The effects of rigid motions on elastic network model force constants

Timothy R. Lezon[1]
Department of Computational and Systems Biology, University of Pittsburgh, PA

Abstract

Elastic network models provide an efficient way to quickly calculate protein global dynamics from experimentally determined structures. The model's single parameter, its force constant, determines the physical extent of equilibrium fluctuations. The values of force constants can be calculated by fitting to experimental data, but the results depend on the type of experimental data used. Here we investigate the differences between calculated values of force constants _t to data from NMR and X-ray structures. We find that X-ray B factors carry the signature of rigid-body motions, to the extent that B factors can be almost entirely accounted for by rigid motions alone. When fitting to more refined anisotropic temperature factors, the contributions of rigid motions are significantly reduced, indicating that the large contribution of rigid motions to B factors is a result of over-fitting. No correlation is found between force constants fit to NMR data and those fit to X-ray data, possibly due to the inability of NMR data to accurately capture protein dynamics.

Keywords

protein dynamics; coarse-grained; B factors; crystal packing; anisotropic temperature factors

Introduction

Elastic network models (ENMs) are extensively used for investigating and predicting the global dynamics of proteins. These relatively simple models have repeatedly shown high qualitative agreement with protein dynamics inferred from experiments [1–4]. The most common ENMS [5–9] have only two key components: a scaling function that decreases the interaction strength between residues with distance, and a scalable force constant that controls the overall strength of the harmonic interactions. The choice of scaling function qualitatively affects the output by determining the shapes of the normal modes of vibration; the force constant determines the physical amplitude of the motion. Generally the scaling function is taken to be either a step function that reduces the interaction to zero beyond a cutoff distance [5;8;9] or a smooth distance-dependent function that gradually lessens pairwise interaction strengths [6;7]. The force constant is often ignored, leaving ENM users with the shapes of predicted motions, but not their amplitudes.

The dynamical description provided by the ENM is fundamentally incomplete without a good value for the force constant, which sets the energetic cost associated with any structural deformation. An undefined force constant provides no qualitative physical limitation on the extent of motions, enabling the modeler to propose deformations of arbitrary amplitude. This lack of detail is particularly hazardous in applications that use

[1]Corresponding author. Address: Department of Computational and Systems Biology, University of Pittsburgh, Suite 3064 BST3, 3501 Fifth Ave. Pittsburgh, PA 15260, U.S.A. Phone: 412.383.5743 Fax: 412.648.3163 lezon@pitt.edu.

ENMs to calculate experimentally testable quantities. Free-energy surfaces [10;11] and transition pathways [12] can be investigated with ENMs, but only if an appropriate value of the force constant is used. Similarly, properly parameterized ENMs can provide insight into atomic force microscopy measurements of the elastic properties of proteins and their complexes [13;14]. The global motions predicted by ENMs can also be used to steer MD simulations [15;16], and accurate estimation of the force constant could further improve the computational efficiency of such studies. It is becoming increasingly clear that proteins do not reside in a single native state structure under physiological conditions, but that they populate a native conformational ensemble. Estimates of such ensembles can be generated in seconds using an ENM, but their validity rests on accurate parameterization of its force constant.

Commonly, force constants are estimated by comparing ENM-predicted dynamics to X-ray B factors, which are an abundant source of atomic-level information on protein dynamics; however, it has long been known that B factors are influenced by lattice vibrations that are not accounted for by ENMs [17–20]. Studies investigating the effects of crystal packing on ENM prediction of B factors [21–27] have concluded that crystal packing plays a considerable role in protein dynamics, and that rigid body motions contribute to B factors. Alternatively, force constants can be calibrated using other measures of protein dynamics, such as anisotropic displacement parameters (ADPs), NMR ensembles, MD trajectories or ensembles generated from multiple crystal structures. As with B factors, each of these methods of inferring dynamics carries its own bias: ADPs are influenced by lattice vibrations; NMR ensembles are themselves determined using pairwise harmonic restraints similar to those used in ENMs; MD trajectories represent only small fluctuations about local minima on very short time scales; multiple crystal structures represent the ensemble of conformations of interest to crystallographers, not necessarily the ensemble that is sampled in vivo. Not surprisingly, the value obtained for the ENM force constant can vary considerably based on the type of data to which it is fit [26;28], and earlier studies [29–32] investigating force constants provide values that range from 0.1 to 10 kcal/mol/$Å^2$.

Here we ask whether fitting force constants to single-structure X-ray data can produce reliable quantitative predictions from ENMs, and whether rigid motions can account for the discrepancies in force constant values obtained from different experimental methods. Using direct mathematical fitting, we investigate the values of force constants that best account for fluctuations inferred from isotropic B factors, anisotropic temperature factors and NMR ensembles. We find that the paucity of information provided by isotropic X-ray B factors results in great uncertainty in values of ENM force constants, that fitting to B factors overemphasizes the contribution of rigid motion, and that fitting to anisotropic temperature factors decreases the extent to which rigid motions influence the force constants. We note a statistical difference in the fluctuations obtained from X-ray and NMR solved structures, suggesting a role for lattice vibrations in X-ray B factors.

Methods

Protein set

We used 64 proteins that have been solved by both X-ray crystallography and solution NMR, shown in Table 1. Forty-four of these were derived from the set of Garbuzynskiy *et al.* [33], and an additional 20 were selected specifically based on the presence of anisotropic temperature factors using similar criteria. The proteins range in size between 52 and 269 residues, and their NMR ensembles each contain between 10 and 50 models.

ANM

We use the Anisotropic Network Model (ANM) [8;9] to provide a harmonic potential near each protein's equilibrium conformation. In this model, each residue is represented as a point particle of unit mass, and neighboring residues interact through Hookean couplings of a fixed uniform spring constant, γ. If $\mathbf{r}$ is the $3N$-dimensional vector of mass-weighted displacements of the N residues from their equilibrium positions, then the equations of motion are given by

$$\ddot{\mathbf{r}} = -\gamma \mathbf{H} \mathbf{r},$$

(1)

where $\gamma\mathbf{H}$ is the Hessian matrix of second derivatives of the potential with respect to the components of $\mathbf{r}$. The eigenvectors of $\mathbf{H}$ that satisfy Eq. 1 are directions of motion for small oscillations about the protein's equilibrium conformation, and their corresponding eigenvalues are the squares of the oscillatory frequencies in units that are determined by the spring constant, γ.

The variance-covariance matrix describing the residue fluctuations is given analytically by $\mathbf{H}^{-1}$. Because $\mathbf{H}$ has six zero eigenvalues corresponding to rigid-body motions, its inverse does not strictly exist, but is replaced in practice by the pseudoinverse, $\mathbf{H}^{\dagger}$, defined as the inverse in the non-zero eigenspace:

$$\mathbf{H}^{\dagger} = \sum_{k=m+1}^{3N} \frac{\mathbf{v}^{(k)} \tilde{\mathbf{v}}^{(k)}}{\lambda_k}.$$

(2)

Here $\mathbf{v}^{(k)}$ is the kth eigenvector of $\mathbf{H}$, $\tilde{\mathbf{v}}^{(k)}$ is its transpose and λ_k is its eigenvalue. The summation explicitly excludes the first m eigenvectors of $\mathbf{H}$, all of which are associated with an eigenvalue of zero and make no contribution to the internal energy.

