

Anharmonic Normal Mode Analysis of Elastic Network Model Improves the Modeling of Atomic Fluctuations in Protein Crystal Structures

Wenjun Zheng*

Physics Department, University at Buffalo, Buffalo, New York

ABSTRACT Protein conformational dynamics, despite its significant anharmonicity, has been widely explored by normal mode analysis (NMA) based on atomic or coarse-grained potential functions. To account for the anharmonic aspects of protein dynamics, this study proposes, and has performed, an anharmonic NMA (ANMA) based on the C_{α} -only elastic network models, which assume elastic interactions between pairs of residues whose C_{α} atoms or heavy atoms are within a cutoff distance. The key step of ANMA is to sample an anharmonic potential function along the directions of eigenvectors of the lowest normal modes to determine the mean-squared fluctuations along these directions. ANMA was evaluated based on the modeling of anisotropic displacement parameters (ADPs) from a list of 83 high-resolution protein crystal structures. Significant improvement was found in the modeling of ADPs by ANMA compared with standard NMA. Further improvement in the modeling of ADPs is attained if the interactions between a protein and its crystalline environment are taken into account. In addition, this study has determined the optimal cutoff distances for ADP modeling based on elastic network models, and these agree well with the peaks of the statistical distributions of distances between C_{α} atoms or heavy atoms derived from a large set of protein crystal structures.

INTRODUCTION

Conformational dynamics at the atomic level is increasingly recognized as important in protein functions including protein-ligand interactions, protein-protein interactions, and allosteric regulation (1–3). Protein dynamics spans a wide range of timescales, from femtoseconds to seconds. Of particular interest is the long-time dynamics (microseconds to seconds) in large protein complexes, which is far beyond the timescales (nanoseconds to microseconds) of atomistic molecular dynamics (MD) simulations (4) despite fast-advancing computing technology (5). To capture such slow protein dynamics, normal mode analysis (NMA) was developed and has been widely applied (6–10). In a typical NMA of protein dynamics, an atomic or coarse-grained potential function is approximated by a harmonic potential near a minimal-energy conformation, from which a Hessian matrix (the second derivatives of potential function) is calculated. Then, a set of normal modes are solved from the Hessian matrix, and these can be used to describe small-amplitude atomic motions at low temperatures. Under harmonic assumption, the inverse of the eigenvalue of a normal mode is proportional to the mean-squared fluctuation (MSF) of atomic coordinates along the direction of its eigenvector.

To facilitate the application of NMA to large protein complexes, it is often performed based on a coarse-grained protein structural model (11). In recent studies, elastic network models (ENMs), including the anisotropic network model (ANM) (12–14) and its isotropic variation, the Gaussian network model (GNM) (15,16), have been devel-

oped to model protein dynamics at amino acid resolution. The ENM is usually constructed based on a C_{α} -only representation of protein structures, where pairs of residues whose C_{α} atoms or heavy atoms are within a cutoff distance are connected by elastic springs with a uniform (12) or distance-dependent (17,18) force constant. Such dramatic simplification allows the coarse-grained normal modes to be calculated efficiently without energy minimization. Early studies have shown that the large-scale collective motions predicted by NMA of ENMs are insensitive to the dramatic simplification in ENMs (13,14). Indeed, the lowest normal modes calculated from ENMs were found to compare well with many large-scale conformational changes observed by crystallography (13,19). Numerous studies have established ENM as an efficient means of probing the functionally relevant protein dynamics with virtually no limit in timescale or system size (20–25).

Despite the great success of NMA, its underlying assumption of harmonicity remains questionable. It is well known that protein dynamics at physiological or lower temperatures is highly anharmonic thanks to various factors, including solvation effect and the multim minima potential energy function. For example, an early study using NMA and MD simulation found that an isolated bovine pancreatic trypsin inhibitor (BPTI) exhibits marked anharmonic dynamics at temperatures of 100–120 K (26). Therefore, the standard NMA is quantitatively inaccurate to describe the anharmonic atomic fluctuations in protein structures, which often results in an underestimation of the MSF (27). To explore the anharmonic aspects of protein dynamics, the quasiharmonic analysis was developed based on the principal component analysis of an MD trajectory (28). It yields quasiharmonic modes that can be compared with the normal modes from standard NMA. In one simulation study of BPTI, it was

Submitted January 5, 2010, and accepted for publication March 15, 2010.

*Correspondence: wjzheng@buffalo.edu

Editor: Gerhard Hummer.

© 2010 by the Biophysical Society
0006-3495/10/06/3025/10 \$2.00

doi: 10.1016/j.bpj.2010.03.027

found that the first quasiharmonic mode shows barrier crossing events on an anharmonic energy surface, whereas the higher quasiharmonic modes appear to be largely harmonic (27). In another simulation study of BPTI (29), it was shown that the larger the MSF of a quasiharmonic mode, the greater the degree of anharmonicity in its motion. For BPTI, the anharmonic modes represent only 12% of the total number of variables, but they account for 98% of the total MSF (29). Therefore, a proper account of anharmonicity for the lowest quasiharmonic or normal modes is critical to accurate modeling of protein dynamics.

Despite the strong effects of anharmonicity on the magnitude of atomic fluctuations, it has been shown that the eigenvectors of the lowest modes constitute an essential subspace of collective coordinates, which offer good descriptions of collective motions in proteins (30,31). It is argued that the directions (but not the magnitude) of these collective motions as captured by the eigenvectors of the lowest modes are insensitive to details of microscopic interactions and solvent damping (25). Therefore, a promising recipe for modeling anharmonicity within the framework of NMA is to keep the eigenvectors (but not eigenvalues) of standard NMA, and then sample atomic fluctuations/motions along the directions of these eigenvectors using an anharmonic potential function (see Methods). This general approach is here referred to as anharmonic NMA (ANMA). A similar strategy has been adopted in previous studies, where the eigenvectors of lowest normal modes were utilized to enhance MD simulations (32), refine x-ray diffraction data (33), and fit low-resolution electron microscopy maps (34) and crystallographic B factors (35).

This study employs ANMA to model anharmonicity within the framework of ENM. The ENM potential function, despite its quadratic form (Eq. 1), is actually anharmonic in Cartesian coordinate space. This study will evaluate how ANMA improves the modeling of anisotropic atomic fluctuations obtained by x-ray crystallographic studies of protein structures. Atomic fluctuations in protein crystal structures have been traditionally quantified by the isotropic temperature factors (or B factors), which use an isotropic Gaussian distribution to characterize the spread of electron density of each atom (36). Recently, a growing number of high-resolution protein crystal structures have been refined using anisotropic Gaussian distributions, which characterize atomic fluctuations by a symmetric tensor with six independent elements called anisotropic displacement parameters (ADPs) (37,38). Unlike the B factors, the ADPs describe not only the magnitude but also the direction of mean-squared atomic fluctuations. Therefore, they offer richer information of protein dynamics in crystalline states. Modeling of ADPs based on NMA has been conducted in several recent studies (39–44), which only considered harmonic contributions to protein dynamics. This study is, to our knowledge, the first in which anharmonicity is considered in the NMA-based modeling of ADPs.

Based on the modeling of ADPs from a large set of 83 high-resolution protein crystal structures studied previously (39,41,43), the main findings are summarized as follows:

1. ANMA significantly improves the modeling of anisotropic atomic fluctuations described by ADPs, especially for low cutoff distances of ENMs. Unlike the standard NMA, ANMA is not susceptible to the tip effect, which causes anomalously large fluctuations of some protruding structural elements of a globular protein (45). Explicit accounting of the interactions between a protein and its crystalline environment leads to further improvement in the modeling of ADPs.
2. By analyzing the cutoff-distance dependence of ADP modeling, this work determines the optimal cutoff distances for three ENM schemes (ANM_{C α} , ANM_{atom}, and DNM; see Methods). The values of optimal cutoff distances correspond to a secondary peak in the distribution of C α -C α distances, and the upper bound of a broad peak in the distribution of heavy-atom distances, which are derived from a large set of protein crystal structures. This finding hints at the dynamic importance of indirect atomic interactions beyond the range of direct interactions via van der Waals or hydrogen-bond forces.

