ7) are observed in Figure 4, indicating its unsteady performance in DCCM calculation. For ANMs, they generally perform better than the

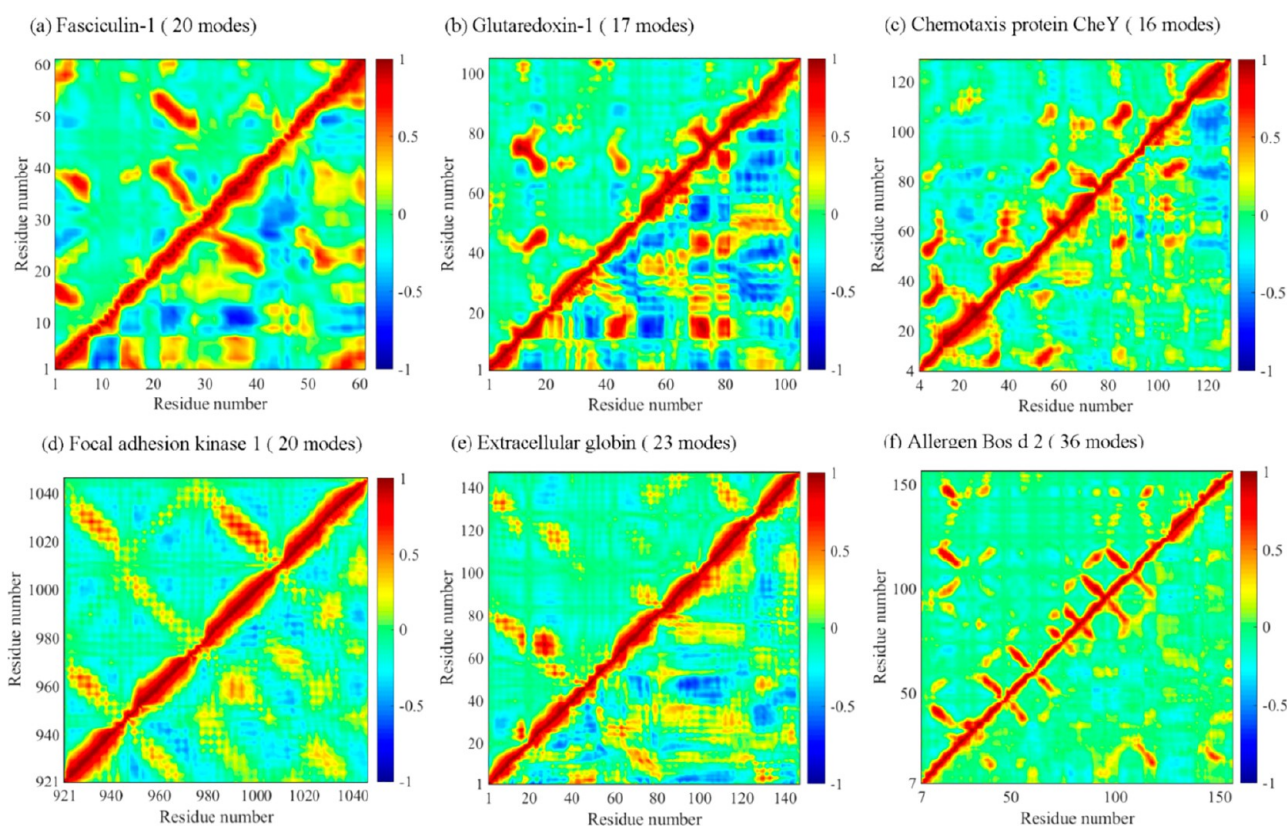


Figure 5. DCCMs obtained from MD ensembles (lower right triangle) and equally weighted mGNM (upper left triangle) at the best *DCCM_PCC* value (corresponding number of motional modes given in parentheses) for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e), and 1BJ7 (f).

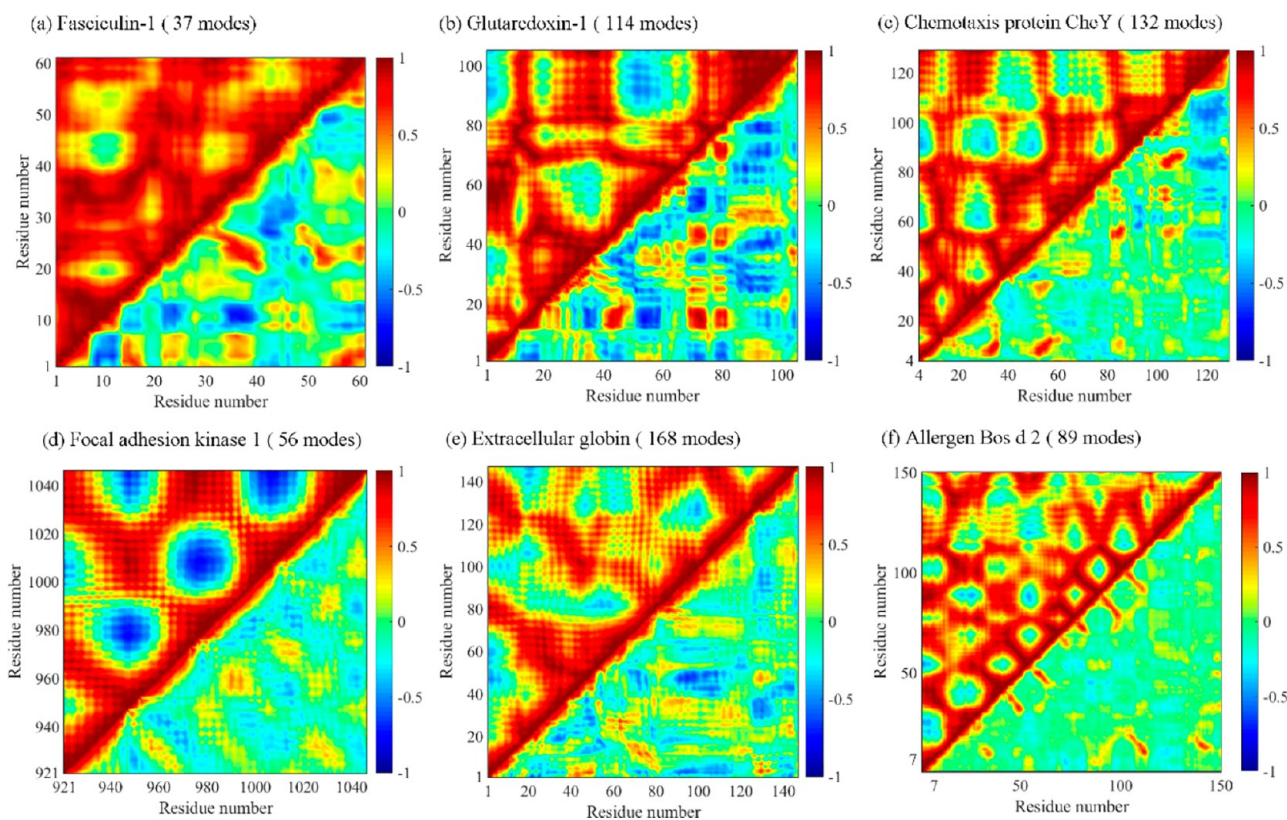


Figure 6. DCCMs obtained from MD ensembles (lower right triangle) and mANM (upper left triangle) at the best *DCCM_PCC* value (corresponding number of motional modes given in parentheses) for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e), and 1BJ7 (f).