The symmetric 3×3 matrices $(\mathbf{H}^{\dagger})_{\mathbf{ii}}$, $i=1\ldots N$ that lie along the diagonal of $\mathbf{H}^{\dagger}$ describe the distributions of the individual residues under the ANM potential. Each matrix $(\mathbf{H}^{\dagger})_{\mathbf{ii}}$ contains six unique elements, corresponding to the variances of the residue fluctuations along the three Cartesian axes, and the covariances of motions along the different axes. The MSF of residue i under the ANM potential is proportional to the trace of $(\mathbf{H}^{\dagger})_{\mathbf{ii}}$:

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle = \frac{k_B T}{\gamma} Tr(\mathbf{H}^{\dagger})_{\mathbf{ii}},$$

(3)

where k_B is Boltzmann's constant and T is the absolute temperature.

Fitting fluctuations

Residue fluctuations are straightforwardly obtained from an ensemble of aligned structures or from the temperature factors that are listed in individual PDB files. The isotropic B factors associated with structures solved via X-ray crystallography relate to the MSFs of the individual atoms as

$$B_i = \frac{8\pi^2}{3} \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle.$$

(4)

For structures of sufficiently high resolution, the spatial anisotropy of atomic fluctuations can be determined, and each atom's spatial distribution is described by a set of six ADPs.

The ADPs of atom i define the trivariate Gaussian distribution and correspond to the upper-triangle of $\frac{k_B T}{\gamma}(\mathbf{H}^\dagger)_{\mathbf{ii}}$.

The ANM spring constants that best account for the experimental values are found by minimizing a distance between experimental and theoretical fluctuations. The distance metric is

$$f = \sum_{i=1}^{n} (b_i - a_i)^2, \tag{5}$$

where $\mathbf{b}$ and $\mathbf{a}$ are the n-vectors of experimental and theoretical fluctuations, respectively. When fitting MSFs, the N components of $\mathbf{a}$ are calculated from Eq. 3, and f is minimized when

$$\frac{\gamma}{k_B T} = \frac{\sum_j [Tr(\mathbf{H}^\dagger)_{\mathbf{jj}}]^2}{\sum_i Tr(\mathbf{H}^\dagger)_{\mathbf{ii}} b_i}. \tag{6}$$

Similarly, when fitting ADPs, the $6N$ components of $\mathbf{a}$ are taken from the diagonal super-elements of $\mathbf{H}^\dagger$, the $6N$ components of $\mathbf{b}$ are the corresponding anisotropic temperature factors from the PDB file, and the force constant that minimizes f is

$$\frac{\gamma}{k_B T} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{6} [(\mathbf{H}^\dagger)_{\mathbf{ii}}]_j^2}{\sum_{i=1}^{N} \sum_{j=1}^{6} [(\mathbf{H}^\dagger)_{\mathbf{ii}}]_j b_{ij}}. \tag{7}$$

Here $[\mathbf{A}]_j$ indicates the jth element of the upper triangle of the 3×3 matrix $\mathbf{A}$, and b_{ij} is its experimental analogue.

The formulas in Eqs. 6 and 7 are valid so long as all of the motion captured by the experiment arises from the internal modes of the molecule. When fitting to B factors or ADPs, the possibility exists that some of the measured fluctuations are accounted for by rigid-body motions, necessitating another approach to finding the force constants from these values.

Accounting for rigid-body motions

Contributions to B factors arising from fluctuations of the crystal lattice do not factor into Eq. 6 because $\mathbf{H}^\dagger$ contains no information on rigid-body motions. The experimental B factors can be fit using all modes by allowing each of the rigid modes to contribute the the fluctuations. For the general case, suppose that there are m degenerate rigid modes (typically $m=6$). We define the vector $\mathbf{a}$ of theoretical MSFs as

$$\mathbf{a} = \mathbf{Gc}, \tag{8}$$

where **c** is a vector of $m+1$ constants to be fit and the elements of the $N \times (m+1)$ matrix **G** have the form

$$G_{ik} = Tr[(\mathbf{v}^{(k)} \tilde{\mathbf{v}}^{(k)})_{ii}] \quad 1 \leq k \leq m \tag{9}$$

$$G_{ik} = Tr[(\mathbf{H}^{\dagger})_{\mathbf{ii}}] \quad k = m+1. \tag{10}$$

Here we have used the notation that the eigenvectors $\mathbf{v}^{(1)} \ldots \mathbf{v}^{(m)}$ correspond to rigid motions that are fit with parameters $c_1 \ldots c_m$. The parameter c_{m+1} fits the internal motions, scaled by their respective eigenvalues. As **H** is dimensionless, **c** has units of squared distance. Alternatively, when fitting ADPs, **G** is $6N \times (m+1)$, and its elements are

$$G_{6i+j,k} = [(\mathbf{v}^{(k)} \tilde{\mathbf{v}}^{(k)})_{ii}]_j \quad 1 \leq k \leq m \tag{11}$$

$$G_{6i+j,k} = [(\mathbf{H}^{\dagger})_{\mathbf{ii}}]_j \quad k = m+1. \tag{12}$$

The c_k values that minimize Eq. 5 are given by

$$\mathbf{c} = (\tilde{\mathbf{G}}\mathbf{G})^{-1}\tilde{\mathbf{G}}\mathbf{b}, \tag{13}$$

and the corresponding spring constant is then $\gamma = k_B T / c_{m+1}$. A unique solution exists to Eq. 13 as long as **G** is rank $m+1$. This condition is generally met when $N \gg m$, even though the rows of **G** are not necessarily orthogonal. If **G** has rank less than $m+1$, as is the case when two of its rows correspond to fluctuations induced by rigid translations along orthogonal coordinate axes, then the number of parameters that require fitting is effectively reduced. As long as column $m+1$, corresponding to internal fluctuations, is linearly independent of the other columns, a well-defined force constant is obtained. If this column is not linearly independent from the others, then it is impossible to separate internal fluctuations from rigid-body motions.

If any component of **c** is negative, the solution is unphysical, corresponding to a negative amplitude along one of the modes. In such cases the global minimum is used as a starting point for an iterative projected Landweber non-negative least squares fit, which numerically provides a physical solution to Eq. 13. It is possible that the coefficient c_{m+1} will be zero, indicating that the experimental B factors are best described by rigid-body motions, and that internal dynamics make no contribution to the fit.