METHODS

ENM for an isolated protein structure

An ENM is constructed based on the C α coordinates of a protein crystal structure. The potential energy of an ENM is

$$E = \frac{1}{2} \sum_{i < j} C_{ij} (d_{ij} - d_{ij,0})^2, \quad (1)$$

where C_{ij} is the force constant of the spring connecting C α atoms i and j , d_{ij} is the atomic distance between C α atoms i and j , and $d_{ij,0}$ is the value of d_{ij} given by the crystal structure. Three schemes of force constant assignments are considered here:

1. ANM_{C α} :

$$C_{ij} = \begin{cases} C & \text{if } d_{ij,0} < R_c, \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

where R_c is a cutoff distance that varies between 7 and 20 Å, and the value of constant C is determined by fitting the experimental ADPs (see below).

2. ANM_{atom}:

$$C_{ij} = C \sum_{i_1 j_1} \theta(R_{\text{atom}} - d_{i_1 j_1,0}), \quad (3)$$

where $\theta()$ is the Heaviside function, and the summation is over all pairs of heavy atoms of residues i and j within a cutoff distance R_{atom} (R_{atom} varies between 4 and 15 Å). Here, i_1 and j_1 are indices for the heavy atoms of residues i and j .

3. DNM:

$$C_{ij} = C \sum_{i_1 j_1} \frac{\theta(R_{\text{atom}} - d_{i_1 j_1,0})}{d_{i_1 j_1,0}^2}. \quad (4)$$

The DNM was proposed by Kondrashov et al. (39), who set force constants for several distance ranges to the reciprocal of the total number of atomic contacts in each range. Because the number of atomic contacts

grows quadratically with distance (assuming the atomic density is constant), this work's DNM formulation is essentially a continuous counterpart of the original DNM (39).

The Hessian matrix, H , is calculated as the second derivative of potential energy E (see Eq. 1) with respect to the Cartesian coordinates of C_α atoms. For a protein structure with N residues, H contains $N \times N$ superelements (called H_{ij}) with size 3×3 :

$$H_{ij} = \begin{bmatrix} \frac{\partial^2 E}{\partial x_i \partial x_j} & \frac{\partial^2 E}{\partial x_i \partial y_j} & \frac{\partial^2 E}{\partial x_i \partial z_j} \\ \frac{\partial^2 E}{\partial y_i \partial x_j} & \frac{\partial^2 E}{\partial y_i \partial y_j} & \frac{\partial^2 E}{\partial y_i \partial z_j} \\ \frac{\partial^2 E}{\partial z_i \partial x_j} & \frac{\partial^2 E}{\partial z_i \partial y_j} & \frac{\partial^2 E}{\partial z_i \partial z_j} \end{bmatrix}, \quad (5)$$

where x_i, y_i , and z_i are the Cartesian coordinates of the C_α atom i .

A subset of total $3N$ normal modes is then solved. For each mode m , its eigenvalue (represented by λ_m) and eigenvector (represented by V_m) are obtained by solving $HV_m = \lambda_m V_m$.

Anharmonicity of the ENM potential function

Although the ENM potential function in Eq. 1 has a quadratic form, it is actually anharmonic in terms of Cartesian coordinates. In fact, the NMA of ENM is based on the Taylor expansion of Eq. 1 truncated at the second order, with higher-order anharmonic terms ignored. Let us consider the elastic interaction energy between C_α atoms i and j after they undergo three-dimensional displacements $\Delta \vec{r}_i$ and $\Delta \vec{r}_j$ from their equilibrium positions. This energy is expanded to the fourth order as follows (for a detailed derivation, see Supporting Material):

$$\begin{aligned} & \frac{1}{2} C_{ij} (d_{ij} - d_{ij,0})^2 \\ &= \frac{1}{2} C_{ij} \left[\sqrt{(d_{ij,0} + \Delta r_{ij,L})^2 + \Delta r_{ij,T}^2} - d_{ij,0} \right]^2 \\ &\sim \frac{C_{ij} d_{ij,0}^2}{2} \left[a^2 + ab^2 + \frac{b^4}{4} - a^2 b^2 \right. \\ &\quad \left. + \text{fifth - or higher - order terms} \right], \end{aligned} \quad (6)$$

where the relative displacement ($\Delta \vec{r}_j - \Delta \vec{r}_i$) is partitioned into a longitudinal component, $\Delta r_{ij,L} = (\Delta \vec{r}_j - \Delta \vec{r}_i) \cdot \vec{n}_{ij}^0$, and a transverse component, $\Delta r_{ij,T} = \sqrt{|\Delta \vec{r}_j - \Delta \vec{r}_i|^2 - \Delta r_{ij,L}^2}$, and \vec{n}_{ij}^0 is the unit vector pointing from the equilibrium position of C_α atom i to that of C_α atom j , $a = \Delta r_{ij,L}/d_{ij,0}$, and $b = \Delta r_{ij,T}/d_{ij,0}$. If the Taylor expansion in Eq. 6 is truncated at the second order, only the longitudinal component, $\Delta r_{ij,L}$, contributes, whereas the transverse component, $\Delta r_{ij,T}$, does not. Therefore, a proper modeling of transverse fluctuations between pairs of C_α atoms requires third- or higher-order terms of Eq. 6. This cannot be achieved in standard NMA, which ignores these anharmonic terms. To evaluate the effect of anharmonic terms on atomic fluctuations, I sample along the direction of the eigenvector of each mode using the exact form of ENM potential energy (Eq. 1).

To demonstrate the anharmonicity of ENM potential energy, this study has shown the ENM energy profiles for the displacements along the eigenvectors of the lowest 10 modes, which are solved for an *Escherichia coli* hppk crystal structure (using ANM $_{C_\alpha}$ with $R_c = 10$ Å) (see Fig. S1 in the Supporting Material). Significant deviation from a harmonic potential is indeed observed, accompanied by slight asymmetry between positive and negative displacements (Fig. S1).

ENM for a protein structure interacting with a crystalline environment

To model the effects of crystal packing on protein dynamics, I construct a C_α -only ENM consisting of two components: the C_α atoms of a main

protein structure (corresponding to an asymmetric unit of a protein crystal) and an environment that includes the C_α atoms of neighboring protein molecules in a crystal. To reduce system size, the environment is truncated by keeping the C_α atoms of neighboring molecules within 25 Å from the C_α atoms of the main protein structure (43). The C_α coordinates of such a truncated environment are generated using the What If webserver (<http://swift.cmbi.ru.nl/servers/html/index.html>). To reduce the computing cost, one assumes that only the C_α atoms of the main protein structure can move, whereas those of the environment are fixed (43).