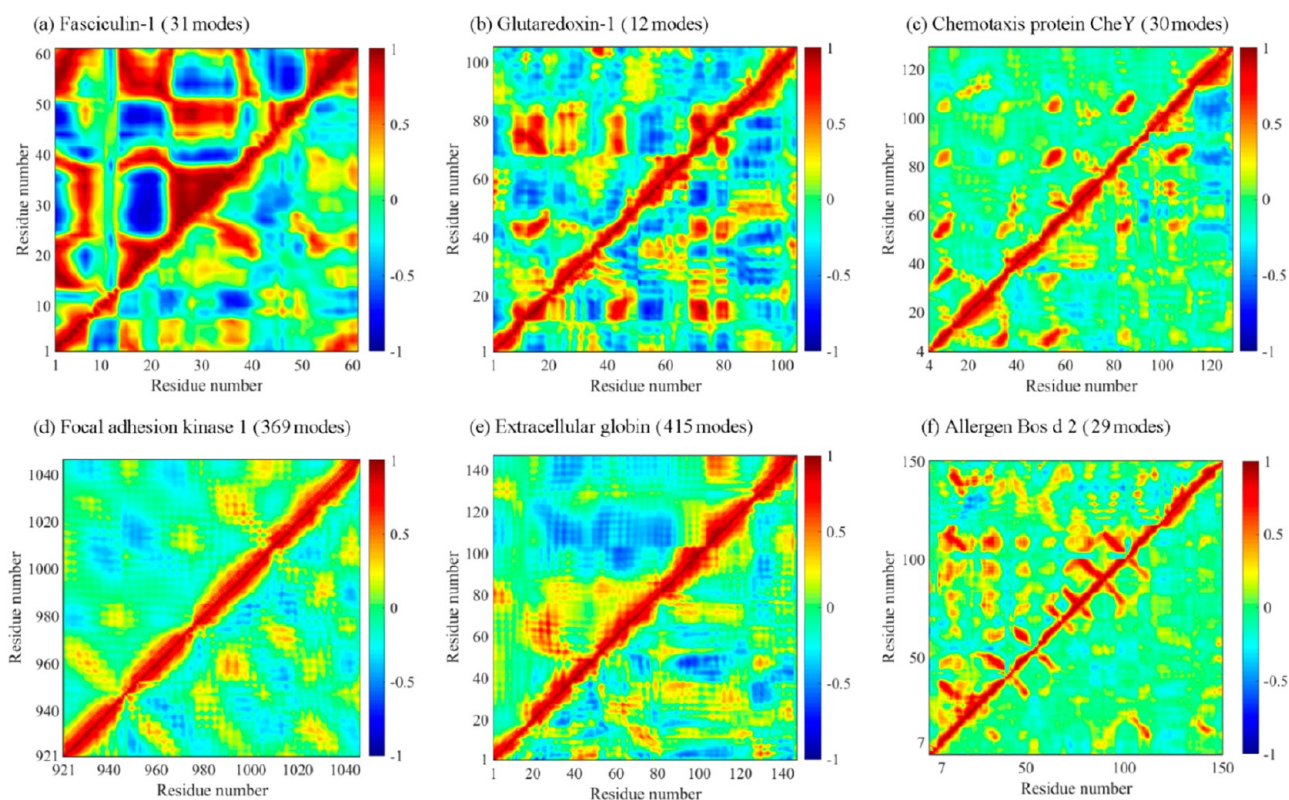


Figure 7. DCCMs obtained from MD ensembles (lower right triangle) and equally weighted mANM (upper left triangle) at the best $DCCM_PCC$ value (corresponding number of motional modes given in parentheses) for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e), and 1BJ7 (f).

Table 3. $DCCM_PCC$ Values between DCCMs Obtained from MD Ensembles and Four Kinds of ENMs Using a Certain Number of Low Frequency Modes That Contribute Just More Than 50% to Residue Fluctuations^a

| | | Protein ID | | | | | |
|------------------------------|-----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Model | | 1FAS | 1KTE | 1CHN | 1K40 | 1ASH | 1BJ7 |
| traditional ENM ^b | traditional GNM | 0.56 (65%) | 0.48 (77%) | 0.61 (50%) | 0.57 (51%) | 0.55 (59%) | 0.51 (66%) |
| | traditional ANM | 0.38 (98%) | 0.79 (51%) | 0.53 (56%) | 0.68 (58%) | 0.66 (52%) | 0.61 (50%) |
| pfENM | pfGNM | 0.65 (51%) | 0.61 (81%) | 0.72 (50%) | 0.46 (75%) | 0.61 (64%) | 0.65 (54%) |
| | pfANM | 0.67 (51%) | 0.76 (51%) | 0.65(50%) | 0.74 (63%) | 0.70 (51%) | 0.74 (56%) |
| mENM | mGNM | 0.56 (85%) | −0.01 (100%) | 0.63(52%) | 0.70 (53%) | 0.49 (52%) | 0.05 (51%) |
| | mANM | 0.35 (97%) | 0.17 (100%) | 0.27(78%) | 0.27 (92%) | 0.30 (96%) | 0.32 (84%) |
| equally weighted mENM | equally weighted mGNM | 0.56 (85%) | 0.31 (69%) | 0.75 (56%) | 0.51 (59%) | 0.49 (50%) | 0.49 (75%) |
| | equally weighted mANM | 0.38 (80%) | 0.80 (51%) | 0.76 (52%) | 0.72 (66%) | 0.53 (61%) | 0.66 (50%) |

^aTwo highest values obtained by GNMs and ANMs, respectively, for each protein are shown in bold. The number of modes considered are shown in parentheses.

corresponding GNMs except for mENM. Like pfGNM and equally weighted mGNM, pfANM and equally weighted mANM perform generally closest to MD simulations, as seen in Figure S6 and Figure 7. Still, like mGNM, mANM performs worst among the four ANMs, which can be seen in Figure 6.

In the actual study, we do not know in priori how many low frequency modes from ENMs should be selected to analyze the dynamical cross-correlations. Generally, the first few low frequency modes that contribute just more than 50% to residue fluctuations are selected to calculate DCCMs.^{40,41} Table 3 shows $DCCM_PCC$ values between the DCCMs obtained from MD ensembles and GNMs or ANMs using a certain number of low frequency modes that contribute just more than 50% to residue fluctuations. For GNMs, still pfGNM performs generally closest to MD simulations. There

exist four cases having the highest values (0.65 for 1FAS, 0.61 for 1KTE, 0.61 for 1ASH, and 0.65 for 1BJ7). Traditional GNM and equally weighted mGNM follow closely behind pfGNM. Still, mGNM performs unstably with one case having the highest value (0.70 for 1K40), whereas four lowest values (0.56 for 1FAS, −0.01 for 1KTE, 0.49 for 1ASH, and 0.05 for 1BJ7) are observed. However, in B-factor computation, mGNM has a remarkable performance on these cases (0.74 for 1FAS, 0.67 for 1KTE, 0.76 for 1ASH, and 0.79 for 1BJ7). The big difference suggests that a good performance of an ENM model in a B-factor calculation does not mean it has a good performance in other dynamical property computations. For ANMs, again pfANM performs generally closest to MD simulations. There are four cases having the highest values (0.67 for 1FAS, 0.74 for 1K40, 0.70 for 1ASH, and 0.74 for

