

# The Gaussian Network Model: Precise Prediction of Residue Fluctuations and Application to Binding Problems

Burak Erman

Department of Chemical and Biological Engineering, Koç University, Sariyer, 34450 İstanbul, Turkey

**ABSTRACT** The single-parameter  $\Gamma$  matrix of force constants proposed by the Gaussian Network Model (GNM) is iteratively modified to yield native state fluctuations that agree exactly with experimentally observed values. The resulting optimized  $\Gamma$  matrix contains residue-specific force constants that may be used for an accurate analysis of ligand binding to single or multiple sites on proteins. Bovine Pancreatic Trypsin Inhibitor (BPTI) is used as an example. The calculated off-diagonal elements of the  $\Gamma$  matrix, i.e., the optimized spring constants, obey a Lorentzian distribution. The mean value of the spring constants is  $\sim -0.1$ , a value much weaker than  $-1$  of the GNM. Few of the spring constants are positive, indicating repulsion between residues. Residue pairs with large number of neighbors have spring constants around the mean,  $-0.1$ . Large negative spring constants are between highly correlated pairs of residues. The fluctuations of the distance between anticorrelated pairs of residues are subject to smaller spring constants. The importance of the number of neighbors of residue pairs in determining the elements of the  $\Gamma$  matrix is pointed out. Allosteric effects of binding on a single or multiple residues of BPTI are illustrated and discussed. Comparison of the predictions of the present model with those of the standard GNM shows that the two models agree at lower modes, i.e., those relating to global motions, but they disagree at higher modes. In the higher modes, the present model points to the important contributions from specific residues whereas the standard GNM fails to do so.

## INTRODUCTION

Residues of a protein in the native state exhibit large-scale fluctuations about their equilibrium positions. The extent of the fluctuations of a given residue depends predominantly on the number of its closest spatial neighbors. The mean-square fluctuation of a residue is, in general, smaller than that of another residue with a smaller number of neighbors. This observation forms the basis of the Gaussian network model (GNM) of proteins (1), which predicts the residue fluctuations in native proteins, in simple analogy with fluctuations of junction positions in Gaussian elastomeric networks (2). The fact that the size of the fluctuation domain of a junction in a Gaussian network varies inversely with the number of other junctions that share this domain is now well established (3), and serves as a plausible analogy for the protein fluctuations. The second simplifying assumption of the GNM was based on an earlier postulate (4) that because of the central limit theorem, the large-scale fluctuations of residues could be characterized by a single-parameter Gaussian energy function. According to this approximation, all the  $C^\alpha$ - $C^\alpha$  interactions as well as the strength of the covalent bonds are assumed identical. The simplification introduced by adopting a single-parameter representation of fluctuations by Tirion (4) and Bahar et al. (1) is notable. Several articles (5–10) following the original GNM article showed that a simple harmonic potential with a single interaction-parameter indeed captures the basic physics underlying the equilibrium fluctuations in proteins. However, a closer and more careful

examination of the articles comparing experiment with GNM predictions (7,9,10) shows that if the single-parameter potential is replaced by a potential that somehow reflects the environment of a given residue in more detail, the agreement between theory and experiment will be further improved. The specific aim of this article is to introduce a simple method for calculating the environment-dependent interaction parameters for a protein when its B-factors are given. We do this by iteratively modifying the residue-residue interaction parameters, until the recalculated  $\Gamma$  matrix of the system yields the experimentally observed mean-square residue fluctuations. The starting  $\Gamma$  matrix is that of the GNM. The parameters of the optimized  $\Gamma$  matrix then give the strength of the residue-specific pairwise interactions, which are corrections to the single-parameter GNM.

Micheletti et al. (8) used a self-consistent Gaussian model to study the equilibrium behavior of proteins in which the pairwise interactions are not equivalent, but are amino-acid specific. Starting with a Hamiltonian similar to that of the GNM, they introduced an iterative self-consistent approach for calculating the equilibrium probabilities of the contacting pairs of residues. Their work clearly shows the importance of differentiating pairwise interactions in a protein in the native state and gives the method of calculation. The approach of this article is similar to their iterative method.

In the next section, we critically review the GNM and point out to what may be missing in the model. In Theory, below, we describe the computational scheme for reevaluating the spring constants to match experimental data. As an application, we determine the spring constants for the protein Bovine Pancreatic Trypsin Inhibitor (BPTI; PDB code No. 5PTI) and present an extensive discussion of the

*Submitted June 9, 2006, and accepted for publication August 8, 2006.*

Address reprint requests to B. Erman, Tel.: 90-212-338-1704; E-mail: berman@ku.edu.tr.

© 2006 by the Biophysical Society

0006-3495/06/11/3589/11 \$2.00

doi: 10.1529/biophysj.106.090803

residue-specific spring constants that lead to precise description of the observed B-factors. The present model gives a consistent theoretical description of the B-factors. A precise and consistent description of the fluctuations in proteins is of great consequence for a quantitative understanding of protein function, ligand binding, and protein-protein interactions. We discuss different possible applications of the model using the optimized values of the spring constants.

## THEORY

### Review of the Gaussian theory of fluctuations in native proteins

The equilibrium fluctuations in a protein are related to the experimentally measured Debye-Waller factors, also referred to as the temperature or B-factors, by the relation

$$\langle \Delta R_i^2 \rangle = \frac{3}{8\pi^2} B_i, \quad (1)$$

where  $\langle \Delta R_i^2 \rangle$  is the mean-square fluctuation of the  $i^{\text{th}}$  residue and  $B_i$  is its Debye-Waller factor in  $\text{\AA}^2$ . The Hamiltonian,  $H$ , for the native protein is usually assumed to consist of Lennard-Jones type pair interactions

$$\beta H = \frac{\beta}{2} \sum_{i>j} E_{ij} \left[ \left( \frac{\bar{R}_{ij}}{R_{ij}} \right)^{12} - 2 \left( \frac{\bar{R}_{ij}}{R_{ij}} \right)^6 \right], \quad (2)$$

where  $R_{ij}$  and  $\bar{R}_{ij}$  are the instantaneous and time-averaged distances between the  $i^{\text{th}}$  and  $j^{\text{th}}$  residues,  $\beta = 1/kT$ ,  $k$  and  $T$  being the Boltzmann constant and the absolute temperature, respectively. The value  $E_{ij}$  is the energy parameter for the  $ij^{\text{th}}$  pair, a positive quantity for attractive interactions. Expanding Eq. 2 in Taylor's series and keeping the first two terms leads to the Gaussian approximation

$$\beta H = -\frac{\beta}{2} \sum_{i>j} E_{ij} + \frac{\beta}{2} \sum_{i>j} \left( \frac{36E_{ij}}{\bar{R}_{ij}^2} \right) (R_{ij} - \bar{R}_{ij})^2. \quad (3)$$

Replacing  $(R_{ij} - \bar{R}_{ij})^2$  by the equivalent expression  $(\Delta R_i - \Delta R_j)^2$  in Eq. 3, where  $\Delta R_i$  is the instantaneous fluctuation of the  $i^{\text{th}}$  residue from its time-averaged position, the Hamiltonian may be recast into the form

$$\beta H = \frac{\beta}{2} H_0 + \frac{\beta}{2} \Delta \mathbf{R}^T \mathbf{\Gamma} \Delta \mathbf{R}, \quad (4)$$

where  $\Delta \mathbf{R}$  is the column vector of  $\Delta R_i$  values,  $H_0 = -\sum_{i>j} E_{ij}$ , and  $\mathbf{\Gamma}$  is given as

$$\Gamma_{ij} = \begin{cases} -\frac{36E_{ij}}{\bar{R}_{ij}^2} & i \neq j \\ \sum_k \frac{36E_{ik}}{\bar{R}_{ik}^2} & i = j \neq k \end{cases}, \quad (5)$$

the partition function for a protein of  $n$  residues may be written as

$$Z = \exp \left[ -\frac{\beta}{2} H_0 \right] \int \exp[-\Delta \mathbf{R}^T \mathbf{\Gamma} \Delta \mathbf{R}] d\{\Delta \mathbf{R}\} = C_0 \left( \frac{\pi^n}{\det \mathbf{\Gamma}} \right)^{3/2}, \quad (6)$$

where  $d\{\Delta \mathbf{R}\} \equiv d\Delta R_1 d\Delta R_2 \dots d\Delta R_n$ , and  $C_0 = \exp[-\frac{\beta}{2} H_0]$ .

