Statistical thermodynamics of residue fluctuations in native proteins

Osman N. Yogurtcu, Mert Gur, and Burak Erman[a)]

Center for Computational Biology and Bioinformatics, Koc University, Sariyer 34450, Istanbul, Turkey

Statistical thermodynamics of residue fluctuations of native proteins in a temperature, pressure, and force reservoir is formulated. The general theory is discussed in terms of harmonic and anharmonic fluctuations of residues. The two elastic network models based on the harmonic approximation, the anisotropic network and the Gaussian network models are discussed as the limiting cases of the general theory. The heat capacity and the correlations between the energy fluctuations and residue fluctuations are obtained for the harmonic approximation. The formulation is extended to large fluctuations of residues in order to account for effects of anharmonicity. The fluctuation probability function is constructed for this purpose as a tensorial Hermite series expansion with higher order moments of fluctuations as coefficients. Evaluation of the higher order moments using the proposed statistical thermodynamics model is explained. The formulation is applied to a hexapeptide and the fluctuations of residues obtained by molecular dynamics simulations are characterized in the framework of the model developed. In particular, coupling of two different modes in the nonlinear model is discussed in detail. © *2009 American Institute of Physics*. [DOI: 10.1063/1.3078517]

I. INTRODUCTION

A protein in aqueous solution constitutes a system whose atoms exhibit fluctuations over time about well defined mean positions. The aqueous medium forms the reservoir at constant temperature and pressure. The magnitude of fluctuations may be large relative to atomic radii as indicated by experiment. Fluctuations in atomic coordinates are well characterized by experiments.[1] In theory, fluctuations are studied at various levels of approximation, ranging from all-atom to coarse-grained scales. Studying the fluctuations of the α-carbons is a convenient approximation where each successive α-carbon pair is assumed to be connected by a virtual bond of fixed length and only interactions between residues, represented by their α-carbons, are considered. In the present study, we adopt this level of approximation.

Coarse-grained models of fluctuations started with the important observation that the large amplitude fluctuations of the protein G-actin could be described in the harmonic approximation by a single parameter only.[2] Based on this simple picture of the elastic fluctuations of a protein, the Gaussian network model (GNM) was proposed,[3,4] according to which the C^{α}'s were assumed analogous to the junctions of an amorphous network whose fluctuations were similar to those given in the random amorphous network model proposed by Flory.[5,6] As in the random network model, the GNM is based on an isotropic description of residue fluctuations where only the number of neighbors of a given residue is important. The anisotropic network model (ANM) was then introduced to estimate the directions of fluctuations.[7,8] The GNM and models that followed it, collectively referred to as the elastic network models (ENMs) are found to provide important insights for understanding the structure-function relations of proteins. For this reason, and because of their immediate applicability to all kinds of proteins without size restrictions, they found wide use during the past decade.[4,9–12] In general, these studies and several others that are cited by them, elaborate on different levels of approximation of the ENMs. They try to identify the force constants associated with the models, compare the different models, associate the models with NMR data, optimize the model parameters over databases, and apply the models to drug design problems and prediction of binding sites, folding cores, allosteric effects, and hot residues. In addition to work in harmonic fluctuations cited here, anharmonicities of protein fluctuations[13,14] in the form of nonlinear modes that are localized in certain regions of the protein play important roles in protein function.[15,16] In this respect, coupling of fast and slow modes resulting in energy flow is the most important process responsible for the protein's function.[17]

Despite this wide range of interest, a general statistical mechanical treatment of fluctuations that describes the theoretical basis of harmonic as well as anharmonic behavior is missing in literature. The specific aim of the present paper is to give a statistical thermodynamic interpretation of fluctuations in native proteins that covers both harmonic and anharmonic behavior.

The paper consists of three major parts: In the first part, we introduce the statistical thermodynamics basis of fluctuations in native proteins. We discuss, in some detail, the pairwise inter-residue energies that play significant role in the model. In the second part, we obtain the harmonic approximation as a special case of the general formalism for fluctuations, and discuss the two most widely used models, ANM and GNM. We also discuss two simple applications of the thermodynamic formalism by deriving the heat capacity and correlations of energy and residue fluctuations of the GNM. In the third part, the effects of anharmonicities are introduced into the probability function of fluctuations, in

---

[a)]Electronic mail: berman@ku.edu.tr.

87 terms of moment based tensorial Hermite series expansion.
88 As a simple application of the series formalism, anharmonic
89 fluctuations of a hexapeptide are obtained by molecular dy-
90 namics (MD) simulations, and the distribution functions for
91 the fluctuations of C^α's are determined. Contributions from
92 different modes of fluctuations to a given mode through
93 mode coupling are discussed in terms of the Hermite series
94 expansion.

95 II. THEORY

96   In this section, we present the thermodynamic and sta-
97 tistical basis of fluctuations in native proteins. We use the
98 entropy representation for the fundamental relation,[18]

99    $$S = S(U, V, \boldsymbol{R}), \tag{1}$$

100 where S, U, V, and $\boldsymbol{R}$ are the mean (thermodynamic) values
101 of the entropy, energy, volume, and position vectors of C^α's,
102 respectively. Water is not shown explicitly in the fundamen-
103 tal relation and only a single protein molecule is considered.
104 The protein is in diathermal contact with the surrounding
105 water. Similarly, the protein is in contact with a pressure (P)
106 and a force ($\boldsymbol{F}$) reservoir, as a result of which the energy,
107 volume and the positions of residues exhibit fluctuations.
108 Other proteins are present in the surroundings but they do
109 not influence the energy levels of the given protein. We call
110 the protein and the surrounding water as an element. The
111 collection of all elements of the system constitutes the en-
112 semble. Statistical mechanics is applicable to a single ele-
113 ment. Thermodynamics applies only to an ensemble of the
114 elements. The ensemble of elements with its extensive prop-
115 erties constitutes a macroscopic system.[18,19] The thermody-
116 namic variables $S, U, V, \boldsymbol{R}$ are obtained from the ensemble.
117 For each element, these variables exhibit fluctuations about
118 their native values. The distribution $f(\hat{U}, \hat{V}, \hat{\boldsymbol{R}})$ of the instan-
119 taneous extensive variables $\hat{U}, \hat{V}, \hat{\boldsymbol{R}}$ are given by the relation

120
121
$$f(\hat{U}, \hat{V}, \hat{\boldsymbol{R}}) = \exp\left\{ -k^{-1} S\left[\frac{1}{T}, \frac{P}{T}, \frac{\boldsymbol{F}}{T} \right] \right. \\ \left. - k^{-1}\left(\frac{1}{T}\hat{U} + \frac{P}{T}\hat{V} - \frac{\boldsymbol{F}}{T} \cdot \hat{\boldsymbol{R}} \right) \right\}, \tag{2}$$

122 where k is the Boltzmann constant and $S[1/T, P/T, \boldsymbol{F}/T]$ is
123 the Massieu transform of the entropy, which for the specified
124 thermodynamic variables chosen reads as

125
$$S\left[\frac{1}{T}, \frac{P}{T}, \frac{\boldsymbol{F}}{T} \right] = S - \frac{U}{T} - \frac{P}{T}V + \frac{\boldsymbol{F}}{T} \cdot \boldsymbol{R}. \tag{3}$$

126 The distribution now takes the explicit form

127
128
$$f(\hat{U}, \hat{V}, \hat{R}) = \exp\left\{ -k^{-1}\left(S - \frac{U}{T} - \frac{P}{T}V + \frac{\boldsymbol{F}}{T} \cdot \boldsymbol{R} \right) \right. \\ \left. - k^{-1}\left(\frac{\hat{U}}{T} + \frac{P}{T}\hat{V} - \frac{\boldsymbol{F}}{T} \cdot \hat{\boldsymbol{R}} \right) \right\}. \tag{4}$$

129 In Eq. (4), provided that the system remains around the given
130 equilibrium point, i.e., a point on the thermodynamic surface
131 $S = S(U, V, \boldsymbol{R})$, there are no restrictions on the degree of de-
132 parture of the system, i.e., the magnitude of fluctuations,

133 from the average thermodynamic variables. If the fluctua-
134 tions are large, the fluctuations may be anharmonic or may
135 induce a jump from one local minimum to another. The ap-
136 plicability of results derived from Eq. (4) is discussed in
137 detail in Secs. III–V.   **AQ:** **#1**
138   The correlation of fluctuations of the ith and jth residues
139 may now be obtained from

140
$$\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^T \rangle = \sum (\hat{\boldsymbol{R}}_i - \boldsymbol{R}_i)(\hat{\boldsymbol{R}}_j - \boldsymbol{R}_j)^T f(\hat{U}, \hat{V}, \hat{\boldsymbol{R}}), \tag{5}$$

141 where the superscript T denotes transpose and the summation
142 is over all allowable states.
143   Using Eq. (4) in Eq. (5) leads to

144
$$\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^T \rangle = kT \left(\frac{\partial \boldsymbol{R}_i}{\partial \boldsymbol{F}_j} \right)_{T,P,F \; i\neq j}, \tag{6}$$

145 where the variables to be kept fixed are indicated as sub-
146 scripts. The equation is valid when the system is in or close
147 to equilibrium. The derivation of Eq. (6) is given by Callen,[18]
148 which is outlined briefly in Appendix A.
149   In general, if Φ_k represents any of the extensive vari-
150 ables ΔU, ΔV, ΔR, and Ψ_k represent the conjugate variables
151 $1/T$, P/T, $-F/T$, then, in principle, all higher moments can
152 be derived iteratively according to the rule[18]

153
$$\langle \phi \Delta \Phi_k \rangle = -k \frac{\partial}{\partial \Psi_k} \langle \phi \rangle - k \left\langle \frac{\partial \phi}{\partial \Psi_k} \right\rangle, \tag{7}$$

154 where ϕ is a higher order product of the fluctuations of the
155 extensive variables, ΔU, ΔV, ΔR. The product of the form
156 $\Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j \Delta \boldsymbol{R}_k \cdots$ is an example to ϕ that leads to higher order
157 moments of residue position fluctuations.
158   Equation (6) forms the statistical mechanical basis of all
159 ENMs for fluctuations in native proteins. Assuming that the
160 protein is in equilibrium, the right-hand side of Eq. (6) may
161 be evaluated if the energy of the system is known as a func-
162 tion of residue positions. For the case of pairwise potentials,
163 the most general form of this relation is

164
$$E_{ij} = E_{ij}{}^0 f_{ij}\left(\frac{R_{ij}}{R_{ij}{}^0} \right), \tag{8}$$

165 where R_{ij} is the distance between residues i and j and f_{ij} is a
166 dimensionless function. $E_{ij}{}^0$ is the reference interaction en-
167 ergy, and $R_{ij}{}^0$ is a reference length, both of which will be
168 discussed in detail below.