The potential energy of the two-component ENM is

$$E = \frac{1}{2} \sum_{i < j} C_{ij} (d_{ij} - d_{ij,0})^2 + \frac{1}{2} f_{env} \sum_{i,l} C_{il} (d_{il} - d_{il,0})^2, \quad (7)$$

where i and j (l) are indices for C_α atoms in the main structure (environment), C_{ij} (C_{il}) is the force constant of the spring connecting C_α atoms i and j (l), d_{ij} (d_{il}) is the atomic distance between C_α atoms i and j (l), and $d_{ij,0}$ ($d_{il,0}$) is the value of d_{ij} given by the crystal structure. A parameter f_{env} within the range $[0, 1]$ is introduced to tune the strength of protein-environment interactions. $f_{env} = 0$ corresponds to the case of an isolated protein structure (see Eq. 1). $f_{env} = 1$ if one assumes equal strength of interprotein and intra-protein interactions. As shown in a recent work, a small f_{env} (≤ 0.05) should be used to accurately model protein dynamics in crystalline states (43). A small f_{env} implies significantly weaker interactions between neighboring proteins than within a protein (for details, see Hafner and Zheng (43)).

Calculation of ADPs and B factors using ANMA and NMA

Based on NMA, the covariance matrix of C_α coordinates of the main protein structure can be constructed as

$$\langle uu^T \rangle = \sum_m w_m V_m V_m^T, \quad (8)$$

where w_m gives the magnitude of MSF along the direction of eigenvector V_m of mode m , and the summation is over a subset of lowest normal modes (see below). Under the harmonic assumption of NMA, $w_m = k_B T / \lambda_m$, where k_B is the Boltzmann constant and T is temperature. In ANMA, w_m is determined by uniformly sampling the ENM potential function (see Eqs. 1 and 7) along the direction of V_m :

$$w_m = \frac{\sum_{l=-L}^L |\Delta X_{ml}|^2 e^{-E(\Delta X_{ml})/k_B T}}{\sum_{l=-L}^L e^{-E(\Delta X_{ml})/k_B T}}, \quad (9)$$

where $\Delta X_{ml} = (lA_m/L)V_m$ is the l th displacement vector of C_α atoms along the direction of V_m , the maximal magnitude A_m is chosen such that $E(\Delta X_{ml}) \geq 5k_B T$, and L is the number of sample displacements in the direction of V_m (or $-V_m$). To ensure good convergence, L is set to 20.

For an isolated protein, the potential energy function of Eq. 1 is used to sample along the directions of eigenvectors of a selected subset of normal modes, which include the lowest 5% of modes with positive eigenvalue, and nontranslation/nonrotation zero modes, if present. The nontranslation/nonrotation zero modes often result from the low connectivity of a subset of residues to the rest of the ENM, which tends to occur when a relatively small cutoff distance, R_c , of ENM is used. These zero modes cannot be properly considered in standard NMA because of their anomalous contributions to atomic fluctuations. The translation/rotation zero modes are excluded because their contributions to atomic fluctuations are anomalous for an isolated protein. For a protein interacting with its environment, the potential energy function of Eq. 7 (with $f_{env} = 0.05$) is used to sample along the directions of eigenvectors of a selected subset of normal modes, which includes the lowest 5% of modes with positive eigenvalue, and all zero modes

(including six translation/rotation modes). The translation/rotation modes can be included because their contributions to atomic fluctuations are no longer anomalous in the presence of protein-environment interactions.

The anharmonicity of mode m is assessed by the parameter $AH_m = \lambda_m w_m / k_B T (AH_m < 1$ if the anharmonicity reduces the MSF along mode m ; $AH_m > 1$ if the anharmonicity increases the MSF along mode m). It is noted that $AH_m = 1$ does not necessarily mean perfect harmonicity. AH_m values are averaged over the above selected subset of normal modes to assess their overall anharmonicity.

The i th 3×3 diagonal block of the covariance matrix $\langle uu^T \rangle$ gives the theoretical prediction of the ADP tensor for C_α atom i :

$$\langle uu^T \rangle_{ii} = \begin{bmatrix} \langle \delta x_i^2 \rangle & \langle \delta x_i \delta y_i \rangle & \langle \delta x_i \delta z_i \rangle \\ \langle \delta x_i \delta y_i \rangle & \langle \delta y_i^2 \rangle & \langle \delta y_i \delta z_i \rangle \\ \langle \delta x_i \delta z_i \rangle & \langle \delta y_i \delta z_i \rangle & \langle \delta z_i^2 \rangle \end{bmatrix} = \begin{bmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{bmatrix}, \quad (10)$$

where the diagonal elements U_{11} , U_{22} , and U_{33} give the MSF of C_α atom i along the x , y , and z directions, and the off-diagonal elements U_{12} , U_{13} , and U_{23} describe the covariance among the displacements of C_α atom i along the x , y , and z directions. Together, the six ADP elements determine a three-dimensional Gaussian distribution function that describes both the direction and magnitude of the atomic fluctuations (38). For fixed probability value, the distribution is ellipsoidal, with a directional preference along the long axis, which is given by the eigenvector of the ADP tensor with the largest eigenvalue. The anisotropy of the Gaussian distribution is defined as the ratio of the smallest to the largest eigenvalue of the ADP tensor.

The B factor is related to the trace of the ADP tensor as

$$B = 8\pi^2(U_{11} + U_{22} + U_{33})/3. \quad (11)$$

Comparison between theoretical and experimental ADPs

To fit the theoretical ADPs (Eq. 10) to experimental ADPs for a protein structure with N C_α atoms (one ADP tensor per C_α atom), the force constant parameter, C (see Eqs. 2–4), is adjusted so that the sum of $3N$ diagonal elements matches up between the theoretical and experimental ADPs. Three metrics are used to assess the similarity between two ADP tensors (represented as U and V).

Real-space correlation coefficient

The real-space correlation coefficient is calculated to evaluate the overlap integral of two three-dimensional Gaussian distributions given by U and V (38):

$$cc(U, V) = \frac{(\det U^{-1} \det V^{-1})^{1/4}}{[\det(U^{-1} + V^{-1})/8]^{1/2}}. \quad (12)$$

Based on the real-space correlation coefficient, the modified correlation coefficient (cc_{mod}) is introduced to evaluate the directional similarity of two ADPs:

$$cc_{\text{mod}}(U, V) = \frac{cc(U, V) - cc(U, V^*)}{1 - cc(U, V^*)}, \quad (13)$$

where V^* is a 3×3 tensor generated by taking the eigenvectors of U and using the eigenvalues of V , with the largest and smallest switched, to define the two ellipsoids with perfect misalignment (39). cc_{mod} is 1.0 (0) if the two ellipsoids are perfectly aligned (misaligned).

Kullback-Leibler distance

The Kullback-Leibler (KL) distance (46) evaluates the difference between the three-dimensional Gaussian distributions a and b as defined by U and V (40). The KL distance can be expressed in terms of the eigenvalues

(d_{ak} and d_{bk} , $k = 1, 2, 3$) and eigenvectors (v_{ak} and v_{bk} , $k = 1, 2, 3$) of U and V as follows:

$$D_{ab} = -\frac{3}{2} + \frac{1}{2} \sum_{k=1}^3 \ln \frac{d_{bk}}{d_{ak}} + \frac{1}{2} \sum_{k=1}^3 \sum_{l=1}^3 \frac{d_{ak}}{d_{bl}} |v_{ak}^T v_{bl}|^2. \quad (14)$$

Since the KL distance is asymmetrical ($D_{ab} \neq D_{ba}$), the arithmetic average $D_{ab} + D_{ba}/2$ was calculated (40). Note that D_{ab} diverges if the distribution b is highly anisotropic (with a near-zero eigenvalue). To avoid such divergence, this study uses $\min\{D_{ab}, D_{ba}\}$ instead of $(D_{ab} + D_{ba})/2$ as this study's KL distance metric (43).

