Statistical Thermodynamics of Residue Fluctuations in Native Proteins

by

Osman N. Yogurtcu

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science
in
Computational Science and Engineering

Koc University
August 2008

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Osman N. Yogurtcu

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Burak Erman, Prof. (Advisor)

_____

Attila Gürsoy, Assoc. Prof.

_____

Özlem Keskin, Assoc. Prof.

Date: _____

ABSTRACT

We have formulated the statistical thermodynamics of residue fluctuations of native proteins at constant temperature and pressure. The underlying assumptions of the two elastic network models, the Gaussian Network Model (GNM) and the Anisotropic Network Model (ANM) are studied and their limits of validity are discussed. The statistical mechanical model adopted allows generalization of the elastic network models. We have validated our results by using trajectories obtained from extensive molecular dynamics simulations. Analysis of the trajectories shows that the principal axes of the fluctuation correlation matrix coincide with the principal axes of the radius of gyration tensor. This establishes the connection between residue fluctuations and protein geometry.

ÖZETÇE

Tezimizde, proteinlerdeki aminoasitlerin hareketlerinin izobarik ve izotermal ortamdaki istatistiksel termodinamiği formüle edildi. Aminoasit hareketliliğini inceleyen iki elastik ağ modeli (ENM), Gaussian Ağ Modeli (GNM) ve Anisotropik Ağ Modeli'nin (ANM) temelleri bu formülasyon üzerinden incelendi, gerçeğe yakınlıkları tartışıldı. Buradan hareketle genelleştirilmiş bir elastik ağ modelinin nasıl olması gerektiği istatistiksel mekanik diliyle sunuldu. Sonuçlarımız moleküler dinamik simülasyonları ile doğrulandı. Buna ek olarak, proteinlerin, genel olarak, ana salınım ekseni ile ana yapısal ekseninin çakışık olduğu gözlemlendi ve böylelikle tümel aminoasit hareketleri ile protein geometrisi arasında bir bağlantı olduğu gösterilmiş oldu.

PREFACE

During the course of my Master's thesis studies, I have been many times posed the question "What do you research?" This happened so many times that I had to develop an automated response to this question which must be clear but striking as well; otherwise, my friendly chats would never progress any further. So, often times, my educated answer went like this: "I study the functionality of protein fluctuations." This answer saved a lot of my conversations. Some even said, "gees, do the proteins fluctuate?" and I replied "oh, sure they do."

I lived happily sometime until Prof. Alper Erdogan, being an excellent scientist, asked me "What makes the proteins fluctuate? What forces are there?" I was stunned and could not answer. It was the time when I realized that I was lacking the most important trait on the road to becoming a good scientist and that was perseverently asking the right questions (and producing the best answers). I think, experience is the key to success in science –it is maybe even more important than in other professions- and I hope I can learn my lessons as fast/painless as I could.

Surely, this thesis will not elucidate the hidden secrets of life and livings. However, please do consider the amount of work done as a first step in my scientific career (hopefully).

A journey of a thousand miles
begins with a single step.

Chinese Proverb

ACKNOWLEDGEMENTS

Of course, my utmost gratitude goes to my thesis advisor, **Prof. Burak Erman** whose ingenuity, creativity and patience added significantly to my graduate experience. I must admit that, at times, I really wonder how he manages to *be younger* and more diligent than his advisees.

I would like to thank the other members of my committee profusively: **Assoc. Prof. Attila Gürsoy** and **Assoc. Prof. Ozlem Keskin** for critical reading of this thesis and for their valuable comments. I thank **Bora Erdemli** for his help in obtaining the molecular dynamics trajectories of some of the proteins used in this study, and **Prof. Ivet Bahar** for helpful comments. I would like to thank also the Scientific and Technological Research Council of Turkey (**TÜBİTAK**) for their financial support during my MS study.

I am also indebted to professors **Engin Erzin, Mehmet Sayar, Zeynep Direk, Alper Demir, Serdar Kozat, Metin Turkay, Oguz Sunay, Murat Tekalp, Halil Kavakli, Deniz Yuret, Alper Erdogan, Cagatay Basdogan** and **Ipek Basdogan** for their unconditional help, sincerity and for being great role models for me over the years.