The average quantity  $\langle \Delta R_i \cdot \Delta R_j \rangle$  is obtained from Eq. 6 according to the known operations (2) as

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \frac{3}{2} (\mathbf{\Gamma}^{-1})_{ij}. \quad (7)$$

The diagonal elements of Eq. 7 express the connection between fluctuations,  $\langle \Delta R_i^2 \rangle$ , and the residue-residue interaction energy parameters,  $E_{ij}$ , in the Gaussian approximation. Combining Eqs. 1 and 7 leads to

$$(\mathbf{\Gamma}^{-1})_{ii} = \frac{1}{4\pi^2} B_i. \quad (8)$$

The  $ij^{\text{th}}$  off-diagonal element of the matrix  $\mathbf{\Gamma}$  defined by Eq. 5 shows the strength of interaction between residues  $i$  and  $j$ . The matrix is simplified in the GNM by assuming that the energy parameter  $E_{ij}$  equates to a constant  $\gamma^*$  if the residues  $i$  and  $j$  are separated by less than a cutoff distance  $r_c$ , and to zero otherwise:

$$\Gamma_{ij} = \begin{cases} -\gamma^* & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_k \Gamma_{ij} & \text{if } i = j \neq k \end{cases}. \quad (9)$$

Defined in this manner, the off-diagonal elements of the  $\mathbf{\Gamma}$  matrix give the contact map of the native protein if the single-parameter  $\gamma^*$  is taken as unity. The single-parameter  $\gamma^*$  may be regarded as a weighting factor. It weights each contact equally in the  $\mathbf{\Gamma}$  matrix. The value of the  $i^{\text{th}}$  diagonal element of  $\mathbf{\Gamma}$  equates to the total number of its contacts, weighted with  $\gamma^*$ .

The  $\mathbf{\Gamma}$  matrix may be written as  $\mathbf{\Gamma} = \mathbf{D} + \mathbf{U}$ , where  $\mathbf{D}$  and  $\mathbf{U}$  are the matrices of the diagonal and off-diagonal elements, respectively. The inverse  $\mathbf{\Gamma}^{-1} = (\mathbf{D} + \mathbf{U})^{-1}$  may be written for small off-diagonal terms by Taylor series expansion up to the linear term in  $\mathbf{U}$  as

$$\mathbf{\Gamma}^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1} \mathbf{U} \mathbf{D}^{-1} + \dots \quad (10)$$

The diagonal component  $\mathbf{D}^{-1}$  shows the contribution of the local packing density to  $\mathbf{\Gamma}^{-1}$ . The second term,  $\mathbf{D}^{-1} \mathbf{U} \mathbf{D}^{-1}$ , shows the contributions resulting from positional correlations among different residue pairs. Thus, the off-diagonal terms carry information on the spatial connectivity of the protein. Depending on the strength of these latter correlations, the contributions of the off-diagonal terms of  $\mathbf{\Gamma}$  to the fluctuations may be significant. These effects are included in the GNM. Some time ago, Halle (7) proposed the local density model (LDM) where only the contribution of diagonal terms,  $\mathbf{D}^{-1}$ , are considered, and all pairs of nonhydrogen atoms within a cutoff distance are counted in  $\mathbf{D}$ . Accordingly, the mean-squared fluctuations of atoms are represented as

$$\langle \Delta R_i^2 \rangle = \frac{1}{D_i}. \quad (11)$$

Halle chose 38 nonhomologous proteins and showed that the LDM gives excellent agreement with the experimental values of the B-factors. However, a closer inspection of the LDM shows that whenever LDM is in good agreement with experiment, GNM also is in good agreement, and whenever LDM fails, GNM also fails. The main source of the failure of LDM relates to the absence of the off-diagonal contributions and to the choice of the same spring constant for all pair interactions, which in turn affects the contributions from the off-diagonal terms.

Recently, Kundu et al. (9) published an important article where they studied possible improvements in the GNM by using an extensive set of 113 proteins as their data. On the average, the predictions of the model in the form it was first proposed (1) gave satisfactory results. To obtain better agreement of the theory with experiment, they varied the spring lengths, including the possible interactions between proteins that are adjacent to one another in the crystal structure. With all these improvements, the best correlation coefficient that measures the relative agreement between B-factors and the GNM was 0.662. Although this correlation coefficient may be accepted as satisfactory for the complex systems at hand, it is rather low for quantitative use of the model. Further attempts to improve the comparisons by using an anisotropic version of the GNM failed, showing that directional correlations are not the significant factors affecting the considered variables. Kundu et al. (9) also showed that there are no large systematic contributions of lattice disorder to crystallographic B-factors.

Three factors, that may be important, are missing in the GNM. First, each spring constant  $\gamma_{ij}$  connecting two neighboring residues  $i$  and  $j$  is taken as equal. This is an oversimplification and deviations from this single-peaked distribution of spring constants may be significant. Secondly, the proteins are situated on a lattice and crystal packing effects are nonnegligible as shown by Kundu et al. (9). Thirdly, non-Gaussian or anharmonic effects may make nonnegligible contributions to the thermal fluctuations, and therefore B-factors, and therefore a purely harmonic model may underestimate the atomic fluctuations. The specific and practical aim of the present work is to express protein fluctuations precisely, by keeping the simple Gaussian structure and systematically readjusting or optimizing the spring constants. This is done by iteratively renormalizing the quadratic Hamiltonian. In this way a distribution of spring constants are obtained that lead to fluctuations that are in agreement with experimentally observed data. The second effect cited above, i.e., contributions from crystal packing effects are present, a posteriori, in the spring constants calculated with the present model. An increase in the number of neighbors of a surface residue coming from crystal packing decreases the fluctuations of that residue and this decrease is in turn represented in the present model by an increased value of the spring constant. Therefore, the effects of crystal packing are implicitly contained in the present model, accurately, if they lie in the harmonic range. Non-Gaussian effects constitute a problem of higher complexity, and currently it is not possible to incorporate such effects directly into the GNM or to any harmonic model. In this respect, the calculated spring constants are to be regarded as effective spring constants that are renormalized to reflect anharmonic effects using an harmonic model. It should be noted, however, that matching experimental data by adjusting the spring constants using an harmonic model may lead to systematic errors, and the present model should be considered with care in this respect. As an example, we cite the modal decomposition of fluctuation trajectories. Within the harmonic approximation the modes are uncoupled and energy imparted to any mode remains forever in that mode, whereas with an anharmonic potential, energy flows to other modes, as has been shown earlier (11,12).

## Determination of neighbor-dependent spring constants

### The model

The protein is represented in its  $C^\alpha$  form. The starting Hamiltonian of the iterative scheme is that of the GNM. The strength of the interaction between all covalently bonded pairs, i.e., the spring constants, of  $C^\alpha$ -values along the chain backbone is chosen as  $\gamma^*$ , and kept fixed throughout the iterations. The constancy of this bond strength follows (8) from the fact that the backbone bonds are formed at the outset and remain in that state at all times. The initial strength of interactions between all pairs of nonbonded residues that are within a cutoff distance of  $r_c$  is taken as  $\gamma^*$  and is varied at each iteration for each residue pair. A Monte Carlo renormalization scheme is employed for evaluating the Hamiltonian of the system iteratively. The iterative computational scheme starting with the single initial interaction parameter  $\gamma^*$  is as follows: The matrix  $\Gamma$  is formed according to

$$\Gamma_{ij} = \begin{cases} -\gamma^* & \text{if } |i-j| = 1 \\ -\gamma^* c_{ij} & \text{if } |i-j| > 1 \text{ and } R_{ij} < r_c \\ 0 & \text{if } |i-j| > 1 \text{ and } R_{ij} > r_c \\ -\sum_k \Gamma_{ik} & \text{if } i = j \neq k \end{cases} \quad (12)$$

In the first step, the  $c_{ij}$ -values are taken as unity (but they are modified in subsequent steps according to Eq. 13 below). The  $\Gamma$  matrix is then inverted and the diagonal elements of  $\Gamma^{-1}$  are compared with experimental  $\langle \Delta R_i^2 \rangle$  using Eq. 8. A residue  $i$  is then chosen randomly, and its interaction with all of its first neighbor residues (excluding the covalently bonded ones) is updated according to

$$\Gamma_{ij}(\text{new}) = \Gamma_{ij}(\text{old}) + \varepsilon \left[ (\Gamma^{-1}(\text{old}))_{ij} - \frac{1}{4\pi^2} B_i \right], \quad (13)$$

where  $\varepsilon$  is a small positive number, and  $j$  ( $|j-i| > 1$ ) goes from 1 to  $n$  (total number of residues). The  $\Gamma$  matrix is then symmetrized, and its new diagonal elements are calculated. The correction introduced in Eq. 13 modifies the spring constants, or the  $c_{ij}$ -values, between the  $i^{\text{th}}$  residue and all of its contacting neighbors,  $j$ . Upon inversion of  $\Gamma$ , the correction introduced to the  $ij$  pairs propagate to all residues that are affected by the fluctuations of the  $i^{\text{th}}$  residue. The iterative scheme outlined above is repeated in this manner, until the experimental and theoretical values of  $\langle \Delta R_i^2 \rangle$  converge. At the end of the iterations, a different value of the interaction parameter for each pair of contacting residue is obtained.