Pearson correlations

The above two metrics evaluate only the directional similarity of two ADPs. To include the magnitude of ADPs in the comparison, this study computes the Pearson correlation (PC_{all}) between two sets of ADPs as two $6N'$ -dimensional vectors \tilde{U} and \tilde{V} (40) (N' is the number of ADPs):

$$PC = \frac{\sum_{j=1}^{6N'} (\tilde{U}_j - \langle \tilde{U} \rangle) (\tilde{V}_j - \langle \tilde{V} \rangle)}{\sqrt{\sum_{j=1}^{6N'} (\tilde{U}_j - \langle \tilde{U} \rangle)^2 \sum_{j=1}^{6N'} (\tilde{V}_j - \langle \tilde{V} \rangle)^2}}. \quad (15)$$

The Pearson correlations is also calculated for $3N'$ diagonal and $3N'$ off-diagonal ADP elements separately (PC_{diagonal} and $PC_{\text{offdiagonal}}$, respectively), as well as the Pearson correlation between theoretical and experimental B factors (PC_B) (40).

Crystallographic dataset for model evaluation

The modeling of ADPs is evaluated using a set of 83 high-resolution crystal structures previously studied in Kondrashov et al. (39). From the ANISOU records of these PDB structures, 16,852 usable ADPs are collected for those C_α atoms with occupancy of 1.0 (though all C_α atoms are included in the construction of ENM).

Following Kondrashov et al. (39), for the evaluation of four Pearson correlations, all 16,852 ADPs are used; for the evaluation of two directional metrics (cc_{mod} and KL distance), a subset of 6784 ADPs with anisotropy ≤ 0.5 are used.

Evaluation of the statistical significance of improvement in ADP modeling

The ADP similarity metrics defined above are calculated for 6784 ADPs from 83 PDB structures using both NMA and ANMA. For a given metric S , the difference between NMA and ANMA, $S_{\text{ANMA}} - S_{\text{NMA}}$, is calculated for each usable ADP (if S is cc_{mod} or KL distance) or PDB structure (if S is PC_{all} , PC_B , PC_{diagonal} , or $PC_{\text{offdiagonal}}$). To evaluate the statistical significance of this difference, its average ($\langle S_{\text{ANMA}} - S_{\text{NMA}} \rangle$) and standard deviation ($\sigma_{\delta S}$) over 6784 usable ADPs or 83 PDB structures is calculated. Then, a Z score for S is defined as

$$Z_S = \langle S_{\text{ANMA}} - S_{\text{NMA}} \rangle / (\sigma_{\delta S} / \sqrt{N_{\text{samp}}}), \quad (16)$$

where $N_{\text{samp}} = 6784$ if S is cc_{mod} or KL distance, and $N_{\text{samp}} = 83$ if S is PC_{all} , PC_B , PC_{diagonal} , or $PC_{\text{offdiagonal}}$. (For KL distance, a minus sign is added to Eq. 16 to make Z_S positive). A large Z score indicates high statistical significance for the improvement in ADP modeling from NMA to ANMA.

RESULTS

To explore how the anharmonicity of the potential energy function affects NMA-based modeling of atomic fluctuations

in protein crystal structures, this study has proposed and performed an anharmonic NMA (ANMA) in the absence and presence of interactions between a protein and its crystalline environment. This study focuses on the anharmonicity of ENM potential energy function. ANMA has been systematically compared with standard NMA based on the modeling of a large set of ADPs from 83 high-resolution protein crystal structures (39). This work now presents the findings based on the modeling results.

ANMA improves ENM-based ADP modeling

In previous ENM-based studies of atomic fluctuations in protein crystal structures (17,39–41,43,44,47,48), it was implicitly assumed that the atomic fluctuations are harmonic. Thus, the MSF along the direction of the eigenvector of a normal mode is proportional to the inverse of its eigenvalue. This is not a realistic assumption for protein dynamics at physiological temperature or crystallographic temperature (26). There are two main sources of anharmonicity in protein dynamics: the presence of multiple minima in a potential energy function and the deviation of a single-minimum potential function from a harmonic well. Many previous studies have addressed the former case using MD simulations and quasiharmonic analysis (27,29). Here, the latter case is addressed. In particular, this study explores how the anharmonicity of the ENM potential function affects ENM-based modeling of atomic fluctuations in protein crystal structures. To this end, this study proposes and evaluates the ANMA protocol (for details, see Methods): 1) I solve a subset of lowest normal modes for an isolated protein structure modeled by ENM; 2) I sample the ENM potential energy function along the direction of the eigen-

vector of each mode, and evaluate the MSF along these directions; and 3) I construct a covariance matrix to calculate the ADP tensors. As a control, ADP tensors using standard NMA (43) are calculated. The theoretical ADPs are compared with experimental ADPs using various metrics to assess the modeling quality (see Methods). To ensure the robustness of the modeling results, I consider three ENM schemes: ANM_{C α} , ANM_{atom}, and DNM (see Methods).

As an example, this work shows the results of ADP modeling using ANM_{C α} (with $R_c = 7 \text{ \AA}$) for an *E. coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (hpk) crystal structure (PDB code: 1F9Y) (Fig. 1 *a*). ANMA gives better agreement between theoretical and experimental ADPs than does NMA—the Pearson correlations for all, diagonal, and off-diagonal ADP elements and for B factors (PC_{all}, PC_{diagonal}, PC_{offdiagonal}, PC_B; see Methods) increase from 0.27, 0.18, 0.06, and 0.25 to 0.58, 0.48, 0.30, and 0.60, respectively. For the directional comparison of experimental and theoretical ADPs, this study focuses on 85 of 158 experimental ADPs with anisotropy ≤ 0.5 (41). Two metrics for directional similarity between theoretical and experimental ADPs (cc_{mod} and KL distance; see Methods) are calculated, which also indicate improvement from NMA to ANMA—the average cc_{mod} increases from 0.33 to 0.43 (Fig. 1 *c*) and the average KL distance decreases from 0.91 to 0.58 (Fig. 1 *d*). Compared with NMA, ANMA has suppressed two pronounced peaks in the theoretical B factors (Fig. 1 *b*) and KL distances (Fig. 1 *d*), which correspond to two floppy loops (residues 45–50 and 85–90) (Fig. 1 *a*). The account of anharmonicity by ANMA improves the modeling of atomic fluctuations in these dangling loops, as suggested by the higher cc_{mod} (Fig. 1 *c*) and lower KL distance (Fig. 1 *d*).