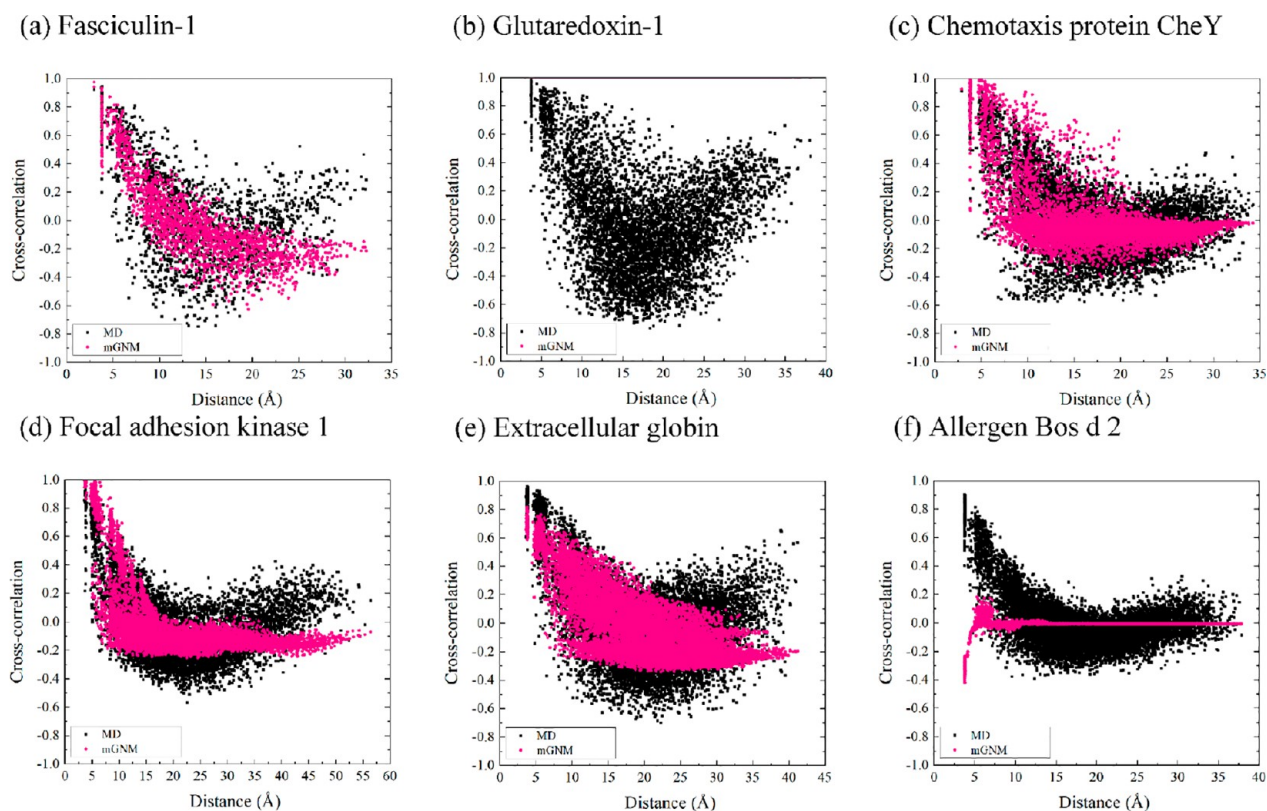


Figure 8. Distributions of dynamical cross-correlations obtained from MD ensembles (black) and mGNM (pink) (at the best DCCM_PCC value) with respect to the inter-residue distance for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e), and 1BJ7 (f).

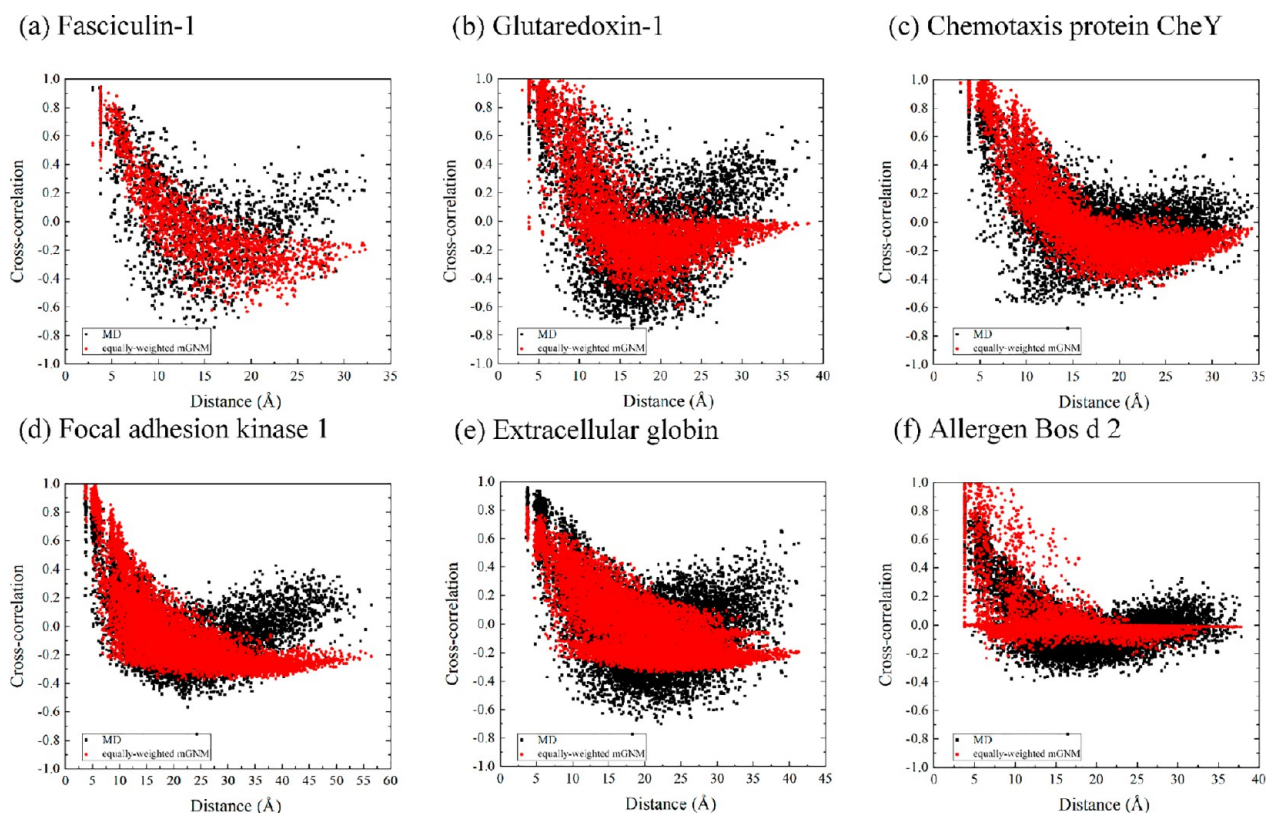


Figure 9. Distributions of dynamical cross-correlations obtained from MD ensembles (black) and equally weighted mGNM (red) (at the best DCCM_PCC value) with respect to the inter-residue distance for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e), and 1BJ7 (f).