I am thankful to all my friends at Koc University, especially to my extended circle of officemates and homemates: **Cengiz Ulubas, Sefer Baday, Fatih Toy, Ozkan Egri, Ramazan Sancak, Bahar Ondul, Ozge Engin, Besray Unal, Nurcan Tuncbag, Ekin Tuzun, Gozde Kar, Abdullah Turan, Huseyin Seren, Baybora Baran, Mert Gur, Aslıhan Aslan, Yasemin Demir, Serhan Isikman, Elif Isikman, Semih Afyon, Emre Guney, Bora Erdemli, Ahmet Bakan, Murat Tugrul, Halil Bisgin** and **Gulay Ergul** with whom I really enjoyed the time I spent and from whom I have learnt a lot.

Above all, I would like to thank my family, **Bedil** and **Mahmut Yogurtcu**, who were always with me. *They* provided the encouragement, support and care when most needed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

GNM	Gaussian Network Model
ANM	Anisotropic Network Model
MD	Molecular Dynamics
fs	femtoseconds
S	Entropy
U	Internal Energy
V	Volume
R	Position Vectors of alpha Carbon atoms
F	Force Vector acting on alpha Carbon atoms
P	Pressure
T	Temperature
k	Boltzmann Constant

Chapter 1

INTRODUCTION

Proteins are chemical compounds which bear crucial importance to the maintenance of cell homeostasis and they constitute at least 50% of the dry mass of the cells. They actively participate in numerous cellular processes including catalysis of biochemical reactions, cell division and motility. Proteins are composed of amino acids. There are 20 standard amino acids. The amino acids polymerize by means of peptide bonds and form the proteins (**Figure 1.1**). After peptidization, the amino acids in the protein molecule are called residues.

A protein molecule can be studied in four structural levels. The sequence of residues in a protein gives the primary structure. The spatial order of proteins without considering the radical (R) groups in amino acids is the secondary structure. The α-helices, β strands and turns are of secondary structure. The tertiary structure includes the geometric coordinates of all the atoms of a protein. If multi-tertiary structures merge and form a super complex, then this complex is called the quaternary structure. Hydrogen bonds, van der Waals, ionic and hydrophobic interactions and disulphide bonds play an important role in making the tertiary and quaternary structure stable; while; only the hydrogen bonds and hydrophobic interactions stabilize the secondary structure (**Figure 1.2**).