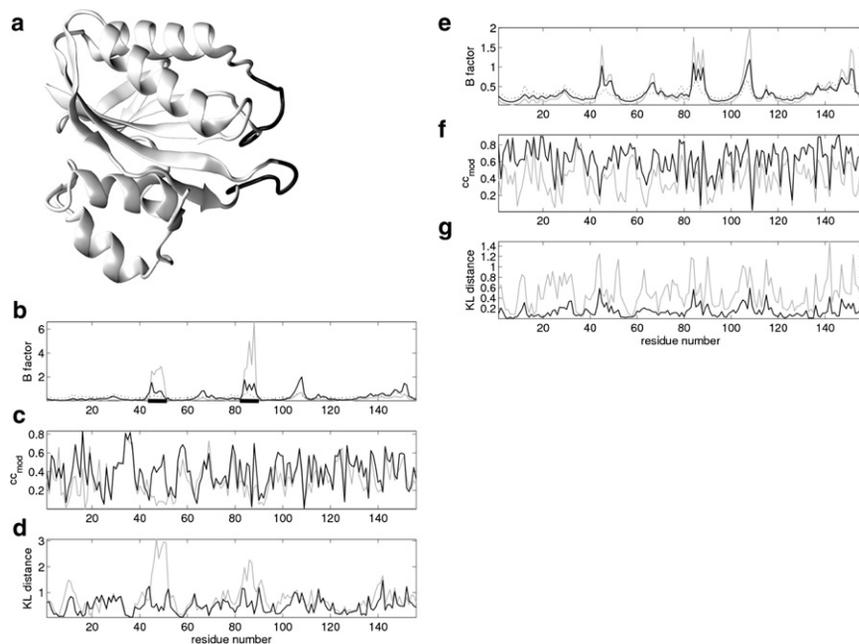


FIGURE 1 The results of ADP modeling for an *E. coli* hpk crystal structure (PDB code: 1F9Y). (a) A cartoon representation of 1F9Y with two floppy loops (residues 45–50 and 85–90) colored black. (b) B factors (rescaled by $3/8\pi^2$) from crystallography (dotted line), modeled by NMA (gray line) and by ANMA with $f_{\text{env}} = 0$ (black line) (the positions of the two loops shown in *a* are underscored). (c and d) cc_{mod} and KL distance, respectively, for NMA (gray lines) and ANMA with $f_{\text{env}} = 0$ (black lines). (e) B factors (rescaled by $3/8\pi^2$) from crystallography (dotted line), modeled by ANMA with $f_{\text{env}} = 0$ (gray line) and by ANMA with $f_{\text{env}} = 0.05$ (black line). (f and g) cc_{mod} and KL distance, respectively, for ANMA with $f_{\text{env}} = 0$ (gray lines) and ANMA with $f_{\text{env}} = 0.05$ (black lines).

The above ADP modeling was conducted for 16,852 ADPs collected from 83 protein crystal structures (39). To assess the performance of ADP modeling, this study averages four Pearson correlations (PC_{all} , PC_{B} , PC_{diagonal} , and $PC_{\text{offdiagonal}}$) over 83 structures, and two directional metrics (cc_{mod} and KL distance) over a subset of 6784 ADPs with anisotropy ≤ 0.5 . These average metrics are calculated as a function of cutoff distance for ANMA and NMA based on three ENM schemes ($ANM_{C\alpha}$, ANM_{atom} , and DNM; see Methods). The results are shown in Figs. 2–4 and Tables S1–S3. Z scores for the above metrics are calculated to assess the statistical significance of the improvement from NMA to ANMA (see Methods).

For $ANM_{C\alpha}$, significant improvement was found in ADP modeling from NMA to ANMA, especially for $R_c < 10 \text{ \AA}$ (see Table S1 in the Supporting Material and Fig. 2, *a* and *b*). For example, for $R_c = 7 \text{ \AA}$, the average Pearson correlations (PC_{all} , PC_{diagonal} , $PC_{\text{offdiagonal}}$, and PC_{B}) increase from 0.31, 0.22, 0.14, and 0.28 to 0.54, 0.39, 0.26, and 0.47; the average cc_{mod} increases from 0.45 to 0.50; and the average KL distance decreases from 0.76 to 0.48. These improvements are highly significant, as indicated by large Z scores > 10 (Table S1). For $R_c \geq 10 \text{ \AA}$, the improvement from NMA to ANMA is smaller. For example, for $R_c = 10 \text{ \AA}$, the average Pearson correlations (PC_{all} , PC_{diagonal} , $PC_{\text{offdiagonal}}$, and PC_{B}) increase from 0.43, 0.36, 0.24, and 0.43 to 0.48, 0.40, 0.28, and 0.47; the average cc_{mod} increases from 0.50 to 0.52; and the average KL distance decreases from 0.48 to 0.43. This smaller improvement remains statistically significant, as the Z scores for all metrics are > 5 (Table S1).

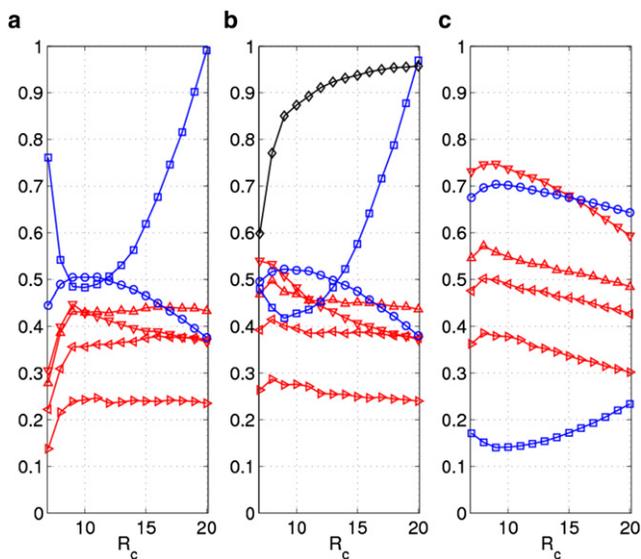


FIGURE 2 The results of ADP modeling averaged over 83 protein crystal structures (39) using $ANM_{C\alpha}$ and (a) NMA, (b) ANMA with $f_{\text{env}} = 0$, and (c) ANMA with $f_{\text{env}} = 0.05$. Shown here are Pearson correlations of diagonal (\triangleleft), off-diagonal (\triangleright), all elements (∇) of ADPs and B factors (\triangle), and directional metrics including cc_{mod} (\circ) and KL distance (\square) as a function of cutoff distance R_c . Also shown in (b) is the average anharmonicity AH (\diamond).

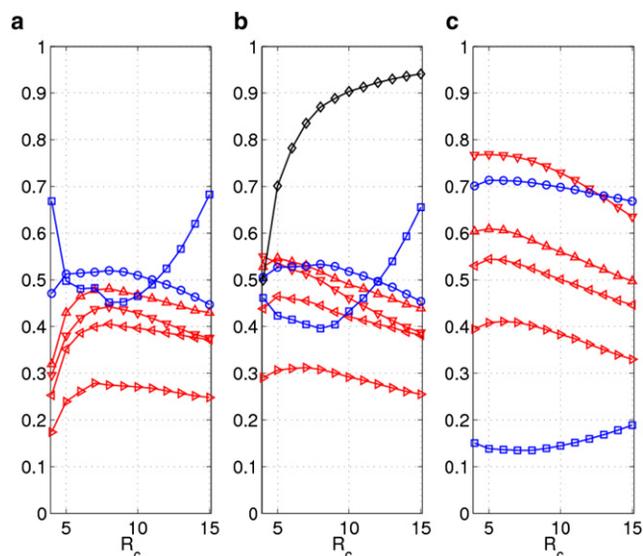


FIGURE 3 The results of ADPs modeling averaged over 83 protein crystal structures (39) using ANM_{atom} and (a) NMA; (b) ANMA with $f_{\text{env}} = 0$; (c) ANMA with $f_{\text{env}} = 0.05$. Shown here are Pearson correlations of diagonal (\triangleleft), off-diagonal (\triangleright), and all elements (∇) of ADPs, and B factors (\triangle), and directional metrics including cc_{mod} (\circ) and KL distance (\square) as a function of cutoff distance R_{atom} . Also shown in *b* is the average anharmonicity, AH (\diamond).

For ANM_{atom} and DNM, significant improvement was also found in ADP modeling from NMA to ANMA, especially at the low cutoff distance $R_{\text{atom}} < 8 \text{ \AA}$ (for details, see Table S2, Table S3, and Figs. 3, *a* and *b*, and 4, *a* and *b*). Therefore, the finding that ANMA improves ENM-based ADP modeling is robust and does not rely on details of ENM schemes.