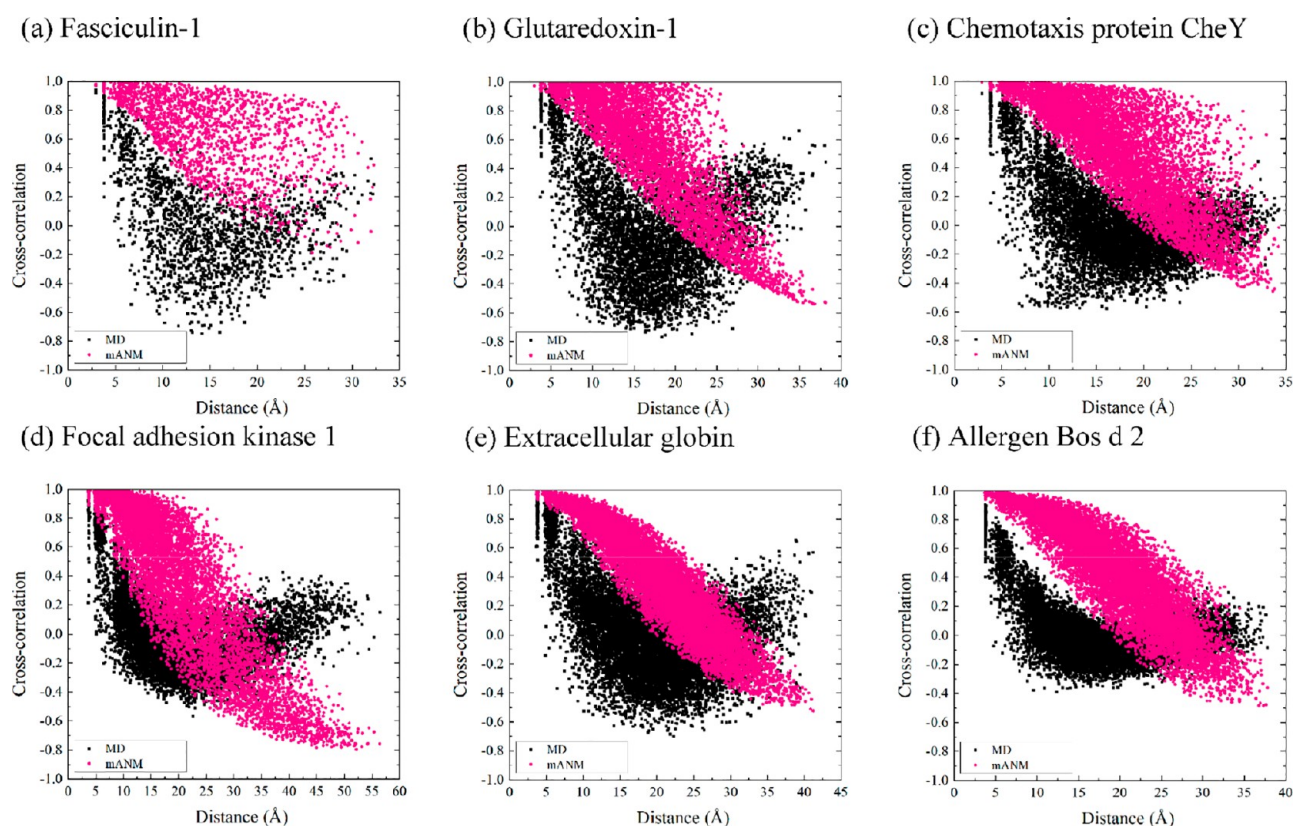


Figure 10. Distributions of dynamical cross-correlations obtained from MD ensembles (black) and mANM (pink) (at the best DCCM_PCC value) with respect to the inter-residue distance for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e), and 1BJ7 (f)

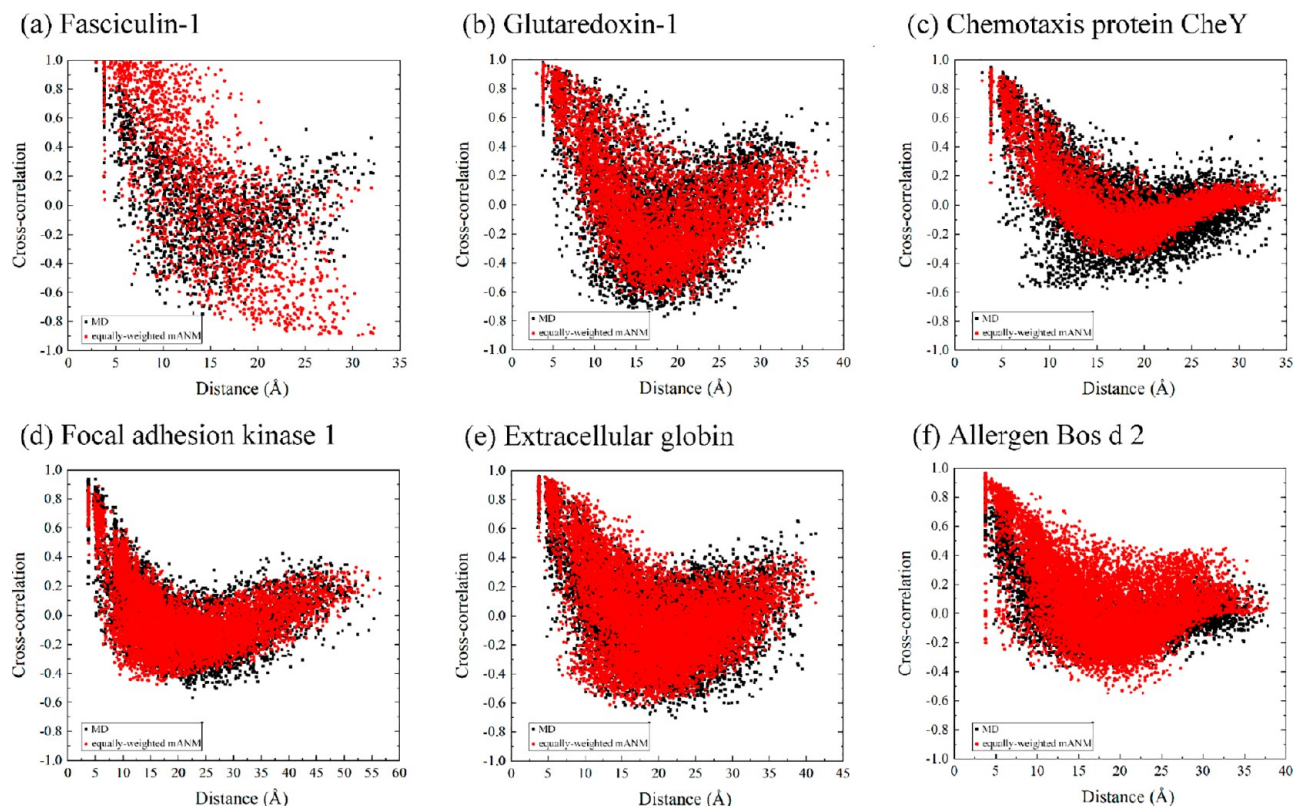


Figure 11. Distributions of dynamical cross-correlations obtained from MD ensembles (black) and equally weighted mANM (red) (at the best DCCM_PCC value) with respect to the inter-residue distance for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e), and 1BJ7 (f)

1BJ7). The equally weighted mANM follows closely behind pfANM with two cases having the highest values (0.80 for 1KTE and 0.76 for 1CHN). Still, mANM has generally the worst performance among all ENM models with all the values less than 0.4.

In conclusion, for GNMs, traditional GNM, pfGNM, and equally weighted mGNM have similar performances in dynamical cross-correlation calculations with the latter two being better. Generally, ANMs perform better than the corresponding GNMs except for the mENM. Thus, pfANM and equally weighted mANM, especially the former, have excellent performances in dynamical cross-correlation calculations.

Comparing Distributions of Cross-Correlations from ENMs and MD Ensembles. Since long-range couplings play an important role in protein folding⁴² and allosteric transitions,^{43,44} we want to explore the distributions of cross-correlations from ENMs. Figures 8–11 and Figures S7–S10 give the distributions (at best DCCM_PCC value) calculated with mENM and equally weighted mENM and traditional ENM and pfENM, respectively, with the results from MD ensembles shown for comparison. For GNMs, the traditional GNM, pfGNM, and equally weighted mGNM produce similar distribution patterns almost for each of the proteins as shown in Figures S7 and S8 and Figure 9, respectively. The pattern shows that the largest positive and negative correlations occur between the residues with the smallest distances apart and half of the maximum distances apart, respectively, which is consistent with the result obtained by Tekpinar and Yildirim on six protein complexes.³⁸ Additionally, at the maximal distance, the correlations show a slightly negative value. However, strangely the distributions from mGNM tend to be a fixed value for 1KTE and 1BJ7, as shown in Figure 8. For them, when the weights are equal and other parameters are kept unchanged, their distributions turn to be normal basically. This means that the optimization of the weights with only the diagonal elements of the Kirchhoff matrix considered is of limitation for calculating DCCMs, although a good performance is obtained in B-factor calculations.

For ANMs, similarly, the traditional ANM, pfANM, and equally weighted mANM produce similar distributions almost for each of the proteins, as shown in Figures S9 and S10 and Figure 11, respectively. Compared with the pattern from GNMs, the main difference from ANMs is that quite a number of positive correlations are captured by ANMs for the residue pairs nearly the largest distances apart, whose correlations are a little negative in the pattern from GNMs, which is probably due to the anisotropy consideration in ANMs. For mANM, the distribution displays an approximately linearly declining pattern almost for each of the proteins, as shown in Figure 10.

In conclusion, the traditional ANM, pfANM, and equally weighted mANM can produce similar distributions of cross-correlations to that from MD simulations, and they can capture good long-range positive correlations for the residue pairs nearly the largest distances apart, which the GNM models cannot achieve.

