Published in final edited form as:
Proteins. 2009 July ; 76(1): 164–175. doi:10.1002/prot.22328.

Comparisons of Experimental and Computed Protein Anisotropic Temperature Factors

Lei Yang[1,2,4,*], **Guang Song**[1,3,4], and **Robert L. Jernigan**[1,2,4]

[1]Program of Bioinformatics and Computational Biology Iowa State University, Ames, IA 50011, USA

[2]Department of Biochemistry, Biophysics and Molecular Biology Iowa State University, Ames, IA 50011, USA

[3]Department of Computer Science Iowa State University, Ames, IA 50011, USA

[4]L. H. Baker Center for Bioinformatics and Biological Statistics Iowa State University, Ames, IA 50011, USA

Abstract

Because of its appealing simplicity, the anisotropic network model (ANM) has been widely accepted and applied to study many molecular motion problems: such as ribosome motions, the molecular mechanisms of GroEL-GroES function, allosteric changes in hemoglobin, motor-protein motions, and conformational changes in general. However, the validity of the ANM has not been closely examined. In this work, we use ANM to predict the anisotropic temperature factors of proteins obtained from X-ray and NMR data. The rich, directional anisotropic temperature factor data available for hundreds of proteins in the Protein Data Bank (PDB) are used as validation data to closely test the ANM model. The significance of this work is that it presents a timely, important evaluation of the model, shows the extent of its accuracy in reproducing experimental anisotropic temperature factors, and suggests ways to improve the model. An improved model will help us better understand the internal dynamics of proteins, which in turn can greatly expand the usefulness of the models, which has already been demonstrated in many applications.

Keywords

anisotropic displacement parameters (ADPs); *anisotropic network model (ANM)*; *anisotropic temperature factors*; *Gaussian network model (GNM)*; *protein dynamics*

INTRODUCTION

Functional proteins are not static structures and most of their functions are generally realized through protein motions. It is of great interest to know how these bio-machines work. Understanding the underlying detailed mechanisms can have a broad practical impact.

One of the most intuitive approaches for the study of molecular motions is molecular dynamics (MD)1›2. By using a force field to approximate the atomic interactions of a given protein, MD can compute the time-dependent behavior of the molecular system and provide much detail about the atomic fluctuations and conformational changes of the molecular

*Correspondence to: Lei Yang, leiyang@iastate.edu..

system being studied. It is an important tool and has been used extensively in protein structure determination and refinement, simulating (un)folding pathways, dynamics and fluctuations of folded proteins, etc. The major challenge in applying MD to study the motions of large macromolecules is the limits of computational power. In general, there is a huge gap between the feasible simulation time duration and the time required for a real biological process to take place, e.g., the folding of a moderately large protein. Moreover, MD is governed by the interactions among the individual atoms and does not explicitly consider the overall concertedness in motion which is commonly seen in the dynamics of folded proteins.

Atomic normal mode analysis (NMA) is an ideal alternative method for the study of the collective motions of proteins. Basically, NMA represents simple harmonic oscillations about a local energy minimum. To apply NMA, an energy minimization has to be first applied to the input structure. The new, energy-locally-minimized structure may make significant changes from the original structure. After the minimization, the second derivative of the potential energy, the Hessian matrix, has to be calculated and then diagonalized. But there are problems with NMA too, especially with large systems. The necessary initial energy minimization process not only requires time and memory but also can distort the input structure significantly, which casts doubt on the validity of the analysis or the structure. In addition, the diagonalization of the Hessian matrix can become prohibitive as the size of the system increases.

Therefore, a more efficient method was needed in order to study the collective motions of larger systems. Tirion3 showed that a single-parameter Hookean potential for all the pairwise interactions between atoms, without the energy minimization step, is able to produce similar low frequency modes to those from the original NMA. This was a big step forward since it allowed the direct analysis of crystal coordinates. Bahar et al.4‚5 and Hinsen6 took the simplification one important step further. They demonstrated that a single parameter harmonic potential together with a simplified protein model that represents each residue by a point mass was able to produce the correct low frequency normal modes and predict reasonably well the equilibrium isotropic fluctuations of several proteins. Such models are referred to as elastic network models (ENM). Specifically, the ENM for isotropic fluctuations is usually called the Gaussian network model (GNM)5, where only the magnitudes of the fluctuations are computed. Its anisotropic counterpart, where the directions of the collective motions are examined, is called the anisotropic network model (ANM)7.

Because of its appealing simplicity and efficiency, ANM has been widely accepted and applied to study many motion problems: such as ribosome motions8, the molecular mechanisms of the GroEL-GroES function9, allosteric changes in hemoglobin10, motor-protein motions11, and conformational changes in general12¯14.

However, the validity of ANM has not been sufficiently examined. In reproducing the isotropic B-factors, it had been noticed that ANM actually performs slightly worse than GNM15, which raised a warning signal. ANM was also used to interpret conformational changes for some proteins12, but the data about the conformational changes alone was insufficient to fully verify the model.

In the present work, we use ANM to predict the anisotropic temperature factors of proteins. The dataset containing hundreds of proteins with directional anisotropic temperature factors can be used as validation data to closely test the ANM model. The significance of this work is that it presents a timely, important evaluation of the model and shows how accurately the experimental anisotropic temperature factors can be reproduced. It also draws attention to

the need for an improved model to help us better understand the internal dynamics of proteins and expand the usefulness of the model, which has already been seen in many applications.

Anisotropic B-factors, or anisotropic displacement parameters (ADPs), have become available recently thanks to improvements in crystallographic data collection techniques that make the determination of atomic or near atomic resolution structures (resolution better than 1.2 Å) available. In the PDB file, these are denoted with ANISOU, followed by six numerical values that are the elements of a symmetric tensor, see PDB data format for details16 (http://www.pdb.org). As of December 1997, there were only 10 protein structures in the PDB with such entries17. By now, however, there are hundreds of protein structures with ANISOU entries. Some recent works have shown the usefulness of normal mode-based methods for predicting18,19 and refinement of anisotropic thermal motions in X-ray structures20,21.

Besides the high-resolution X-ray structures, NMR ensembles provide another good resource of structural and dynamic information. Recent studies have shown that the dynamics from NMR ensembles is less tainted by the surroundings and agrees better with computational results than that from X-ray data22,23, the latter of which may be strongly affected by crystal packing. Here we thus include in our study a dataset containing hundreds of NMR ensembles as well.

METHODS

X-ray dataset

We choose to include in our dataset all protein crystals with atomic or near-atomic resolution (resolution equal or better than 1.2 Å) currently available in the Protein Data Bank (PDB) that have anisotropic temperature factors, or ANISOU entries. There were 190 such structures in our dataset after removing the structures with more than 50% sequence similarity.

