Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential

Ivet Bahar, Ali Rana Atilgan and Burak Erman

Background: An elastic network model is proposed for the interactions between closely (≤ 7.0 Å) located α-carbon pairs in folded proteins. A single-parameter harmonic potential is adopted for the fluctuations of residues about their mean positions in the crystal structure. The model is based on writing the Kirchhoff adjacency matrix for a protein defining the proximity of residues in space. The elements of the inverse of the Kirchhoff matrix give directly the auto-correlations or cross-correlations of atomic fluctuations.

Results: The temperature factors of the C^{α} atoms of 12 X-ray structures, ranging from a 41 residue subunit to a 633 residue dimer, are accurately predicted. Cross-correlations are also efficiently characterized, in close agreement with results obtained with a normal mode analysis coupled with energy minimization.

Conclusions: The simple model and method proposed here provide a satisfactory description of the correlations between atomic fluctuations. Furthermore, this is achieved within computation times at least one order of magnitude shorter than commonly used molecular approaches.

Address: Polymer Research Center and School of Engineering, Bogazici University, and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey.

Correspondence: Ivet Bahar
E-mail: bahar@prc.bme.boun.edu.tr

Key words: cross-correlations, Kirchhoff adjacency matrix, nonbonded interactions, temperature factors, X-ray structures

Received: 02 Jan 1997
Revisions requested: 06 Feb 1997
Revisions received: 21 Feb 1997
Accepted: 17 Mar 1997

Published: 07 May 1997
Electronic identifier: 1359-0278-002-00173

Folding & Design 07 May 1997, 2:173–181

Introduction

With increasing numbers of X-ray or NMR elucidated protein structures, efforts have been directed to extracting residue-specific potentials stabilizing native structures as an important step towards solving the protein folding problem (for review, see e.g. [1,2]). On the other hand, some recent studies have suggested that simple energy functions that incorporate the two most important characteristics of amino acids, mainly hydrophobicity and hydrogen bond formation capacity, may adequately account for the selection of some supersecondary structures or tertiary folds in small proteins [3,4]. So, the question "how much detail in interresidue potentials is needed for describing the behavior of proteins at a given level of approximation?" has not yet found a unanimous answer.

Recently, a single-parameter Hookean potential was adopted for the pairwise interaction of all atoms in X-ray crystallographic structures in a normal mode analysis (NMA) of large-amplitude elastic motions [5]. This approximation is based on a Gaussian distribution of interatomic distances about their equilibrium values. The crystallographic temperature factors obtained by this method for G-actin bound with ADP and Ca^{2+} showed full agreement with those obtained with a detailed potential introduced by Levitt [6]. The simplicity of the postulated single-parameter Gaussian model and its success in predicting results for a complex system may have far-reaching consequences in understanding protein structures.

In the present report we show, by analyzing the Kirchhoff adjacency matrix of nonbonded interactions, that a single-parameter harmonic potential does successfully describe the temperature factors and cross-correlations in folded proteins. With the knowledge of the crystal structures only, and adopting a single parameter per protein, we have satisfactorily reproduced the temperature factors of the C^{α} atoms of 12 proteins from the PDB. The method is strikingly simple and efficient: no knowledge of detailed residue-specific potentials is required, and the time-consuming energy minimization algorithms and molecular dynamics (MD) simulations with complex potentials as in NMA of X-ray structures [7–14] are avoided. Instead, a highly efficient analytical method is used, which finds its roots in the elasticity theory of random polymer networks [15–17].

The basic postulate adopted in the present study is that the protein in the folded state is equivalent to a three-dimensional elastic network. In the classic theories [15], the junctions of the network undergo Gaussian-distributed fluctuations under the potential of the pendant chains. Here, the C^{α} atoms are identified with the junctions of the network, and they fluctuate under the potentials of their near neighbors. Thus, the interaction potential between closely located pairs of C^{α} atoms substitutes for the harmonic potential constraining the end-to-end separation of the classic network chains. We note that in addition to nonbonded interactions, the effect of chain connectivity is also taken into account in the present theory, as the model

automatically includes the constraints imposed by the first neighboring Cα atoms along the backbone.

Theory

Following the model network of randomly fluctuating junctions postulated above, the fluctuations ΔR_{ij} in the separation $R_{ij} = |R_j - R_i|$ between the ith and jth Cα atoms in the folded protein are assumed to obey a Gaussian distribution:

$$W(\Delta R_{ij}) = (\gamma^*/\pi)^{3/2} \exp(-\gamma^* \Delta R_{ij}^2) \quad (1)$$

Here, the normalization constant γ^* is the counterpart of the single parameter in the Hookean potential adopted by Tirion [5]. In the statistical theory of polymer networks, the distribution function $W(\Delta R_{ij})$ is substituted into the expression $\Delta A = -k_B T \ln W(\Delta R_{ij})$ for the elastic free energy change associated with the fluctuation ΔR_{ij} in the end-to-end vector connecting junctions i and j. This yields the harmonic potential $k_B T \gamma^* \Delta R_{ij}^2$ and a Hookean force constant equal to $2k_B T\gamma^*$. The latter replaces the single parameter C used by Tirion. The configurational factor for a protein of N residues may then be expressed, with analogy to the theory of random Gaussian networks [15–17], by equation 2:

$$Z_N = K \exp(-\{\Delta R^T\}\, \Gamma\, \{\Delta R\}) \quad (2)$$

Here, $\{\Delta R\}$ is the N dimensional column vector formed by the fluctuations $\{\Delta R_1, \Delta R_2, \ldots, \Delta R_N\}$ of the Cα atoms, the superscript T denotes the transpose, K is a constant, and Γ is a symmetric matrix known as the Kirchhoff or valency-adjacency matrix [18] in graph theory. The elements of Γ are given by equation 3:

$$\Gamma_{ij} = \begin{cases} -\gamma^* \text{ if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 \text{ if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{i \neq j} \Gamma_{ij} \text{ if } i = j \end{cases} \quad (3)$$