An ENM with low cutoff distance is prone to the so-called tip effect, which involves some pathological modes describing anomalously large motions of some protruding structural elements (such as a dangling loop) while the rest of the protein remains essentially static (45). The large local motions in these modes dramatically reduce the accuracy of harmonic approximation to the ENM potential function (see Eq. 6), therefore making those modes highly anharmonic. Indeed, for $ANM_{C\alpha}$ (ANM_{atom} and DNM) the average anharmonicity AH (see Methods) decreases sharply as R_c (R_{atom}) drops below 10 \AA (8 \AA) (see Figs. 2 *b*, 3 *b*, and 4 *b*). Therefore, it is not surprising that the proper account of anharmonicity by ANMA significantly improves ADP modeling at low cutoff distance. Thus, an accurate modeling of protein atomic fluctuations demands a proper account of the anharmonicity of the ENM potential function.

Account of protein-environment interactions further improves ADP modeling by ANMA

Previous studies have shown the significant effect of crystalline environment on protein dynamics in crystalline states (41,43,47,48). In particular, the incorporation of protein-environment interactions allows proper modeling of the

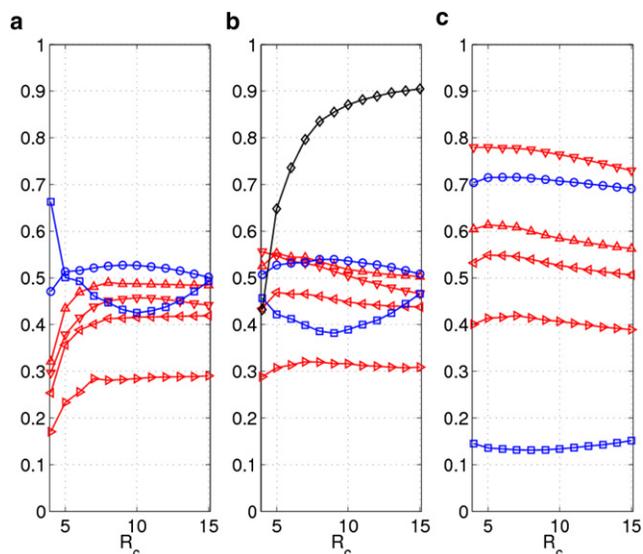


FIGURE 4 The results of ADPs modeling averaged over 83 protein crystal structures (39) using DNM and (a) NMA; (b) ANMA with $f_{env} = 0$; (c) ANMA with $f_{env} = 0.05$. Shown here are Pearson correlations of diagonal (◁), off-diagonal (▷), all elements (▽) of ADPs and B factors (△), and directional metrics including cc_{mod} (○) and KL distance (□) as a function of cutoff distance R_{atom} . Also shown in *b* is the average anharmonicity, AH (◇).

contributions of rigid-body motions of an entire protein to ADPs (41,43,48), which cannot be easily attained for an isolated protein structure. Therefore, it will be interesting to incorporate protein-environment interactions in the ANMA formulation in this study. To this end, a two-component ENM comprised of a protein structure and its fixed crystalline environment has been constructed (see Methods). I have recently found the interactions between a protein and its crystalline environment to be much weaker than the intraprotein interactions (43). ANMA was performed as follows to account for the protein-environment interactions: first, solve a subset of lowest normal modes for an isolated protein structure; then sample the ENM potential energy function (with the protein-environment interactions included; see Eq. 7) along the direction of the eigenvector of each mode; then calculate the ADPs based on the sampling results (see Methods).

As an example, this study shows and compares the results of ADP modeling by ANMA in the absence and presence of protein-environment interactions (using $ANM_{C\alpha}$ with $R_c = 7$ Å) for an *E. Coli* hppk crystal structure (PDB code: 1F9Y) (Fig. 1 a). Large improvement in ADP modeling is found in terms of the four Pearson correlations— PC_{all} , $PC_{diagonal}$, $PC_{offdiagonal}$, and PC_B change from 0.58, 0.48, 0.30, and 0.60 to 0.75, 0.56, 0.39, and 0.65. Furthermore, improvement is found in terms of the directional similarity in ADPs—the average cc_{mod} increases from 0.43 to 0.66 (Fig. 1 f), and the average KL distance decreases from 0.58 to 0.18 (Fig. 1 g).

The results of ADP modeling in the presence of protein-environment interactions for 83 protein crystal structures

(39) are shown in Figs. 2 c, 3 c, and 4 c, and Table S1, Table S2, and Table S3, which are compared with the results for isolated protein structures (see Figs. 2 b, 3 b, and 4 b, and Table S1, Table S2, and Table S3). The findings are summarized as follows.

For $ANM_{C\alpha}$, a significant improvement was found in modeling of the direction of ADP tensors with the addition of protein-environment interactions (Fig. 2, b and c)—for example, for $R_c = 10$ Å, the average cc_{mod} increases from 0.52 to 0.70 and the average KL distance decreases from 0.43 to 0.14. Furthermore, there is large improvement in the modeling of both magnitude and direction of ADP tensors as assessed by the four Pearson correlations.

For ANM_{atom} and DNM, a significant improvement was also found in the modeling of ADP tensors (see Figs. 3, b and c, and 4, b and c). Therefore, this finding is robust and does not rely on details of ENM schemes.

Optimal cutoff distances of ENM for ADP modeling by ANMA

In previous studies, it was found that the optimal fitting of B-factors by $ANM_{C\alpha}$ is attained at a high cutoff distance $R_c = 15$ – 24 Å (49), which is beyond the range of C_{α} – C_{α} distances between contacting residues (4.4–12.8 Å (50)). Some recent studies have addressed this inconsistency in ENM parameterization (43,51). To further address this issue, this study evaluates the quality of ADP modeling by NMA and ANMA as a function of cutoff distance using three ENM schemes (R_c for $ANM_{C\alpha}$, R_{atom} for ANM_{atom} , and DNM).

The R_c dependence for $ANM_{C\alpha}$ is summarized in Fig. 2. For NMA (Fig. 2 a), the R_c dependence of various ADP comparison metrics diverges: $PC_{diagonal}$, $PC_{offdiagonal}$, and PC_B saturate as R_c increases from 9 to 20 Å, whereas PC_{all} peaks at $R_c \sim 9$ Å; cc_{mod} peaks at $R_c \sim 10$ Å; and the KL distance is minimal at $R_c \sim 10$ Å. Such divergence was also found in a previous study (43), which makes it hard to parameterize ENM by optimizing the fitting of ADPs.

For ANMA (Fig. 2 b), the divergence of the R_c dependence of various ADP comparison metrics is reduced: $PC_{diagonal}$, $PC_{offdiagonal}$, PC_B and PC_{all} peak at $R_c \sim 8$ Å, cc_{mod} peaks at $R_c \sim 9$ Å, and the KL distance is minimal at $R_c \sim 9$ Å. Thus, a consensus-based optimization of various ADP comparison metrics points to an optimal R_c of ~ 8 – 9 Å, which falls well within the range of C_{α} – C_{α} distances between contacting residues (4.4–12.8 Å (50)). Note that the addition of protein-environment interactions does not change the optimal R_c values (Fig. 2 c).

The R_{atom} dependence for ANM_{atom} and DNM is summarized in Figs. 3 and 4, respectively. For ANMA based on ANM_{atom} (Fig. 3 b), the R_c dependence differs among various ADP comparison metrics: $PC_{diagonal}$, PC_{all} and PC_B peak at $R_{atom} \sim 5$ Å, $PC_{offdiagonal}$ peaks at $R_{atom} \sim 7$ Å, whereas cc_{mod} peaks at $R_{atom} \sim 8$ Å, and the KL distance is minimal at $R_{atom} \sim 8$ Å. The optimal R_{atom} values are

only slightly shifted with the addition of protein-environment interactions (Fig. 3 c).