NMR dataset

Based on our previous study of NMR ensembles (Yang, Song and Jernigan, unpublished), we select NMR ensembles whose conformers are representative and sufficient in covering the conformational space. Technically, for each ensemble, we check the correlations between the first three principal components (PCs) calculated from all the conformers and those from a reduced number of conformers (half as many, randomly chosen), and only keep the ensembles that have high enough correlations for all the three PCs (PC1: > 0.9, PC2: > 0.8, PC3: > 0.7). This is to ensure the quality of the NMR ensemble, i.e., so that the models in the ensemble provide a good representation of the conformational space. We also set the criteria that the number of conformers in each ensemble is no less than 20 and the protein size is no less than 50. Finally, after removing the structures with more than 50% sequence similarity we obtain 436 ensembles to form our NMR dataset.

Isotropic and anisotropic B-factors

X-ray diffraction data of a protein crystal usually provide information about protein dynamics in the form of isotropic temperature factors B_i, which relate to the mean-square fluctuation $\langle \Delta R_i \Delta R_i \rangle$ of atom i from its average coordinate by:

$$B_i = \frac{8\pi^2}{2} \langle \Delta R_i \Delta R_i \rangle \quad (1)$$

The B_i's, one for each non-hydrogen heavy atom, is determined by fitting the X-ray diffraction data during the structural determination and refinement process. The fluctuation of atoms, as we know, is generally not isotropic. A more accurate description of the fluctuations is to use the anisotropic B-factors, or anisotropic displacement parameters (ADPs). Anisotropic B-factors B^{aniso} are represented as a 3×3 symmetric tensor U to represent both the magnitude and the directionality of the fluctuations, i.e.,

$$B^{anisou}=\begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{21} & U_{22} & U_{23} \\ U_{31} & U_{32} & U_{33} \end{pmatrix} \tag{2}$$

In essence, these describe the probability density distribution for the nuclear positions using a 3-dimensional Gaussian function24. For a fixed probability value, the distribution is ellipsoidal and has a directional preference. The more deformed the shape is from a sphere, the more anisotropic is the fluctuation. We will measure this using a term called anisotropy, to be defined later.

Similarly to the isotropic B-factors, B^{aniso} relates to the fluctuation $<\Delta R_i \Delta R_i>$ of atom i as:

$$B^{anisou}=8\pi^2 \langle \Delta R_i \Delta R_i \rangle \tag{3}$$

where

$$\langle \Delta R_i \Delta R_i \rangle = \begin{pmatrix} \Delta x_i^2 & \Delta x_i \Delta y_i & \Delta x_i \Delta z_i \\ \Delta x_i \Delta y_i & \Delta y_i^2 & \Delta y_i \Delta z_i \\ \Delta x_i \Delta z_i & \Delta y_i \Delta z_i & \Delta z_i^2 \end{pmatrix} \tag{4}$$

From the anisotropic B-factors, we can obtain the corresponding isotropic B-factors, since they are related by:

$$B_i=\frac{1}{3} trace\left(B_i^{anisou}\right) \tag{5}$$

For NMR ensembles, the pseudo anisotropic B-factors are calculated by averaging the residue fluctuations between all conformer pairs in the ensemble, using Eq. (3).

GNM and ANM

Given a protein structure, GNM4 simplifies the system by modeling it with its alpha carbons only and attaching springs with uniform constants to all contacting alpha carbon pairs. Alpha carbon pairs are considered to be in contact when their separation distance is smaller than a preset cutoff distance, usually 7 to 8 Å. All springs are set at equilibrium for the input structure. One advantage of this approach is that the fluctuations of each carbon around its equilibrium position and their cross-correlations can be expressed in analytical forms. To determine the atomic fluctuations, we first write down the Kirchhoff matrix based on the contact information,

$$\Gamma = \begin{cases} -1 & if\, i \neq j\, and\, r_{ij} \leq r_c \\ 0 & if\, i \neq j\, and\, r_{ij} > r_c \\ -\Sigma_{i,j\neq i}\Gamma_{ij} & if\, i = j \end{cases} \quad (6)$$

where r_{ij} is the distance between atoms i and j, and r_c is the cutoff distance. The mean square fluctuations of each atom and the theoretical B-factors can be conveniently expressed as:

$$\langle \Delta R_i \Delta R_i \rangle = \frac{3k_B T}{\gamma} \left[\Gamma^{-1} \right]_{ii} \quad (7)$$

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

where γ is the spring constant.

In ANM7, the counterpart of the $N \times N$ Kirchhoff matrix Γis a $3N \times 3N$ Hessian matrix H (here N is the number of alpha carbons in the structure). As a result, the inverse of H contains $N \times N$ super-elements, whereas the ii^{th} super-element of H^{-1}, a 3×3 matrix, describes the self correlations between the components of ΔR_i, i.e.,

$$\langle \Delta R_i \Delta R_i \rangle = \frac{3k_B T}{\gamma} H_{ii}^{-1} \quad (9)$$

The coarse-grained alpha-carbon model is normally used for both ANM and GNM. In this work, we set the cutoff distance to be 13 Å for ANM7 and 7.3 Å for GNM15.

Now it is straightforward to extend the method to the all atom model, even though there might not be much gain with the increased complexity25. It is also easy to treat the backbone contacts, which are covalent bonds, differently by assigning them a larger spring constant. Though it has been shown that this has little effect in reproducing isotropic B-factors4, it is possible that it might give a more pronounced effect when using ANM to produce anisotropic B-factors. For the NMR ensembles, the GNM/ANM is applied to the reference structure which is chosen as the one closest to the average among all the conformers in the ensemble23.

Calculating anisotropic B-factors from ANM

From Eq. (9), it is straightforward to obtain theoretical anisotropic B-factors B^{theo}_i by:

$$B_i^{theo} = 8\pi^2 \langle \Delta R_i \Delta R_i \rangle = 8\pi^2 \frac{3k_B T}{\gamma} H_{ii}^{-1} \quad (10)$$

The single parameter γ will serve as a scaling factor.

Comparing theoretical anisotropic B-factors with experimental data

Isotropic B-factors are scalars. The most commonly used method for comparing experimental and calculated isotropic B-factors is the correlation between these two arrays. However, anisotropic B-factors are tensors. The comparison of tensors is more complex. A naive comparison of two tensors by converting them to arrays and then calculating their correlation is not appropriate, since the elements of the tensor are not independent18. Instead, the following approach is used. Each tensor represents a 3-dimensional distribution, which can be visualized as an ellipsoid. Therefore, comparing two tensors can be done by comparing the two corresponding ellipsoids. We want to compare their size (or magnitude), their shape, and their orientation. To do this, we first diagonalize the tensors. The magnitude and shape are represented by the eigenvalues, while the directional preferences of the fluctuations are captured by the eigenvectors. The three eigenvectors of a tensor represent an orthonormal frame, and the orientations of two ellipsoids can be compared by measuring how the two corresponding orthonormal frames align with one another.

We use five metrics we use in comparing the anisotropic B-factors:

- The magnitude of the fluctuation. For this, we use the trace of the tensors, which is related to the isotropic B-factors.
- The shape of the ellipsoids, or how anisotropic they are. To this end we define two terms: (1) *first anisotropy* κ, which is the ratio of the smallest eigenvalue to the largest eigenvalue. The ratio ranges from 0 to 1, with 1 being spherical and 0 being extremely non-spherical; and (2) *second anisotropy* χ, which is the ratio of the middle eigenvalue to the largest eigenvalue.
- The orientation of the ellipsoids, or the directional preference of the fluctuations. For this, we use polar angles: (1) the angle θ, which is the angle between the first principal axes of the two tensors being compared (see Fig. 1); (2) the angle ¢: the angle between the second principal axes after the first are aligned (see Fig. 1).