## RESULTS

### Evaluation of the modified interaction parameters for BPTI and comparison with experiment

Here, we apply the method of the preceding section to the widely studied protein Bovine Pancreatic Trypsin Inhibitor, BPTI, which has 58 residues. The choice of this protein is only because it is one of the most widely studied proteins and its native structure is known to within an RMSD of 1 Å.

In the calculations, the cutoff distance is taken as 7.0 Å. Iterative calculations were made according to Eq. 13 with  $\varepsilon = 0.01$ . Iterations were continued until the mean-squared error between the calculated and experimental B-factors reached a steady low value. Initially, the GNM gave a mean-squared error of 7 Å<sup>2</sup>. At the end of 5000 steps the mean-square deviation decreased to and remained at a steady value of 0.8 Å<sup>2</sup>. The value of the scaling factor  $\gamma^*$  was obtained as 18.14.

In Fig. 1 *a*, predictions of GNM are compared with experimental B factors. Although the fluctuation patterns of various domains are predicted well, there are significant deviations for individual residues. For example, the decrease in fluctuations in going from residue 1 to 5, the minimum about residue 10, the peak about residue 15, the minimum about residue 20, the peak around residue 40, and the two peaks around residues 48 and 54 are all predicted. However, individual peaks at residues 8, 14, 21, 26, 30, 38, 42, 47, 51, and 55 exhibit significant deviations from experiment. Normal mode decomposition of fluctuations to investigate structure-function relations, as have been the common practice in interpreting the GNM results, is most satisfactory in the low frequency modes relating to the domain motions. For the higher modes to be meaningful, precise agreement between experiment and theory is needed. This is established with the present model. In Fig. 1 *b* the agreement of the results of the model and experiment are clearly seen.

The spring constants are all equal in the GNM model, hence their distribution is a spike at  $-1$ . The present model transforms this distribution into a single-peaked Lorentzian. This is elucidated in Fig. 2. The ordinate in the figure shows the range of  $\gamma$ -values obtained as a result of the iterative procedure of the present model. The negative values correspond to attractive forces between the corresponding residues. The ordinate represents the fraction of the  $\gamma_{ij}$ -values corresponding to the indicated values of the abscissa. Some of the spring

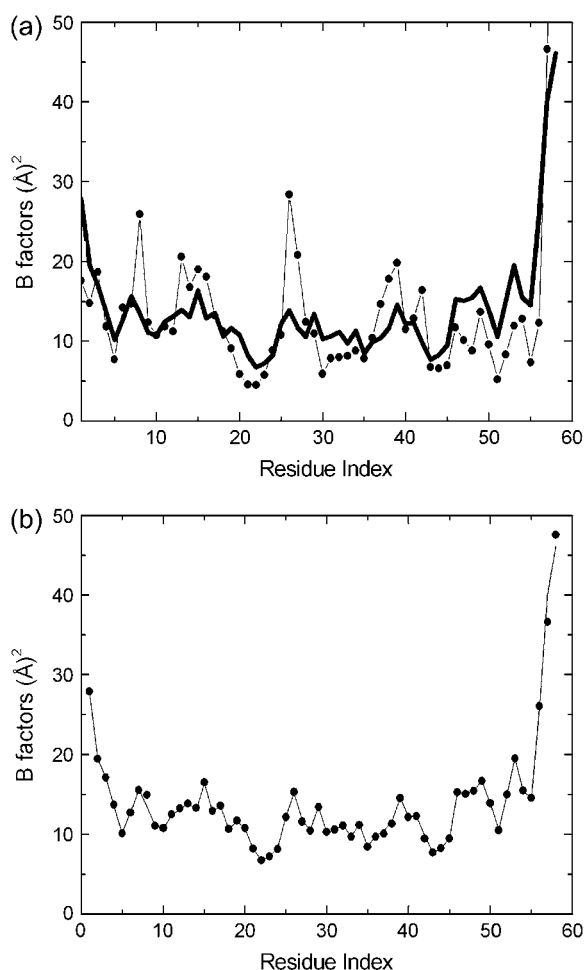


FIGURE 1 (a) Experimental B factors for BPTI (thick curve) compared with the GNM prediction (light curve and solid circles), (b) Experimental B factors (light curve), which are in agreement with calculated values (solid circles).

constants are positive, indicating repulsive forces between residue pairs that are spatially too close to each other. This last statement will be further discussed below. The solid curve is the best fitting Lorentzian that has the equation

$$f_{ij} = \frac{A}{2\pi} \left( \frac{\omega}{4(\gamma_{ij} - \gamma_c)^2 + \omega^2} \right), \quad (14)$$

where  $f_{ij}$  is the fraction of  $\gamma_{ij}$ -values, and the parameters  $A$ ,  $\gamma_c$ , and  $\omega$  are obtained for the fit shown in Fig. 2 as  $A = 0.426$ ,  $\gamma_c = -0.0887$ , and  $\omega = 0.174$ . Thus, the single peak at  $\gamma_{ij} = -1.0$  for the GNM is now shifted to  $\gamma_c = -0.0887$ , and the distribution is slightly diffused around this value as seen in the figure. Calculations for several other proteins along the same lines also transform the spring constant distribution into a Lorentzian. The Lorentzians for all the proteins studied may be superposed into a single curve with proper scaling. A detailed analysis of this feature is in progress in our lab.

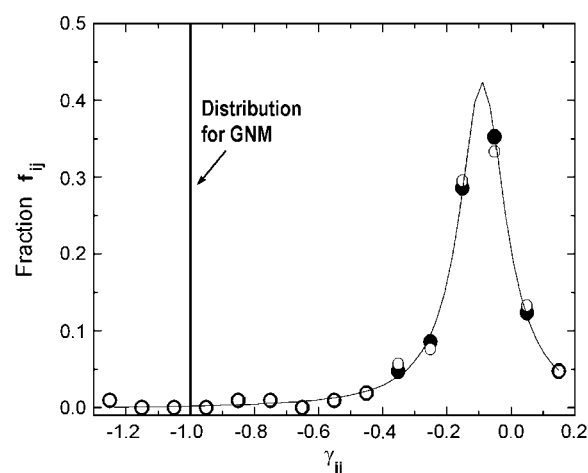


FIGURE 2 The fraction of  $\gamma_{ij}$ -values obtained as a result of the present iterative model.

The distribution shown in Fig. 2 is obtained by randomly choosing a residue and modifying its interactions with all of its neighbors. The set of spring constants obtained in this manner should be independent of the random choice of the residues. The solid points in Fig. 2 show results obtained with another initial choice of the spring constants. Only those points that exhibit sufficient deviation from the original distribution are visible as solid points in the figure, the others being essentially identical and masked by the original open circles. The values of  $\gamma_{ij}$  are presented in Table 1.