A. The forces   169

170   The right-hand side of Eq. (3) can be expressed in terms
171 of the independent variables T, P, and $\boldsymbol{F}$ as $-\Psi(T, P, \boldsymbol{F})/T$.
172 Knowing this relationship leads to the following five equa-
173 tions:

174
175
$$\Psi = \Psi(T, P, F), \quad \Psi = U - TS + PV - FR, \tag{9}$$

176
$$S = -\frac{\partial \Psi}{\partial T}, \quad V = \frac{\partial \Psi}{\partial P}, \quad R = -\frac{\partial \Psi}{\partial F},$$

177 where $\boldsymbol{F}$ and $\boldsymbol{R}$ are $3N$ dimensional, but here we represented
178 them as scalars for the clarity of the discussion. The four
179 variables Ψ, T, P, and F may be eliminated among these five

180 equations to yield $U = U(S, V, R)$. The forces are then ob-
181 tained from U according to the relation $F = -\partial U(S, V, R)/\partial R$.
182 Considering pairwise potentials E_{ij} and concentrating on the
183 position variables only, i.e., neglecting S and V dependence,
184 the forces may be written as

$$F_i = -\nabla_{R_i} \sum_j E_{ij}. \tag{10}$$

185

186 The gradient $\partial E_{ij}/\partial R_j$ in Eq. (10) is obtained by the chain
187 rule

$$\nabla_{Ri} E_{ij} = \left(\frac{\partial E_{ij}}{\partial R_{ij}^2} \right) \frac{\partial R_{ij}^2}{\partial R_j}$$

188

$$= \left(\frac{\partial E_{ij}}{\partial R_{ij}^2} \right) \frac{\partial (R_i \cdot R_i - 2R_i \cdot R_j + R_j \cdot R_j)}{\partial R_j}$$

189

$$= 2 \left(\frac{\partial E_{ij}}{\partial R_{ij}^2} \right) (R_j - R_i). \tag{11}$$

190

191 The term in the first parenthesis in the second line is the
192 slope of E_{ij} with respect to R_{ij}^2. For a given i and j, it is a
193 scalar quantity whose value depends solely on the type of the
194 energy function used. The vectorial property of the force
195 comes as the term $R_j - R_i$ in the second parenthesis. Equa-
196 tions (10) and (11) may be arranged in matrix form as

$$197 \quad F = \Lambda^{(3N)} R, \tag{12}$$

198 where $\Lambda^{(3N)}$ is the $3N \times 3N$ matrix defined as

$$\Lambda_{ij} = \begin{cases} -2 \dfrac{\partial E_{ij}}{\partial R_{ij}^2} & i \neq j \\ -\displaystyle\sum_{i \neq k} \Lambda_{ik} & i = j. \end{cases} \tag{13}$$

199

200 Two different ordering of the $\Lambda^{(3N)}$ matrix is used in the
201 study of ENMs. We name them as block representation and
202 standard MD representation. For details see Appendix B. In
203 the block representation described in Appendix B, Eq. (11)
204 reads as

$$\begin{bmatrix} F_X \\ F_Y \\ F_Z \end{bmatrix} = \begin{bmatrix} \Lambda_X^{(N)} & 0 & 0 \\ 0 & \Lambda_Y^{(N)} & 0 \\ 0 & 0 & \Lambda_Z^{(N)} \end{bmatrix} \begin{bmatrix} R_X \\ R_Y \\ R_Z \end{bmatrix}. \tag{14}$$

205

206 Here, the subscripts denote the X, Y, or Z components, and
207 $\Lambda_X^{(N)}$, $\Lambda_Y^{(N)}$, and $\Lambda_Z^{(N)}$ are $N \times N$. It is to be noted that
208 irrespective of the form of the energy function, the three
209 submatrices in Eq. (14) are identical, as can be seen from the
210 definition of the derivative given in Eq. (11). The force and
211 position vectors in Eq. (14) are thermodynamic quantities,
212 i.e., average values, and at equilibrium, $F = \Lambda^{(3N)} R^{eq} = 0$.
213   Secs. III and IV, we will use the block representation.
214 The order of the matrices, $3N \times 3N$ or $N \times N$, will be self-
215 evident and will not be shown explicitly unless needed for
216 clarity.

AQ: #2

B. The correlation matrix

217
218   Correlations among the fluctuations of residues are given
219 by Eq. (6). which requires the evaluation of the derivative
220 $\partial F_i/\partial R_j$. Using Eq. (14), this derivative is written as

$$\frac{\partial F_i}{\partial R_k} = \Lambda_{ij} \frac{\partial R_j}{\partial R_k} + \frac{\partial \Lambda_{ij}}{\partial R_k} R_j \equiv \Lambda_{ik} + \Lambda'_{ik} \equiv \Gamma_{ik}. \tag{15}$$

221

222 Here, $\Lambda'_{ik} = (\partial \Lambda_{ij}/\partial R_k) R_j$ where $\partial \Lambda_{ij}/\partial R_k$ is third order, and
223 its inner product with the position vector R_j gives a second
224 order matrix that has the following block form:

$$\begin{bmatrix} \Lambda'_{XX} & \Lambda'_{XY} & \Lambda'_{XZ} \\ & \Lambda'_{YY} & \Lambda'_{YZ} \\ & & \Lambda'_{ZZ} \end{bmatrix}, \tag{16}$$

225

226 where the symmetric lower half is not shown. The block
227 matrices are of dimensions $N \times N$, with

$$\Lambda'_{XX} = \begin{cases} \dfrac{\partial \Lambda_{ij}}{\partial R_{ij}^2}(X_j - X_i)^2 = -2 \dfrac{\partial^2 E_{ij}}{\partial (R_{ij}^2)^2}(X_j - X_i)^2 & i \neq j \\ -\displaystyle\sum_{k \neq j} \Lambda'_{jk} & i = j, \end{cases} \tag{17}$$

228

229 where X_i and X_j are the X-components of the ith and jth
230 residues, respectively. The terms for Λ'_{YY} and Λ'_{ZZ} are ob-
231 tained similarly where Y and Z replaces the X's respectively.
232 The first off-diagonal term Λ'_{XY} is obtained as

$$\Lambda'_{XY} = \begin{cases} \dfrac{\partial \Lambda_{ij}}{\partial R_{ij}^2}(X_j - X_i)(Y_j - Y_i) \\ = -2 \dfrac{\partial^2 E_{ij}}{\partial (R_{ij}^2)^2}(X_j - X_i)(Y_j - Y_i) & i \neq j \\ -\displaystyle\sum_{k \neq j} \Lambda'_{jk} & i = j. \end{cases} \tag{18}$$

233

234 The terms for the other off-diagonal blocks are written simi-
235 larly by replacing the variables in Eq. (18) accordingly. The
236 components of the Λ' matrix may thus be written in compact
237 form as

$$\Lambda'_{\xi\eta} = \begin{cases} -2 \dfrac{\partial^2 E_{ij}}{\partial (R_{ij}^2)^2}(\xi_j - \xi_i)(\eta_j - \eta_i) & i \neq j \\ -\displaystyle\sum_{k \neq j} \Lambda'_{jk} & i = j, \end{cases} \tag{19}$$

238

239 where ξ and η represent the coordinates, X, Y, or Z at the
240 given equilibrium or reference state.
241   The gradient of Λ' given in Eq. (15) is obtained by first
242 taking the first column of the Λ matrix, taking its gradient,
243 which gives a vector and then dot this with R and obtain the
244 first column of Λ'. Applying the same operation to the re-
245 maining columns of Λ leads to the $3N \times 3N \Lambda'$ matrix. Re-
246 arranging the terms leads to Eq. (19). Substituting Eq. (15) in
247 Eq. (6) leads to

$$\langle \Delta R_i \Delta R_j^T \rangle = kT([\Lambda + \Lambda']^{-1})_{ij} = kT(\Gamma^{-1})_{ij}. \tag{20}$$

248

249 Equation (20) is the fundamental relation expressing the cor-
250 relations in terms of the inverse of the Γ matrix.

C. Comments on the energy function and the applicability of Equation (6) in general

In its most general form, the pairwise additive energy of the system in the coarse-grained approximation from Eq. (8) is

$$U = \frac{1}{2}\sum_{i,j} E_{ij}{}^0 f_{ij}\left(\frac{R_{ij}}{R_{ij}{}^0}\right).\tag{21}$$

Here, $E_{ij}{}^0$ and $R_{ij}{}^0$ are the energy and distance parameters that characterize the native state such that $dU = 0$ and $d^2U > 0$. The first condition gives

$$dU = \sum_i \nabla_{\boldsymbol{R}i} U \cdot d\boldsymbol{R}_i = \frac{1}{2}\sum_i \left[\nabla_{\boldsymbol{R}i}\sum_j E_{ij}{}^0 f_{ij}\left(\frac{R_{ij}}{R_{ij}{}^0}\right)\right]\cdot d\boldsymbol{R}_i$$

$$= 0.\tag{22}$$

There is a special functional form of $f_{ij}(R_{ij}/R_{ij}{}^0)$ in Eq. (22) where the reference distance $R_{ij}{}^0$ is chosen equal to the equilibrium distance $R_{ij}{}^{eq}$ between i and j in the native state in the presence of all other residues. In this case, at equilibrium the gradient in the square brackets will equate to zero independently for each term in the summation. We call this form of the energy function "the standard form" because at equilibrium, the term in the square brackets is a minimum for all i and j. This advantage of the standard form rests on the *a priori* knowledge of the equilibrium state. However, the standard form may not always be readily available for proteins at equilibrium in an aqueous environment because the coordinates of the native protein are usually obtained from crystal structure x-ray data, which may differ significantly from those at equilibrium in water. In this case, MD simulations of the protein in water are needed to establish $R_{ij}{}^{eq}$ and $E_{ij}{}^{eq}$ for recovering the standard form of the energy function. An example of the standard form is given in Sec. III on the harmonic approximation. It is to be noted here that when the energy function is given in the standard form, $\Lambda_{ij} \equiv 0$ and only the Λ' matrix contributes to the correlation of fluctuations. The Γ matrix, the inverse of which gives the second order correlations, has first and second derivatives of the energy with respect to position, which corresponds to information on the local structure of the energy function only. Higher order correlations that can be obtained by Eq. (7) require higher order derivatives of the pair potential. This equation contains the n-1st derivative of the Γ matrix, or the nth derivative of the pair potential, when it is used to obtain the nth order correlation. Characterization of the full statistical features of fluctuations with the proposed model requires the knowledge of all order derivatives of the pair potential which is equivalent to the knowledge of the full probability distribution function. The second order correlations $\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j{}^T\rangle$ containing the effects of anharmonicities can only be evaluated when $\Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j{}^T$ is averaged using the full probability function. The full probability density function is presented below in the form of a tensorial Hermite polynomial, the coefficients of which are given by the general theory presented here. However, it is also possible, by using a perturbation scheme, to evaluate the second order correlation matrix without resort to the full distribution function. In the remaining part of this section, we show that the Γ matrix can consistently be renormalized to include the effects of anharmonicities of the pair potentials.