The comparison process between the theoretical and experimental anisotropic B-factors of a given protein can be summarized as follows:

1. retrieve experimental anisotropic B-factors from the PDB (ANISOU entries) and calculate the theoretical anisotropic B-factors using ANM [Eq. (10)];
2. for each residue i ($1 \le i \le N$), based on its experimental and theoretical anisotropic tensors, determine B_i, κ_i, χ_i for both experiment and theory, and θ_i and ϕ_i.
3. for isotropic B-factors, or B_i's, calculate the correlation between experiment and theory by:

$$corr\left(B^{exp}, B^{theo}\right) = \frac{B^{exp} - \langle B^{exp}\rangle\, B^{theo} - \left\langle B^{theo}\right\rangle}{\|B^{exp} - \langle B^{exp}\rangle\| \left\|B^{theo} - \langle B^{theo}\rangle\right\|} \quad (11)$$

   A perfect correlation between two vectors gives a value of 1 while a perfect anti-correlation gives -1. Others fall in between.

4. for the first anisotropy κ_i's and second anisotropy χ_i's, calculate the difference between experiment and theory, i.e., set $\Delta\kappa_i = \kappa^{exp}_i - \kappa^{theo}_i$, and likewise χ_i (i is the residue index). To measure how well overall the first anisotropy (and second anisotropy) is predicted by theory for a given protein, we use $<\Delta\kappa> = \text{mean}(\Delta\kappa_i)$ and its standard deviation $\sigma(\Delta\kappa) = \text{sd}(\Delta\kappa_i)$, and express the difference as $<\Delta\kappa> \pm \sigma(\Delta\kappa)$. Similarly $<\Delta\chi> \pm \sigma(\Delta\chi)$ is used for the second anisotropy.

5. Similarly, we use <θ> and <ø> to measure how well overall the directions of the fluctuations are predicted for a protein.

Another measure - the correlation coefficients for comparison of anisotropic B-Factors

Besides the above comparison of experimental and predicted anisotropic B-factors, there is another method to compare two tensors26. Let U and V be two tensors (anisotropic B-factors), the correlation coefficient between them is derived from their electron-density maps as follows:

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

The normalized correlation coefficient is given by:

$$ncc(U,V)=\frac{CC\left[U,\left(U_{eq}/V_{eq}\right)V\right]}{cc(U,U_{iso})\,cc(V,V_{iso})} \tag{13}$$

where U_{iso} and V_{iso} describe a pair of isotropic atoms, with U 11 *iso* = U 22 *iso* = U 33 *iso* = U_{eq} = trace(U)/3 and similarly for V_{iso}. This normalized correlation coefficient *ncc* will be greater than 1 if two atoms described by U and V are more similar to each other than to an isotropic atom, and will be no more than 1 otherwise. Thus, the *ncc* provides an excellent measure to compare the size, orientation and direction of two tensors. In practice, a simple ratio of how many atoms in a structure have their normalized correlation coefficient values larger than 1 and the total number of atoms would give a good measure of the quality of an anisotropic B-factor prediction.

RESULTS

As we discussed in the Methods section, the anisotropic B-factors, or anisotropic displacement parameters (ADPs), are symmetric tensors for each atom. We diagonalize the tensors to find the eigenvalues and principal axes (eigenvectors). The eigenvalues indicate the magnitude of the fluctuations and the shape of the atom displacements, which in general are anisotropic and therefore ellipsoidal instead of spherical. On the other hand, the eigenvectors of a given ADP tensor tell us the directionality of the fluctuation. The fluctuation is usually not isotropic and is biased toward the direction of the principal axis corresponding to the largest eigenvalue (in other words, along the longest axis of the ellipsoid).

For the magnitude/shape of the fluctuation, we look at three terms: the magnitude, which is equivalent to the isotropic B-factors B_i's; the first anisotropy κ_i and the second anisotropy χ_i, which measure the shape of the atomic displacements.

We perform these comparisons for all the proteins in our dataset and give the results below.

Magnitude and anisotropy prediction using ANM

Isotropic B-factors—The correlation between experimental and calculated isotropic B-factors gives us a good measure of how well a model can reproduce/predict these values. As shown in Fig. 2, the quality of prediction using ANM is comparable to that from GNM. For the X-ray dataset, the mean correlation obtained by using ANM is about 0.51, which is

slightly lower than what is obtained with GNM, about 0.58. For the NMR dataset, it is 0.73 and 0.79 for ANM and GNM respectively. Using either model (ANM or GNM), significantly better correlations are found with NMR dataset, as observed in previous studies22,23.

Anisotropy prediction—Fig. 3(A) shows the mean first anisotropy difference $\langle\Delta\kappa\rangle$ (see Methods section) between experiment and calculation, for the X-ray dataset. From the figure we can see that ANM on average is able to predict fairly well the overall level of the first anisotropy. For most proteins, $\langle\Delta\kappa\rangle$ is within the range of [-0.2,0.2]. However, we see the standard deviation $\sigma(\Delta\kappa)$ is fairly large, about 0.2, and is strikingly similar for all the proteins. This means that for an individual residue, the first anisotropy predicted by ANM on average deviates by about 0.2 from experimental values, for all these proteins. The results for the second anisotropy χ are similar [see Fig. 3(B)] - the second anisotropy predicted by ANM also deviates by about 0.2. For most proteins, since the anisotropy distribution among all residues/atoms is roughly normal with a mean value around 0.517 and the mean value for the second anisotropy is about 0.7 based on our calculations, the discrepancy of 0.2 means that the anisotropy predictions of ANM differs from experimental values by about 0.2/0.5 = 40% for the first anisotropy and 0.2/0.7 = 30% for the second anisotropy. For the NMR dataset, the results of first and second anisotropy prediction are shown in Fig. 3(C) and (D). The results are slightly better than those of the X-ray dataset. It is also noted that in the NMR dataset, the anisotropic difference between experimental and predicted values tends to be negative, which is the opposite of the X-ray dataset. This indicates that there is an intrinsic difference between the anisotropic fluctuations found in X-ray structures and those in NMR ensembles. The fluctuations are less anisotropic (higher anisotropy values) in X-ray and more anisotropic (lower anisotropy value) in NMR. The predicted anisotropy level from the ANM falls in the middle. This intrinsic difference between X-ray and NMR structures is likely due to the fact that for the X-ray structures, the atomic fluctuations and thus anisotropy are underestimated27, while for the NMR structures, the anisotropy may be overestimated, particularly for regions having fewer NOE constraints.