The summation for evaluating Γ_{ii} is performed over all off-diagonal elements on the ith column (or row). r_c is the cutoff separation defining the range of nonbonded contacts. The equilibrium correlation between the fluctuations of two sites k and l is obtained from [16,17]:

$$\langle \Delta R_k \cdot \Delta R_l \rangle = (1/Z) \int \Delta R_k \cdot \Delta R_l e^{-V/kT} \mathrm{d}\{\Delta R\}$$

$$= \int \Delta R_k \cdot \Delta R_l e^{-V/kT} \mathrm{d}\{\Delta R\} / \int e^{-V/kT} \mathrm{d}\{\Delta R\} = \left[\Gamma^{-1}\right]_{kl} \quad (4)$$

where $V = \{\Delta R^T\}\, \Gamma\, \{\Delta R\}$ is the potential associated with the vibrations of the Cα atoms, Z is the partition function, the angular brackets designate the ensemble average over all fluctuations, $\mathrm{d}\{\Delta R\} \equiv \mathrm{d}\Delta R_1\, \mathrm{d}\Delta R_2 \ldots \mathrm{d}\Delta R_N$, and $[\Gamma^{-1}]_{kl}$ is the klth element of the inverse of Γ. Γ may be regarded as the atomistic counterpart of the stiffness matrix in the analysis of elastic bodies. The mean-square fluctuations of the Cα atoms are readily evaluated from the diagonal elements of Γ^{-1} using $\langle \Delta R_k^2 \rangle = [\Gamma^{-1}]_{kk}$ and the cross-correlations between the fluctuations of the Cα atoms are found from the off-diagonal entries of Γ^{-1}.

Results and discussion

Comparison of predicted mean-square fluctuations with crystallographic temperature factors

Results from calculations for 12 proteins of different sizes and structural classes (Table 1) taken from the PDB [19,20] are displayed in Figures 1–3. Here, the mean-square fluctuations of Cα atoms calculated from equation 4 are compared with the Debye–Waller or temperature factors measured for the same atoms by X-ray crystallography:

$$B_k = 8\pi^2 \langle \Delta R_k \cdot \Delta R_k \rangle / 3 \quad (5)$$

Results are plotted as a function of residue index k, $1 \leq k \leq N$. The only adjustable parameter in our theory is γ^*, which is determined by normalizing the theoretical distribution with respect to the experimental, so as to match the areas enclosed by the two curves in each figure. The γ^* value used for each protein is listed in Table 1.

In Figure 1, relatively small size proteins ($40 < N < 90$) are considered (PDB codes are indicated in the figures). Details of these structures are given in Table 1. In Figure 2, intermediate size proteins ($90 < N < 130$) are presented, and larger size proteins ($N > 160$) are displayed in Figure 3.

Despite the simplicity of the model and method, theoretical and experimental curves are found to be strikingly similar. We note that B values include not only temperature-dependent vibrations of the atoms (as represented by eq. 5), but also other effects such as static disorders [21]. Yet, most of the peaks and minima of the experimental curves are accurately reproduced by the theory.

A few general features follow from the examination of the curves in Figures 1–3. Firstly, the more localized atoms, i.e. those exhibiting relatively low B values, are generally those participating in secondary structures, neighboring disulfide bridges, etc. Thus, several portions of the distributions resembling parabolic forms may be identified with α-helices or β-strands. Substrate binding regions are also accurately predicted, in general. Secondly, chain termini exhibit fluctuation peaks in several proteins due to their inherent enhanced flexibility. Also, surface atoms generally exhibit larger fluctuations, compared to core atoms. Thirdly, if the protein is a multimer, theoretical curves fall above the experimental curves at the regions of subunit–subunit interfaces, due to the absence of interactions between neighboring subunits. For example, residues 87–90, 134–151, 159–163 and 196–212 form the contacting region of the subunits of the trimer purine nucleoside phosphorylase (1ula) [22], which explains the discrepancy between theory and experiments at these specific regions. On the other hand, the theoretical curve

Table 1

Description of proteins examined in the present study.

PDB code*	Size†	Name	Resolution (Å)	Structural class	γ*	References
1ltsc	41	Heat-labile enterotoxin	1.95	α	2.55	[35]
1crn	46	Crambin	1.5	–	0.99	[36,37]
5pti	58	Trypsin inhibitor	1.0	–	1.94	[38]
1aba	87	Glutaredoxin	1.45	α/β	1.71	[39,40]
1vaab	99	MHC‡	2.3	β	3.00	[41]
1tho	108	*E. coli* thioredoxin	1.68	α+β	2.66	[42]
5rsa	124	Ribonuclease A	2.0	α+β	1.41	[43]
2ccya	127	Cytochrome *c′*	1.67	α	3.95	[44]
3lzm	164	T4 lysozyme	1.7	–	2.05	[45,46]
1ula	289	Purine nucleoside phosphorylase	2.75	α/β	3.00	[22]
1omf	340	Matrix porin	2.4	β	3.91	[24]
1atna	373	DNase I–actin§	2.8	–	2.70	[23]

*If there is more than one subunit, an extra letter designates the subunit. †Number of residues. ‡MHC class I H-2K^b complexed with a peptide antigen from vesicular stomatitis virus nucleocapsid protein. §Calculations for the subunit A of endodeoxyribonuclease I complexed with actin were performed by considering subunits A and D (a total of 633 residues) reported in the PDB.

in Figure 3d was obtained by taking into consideration the two subunits (a total of 633 residues) of DNase I–actin [23], which shows good agreement with experimental temperature factors. Fourthly, in the present model, only α-carbons are taken into consideration in evaluating root-mean-square fluctuations. Constraints imposed by highly specific sidechain–sidechain interactions are ignored. Consequently, the fluctuations of the corresponding α-carbons are overestimated. An example is the membrane protein 1omf. The channel-forming β-barrel in each subunit of this porin is pointed out to have a constriction region that is delimited by a row of three arginines, Arg42, Arg82 and Arg132, and Lys16 [24]. The guanidinium groups of the arginines are stacked and presumably have favorable π–π interactions. This unusual cluster is further stabilized by the sidechains of Asn64 and Glu62. These sidechains determine the size limitation and ion selectivity of the pore [24]. So, in this region, stability is imparted by sidechains, and examination of the C$^\alpha$ atoms leads to an underestimation of the constraints imposed by the specific segregation of the pendant charged sidechains in the constriction zone, as may be verified from Figure 3c.