The pair potentials, and therefore the full energy of the system, do not need to possess a unique minimum. Indeed, amino acid pair potentials with multiple minima are common.[20] The only requirement for Eq. (6) to be valid is that the fluctuations should not move the system too much away from equilibrium. If the latter is the case, then the variables U, V, and $\boldsymbol{R}$ will not be sufficient to describe the behavior of the system, and additional independent variables will be necessary. Several factors contribute to the deviation of a given protein from the minimum energy configuration. The manifold of internal constraints that the protein is subject to around the native state sets the mechanisms by which a protein may move from one stable energy state to another during its fluctuations. Existence of a pair of neighboring residues in one or the other pair-energy minima is of this type. Given the state of the protein, Eq. (6) then gives the correlations. The chain relations given in Eq. (11) characterize the energy surface at the given conformation of the protein, but also leads to the question of whether the elements of the matrix Γ consists of only the local slope and curvatures of pair potentials [see Eqs. (13) and (19)]. If this were the case, then the general result given in Eq. (6) would not reflect the effects of anharmonicities on second order correlations. As stated above, introduction of anharmonicities into the second order correlations needs the construction of the full probability distribution function, Eq. (2). As an alternative route, in the remainder of this section, we show, by a perturbation scheme, that any feature of the pair potential may be incorporated into the second order correlation matrix $\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j{}^T\rangle$.

For illustrative purposes, let us assume that the potential between every pair of residues is strictly harmonic, except one pair, say residues m and n, for which the potential differs from E_{mn} by a small amount $\Delta E_{mn}(\varepsilon)$, where ε is a measure of deviation from harmonicity. This will change the elements mm, nn, mn, and nm of the Γ matrix. which will take the following form:

$$\Gamma(\varepsilon) = \Gamma(0) + \varepsilon V,\tag{23}$$

where we denote the unperturbed state by zero. ε is small and V is the matrix all of whose elements are zero except mm, nn, mn, and nm, such that[21]

$$V = \frac{\partial \Gamma(\varepsilon)}{\partial \varepsilon}.\tag{24}$$

The eigenvalues and eigenvectors of $\Gamma(0)$ are $\lambda_i(0)$ and $u_i(0)$. The perturbed eigenvalues and eigenvectors are as follows:

$$\lambda_k(\varepsilon) = \lambda_k(0) + \varepsilon u_k(0)^T V u_k(0) + \varepsilon^2 \sum_{l\neq k}\frac{u_l(0)^T V u_k(0)}{\lambda_k(0) - \lambda_l(0)},$$

$$\tag{25}$$

$$u_k(\varepsilon) = u_k(0) + \varepsilon \sum_{l \neq k} \frac{u_l(0)^T V u_k(0)}{\lambda_k(0) - \lambda_l(0)} u_l(0). \tag{26}$$

The perturbed correlation matrix will be written as

$$\langle \Delta R_i(\varepsilon) \Delta R_j(\varepsilon) \rangle = \frac{3\varepsilon}{2\beta} \sum_k \frac{1}{\lambda_k(0) + \varepsilon u_k(0)^T V u_k(0) + \varepsilon^2 \sum_{l \neq k} \frac{u_k(0)^T V u_k(0)}{\lambda_k(0) - \lambda_l(0)}} \cdot \left\{ \left[(u_k(0))_l + \varepsilon \sum_{l \neq k} \frac{u_l(0)^T V u_k(0)}{\lambda_k(0) - \lambda_l(0)} (u_l(0))_i \right] \right.$$

$$\left. \times \left[(u_k(0))_l + \varepsilon \sum_{l \neq k} \frac{u_m(0)^T V u_k(0)}{\lambda_k(0) - \lambda_l(0)} (u_m(0))_i \right] \right\}. \tag{28}$$

Keeping only the first order terms in ε leads to

$$\langle \Delta R_i(\varepsilon) \Delta R_j(\varepsilon) \rangle = \langle \Delta R_i(0) \Delta R_j(0) \rangle + \frac{3\varepsilon}{2\beta} \sum_k \frac{1}{\lambda_k(0)} \left[2 \sum_{l \neq k} \frac{u_l(0)^T V u_k(0)}{\lambda_k(0) - \lambda_l(0)} (u_l(0))_i (u_k(0))_j - \frac{1}{\lambda_k(0)} u_k(0)^T V u_k(0) (u_k(0))_i (u_k(0))_j \right]. \tag{29}$$

$$\langle \Delta R_i(\varepsilon) \Delta R_j(\varepsilon) \rangle = \frac{3}{2\beta} \sum_k \frac{1}{\lambda_k(\varepsilon)} (u_k(\varepsilon))_i (u_k(\varepsilon))_j. \tag{27}$$

Substituting for the perturbed eigenvalues and eigenvectors

Thus, the anharmonicity introduced to a pair potential propagates to the full second order correlation matrix. One can iterate this process, by replacing the unperturbed $\boldsymbol{\Gamma}(0)$, $\boldsymbol{\lambda}(0)$, and $\boldsymbol{u}(0)$ by $\boldsymbol{\Gamma}(\varepsilon)$, $\boldsymbol{\lambda}(\varepsilon)$, and $\boldsymbol{u}(\varepsilon)$ and repeating the above analysis. This introduces the effects of anharmonicities to the $\boldsymbol{\Gamma}$ matrix, including those resulting from multiple pair-energy minima. Of course, effects of anharmonicities can be introduced in this way not only to the second order correlation function, but to all higher order moments of the fluctuation vectors using Eq. (7). The higher order moments obtained in this manner may then be used to characterize the probability distribution function $f(\hat{U}, \hat{V}, \hat{R})$ in terms of moment based Hermite polynomials, which we present following the discussion on the harmonic approximation.

III. THE HARMONIC APPROXIMATION

The harmonic approximation is based on expanding the energy function into Taylor series and keeping the quadratic terms. Then, using the differentiations indicated in Eqs. (15), (17), and (18), the matrix $\boldsymbol{\Gamma}$ is obtained, the inverse of which gives the correlation matrix. This approach, which is called the ANM was introduced by Hinsen[7] and applied to proteins by several authors. The model is based on the expansion of the standard form of the energy expression

$$E = \frac{1}{2} \sum_{i,j} E_{ij}(R_{ij}{}^{\text{eq}})(R_{ij} - R_{ij}{}^{\text{eq}})^2. \tag{30}$$

Since this expression is in standard form, the $\boldsymbol{\Lambda}$ matrix vanishes when $R_{ij} = R_{ij}{}^{\text{eq}}$ and the front factor $\partial^2 E_{ij}/\partial(R_{ij}{}^2)^2$ in the $\boldsymbol{\Lambda}'$ matrix in Eq. (21) becomes $E_{ij}(R_{ij}{}^{\text{eq}})/(R_{ij}{}^{\text{eq}})^2$. It is to be noted, however, that when $R_{ij}{}^0$ is not chosen as $R_{ij}{}^{\text{eq}}$, the $\boldsymbol{\Lambda}$ matrix will be nonzero.

When the term in parenthesis in Eq. (30), $(\Delta R_{ij}{}^2 + 2R_{ij}{}^{\text{eq}} \cdot \Delta R_{ij} + R_{ij}{}^{\text{eq}^2})^{1/2} - R_{ij}{}^{\text{eq}}$ is expanded into Taylor series, and the leading term is taken for infinitesimal fluctuations, the resulting expression is obtained,

$$E = \frac{1}{2} \sum_{i,j} \frac{\partial^2 E_{ij}}{\partial(R_{ij}{}^2)^2} [(\Delta R_i - \Delta R_j) \cdot u_{ij}]^2, \tag{31}$$

where $\Delta R_{ij} = \Delta R_j - \Delta R_i$ and $u_{ij} = (R_i{}^{\text{eq}} - R_j{}^{\text{eq}})/R_{ij}{}^{\text{eq}}$ is the unit vector along $R_i{}^{\text{eq}} - R_j{}^{\text{eq}}$. Equation (31) may now be expressed in matrix form as

$$E = \frac{1}{2} \Delta R^T \boldsymbol{\Gamma} \Delta R, \tag{32}$$

where

$$\Gamma_{ij} = \begin{cases} -\dfrac{\partial^2 E_{ij}}{\partial(R_{ij}{}^2)^2} \cos^2 \alpha_{ij} & i \neq j \\ -\sum_k \Gamma_{ij} & i = j \neq k. \end{cases} \tag{33}$$

Here, α_{ij} is the angle between $\Delta R_i - \Delta R_j$ and $R_i{}^{\text{eq}} - R_j{}^{\text{eq}}$.

The energy expression given by Eq. (32), together with the definition of the $\boldsymbol{\Gamma}$ matrix given by Eq. (33), is the basis of the ENMs which hold for infinitesimally small fluctuations. Formulations of the ENMs based on Eq. (32) are outlined in several papers in the recent book of Cui and Bahar.[10]

In the GNM, the matrix $\boldsymbol{\Gamma}$ is assumed to be of the following form:

$$\boldsymbol{\Gamma}_{ij} = \begin{cases} -\gamma^* & i \neq j \quad \text{and} \quad R_{ij} \leq r_{\text{cutoff}} \\ 0 & i \neq j \quad \text{and} \quad R_{ij} > r_{\text{cutoff}} \\ -\sum_k \gamma^* & i = j \neq k. \end{cases} \tag{34}$$

Here, R_{ij} is the distance between the ith an jth C^α's that are within an interaction distance of r_{cutoff}, γ^* is the force constant representing this interaction. Residues separated by a

distance larger than r_{cutoff} are assumed not to interact. The $\mathbf{\Lambda}$ matrix, or the $\mathbf{\Gamma}$ matrix by Eq. (20) when $\mathbf{\Lambda}' = 0$, is $N \times N$, and is identical for the X, Y, and Z components.