In the ANM, the six rigid modes correspond to three translations and three rotations; however, arbitrary linear combinations of these modes also yield valid rigid motions, such as the screw motions in the TLS model [18]. Because each rigid mode contributes non-linearly to the columns of **G**, the result of Eq. 13 will depend on the choice of basis $\mathbf{v}^{(1)} \ldots \mathbf{v}^{(m)}$. To assess the sensitivity of the force constants to the form of the rigid motions, we generate 10,000 randomly oriented basis sets and we solve Eq. 13 using each unique basis set. As a result, several arbitrary forms are examined for the lattice vibrations, and the force constants governing the internal motions are fit to the residual fluctuations for each.

Accounting for crystal contacts

To investigate the influence crystal contacts on ENM results, we employed a method introduced by Kim and coworkers [34] for generating the Hessian using symmetry transformations. The underlying idea is that, in addition to the contacts formed between residues in the protein, contacts may additionally be formed with neighboring proteins within the crystal. The modes of all molecules within the crystal are then identical upon application of symmetry transformations.

Results and Discussion

Force constants determined by fitting internal motions to isotropic B factors show no correlation with those determined using NMR ensembles

The results are summarized in Table 1. In units of $k_B T / \text{Å}^2$, the force constants calculated from NMR MSFs are distributed around 0.952±1.445, and those found by fitting to X-ray B factors without any rigid-body motions are distributed around 1.334±2.272. Even though the two sets produce force constants that are on the same order of magnitude, their values do not correlate well: The Pearson correlation between the two sets is found to be −0.078 (see Fig. 1A). Interestingly, the agreement between force constants fit by NMR and X-ray does not change when crystal contacts are considered. The correlation between the force constants fit to isolated X-ray structures and those fit using crystal contacts is 0.758 (Fig. 1B), indicating that including crystal contacts is not an essential part of fitting force constants. We further find that including crystal contact information does not affect the correlation of ENM-predicted fluctuations with B factors, in contrast to previous observations [21;24;25]. The average correlation with B factors is 0.458 for isolated structures and 0.447 for structures in crystal. As illustrated in Fig. 2, inclusion of crystal contacts attenuates excessive fluctuations of flexible loops but does not otherwise significantly impact the protein dynamics.

Freely fitting the rigid-body motions completely alters the calculated force constants as well as the correlation with experiment

There is a weak correlation (r= 0.530) between force constants fit to B factors using all modes and those fit using internal modes only; however, there remains no correlation (r= −0.022) between force constants fit to B factors using all modes and those fit to NMR MSFs (Fig. 1C,D).

When rigid-body motions are allowed to contribute to the MSFs, the contribution from internal motions decreases and the springs controlling these motions become stiffer. If a combination of rigid-body motions by itself fully accounts for the MSFs, then the coefficient c_{m+1} diminishes to zero and the force constant blows up owing to its reciprocal relationship with c_{m+1}. For isolated X-ray structures, inclusion of rigid motions is a valid assumption, as rigid motions may contribute to B factors. Similarly, for certain crystal lattices, limited rigid motions are permitted by symmetry, and these may contribute to observed B factors. In contrast, the MSFs calculated from NMR ensembles contain no rigid-body components, and therefore any calculated rigid-body contribution to the MSF indicates systematic error. Thus, by including rigid-body motions in fitting to NMR MSFs, we can benchmark the error in fitting B factors using the present method.

The best fit to experimental MSFs is achieved using rigid-body motions alone for 29 (45%) of the isolated X-ray proteins and 31 (48%) of the NMR proteins. This surprising result indicates that MSFs can frequently be interpreted without any internal motions whatsoever, even if there truly is no rigid component, as is the case with NMR. We fit the spring constants for each protein using 10,000 random combinations of rigid-body motions (see Methods) and find that, for many proteins, certain combinations of rigid-body motions cause

all the springs to become completely stiff. Among the isolated X-ray structures, we find only two proteins (1svn, 2fe5) for which no sampled combination of rigid motions results in an infinite force constant; 21 such proteins are found among the NMR structures. Thus, X-ray B factors are almost exclusively best accounted for with a combination of rigid-body and internal motions, whereas NMR MSFs may often be best accounted for using internal motions alone.

For the best-fit model, that is, the one that gives the lowest distance as defined by Eq. 5, the fraction of the MSF attributed to rigid-body motions was calculated as $\sum_{k=1}^{m} c_k / \sum_{i=1}^{N} \langle \Delta \mathbf{R}_i^2 \rangle$ (see Eq. 8). In the 29 NMR proteins for which the best fit was obtained with a combination of rigid and internal motions, the average contribution of rigid-body motions to MSFs was 59%. Of the 43 X-ray structures that were best fit using rigid and internal motions, rigid modes accounted for an average of 91% of the MSFs. These results indicate that rigid body motions, when included in the fit to MSFs, contribute less to NMR fluctuations than they do to X-ray fluctuations. As we know that NMR MSFs contain no rigid component, we would expect NMR and X-ray results to be similar if B factors also contained no rigid components. This is not the case (the K-S probability that the two are from the same distribution is $7.6 \cdot 10^{-7}$), in agreement with previous observations that B factors are composed of lattice vibrations as well as internal dynamics of proteins.

This trend is reflected in correlations between MSFs calculated from ANM and those derived from experimental data. Although minimizing the distance in Eq. 5 does not necessarily maximize the correlation between experimental and theoretical MSFs, the correlations are generally enhanced by including rigid motions. As shown in Table 1, rigid motions increase average correlations with B factors from 0.458 to 0.801. Similarly including rigid motions in fitting to NMR MSFs increases correlations from 0.598 to 0.720. The large increase in correlations with B factors upon inclusion of rigid motions further suggests that these motions play a more important role in B factors than they do in NMR MSFs.

The different role of rigid modes in B factors and NMR MSFs is partially explained by the shapes of the profiles. Typical results are illustrated in Fig. 2, which shows experimental and theoretical MSF profiles for three of the proteins in the set. Each panel shows results from X-ray (top graphs) and NMR (bottom graphs) structures for one protein. In comparison to experimental MSFs (black curves), the red curves show that ANM tends to produce excessive motion around flexible loops, as seen around residue 85 in 1bsy (Fig. 2A) and residue 92 in 1fil (Fig. 2C). This motion is damped in the crystal environment (green curves), which also tends to accentuate mobility in other regions, such as the helix in residues 54 to 61 of 1fil. Upon including rigid motions, the theoretical profiles (blue curves) fall much closer to their experimental counterparts for X-ray structures; for the NMR structures (bottom graphs, Fig. 2), the contribution of rigid modes does not generate as significant an improvement. The experimental MSF profiles calculated from X-ray and NMR data are qualitatively different: The NMR profiles are punctuated by concentrated regions of high local mobility, whereas the B factor derived mobilities are smoother and more gradually varying. Rigid-body motions, which also vary smoothly, are therefore expected to contribute more to B factors than to NMR MSFs.