In the interest of clarifying the differences in γ* values of different proteins, we did a systematic analysis of the results obtained for the proteins listed in Table 1. This indicated that γ* increases with increasing secondary structure content of the examined protein, and with the size of the protein. These two factors have the common effect of enhancing the fraction of residues with a high packing density. Thus, the differences in the γ* values of different proteins originate from the differences in their local packing density distributions. Those proteins offering a hydrophobic core with an efficient packing of residues, or possessing a great deal of secondary structural elements which constrain atomic fluctuations, generally exhibit larger γ* values. This is also consistent with the fact that the root-mean-square fluctuations scale with γ^{*-1}.

A further examination showed that the consideration of larger ranges of interactions (i.e. taking $r_c > 7.0$ Å) also leads to an increase in γ*. This may be attributed to the fact that an increase in the interaction ranges automatically implies an increase in the strength of constraints controlling atomic fluctuations. And this leads to a decrease in the overall amplitudes of motions reflected by higher γ* values.

Cross-correlations between atomic fluctuations

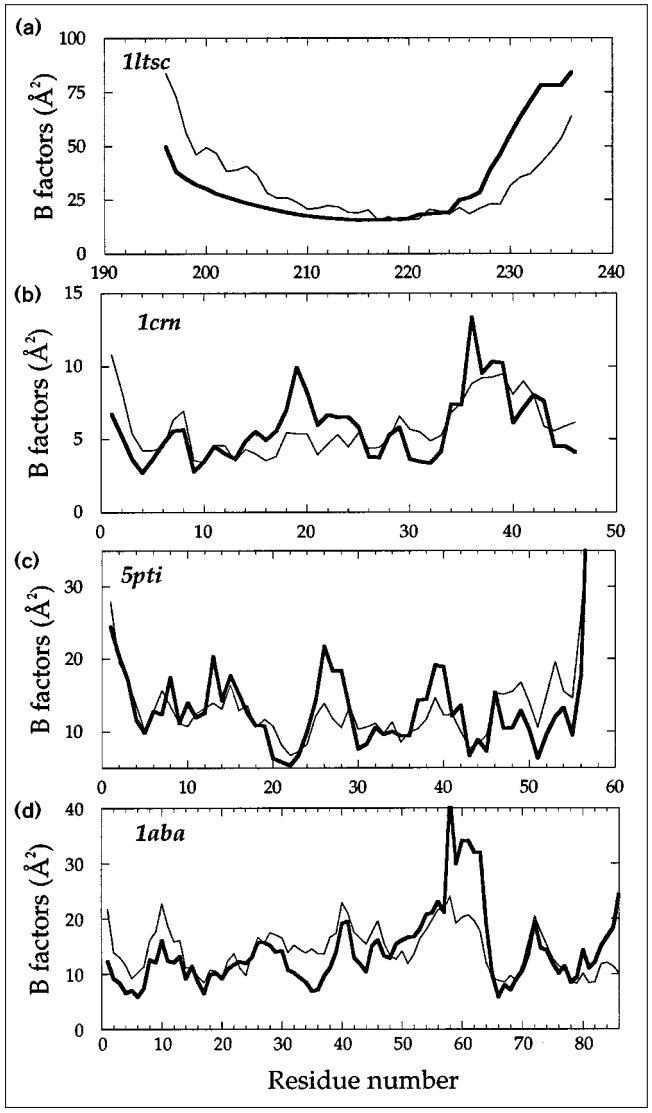
Figure 4 shows cross-correlations between the fluctuations of C$^\alpha$ atoms obtained for trypsin inhibitor (5pti). Cross-correlations are evaluated from the off-diagonal elements of Γ^{-1}, as outlined above. These are normalized with respect to the auto-correlations as:

$$C(k,l) = \frac{\langle \Delta R_k \cdot \Delta R_l \rangle}{\left[\langle \Delta R_k \cdot \Delta R_k \rangle \langle \Delta R_l \cdot \Delta R_l \rangle \right]^{1/2}} \qquad (6)$$

The lower-left and upper-right triangular portions of the correlation map display the correlated ($C(k,l) > 0$) and the anticorrelated ($C(k,l) < 0$) pairs, respectively. Four equally spaced contours are displayed in both parts, with respective values 0.21, 0.32, 0.43 and 0.54 for the correlated, and –0.14, –0.18, –0.22 and –0.26 for the anticorrelated regions. Innermost contours indicate regions with strongest cross-correlations.