To more strictly assess the difference between computed anisotropies and the experimental values, we conducted a simulation test. For each structure, we generated a random anisotropic tensor for each residue and compared it with the corresponding experimental tensor. Each random anisotropic tensor represents a randomly oriented ellipsoid with radii that are sampled randomly between (0, 1]. For the X-ray dataset, it is found that the mean and standard deviation of the first anisotropy differences between random and experimental tensors are much larger than those found using the ANM (0.20±0.26 vs. 0.04±0.17), while for the second anisotropy differences, the results are less striking (0.06±0.27, 0.06±0.19), though the ANM predictions clearly have a smaller standard deviation. These results show that the ANM predicted tensors are much more similar to the experimental tensors than the randomly generated tensors are and thus demonstrate the ANM predictions are significant.

Fig. 4 shows, at the residue level, the difference between the experimental anisotropies and the values predicted by ANM. For the X-ray structure of the rubredoxin (PDB id: 1IR0), it is seen from Fig. 4(A) and (B) that the shape of the fluctuation of each residue is reproduced reasonably well (in terms of the first anisotropy and second anisotropy values). The results for the NMR ensemble of the poxvirus complement control protein (PDB id: 1E5G) are shown in Fig. 4(C) and (D). Again, the residue fluctuations are well reproduced.

Motion directions predicted by ANM

Anisotropic B-factors (or ADPs) provide not only the magnitude, but also, of even greater interest, directional information about atomic fluctuations. Direct comparison between a

model and experimental data can help uncover some further details about atomic fluctuations and identify collective modes of motion that could be important for function. The experimental anisotropic B-factors thus can provide more extensive experimental validation of a model, such as ANM. If good agreement is found, such validations can provide justification for applying a model to study other aspects of protein dynamics, in order to understand how large scale protein conformation transitions take place.

As we defined earlier (see Methods section), the θ value measures the angle between the experimental and calculated directions of fluctuations, while the φ value measures the rotation needed to align the two sets of principal axes after their largest axes are aligned (see Fig. 1, here we considered the absolute values of the two angles). The $\langle\theta\rangle$ value thus gives an overall estimation of the performance of the model (here ANM) in predicting the directions of fluctuations for a given protein.

Fig. 5(A) shows the $\langle\theta\rangle$ values for the proteins in the X-ray dataset. It is seen from the figure that $\langle\theta\rangle$ and $\langle\sigma\rangle$ are consistently quite large, around 50°. Slightly better results are obtained for the NMR dataset, as shown in Fig. 5(B), where the average of $\langle\theta\rangle$ is about 40°. The standard deviations for θ and σ angles are similar for the X-ray and NMR datasets - the mean standard deviations are about 22° for both datasets.

Using one protein (again the rubredoxin, PDB id: 1IR0) as an example, Fig. 6(A) shows the θ and σ values of individual residues, specifically the alpha carbons. Since θ and σ measure how well the directions of the fluctuations are predicted and the lower the θ and σ values, the better the prediction, Fig. 6(A) indicates that the quality of the prediction for the directions of the atomic fluctuations varies significantly from residue to residue. While for some residues the directions of fluctuations calculated from ANM match well with those deduced from experimental anisotropic B-factor data, for many other residues the predictions are quite poor, some even differing by nearly 90°. A possible explanation for the latter is that the first principal direction of the fluctuations predicted by ANM might be aligned with the second principal direction of the experimental fluctuations (or vice versa), which may occur especially when it is hard to differentiate the (first) principal direction from the second (i.e., when their corresponding eigenvalues are the same). Better results are obtained for the NMR ensemble of the poxvirus complement control protein (PDB id: 1E5G), for which the θ values for most of its residues are below 30° [see Fig. 6(B)].

Correlation coefficients between experimental and theoretical results

The unnormalized and normalized correlation coefficients [*cc* and *ncc*, see Eqs. (12) and (13)] are used to compare the experimental anisotropic temperature factors with those predicted by ANM. From Fig. 7 we can see that for most X-ray structures, the percentage of residues with *ncc* above 1 (which means the prediction is good) is quite high, with an average value of about 68%. For the NMR dataset, the results are significantly better, with an average value of about 89%. These results demonstrate that there exists high similarity between the experimental (or derived, for NMR case) anisotropic B-factors and ANM predicted ones. And in general, the prediction is more successful for NMR data than for X-ray. The difference likely comes from the crystal packing effects that are not accounted for in the ANM model but exist for X-ray structures15. Using X-ray structure (1IRO) and NMR structure (1E5G) again as examples, Fig. 8 shows the *cc* and *ncc* distributions at the residue level. It is seen that most residues have *ncc* values above 1 for the chosen X-ray structure, while for the selected NMR structure, all of its residues have *ncc* values larger than 1.

DISCUSSIONS

From the comparisons between results from ANM and experimental data shown above, we see that the ANM is able to predict moderately well the relative fluctuation magnitudes of individual residues and even their anisotropies. Its prediction of the directional aspect of the fluctuations using the θ and φ measures, on the other hand, appears to deviate quite significantly from experimental values. However, it is quite likely that many of these deviations result from the artifact of the misalignment of the principal axes, as discussed in the previous section. Indeed, the results using normalized correlation coefficient as the measure show that most of the predictions are correct. In particular, the results from NMR dataset are consistently better than those from X-ray data.

To assess the effect of possible misalignment of the principal axes, we calculated θ and φ angles after switching the order of the two principal axes that have similar lengths in the computed tensors. These θ and φ angles were then compared with the original θ and φ angles and the smaller ones were used as the new θ and φ angles. After comparing the new θ and φ angles with the original ones for all the structures in the X-ray dataset, we found that the <θ> decreases from 50° to 46°, while the <φ> is almost the same (about 55°). For the NMR dataset, we obtained similar results (the <θ> decreases from 39° to 37°, while the <φ> is almost the same, around 51°). These results indicate that the misalignment of the principal axes of the tensors has a slight effect on the θ angles, but has little effect on the φ angles.

A natural question to ask is why this occurs, and then what can be done to improve the model. First, experimental anisotropic B-factors, or anisotropic displacement parameters (ADPs), are found by fitting X-ray diffraction data of protein crystals. These parameters thus may describe static disorder (atomic coordinate differences between unit cells), dynamic disorder (since the diffraction data represent a time average of protein motion), rigid-body motion of the protein, internal motion of the protein, and lastly, refinement errors and uncertainties, as pointed out in several recent papers[28-30]. And it is not clear how much the internal motion contributes to the total observed fluctuations. It has been proposed that the external rigid-body motions of proteins may contribute up to 60% of the total fluctuations[31,32]. If this is true, the ANM, as a coarse-grained model that only considers the internal motion of a protein, may have missed this important component in the comparison with experimental data. However, proper inclusion of rigid-body motion is not trivial, as it usually involves introducing many parameters and fitting, and potentially over-fitting, to the experimental data, which is not desired. For example, the translation libration skew (TLS) model[33,34] has 20 parameters, as compared to one parameter in the ANM. Soheilifard et al. have evaluated the capability of the ANM and TLS model in predicting the protein isotropic B-factors[30]. Here we represent the internal motions by the ANM and the rigid-body motions by the TLS. The combined motions are then fitted to the experimental tensors using least-square fitting. Compared with the original results from X-ray structures, it is found that the fitted tensors give better agreements with the experimental tensors. The mean deviations of anisotropies using the ANM/TLS fitted tensors remain small, while the means of θ and φ angles are smaller than those from the ANM alone (35° vs. 50° for <θ> and 50° vs. 55° for <φ>). For the measure of percentage of residues with *ncc* above 1, the result from the fitted tensors is also better (91% vs. 69%). As fitting can always improve results, the above improvement is not that surprising. But the results do confirm that the inclusion of rigid-body motions has the potential to improve a model's predictions of anisotropic tensors. What is difficult is how to set proper limits on the parameters, so that over-fitting is minimized. This will be explored in future work.