Comparing ANM Modes with Motions Present in MD Ensembles. It is worth studying how the correlations are between motional modes from ANMs and the principle components (PCs) of motions sampled by MD simulations. For each protein, we calculated the maximum and cumulative overlaps between the first 20 modes from ANMs and each of PC1, PC2, and PC3, respectively (denoted as O_1^{\max} , O_2^{\max} , O_3^{\max}

and CO_1^{20} , CO_2^{20} , CO_3^{20}), and the root mean square inner products (RMSIPs) between the first 20 ANM modes and sets of the first 3, 6, 10, and 20 PCs (denoted as $RMSIP_3^{20}$, $RMSIP_6^{20}$, $RMSIP_{10}^{20}$, and $RMSIP_{20}^{20}$) (see Materials and Methods).

Table 4 shows the average values of each of the 10 performance metrics for six proteins, with the detailed data for

Table 4. Average Values of overlaps and RMSIPs between Motional Modes from ANMs and Principle Components of Motions Sampled by MD Simulations for Six Proteins^a

| Metrics | Model | | | |
|-------------------|--------------------|--------------------|-------------|-----------------------|
| | traditional ANM | pfANM | mANM | equally weighted mANM |
| O_1^{\max} | 0.31 (0.14) | 0.34 (0.14) | 0.24 (0.15) | 0.32 (0.10) |
| O_2^{\max} | 0.27 (0.08) | 0.29 (0.11) | 0.18 (0.12) | 0.29 (0.05) |
| O_3^{\max} | 0.35 (0.09) | 0.30 (0.06) | 0.19 (0.13) | 0.31 (0.07) |
| CO_1^{20} | 0.58 (0.20) | 0.60 (0.16) | 0.38 (0.22) | 0.59 (0.13) |
| CO_2^{20} | 0.54 (0.12) | 0.51 (0.14) | 0.34 (0.21) | 0.56 (0.08) |
| CO_3^{20} | 0.59 (0.07) | 0.57 (0.09) | 0.36 (0.25) | 0.63 (0.07) |
| $RMSIP_3^{20}$ | 0.57 (0.10) | 0.57 (0.10) | 0.36 (0.22) | 0.60 (0.04) |
| $RMSIP_6^{20}$ | 0.31 (0.03) | 0.31 (0.03) | 0.23 (0.09) | 0.32 (0.01) |
| $RMSIP_{10}^{20}$ | 0.54 (0.07) | 0.54 (0.08) | 0.36 (0.21) | 0.57 (0.04) |
| $RMSIP_{20}^{20}$ | 0.50 (0.07) | 0.50 (0.05) | 0.36 (0.20) | 0.53 (0.03) |

^aThe highest value for each of the 10 metrics is shown in bold. Standard deviations are given in parentheses.

each protein shown in Table S3. From Table 4, evidently, the performance of the equally weighted mANM is excellent with seven (O_2^{\max} , CO_2^{20} , CO_3^{20} , $RMSIP_3^{20}$, $RMSIP_6^{20}$, $RMSIP_{10}^{20}$, and $RMSIP_{20}^{20}$) out of 10 metrics having the highest values, and the other three (O_1^{\max} , O_3^{\max} , and CO_1^{20}) all ranked second. Furthermore, their individual standard deviations are almost the lowest among the four models. The following one is pfANM which achieves the highest values in three metrics (O_1^{\max} , O_2^{\max} , and CO_1^{20}) and slightly lower values in the other seven metrics (O_3^{\max} , CO_2^{20} , CO_3^{20} , $RMSIP_3^{20}$, $RMSIP_6^{20}$, $RMSIP_{10}^{20}$, and $RMSIP_{20}^{20}$) than the corresponding highest values. For traditional ANM, only one (O_3^{\max}) has the highest value, and for mANM, each of the 10 has the lowest value. Thus, taking MD simulations as the standard, the performance power of ANMs in producing the low frequency functional motional modes is the equally weighted as mANM > pfANM > traditional ANM > mANM.

Overall, the equally weighted mANM and pfANM, especially the former, perform excellently in reproducing protein motions in MD ensembles. This suggests that the consideration of long-range interactions is critical for ANM models to capture protein functional motions.

As a summary, we compiled a map (Figure 12) about the performances of the ENM models in the calculations of B-factors, DCCMs, and motional modes, which shows that the equally weighted mENM achieves the ideal performance balance in these dynamical property calculations.

Additionally, it is noted that with only the weights of Kirchhoff/Hessian matrixes being equal, the equally weighted mENM displays a remarkable improvement compared with the original mENM in calculating cross-correlations and motional modes and meanwhile has a good performance in B-factor reproduction. The reason could be that the diagonal element (negative sum of the off-diagonal elements representing the

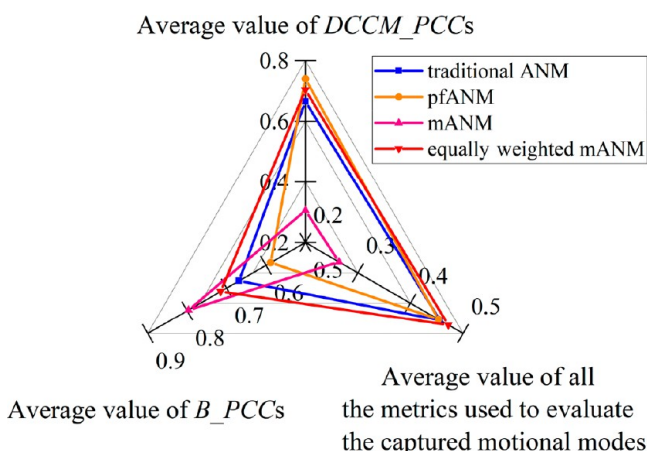


Figure 12. Performance comparison among traditional ANM, pfANM, mANM, and equally weighted mANM in the calculations of B-factor, DCCM, and motional mode with the average values of B_PCC , $DCCM_PCC$, and all the metrics describing the correlations between motional modes from ANMs and PCs of motions sampled by MD simulations over the six proteins. The three axes extend in the positive direction from the origin. The lines connect the values obtained from the same model. Traditional ANM, pfANM, mANM, and equally weighted mANM are colored in blue, orange, pink, and red, respectively.

pairwise interactions) in the Kirchhoff/Hessian matrix reflects the constraints (i.e., rigidity) from other residues in the system on the one the diagonal element represents, and thus, it is not difficult to understand that the optimization of the weights only with the diagonal elements considered can produce a better B-factor calculation. However, it is possible to have different off-diagonal elements and yet obtain the same diagonal elements, which means that the details of pairwise interactions are neglected in optimizing the weights. The optimization way in mENM abolishes the “network” contribution in the ENM model and maybe is the main reason why mENM produces a not ideal performance in calculating cross-correlations and motional modes although a better B-factor computation. Probably, it is better to get an optimal set of parameters through minimizing the differences between the diagonal elements of the pseudoinverse Kirchhoff/Hessian matrix and the corresponding experimental B-factors, which is now underway.

Comparison of Equally Weighted mGNM and mGNM in B-Factor Calculation on a Large Data Set. From above, we can see that with an equal weight for different scales there is a better balance of performance in B-factor calculations and dynamic cross-correlation calculations on the six systems. In order to detect the performance of equally weighted mGNM in B-factor calculations on a large data set, we give B_PCC values of the model on the relatively small- (33 systems), medium-,³⁶ and large-sized³⁵ sets of structures used by the Wu group⁴⁵ and Wei group in their comparative study⁴⁶ in Figure 13 (see Table S4 for detailed data), with the corresponding results from mGNM also shown for comparison. For each of the two models, the optimal η parameters are systematically searched in the range of [1 Å, 26 Å] with step size of 1 Å, and κ is set to 1.²⁴ From Figure 13, for all the sets, generally, the B_PCC values from equally weighted mGNM are the same with or slightly lower than those from mGNM except for one small-sized structure 1USE and two medium-sized ones 2PKT and 1R7J. Poor performances are observed for equally weighted