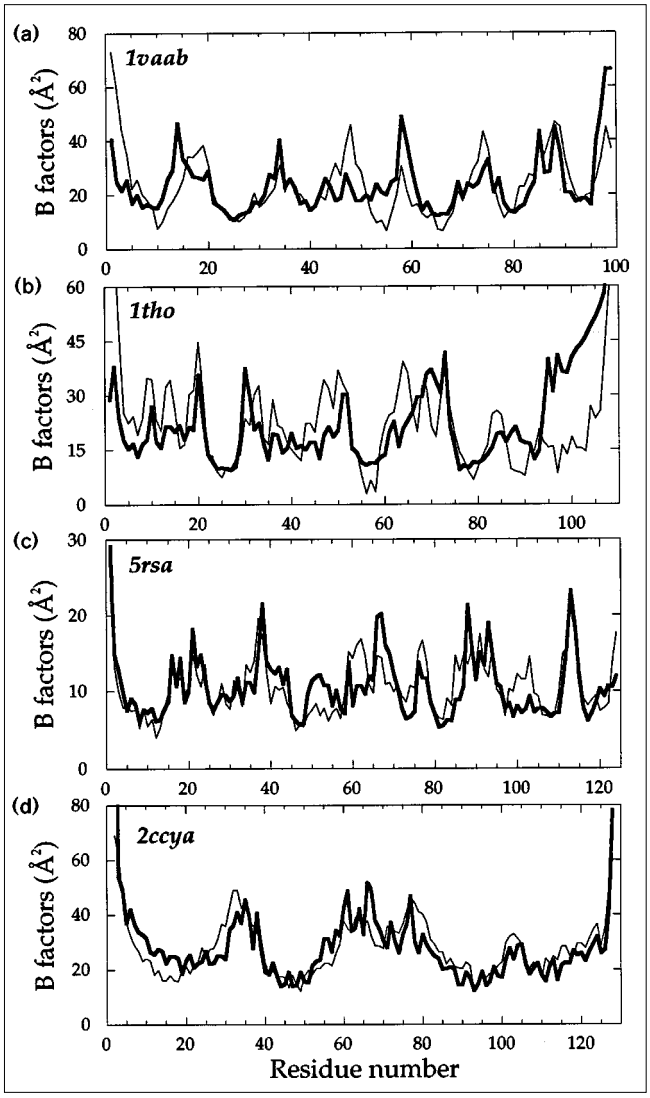
Let us first consider the regions of the molecule undergoing strongly correlated fluctuations, i.e. examine the lower-left triangular portion of the map displayed in

Figure 1



Temperature factors (*B* factors) for (a) 1ltsc, (b) 1crn, (c) 5pti and (d) 1aba, as a function of residue index. Curves shown in bold are obtained with the present theory, from the inversion of the Kirchhoff matrices (eq. 3) corresponding to the four X-ray structures reported in the PDB. Curves drawn as thin lines represent experimental data. See Table 1 for the description of the proteins.

Figure 2

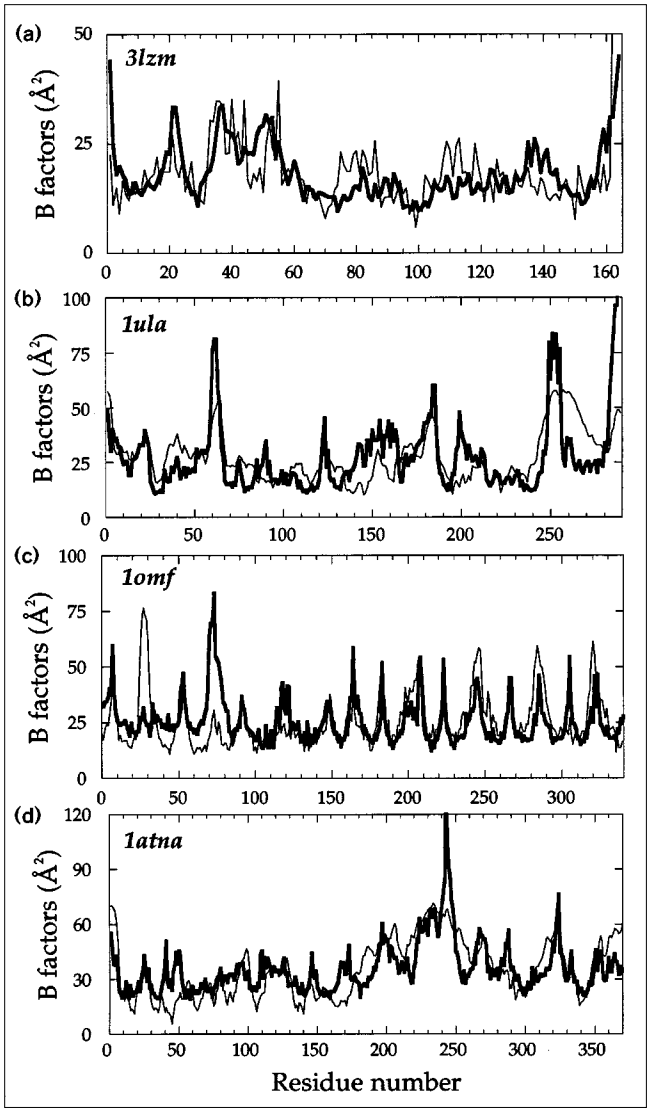


Temperature factors (*B* factors) for (a) 1vaab, (b) 1tho, (c) 5rsa and (d) 2ccya. Curves shown in bold are obtained with the present theory. Curves drawn as thin lines represent experimental data. See Table 1 for the description of the proteins.

Figure 4. As expected from chain connectivity, the positively correlated regions include the diagonal elements of the map, and a few other regions indicated by the contours. For example, two antiparallel β-strands, Ile16–Asn24 and Leu29–Tyr35, correlate strongly. Likewise, a strong correlation is observed at the C terminus among residues belonging to the helix extending between Ser47 and Gly56. We note that these results are in perfect agreement with those obtained by Levitt *et al.* [8] from NMA of the same protein. For comparative purposes, we marked on the map with 'X' signs the centers of the regions that were reported by Levitt *et al.* to exhibit the strongest correlations ($C(k,l) \geq 0.6$), omitting the diagonal portions of the map.