This effect can be quantified using scaled entropies of MSFs. If the MSF of residue i is b_i, then the fraction of the total MSF that is contributed by residue i is $p_i = b_i / \sum_{j}^{N} b_j$. The scaled

entropy of the MSFs is $S = -\sum_{i}^{N} p_i \ln p_i / \ln N$, which ranges between 0 and 1. Rigid motions have the highest entropy (0.991), or are most uniform, followed by B factors (0.981), ANM MSFs (0.946) and NMR MSFs (0.782). Adding highly entropic rigid motions to ANM fluctuations will increase the entropy, providing better agreement with B factors but worse agreement with the low-entropy NMR fluctuations.

The effects of rigid motions are diminished when fitting to anisotropic temperature factors

When only internal motions are considered, the force constants calculated by fitting to X-ray ADPs are nearly identical to those calculated by fitting to isotropic B factors (see Fig. 3A), and those calculated using NMR-derived ADPs are very much like their counterparts calculated using NMR-derived MSFs (Fig. 3B). Such similarity is the expected result of fitting to internal motions alone: Changing the force constant does not alter the shapes of the thermal ellipsoids that are predicted by the internal ANM modes, but it affects their volumes. Altering the shapes of the thermal ellipsoids requires additional modes, such as those from an external source. In fitting internal modes to experimental ADPs, the volumes of the theoretical thermal ellipsoids are adjusted through the force constant to produce maximum overlap with those from experiment.

Upon inclusion of rigid modes, the force constants fit to X-ray ADPs increase by less than one order of magnitude, from 2.011±3.136 to 7.564±11.863 (Fig. 3C). This mild change is in sharp contrast to the dramatic increase in force constants observed upon including rigid motions in the fit to B factors (see Fig. 1C and Table 1). This effect is not universal: For two of the proteins studied (PDB codes 2igd and 2iim) no solution is found in which internal motions contribute favorably to the ADPs. Nonetheless, the contribution of rigid modes to ADPs is on the whole less than the apparent rigid contribution to B factors. Of the 24 proteins with ADP data, 12 were best fit to B factors using rigid-body motions alone. When the same set is fit to ADPs, only 4 proteins have their best agreement with experiment when only rigid-body motions are included. This reduction indicates that the spatial information included in ADPs favors internal motions, and not lattice vibrations. Furthermore, in 15 of these proteins, no basis set was found that could account for the experimental ADPs without some contribution of rigid motions. Finally, in the 20 X-ray structures that are best fit to ADPs using all modes, the contribution of the rigid modes decreases from 92% of the total MSF to 79%.

B factors are an uninformative measure of protein dynamics

The increased precision associated with fitting force constants to ADPs rather than to B factors partially stems from the inclusion of directional information in ADPs. Unlike B factors, ADPs are fit well only by motions that capture the favored directions of motion of all atoms, and they are therefore more selective. B factors, on the other hand, indicate the size, but not the shape, of each atom's motion relative to the other atoms. Any model that produces fluctuations of the proper magnitude will provide a good fit to B factors, regardless of the directions of the atomic motions. Fig. 4 illustrates how the excessive use of rigid modes to correctly fit B factors actually produce incorrect dynamics. This flexibility is what allows calculated values of force constants to vary wildly, depending on the assumed form of lattice vibrations. If one set of lattice vibrations produces MSFs that agree well with the B factors, then the internal motions will be minimized, resulting in a large value of the force constant. Equivalently, if a model's internal dynamics alone produce MSFs that agree very well with B factors, then the contribution of the rigid modes will be negligible. The large variations in force constants that are fit to B factors therefore results in part to the mediocre agreement between ANM dynamics and B factors.

Interestingly, even though rigid motions play a dominant role in fitting to isotropic B factors, they play only a minimal role in fitting to anisotropic temperature factors. As a result, the force constants that are calculated by fitting B factors using internal motions alone correlate well with those calculated by fitting ADPs using both internal and rigid motions. Accordingly, B factors fit in the absence of internal motions provide better estimates of the true force constants than do B factors fit using internal motions.

Although rigid motions account in part for the discrepancy between X-ray and NMR-inferred dynamics, a second source of uncertainty is the bias of the NMR ensembles. A key strength of solution NMR is its ability to directly probe protein dynamics [35;36] and various methods exist for inferring native state protein dynamics from NMR ensembles [37–42]. The agreement between covariances calculated from NMR ensembles and covariances of multiple crystal structures [42–45], B factors [28;46] and MD trajectories [47], suggest that NMR ensembles are a valid source of dynamical information; however, they also contain information on the uncertainty of the experimental method. Unlike the effects of rigid motion on crystal structures, the variations caused by structural uncertainties cannot be separated from those caused by dynamics. The present lack of agreement between dynamics inferred from X-ray and NMR hints that NMR ensembles may not be an optimal source of data on protein dynamics.

Although the force constants calculated from NMR MSFs and B factors do not agree well, they are on the same order of magnitude when rigid-body motions are neglected, and some of the residual variation can be taken up by the difference in temperatures at which NMR and crystallography are performed. It is easy to calculate the acceptable extent of motion predicted by the ANM. Assuming thermal fluctuations of $k_B T$, half of which may be reasonably contributed toward potential energy, the extent of motion along a single ANM mode is $d_{ANM} = \left(\dfrac{T}{T_0 \lambda \gamma} \right)^{1/2}$, where d_{ANM} is in Ångstroms, T is physiological temperature, T_0 is the experimental temperature, λ is the ANM mode eigenvalue and γ is the force constant as reported in Table 1. Considering the slowest mode of the protein (1zw6) that is shown in Fig. 4 as an example, we have $\lambda=1.37$, $\gamma_{NMR}=0.317$ and $\gamma_{Xray}=3.029$. Using $T_0=100K$ for X-ray and $T_0=300K$ for NMR, we have $d_{NMR}=1.517$Å and $d_{Xray}=0.850$Å. Rare events requiring larger displacements may be postulated, but one should note that the calculated force constants provide a somewhat small upper limit on the extent of ANM fluctuations.

Conclusion

We set out to find whether crystal contacts or rigid-body motions could bridge the gap between ENM force constants calculated with X-ray B factors and those calculated from NMR ensembles. We found that, whereas crystal contacts have only a small influence on force constants, rigid body motions by themselves can account for a significant fraction of B factors. In some cases, the contributions of rigid modes overshadowed those of internal dynamics. The influence of rigid modes on force constants is significantly reduced when fitting to X-ray ADPs instead of B factors. Regardless of the form of fitting, no agreement is found between force constants fit to separately to X-ray and NMR data.