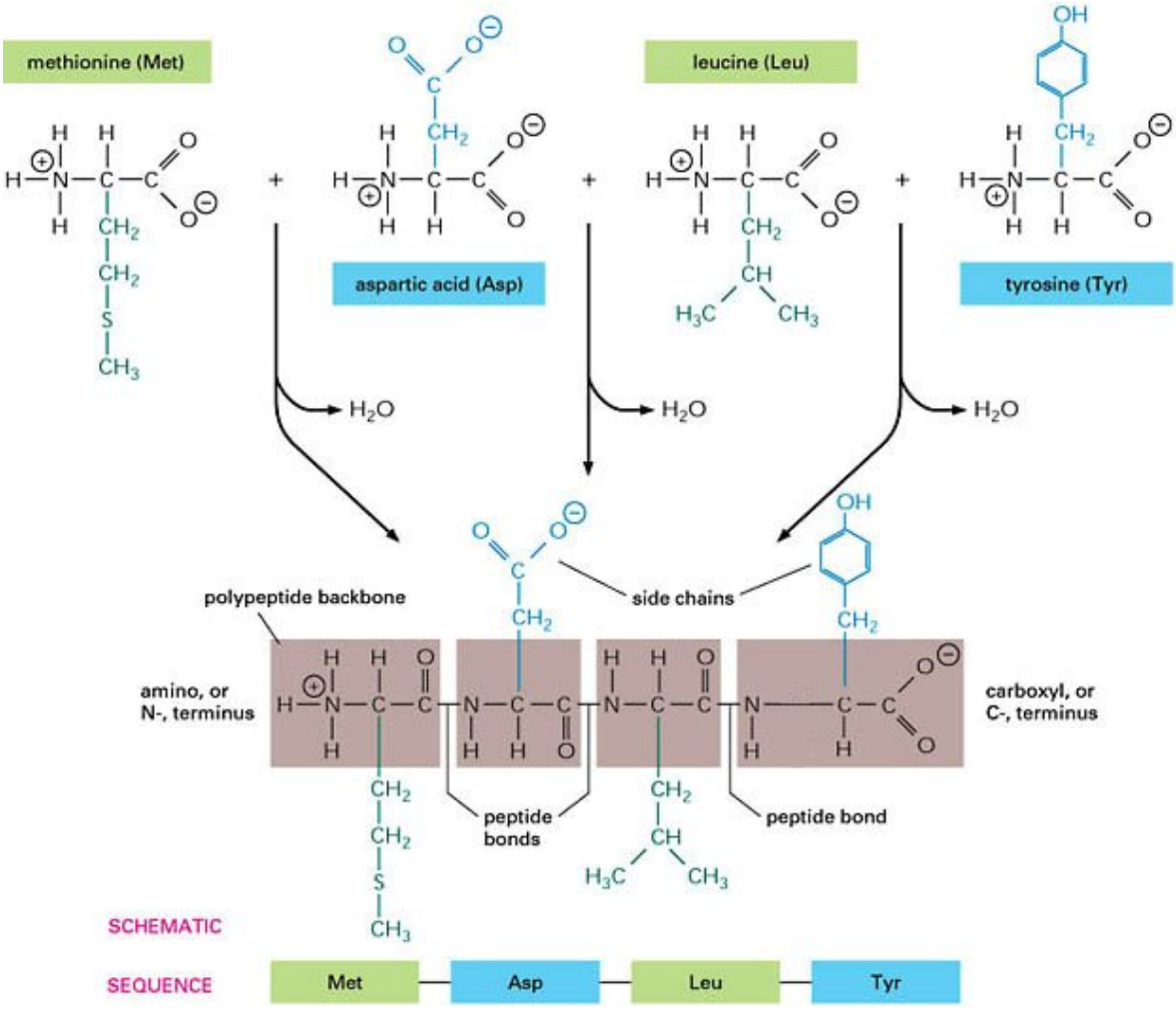


Figure 1.1: The covalent bond formed between the α-corboxyl and α-amine groups of amino acids is called the peptide bond. 1 mole of H_2O is released after the formation of the peptide bond [1].

Protein molecules are nanoscopic structures and naturally cannot be seen by the naked eye. Three dimensional data about the proteins are gathered through elaborate imaging techniques, such as X-Ray Crystallography, Nuclear Magnetic Resonance (NMR)

Spectroscopy, electron microscopy and Neutron Diffraction. These data are then deposited to protein databanks e.g. PDB. A quick recent search on PDB website yielded that the longest globular protein (PDB ID: 2uva) has an approximate height of 27 nm and a width of 25 nm.

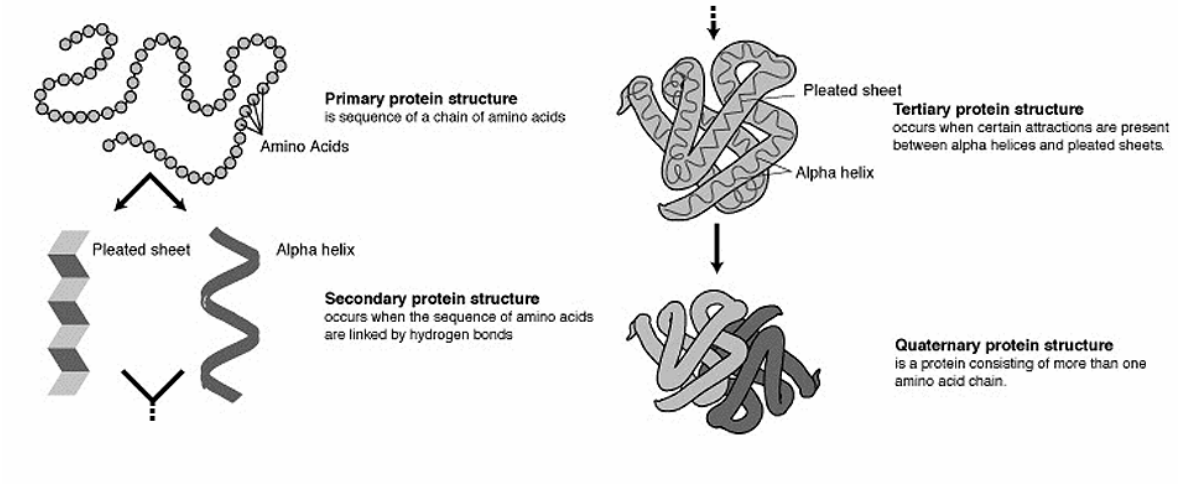


Figure 1.2: Protein structure, from primary to quaternary structure [1]

Proteins are flexible structures: their atoms exhibit fluctuations over time about well defined mean positions; and particularly, the backbone covalent bonds have the rotational freedom. The complexity of the protein structure can be better understood with the Levinthal Paradox according to which, for example, a 100-residue protein can assume $3^{100} = 5.10^{47}$ different conformations (considering the twist and turn of backbone covalent bonds). It is this flexibility that paves the way for proteins to perform various biological functions, including muscular contraction, cellular transportation.