On the other hand, the coarse-grained nature of the ANM itself may account for some of the differences between experimental and theoretical results shown earlier. The ANM normally

simplifies each residue by representing it with its alpha carbons. It ignores the other atoms on the backbone and even the side chains, which likely strongly influences how atoms fluctuate locally. It also normally does not take any bound ligands into account. The ANM uses a uniform spring constant and cutoff distance for every residue/atom. While in reality, the interaction strength may be residue specific and distance/orientation dependent. Therefore, these details may contribute significantly to the anisotropy of the atomic fluctuations. For example, the ANM uses the same spring constant for the backbone contacts as for the rest of the contacts. While this has been shown not to affect the isotropic B-factors much, i.e., the fluctuation magnitude4, one may wonder whether it might have a more pronounced effect on the directional aspects of the fluctuations. Even though a protein molecule in a crystal can be in close contact with other molecules in neighboring cells, the ANM usually treats a protein as an isolated molecule and ignores any effects of the crystal environment. It has been shown15,35 that including some neighboring effects helps improve such models to some extent.

It is notable that the prediction is better for the NMR data than for the X-ray data. One possible explanation is that the NMR structures are determined in a solution environment that is free from the crystal packing effects that X-ray structures have. Similar phenomena are observed in our previous comparison of NMR and X-ray structures of HIV-1 protease23 and also reported in Yang et al.'s work22.

CONCLUSIONS

In this work, we have used the ANM to compute the anisotropic temperature factors of a large set of high resolution protein structures. The rich experimental anisotropic temperature factor data in turn are used as validation data to closely test the ANM model. We employed five terms to compare the experimental and theoretical anisotropic tensors: (1) isotropic B-factors, (2) first anisotropy κ, (3) second anisotropy χ, and (4) and (5) directional preferences θ and φ. As a separate measure of similarity, we also calculated the (normalized) correlation coefficients between the experimental and calculated anisotropic tensors. Our results show that for the X-ray data: (1) the correlation for isotropic B-factors predicted by the ANM is about 0.51, (2) the anisotropy predictions differ from experimental values by about 30% to 40%, (3) the directions of fluctuations are different by about 50° on average, and (4) using normalized correlation coefficient as the measure, over 68% of the residue anisotropic tensors are predicted well by the ANM. For the NMR data, the prediction results are even better. These results further confirm the validity of the ANM for predicting the anisotropic temperature factors. On the other hand, there still exist some differences between the experimental and predicted results, indicating improvements to the model are needed to resolve these differences and to obtain a more accurate understanding of protein motions and dynamics.

Acknowledgments

We would like to thank the two anonymous reviewers for their helpful advice on this paper. RLJ acknowledges the financial support provided by NIH grant 1R01GM072014 and NSF grant CNS-0521568.

REFERENCES

1. Frenkel, D.; Smit, B. Understanding molecular simulation. Academic Press; 2001.
2. McCammon, JA.; Harvey, SC. Dynamics of proteins and nucleic acids. Cambridge University Press; 1987.
3. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett. 1996; 77:1905–1908. [PubMed: 10063201]

4. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des. 1997; 2:173–181. [PubMed: 9218955]
5. Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. Phys Rev Lett. 1997; 79:3090–3093.
6. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins. 1998; 33:417–429. [PubMed: 9829700]
7. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J. 2001; 80:505–515. [PubMed: 11159421]
8. Wang Y, Rader AJ, Bahar I, Jernigan RL. Global ribosome motions revealed with elastic network model. J Struct Biol. 2004; 147:302–314. [PubMed: 15450299]
9. Keskin O, Bahar I, Flatow D, Covell DG, Jernigan RL. Molecular mechanisms of chaperonin GroEL-GroES function. Biochemistry. 2002; 41:491–501. [PubMed: 11781087]
10. Xu C, Tobi D, Bahar I. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T<->R2 transition. J Mol Biol. 2003; 2333:153–168. [PubMed: 14516750]
11. Zheng W, Doniach S. A comparative study of motor-protein motions by using a simple elastic network model. Proc Natl Acad Sci USA. 2003; 100:13253–13258. [PubMed: 14585932]
12. Tama F, Gadea FX, Marques O, Sanejouand YH. Building-block approach for determining low-frequency normal modes of macromolecules. Proteins. 2000; 41:1–7. [PubMed: 10944387]
13. Kim MK, Jernigan RL, Chirikjian GS. Rigid-cluster models of conformational transitions in macromolecular machines and assemblies. Biophys J. 2005; 89:43–55. [PubMed: 15833998]
14. Yang L, Song G, Jernigan RL. How well can we understand large-scale protein motions using normal modes from elastic network models? Biophys J. 2007; 93:920–929. [PubMed: 17483178]
15. Kundu S, Melton JS, Sorensen DC, Phillips GN Jr. Dynamics of proteins in crystals: comparison of experiment with simple models. Biophys J. 2002; 83:723–732. [PubMed: 12124259]
16. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhatand TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]
17. Merritt EA. Expanding the model: anisotropic displacement parameters in protein structure refinement. Acta Cryst D. 1999; 55:1109–1117. [PubMed: 10329772]
18. Eyal E, Chennubhotla C, Yang LW, Bahar I. Anisotropic fluctuations of amino acids in protein structures: insights from X-ray crystallography and elastic network models. Bioinformatics. 2007; 23:i175–i184. [PubMed: 17646294]
19. Kondrashov DA, Van Wynsberghe AW, Bannen RM, Cui Q, Phillips GN Jr. Protein structural variation in computational models and crystallographic data. Structure. 2007; 15:637–637.
20. Chen X, Poon BK, Dousis A, Wang Q, Ma J. Normal-mode refinement of anisotropic thermal parameters for potassium channel KcsA at 3.2 Å crystallographic resolution. Structure. 2007; 15:955–962. [PubMed: 17698000]
21. Poon BK, Chen X, Lu M, Vyas NK, Quiocho FA, Wang Q, Ma J. Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-Å crystallographic resolution. Proc Natl Acad Sci USA. 2007; 104:7869–7874. [PubMed: 17470791]
22. Yang LW, Eyal E, Chennubhotla C, Jee J, Gronenborn AM, Bahar I. Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. Structure. 2007; 15:741–749. [PubMed: 17562320]
23. Yang L, Song G, Carriquiry A, Jernigan RL. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. Structure. 2008; 16:321–330. [PubMed: 18275822]
24. Scheringer C. On the interpretation of anisotropic temperature factors. Acta Cryst. 1977; A33:879–884.
25. Sen TZ, Feng Y, Garcia JV, Kloczkowski A, Jernigan RL. The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. J Chem Theo Comp. 2006; 2:696–704.
26. Merritt EA. Comparing anisotropic displacement parameters in protein structures. Acta Cryst D. 1999; 55(Pt 12)