The ability of rigid modes to account for B factors complicates the calculation of force constants. The smallest force constant for a given molecule is that which is fit using internal modes alone, and the largest is the infinite force constant found when only rigid motions are used in the fit. For most of the proteins in the set, the best-fit force constant is somewhere in between; however, this fit is based on our choice of metric (Eq. 5). We should not necessarily assume that the combination of motions that gives the smallest distance between experimental and theoretical MSFs also gives the best force constant. Indeed, it gives the

best force constant using this model, but small errors in the theoretical internal modes or in the experimentally determined MSFs impact the force constant non-linearly, making it sensitive to both the experiment and the model.

As a final note, we caution investigators against placing too much emphasis on correlations with MSFs. This study demonstrates that properly selected rigid-body motions can provide high correlation with experimental measures of MSFs; the implication is that improperly selected internal motions may also produce high correlations with MSFs. The root of this issue is that we are using *N*-dimensional data to fit dynamics from a 3*N*-dimensional space, so multiple solutions can be expected. We suggest that, whenever possible, ENM parameters should be fit to descriptive data such as anisotropic fluctuations or the complete matrix of dynamical covariances.

References

1. Ma J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure. 2005; 13 (3):373–380. [PubMed: 15766538]

2. Tama F, Brooks C. Symmetry, form, and shape: Guiding principles for robustness in macromolecular machines. Ann Rev Biophys Biomolecular Struct. 2006; 35:115–133.

3. 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 (8):2760–2767. [PubMed: 16891367]

4. Bahar I, Lezon TR, Bakan A, Shrivastava IH. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. Chem Rev. 2010; 110 (3):1463–1497. [PubMed: 19785456]

5. Bahar I, Atilgan A, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Folding & Des. 1997; 2 (3):173–181.

6. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins. 1998; 33 (3):417–429. [PubMed: 9829700]

7. Hinsen K, Thomas A, Field M. Analysis of domain motions in large proteins. Proteins. 1999; 34 (3): 369–382. [PubMed: 10024023]

8. Doruker P, Atilgan A, Bahar I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to alpha-amylase inhibitor. Proteins. 2000; 40 (3):512–524. [PubMed: 10861943]

9. Atilgan A, Durell S, Jernigan R, Demirel M, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J. 2001; 80 (1):505–515. [PubMed: 11159421]

10. Maragakis P, Karplus M. Large amplitude conformational change in proteins explored with a plastic network model: Adenylate kinase. J Mol Biol. 2005; 352 (4):807–822. [PubMed: 16139299]

11. Chu JW, Voth GA. Coarse-grained free energy functions for studying protein conformational changes: A double-well network model. Biophys J. 2007; 93 (11):3860–3871. [PubMed: 17704151]

12. Yang Z, Majek P, Bahar I. Allosteric transitions of supramolecular systems explored by network models: Application to chaperonin groel. PLoS Comp Biol. 2009; 5(4)

13. 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 (9):3424–3435. [PubMed: 18223005]

14. Yang Z, Bahar I, Widom M. Vibrational dynamics of icosahedrally symmetric biomolecular assemblies compared with predictions based on continuum elasticity. Biophys J. 2009; 96 (11): 4438–4448. [PubMed: 19486668]

15. Sonne J, Kandt C, Peters GH, Hansen FY, Jensen MO, Tieleman DP. Simulation of the coupling between nucleotide binding and transmembrane domains in the atp binding cassette transporter btucd. Biophys J. 2007; 92 (8):2727–2734. [PubMed: 17208973]

16. Isin B, Schulten K, Tajkhorshid E, Bahar I. Mechanism of signal propagation upon retinal isomerization: Insights from molecular dynamics simulations of rhodopsin restrained by normal modes. Biophys J. 2008; 95 (2):789–803. [PubMed: 18390613]

17. Cruickshank D. The analysis of the anisotropic thermal motion of molecules in crystals. Acta Crystallogr, B. 1956; 24:754–756.

18. Schomaker V, Trueblood K. On the rigid-body motion of molecules in crystals. Acta Cryst. 1968; B24:63–76.

19. Kidera A, Gō N. Normal mode refinement: crystallographic refinement of protein dynamic structure i. theory ant test by simulated diffraction data. J Mol Biol. 1992; 225:457–475. [PubMed: 1593630]

20. Kidera A, Inaka K, Matsushima M, Gō N. Normal mode refinement: crystallographic refinement of protein dynamic structure ii. application to human lysozyme. J Mol Biol. 1992; 225:477–486. [PubMed: 1593631]

21. Kundu S, Melton J, Sorensen D, Phillips G. Dynamics of proteins in crystals: Comparison of experiment with simple models. Biophys J. 2002; 83 (2):723–732. [PubMed: 12124259]

22. Song G, Jernigan RL. vgnm: A better model for understanding the dynamics of proteins in crystals. J Mol Biol. 2007; 369 (3):880–893. [PubMed: 17451743]

23. 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 (13):I175–I184. [PubMed: 17646294]

24. Hinsen K. Structural flexibility in proteins: impact of the crystal environment. Bioinformatics. 2008; 24 (4):521–528. [PubMed: 18089618]

25. Riccardi D, Cui Q, Phillips GN Jr. Application of elastic network models to proteins in the crystalline state. Biophys J. 2009; 96 (2):464–475. [PubMed: 19167297]

26. 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(2)

27. Bahar, I.; Lezon, TR.; Yang, LW.; Eyal, E. Ann Rev Biophys, vol 39, vol 39 of Annual Review of Biophysics. 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0897 USA: Ann. Rev; 2010. Global dynamics of proteins: Bridging between structure and function; p. 23-42.

28. 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 (6):741–749. [PubMed: 17562320]

29. Hinsen K, Petrescu A, Dellerue S, Bellissent-Funel M, Kneller G. Harmonicity in slow protein dynamics. Chem Phys. 2000; 261 (1–2, SI):25–37.

30. Ming D, Wall M. Allostery in a coarse-grained model of protein dynamics. Phys Rev Lett. 2005; 95(19)

31. Moritsugu K, Smith JC. Coarse-grained biomolecular simulation with reach: Realistic extension algorithm via covariance hessian. Biophys J. 2007; 93 (10):3460–3469. [PubMed: 17693469]

32. Lezon TR, Bahar I. Using entropy maximization to understand the determinants of structural dynamics beyond native contact topology. PLoS Comp Biol. 2010; 6(6)

33. Garbuzynskiy S, Melnik B, Lobanov M, Finkelstein A, Galzitskaya O. Comparison of x-ray and nmr structures: Is there a systematic difference in residue contacts between x-ray and nmr-resolved protein structures? Proteins. 2005; 60 (1):139–147. [PubMed: 15856480]

34. Kim M, Jernigan R, Chirikjian G. An elastic network model of hk97 capsid maturation. J Struct Biol. 2003; 143 (2):107–117. [PubMed: 12972347]

35. Kruschel D, Zagrovic B. Conformational averaging in structural biology: issues, challenges and computational solutions. Mol Biosyst. 2009; 5 (12):1606–1616. [PubMed: 20023721]