This thesis has three chapters. In **Chapter 2**, we present our theoretical approach to generalize the elastic network models which provide important insights for understanding the flexibility-function relations of proteins. In **Chapter 3**, we discuss our results.

Chapter 2

STATISTICAL THERMODYNAMICS OF RESIDUE FLUCTUATIONS IN NATIVE PROTEINS

2.1 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 [2]. 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, C^{α}, is a convenient approximation where each successive α-carbon pair is connected by a virtual bond of fixed length. 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 [3]. Based on this simple picture of the elastic fluctuations of a protein, the Gaussian Network Model, GNM, was proposed [4,5], 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 [6,7]. As in the random network model, the GNM was based on

an isotropic description of residue fluctuations where only the number of neighbors of a given residue is important. Another model was then introduced to estimate the directions of fluctuations [8,9]. The latter is referred to as the Anisotropic Network Model, ANM. The GNM and models that followed it, collectively referred to as the Elastic network Models, ENM, 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 [5,8-43]. In general, these studies and several others that are cited by them, elaborate on different levels of approximation of the ENM's. 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, apply the models to drug design problems and prediction of binding sites, folding cores, allosteric effects and hot residues.

In this thesis, we present the statistical thermodynamics of fluctuations in the C^{α} based coarse-grained approximation and investigate the statistical basis of the two elastic network models, GNM and ANM. Specifically, we elaborate on their general features and limits of validity of the assumptions on which they are based. We validate our statements and conclusions by direct comparison with 6-40 ns molecular dynamics trajectories on ten different proteins [44].

2.2 Theory

In this section, we present the thermodynamic and statistical basis of fluctuations in native proteins. We use the entropy representation for the fundamental relation [45],

$$\boldsymbol{S} = \boldsymbol{S}(U, V, \boldsymbol{R}) \tag{2.1}$$

where $S,U,V,\boldsymbol{R}$ are the mean (thermodynamic) values of the entropy, energy, volume, and position vectors of C^{α}'s, respectively. Water is not shown explicitly in the fundamental relation and only a single protein molecule is considered. The protein and its environment constitute a small system and the fundamental relation and its arguments are regarded in this sense. In physics, the extensive variables are those that change proportionately to the size of the system while the intensive variables do not depend on the system size [46]. The extensive variables exhibit fluctuations about their native values. The distribution $\boldsymbol{f}\left(\hat{U},\hat{V},\hat{\boldsymbol{R}}\right)$ of the instantaneous extensive variables $\hat{U},\hat{V},\hat{\boldsymbol{R}}$ are given by the relation,

$$f\left(\hat{U},\hat{V},\hat{\boldsymbol{R}}\right)=exp\left\{-k^{-1}\ S\left[\frac{1}{T},\frac{P}{T},\frac{\boldsymbol{F}}{T}\right]-k^{-1}\left(\frac{1}{T}\hat{U}+\frac{P}{T}\hat{V}-\frac{\boldsymbol{F}}{T}\cdot\hat{\boldsymbol{R}}\right)\right\} \tag{2.2}$$

where k is the Boltzmann constant and $\mathrm{S}\left[\frac{1}{\mathrm{T}},\frac{\mathrm{P}}{\mathrm{T}},\frac{\mathbf{F}}{\mathrm{T}}\right]$ is given as:

$$\boldsymbol{S}\left[\frac{1}{T},\frac{P}{T},\frac{\boldsymbol{F}}{T}\right]=S-\frac{U}{T}-\frac{P}{T}V+\frac{\boldsymbol{F}}{T}\cdot\boldsymbol{R} \tag{2.3}$$

and P is pressure (defined as $-\frac{\partial U}{\partial V}$) and $\mathbf{F}$ (defined as $-\frac{\partial U}{\partial \boldsymbol{R}}$) is the force vector that contains the forces acting on the C^{α}'s. The distribution now takes the explicit form

$$f\left(\hat{U},\hat{V},\hat{\boldsymbol{R}}\right)=exp\left\{-k^{-1}\left[S-\frac{U}{T}-\frac{P}{T}V+\frac{\boldsymbol{F}}{T}\cdot\boldsymbol{R}\right]-k^{-1}\left(\frac{\hat{U}}{T}+\frac{P}{T}\hat{V}-\frac{\boldsymbol{F}}{T}\cdot\hat{\boldsymbol{R}}\right)\right\} \tag{2.4}$$