The harmonic approximation may also be obtained, equivalently, by using the Gaussian distribution $W(\Delta \mathbf{R})$ of fluctuations

$$W(\Delta \mathbf{R}) = (2\pi)^{-3N/2} (\det \mathbf{\Gamma}/kT)^{-1/2}$$

$$\times \exp\left[-\tfrac{1}{2}\Delta \mathbf{R}^T(\mathbf{\Gamma}/kT)\Delta \mathbf{R}\right]. \tag{35}$$

Here, $\Delta \mathbf{R}$ is the $3N$ dimensional fluctuation vector, and $\mathbf{\Gamma}$ is the spring constant matrix. Multiplying Eq. (35) with $\Delta \mathbf{R} \Delta \mathbf{R}^T$ and integrating over all possible states of fluctuations leads to

$$\langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle \equiv \int \Delta \mathbf{R} \Delta \mathbf{R}^T W(\Delta \mathbf{R}) d\{\Delta \mathbf{R}\}$$

$$= (2\pi)^{-3N/2} (\det \mathbf{\Gamma})^{-1/2} \int \Delta \mathbf{R} \Delta \mathbf{R}^T$$

$$\times \exp\left[-\frac{1}{2}\Delta \mathbf{R}^T \mathbf{\Gamma} \Delta \mathbf{R}\right] d\{\Delta \mathbf{R}\}, \tag{36}$$

where

$$d\{\Delta \mathbf{R}\} = d\Delta X_1 d\Delta X_2 \cdots d\Delta X_N d\Delta Y_1 d\Delta Y_2 \cdots$$

$$d\Delta Y_N d\Delta Z_1 d\Delta Z_2 \cdots d\Delta Z_N. \tag{37}$$

Carrying out the integration leads to

$$\langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle = kT\mathbf{\Gamma}^{-1}, \tag{38}$$

where the elements of the spring constant matrix $\mathbf{\Gamma}$ are chosen such that Eq. (38) is consistent with Eq. (20). This result, which forms the basis of the GNM has been first shown for Gaussian networks by Kloczkowski et al.[6] In that work, a factor of 3/2 on the right-hand side was present due to the different constants of proportionality adopted in the definition of the elements of the $\mathbf{\Gamma}$ matrix.

In this section, we rederived the equations of the harmonic ENMs starting from the general statistical thermodynamics formalism. A statistical thermodynamics rendition of the ENMs has not been elaborated in previous studies. This general approach has several advantages over previous mechanistic approaches as may be apparent in the two examples below.

A. The contribution of harmonic fluctuations to heat capacity

The heat capacity of a native protein can be obtained from Eq. (7) by letting $\phi = \Delta \Phi_k = \Delta E$ and $\Psi_k = 1/T$. With these substitutions, Eq. (7) takes the form

$$\langle (\Delta E)^2 \rangle = -\frac{\partial U}{\partial \beta} = kT^2 C_v, \tag{39}$$

where $\beta = 1/kT$. The contribution from the fluctuations of residues of a protein of N residues comes from the mean energy[11]

$$U = \left\langle \tfrac{1}{2}\Delta \mathbf{R}^T \mathbf{\Gamma} \Delta \mathbf{R} \right\rangle = \tfrac{3}{2}(N-1)kT. \tag{40}$$

Differentiating U with respect to β leads to the heat capacity

$$C_v = \tfrac{3}{2}(N-1)k. \tag{41}$$

The term $N-1$ rather than N appears in Eq. (41) because one degree of freedom is suppressed against translation in each coordinate direction.[22] The present statistical thermodynamics model from which the fluctuations are derived is that of a solid where the C^α's fluctuate around their equilibrium positions. In this case, one would expect $C_v = 3(N-1)k$ rather than Eq. (41). This result is obtained because only the potential energy associated with fluctuations is considered in the derivation. The vibrational component, which is not included here, contributes another $\tfrac{3}{2}(N-1)k$, resulting in the heat capacity of a monatomic solid. This is the high temperature limit where all modes of motion are excited. Recently, Yuan et al.[23] obtained the harmonic contributions to the heat capacity of native proteins using the GNM and associating each mode λ_i with a frequency $\omega_i = \sqrt{\gamma \lambda_i}$ and using the Einstein relation

$$E = \sum_{i=2}^{N} \frac{h\omega_i}{\exp[h\omega_i/kT] - 1}$$

for the energy. In this sense, the vibrational modes of the protein are assumed not fully excited at finite temperatures.

B. Coupling between energy fluctuations and the fluctuations of residue positions

In the statistical mechanics model presented here, the protein exchanges energy with its surroundings. This exchange is the source of $\langle (\Delta E)^2 \rangle$ given in Eq. (39). It is also the driving potential for the fluctuations of residue positions. In this section, we discuss how the fluctuations in energy are correlated with the fluctuations of residue positions in a concerted way, as a consequence of which the protein performs its function.

Using Eq. (7) for correlating ΔU, ΔR_i, and ΔR_j, we obtain

$$\langle \Delta U \Delta \mathbf{R}_i \Delta \mathbf{R}_j^T \rangle = (kT)^2 \left(\frac{\partial^2 U}{\partial \mathbf{F}_j \, \partial \mathbf{F}_k} \right). \tag{42}$$

Performing the differentiation shown in Eq. (42) and using the relations $(\partial/\partial \mathbf{F}_j)(\partial U/\partial \mathbf{F}_k) = (\partial \mathbf{R}_k/\partial \mathbf{F}_j) = \mathbf{\Gamma}^{-1}$ leads to the expression

$$\langle \Delta U \Delta \mathbf{R}_i \Delta \mathbf{R}_j^T \rangle = (kT)^2 (\mathbf{\Gamma}^{-1})_{ij} = kT \langle \Delta \mathbf{R}_i \Delta \mathbf{R}_j^T \rangle. \tag{43}$$

Thus, fluctuations of energy are distributed to the residues in proportion to the correlations of fluctuations. The diagonal elements $\langle \Delta \mathbf{R}_i \Delta \mathbf{R}_i^T \rangle$ are positive by definition. Therefore, the average $\langle \Delta U (\Delta \mathbf{R}_i)^2 \rangle$ must be positive for each i, if Eq. (43) has to hold. In order for the average $\langle \Delta U (\Delta \mathbf{R}_i)^2 \rangle$ to be positive, a positive value of ΔU must couple with large values of $(\Delta R_i)^2$ and a negative ΔU must couple with small values of $(\Delta R_i)^2$. For the off-diagonal terms, the same pattern holds. If $\langle \Delta \mathbf{R}_i \Delta \mathbf{R}_j^T \rangle > 0$, then positive energy fluctuations pick up the large positive $\Delta \mathbf{R}_i \Delta \mathbf{R}_j^T$'s. Conversely, if

$\langle \Delta \mathbf{R}_i \Delta \mathbf{R}_j{}^T \rangle < 0$, then positive energy fluctuations pick up the large negative $\Delta \mathbf{R}_i \Delta \mathbf{R}_j{}^T$'s. The exchange of energy of a protein with its surroundings is expected to have major role in protein-ligand binding. Recent work[12,24] shows that highest modes of $\langle \Delta U (\Delta \mathbf{R}_i)^2 \rangle$ locate the binding sites of ligands on proteins.

IV. DEVIATIONS FROM THE HARMONIC POTENTIAL

When the energy function is expressed in standard form, the thermodynamic model proposed leads to fluctuations that depend on the curvature of the energy surface, as can be seen from Eq. (19). However, proteins in general exhibit large scale fluctuations, and the dynamics is strongly dependent on the anharmonicity of the energy landscape.[25] The fact that native proteins exhibit large fluctuations about the equilibrium configuration necessitates the introduction of an improvement in the model that accounts for departures from harmonicity. In Secs. I–III we introduced an iterative perturbation scheme to characterize the effects of anharmonicities in the $\mathbf{\Gamma}$ matrix. In this section, we further elaborate on this problem.

We approximate the probability function $f(\hat{U}, \hat{V}, \hat{R})$ in the presence of large scale fluctuations by a tensorial Hermite series[26]

$$f(\Delta \mathbf{R}) = (2\pi)^{-3N/2} (\det \langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle)^{-1/2}$$

$$\times \exp \left[-\frac{1}{2} \Delta \mathbf{R}^T \langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle^{-1} \Delta \mathbf{R} \right]$$

$$\cdot \left[1 + \sum_{\nu=3}^{\infty} (\nu!)^{-1} \langle H_\nu \rangle \cdot H_\nu (\langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle^{-1/2} \Delta \mathbf{R}) \right].$$

$$(44)$$

On the left-hand side, we dropped the arguments $\hat{U}$ and $\hat{V}$, and used $\Delta \mathbf{R}$ instead of $\mathbf{R}$. The leading term of the distribution function is the Gaussian as given by Eq. (35). Nonlinear terms are introduced as corrections in terms of the Hermite polynomials. These correction terms become unimportant as the fluctuations become small, and/or the system approaches a harmonic one. The first few polynomials, H_ν, are as follows:

$$H_1(\Delta \mathbf{R}) = \Delta R_i,$$

$$H_2(\Delta \mathbf{R}) = \Delta R_i \Delta R_j - \delta_{ij},$$

$$H_3(\Delta \mathbf{R}) = \Delta R_i \Delta R_j \Delta R_k - (\Delta R \delta)_{ijk},$$

$$H_4(\Delta \mathbf{R}) = (\Delta R^4 - \Delta R^2 \delta + \delta^2)_{ijkl},$$

$$(45)$$

$$H_5(\Delta \mathbf{R}) = (\Delta R^5 - \Delta R^3 \delta + \Delta R \delta^2)_{ijklm},$$

$$H_6(\Delta \mathbf{R}) = (\Delta R^6 - \Delta R^4 \delta + \Delta R^2 \delta^2 - \delta^3)_{ijklmn},$$

where δ_{ij} is the Kronecker delta, and $(\Delta R \delta)_{ijk}$ in the expression for H_3 is a short-hand notation for $\Delta R_i \delta_{jk} + \Delta R_j \delta_{ik} + \Delta R_k \delta_{ji}$, with similar expressions for the remaining terms in Eq. (45). For example,

$$(\Delta R^2 \delta)_{ijkl} = \Delta R_i \Delta R_j \delta_{kl} + \Delta R_i \Delta R_k \delta_{jl} + \Delta R_i \Delta R_l \delta_{ki}$$

$$+ \Delta R_j \Delta R_k \delta_{il} + \Delta R_j \Delta R_l \delta_{ik} + \Delta R_k \Delta R_l \delta_{ij}.$$