27. Levin EJ, Kondrashov DA, Wesenberg GE, Phillips GN Jr. Ensemble refinement of protein crystal structures: validation and application. Structure. 2007; 15:1040–1052. [PubMed: 17850744]
28. Kondrashov DA, Cui Q, Phillips GN Jr. Optimization and evaluation of a coarse-grained model of protein motion using X-ray crystal data. Biophys J. 2006; 91:2760–2767. [PubMed: 16891367]
29. Eyal E, Bahar I. Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. Biophys J. 2008; 94:3424–3435. [PubMed: 18223005]
30. Soheilifard R, Makarov DE, Rodin GJ. Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors. Phys Biol. 2008; 5:26008. [PubMed: 18577808]
31. Stec B, Zhou R, Teeter MM. Full-matrix refinement of the protein crambin at 0.83 Å and 130 K. Acta Cryst D. 1995; 51:663–681. [PubMed: 15299796]
32. Diamond R. On the use of normal modes in thermal parameter refinement: theory and applicationto the bovine pancreatic trypsin inhibitor. Acta Cryst A. 1990; 46:625–635. [PubMed: 2206485]
33. Cruickshank DWJ. The analysis of the anisotropic thermal motion of molecules in crystals. Acta Cryst. 1956; 9:754–756.
34. Schomaker V, Trueblood KN. On the rigid-body motion of molecules in crystals. Acta Cryst. 1968; B24:63–76.
35. Song G, Jernigan RL. vGNM: a better model for understanding the dynamics of proteins in crystals. J Mol Biol. 2007; 369:880–893. [PubMed: 17451743]

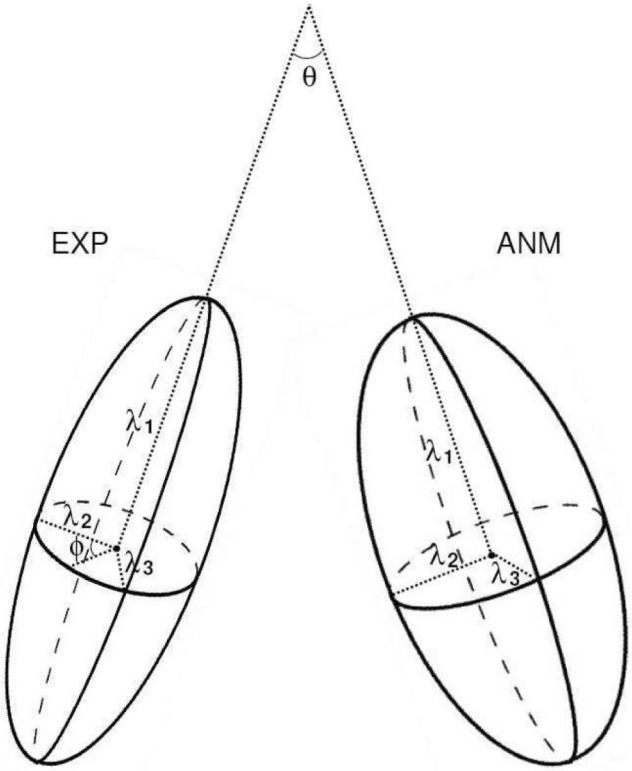


Figure 1.
Comparison of anisotropic B-factor tensors from experiment and theory. The first anisotropy κ is defined by the ratio λ_3/λ_1; the second anisotropy χ is defined as λ_2/λ_1; θ is the angle between the first principal axes of the two tensors; φ is the angle between the second principal axes after the first principal axes are aligned.

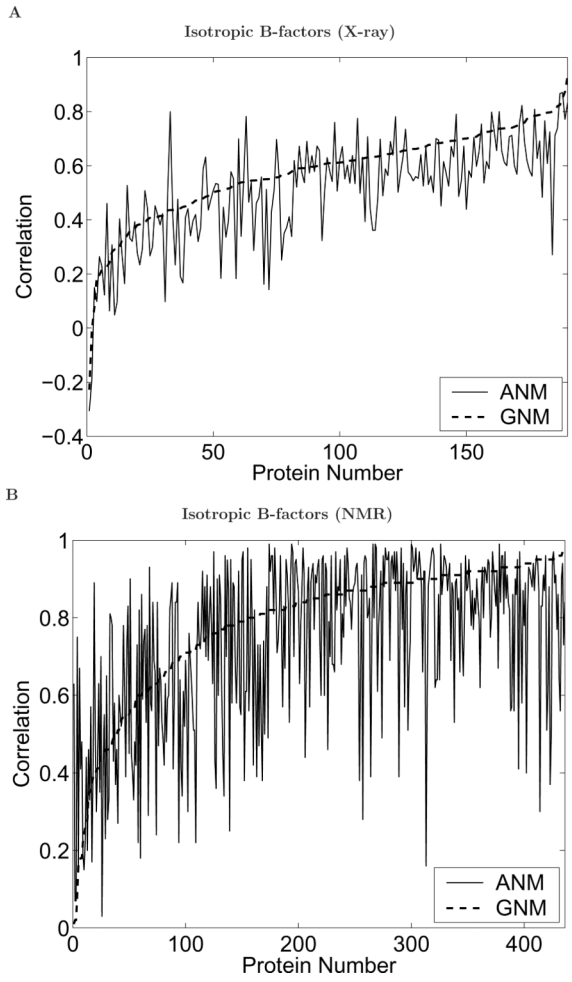


Figure 2.

The correlations between experimental isotropic B-factors and those predicted by GNM (shown in dashed line) and ANM (shown in solid line), for all the proteins in the (A) X-ray dataset and (B) NMR dataset (the results are sorted by the GNM correlation values). The quality of prediction using ANM is quite similar to that of GNM.