36. Mittermaier AK, Kay LE. Observing biological dynamics at atomic resolution using nmr. Trends Biochem Sci. 2009; 34 (12):601–611. [PubMed: 19846313]

37. Spronk C, Nabuurs S, Bonvin A, Krieger E, Vuister G, Vriend G. The precision of nmr structure ensembles revisited. J Biomol NMR. 2003; 25 (3):225–234. [PubMed: 12652134]

38. Lindorff-Larsen K, Best R, DePristo M, Dobson C, Vendruscolo M. Simultaneous determination of protein structure and dynamics. Nature. 2005; 433 (7022):128–132. [PubMed: 15650731]

39. Rieping W, Habeck M, Nilges M. Inferential structure determination. Science. 2005; 309 (5732): 303–306. [PubMed: 16002620]

40. Richter B, Gsponer J, Varnai P, Salvatella X, Vendruscolo M. The mumo (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. J Biomol NMR. 2007; 37 (2):117–135. [PubMed: 17225069]

41. Laughton CA, Orozco M, Vranken W. Coco: A simple tool to enrich the representation of conformational variability in nmr structures. Proteins. 2009; 75 (1):206–216. [PubMed: 18831040]

42. Lange OF, Lakomek NA, Fares C, Schroeder GF, Walter KFA, Becker S, Meiler J, Grubmueller H, Griesinger C, de Groot BL. Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution. Science. 2008; 320 (5882):1471–1475. [PubMed: 18556554]

43. Bakan A, Bahar I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. Proc Natl Acad Sci USA. 2009; 106 (34): 14349–14354. [PubMed: 19706521]

44. Friedland GD, Lakomek NA, Griesinger C, Meiler J, Kortemme T. A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. PLoS Comp Biol. 2009; 5(5)

45. Liu L, Koharudin LMI, Gronenborn AM, Bahar I. A comparative analysis of the equilibrium dynamics of a designed protein inferred from nmr, x-ray, and computations. Proteins. 2009; 77 (4): 927–939. [PubMed: 19688820]

46. Yang LW, Eyal E, Bahar I, Kitao A. Principal component analysis of native ensembles of biomolecular structures (pca_nest): insights into functional dynamics. Bioinformatics. 2009; 25 (5):606–614. [PubMed: 19147661]