The third term in the series represents the first deviation from the harmonic potential and contains the average $\langle \Delta R_i \Delta R_j \Delta R_k \rangle$. According to the present model, this average is given, by the application of Eq. (7), as[18]

$$\langle \Delta R_i \Delta R_j \Delta R_k \rangle = (kT)^2 \frac{\partial^2 R_i}{\partial F_j \, \partial F_k}$$

$$= (kT)^2 \sum_m \left[\frac{\partial}{\partial R_m} (\mathbf{\Gamma}^{-1})_{ik} \right] (\mathbf{\Gamma}^{-1})_{mj}, \quad (46)$$

where the second line is obtained by using Eq. (20) for the force relation, as

$$\frac{\partial^2 R_i}{\partial F_j \, \partial F_k} = \frac{\partial}{\partial F_j} \left(\frac{\partial R_i}{\partial F_k} \right) = \frac{\partial}{\partial F_j} (\mathbf{\Gamma}^{-1})_{ik}$$

$$= \sum_m \frac{\partial}{\partial R_m} (\mathbf{\Gamma}^{-1})_{ik} \frac{\partial R_m}{\partial F_j}$$

$$= \sum_m \left[\frac{\partial}{\partial R_m} (\mathbf{\Gamma}^{-1})_{ik} \right] (\mathbf{\Gamma}^{-1})_{mj}.$$

The second line of Eq. (46) contains the derivative of $\mathbf{\Gamma}^{-1}$, which can be carried out if the energy function is known.

As an alternative to Eq. (46), higher order moments, of course, can be evaluated from MD trajectories. In this case, long trajectories are needed for the molecule to populate all the accessible states.[25] The example worked out in Sec. V of this paper derives the averages from MD trajectories.

Equation (44) may take a simpler form if it is presented in terms of the transformed fluctuations Δr that is related to $\Delta \mathbf{R}$ by the transformation

$$\Delta r = \langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle^{-1/2} \Delta \mathbf{R}. \quad (47)$$

With this transformation, the correlation matrix $\langle \Delta r \Delta r^T \rangle$ is written as

$$\langle \Delta r \Delta r^T \rangle = \langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle^{-1/2} \langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle \langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle^{-1/2^T} \equiv E,$$

$$(48)$$

where E is the identity matrix, and the last equality follows because the inverse square root of a symmetric matrix is symmetric. We let V represent the eigenvector matrix that diagonalizes $\langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle$, and λ represent the eigenvalues. Then, for Eq. (48) to be the unit matrix, we must have

$$\langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle^{-1/2} = \text{diag } \lambda^{-1/2} V^T. \quad (49)$$

With these equalities, we see that the fluctuations Δr are the fluctuations in the mode space spanned by the eigenvectors, V.[11]

The linear transformation given by Eq. (47) is the Karhunen Loeve or the principal component analysis widely used in the analysis of MD trajectories.[14,27,28] Equation (44) may now be written in mode space as

594
$$f(\Delta \boldsymbol{r}) = (2\pi)^{-3N/2}\exp\left[-\frac{1}{2}\Delta \boldsymbol{r}^2\right]$$

595
$$\cdot \left[1 + \sum_{\nu=3}^{\infty}(\nu!)^{-1}\langle H_\nu\rangle \cdot H_\nu(\Delta \boldsymbol{r})\right], \tag{50}$$

596 where the average Hermite polynomials are defined as

597
$$\langle \boldsymbol{H}_v\rangle = \int_{-\infty}^{\infty}\boldsymbol{H}_v(\Delta \boldsymbol{r})W(\Delta \boldsymbol{r})d\Delta \boldsymbol{r}. \tag{51}$$

598 Equation (50) represents the distribution of coordinates in
599 modal space. The elements of $\langle H_v\rangle$ now contain products of
600 modal coordinates. For example the third order terms are
601 now $\langle \Delta r_i\Delta r_j\Delta r_k\rangle$, and are measures of the extent of mode
602 coupling. Obviously, the second order modes are decoupled
603 since $\langle \Delta r_i\Delta r_j\rangle = \delta_{ij}$.
604 Let us consider the distribution of the first mode $f(\Delta r_1)$,
605 for example. There are three types of terms in $\langle H_v\rangle$'s in Eq.
606 (50): (i) Terms that contain Δr_1 only, (ii) terms that are com-
607 binations of Δr_1 and other modes, and (iii) terms that do not
608 contain Δr_1. Terms of type (i) are pure first mode contribu-
609 tions to $f(\Delta r_1)$. Terms of type (ii) indicate the extent of mode
610 coupling on the distribution of Δr_1, and terms of type (iii)
611 have no contribution to mode 1. The part given by (ii) shows
612 the contributions from the coupling of other modes to $f(\Delta r_1)$.
613 Obviously, same argument is valid for any mode other than
614 mode 1, and may serve as a suitable approach to understand
615 the effects of mode coupling in proteins.

616 ## A. Transformation from modal space to real space

617 Having determined the statistical features of fluctuations
618 in modal space, it is straightforward to study the properties
619 of these correlations and couplings in real space on residue
620 basis with the help of the transformation $\Delta \boldsymbol{R}$
621 $= \boldsymbol{V}$ diag $\lambda^{1/2}\Delta \boldsymbol{r}$. For example, how the fluctuations $\langle \Delta R_i\Delta R_j\rangle$
622 of residues i and j are affected due to the coupling of the
623 modes u and v is given by the expression

624
$$\langle \Delta \boldsymbol{R}\Delta \boldsymbol{R}^T\rangle = \boldsymbol{V} \text{ diag } \lambda \boldsymbol{V}^T, \tag{52}$$

625 where all the elements of diag λ are set equal to zero except
626 λ_u and λ_v. Similarly, for the third order correlations, how the
627 fluctuations $\langle \Delta R_i\Delta R_j\Delta R_k\rangle$ of residues i, j, and k are affected
628 due to the coupling of the modes u, v, and w is given by the
629 expression

$$\langle \Delta \boldsymbol{R}_i\Delta \boldsymbol{R}_j\Delta \boldsymbol{R}_k\rangle = \sum_{p,q,r}(\lambda_p\lambda_q\lambda_r)^{1/2}V_{ip}V_{jq}V_{kr}\langle \Delta r_p\Delta r_q\Delta r_r\rangle,$$

630
$$\tag{53}$$

631 where the set (p,q,r) is the permutation of the values
632 (u,v,w).
633 **Example:** Calculations of harmonic and anharmonic
634 contributions to residue correlations for a hexapeptide
635 We use a randomly chosen hexapeptide of sequence
636 ASN-ASP-MET-PHE-ARG-LEU. This is a toylike protein
637 chosen for illustrative purposes only. Initially, a random con-
638 formation was chosen and the energy of the system was
639 minimized for a sufficiently long time until no large scale

640 conformational changes took place. This conformation is
641 taken as the "native" state of the peptide. The fluctuations of
642 residue positions about this conformation are determined by
643 MD simulations. Simulations were performed in explicit sol-
644 vent (water) using NAMD 2.5 package with CHARMM27
645 force field. All simulations were performed at constant tem-
646 perature (300 K) in a periodic water box with a 20 Å cush-
647 ion. To evaluate the nonbonded interaction, cutoff distance
648 was set to 12 Å. The particle Ewald sum was used as a way
649 of calculating long-range forces in the periodic systems,
650 thereby minimizing the error introduced by truncation due to
651 the cutoff distance. Integration time step was set to 2 fs and
652 structure was recorded at 2000 step (4 ps in MD) for a 22 ns
653 long simulation. Only the final 14 ns part of the trajectory
654 was used for the present calculations.
655 We recorded only the C^α positions for the trajectory. The
656 results of calculations reported here are based on the C^α
657 coordinates, which lead to 18 degrees of freedom. The tra-
658 jectory consists of an 18×1 fluctuation vector $\Delta \boldsymbol{R}$ recorded
659 for each time step. The block representation is used in re-
660 cording each fluctuation vector. The three degrees of free-
661 dom due to rigid body rotation and three due to the transla-
662 tion of the centroid are removed from the trajectories,
663 leading to 12 degrees of freedom only. Removal of the six
664 degrees of freedom was performed using VMD's RMSD **AQ:**
665 Tool plug in. Atom selection for alignment was set to C^α **#5**
666 atoms and all structures were aligned using the first structure
667 of the trajectory as reference.
668 The second order correlation $\langle \Delta \boldsymbol{R}\Delta \boldsymbol{R}^T\rangle$ matrix is deter-
669 mined for the trajectory. The matrix has 12 nonzero eigen-
670 values and eigenvectors. The modal coordinates are obtained
671 by Eq. (47), and the various averages $\langle \Delta r_i\Delta r_j\cdots\Delta r_m\rangle$ are
672 calculated from the trajectory. The distributions $f(\Delta r_i)$ for
673 Δr_i, irrespective of the values of all other Δr_j are also calcu-
674 lated from the trajectory. Results are shown by the filled
675 circles in Fig. 1 for all modes, 1–12. The solid curve in each
676 figure is the Gaussian approximation obtained from Eq. (50)
677 with the Hermite series terms equated to zero. The difference
678 between the solid curve and the calculated points is the con-
679 tribution of anharmonicities to each mode. Only the slowest
680 two modes show significant deviations from harmonicity.
681 The shapes of the modes 3–12 may be approximated rela-
682 tively well by Gaussians, although there are significant de-
683 viations in the maximum anharmonic amplitudes from har-
684 monics, as may readily be verified from Fig. 1.
685 We now search an answer to the important question of
686 what fraction of the deviations from harmonicity in a given
687 mode results from coupling with other modes. We will then
688 search the answer to which other mode couples most
689 strongly to the given mode. The proposed tensorial Hermite
690 series expansion is capable of providing answers to these
691 questions.
692 The coupling of different modes to a given mode i re-
693 sults from the nonzero averages of mixed terms in the Her-
694 mite series expansion, such as $\langle \Delta r_i^p\Delta r_j^q\cdots\Delta r_m^s\rangle$. Anhar-
695 monic contributions to the distribution $f(\Delta r_i)$ purely from
696 mode i will be from the moments $\langle \Delta r_i^p\rangle$ only. In Fig. 2, we
697 present results of Hermite series expansions up to the 17th
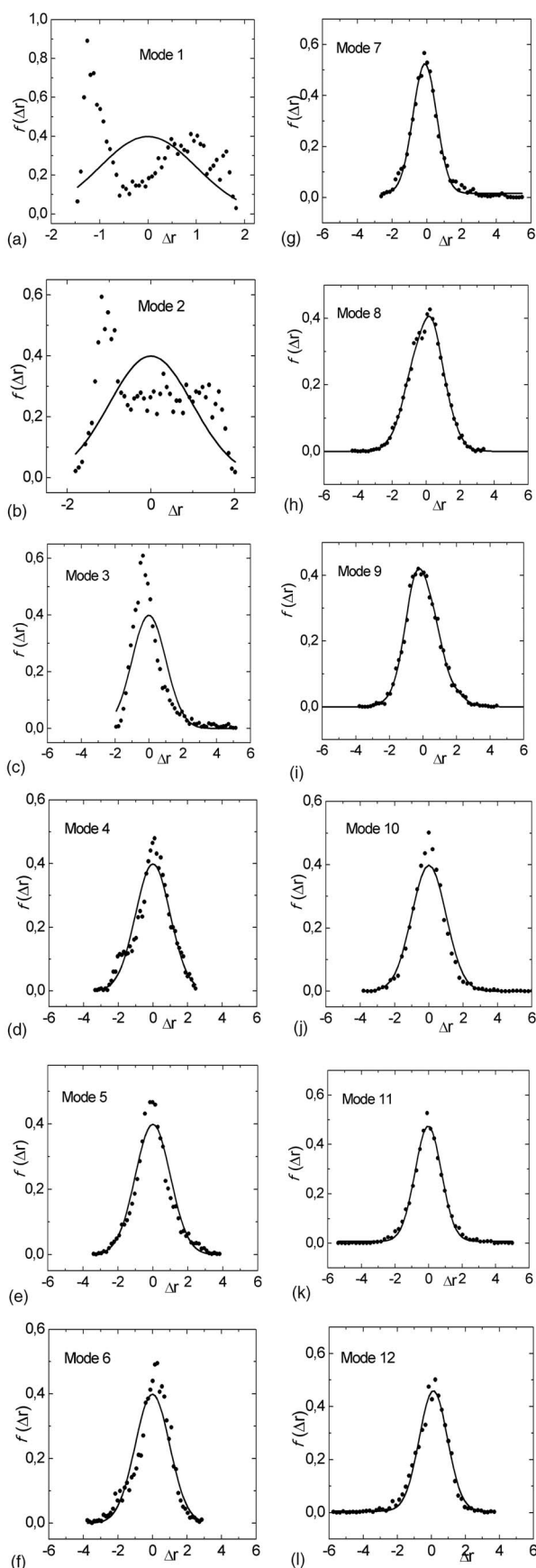**698** order moments, obtained by equating the mixed terms