The correlation of fluctuations of the i^{th} and j^{th} residues may now be obtained from

$$\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j{}^T \right\rangle = \sum \left(\hat{\boldsymbol{R}}_i - \boldsymbol{R}_i \right) \left(\hat{\boldsymbol{R}}_j - \boldsymbol{R}_j \right)^T \; f\left(\hat{U}, \hat{V}, \hat{\boldsymbol{R}} \right) \tag{2.5}$$

where the superscript T denotes transpose and the summation is over all allowable microstates.

Using Eq. 2.4 in Eq. 2.5 leads to

$$\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j{}^T \right\rangle = -kT \left(\frac{\partial \boldsymbol{R}_i}{\partial \boldsymbol{F}_j} \right)_{T,P,F \; i \neq j} \tag{2.6}$$

where the variables to be kept fixed are indicated as subscripts. The derivation of Eq. 2.6 is given in Callen [44], which is outlined briefly in Appendix A.1.

Equation 2.6 forms the basis of all elastic network models for fluctuations in native proteins. The right-hand side of Eq. 2.6 may be evaluated if the energy of the system is known as a function of residue positions. For the case of pairwise potentials, the most general form of this relation is

$$E_{ij} = E_{ij}{}^0 \; f_{ij} \left(\frac{R_{ij}}{R_{ij}{}^c} \right) \tag{2.7}$$

where, $E_{ij}{}^0$ is the interaction energy of the residues i and j, $R_{ij}{}^c$ is the sum of the van der Waals radii which we define as the contact distance, i.e., the separation below which residues i and j repel each other, and f_{ij} is a dimensionless function of its arguments. A representative functional form for f_{ij} is the Mie potential which is given in Appendix A.2. and which we use for illustrative purposes in order to give an explicit interpretation of Eq.

2.6. For more detailed discussion of potentials the reader is referred to the work of Micheletti et. al [41].

The right-hand side of Eq. 2.3 can be expressed in terms of Ψ, the Euler form for generalized thermodynamic potential of the independent variables T, P, and $\boldsymbol{F}$, as $-\frac{\Psi(T,P,\boldsymbol{F})}{T}$. Knowing this relationship leads to five equations:

$$\begin{aligned} \Psi &= \Psi(T,P,\boldsymbol{F}) \\ \Psi &= U - TS + PV - \boldsymbol{FR} \\ S &= -\frac{\partial \Psi}{\partial T} \\ V &= \frac{\partial \Psi}{\partial P} \\ \boldsymbol{R} &= -\frac{\partial \Psi}{\partial \boldsymbol{F}} \end{aligned} \tag{2.8}$$

where $\boldsymbol{F}$ and $\boldsymbol{R}$ are 3N dimensional, but here we represented them as scalars for the clarity of the discussion. The four variables Ψ, T, P, and $\boldsymbol{F}$ may be eliminated among these five equations to yield $U = U(S,V,R)$. The forces are then obtained from U according to the relation

$$F = -\frac{\partial U(S,V,R)}{\partial R}$$

Considering pairwise potentials E_{ij} and concentrating on the position variables only, i.e., neglecting S and V dependence, the forces may be written as

$$\boldsymbol{F_i} = -\nabla_{\boldsymbol{R_i}} \sum \boldsymbol{E_{ij}} = -\sum_j \frac{\partial E_{ij}}{\partial \boldsymbol{R}_j} \tag{2.9}$$

Thus, the 3N dimensional force vector is obtained as a function of the position vectors of all the α-carbons. Performing the differentiations, shown in Appendix A.2., for the Mie potential, the following general relation is obtained:

$$\boldsymbol{F} = \boldsymbol{\Gamma}^{(3N)} \boldsymbol{R} \tag{2.10}$$

where, $\boldsymbol{\Gamma}^{(3N)}$ is a 3Nx3N matrix. Two different ordering of the $\boldsymbol{\Gamma}^{(3N)}$ matrix is used in the study of elastic network models. We name them as block representation and standard MD representation. For details see Appendix A.3. In the block representation described in Appendix A.3., eq 2.10 reads as:

$$\begin{bmatrix} \boldsymbol{F_X} \\ \boldsymbol{F_Y} \\ \boldsymbol{F_Z} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_X^{(N)} & 0 & 0 \\ 0 & \boldsymbol{\Gamma}_Y^{(N)} & 0 \\ 0 & 0 & \boldsymbol{\Gamma}_Z^{(N)} \end{bmatrix} \begin{bmatrix} \boldsymbol{R_X} \\ \boldsymbol{R_Y} \\ \boldsymbol{R_Z} \end{bmatrix} \tag{2.11}$$