For ANMA based on DNM, the results are similar. For example, cc_{mod} peaks at $R_{\text{atom}} \sim 9 \text{ \AA}$, and the KL distance is minimal at $R_{\text{atom}} \sim 9 \text{ \AA}$ (Fig. 4 b), which agrees with the results of Riccardi et al. (41). The optimal R_{atom} values are only slightly shifted with the addition of protein-environment interactions (Fig. 4 c).

This work's finding that the optimal modeling of the direction of ADPs by ANM_{atom} (DNM) is at $R_{\text{atom}} \sim 8 \text{ \AA}$ (9 \AA) is surprising, because this heavy-atom distance is beyond the range of van der Waals interactions ($<4 \text{ \AA}$ as estimated based on van der Waals radii (52)) or hydrogen-bond interactions ($<4.5 \text{ \AA}$ as defined in the CONGEN program (53)).

Statistical distribution of atomic distances agrees with optimal cutoff distances

To rationalize the finding of an optimal cutoff distance of $\sim 10 \text{ \AA}$ for $\text{ANM}_{C_{\alpha}}$, this work analyzes the statistical distribution, $\rho(r)$, of distances between the C_{α} atoms of nonbonded residues of 83 protein crystal structures (39). $\rho(r)$ is given by $f(r)/4\pi r^2$, where $f(r)$ is the percentage of all C_{α} - C_{α} distances within the range $(r - 0.5, r + 0.5)$, and r is sampled at integer values (in \AA). It is interesting that this study found that $\rho(r)$ has a primary peak around 5–6 \AA and a secondary peak at 10 \AA (Fig. 5 a). These two peaks correspond to two minima of a statistical potential function based on C_{α} - C_{α} distances. The coincidence of the secondary peak of $\rho(r)$ and the optimal R_c for $\text{ANM}_{C_{\alpha}}$ hints at the presence of effective interactions between residues separated by a C_{α} - C_{α} distance of $\leq 10 \text{ \AA}$, which is important to protein dynamics.

In a similar way, to rationalize the finding of an optimal atomic distance of $\sim 8 \text{ \AA}$ (9 \AA) for ANM_{atom} (DNM), this work analyzes the statistical distribution of distances between the heavy atoms of nonbonded residues of 83 protein crystal structures (39). The distribution shows a broad peak ranging from 5 to 10 \AA (Fig. 5 b), which suggests the presence of statistical interactions between heavy atoms from different residues separated by $\leq 10 \text{ \AA}$. This value is beyond the range of van der Waals or hydrogen-bond interactions. The idea that long-range electrostatic interactions may be the cause is ruled out, because the same distribution is obtained after charged residues are excluded from the statistical analysis. This observation roughly agrees with this study's finding of $R_{\text{atom}} \sim 8 \text{ \AA}$ (9 \AA) as the optimal atomic distance for ANM_{atom} (DNM). Taken together, these findings support the presence of indirect interactions between atoms separated by $\leq 10 \text{ \AA}$ and the importance of these interactions to protein dynamics. Such effective interactions may arise as a result of averaging over some hidden variables (such as multiple substates of a protein native state).

DISCUSSION AND CONCLUSION

To account for the anharmonic aspects of protein dynamics, this study has used ANMA to model the ADPs from a set of

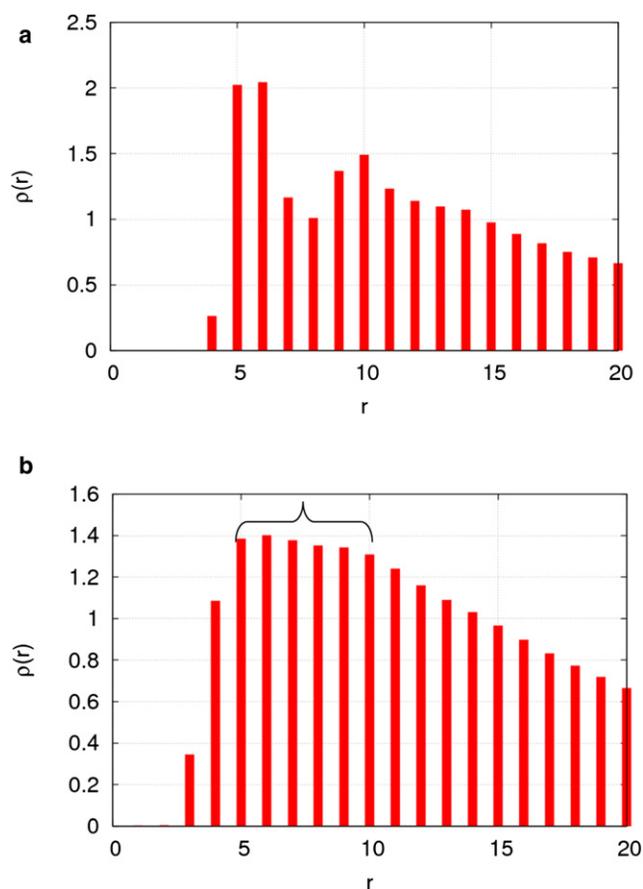


FIGURE 5 The density functions for the statistical distributions of distances between (a) C_{α} atoms and (b) heavy atoms of nonbonded residues in 83 protein crystal structures (39). $\rho(r)$ is rescaled by the density function ρ_0 for a uniform distribution. The same statistical distributions (not shown here) are obtained from a much larger set of >2000 protein crystal structures (cullpdb_pc30_res1.6_R0.25_d081223_chains2039, downloaded from <http://dunbrack.fccc.edu/PISCES.php>).

83 high-resolution protein crystal structures. Significant improvement was found in the modeling of ADPs by ANMA compared with standard NMA. Further improvement in the modeling of ADPs is attained if the interactions between a protein and its crystalline environment are taken into account. In addition, this work has determined the optimal cutoff distances for ENM-based ADP modeling, which agree well with the peaks of the statistical distributions of distances between C_{α} atoms or heavy atoms deduced from a large set of protein crystal structures.

This work applied ANMA to explore the anharmonicity of the ENM potential function (Eq. 1). It is straightforward to apply ANMA to any other anharmonic potential function (such as a Lennard-Jones potential). In fact, I applied ANMA to the generalized potential function $E_n(d_{ij}) = 0.5C_{ij} d_{ij,0}^2/n^2(1 - d_{ij,0}^n/d_{ij}^n)^2$, which is reduced to the ENM potential in Eq. 1 when $n = -1$ and to a Lennard-Jones potential when $n = 6$. The application of ANMA to this generalized potential function, however,

does not lead to further improvement of ADP modeling (data not shown).

In this study, the anharmonicity, AH, is found to be < 1 (see Figs. 2 b, 3 b, and 4 b). Therefore, anharmonicity of the ENM potential function results in the stiffening of a normal mode (or reduction of the MSF along the direction of its eigenvector). This finding complements the early findings of a softening of normal modes due to the presence of multiple potential energy minima (27,29).

By using a fixed set of eigenvectors solved by standard NMA, this study focuses on the anharmonic effect on the magnitude of atomic fluctuations. It will be interesting to explore in future studies the anharmonic effect on the direction of atomic fluctuations.

This study offers a new avenue to the accurate parameterization of ENM beyond the limit of harmonic approximation. The correlation of optimal ENM parameters with statistical distributions of atomic distances may also offer a systematic way to develop ENM parameters based on residue type. A more accurate ENM is essential to the development and refinement of ENM-based techniques that probe protein dynamics of functional relevance (54–57).

SUPPORTING MATERIAL

One figure and three tables are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(10\)00359-0](http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)00359-0).

The author thanks the American Heart Association (0835292N) for funding support.