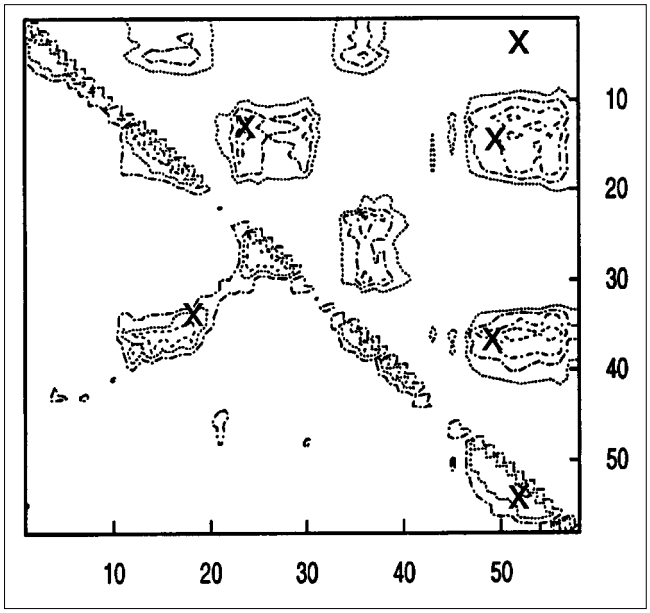
Finally, the regions of the molecule exhibiting strong anticorrelations, displayed in the upper-right triangular part of the map, also show close agreement with those identified by Levitt *et al.* [8]. The centers of the regions that were found by NMA to be subject to the strongest anticorrelation ($C(k,l) \leq -0.3$) are again indicated by the 'X' signs [8]. Four such regions were identified by NMA, three of which conform with present calculations. Only the anticorrelation between the fluctuations of the chain termini observed by

Figure 3



Temperature factors (*B* factors) for (a) 3lzm, (b) 1ula, (c) 1omf and (d) 1atna. Curves shown in bold are obtained with the present theory. Curves drawn as thin lines represent experimental data. See Table 1 for the description of the proteins.

Figure 4



Cross-correlations $C(k,l)$ between the fluctuations of C^{α} atoms k and l evaluated for trypsin inhibitor (5pti) using equation 4. The lower-left and upper-right triangular portions of the correlation map display the correlated and the anticorrelated pairs, respectively. Four equally spaced contours are shown in each region, with values of 0.21, 0.32, 0.43 and 0.54 for the correlated, and –0.14, –0.18, –0.22 and –0.26 for the anticorrelated regions. Innermost contours indicate regions with strongest cross-correlations. 'X' signs indicate the centers of the regions determined by Levitt *et al.* [8] to correlate strongly ($C(k,l) \geq 0.6$ for the lower-left triangular and $C(k,l) \leq -0.3$ for the upper-right triangular parts). The axes represent the residue numbers.

NMA, indicated by the 'X' sign in the upper-right corner of the map, could not be observed in the present analysis.

In sum, the satisfactory agreement between the present loci of cross-correlations and those extracted from NMA is quite significant, in view of the greater simplicity and computational efficiency of the present approach compared to NMA/energy minimization algorithms with complex atomic potentials.

The effect of chain connectivity

The observed fluctuations of α-carbons result from the additive contribution of two effects: chain connectivity, as accounted for by the automatic inclusion of the first neighboring α-carbons that are separated by approximately 3.8 Å into the adjacency matrix, and nonbonded interactions between all residue pairs that are not necessarily close along the backbone but are separated by ≤ 7.0 Å in space. These two contributions may be separated from each other by expressing the Kirchhoff matrix as the sum of two matrices, as in equation 7:

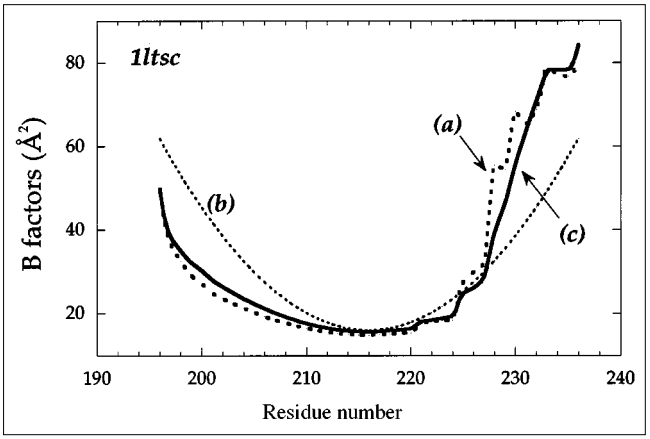
$$\Gamma = \Gamma_{nb} + \Gamma_{cc} \tag{7}$$

where the subscripts refer to the respective contributions of nonbonded interactions (*nb*) and chain connectivity (*cc*). Γ_{cc} is identical in form to the classic Rouse matrix [25] of polymer dynamics. Its *ij*th element is defined as:

$$\left[\Gamma_{cc}\right]_{ij} = \gamma^{*}(2\delta_{ij} - \delta_{i,j+1} - \delta_{i,j-1}) \tag{8}$$

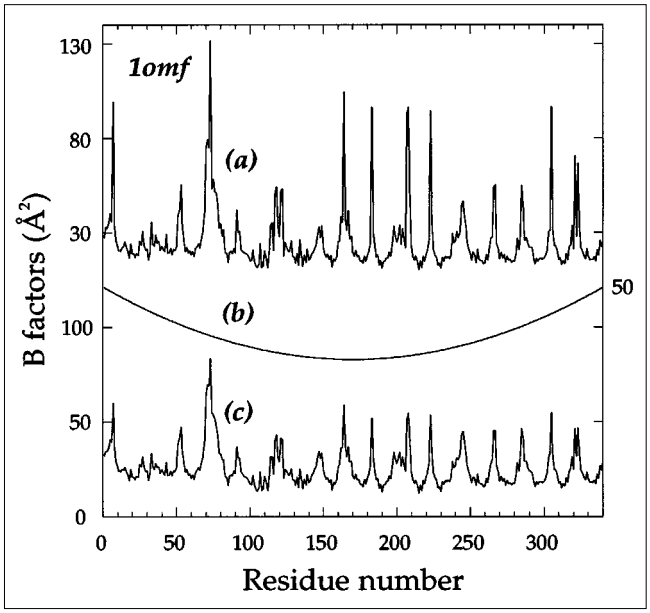
where δ_{ij} is the Kronecker delta, and γ^{*} is chosen to be identical to the one in equation 1. Without factor γ^{*}, Γ_{cc} is also known as the Kirchhoff matrix of the circle graph without chords [26].