In Fig. 3 we compare the  $\gamma$ -values obtained by two different iterations. The random numbers used for choosing the residue pairs in the calculations were different in the two sets. The abscissa and ordinate labeled as  $\gamma_1$  and  $\gamma_2$ , respectively,

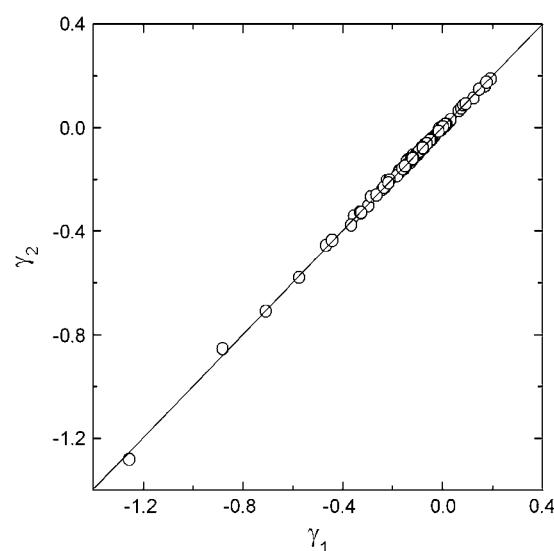


FIGURE 3 Comparison of the calculated spring constants  $\gamma_{ij}$  for two different Monte Carlo runs.

indicate the two sets of the parameters  $\gamma_{ij}$  obtained by the two independent runs. The points collapse perfectly on a 45° line that passes through the origin, indicating that the scheme is independent of the randomness inherent in the Monte Carlo scheme employed.

The dominant factor that leads to the Lorentzian distribution shown in Fig. 2 is the average number,  $n_{ij}$ , of residues in the domains of fluctuation of the residues  $i$  and  $j$ , defined as

$$n_{ij} = (n_i \cdot n_j)^{1/2}, \quad (15)$$

where  $n_i$  is the number of neighbors of residue  $i$ . In Fig. 4, the average number of residues  $n_{ij}$  are presented as a function of  $\gamma_{ij}$ . The shaded circles are obtained by counting the number of neighbors  $n_i$  and  $n_j$  that are within a cutoff distance of 7 Å of residue  $i$  and  $j$ , respectively, and using Eq. 15. The vertical dotted line shows the values of  $\gamma$  for the GNM and is drawn for reference. The solid vertical line locates the zero of  $\gamma$  and is drawn to guide the eye. The average number of junctions obtained by using Eq. 15 varies between 4 and 12. The calculated values of  $\gamma_{ij}$  are shifted to larger values, but still mostly negative, indicating that the attractive forces between pairs of residues are diminished relative to that of GNM. There are, however, few positive values that represent repulsive forces between pairs. Pairs of residues in crowded environments represented by large values of  $n_{ij}$  correspond to small values of  $\gamma_{ij}$ . Stated in another way, these pairs are weakly connected to each other. The solid circles are obtained by averaging the values of  $n_{ij}$  in a given interval of  $\gamma_{ij}$ . For negative values of  $\gamma_{ij}$ , averaging is done over equal intervals of 0.25. For positive values of  $\gamma_{ij}$ , the interval is taken as 0.05 since there are fewer points in the positive region and their range is smaller. The line connects the solid points to guide the eye. The peak of the curve representing the averages is  $\sim \gamma = -0.1$ .

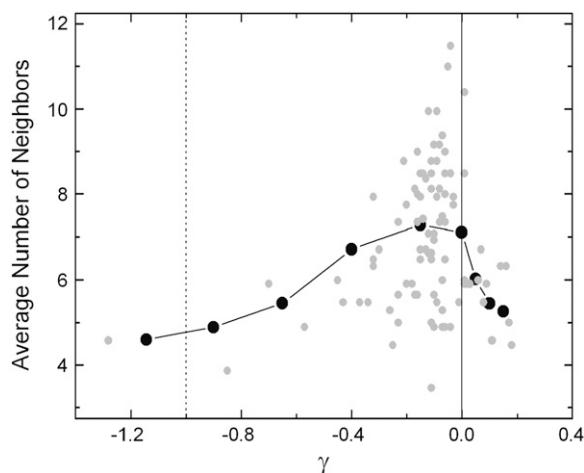


FIGURE 4 Relationship of  $\gamma_{ij}$ -values on the average number of neighbors of residues  $i$  and  $j$ .

The present model is applied to 12 different proteins of different sizes (B. Erman, unpublished) and the magnitudes of attractive spring constants are observed to scale as

$$\gamma_{ij} \sim \frac{1}{\langle \Delta R_{ij}^2 \rangle^m}, \quad (16)$$

where  $m$  is in the order of 1.6 and  $\langle \Delta R_{ij}^2 \rangle$  is the mean-square fluctuation of the distance between the  $i^{\text{th}}$  and  $j^{\text{th}}$  residues, defined in terms of the mean-square residue fluctuations as

$$\langle \Delta R_{ij}^2 \rangle = \langle \Delta R_i^2 \rangle - 2\langle \Delta R_i \cdot \Delta R_j \rangle + \langle \Delta R_j^2 \rangle. \quad (17)$$

The value of  $\langle \Delta R_{ij}^2 \rangle$  is affected by two factors: First, if the mean-square fluctuations of the residues are small, then  $\langle \Delta R_{ij}^2 \rangle$  will be small, leading to a large value of the spring constant. Thus, residues in crowded regions where  $\langle \Delta R_{ij}^2 \rangle$  are small, are joined by stiffer spring constants. Secondly, for residues in less crowded regions, for anticorrelated fluctuations, the dot product in the middle term in Eq. 17 will be negative and consequently  $\langle \Delta R_{ij}^2 \rangle$  will be large, leading to a small value of  $\gamma_{ij}$ . For correlated motions, the dot product will be positive, and  $\langle \Delta R_{ij}^2 \rangle$  may be small, leading possibly to a large value of the spring constant.

With the expectation of decreasing the scatter in the calculated shaded points in Fig. 4, another run was made where the starting  $\gamma_{ij}$ -values were not equated to  $-1$  at the outset but their values were assigned according to  $1/2n_{ij}$ . At the end of the iterations, the values of  $\gamma_{ij}$ -values satisfied the 45° relation of Fig. 3. This shows that 1), the computational scheme is robust; and 2), there is an underlying effect that consistently leads to a unique set of  $\gamma_{ij}$ -values.

## Effects of binding on fluctuations

A unique set of spring constants for a protein that gives a precise description of fluctuations may suitably be used for the investigation of binding effects on them. Binding of a ligand on a single residue, say  $i^{\text{th}}$ , has the effect of increasing the number of neighbors in the domain of fluctuation of the residue. This changes the number  $n_{ij}$ , thereby affecting the fluctuations of the  $j^{\text{th}}$  residue, and the effect propagates throughout the protein. For some residues, this effect may propagate further into the protein and for others it may die out fast. To describe and study these effects, a detailed and a precise Hamiltonian is needed and the present method of improved  $\Gamma$  matrix is suitable for this. Without an accurate description of fluctuations, changes caused in them by binding can only be studied qualitatively and only in the slow modes. In this section, we formulate the GNM with binding and apply it to the analysis of ligand binding on various residues of BPTI.

### Binding of a ligand on a single residue

We assume that a ligand binds on the  $i^{\text{th}}$  residue of a protein of  $n$  residues. The  $\Gamma$  matrix of the new system, i.e., the protein plus the ligand, will be

$$\Gamma = \begin{bmatrix} \left(1 - \sum_{k=3}^n \gamma_{1k}\right) & -1 & \gamma_{13} & \gamma_{14} & \cdot & \cdot & \cdot & \cdot & \gamma_{1n} & 0 \\ -1 & \left(2 - \sum_{k=4}^n \gamma_{1k}\right) & -1 & \gamma_{24} & \gamma_{25} & \cdot & \cdot & \cdot & \gamma_{2n} & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & -1 & \left(2 - \gamma_{i,n+1} - \sum_{k=1}^{i-2} \gamma_{ik} - \sum_{k=i+2}^n \gamma_{ik}\right) & -1 & \gamma_{i,i+2} & \cdot & \cdot & \gamma_{in} & \gamma_{i,n+1} \\ 0 & 0 & 0 & \gamma_{i,n+1} & 0 & \cdot & \cdot & \cdot & -\gamma_{i,n+1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (18)$$

The value of  $\gamma_{i,n+1}$  measures the strength of binding of the ligand to the residue. A value of  $-1$  makes it equivalent to a covalent bond. In the calculations below, we adopt this value.