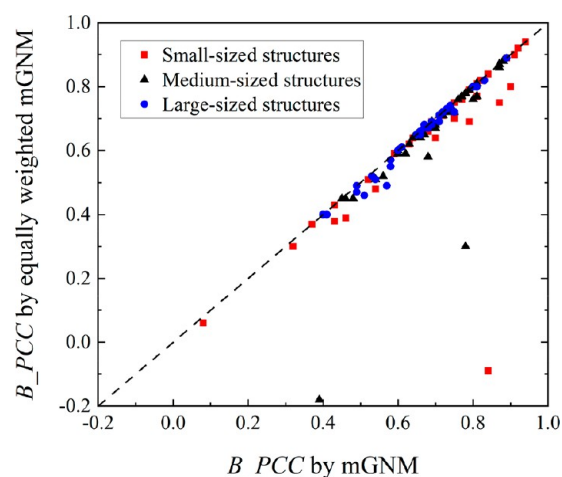


Figure 13. Head-to-head comparison between B_PCC s obtained by equally weighted mGNM and those by mGNM over relatively small-, medium-, and large-sized sets of structures.

mGNM for the three cases (with values being -0.09 , -0.18 , and 0.30), and interestingly, the traditional GNM also gives similar results (-0.14 , -0.19 , and 0.37) reported by Wei group.⁴⁶ The reason for this may be their loosely packed structures. These results suggest again that the performance of equally weighted mGNM is close to that of mGNM in B-factor calculations, consistent with the previous data on the six proteins.

DISCUSSION

Usually, the ENM model is constructed with the C_α atom representing each residue in the protein. The ENM methods with other atoms (not only C_α) representing each residue have also been explored in the past years. Zhang and Kurgan introduced a linear regression-based, parameter-free, sequence-derived GNM (L-pfSeqGNM) whose force constants are from the predicted residue contact maps (with C_β atoms representing residues) to study motions for millions of the readily available sequences.⁴⁷ Additionally, Wu et al. compared the performances of GNMs with backbone atoms C_α , C , N , and C_β representing each residue, respectively, and validated that the model using C_α as the representative atom performs the best in flexibility predictions for a large set of 104 proteins.⁴⁵

For the determination of the force parameters in ENM, fitting theoretical with experimental B-factors is a relatively easy way as the latter can be obtained directly from X-ray structures. In fact, the dynamic cross-correlation can also be used to determine the force constants, which need the MD simulation or NMR ensemble for the studied structure. The REACH method identifies the force constants through an inversion of a covariance matrix (dynamic cross-correlation) derived from MD simulations.⁴⁸ Similarly, the heteroENM utilizes an iterative algorithm to fit the force constants with MD-derived covariances.⁴⁹ Yang et al. directly computed the force constants with the inverse covariance estimation using the ROPE method⁵⁰ on a large-scale set NMR ensemble.⁵¹ Generally, these methods achieve a trade-off on concurrently improving the predictions of residue fluctuation and dynamic cross-correlation compared with the traditional GNM. An additional advantage of using MD-derived covariances is that MD simulations permit in silico alterations to the system under

study, allowing one to find effective force constants that are specific to any environment that can be simulated.

As for the evaluation of ENMs, in fact, the crystallographic B-factor may not be a best index as it is often influenced by crystallization conditions, the refinement method, and importantly, the molecular environment of the crystal structure.⁵² Hence, other experimental data (for example NMR⁵¹) or MD simulations⁵³ should be also considered for ENMs' evaluation. Additionally, Kmiecik et al. developed a Monte Carlo sampling-based coarse-grained approach (CABS-flex) for predicting structure flexibility from a single protein model, reproducing the residue fluctuations well correlated to those of NMR ensembles.⁵²

Usually, the ENM model is constructed based on one reference structure of the protein under study. However, the ENM approaches based on more than one reference structure have been proven to be a good scheme to describe protein dynamics especially for those with evident conformational transitions.^{48,54,55} In 2007, the REACH method proposed by Moritsugu and Smith where the force constants are calculated from MD-derived covariances was applied to monomeric and dimeric myoglobin, reproducing well the residue fluctuations involved in the former intramolecular motions and the latter intermolecular motions.⁴⁸ In 2010, Lezon and Bahar presented a technique for optimizing the distance- and contact order-dependent force constants (OFC-GNM) based on the statistical analyses of the auto- and cross-correlations in residue fluctuations for the NMR ensembles of 68 proteins, achieving remarkable improvements compared with traditional GNM especially in reproducing cross-correlations between residues located far apart in the structure.⁵⁴ In 2015, Katebi et al. constructed the GNM for HIV-1 protease based on 329 X-ray structures representing different states besides "open" and "closed" ones by relating spring constants to the inverse of the variance of internal distance changes of residue pairs, revealing well HIV-1 protease's dynamics, mechanisms, and allostery buried within the structures.⁵⁵

Additionally, many freely available ENM servers have provided the data describing structure dynamics. The MAVEN server (<http://maven.sourceforge.net>) developed by Jernigan et al. incorporates many ENM methods including traditional ANM and GNM and pfANM to facilitate useful analyses on the same platform.⁵⁶ The ANM server (<http://anm.csb.pitt.edu/>) introduced by Eyal et al. is more accessible to nonsophisticated users, and its updated version allows inclusion of nucleic acids and ligands in the ANM model and offers the flexibility of defining network nodes, interaction types, and cutoffs.²⁰ The user-friendly interface and database iGNM (<http://gnmdb.csb.pitt.edu/>) developed by Bahar et al. covers more than 95% of the structures available in PDB, which makes it a useful resource for establishing the bridge between structure, dynamics, and function. Advanced search and visualization capabilities permit users to retrieve information on inter-residue and interdomain cross-correlations, cooperative motion modes, and locations of hinge sites.¹⁵

At last, as for the application of ENMs, as a good complementarity, Kolinski et al. developed efficient coarse-grained (C_α , C_β , side-chain (CABS) CG model) modeling methods (CABS MC and CABS-flex) based on the Monte Carlo (MC) technique for protein structure and dynamics simulations. For the structural fluctuation, ENM and CABS-flex provide dynamic profiles very close to those from NMR;⁵²

For large-scale structural transitions, ENM methods are well suited only for certain problems, such as protein dynamics around the native state or conformational transitions between "open" and "closed" states, while the CABS MC simulations can be used for simulations of general structural transitions such as entire folding/unfolding processes.^{57,58}

CONCLUSIONS

In the present study, we systematically compare the performances of the widely used traditional ENM and pfENM models, and the recently proposed mENM model with multiscale interactions considered in the calculations of B-factors, dynamical cross-correlation maps (DCCMs, and motional modes on the six representative proteins with MD simulations taken as standard for the evaluation of the latter two dynamical property calculations. Here, we introduce an equally weighted mENM model based on the original one.

In the B-factor calculation, mENM has a remarkable performance, better than the traditional ENM and pfENM models. Interestingly, with the weights of Kirchhoff/Hessian matrixes being equal, the equally weighted mENM still has a good performance with the equally weighted mGNM better than the corresponding mANM, following closely after mENM. In the DCCM calculation, compared with GNMs, generally the corresponding ANMs (except for mENM) produce more similar DCCMs to those from MD ensembles regardless of the certain number of low frequency modes considered in ANMs that contribute just more than 50% to residue fluctuations or at the best pearson's correlation coefficient between the two DCCMs. This may be due to the anisotropy consideration in ANM models. pfANM and the equally weighted mANM models have a very good performance. However, mGNM gives an unstable performance, and mANM fails in the six cases. For motional mode extraction, encouragingly, the equally weighted mANM has a best performance with the results being closest to those from MD simulations, which is followed by pfANM, which suggests that the consideration of long-range interactions plays a critical role in capturing the collective functional modes encoded in protein structures. However, still the mANM fails in the six cases.

Considering that the six proteins are not very big (residue numbers ranging 61–150) and have no evidently large conformational changes, the 10 ns simulation time may be acceptable for extracting dynamic cross-correlations. In the future, the method will be tested on the bigger proteins with evident conformational changes, which should require longer simulation times to sample the conformational space. This work is helpful for strengthening the understanding of the physical bases of ENM methods and guiding researchers to select suitable ENM models to explore the function-related dynamics of proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.0c01178>.