FIG. 1. The distribution functions $f(\Delta r)$ for each of the 12 modal coordinates, irrespective of the others. The filled circles are calculated as histograms from the MD trajectory. The solid curves are the Gaussians obtained from Eq. (50), with the Hermite terms equated to zero.
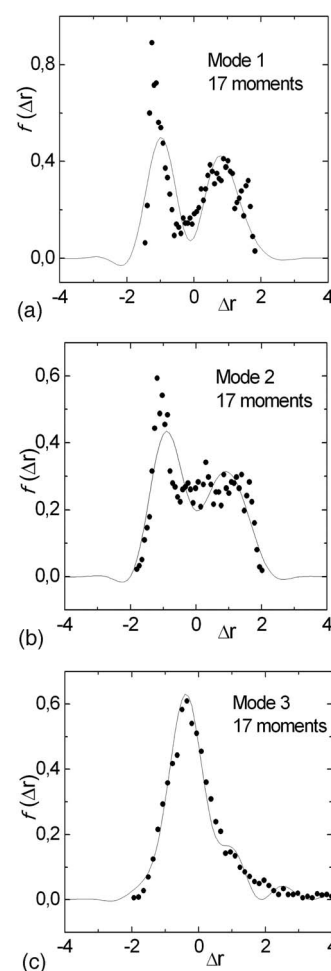

FIG. 2. Comparison of the MD histograms with Hermite series of 17 terms, for modes 1–3.

$\langle \Delta r_i^{\,p} \Delta r_j^{\,q} \cdots \Delta r_m^{\,s} \rangle$ to zero, and keeping moments of type $\langle \Delta r_i^{\,p} \rangle$ for each mode i, only. The reason for going up to the 17th term is that full convergence was observed only at this order. The modal coordinates are nondimensional. Anharmonic contributions in the absence of mode coupling, for modes 1–3 are shown by the solid curves in Fig. 2. The filled circles are obtained from MD histograms. The difference between the points and the curve for a given mode i comes from coupling of other modes to the ith. For mode 1, the shape and the skewness of the distribution is ~~well is~~ obtained by purely anharmonic contributions from mode 1, and contributions from coupling with other modes affect only the peak at $\Delta r = -1.2$. A similar trend is seen also for mode 2. Mode coupling affects only the peak values of the distributions. For mode 3, small deviations at the tails are observed. Modes 4–12 were predicted almost perfectly with pure terms and effects of mode couplings on the distributions are small. The unrealistic negative values of $f(\Delta r_1)$ seen in the first of Fig. 1 result from an artifact of the Hermite series expansion. However, the negative values are insignificant, as seen from the figure. When sufficiently large number of terms are used in the expansion, the negative values become insignificantly small.

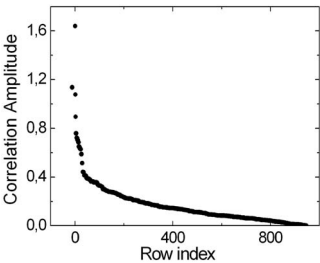We now try to answer the question of which mode

FIG. 3. The values of the mixed third order terms, sorted in descending order.

723 couples to which mode in the most significant way. This
724 question has to be answered for each order of the moment
725 separately. After analyzing moments of all orders, one can
726 conclude on the strongest coupling in the system. Here, for
727 illustrative purposes, we consider only third order moments,
728 $\langle \Delta r_i \Delta r_j \Delta r_k \rangle$, where i, j, and k take values from 1 to 12, and
729 all three are not equal to each other because those terms do
730 not represent mode coupling. In Fig. 3, we present the values
731 of $\langle \Delta r_i \Delta r_j \Delta r_k \rangle$, which we term "correlation amplitudes," of
732 the third moments. We present only the positive values. They
733 are sorted in descending order.
734   It is worth noting that there is one value that is much
735 larger than all the others. Also, most of the correlation am-
736 plitude values are below 0.5. In order to give an idea on
737 which point in Fig. 3 corresponds to which triplet of modes,
738 the three modes and the corresponding correlation ampli-
739 tudes are presented in Table I for values larger than 0.5. The
740 values in bold point in the table are for pure modes, the
741 others are for mixed modes. Pure mode values are also in-
742 cluded in the table for comparison of their magnitudes with
743 those of the mixed modes.
744   In Fig. 4, we present the degree of coupling of the first
745 mode with the two other modes in third order correlations,
746 i.e., in $\langle \Delta r_1 \Delta r_j \Delta r_k \rangle$, where j and k are the mode indices
747 shown along the abscissa and the ordinate in Fig. 4. Darker
748 regions in the figure indicate stronger correlations. In gen-
749 eral, positive correlations are stronger when compared with

TABLE I. Largest correlation amplitudes for third order moments.

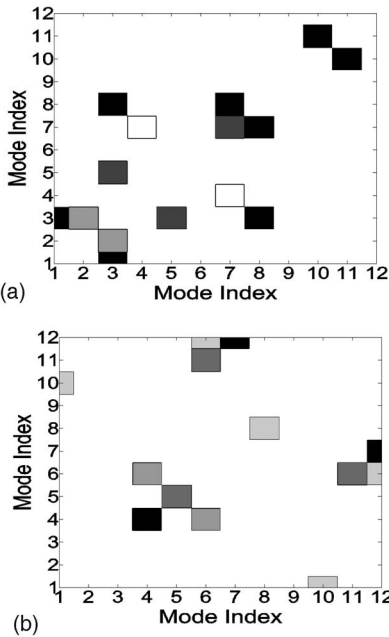| Mode i	Mode j	Mode k	Mode amplitude
3	3	3	1.640
7	7	7	1.077
10	10	10	0.894
12	12	12	−0.724
2	3	3	−0.763
3	3	10	0.759
3	11	11	0.723
7	7	9	0.706
3	10	10	0.685
3	3	9	0.650
5	7	7	0.638
3	12	12	0.629
7	7	12	−0.609
9	11	11	0.588
2	7	7	0.518
7	11	11	0.514

FIG. 4. Coupling of the first mode to the other two modes in third order correlations. The two modes are indicated by the corresponding mode indices along the abscissa and ordinates. Darker regions indicate stronger correlations.

the negative ones. The coupling of the first mode is not con- **750**
fined only to its neighboring modes. For example the cou- **751**
pling $\langle \Delta r_1 \Delta r_{10} \Delta r_{11} \rangle$ between the first mode and the 10th and **752**
11th is one of the strongest positive couplings. Similarly, the **753**
negative coupling $\langle \Delta r_1 \Delta r_7 \Delta r_{12} \rangle$ is also among the strongest. **754**

V. DISCUSSION   **755**

     The statistical thermodynamics treatment of native pro- **756**
teins presented here points out to the fact that nonbonded **757**
pair potentials between residues play the key role in deter- **758**
mining the fluctuations, or in general, the full thermostatis- **759**
tics. The Λ' matrix of Eq. (20) contains the second deriva- **760**
tives of the pair potentials. Higher order derivatives of pair **761**
potentials are present in the distribution function given by **762**
Eq. (44). The forms of the short and long-range inter-residue **763**
potentials evaluated by Bahar et al.,[20,29] over a databank of **764**
302 globular proteins show that inter-residue potentials for **765**
some pairs may be expressed as Lennard-Jones type, or more **766**
generally as a Mie potential.[30] For such cases, neglecting the **767**
asymmetry at the potential minima, the harmonic represen- **768**
tation suffices for characterizing the system for small fluc- **769**
tuations. However, the first and important deviation from a **770**
symmetric harmonic potential comes from to the asymmetry **771**
of the potentials. The pair potential rises steeply when the **772**
distance between two residues becomes smaller than the po- **773**
tential minimum, but rises mildly when the distance becomes **774**
larger. The evaluated coarse-grained inter-residue potentials **775**
for some other residue pairs depart strongly from the Mie- **776**
type potential in that they either do not have minima, or have **777**
a multitude of minima with increasing inter-residue separa- **778**
tion. These differences establish the specificity of pair poten- **779**

tials and have to be taken into consideration, especially for larger fluctuations. Piazza and Sanejouand[16] used an energy function of the form

$$E_{ij} = \sum_{p=1}^{4} \frac{k_p}{p}(R_{ij} - R_{ij}{}^{\text{eq}})^p$$

to investigate the effects of nonlinearities in the potential. The tensorial Hermite polynomials formalism presented here is an alternative rational scheme of introducing the deviations of pair potentials from the harmonic.