In Fig. 5, we show the changes in the fluctuations of residue  $j$  when a ligand binds on ligand  $i$ . The solid and dark-shaded contour regions indicate a decrease of fluctuations of the corresponding residues  $j$ , and the open and light-shaded regions indicate an increase. For example, binding on residue 15 decreases the fluctuations of residue 37 that falls in the solid contour. The solid regions indicate a decrease of the indicated residues along the ordinate by 3–5% relative to the unbound state. The open regions indicate an increase of fluctuations by 1–2%. The contours indicate a strong symmetry with respect to exchange of axes. This shows

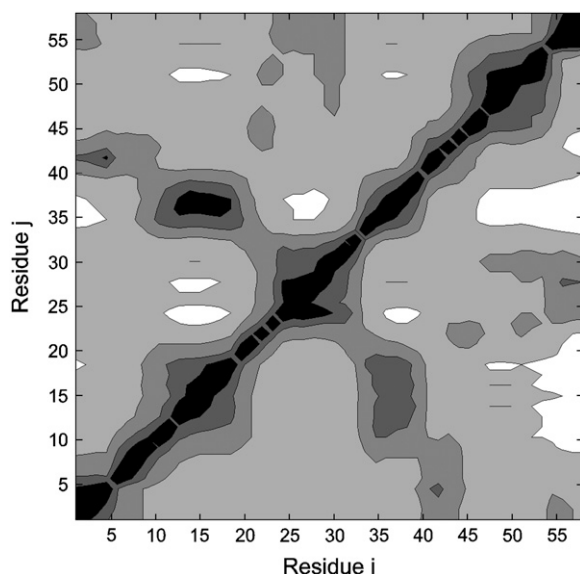


FIGURE 5 Contour map of perturbation of a residue  $j$  when binding takes place on a residue  $i$ .

that the effect on residue  $j$  when binding is on residue  $i$  is similar to the effect on residue  $i$  when binding is on residue  $j$ . Response of residues to perturbation has previously been formulated and analyzed for several proteins (14–16).

In Fig. 6, effects of binding on Lys<sup>15</sup> and Gly<sup>37</sup> are compared. The solid curve shows the percent change in the fluctuations of other residues when binding is on residue Lys<sup>15</sup>. The fluctuations of the residues 9–20 decrease upon this binding. Also, the fluctuations of Gly<sup>37</sup> are decreased significantly, as observed from the second minimum in the solid curve. Binding also increases the fluctuations of some residues, specifically, those of 21–33 and 41–58. It is to be noted that the decrease in fluctuations of Gly<sup>37</sup> is a direct consequence of the fact that Lys<sup>15</sup> and Gly<sup>37</sup> are close spatial neighbors. The thin line shows the changes taking place when a ligand binds on Gly<sup>37</sup>. Thus the effects induced by binding on Gly<sup>37</sup> are similar to those of binding on Lys<sup>15</sup>, indicating the approximate reciprocity of binding to Lys<sup>15</sup> and Gly<sup>37</sup>.

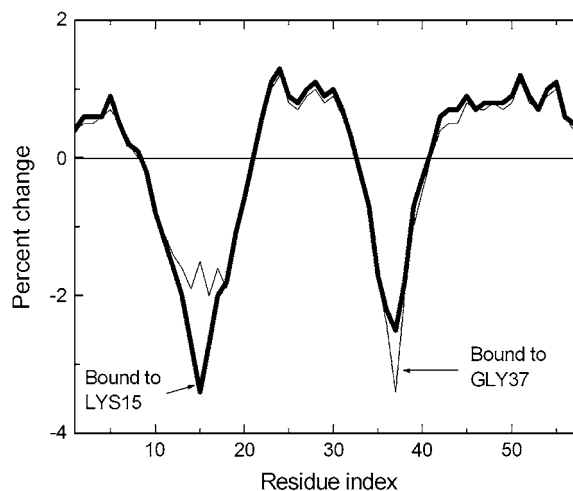
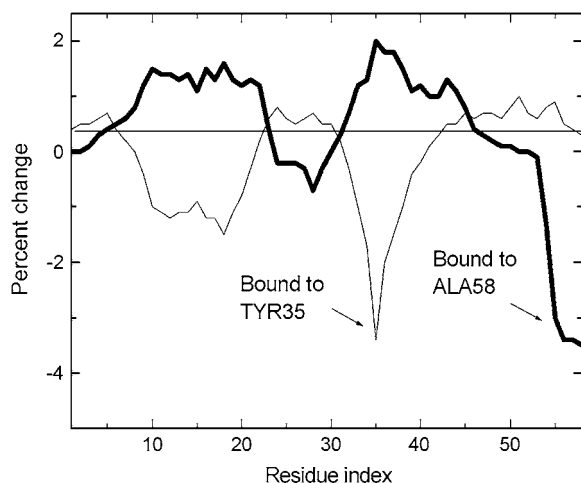
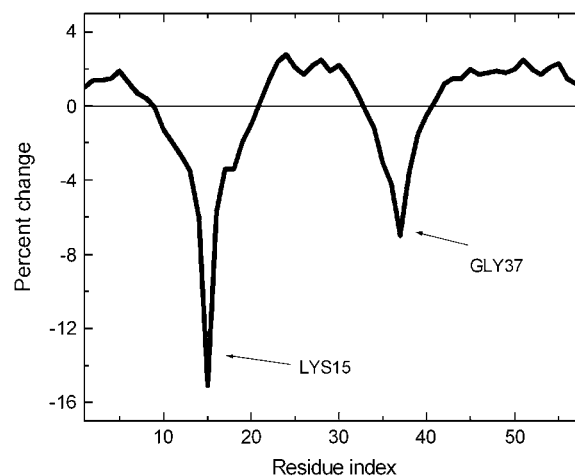


FIGURE 6 Binding on Lys<sup>15</sup> and on Gly<sup>37</sup>, separately.

FIGURE 7 Comparison of the effects of binding on Tyr<sup>35</sup> and Ala<sup>58</sup>.

When the two residues are not neighbors in space, binding on one effects the fluctuations of the other, but the reciprocity stated above does not necessarily hold. As an example, in Fig. 7, effects of binding on residues Tyr<sup>35</sup> and Ala<sup>58</sup> are shown. Binding on Tyr<sup>35</sup> and Ala<sup>58</sup> is indicated by the thin and thick lines, respectively. Binding on Ala<sup>58</sup> induces an

FIGURE 8 Binding on Lys<sup>15</sup> and Gly<sup>37</sup>.

$$[\sigma]_{ij} = \begin{cases} \gamma_{ij} & \text{if } j^{\text{th}} \text{ point of the ligand is bound to } i^{\text{th}} \text{ residue} \\ 0 & \text{for } j = 1, \dots, m \text{ if no binding to } i^{\text{th}} \text{ residue} \end{cases} \quad (20)$$

The matrix  $[\gamma\sigma]_{m,m}$  has the form

$$[\gamma\sigma]_{m,m} = \begin{bmatrix} 1 - \gamma_{i1} & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 - \gamma_{j2} & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 - \gamma_{k3} & -1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & -1 & 2 - \gamma_{p,m-2} & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 - \gamma_{q,m-1} & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 - \gamma_m \end{bmatrix}, \quad (21)$$

increase in the fluctuations of Tyr<sup>35</sup>, but binding on Tyr<sup>35</sup> does not have any effect on the fluctuations of Ala<sup>58</sup>. Furthermore, binding on Ala<sup>58</sup> induces an increase in the fluctuations of residues 8–21, whereas binding on Tyr<sup>35</sup> induces a decrease of fluctuations for these residues.

#### Binding to multiple sites on the protein

The  $\Gamma$  matrix defined by Eq. 16 may be extended to the case of multiple binding to different residues of a protein. For a protein of  $n$ -residues and a ligand that binds to  $m$ -sites, Eq. 16 may be written in block form as

$$\Gamma = \begin{bmatrix} [\gamma]_{n,n} & [\sigma]_{n,m} \\ [\sigma]_{m,n} & [\gamma\sigma]_{m,m} \end{bmatrix}. \quad (19)$$

Here,  $[\sigma]_{n,m}$  is the matrix that has  $n$ -rows and  $m$ -columns defined as

where  $m$ -bindings have taken place on residues  $i, j, k, \dots, p, q, r$ . The ligand is assumed to constitute a linear chain of  $m$ -binding sites, and the linear connectivity is acknowledged by  $-1$  values along the first off-diagonal terms of  $[\sigma\gamma]_{m,m}$ .

In Fig. 8, the effects of simultaneous binding on Lys<sup>15</sup> and Gly<sup>37</sup> are shown. Compared to Fig. 6, binding simultaneously on both of these residues causes a fourfold-larger decrease than if binding took place on Lys<sup>15</sup> only, or Gly<sup>37</sup> only.