47. Abseher R, Horstink L, Hilbers C, Nilges M. Essential spaces defined by nmr structure ensembles and molecular dynamics simulation show significant overlap. Proteins. 1998; 31 (4):370–382. [PubMed: 9626697]
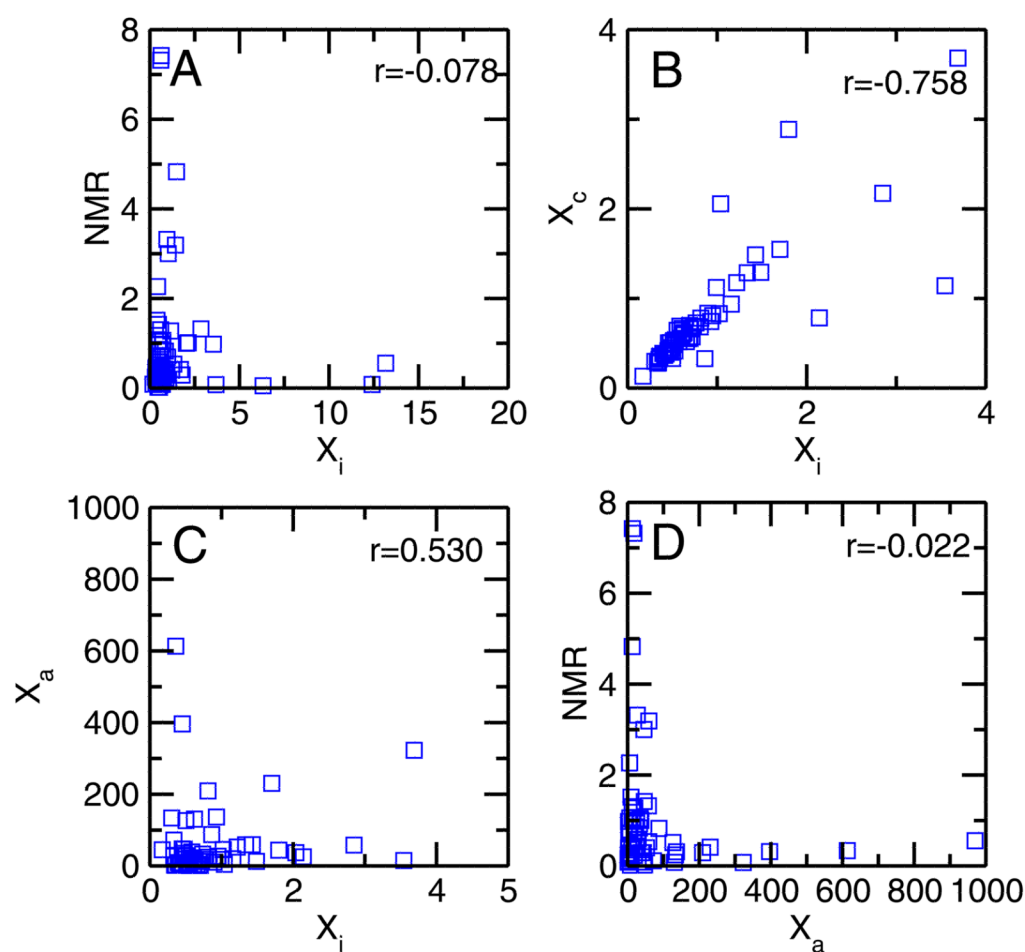
Figure 1.
Scatter plots of force constants fit to various experimental values. When fitting to X-ray B factors, X_i indicates only internal modes are used, X_a indicates that all modes are used and X_c indicates that crystal contacts are used. Force constants fit using NMR MSFs do not correlate with those calculated using X-ray B factors (A and D), and the only significant correlation is seen between X_i and X_c (B).
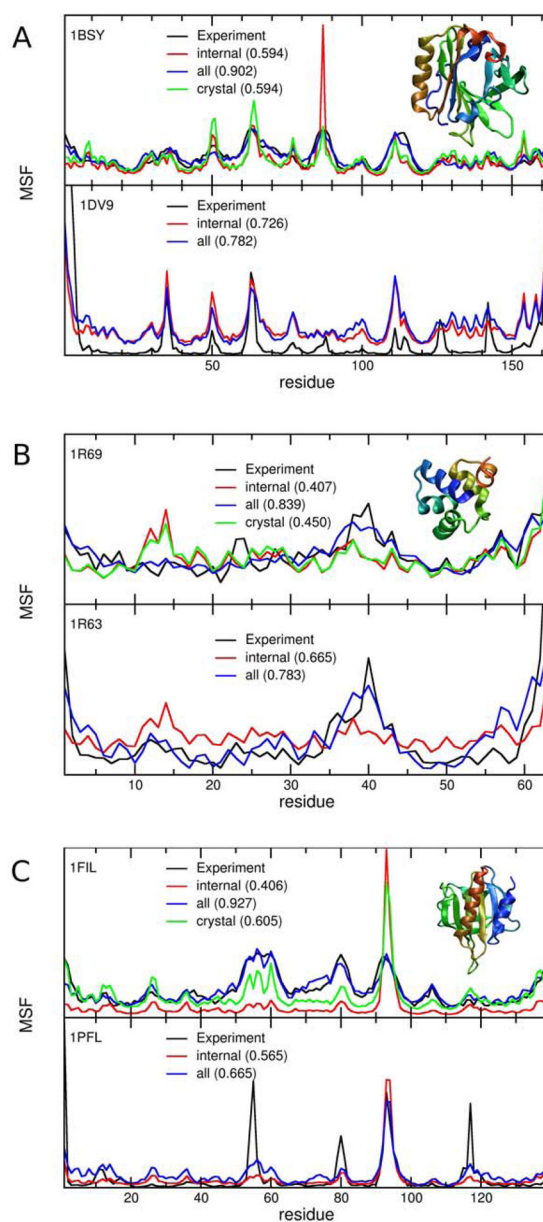
Figure 2.
Profiles of MSFs calculated using different models for three proteins. See text for details.
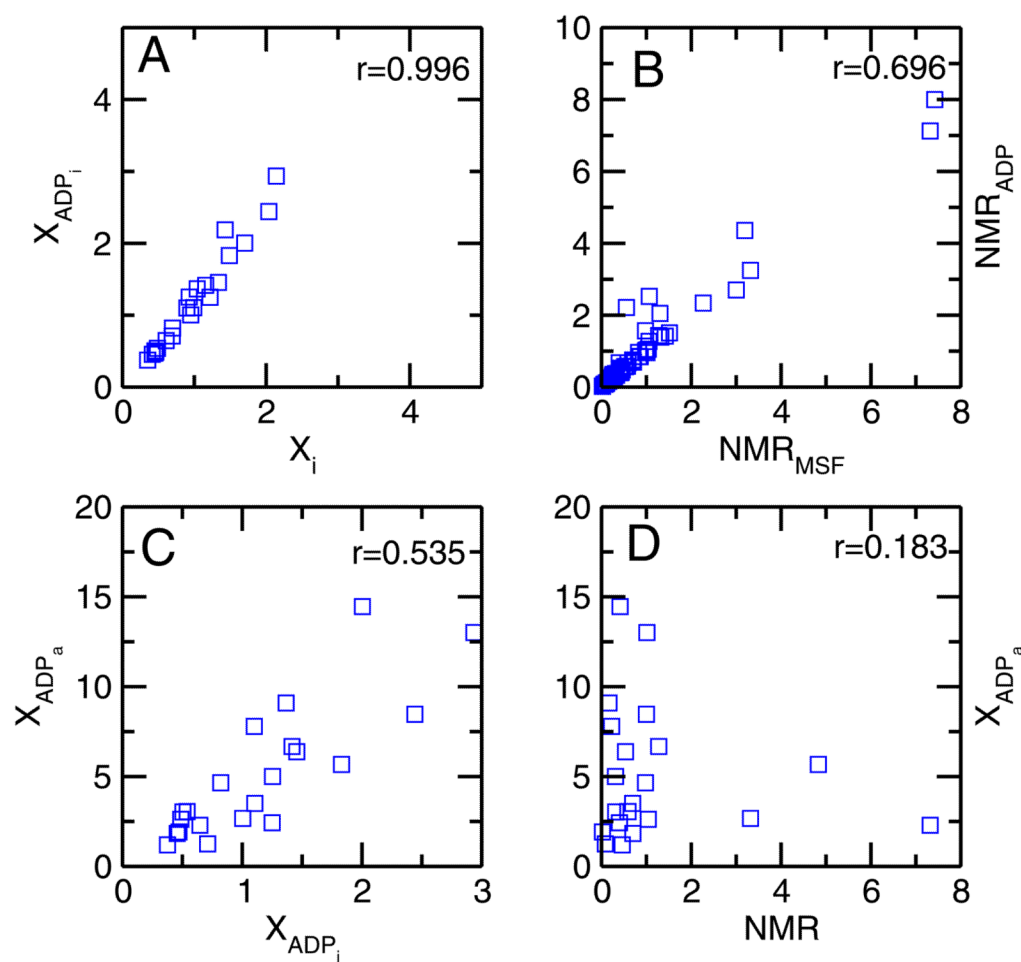
Figure 3.
Scatter plots of force constants calculated using ADPs. When internal modes only are used, force constants fit to X-ray ADPs are nearly identical to those fit to B factors (A); similarly for NMR (B). The force constants best fit to ADPs when all modes are considered show a mild correlation with those fit using only internal motions (C), indicating that rigid motions are less prominent in ADPs than they appear to be in B factors. There remains no substantial correlation between force constants fit to NMR MSFs and those fit using X-ray ADPs (D).
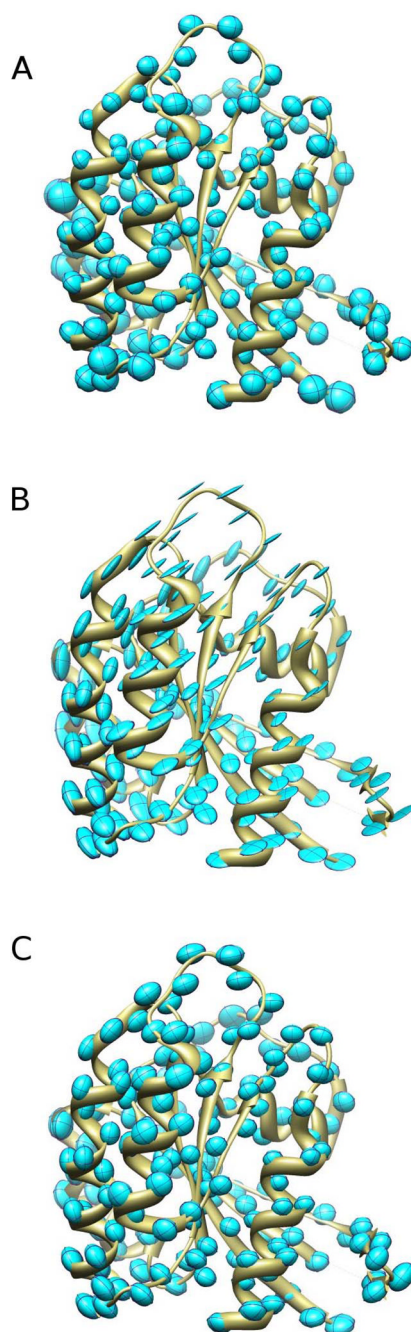
Figure 4.
An illustrative example of the overestimation of rigid motions to B factors. (A) The experimentally determined ADPs of protein 1zw6 appear to be spherical, but the B factors are calculated by summing the principal axes, discarding any directional information. (B) The best-fit model to B factors shows a large rotational component that flattens the thermal ellipsoids on the right to increase the volume of those on the left. (C) When the ENM motions are fit to the ADPs, the directional information retains its relevance, and the thermal ellipsoids better represent those taken from the PDB file.