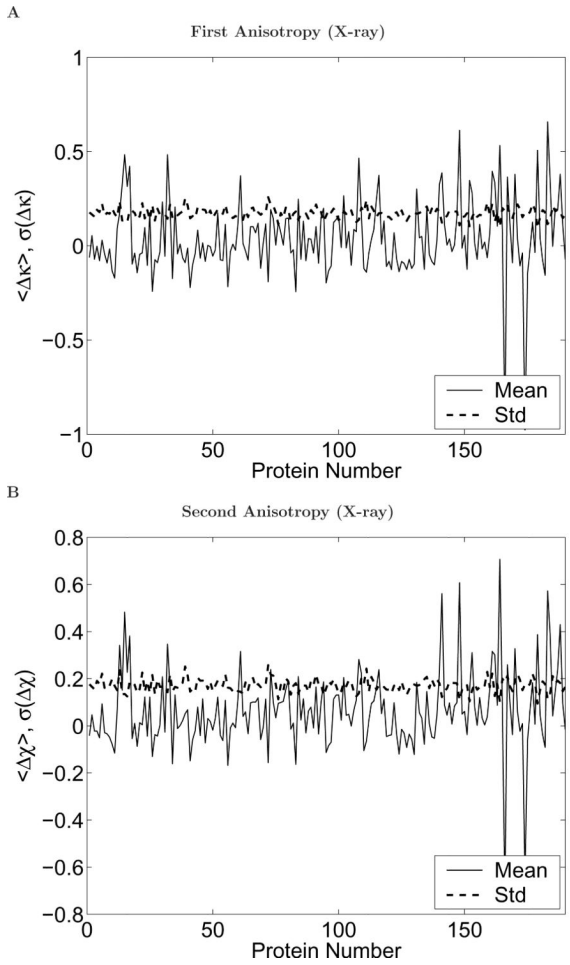
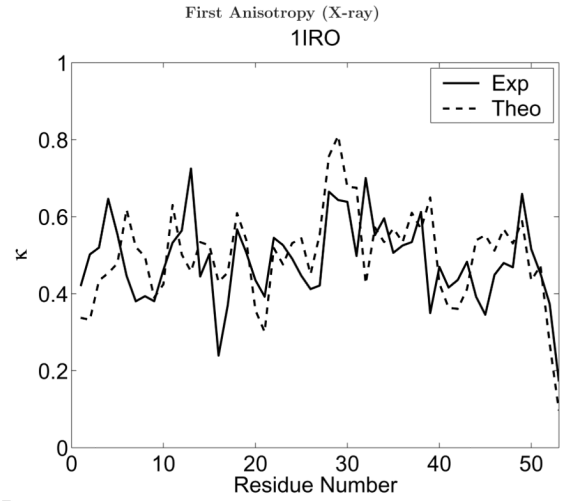


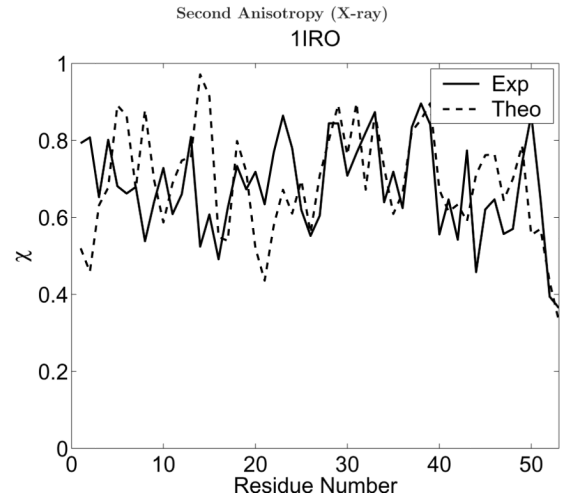
Figure 3.

The anisotropy differences between experimental values and predictions: (A) the first anisotropy difference Δκ for X-ray dataset, (B) the second anisotropy difference Δχ for X-ray dataset, (C) the first anisotropy difference Δκ for NMR dataset, and (D) the second anisotropy difference Δχ for NMR dataset.

A



B



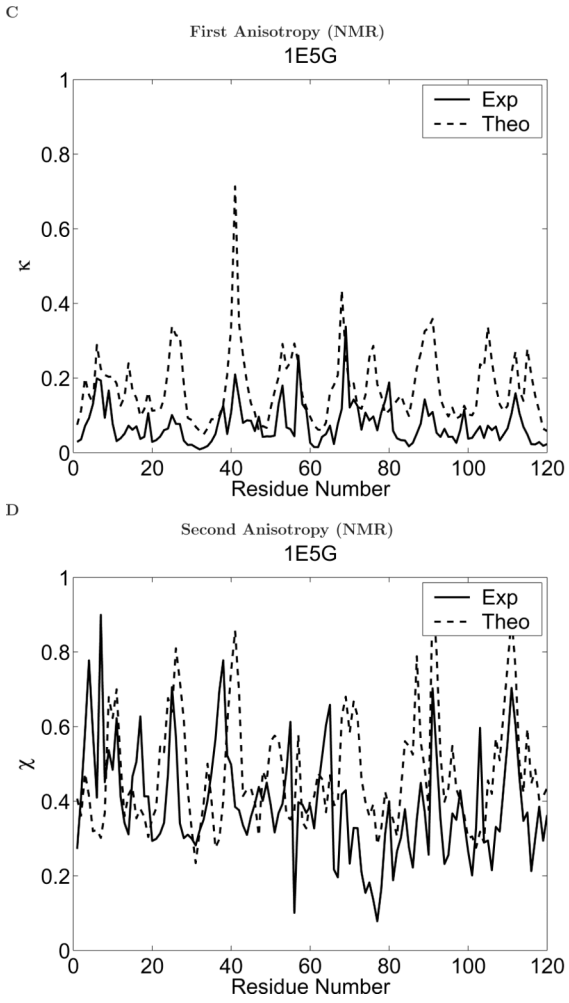


Figure 4.

The values of (A) the first anisotropy κ and (B) the second anisotropy χ for the alpha carbons of rubredoxin (X-ray structure, PDB id: 1IRO), and (C) the first anisotropy κ and (D) the second anisotropy χ for the alpha carbons of poxvirus complement control protein (NMR ensemble, PDB id: 1E5G). The experimental values are shown in solid lines, while the values predicted by ANM are shown in dashed lines. In ANM, each residue is represented by its alpha carbon.

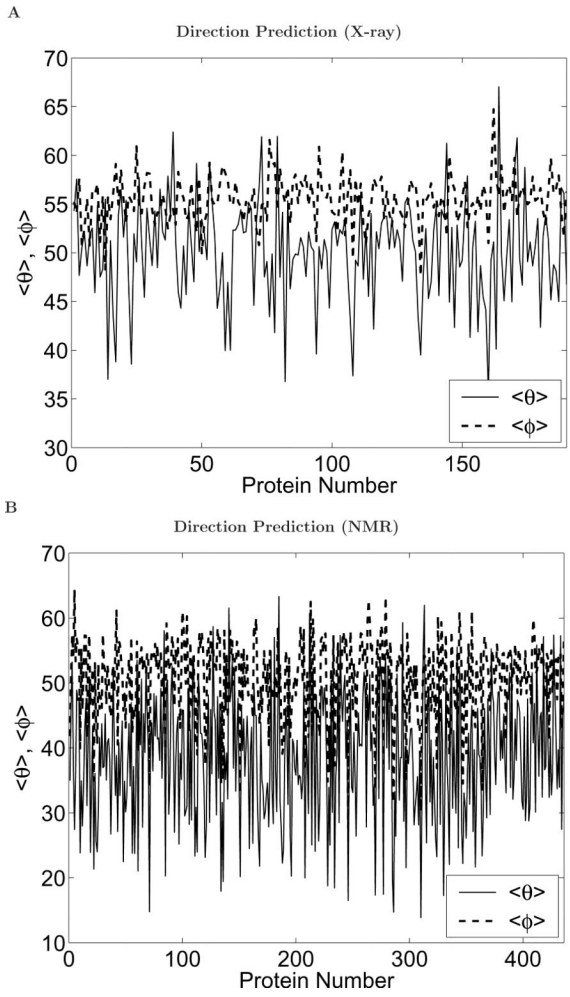


Figure 5.