The effects of binding two independent ligands to two residues are significantly different than when the two ligands are connected to form a single molecule. In Fig. 9, effects of simultaneous and independent binding on Gly<sup>28</sup> and Ala<sup>58</sup> are compared. The two residues are 7.1 Å apart, and hence not within the cutoff distance of 7.0 Å.

The thin line represents the effects when the spring constant between the two ligands are taken as 0 that corresponds to independent binding, i.e., two independent ligands binding

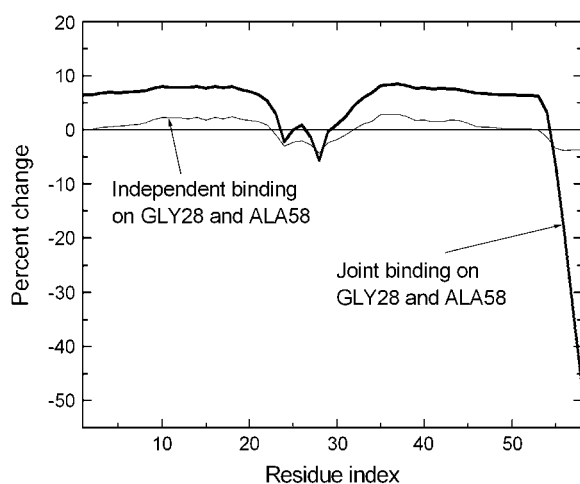


FIGURE 9 Effects of simultaneous binding on Gly<sup>28</sup> and Ala<sup>58</sup>.

on these two sites. The heavy solid curve is obtained when this spring constant between the two ligands is equated to  $-1$ . This makes the ligand behave as a single entity of two binding sites. The fluctuations of Gly<sup>28</sup> are not affected much upon this modification of the ligand. However, the fluctuations of Ala<sup>58</sup> are significantly reduced, and the fluctuations of the rest of the protein between residues 1–23 and 29–52 increase significantly. In Fig. 10, effects of binding at five points on the helix Ser<sup>47</sup>–Gly<sup>56</sup> are shown by the solid curve. The light-shaded curve indicates effects of binding to only one residue, Glu<sup>49</sup>, on the helix. Comparison of the two curves shows the magnification of the effect of simultaneous binding on several successive residues. The figure also shows the allosteric effects of binding, according to which binding on one part induces strong changes on another part of the protein that is far from the binding site.

## DISCUSSION AND CONCLUSION

This article consists of two parts. In the first part, the originally proposed GNM is modified to obtain an exact

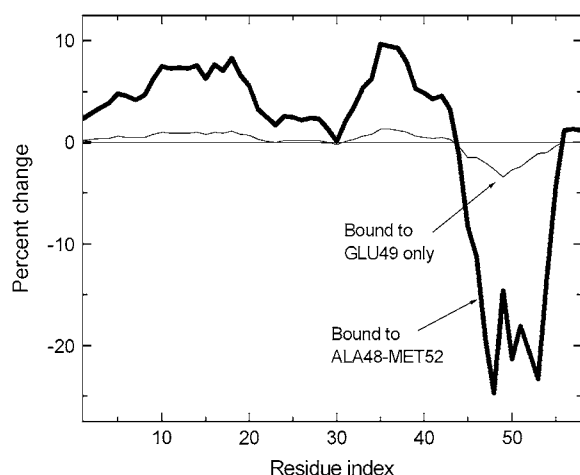


FIGURE 10 Binding on Ala<sup>48</sup>–Met<sup>52</sup> of the helix.

match between experimental and predicted values of residue fluctuations. This improvement in the model is important because it provides an exact description of fluctuations in a consistent way, with the aid of which residue-specific events relating to fluctuations can be analyzed in greater detail and accuracy. The second part of the article involves the application of the model to a specific protein. In this part, we showed that an accurate description of fluctuations is indeed useful in understanding the detailed behavior of the protein. A wide range of properties of proteins relating to fluctuations have been addressed successfully with the original version of the GNM, which showed remarkable agreement with experiment at the coarse-grained level. The improvement introduced here allows for the analysis of specific details very accurately at the residue level. Corrections to an already successful model have to be justified carefully. First, it should be robust, which in turn requires the model to yield the same results irrespective of the initial distribution of  $\gamma_{ij}$ . This has been shown to be the case for BPTI. Calculations carried out on several other proteins but not reported here also show that the  $\gamma_{ij}$ -values converge to a fixed distribution, irrespective of the starting distribution of  $\gamma_{ij}$ -values. A few patterns on the magnitudes of  $\gamma_{ij}$ -values may be extracted from the results on BPTI. Firstly, the distribution of  $\gamma_{ij}$ -values obey a Lorentzian distribution, which has a pronounced peak at the small value of  $-0.088$  (compared to  $-1$  of GNM), and there are a few positive  $\gamma_{ij}$ -values. The repulsive springs are required in the cases where a cluster of neighboring attractive springs tend to bring certain pairs of residues close to each other, and the repulsive springs are needed to prevent the collapse of these residues onto each other. Examples of this are given below. It is to be noted that the number and strength of the repulsive springs are much smaller than those that would possibly lead to the instability of the protein. Calculations were carried out by allowing only attractive springs and equating the spring constant to zero when the simulation indicated a repulsive spring. However, convergence of the calculated fluctuations to the experimental ones was not possible in that case, and the model was not accurate. Residue pairs that have fewer neighbors are located at the surface. These are the important residue pairs in the sense that the absolute values of their  $\gamma_{ij}$ -values are large. These pairs are either strongly attracted together (pairs connected with a stiff attractive spring, or high negative value of  $\gamma_{ij}$ ) or repel each other strongly (pairs connected with a stiff repulsive spring, or high positive value of  $\gamma_{ij}$ ). An analysis of the spring constants given in Table 1 shows that the locations of these residues are important for the stability and/or function of the protein. For example, the largest value  $\gamma_{ij} = -1.09$  is for the pair Asn<sup>24</sup>–Ala<sup>27</sup>, both of which are located on a tight turn at the surface. The next important pair is also on the same tight turn, Asn<sup>24</sup>–Gly<sup>28</sup> with  $\gamma_{ij} = -0.68$ . The pair Ala<sup>16</sup>–Gly<sup>37</sup> with  $\gamma_{ij} = -0.48$  joins the turns of two major loops at the surface of the protein. Phe<sup>4</sup>–Arg<sup>42</sup> pair with the next highest  $\gamma_{ij} = -0.389$



is also located at the center, joining the tail of the chain to a point on the body. The locations of these important interactions are presented in Fig. 11 *a*. The pair Arg<sup>53</sup>-Gly<sup>56</sup> with the highest positive  $\gamma_{ij} = 0.193$  is located at the end of the chain, where Gly<sup>56</sup> is on the free unstructured tail of the helix. Arg<sup>53</sup> is situated on the helix. The repulsive spring prevents Gly<sup>56</sup> from collapsing on the helix and keeps it protruding out from the surface. Similarly, each residue of the pair Gly<sup>12</sup>-Arg<sup>39</sup> with  $\gamma_{ij} = 0.113$  is on the surface and located at the midpoints of the two neighboring long coils of the protein. The repulsive spring between them keeps the two coils from collapsing onto each other.

The locations of these pairs on the surface of the protein are shown in Fig. 11 *b*. These examples indicate that the factors affecting the magnitudes of the spring constants depend on diverse structural features of the molecule probably relating to stability and simultaneously to function. However, inspection of Fig. 4 shows that the majority of  $\gamma_{ij}$ -values lie in a narrow region  $\sim \gamma_{ij} = -0.088$ . The outliers are those pairs at the surface of the protein.