Table 1

The structures used in this study, the calculated force constants and the correlations with experiment.

PDB Code		Force Constant (γ)				Pearson Correlation		
NMR	Xray	NMR	X_i	X_a	X_{ADP}	NMR	X_i	X_a
1e8j	2dsx	0.535	1.336	59.1	6.379	0.633	0.260	0.740
2ql0	1rb9	0.399	1.218	52.7	2.431	0.736	0.441	0.851
1m8c	2gkr	0.311	0.928	137.	5.002	0.965	−0.158	0.423
1tur	2ovo	1.065	0.720	12.9	-	0.959	0.436	0.874
3gb1	1pgb	7.422	0.631	13.3	-	0.176	0.613	0.805
1awo	1abq	0.516	0.503	127.	-	0.702	0.501	0.877
1x3q	3deo	0.167	1.039	4.652	9.089	0.602	0.703	0.866
1aey	1shg	0.439	0.333	27.1	-	0.611	0.583	0.875
1f2g	1fxd	0.836	0.735	33.2	-	0.463	0.577	0.887
1kun	1kth	0.224	0.896	10.4	7.789	0.734	0.390	0.822
2igh	2igd	0.080	12.4	-	-	0.821	0.143	-
1fra	3ebx	1.322	2.849	58.2	-	0.591	0.191	0.817
1h92	2iim	0.077	3.688	323.	-	0.598	0.049	0.632
1ijc	1l94	3.190	1.428	59.3	59.426	0.182	0.141	0.804
1r63	1r69	0.577	0.509	28.1	-	0.665	0.407	0.838
3ci2	2ci2	0.284	0.658	4.768	-	0.788	0.757	0.940
1lqi	2asc	0.979	0.697	2.595	4.652	0.699	0.758	0.965
1nha	1i27	0.413	1.701	230.	14.458	0.462	0.586	0.718
1ikm	3il8	0.079	0.622	131.	-	0.748	0.635	0.928
3mef	1mjc	0.115	0.335	72.0	-	0.800	0.654	0.919
1k3g	1c75	3.319	0.950	27.0	2.669	0.533	0.442	0.868
2orc	1dl1	0.979	3.543	15.1	-	0.953	0.718	0.893
1cdn	3icb	0.234	0.765	18.5	-	0.690	0.315	0.803
1d3z	1ubi	0.554	13.2	970.	-	0.989	0.489	0.757
2pac	451c	0.292	0.811	209.	-	0.601	0.342	0.645
1hdn	1cm2	0.428	0.692	12.1	-	0.293	0.079	0.551

PDB Code		Force Constant (γ)				Pearson Correlation		
NMR	Xray	NMR	X_i	X_a	X_{ADP}	NMR	X_i	X_a
2abd	1hb6	0.421	0.818	24.3	-	0.788	0.036	0.936
1ced	1ctj	0.461	0.350	3.154	1.198	0.467	0.560	0.936
1afh	1mzl	0.337	0.361	613.	-	0.433	0.319	0.507
1bmw	1who	0.235	0.481	19.0	-	0.681	0.696	0.898
2awt	1wm3	0.021	0.451	47.3	1.915	0.312	0.389	0.880
1c54	1t2i	1.275	1.157	20.1	6.670	0.279	0.707	0.893
1jnj	1lds	0.291	1.799	44.1	-	0.544	0.292	0.841
2lfb	1lfb	0.052	6.318	4393	-	0.390	0.469	0.823
1ot4	2c9q	0.293	0.584	39.0	-	0.516	−0.280	0.658
1ygw	4rnt	0.440	0.690	3.341	-	0.497	0.583	0.826
1trv	2hsh	7.319	0.608	17.6	2.291	0.646	0.652	0.841
4trx	1erv	1.514	0.395	9.454	-	0.481	0.525	0.702
1it1	2cdv	3.002	1.023	45.7	-	0.818	0.393	0.863
1xoa	2tir	0.861	0.572	22.5	-	0.283	0.354	0.720
1bnr	1rnb	0.243	0.304	134.	-	0.919	0.401	0.801
1um7	2fe5	4.828	1.487	12.9	5.674	0.773	0.703	0.855
1kot	1gnu	1.419	0.478	46.5	-	0.354	0.256	0.725
1ly7	1ekg	0.257	0.359	1.761	-	0.372	0.738	0.795
1bvm	1g4i	0.693	0.991	18.2	3.504	0.255	0.637	0.806
2aas	1kf3	1.011	2.141	25.6	13.009	0.923	0.566	0.833
3phy	1nwz	0.084	0.692	1.50	1.253	0.846	0.475	0.881
1djm	1chn	1.303	0.600	12.9	-	0.769	0.612	0.836
1e8l	2vb1	1.000	2.036	37.0	8.467	0.419	0.351	0.647
1blr	2fr3	0.589	0.487	29.5	3.046	0.283	0.559	0.890
1pfl	1fil	0.825	0.863	87.6	-	0.565	0.406	0.928
2def	1dff	2.266	0.430	5.680	-	0.651	0.620	0.902
1vre	1jf4	0.712	0.609	3.223	-	0.452	0.383	0.624
1jor	1ey4	0.395	0.344	2.671	-	0.794	0.708	0.855

PDB Code		Force Constant (γ)				Pearson Correlation		
NMR	Xray	NMR	X_i	X_a	X_{ADP}	NMR	X_i	X_a
1myf	1duk	0.190	0.512	5.244	-	0.380	0.371	0.530
1eq0	1hka	1.058	0.528	6.353	-	0.849	0.758	0.904
1btv	1bv1	0.260	0.457	9.563	-	0.345	0.322	0.599
1ax3	1gpr	0.017	0.534	8.030	-	0.498	0.608	0.822
1dv9	1bsy	0.087	0.172	45.2	-	0.726	0.595	0.902
1oca	1w8m	0.699	0.414	6.535	1.833	0.429	0.420	0.632
1crp	1zw6	0.317	0.451	396.	3.029	0.789	0.366	0.880
1yho	1kmv	1.041	0.463	35.0	2.619	0.634	0.419	0.844
1dgq	1mjn	0.160	0.704	1.184	-	0.359	0.594	0.833
1ah2	1svn	0.112	0.553	1.210	-	0.739	0.680	0.737
μ		0.952	1.33	141.	7.564	0.598	0.458	0.801
σ		1.445	2.272	561.	11.863	0.211	0.219	0.118