The <θ> and <φ> values for all the proteins in (A) the X-ray dataset and (B) the NMR dataset. The <θ> and <φ> measure how well overall the directions of the fluctuations are predicted for a protein. A perfect prediction renders both <θ> and <φ> as 0.

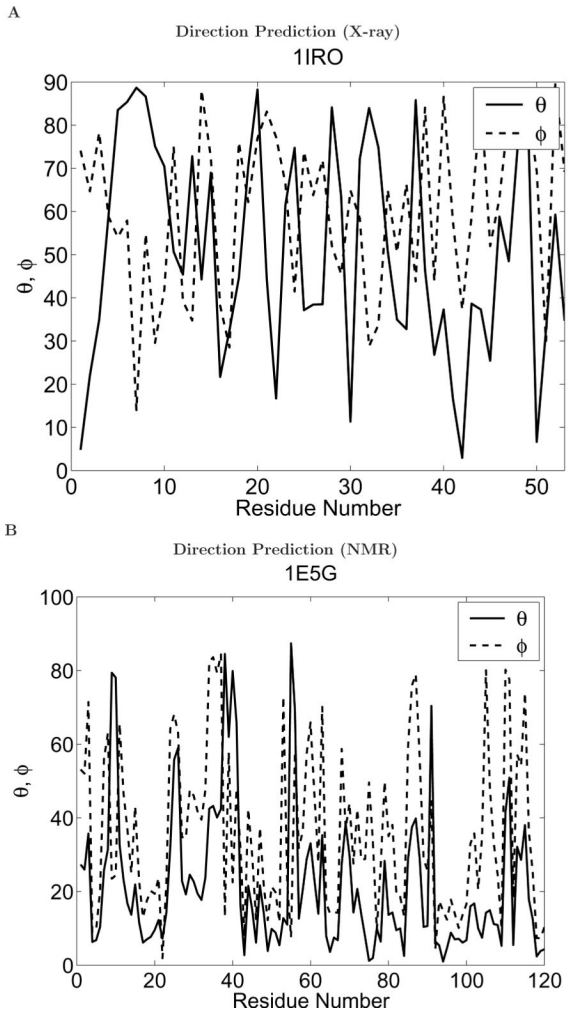


Figure 6.

The θ and φ values at the residue level: for the alpha carbons of (A) the X-ray structure of rubredoxin (PDB id: 1IRO) and (B) the NMR ensemble of poxvirus complement control protein (PDB id: 1E5G). The θ and φ measure how well the directions of the fluctuations are predicted for a given atom/residue. A perfect prediction renders both θ and φ as 0.

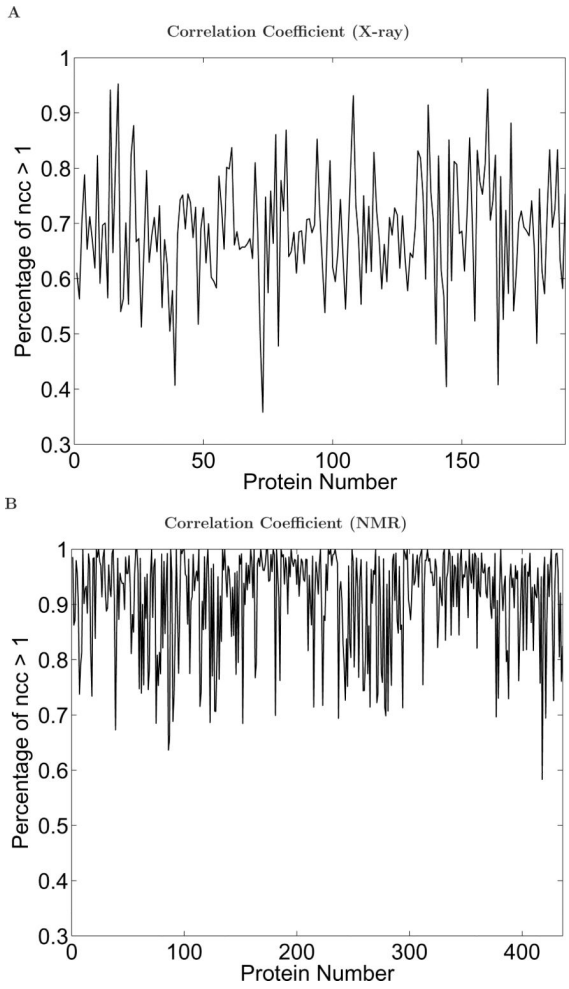


Figure 7.
The percentage of residues with *ncc* above 1 for all the proteins in (A) the X-ray dataset and (B) the NMR dataset.

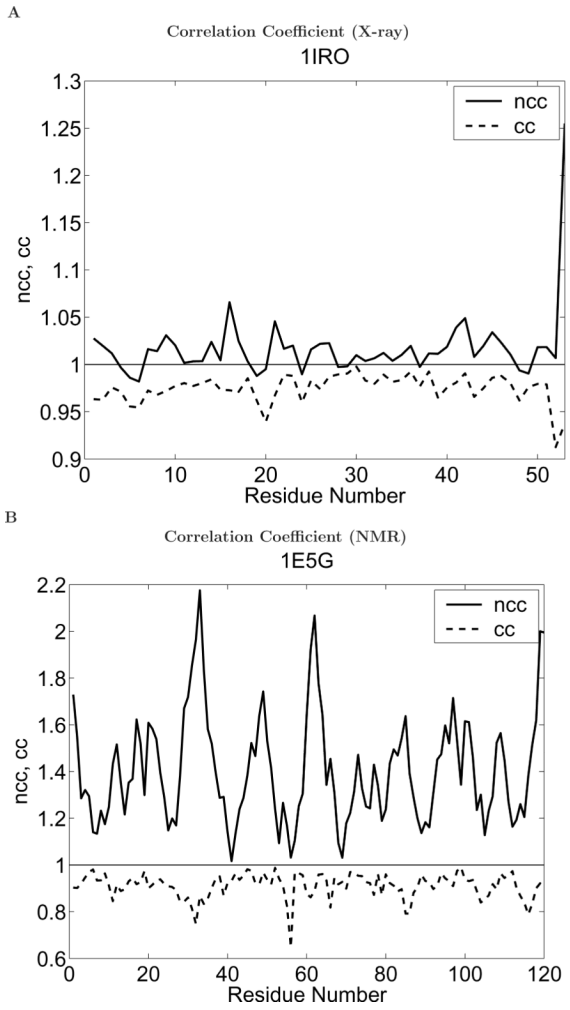


Figure 8.

The *cc*/*ncc* values at the residue level: for the alpha carbons of (A) the X-ray structure of rubredoxin (PDB id: 1IRO) and (B) the NMR ensemble of poxvirus complement control protein (PDB id: 1E5G).