In Eq. 2.11, $\boldsymbol{\Gamma}_X^{(N)}$, $\boldsymbol{\Gamma}_Y^{(N)}$ and $\boldsymbol{\Gamma}_Z^{(N)}$ are NxN. In the remaining sections, we will use the block representation. The order of the matrices, 3Nx3N or NxN, will be self-evident and will not be shown explicitly unless needed for clarity.

The derivative $\frac{\partial \boldsymbol{F}_i}{\partial \boldsymbol{R}_j}$ has to be evaluated for the correlation of fluctuations defined in Eq. 2.6.

This derivative is written as:

$$\frac{\partial \boldsymbol{F}_i}{\partial \boldsymbol{R}_k} = \boldsymbol{\Gamma}_{ij} \frac{\partial \boldsymbol{R}_j}{\partial \boldsymbol{R}_k} + \frac{\partial \boldsymbol{\Gamma}_{ij}}{\partial \boldsymbol{R}_k} \boldsymbol{R}_j \equiv \boldsymbol{\Gamma}_{ik} + \boldsymbol{\Lambda}_{ik} \tag{2.12}$$

The last equality in Eq. 2.12 defines the matrix $\boldsymbol{\Lambda}_{ik} = \frac{\partial \boldsymbol{\Gamma}_{ij}}{\partial \boldsymbol{R}_k} \boldsymbol{R}_j$ where $\frac{\partial \boldsymbol{\Gamma}_{ij}}{\partial \boldsymbol{R}_k}$ is third order, and its inner product with the position vector $\boldsymbol{R}_j$ gives a second order matrix that has the following block form:

$$\begin{bmatrix} \boldsymbol{\Lambda}_{XX} & \boldsymbol{\Lambda}_{XY} & \boldsymbol{\Lambda}_{XZ} \\ & \boldsymbol{\Lambda}_{YY} & \boldsymbol{\Lambda}_{YZ} \\ & & \boldsymbol{\Lambda}_{ZZ} \end{bmatrix} \tag{2.13}$$

where the symmetric lower half is not shown. The block matrices are of dimensions NxN, with

$$\boldsymbol{\Lambda}_{XX} = \begin{cases} 2\frac{\partial \boldsymbol{\Gamma}_{ij}}{\partial R_{ij}^{\,2}} \left(X_j - X_i\right)^2 & i \neq j \\ \\ -\sum_{k \neq j} \boldsymbol{\Lambda}_{jk} & i = j \end{cases} \tag{2.14}$$

where X_i and X_j are the X-components of the ith and jth residues, respectively. The terms for $\boldsymbol{\Lambda}_{YY}$ and $\boldsymbol{\Lambda}_{ZZ}$ are obtained similarly where Y and Z replaces the X's respectively. The first off-diagonal term $\boldsymbol{\Lambda}_{XY}$ is obtained as:

$$\boldsymbol{\Lambda}_{XY} = \begin{cases} 2\dfrac{\partial \boldsymbol{\Gamma}_{ij}}{\partial R_{ij}^{\,2}}\left(X_j - X_i\right)\left(Y_j - Y_i\right) & i \neq j \\ -\displaystyle\sum_{k \neq j} \boldsymbol{\Lambda}_{jk} & i = j \end{cases} \tag{2.15}$$

The terms for the other off diagonal blocks are written similarly, by replacing the variables in Eq. 2.15 accordingly. The derivation of Eqs. 2.14 and 2.15 are described in more detail in Appendix A.2.