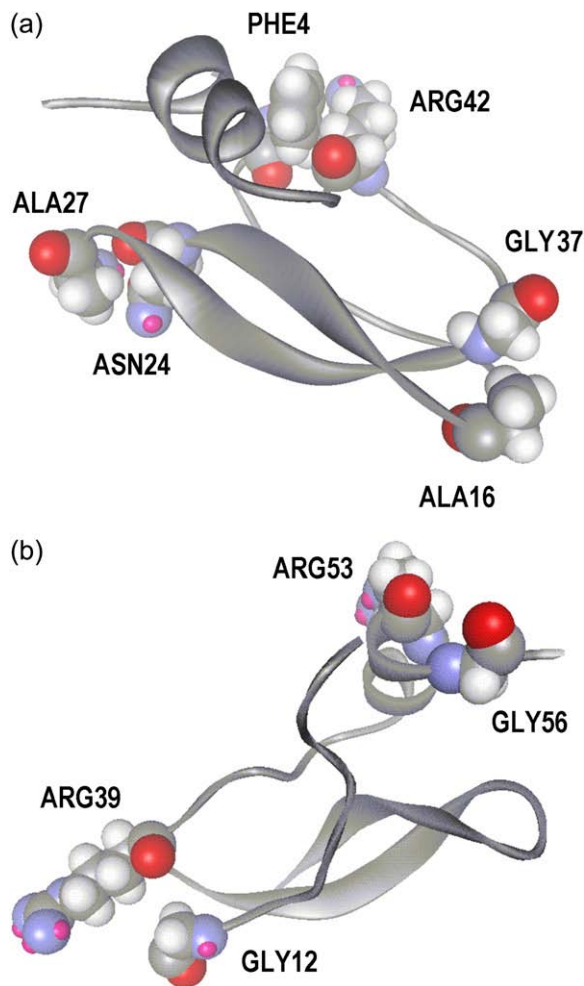


FIGURE 11 (a) Pairs with large attractive spring constants. (b) Pairs with large repulsive spring constants.

The present model describes the fluctuations of residues in more detail than the standard GNM. To clarify the predictions of the present model relative to those of the standard GNM, we conducted a detailed modal decomposition of the present model according to the expression (17)

$$\langle \Delta R_i^2 \rangle_k = \left( \frac{3kT}{\gamma} \right) \lambda_k^{-1} [u_k]_i^2. \quad (22)$$

Here,  $\langle \Delta R_i^2 \rangle_k$  denotes the  $k^{\text{th}}$  component of the fluctuation of the  $i^{\text{th}}$  residue,  $\lambda_k$  is the  $k^{\text{th}}$  eigenvalue, and  $[u_k]_i$  is the  $i^{\text{th}}$  component of the  $k^{\text{th}}$  eigenvector. In Fig. 12 *a*, the collective contribution of the lowest five modes is shown. The dotted line is for the standard GNM, and the solid curve is for the present model, and the two agree more or less perfectly in the lowest five cumulative modes. Fig. 12 *b* compares the highest five modes of the two models. Here, the thick solid line refers to the present model and the thin line to the standard GNM, and significant detail is observed in the present model while it is not present in the standard GNM.

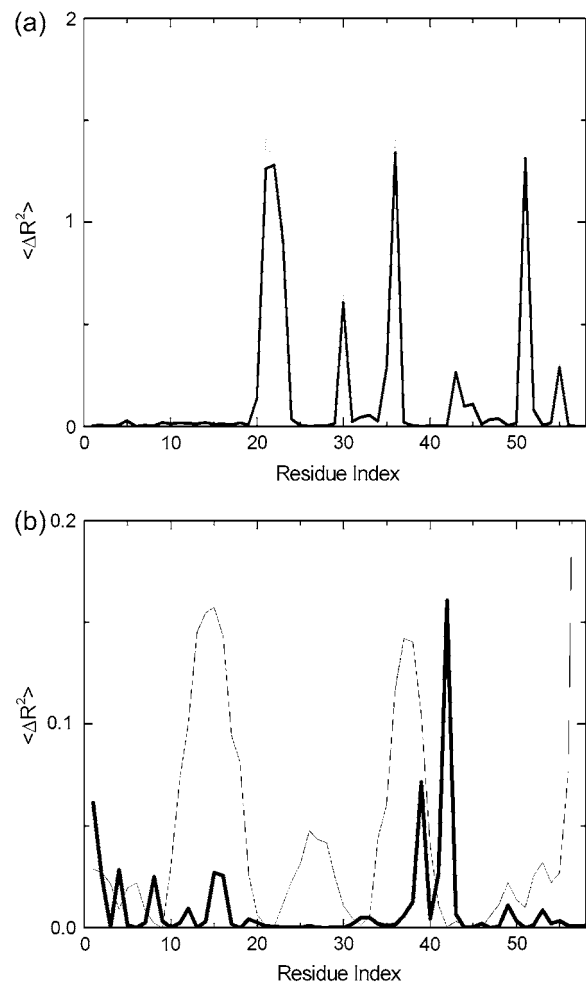


FIGURE 12 (a) Contribution of the lowest five modes to the fluctuations of residues. Solid line shows results of the present model; dotted line that of the standard GNM. (b) Contribution of the highest five modes. Thick solid line, present model; thin solid line, standard GNM.