Figure 5



Theoretical curves for the temperature factors (*B* factors) for the smallest protein examined in the present study, subunit C of heat-labile enterotoxin (1ltsc) [35]. Curves (a) and (b) are found from the inversion of the matrices Γ_{nb} and Γ_{cc}, reflecting the contributions of nonbonded interactions and chain connectivity effects, respectively. Curve (c) is obtained by considering both nonbonded and chain connectivity effects.

Figure 6



Theoretical curves for the temperature factors (*B* factors) for one of the largest proteins examined the matrix porin (1omf) [24]. Curves (a) and (b) are found from the inversion of the matrices Γ_{nb} and Γ_{cc}, reflecting the contributions of nonbonded interactions and chain connectivity effects, respectively. Curve (c) is obtained by considering both nonbonded and chain connectivity effects.

Calculations were repeated by inverting Γ_{nb} and Γ_{cc}, separately, and the distributions of atomic fluctuations obtained from the individual matrices were compared with those resulting from the inversion of the complete adjacency matrix Γ. Γ_{cc} is identical in form for all proteins and yields a parabolic distribution of the form $(i - N/2)^2$ for the ith α-carbon of a protein of N residues, i.e. fluctuations are largest at the ends of the protein and smallest at the middle.

Γ_{nb}, on the other hand, is specific to the investigated proteins, and its inversion is found to reproduce closely most of the peaks and minima obtained with the total Kirchhoff matrix Γ. Figures 5 and 6 display, for example, the distributions obtained for the smallest and one of the largest size proteins examined here. The curves found from Γ_{nb} and Γ_{cc} may be compared to the corresponding theoretical curve obtained from the inversion of Γ for each protein. It may easily be verified that Γ_{nb} plays the major role in determining the resultant curve; however, Γ_{cc} improves Γ_{nb} by either lowering the most pronounced peaks at the inner regions of the chain or by perturbing the terminal portions of the distributions in the required direction.

Correspondence between the present approach and NMA

The analysis introduced above is closely related to the NMA of elastic bodies. In order to establish the connection between the present treatment and NMA, we use the eigenvalue decomposition both for the Kirchhoff matrix Γ and for its inverse Γ^{-1}:

$$\Gamma = U\,\Lambda^2\,U^T$$
$$\Gamma^{-1} = U\,(1/\Lambda^2)\,U^T \qquad (9)$$

where U is an orthogonal matrix whose columns are the eigenvectors of Γ. Λ^2 is a diagonal matrix whose entries λ_i^2 are the eigenvalues of Γ. The $3N$ dimensional column vector Δr consisting of the fluctuations of the position vectors along the eigendirections of Γ is:

$$\Delta r = U^T\,\Delta R \qquad (10)$$

The elements of the column vector Δr are referred to as the normal coordinates in NMA [8,12]. Use of equations 9 and 10 in equation 4 leads to the classic mean-square fluctuation along the normal directions:

$$\langle \Delta r_i^2 \rangle = 1 / \lambda_i^2 \qquad (11)$$

which is identical to the second moments of the amplitude distributions in NMA [12]. The only difference is in the definition of the eigenvalues λ_i^2. In NMA, the eigenvalue problem is defined by the equation $\Gamma U = H\,U\Lambda^2$, where H is the mass matrix weighting the velocities in the kinetic energy expression [8]. The present formulation may be regarded as the zero frequency limit of NMA. A fundamental difference between NMA and the present approach is that the latter requires the linearization of the

nonlinear potential energy expression around the equilibrium state while our Γ matrix is obtained directly from the X-ray structure, and there is no need to know the potentials in our case.

Conclusions

We compared the mean-square fluctuations of C^{α} atoms obtained from the inversion of Kirchhoff matrices as $\langle \Delta R_k^2 \rangle = [\Gamma^{-1}]_{kk}$ with the temperature factors from crystallographic measurements for 12 different proteins. These were selected with reference to their sizes and structural classes, so as to test the predictions of the theory against a broad variety of protein structures. A remarkable agreement between experiments and theory is demonstrated without recourse to simulations and to detailed potentials. The results are very significant because we use an oversimplified potential, mainly a single-parameter harmonic potential based on a Gaussian distribution of interresidue distances about their native states. Thus, no specificity is introduced, and backbone constraints are ignored apart from the connectivity between successive α-carbons.

It should be noted that the experimental temperature factors derive from the rate at which the electron density falls off from the corresponding maximum for each residue [21,27]. A Gaussian decay function of the form $f = \exp(-Br^2)$ is used to fit the smearing out of the electron density f. Here, r is a reciprocal lattice vector. The use of a single B value for each residue as opposed to a six-parameter tensor implicitly assumes an isotropic displacement for that particular residue [27]. Thus, the importance of any non-Gaussian effects cannot be estimated by the present comparison. The presence of anharmonic fluctuations has been reported by Garcia [28], in contrast to the previous NMA of Noguti, Go and co-workers [11,29] in which 80% of collective fluctuations in native proteins obey harmonic potentials. We note that the recent NMA study by Tirion [5] demonstrates that contributions to fluctuations from nonlinearities in the potentials are negligibly small. A direct evidence for this is provided by the superposition of the temperature factor distribution curves obtained by assuming a harmonic potential for all interacting pairs and using the nonlinear residue-specific interaction energies derived by Levitt *et al.* [8].

The success of a model based on a single-parameter interaction energy between all near-neighboring residues irrespective of the types of amino acids suggests that residue-specific effects are of secondary importance in maintaining the overall stability of a protein's native state, but instead a network of uniform interactions between all residue pairs in contact satisfactorily account for the fluctuations near the native state. This observation is in perfect agreement with a recent analysis of the potentials of mean force that stabilize native protein structures [30]. Therein, the interresidue interactions were expressed as a sum of two contributions: a homogeneous part common to all interacting pairs, and a residue-specific part accounting for the departure of specific pairs from the homogeneous behavior. Examination of databank structures revealed that the first contribution, which may be viewed as a generic characteristic of amino acids in native structures, is stronger than the latter by a factor of five, i.e. a dominant part of the potentials of mean force experienced by individual residues is contributed by homogeneous (nonspecific) interactions [30]. This result is consistent with the observation that a single-parameter interresidue potential yields an adequate description of fluctuations in the present study. Thus, one might expect to have a single force constant γ* for all proteins, inasmuch as the latter is representative of the homogeneous interresidue interactions existing in all proteins in their native state. As mentioned above, deviations from a common value of γ* are partly due to differences in local packing densities in different proteins.