REFERENCES

- Henzler-Wildman, K., and D. Kern. 2007. Dynamic personalities of proteins. *Nature*. 450:964–972.
- Boehr, D. D., R. Nussinov, and P. E. Wright. 2009. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5:789–796.
- Bahar, I., C. Chennubhotla, and D. Tobi. 2007. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* 17:633–640.
- Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9:646–652.
- Klepeis, J. L., K. Lindorff-Larsen, ..., D. E. Shaw. 2009. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* 19:120–127.
- Brooks, B. R., and M. Karplus. 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA*. 80:6571–6575.
- Brooks, B. R., D. Janezic, and M. Karplus. 2004. Harmonic analysis of large systems. I. Methodology. *J. Comput. Chem.* 16:1522–1542.
- Go, N., T. Noguti, and T. Nishikawa. 1983. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA*. 80:3696–3700.
- Levitt, M., C. Sander, and P. S. Stern. 1985. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* 181:423–447.
- Cui, Q., and I. Bahar. 2005. Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems. CRC Press, Boca Raton, FL.
- Tozzini, V. 2005. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* 15:144–150.
- Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.
- Tama, F., and Y. H. Sanejouand. 2001. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14:1–6.
- Atilgan, A. R., S. R. Durell, ..., I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.
- Bahar, I., A. R. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* 2:173–181.
- Haliloglu, T., I. Bahar, and B. Erman. 1997. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79:3090–3093.
- Kondrashov, D. A., Q. Cui, and G. N. Phillips, Jr. 2006. Optimization and evaluation of a coarse-grained model of protein motion using x-ray crystal data. *Biophys. J.* 91:2760–2767.
- Yang, L., G. Song, and R. L. Jernigan. 2009. Protein elastic network models and the ranges of cooperativity. *Proc. Natl. Acad. Sci. USA*. 106:12347–12352.
- Krebs, W. G., V. Alexandrov, ..., M. Gerstein. 2002. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins*. 48:682–695.
- Bahar, I., and A. J. Rader. 2005. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* 15:586–592.
- Tama, F., and C. L. Brooks. 2006. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu. Rev. Biophys. Biomol. Struct.* 35:115–133.
- Rader, A. J., D. H. Vlad, and I. Bahar. 2005. Maturation dynamics of bacteriophage HK97 capsid. *Structure*. 13:413–421.
- Wang, Y., A. J. Rader, ..., R. L. Jernigan. 2004. Global ribosome motions revealed with elastic network model. *J. Struct. Biol.* 147:302–314.
- Lezon, T. R., A. Sali, and I. Bahar. 2009. Global motions of the nuclear pore complex: insights from elastic network models. *PLOS Comput. Biol.* 5:e1000496.
- Ma, J. 2005. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*. 13: 373–380.
- Hayward, J. A., and J. C. Smith. 2002. Temperature dependence of protein dynamics: computer simulation analysis of neutron scattering properties. *Biophys. J.* 82:1216–1225.
- Hayward, S., A. Kitao, and N. Go. 1994. Harmonic and anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis. *Protein Sci.* 3:936–943.
- Levy, R. M., A. R. Srinivasan, ..., J. A. McCammon. 1984. Quasi-harmonic method for studying very low frequency modes in proteins. *Biopolymers*. 23:1099–1112.
- Hayward, S., A. Kitao, and N. Go. 1995. Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis. *Proteins*. 23:177–186.
- Hayward, S., and N. Go. 1995. Collective variable description of native protein dynamics. *Annu. Rev. Phys. Chem.* 46:223–250.
- Kitao, A., and N. Go. 2004. Conformational dynamics of polypeptides and proteins in the dihedral angle space and in the cartesian coordinate space: Normal mode analysis of deca-alanine. *J. Comput. Chem.* 12:359–368.
- Zhang, Z., Y. Shi, and H. Liu. 2003. Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophys. J.* 84:3583–3593.
- Delarue, M., and P. Dumas. 2004. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl. Acad. Sci. USA*. 101:6957–6962.

34. Tama, F., O. Miyashita, and C. L. Brooks, 3rd. 2004. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.* 337:985–999.
35. Song, G., and R. L. Jernigan. 2007. vGNM: a better model for understanding the dynamics of proteins in crystals. *J. Mol. Biol.* 369: 880–893.
36. Willis, B. T. M., and A. W. Pryor. 1975. *Thermal Vibrations in Crystallography*. Cambridge University Press, London.
37. Scheringer, C. 1977. On the interpretation of anisotropic temperature factors. III. Anharmonic motions. *Acta Crystallogr. A.* 33:879–884.
38. Merritt, E. A. 1999. Comparing anisotropic displacement parameters in protein structures. *Acta Crystallogr. D Biol. Crystallogr.* D55: 1997–2004.
39. Kondrashov, D. A., A. W. Van Wynsberghe, ..., G. N. Phillips, Jr. 2007. Protein structural variation in computational models and crystallographic data. *Structure.* 15:169–177.
40. Eyal, E., C. Chennubhotla, ..., I. Bahar. 2007. Anisotropic fluctuations of amino acids in protein structures: insights from x-ray crystallography and elastic network models. *Bioinformatics.* 23:i175–i184.
41. Riccardi, D., Q. Cui, and G. N. Phillips, Jr. 2009. Application of elastic network models to proteins in the crystalline state. *Biophys. J.* 96: 464–475.
42. Lu, M., and J. Ma. 2008. A minimalist network model for coarse-grained normal mode analysis and its application to biomolecular x-ray crystallography. *Proc. Natl. Acad. Sci. USA.* 105:15358–15363.
43. Hafner, J. P., and W. Zheng. 2010. Optimal modeling of atomic fluctuations in protein crystal structures for weak crystal contact interactions. *J. Chem. Phys.* 132:014111.
44. Yang, L., G. Song, and R. L. Jernigan. 2009. Comparisons of experimental and computed protein anisotropic temperature factors. *Proteins.* 76:164–175.
45. Lu, M., B. Poon, and J. Ma. 2006. A new method for coarse-grained elastic normal-mode analysis. *J. Chem. Theory Comput.* 2:464–471.
46. Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. *Ann. Math. Stat.* 22:79–86.
47. Kundu, S., J. S. Melton, ..., G. N. Phillips, Jr. 2002. Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophys. J.* 83:723–732.
48. Hinsen, K. 2008. Structural flexibility in proteins: impact of the crystal environment. *Bioinformatics.* 24:521–528.
49. Eyal, E., L. W. Yang, and I. Bahar. 2006. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics.* 22:2619–2627.
50. Cieplak, M., and T. X. Hoang. 2003. Universality classes in folding times of proteins. *Biophys. J.* 84:475–488.
51. Zheng, W. 2008. A unification of the elastic network model and the Gaussian network model for optimal description of protein conformational motions and fluctuations. *Biophys. J.* 94:3853–3857.
52. Bondi, A. 1964. van der Waals volumes and radii. *J. Phys. Chem.* 68:441–451.
53. Bruccoleri, R. E., E. Haber, and J. Novotný. 1988. Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature.* 335:564–568.
54. Zheng, W. 2009. Normal-mode-based modeling of allosteric couplings that underlie cyclic conformational transition in F(1) ATPase. *Proteins.* 76:747–762.
55. Zheng, W., B. R. Brooks, and G. Hummer. 2007. Protein conformational transitions explored by mixed elastic network models. *Proteins.* 69:43–57.
56. Zheng, W., and M. Tekpinar. 2009. Large-scale evaluation of dynamically important residues in proteins predicted by the perturbation analysis of a coarse-grained elastic model. *BMC Struct. Biol.* 9:45.
57. Zheng, W., and D. Thirumalai. 2009. Coupling between normal modes drives protein conformational dynamics: illustrations using allosteric transitions in myosin II. *Biophys. J.* 96:2128–2137.