TABLE 1 Calculated values of the spring constants

Residues	Value	Residues	Value	Residues	Value
Asn <sup>24</sup> -Ala <sup>27</sup>	-1.090	Cys <sup>5</sup> -Tyr <sup>23</sup>	-0.122	Tyr <sup>23</sup> -Gln <sup>31</sup>	-0.047
Asn <sup>24</sup> -Gly <sup>28</sup>	-0.683	Ser <sup>47</sup> -Asp <sup>50</sup>	-0.119	Asp <sup>50</sup> -Arg <sup>53</sup>	-0.046
Ala <sup>16</sup> -Gly <sup>37</sup>	-0.482	Tyr <sup>21</sup> -Phe <sup>45</sup>	-0.119	Cys <sup>30</sup> -Ala <sup>48</sup>	-0.044
Ala <sup>25</sup> -Gly <sup>28</sup>	-0.453	Arg <sup>17</sup> -Tyr <sup>35</sup>	-0.118	Phe <sup>22</sup> -Thr <sup>32</sup>	-0.044
Phe <sup>4</sup> -Arg <sup>42</sup>	-0.389	Tyr <sup>10</sup> -Ala <sup>40</sup>	-0.118	Arg <sup>20</sup> -Phe <sup>45</sup>	-0.042
Ala <sup>16</sup> -Gly <sup>36</sup>	-0.338	Phe <sup>22</sup> -Asn <sup>44</sup>	-0.117	Tyr <sup>21</sup> -Cys <sup>51</sup>	-0.039
Ile <sup>18</sup> -Gly <sup>37</sup>	-0.325	Pro <sup>9</sup> -Phe <sup>22</sup>	-0.113	Tyr <sup>23</sup> -Cys <sup>55</sup>	-0.032
Asn <sup>24</sup> -Cys <sup>30</sup>	-0.315	Ile <sup>19</sup> -Tyr <sup>35</sup>	-0.111	Tyr <sup>21</sup> -Ser <sup>47</sup>	-0.028
Cys <sup>14</sup> -Gly <sup>37</sup>	-0.304	Arg <sup>20</sup> -Tyr <sup>35</sup>	-0.110	Tyr <sup>10</sup> -Lys <sup>41</sup>	-0.019
Asn <sup>24</sup> -Gln <sup>31</sup>	-0.296	Asp <sup>5</sup> -Leu <sup>6</sup>	-0.110	Phe <sup>22</sup> -Cys <sup>30</sup>	-0.016
Cys <sup>14</sup> -Cys <sup>38</sup>	-0.279	Ile <sup>18</sup> -Val <sup>34</sup>	-0.107	Lys <sup>15</sup> -Gly <sup>37</sup>	-0.016
Ile <sup>18</sup> -Tyr <sup>35</sup>	-0.276	Arg <sup>20</sup> -Phe <sup>33</sup>	-0.105	Arg <sup>20</sup> -Asn <sup>44</sup>	-0.015
Pro <sup>13</sup> -Gly <sup>36</sup>	-0.261	Ala <sup>48</sup> -Met <sup>52</sup>	-0.104	Arg <sup>1</sup> -Cys <sup>55</sup>	-0.013
Tyr <sup>10</sup> -Tyr <sup>35</sup>	-0.244	Phe <sup>22</sup> -Asn <sup>43</sup>	-0.099	Gly <sup>12</sup> -Ala <sup>40</sup>	-0.012
Gly <sup>12</sup> -Cys <sup>38</sup>	-0.226	Phe <sup>22</sup> -Phe <sup>45</sup>	-0.096	Phe <sup>22</sup> -Gln <sup>31</sup>	-0.011
Phe <sup>45</sup> -Cys <sup>51</sup>	-0.211	Tyr <sup>21</sup> -Thr <sup>32</sup>	-0.095	Tyr <sup>21</sup> -Ala <sup>48</sup>	-0.007
Pro <sup>9</sup> -Asn <sup>43</sup>	-0.210	Ile <sup>19</sup> -Phe <sup>33</sup>	-0.095	Tyr <sup>21</sup> -Gln <sup>31</sup>	-0.004
Thr <sup>11</sup> -Tyr <sup>35</sup>	-0.194	Glu <sup>49</sup> -Met <sup>52</sup>	-0.093	Arg <sup>20</sup> -Thr <sup>32</sup>	-0.003
Pro <sup>2</sup> -Cys <sup>5</sup>	-0.193	Pro <sup>2</sup> -Cys <sup>55</sup>	-0.090	Ile <sup>19</sup> -Val <sup>34</sup>	-0.002
Gly <sup>28</sup> -Gly <sup>56</sup>	-0.190	Cys <sup>30</sup> -Cys <sup>51</sup>	-0.089	Phe <sup>4</sup> -Asn <sup>43</sup>	0.011
Asn <sup>24</sup> -Leu <sup>29</sup>	-0.188	Cys <sup>51</sup> -Cys <sup>55</sup>	-0.089	Glu <sup>49</sup> -Arg <sup>53</sup>	0.014
Cys <sup>5</sup> -Asn <sup>43</sup>	-0.174	Gly <sup>12</sup> -Gly <sup>36</sup>	-0.086	Tyr <sup>21</sup> -Cys <sup>30</sup>	0.021
Ala <sup>48</sup> -Cys <sup>51</sup>	-0.164	Cys <sup>5</sup> -Cys <sup>55</sup>	-0.085	Tyr <sup>21</sup> -Lys <sup>46</sup>	0.021
Cys <sup>14</sup> -Gly <sup>36</sup>	-0.161	Cys <sup>5</sup> -Ala <sup>25</sup>	-0.084	Arg <sup>17</sup> -Val <sup>34</sup>	0.024
Ala <sup>40</sup> -Asn <sup>44</sup>	-0.154	Met <sup>52</sup> -Cys <sup>55</sup>	-0.083	Ile <sup>19</sup> -Thr <sup>32</sup>	0.028
Ile <sup>18</sup> -Gly <sup>36</sup>	-0.152	Tyr <sup>21</sup> -Asn <sup>44</sup>	-0.080	Thr <sup>11</sup> -Val <sup>34</sup>	0.028
Ser <sup>47</sup> -Cys <sup>51</sup>	-0.146	Thr <sup>11</sup> -Gly <sup>36</sup>	-0.072	Arg <sup>20</sup> -Val <sup>34</sup>	0.041
Tyr <sup>21</sup> -Phe <sup>33</sup>	-0.139	Leu <sup>6</sup> -Ala <sup>25</sup>	-0.066	Arg <sup>20</sup> -Lys <sup>46</sup>	0.050
Gly <sup>12</sup> -Tyr <sup>35</sup>	-0.138	Lys <sup>41</sup> -Asn <sup>44</sup>	-0.059	Met <sup>52</sup> -Gly <sup>56</sup>	0.060
Phe <sup>22</sup> -Phe <sup>33</sup>	-0.136	Cys <sup>30</sup> -Met <sup>52</sup>	-0.059	Tyr <sup>23</sup> -Leu <sup>29</sup>	0.072
Phe <sup>45</sup> -Asp <sup>50</sup>	-0.135	Phe <sup>22</sup> -Cys <sup>51</sup>	-0.058	Phe <sup>4</sup> -Glu <sup>7</sup>	0.103
Cys <sup>51</sup> -Thr <sup>54</sup>	-0.131	Asp <sup>50</sup> -Thr <sup>54</sup>	-0.056	Lys <sup>15</sup> -Gly <sup>36</sup>	0.107
Tyr <sup>23</sup> -Cys <sup>30</sup>	-0.124	Arg <sup>1</sup> -Cys <sup>5</sup>	-0.051	Ala <sup>25</sup> -Leu <sup>29</sup>	0.112
Tyr <sup>23</sup> -Cys <sup>51</sup>	-0.122	Arg <sup>17</sup> -Gly <sup>36</sup>	-0.049	Gly <sup>12</sup> -Arg <sup>39</sup>	0.113
Tyr <sup>23</sup> -Asn <sup>43</sup>	-0.122	Glu <sup>7</sup> -Asn <sup>43</sup>	-0.049	Arg <sup>53</sup> -Gly <sup>56</sup>	0.193

Specifically, the latter predicts a peak for the range of residues 33–42, as observed from Fig. 12 *b*, but the present model resolves this peak to a peak at Arg<sup>39</sup> and Arg<sup>42</sup>, the two important residues of the binding site. Similarly, the standard GNM gives a diffused peak in the range of residues 9–20, whereas the present model points to the significance of residues 12, 15, 16, and 19 in this range. We therefore conclude that the present model is more detailed and more specific in the higher modes. The relevance of this detail to known experimental data for different systems is the subject of future work.

Finally, it is worthwhile to add several recently published articles relating to research reported in this article; for example, Ming and Wall (18), who improved the model by strengthening backbone interactions; Tobi and Bahar (19), who found correlations between intrinsic motions of unbound proteins and structural changes upon binding; and Sen et al. (20), who systematically compared Gaussian Network Models with varying scales of coarse-graining.

It is a great pleasure and an overdue duty to acknowledge the contributions of Dr. Andrzej Kloczkowski to our understanding of the Gaussian Network

Model. His critical appreciation of the work of Flory and especially of Pearson and his clear reformulation of the theory have been crucial in the development of the Gaussian Network Model for proteins.

## REFERENCES

1. Bahar, I., A. R. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* 2:173–181.
2. Kloczkowski, A., J. E. Mark, and B. Erman. 1989. Chain dimensions and fluctuations in random elastomeric networks. I. Phantom Gaussian networks in the undeformed state. *Macromolecules.* 22:1423–1432.
3. Erman, B., and P. J. Flory. 1982. Relationship between stress, strain, and molecular constitution of polymer networks. Comparison of theory with experiments. *Macromolecules.* 15:806–812.
4. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.
5. Bahar, I., A. R. Atilgan, M. C. Demirel, and B. Erman. 1998. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* 80:2733–2736.
6. Bahar, I., B. Erman, R. L. Jernigan, A. R. Atilgan, and D. G. Covell. 1999. Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function. *J. Mol. Biol.* 285:1023–1037.

7. Halle, B. 2002. Flexibility and packing in proteins. *Proc. Natl. Acad. Sci. USA*. 99:1274–1279.
8. Micheletti, C., J. R. Banavar, and A. Maritan. 2001. Conformations of proteins in equilibrium. *Phys. Rev. Lett.* 87:8102–8105.
9. Kundu, S., J. S. Melton, D. C. Sorensen, G. N. Phillips. 2002. Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophys. J.* 83:723–732.
10. Ming, D., Y. Kong, M. A. Lambert, Z. Huang, and J. Ma. 2002. How to describe protein motion without amino acid sequence and atomic coordinates. *Proc. Natl. Acad. Sci. USA*. 99:8620–8625.
11. Moritsugu, K., O. Miyashita, and A. Kidera. 2000. Vibrational energy transfer in a protein molecule. *Phys. Rev. Lett.* 85:3970–3973.
12. Moritsugu, K., O. Miyashita, and A. Kidera. 2003. Vibrational energy transfer in a protein molecule. *J. Phys. Chem. B*. 107:3309–3317.
13. Reference deleted in proof.
14. Yilmaz, L. S., and A. R. Atilgan. 2000. Identifying the adaptive mechanism in globular proteins: fluctuations in densely packed regions manipulate flexible parts. *J. Chem. Phys.* 113:4454–4464.
15. Baysal, C., and A. R. Atilgan. 2001. Elucidating the structural mechanisms for biological activity of the chemokine family. *Proteins*. 43:150–160.
16. Baysal, C., and A. R. Atilgan. 2001. Coordination topology and stability for the native and binding conformers of chymotrypsin inhibitor 2. *Proteins Struct. Funct. Gen.* 45:62–70.
17. Demirel, M. C., A. R. Atilgan, R. L. Jernigan, B. Erman, and I. Bahar. 1998. Identification of kinetically hot residues in proteins. *Protein Sci.* 7:2522–2532.
18. Ming, D., and M. E. Wall. 2005. Allostery in a coarse-grained model of protein dynamics. *Phys. Rev. Lett.* 95:198103–198106.
19. Tobi, D., and I. Bahar. 2005. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA*. 102:18908–18913.
20. Sen, T. Z., Y. Feng, J. V. Garcia, A. Kloczkowski, and R. L. Jernigan. 2006. The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. *J. Chem. Theory Comput.* 2:696–704.