It should be noted that one may totally eliminate the variable γ* by normalizing both the experimental and theoretical results. The experimentally observed absolute amplitudes may indeed be affected by various factors, including static disorder, temperature, or other environmental effects. The relative mobilities of different residues, rather than the absolute amplitudes of fluctuations, are of interest in general. And the present Gaussian model of fluctuations provides an excellent description of the relative amplitudes of fluctuations of the individual residues.

The present model gives the zero frequency limit of conformational fluctuations in proteins. At this limit, biological macromolecules behave like an elastic body [11,31]. The present model treats the elastic body as a collection of points interacting with a single harmonic potential. Torsional bond-stretching and bond-bending potentials are not included in the model. Agreement of calculations with experimental data justifies their omission. Even chain connectivity represented by the matrix Γ_{cc} is of minor importance. The reason for the very good agreement between experiments and our simple model based on Gaussian fluctuations of interresidue distances in the native state is that the dominant vibrational modes involve the cooperative contribution of a large number of atoms. Averaging over these numerous interactions demonstrates a universality due to the central limit theorem [32], irrespective of the details buried in the individual pairwise interactions.

The computation time for a given protein is limited by the inversion of the Kirchhoff matrix. This time is several orders of magnitude shorter than that for typical NMA and MD simulations. For instance, for DNase I–actin with two subunits A and D having a total of 633 residues, 9.4 min CPU time is required for calculating both B factors and

cross-correlations. Obviously, calculations are much faster in the case of smaller proteins. For T4 lysozyme, for example, which comprises 164 residues, a CPU time of 8.2 s is sufficient. These calculations are performed on a Silicon Graphics Power Indigo workstation. Note that in comparison with the CPU times reported by Tirion [5] for NMA with energy minimization, the present computation times are at least one order of magnitude shorter.

Computations with the present model require a single inversion of the $N \times N$ Kirchhoff matrix for a protein of N residues. This forms another advantage of the present model over the conventional computation methods, mainly its applicability to proteins of any size.

It should be remembered that the theory relies on knowledge of atomic coordinates or, more precisely, on the list of residue pairs whose equilibrium C^{α}–C^{α} distance is lower than a cutoff value, say 7.0 Å. These data are provided by experiments. Determination of the atomic coordinates from knowledge of amino acid sequence, on the other hand, is a different and still unresolved issue which cannot be solved by such simple harmonic potentials, but necessitates the development and use of highly specific inter-residue potentials cognate to models of different complexities [30,33]. Yet, it would be challenging to test the predictions of the present simple theory as applied to partially known structures for which some distance restraint data are available. Such data are provided by NMR experiments, for example. Another interesting application would be the characterization of type and strength of correlations on a larger scale, such as those between secondary structural elements or between structural domains.

In sum, the present study shows that, insofar as the correlations between atomic fluctuations are concerned, a single-parameter harmonic potential may be satisfactorily adopted for each protein in its folded state. The temperature factors and equilibrium correlations between atomic fluctuations predicted by the theory show very close agreement with experimental observations or with results of significantly more complex NMA or MD simulations. And these results are computed very efficiently, at least one order of magnitude faster than other atomic approaches developed to date.

Materials and methods

The Kirchhoff matrix for the protein under consideration is formed following the definition given by equation 3. The cutoff separation defining the range of nonbonded contacts r_c is taken as 7.0 Å. This value is representative of the distance range of interactions for pairs of nonbonded amino acids in compact globular proteins [34]. All pairs of C^{α} atoms i and j are taken into consideration in evaluating Γ, which automatically incorporates the two terms Γ_{nb} and Γ_{cc}, contributed by nonbonded interactions ($|j - i| \geq 2$) and bonded neighbors ($|j - i| = 1$), residues being indexed from 1 to N starting from the N terminus. Auto-correlations and cross-correlations between the fluctuations of C^{α} atoms are evaluated from the inversion of the Kirchhoff adjacency matrix, using equation 4.

We note that the determinant of the Kirchhoff matrix is equal to zero, and hence the matrix cannot be inverted directly. Instead, the matrix is subjected to a similarity transformation and reconstructed after eliminating the zero eigenvalue. Results are expressed in terms of the parameter γ^*, the value of which is adjusted for each protein (see Table 1) upon comparison with experimental temperature factors, following equation 5.

Acknowledgement

The authors gratefully acknowledge partial support from Bogazici University Research Fund, project numbers 96P0003 and 96A0430.