Figures S1–S10 and Tables S1–S4. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Chunhua Li – Faculty of Environmental and Life Sciences and Beijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation, Beijing University of Technology, Beijing 100124, ChinaBeijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation and Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China; Email: chunhuali@bjut.edu.cn

Authors

Weikang Gong – Faculty of Environmental and Life Sciences and Beijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation, Beijing University of Technology, Beijing 100124, ChinaBeijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation and Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China; orcid.org/0000-0001-8797-784X

Yang Liu – Faculty of Environmental and Life Sciences and Beijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation, Beijing University of Technology, Beijing 100124, ChinaBeijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation and Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China

Yanpeng Zhao – Faculty of Environmental and Life Sciences and Beijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation, Beijing University of Technology, Beijing 100124, ChinaBeijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation and Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China

Shihao Wang – Faculty of Environmental and Life Sciences and Beijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation, Beijing University of Technology, Beijing 100124, ChinaBeijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation and Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China

Zhongjie Han – Faculty of Environmental and Life Sciences and Beijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation, Beijing University of Technology, Beijing 100124, ChinaBeijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation and Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.0c01178>

Author Contributions

W.G. and C.L. designed the research. W.G. wrote the program of equally weighted mENM models. W.G., Y.L., Y.Z., S.W., and Z.H. performed data analyses. W.G. and C.L. wrote the manuscript. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Qi Shao for helpful discussions. This work was supported by the National Natural Science Foundation of China [31971180, 11474013].

■ REFERENCES

- (1) Kaynak, B. T.; Findik, D.; Doruker, P. RESPEC Incorporates Residue Specificity and the Ligand Effect into the Elastic Network Model. *J. Phys. Chem. B* **2018**, 122 (21), 5347–5355.
- (2) Wang, W. B.; Liang, Y.; Zhang, J.; Wu, Y. D.; Du, J. J.; Li, Q. M.; Zhu, J. Z.; Su, J. G. Energy Transport Pathway in Proteins: Insights from Non-equilibrium Molecular Dynamics with Elastic Network Model. *Sci. Rep.* **2018**, 8 (1), 9487.
- (3) Cheng, M. H.; Bahar, I. Monoamine Transporters: Structure, Intrinsic Dynamics and Allosteric Regulation. *Nat. Struct. Mol. Biol.* **2019**, 26 (7), 545–556.
- (4) Mikulska-Ruminska, K.; Shrivastava, I.; Krieger, J.; Zhang, S.; Li, H.; Bayir, H.; Wenzel, S. E.; VanDemark, A. P.; Kagan, V. E.; Bahar, I. Characterization of Differential Dynamics, Specificity, and Allostery of Lipoxygenase Family Members. *J. Chem. Inf. Model.* **2019**, 59 (5), 2496–2508.
- (5) Zhang, S.; Li, H.; Krieger, J. M.; Bahar, I. Shared Signature Dynamics Tempered by Local Fluctuations Enables Fold Adaptability and Specificity. *Mol. Biol. Evol.* **2019**, 36 (9), 2053–2068.
- (6) Kohn, J. E.; Afonine, P. V.; Ruscio, J. Z.; Adams, P. D.; Head-Gordon, T. Evidence of Functional Protein Dynamics from X-ray Crystallographic Ensembles. *PLoS Comput. Biol.* **2010**, 6 (8), No. e1000911.
- (7) Fenwick, R. B.; van den Bedem, H.; Fraser, J. S.; Wright, P. E. Integrated Description of Protein Dynamics from Room-temperature X-ray Crystallography and NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, 111 (4), E445–E454.
- (8) Fernandez-Leiro, R.; Scheres, S. H. Unravelling Biological Macromolecules with Cryo-electron Microscopy. *Nature* **2016**, 537 (7620), 339–346.
- (9) Zhang, Y.; Gong, W.; Wang, Y.; Liu, Y.; Li, C. Exploring Movement and Energy in Human P-glycoprotein Conformational Rearrangement. *J. Biomol. Struct. Dyn.* **2019**, 37 (5), 1104–1119.
- (10) Li, C.; Zuo, Z.; Su, J.; Xu, X.; Wang, C. The Interactions and Recognition of Cyclic Peptide Mimetics of Tat with HIV-1 TAR RNA: A Molecular Dynamics Simulation Study. *J. Biomol. Struct. Dyn.* **2013**, 31 (3), 276–287.
- (11) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, 77 (9), 1905–1908.
- (12) Zheng, W.; Wen, H. A Survey of Coarse-grained Methods for Modeling Protein Conformational Transitions. *Curr. Opin. Struct. Biol.* **2017**, 42, 24–30.
- (13) Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **2013**, 42, 73–93.
- (14) Han, Z.; Shao, Q.; Gong, W.; Wang, S.; Su, J.; Li, C.; Zhang, Y. Interpreting the Dynamics of Binding Interactions of snRNA and U1A Using a Coarse-Grained Model. *Biophys. J.* **2019**, 116 (9), 1625–1636.
- (15) Li, H.; Chang, Y.; Yang, L.; Bahar, I. iGNM 2.0: the Gaussian Network Model Database for Biomolecular Structural Dynamics. *Nucleic Acids Res.* **2016**, 44 (D1), D415–D422.