In the harmonic approximation, the vibrational modes are independent and the energy of each mode is equal, obeying the equipartition theorem.[11] Independence of modes is destroyed when pair potentials deviate from the harmonic. There is significant interest in the coupling of fluctuational modes in native proteins.[15–17] The interest is mainly an outcome of the belief that biological function is coupled with anharmonic dynamics.[15,16,31] There are two types of anharmonicity associated with a given mode, i: (i) the part of the distribution that results from the higher moments of mode i itself, and (ii) the part of the distribution that results from coupling of mode i with other modes. The example worked out in this paper shows that the shape of the distribution of mode i is well represented by the higher moments of that same mode while the coupling of other modes affects the amplitude of the distribution, specifically the peak values.

Moritsugu et al.,[17] performed a normal mode analysis of myoglobin assuming that energy transfer is due to a weak anharmonicity that can be decomposed into a vibrational energy flow between a pair of normal modes. Nonlinearity was introduced in terms of third order mode coupling that corresponds to the third order moments of the Hermite series. They showed that the vibrational energy was transferred from a normal mode to a very few number of specific normal modes for myoglobin. The magnitude of the coupling coefficient, which corresponds to the third order moments of fluctuations, was estimated by the degree of the geometrical overlap between the coupled modes. The present Hermite series approximation shows that coupling from higher order than the third may play significant role in protein behavior.

Larger fluctuations in native proteins are of significant interest in studying the hopping of residues from one state to the other. In this case, the anharmonicity of the energy landscape needs be taken into consideration. There is growing interest in this direction, and semianalytical models have been used in addition to MD simulations of the anharmonicity.[28,32] The proposed moment based expansion of the fluctuation probability function is capable of characterizing such anharmonicity effects, especially if sufficient number of higher moments is included in the expansion.

The interest in expressing the fluctuations of proteins in normal mode or principal components is not new.[33] In order to describe the internal motions of human lysozyme obtained by MD or Monte Carlo (MC) simulations as motions of normal mode variables, Horiuchi and Go[34] projected the MC and MD trajectories of the protein on its normal mode axes. The idea behind this study was that the harmonic motion predicted by the normal mode analysis could approximately simulate the motion which is in reality highly unharmonic. They showed that the lowest frequency normal mode extracted from the MC and MD simulations correlate very well with the hinge bending motion motions. Amadei et al.,[35] showed that it is possible to separate the configurational space into an essential subspace of few degrees of freedom in which anharmonic motions occur and a harmonic space in which the motion has a narrow Gaussian distribution. The relevance of fluctuation dynamics to energy landscape has been discussed by Hayward and co-workers.[14,32] The present study generalizes these arguments by introducing a tensorial moment based Hermite series form for the well known Karhunen Loeve expansion. Our effort in the present work, unlike those of the previous studies mentioned above, is motivated mainly in identifying the effects of mode coupling which underlies the function of proteins.

APPENDIX A:

The correlation matrix $\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \rangle$ is defined as

$$\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \rangle = \sum (\hat{\boldsymbol{R}} - \boldsymbol{R})(\hat{\boldsymbol{R}} - \boldsymbol{R})^T f. \tag{A1}$$

The gradient of $f(\hat{U}, \hat{V}, \hat{\boldsymbol{R}}, \hat{N})$ with respect to $\boldsymbol{F}/T$ reads

$$\frac{\partial f}{\partial(\boldsymbol{F}/T)} = k^{-1}\left(\hat{R} - \frac{\partial}{\partial(\boldsymbol{F}/T)}S\left[\frac{1}{T}, \frac{P}{T}, \frac{\boldsymbol{F}}{T}, \frac{\mu}{T}\right]\right)f$$

$$= k^{-1}(\hat{\boldsymbol{R}} - \boldsymbol{R})f. \tag{A2}$$

Substituting Eq. (A2) into Eq. (A1), we have

$$\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \rangle = kT \sum (\hat{\boldsymbol{R}} - \boldsymbol{R})\frac{\partial f}{\partial \boldsymbol{F}}$$

$$= kT\frac{\partial}{\partial \boldsymbol{F}}\langle \hat{\boldsymbol{R}} - \boldsymbol{R} \rangle - kT\left\langle \frac{\partial}{\partial \boldsymbol{F}}\langle \hat{\boldsymbol{R}} - \boldsymbol{R} \rangle \right\rangle. \tag{A3}$$

The first term on the right-hand side vanishes, and $\hat{\boldsymbol{R}}$ is statistically independent of $\boldsymbol{F}$, and we have

$$\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \rangle = kT\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{F}}. \tag{A4}$$

Equation (A4) is valid irrespective of system size and is therefore suitable for the study of a single protein.

APPENDIX B:

There are two different representations of the matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}'$ with respect to ordering of the X, Y, and Z coordinates of the N residues. The use of one instead of the other causes confusion. In its full generality, the left-hand side of Eq. (10) consists of the various products of ΔX_i, ΔY_i, ΔZ_i and ΔX_j, ΔY_j, ΔZ_j, expressed with respect to a laboratory fixed coordinate system $OXYZ$. In the block representation, the elements of $\Delta \boldsymbol{R}$ are arranged as $\Delta \boldsymbol{R} = \text{col}[\Delta X_1, \Delta X_2, \ldots, \Delta X_N, \Delta Y_1, \Delta Y_2, \ldots, \Delta Y_N, \Delta Z_1, \Delta Z_2, \ldots, \Delta Z_N]$. In other ENMs the standard MD representation is used according to which, $\Delta \boldsymbol{R}^t = \text{col}[\Delta X_1, \Delta Y_1, \Delta Z_1, \Delta X_2, \Delta Y_2, \Delta Z_2, \ldots, \Delta X_N, \Delta Y_N, \Delta Z_N]$. The correlation matrix $\boldsymbol{C}$ is accordingly written either as $\boldsymbol{C} = \langle \Delta R \Delta R^T \rangle$ or $\boldsymbol{C}^t = \langle \Delta R' \Delta R'^T \rangle$. Both $\boldsymbol{C}$ and $\boldsymbol{C}^t$ are of order $3N \times 3N$, where N

is the number of residues. The passage from one to the other is made by $C = TC'T^T$, where T is a $3N \times 3N$ permutation matrix formed as

$$T_{ij} = \begin{cases} 1, & \text{for } i = 1, 2, \ldots, 3N \text{ and} \\ & j = 3((i-1)\bmod N) + \left\lfloor \dfrac{i-1}{N} \right\rfloor + 1 \\ 0, & \text{otherwise.} \end{cases} \quad \text{(B1)}$$

In the block representation the matrices Λ and Λ' are partitioned into submatrices as

$$\Lambda = \begin{bmatrix} \Lambda(XX) & 0 & 0 \\ 0 & \Lambda(YY) & 0 \\ 0 & 0 & \Lambda(ZZ) \end{bmatrix}$$

$$\Lambda' = \begin{bmatrix} \Lambda'(XX) & \Lambda'(XY) & \Lambda'(XZ) \\ - & \Lambda'(YY) & \Lambda'(YZ) \\ - & - & \Lambda'(ZZ) \end{bmatrix}, \quad \text{(B2)}$$

where each submatrix is $N \times N$. The second submatrix $\Lambda(XY)$, for example, has the mixed products $\Delta X_i \Delta Y_j$.

APPENDIX C:

Here we give the explicit forms of the Hermite polynomials up to the 17th order used in the calculations in the present work. These are the terms for obtaining the contributions to the distribution function by a single mode. For the general case of mixed modes, the definition of the tensorial Hermite polynomials given by Eq. (44) should be used. In the expressions given here, q has two meanings, either Δr or $\langle \Delta r \rangle$ depending whether it is in H or $\langle H \rangle$, respectively,

$$H_1 = 0,$$

$$H_2 = 0,$$

$$H_3 = q^3 - 3q,$$

$$H_4 = q^4 - 3q^2 + 3,$$

$$H_5 = q^5 - 10q^3 + 15q,$$

$$H_6 = q^6 - 15q^4 + 45q^2 - 15,$$

$$H_7 = q^7 - 21q^5 + 105q^3 - 105q,$$

$$H_8 = q^8 - 28q^6 + 210q^4 - 420q^2 + 105,$$

$$H_9 = q^9 - 36q^7 + 378q^5 - 1260q^3 + 945q,$$

$$H_{10} = q^{10} - 45q^8 + 630q^6 - 3150q^4 + 4725q^2 - 945,$$

$$H_{11} = q^{11} - 55q^9 + 990q^7 - 6930q^5 + 17\,325q^3 - 9450q,$$

$$H_{12} = q^{12} - 66q^{10} + 1485q^8 - 13\,860q^6 + 51\,975q^4$$
$$- 62\,370q^2 + 10\,395,$$

$$H_{13} = q^{13} - 78q^{11} + 2145q^9 - 25\,740q^7 + 135\,135q^5$$
$$- 270\,270q^3 + 10\,395q,$$

$$H_{14} = q^{14} - 91q^{12} + 3003q^{10} - 45\,045q^8 + 315\,315q^6$$
$$- 945\,945q^4 + 945\,945q^2 - 135\,135,$$

$$H_{15} = q^{15} - 105q^{13} + 4095q^{11} - 75\,075q^9 + 675\,675q^7$$
$$- 2\,837\,835q^5 + 4\,729\,725q^3 - 2\,027\,025q,$$

$$H_{16} = q^{16} - 120q^{14} + 5460q^{12} - 120\,120q^{10}$$
$$+ 1\,351\,350q^8 - 7\,567\,560q^6 + 18\,918\,900q^4$$
$$- 16\,216\,200q^2 + 2\,027\,025,$$

$$H_{17} = q^{17} - 136q^{15} + 7140q^{13} - 185\,640q^{11}$$
$$+ 2\,552\,550q^9 - 18\,378\,360q^7 + 64\,324\,260q^5$$
$$- 91\,891\,800q^3 + 34\,459\,425q. \quad \text{(C1)}$$