References

1. Jernigan, R.L. & Bahar, I. (1996). Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* 6, 195–209.
2. Sippl, M.J. (1995). Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* 5, 229–235.
3. Srinivasan, R. & Rose, G.D. (1995). LINUS: a hierarchic procedure to predict the fold of a protein. *Proteins* 22, 81–99.
4. Yue, K. & Dill, K.A. (1996). Folding proteins with a simple energy function and extensive conformational searching. *Protein Sci.* 5, 254–261.
5. Tirion, M.M. (1996). Large amplitude elastic motions in proteins from a single-parameter atomic analysis. *Phys. Rev. Lett.* 77, 1905–1908.
6. Levitt, M. (1983). Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* 168, 595–620.
7. Brooks, B. & Karplus, M. (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl Acad. Sci. USA* 80, 6571–6575.
8. Levitt, M., Sander, C. & Stern, P.S. (1985). Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* 181, 423–447.
9. Horiuchi, T. & Go, N. (1991). Projection of Monte Carlo and molecular dynamics trajectories onto the normal mode axes: human lysozyme. *Proteins* 10, 106–116.
10. Kidera, A. & Go, N. (1992). Normal mode refinement: crystallographic refinement of protein dynamic structure I. Theory and test by simulated diffraction data. *J. Mol. Biol.* 225, 457–475.
11. Go, N., Noguti, T. & Nishikawa, T. (1983). Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA* 80, 3696–3700.
12. Case, D.A. (1994). Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* 4, 285–290.
13. Hayward, S., Kitao, A. & Go, N. (1995). Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis. *Proteins* 23, 177–186.
14. Marques, O. & Sanejouand, Y.-H. (1995). Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins* 23, 557–560.
15. Flory, P.J. (1976). Statistical thermodynamics of random networks. *Proc. R. Soc. Lond. A* 351, 351–380.
16. Pearson, D.S. (1977). Scattered intensity from a chain in a rubber network. *Macromolecules* 10, 696–701.
17. Kloczkowski, A., Mark, J.E. & Erman, B. (1989). Chain dimensions and fluctuations in random elastomeric networks. 1. Phantom Gaussian networks in the undeformed state. *Macromolecules* 22, 1423–1432.
18. Eichinger, B.E. (1972). Elasticity theory. I. Distribution functions for perfect phantom networks. *Macromolecules* 5, 496–505.
19. Bernstein, F.C., *et al.*, & Tasumi, M. (1977). The protein databank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535–542.
20. Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. & Weng, J. (1987). In *Crystallographic Databases – Information Content Software Systems, Scientific Applications.* (Allen, F.H., Bergerhoff, G. & Sievers, R., eds). pp. 107. Data Commision of the International Union of Crystallography, Bonn, Cambridge and Chester.
21. Frauenfelder, H., Petsko, G.A. & Tsernoglou, D. (1979). Temperature dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* 280, 558–563.
22. Ealick, S.E., *et al.*, & Bugg, C.E. (1990). Three-dimensional structure of human erythrocytic purine nucleoside phosphorylase at 3.2 Å resolution. *J. Biol. Chem.* 265, 1812–1820.
23. Kabsch, W., Mannherz, H.G., Suck, D., Pai, E.F. & Holmes, K.C. (1990). Atomic structure of the actin:DNase I complex. *Nature* 347, 37–44.

24. Cowan, S.W., *et al*., & Rosenbusch, J.P. (1992). Crystal structures explain functional properties of two *E. coli* porins. *Nature* 358, 727–733.
25. Rouse, P.E. (1953). A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *J. Chem. Phys.* 21, 1272–1280.
26. Eichinger, B.E. (1973). Self-avoiding random walks, circle and star graphs, and a group of singular matrices. *J. Chem. Phys.* 59, 5787–5795.
27. Frauenfelder, H. & Petsko, G.A. (1980). Structural dynamics of liganded myoglobin. *Biophys. J.* 53, 465–483.
28. Garcia, A.E. (1992). Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68, 2696–2699.
29. Noguti, T. & Go, N. (1982). Collective variable description of small amplitude conformational fluctuations in a globular protein. *Nature* 296, 776–778.
30. Bahar, I. & Jernigan, R.L. (1997). Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* 266, 195–214.
31. Ben-Avraham, D. & Tirion, M.M. (1995). Dynamic and elastic properties of F-actin: a normal modes analysis. *Biophys.* J. 68, 1231–1245.
32. Feller, W. (1968). *An Introduction to Probability Theory and its Applications.* Wiley, New York.
33. Bahar, I. & Jernigan, R.L. (1996). Coordination geometry of nonbonded residues in globular proteins. *Fold. Des.* 1, 357–370; 1359-0278-001-00357.
34. Miyazawa, S. & Jernigan, R.L. (1985). Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* 18, 534–552.
35. Sixma, T.K., Pronk, S.E., Kalk, K.H., Van Zanten, B.A.M., Berghuis, A.M. & Hol, W.G. (1992). Lactose binding to heat-labile enterotoxin revealed by X-ray crystallography. *Nature* 355, 561–564.
36. Teeter, M.M. & Hendrickson, W.A. (1979). Highly ordered crystals of the plant seed protein crambin. *J. Mol. Biol.* 127, 219–223.
37. Teeter, M.M. (1984). Water structure of a hydrophobic protein at atomic resolution: pentagon rings of water molecules in crystals of crambin. *Proc. Natl. Acad. Sci. USA* 81, 6014– 6018.
38. Wlodawer, A., Deisenhofer, J. & Huber, R. (1987). Comparison of two highly refined structures of bovine pancreatic inhibitor. *J. Mol. Biol.* 193, 145–156.
39. Sjöberg, B.-M. & Söderberg, B.-O. (1976). Thioredoxin induced by bacteriophage T4: crystallization and preliminary crystallographic data. *J. Mol. Biol.* 100, 415–419.
40. Söderberg, B.-O., Sjöberg, B.-M., Sonnerstam, U. & Brandén, C.-I. (1978). Three-dimensional structure of thioredoxin induced by bacteriophage T4. *Proc. Natl. Acad. Sci. USA* 75, 5827–5830.
41. Stura, E.A., Matsumura, M., Fremont, D.H., Saito, Y., Peterson, P.A. & Wilson, I.A. (1992). Crystallization of murine major histocompatibility complex class I H-2K^{b} with single peptides. *J. Mol. Biol.* 228, 975–982.
42. Katti, S.K., LeMaster, D.M. & Eklund, H. (1990). Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. *J. Mol. Biol.* 212, 167–184.
43. Wlodawer, A., Bott, R. & Sjolin, L. (1982). The refined crystal structure of ribonuclease A at 2.0 Å resolution. *J. Biol. Chem.* 257, 1325–1332.
44. Finzel, B.C., Weber, P.C., Hardman, K.D. & Salemme, F.R. (1985). Structure of ferricytochrome c′ from rhodospirillium molischianum at 1.67 Å resolution. *J. Mol. Biol.* 186, 627–643.
45. Weaver, L.H. & Matthews, B.W. (1987). Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol.* 193, 189–199.
46. Matthews, B.W., Nicholson, H. & Becktel, W.J. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. USA* 84, 6663–6667.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/fad.htm – for further information, see the explanation on the contents pages.