- (16) Yang, L.; Rader, A. J.; Liu, X.; Jursa, C. J.; Chen, S. C.; Karimi, H. A.; Bahar, I. oGNM: Online Computation of Structural Dynamics Using the Gaussian Network Model. *Nucleic Acids Res.* **2006**, *34* (SI), W24–W31.
- (17) Bahar, I. Dynamics of Proteins and Biomolecular Complexes: Inferring Functional Motions from Structure. *Rev. Chem. Eng.* **1999**, *15* (4), 319–347.
- (18) Lv, D.; Gong, W.; Zhang, Y.; Liu, Y.; Li, C. A Coarse-grained Method to Predict the Open-to-closed Behavior of Glutamine Binding Protein. *Chem. Phys.* **2017**, *493*, 166–174.
- (19) Li, C.; Lv, D.; Zhang, L.; Yang, F.; Wang, C.; Su, J.; Zhang, Y. Approach to the Unfolding and Folding Dynamics of Add A-riboswitch upon Adenine Dissociation Using a Coarse-grained Elastic Network Model. *J. Chem. Phys.* **2016**, *145* (1), 014104.
- (20) Eyal, E.; Lum, G.; Bahar, I. The Anisotropic Network Model Web Server at 2015 (ANM 2.0). *Bioinformatics* **2015**, *31* (9), 1487–1489.
- (21) Hamacher, K.; McCammon, J. A. Computing the Amino Acid Specificity of Fluctuations in Biomolecular Systems. *J. Chem. Theory Comput.* **2006**, *2* (3), 873–878.
- (22) Yang, L.; Song, G.; Jernigan, R. L. Protein Elastic Network Models and the Ranges of Cooperativity. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (30), 12347–12352.
- (23) Zhang, H.; Jiang, T.; Shan, G.; Xu, S.; Song, Y. Gaussian Network Model can be Enhanced by Combining Solvent Accessibility in Proteins. *Sci. Rep.* **2017**, *7* (1), 7486.
- (24) Xia, K.; Opron, K.; Wei, G. Multiscale Gaussian Network Model (mGNM) and Multiscale Anisotropic Network Model (mANM). *J. Chem. Phys.* **2015**, *143* (20), 204106.
- (25) Sankar, K.; Mishra, S. K.; Jernigan, R. L. Comparisons of Protein Dynamics from Experimental Structure Ensembles, Molecular Dynamics Ensembles, and Coarse-Grained Elastic Network Models. *J. Phys. Chem. B* **2018**, *122* (21), 5409–5417.
- (26) Kormos, B. L.; Baranger, A. M.; Beveridge, D. L. Do Collective Atomic Fluctuations Account for Cooperative Effects? Molecular Dynamics Studies of the U1A-RNA Complex. *J. Am. Chem. Soc.* **2006**, *128* (28), 8992–8993.
- (27) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Bruschweiler, R.; Blackledge, M. Identification of Slow Correlated Motions in Proteins using Residual Dipolar and Hydrogen-bond Scalar Couplings. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (39), 13885–13890.
- (28) Su, J.; Zhang, X.; Han, X.; Zhao, S.; Li, C. The Intrinsic Dynamics and Unfolding Process of an Antibody Fab Fragment Revealed by Elastic Network Model. *Int. J. Mol. Sci.* **2015**, *16* (12), 29720–29731.
- (29) Zhang, P.; Su, J. Identification of Key Sites Controlling Protein Functional Motions by Using Elastic Network Model Combined with Internal Coordinates. *J. Chem. Phys.* **2019**, *151* (4), 045101.
- (30) Meyer, T.; D'Abramo, M.; Hospital, A.; Rueda, M.; Ferrer-Costa, C.; Perez, A.; Carrillo, O.; Camps, J.; Fenollosa, C.; Repchevsky, D.; Gelpi, J. L.; Orozco, M. MoDEL (Molecular Dynamics Extended Library): A Database of Atomistic Molecular Dynamics Trajectories. *Structure* **2010**, *18* (11), 1399–1409.
- (31) Ledu, M. H.; Marchot, P.; Bougis, P. E.; Fontecilla-Camps, J. C. 1.9-Angstrom Resolution Structure of Fasciculin-I, An Anticetylcholinesterase Toxin From Green Mamba Snake-venom. *J. Biol. Chem.* **1992**, *267* (31), 22122–22130.
- (32) Hayashi, I.; Vuori, K.; Liddington, R. C. The Focal Adhesion Targeting (FAT) Region of Focal Adhesion Kinase is a Four-helix Bundle that Binds Paxillin. *Nat. Struct. Biol.* **2002**, *9* (2), 101–106.
- (33) Yang, J.; Kloek, A. P.; Goldberg, D. E.; Mathews, F. S. The Structure of Ascaris Hemoglobin Domain I at 2.2 Angstrom Resolution: Molecular Features of Oxygen Avidity. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (10), 4224–4228.
- (34) Katti, S. K.; Robbins, A. H.; Yang, Y.; Wells, W. W. Crystal-structure of Thioltransferase at 2.2 Angstrom Resolution. *Protein Sci.* **1995**, *4* (10), 1998–2005.
- (35) Bellolell, L.; Prieto, J.; Serrano, L.; Coll, M. Magnesium Binding to the Bacterial Chemotaxis Protein CheY Results in Large Conformational Changes Involving its Functional Surface. *J. Mol. Biol.* **1994**, *242* (1), 103.
- (36) Rouvinen, J.; Rautiainen, J.; Virtanen, T.; Zeiler, T.; Kauppinen, J.; Taivainen, A.; Mantjarvi, R. Probing the Molecular Basis of Allergy Three-dimensional Structure of the Bovine Lipocalin Allergen Bos d 2. *J. Biol. Chem.* **1999**, *274* (4), 2337–2343.
- (37) Bahar, I.; Rader, A. J. Coarse-grained Normal Mode Analysis in Structural Biology. *Curr. Opin. Struct. Biol.* **2005**, *15* (5), 586–592.
- (38) Tekpinar, M.; Yildirim, A. Only a Subset of Normal Modes is Sufficient to Identify Linear Correlations in Proteins. *J. Chem. Inf. Model.* **2018**, *58* (9), 1947–1961.
- (39) Fuglebakk, E.; Reuter, N.; Hinsén, K. Evaluation of Protein Elastic Network Models Based on an Analysis of Collective Motions. *J. Chem. Theory Comput.* **2013**, *9* (12), S618–S628.
- (40) Li, C.; Yang, Y.; Su, J.; Liu, B.; Tan, J.; Zhang, X.; Wang, C. Allosteric Transitions of the Maltose Transporter Studied by an Elastic Network Model. *Biopolymers* **2014**, *101*, 758–768.
- (41) Zheng, W.; Brooks, B. Identification of Dynamical Correlations within the Myosin Motor Domain by the Normal Mode Analysis of an Elastic Network Model. *J. Mol. Biol.* **2005**, *346*, 745–759.
- (42) Zhang, X.; Rebane, A. A.; Ma, L.; Li, F.; Jiao, J.; Qu, H.; Pincet, F.; Rothman, J. E.; Zhang, Y. Stability, Folding Dynamics, and Long-range Conformational Transition of the Synaptic t-SNARE Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E8031–E8040.
- (43) Gourinchas, G.; Etzl, S.; Gobl, C.; Vide, U.; Madl, T.; Winkler, A. Long-range Allosteric Signaling in Red Light-regulated Diguanylyl Cyclases. *Sci. Adv.* **2017**, *3* (3), No. e1602498.
- (44) Chan, Y. M.; Moustafa, I. M.; Arnold, J. J.; Cameron, C. E.; Boehr, D. D. Long-Range Communication between Different Functional Sites in the Picornaviral 3C Protein. *Structure* **2016**, *24* (4), 509–517.
- (45) Park, J. K.; Jernigan, R.; Wu, J. Coarse Grained Normal Mode Analysis vs. Refined Gaussian Network Model for Protein Residue-Level Structural Fluctuations. *Bull. Math. Biol.* **2013**, *75* (1), 124–160.
- (46) Opron, K.; Xia, K.; Wei, G. Fast and anisotropic flexibility-rigidity index for protein flexibility and fluctuation analysis. *J. Chem. Phys.* **2014**, *140* (23), 234105.
- (47) Zhang, H.; Kurgan, L. Sequence-based Gaussian network model for protein dynamics. *Bioinformatics* **2014**, *30* (4), 497–505.
- (48) Moritsugu, K.; Smith, J. C. Coarse-Grained Biomolecular Simulation with REACH: Realistic Extension Algorithm via Covariance Hessian. *Biophys. J.* **2007**, *93*, 3460–3469.
- (49) Lyman, E.; Pfaendner, J.; Voith, G. A. Systematic Multiscale Parameterization of Heterogeneous Elastic Network Models of Proteins. *Biophys. J.* **2008**, *95*, 4183–4192.
- (50) Kuismin, M. O.; Kemppainen, J. T.; Sillanpää, M. J. Precision Matrix Estimation With ROPE. *J. Comput. Graph. Stat.* **2017**, *26* (3), 682–694.
- (51) Zhang, H.; Shan, G.; Yang, B. Optimized elastic network models with direct characterization of inter residue cooperativity for protein dynamics. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2020**, *1*.
- (52) Jamroz, M.; Kolinski, A.; Kmiecik, S. CABS-flex predictions of protein flexibility compared with NMR ensembles. *Bioinformatics* **2014**, *30* (15), 2150–2154.
- (53) Hu, G.; He, L.; Iacovelli, F.; Falconi, M. Intrinsic Dynamics Analysis of a DNA Octahedron by Elastic Network Model. *Molecules* **2017**, *22* (1), 145.
- (54) Lezon, T. R.; Bahar, I. Using Entropy Maximization to Understand the Determinants of Structural Dynamics beyond Native Contact Topology. *PLoS Comput. Biol.* **2010**, *6* (6), No. e1000816.
- (55) Katebi, A. R.; Sankar, K.; Jia, K.; Jernigan, R. L. The Use of Experimental Structures to Model Protein Dynamics. *Methods Mol. Biol.* **2015**, *1215*, 213–236.
- (56) Zimmermann, M. T.; Kloczkowski, A.; Jernigan, R. L. MAVENs: Motion analysis and visualization of elastic networks and structural ensembles. *BMC Bioinf.* **2011**, *12*, 264.

(57) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116* (14), 7898.

(58) Kmiecik, S.; Kouza, M.; Badaczewska-Dawid, A. E.; Kloczkowski, A.; Kolinski, A. Modeling of Protein Structural Flexibility and Large-Scale Dynamics: Coarse-Grained Simulations and Elastic Network Models. *Int. J. Mol. Sci.* **2018**, *19* (11), 3496.