[1] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000).
[2] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[3] I. Bahar, A. R. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997).
[4] T. Haliloglu, I. Bahar, and B. Erman, Phys. Rev. Lett. **79**, 3090 (1997).
[5] P. J. Flory, Proc. R. Soc. London, Ser. **351**, 351 (1976).
[6] A. Kloczkowski, J. E. Mark, and B. Erman, Macromolecules **22**, 1423 (1989).
[7] K. Hinsen, Proteins **33**, 417 (1998).
[8] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, Biophys. J. **80**, 505 (2001).
[9] M. Delarue and Y. H. Sanejouand, J. Mol. Biol. **320**, 1011 (2002); O. Keskin, I. Bahar, D. Flatow, D. G. Covell, and R. L. Jernigan, Biochemistry **41**, 491 (2002); W. Zheng, B. R. Brooks, and G. Hummer, Proteins (2007); L. Yang, G. Song, and R. L. Jernigan, Biophys. J. **93**, 920 (2007); M. C. Demirel, A. R. Atilgan, R. L. Jernigan, B. Erman, and I. Bahar, Protein Sci. **7**, 2522 (1998); C. Micheletti, P. Carloni, and A. Maritan, Proteins **55**, 635 (2004); Y. , M. Liang, Y. Yang, and Q. Cui, PLOS Comput. Biol. **3**, 2); . , J. Jeong, and M. K. Kim, J. Mol. Graphics Modell. **24**, 296 (200 ; L. Leherte and D. P. Vercauteren, Comput. Phys. Commun. (unpubl ; G. Song and R. L. Jernigan, J. Mol. Biol. **369**, 880 (2007); R. Lavery and S. Sophie Sacquin-Mora, J. Biosci. **32**, 891 (2007); L. Marsella, Proteins: Struct., Funct., Bioinf. **62**, 173 (2006); S. Kundu, S. C. Sorensen, and G. N. Phillips, Proteins: Struct. Funct. Bioinf. **57**, 725 (2004); K. Eom, S. C. Baek, J. Ahn, and S. Na, J. Comput. Chem. **28**, 1400 (2007); P. Doruker, R. L. Jernigan, and I. Bahar, *ibid.* **23**, 119 (2002); T. Haliloglu and I. Bahar, Proteins **31**, 271 (1998); I. Bahar, B. Erman, R. L. Jernigan, A. R. Atilgan, and D. G. Covell, J. Mol. Biol. **285**, 1023 (1999); P. Doruker, A. R. Atilgan, and I. Bahar, Proteins **40**, 512 (2000); E. Eyal, C. Chennubhotla, L.-W. Yang, and I. Bahar, Bioinformatics **23**, 175 (2007); L.-W. Yang, E. Eyal, C. Chennubhotla, J. JunGoo, A. M. Gronenborn, and I. Bahar, Structure **15**, 741 (2007); E. Eyal, L. Yang, and I. Bahar, Bioinformatics **22**, 2619 (2006); L. W. Yang, A. J. Rader, X. Liu, C. J. Jursa, S. C. Chen, and H. A. Karimi, Nucleic Acids Res. **34**, 24 (2006); C. Chennubhotla and I. Bahar, Lect. Notes Comput. Sci. **3909**, 379 (2006); L. W. Yang and I. Bahar, Structure (London) **13**, 893 (■); L. W. Yang, X. Liu, X. Christopher, J. Jursa, M. Holliman, A. J. Rader, H. Karimi, and I. Bahar, Bioinformatics **21**, 2978 (2005); Y. Wang, A. J. Rader, I. Bahar, and R. L. Jernigan, J. Struct. Biol. **147**, 302 (2004); A. J. Rader and I. Bahar, Polymer **45**, 659 (2004); C. Xu, D. Tobi, and I. Bahar, J. Mol. Biol. **333**, 153 (2003); B. Erman and K. A. Dill, J. Chem. Phys. **112**, 1050 (2000); A. Erkip, B. Erman, C. Seok, and K. A. Dill, Polymer **43**, 495 (20 B. Erman, Biophys. J. **91**, 3589 (2006); J. L. Liao and D. N. Beratan, *ibid.* **87**, 1369 (2004); F. Tama and C. L. Brooks, J. Mol. Biol. **318**, 733 (2002); I. Bahar and A. J. Rader, Curr. Opin. Struct. Biol. **15**, 5).

[10] Q. Cui and I. Bahar, *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems* (Chapman and Hall, London, 2006).

[11] I. Bahar, A. R. Atilgan, M. C. Demirel, and B. Erman, Phys. Rev. Lett. **80**, 2733 (1998).

[12] T. Haliloglu, E. Seyrek, and B. Erman, Phys. Rev. Lett. **100**, 228102 (2008).

[13] R. Levy, D. Perahia, and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **79**, 1346 (1982).

[14] S. Hayward, A. Kitao, and N. Go, Proteins **23**, 177 (1995).

[15] B. Juanico, Y. H. Sanejouand, F. Piazza, and D. P. Los Rios, Phys. Rev. Lett. **99**, 238104 (2007).

[16] F. Piazza and Y. H. Sanejouand, Phys. Biol. **5**, 026001 (2008).

[17] K. Moritsugu, O. Miyashita, and A. Kidera, Phys. Rev. Lett. **85**, 3970 (2000); K. Moritsugu, O. Miyashita, and A. Kidera, J. Phys. Chem. B **107**, 3309 (2003).

[18] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd ed. (Wiley, New York, 1985).

[19] T. L. Hill, *Thermodynamics of Small Systems* (Dover, New York, 1994).

[20] I. Bahar and R. L. Jernigan, J. Mol. Biol. **266**, 195 (1997).

[21] B. Erman, I. Bahar, and C. Chennubhotla (unpublished).

[22] The appearance of $3(N-1)$ in Eq. (33) is because the Hamiltonian for the GNM is only rotationally invariant and three degrees of freedom due to rigid body translation are suppressed. For a Hamiltonian that has both translational and rotational invariance, six degrees of freedom will be suppressed leading to a factor of $3(N-2)$.

[23] Y. Yuan, Y. Wu, and J. Zi, J. Phys.: Condens. Matter **17**, 469 (2005).

[24] T. Haliloglu and B. Erman, Phys. Rev. Lett. (unpublished).

[25] F. Pontiggia, G. Colombo, C. Micheletti, and H. Orland, Phys. Rev. Lett. **98**, 048102 (2007).

[26] P. J. Flory and D. Y. Yoon, J. Chem. Phys. **61**, 5358 (1974).

[27] M. A. Balsera, W. Wriggers, Y. Oono, and K. Schulten, J. Phys. Chem. **100**, 2567 (1996.); A. A. Palazoglu, A. Gursoy, Y. Arkun, and B. Erman, J. Comput. Biol. **11**, 1149 (2004).

[28] A. Kitao, in *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, edited by Q. Cui and I. Bahar (Chapman and Hall, London, 2006), p. 233.

[29] I. Bahar, M. Kaplan, and R. L. Jernigan, Proteins **29**, 292 (1997).

[30] G. Mie, Ann. Phys. **1** (1903).

[31] H. Frauenfelder, F. Parak, and R. D. Young, Annu. Rev. Biophys. Biophys. Chem. **17**, 451 (1988); *Nonlinear Excitations in Biomolecules*, edited by M. Peyrard (Springer, Berlin, 1995).

[32] A. Kitao, S. Hayward, and N. Go, Proteins **33**, 496 (1998).

[33] M. Karplus, in *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, edited by Q. Cui and I. Bahar (Chapman and Hall, London, 2006); B. Brooks and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **80**, 6571 (1983).

[34] T. Horiuchi and N. Go, Proteins **10**, 106 (1991).

[35] A. Amadei A. Linssen, and H. J. C. Berendsen, Proteins **17**, 412 (1993).

AUTHOR QUERIES — 054907JCP

#1   AU: please check specific citation of sections
#2   AU: please check specific citation of sections
#3   AU: please check specific citation of sections
#4   AU: please check specific citation of section
#5   AU: please spell out "VMD" and "RMSD" if possible
#6   AU: please supply brief description for appendix heading
#7   AU: equations in the appendix were renumbered. Please check.
#8   AU: please supply brief description for appendix heading
#9   AU: please supply brief description for appendix heading
#10   AU: please check citation of Eq. 44 in place of "Eq. …"
#11   AU: please check entries in Ref. 9
#12   AU: please supply volume and page numbers in Ref. 9 third entry
#13   CrossRef reports the first page should be ″e214″ not ″214″ in the reference ″Haibo, Liang, Yang, Cui, 2007″.
#14   CrossRef reports the author should be ″Huang″ not ″Haibo″ in the reference ″Haibo, Liang, Yang, Cui, 2007″.
#15   CrossRef reports the author should be ″JEONG″ not ″Jay″ in the reference ″Jay, Jeong, Kim, 2006″.
#16   AU: please check authors in Ref. 9 eighth entry
#17   AU: please update Ref. 9 ninth entry b
#18   CrossRef reports the first page should be ″i175″ not ″175″ in the reference ″Eyal, Chennubhotla, Yang, Bahar, 2007″.
#19   CrossRef reports the first page should be ″W24″ not ″24″ in the reference ″Yang, Rader, Liu, Jursa, Chen, Karimi, 2006″.
#20   AU: please check author AU: please supply full author list in Ref. in Ref. 9 22nd entry
#21   AU: please supply year in Ref. 9 24th entry
#22   CrossRef reports the year should be ″2002″ not ″2001″ in the reference ″Erkip, Erman, Seok, Dill, 2001″.
#23   CrossRef reports the first page should be ″586″ not ″1″ in the reference ″Bahar, Rader, 2005″.
#24   AU: please confirm change in page number in Ref. 16
#25   AU: please update Ref. 21
#26   CrossRef reports the author should be ″Palazoglu″ not ″Balsera″ in the reference 27 ″Balsera, Wriggers, Oono, Schulten, 2004″.
#27   CrossRef reports the volume should be ″316″ not ″11″ in the reference 30 ″Mie, 1903″.
#28   AU: please confirm change in volume number in Ref. 32