Substituting Eq. 2.15 in Eq. 2.12 and using Eq. 2.6 leads to

$$\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T} \right\rangle = -\boldsymbol{kT}\left[\boldsymbol{\Gamma}_{ij} + \boldsymbol{\Lambda}_{ij}\right]^{-1} \tag{2.16}$$

Equation 2.16 can be rearranged to give the correlation matrix as the product of a correction matrix $\boldsymbol{C}$ and the GNM result:

$$\left\langle \Delta \boldsymbol{R}\ \Delta \boldsymbol{R}^{\boldsymbol{T}} \right\rangle = -\,\boldsymbol{k}\ \boldsymbol{C}\ \boldsymbol{\Gamma}^{-1} \tag{2.17}$$

where, $\boldsymbol{E}$ is the identity matrix and the correction matrix $\boldsymbol{C}$ is $\boldsymbol{C} = \boldsymbol{(E+\Gamma^{-1}\Lambda)^{-1}}$. When $\boldsymbol{\Lambda}$ vanishes, the GNM result is obtained.

Thus, $\boldsymbol{\Gamma^{-1}}$ represents the isotropic part of fluctuations in which the X, Y, and Z components cannot be identified independently, and there is no explicit dependence on a coordinate system. It consists of three identical sub-matrices on the diagonal, which are the inverses of the matrices given in Eq. 2.11. The three matrices, $\boldsymbol{\Gamma}_X$, $\boldsymbol{\Gamma}_Y$ and $\boldsymbol{\Gamma}_Z$ are identical as may be verified from Eq. A.3.2. This part is an indicator of the neighborhood effect in

fluctuations. The correction matrix, $\boldsymbol{C}$, contains nonzero diagonals as well as off-diagonal sub-matrices, which are functions of X, Y, and Z, and hence reflects the effects of anisotropy in proteins.

The structure of Eqs. 2.10, 11, 14 and 15 shows that any potential, which is a scalar function of the distances between residue pairs will yield identical forms for the $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ matrices. The only difference will arise from the scalar entries of the $\boldsymbol{\Gamma}$ matrix in Eq. 2.11 and the front factors $\frac{\partial \boldsymbol{\Gamma}_{ij}}{\partial R_{ij}^{\,2}}$ in Eq. 2.15.

In the next section, we discuss the various elastic network models in terms of Eq. 2.10.

2.3 Comparison of Different Elastic Network Models

2.3.1 The Gaussian Network Model

In the GNM the correction matrix $\boldsymbol{C}$ is taken as the identity matrix, and Eq. 2.17 takes the simple form $\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T} \right\rangle = \boldsymbol{\Gamma}_{ij}^{\,-1}$, where the matrix $\boldsymbol{\Gamma}$ is defined as

$$\Gamma_{ij} = \begin{cases} -\gamma^* & i \neq j \;\; and \;\; R_{ij} \leq r_{cutoff} \\ 0 & i \neq j \;\; and \;\; R_{ij} > r_{cutoff} \\ -\sum_k \gamma^* & i = j \neq k \end{cases} \tag{2.18}$$

Here, R_{ij} is the distance between the i^{th} an j^{th} 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.

According to the GNM assumption, the potential governing the interactions of a residue with its neighbors is isotropic. This is a consequence of the equality of the three submatrices shown in Eq. 2.11. Thus, the environment of a residue is isotropically smeared out and reference to an external coordinate system is not possible. In the general formulation, orientational correlations are embodied in the off-diagonal block matrices of $\boldsymbol{\Lambda}$. These matrices equate to zero in the GNM. One way to interpret this is that the direction of fluctuation of each residue changes randomly in time, which is the definition of isotropy. This also follows from the condition that the off-diagonal elements of the supermatrix $\boldsymbol{\Gamma}$ in Eq. 2.11 are zero, i.e., the correlations $\langle \Delta X_i \Delta Y_j \rangle$ and the similar mixed terms for $i \neq j$ are identically zero in the GNM. In order to understand this condition in more depth, we consider two limiting cases: Case (i) The fluctuations of residue i and j are totally uncorrelated. Then $\langle \Delta X_i \Delta Y_j \rangle$ averages out to zero. Case (ii) The directions of fluctuations of two residues are correlated, and these directions are in turn correlated with an externally fixed direction, then, and only then, the off-diagonal terms identified by $\langle \Delta X_i \Delta Y_j \rangle / \langle \Delta X_i^2 \rangle^{1/2} \langle \Delta Y_j^2 \rangle^{1/2}$ will be nonzero and equate to a value between -1 and 1. Terms that result from the realization of case (ii) are missing in the GNM. GNM identifies effects coming from the density of neighbors surrounding a given residue. The spatial variation of neighbor density is the essential driving force that identifies the differences in the mean square fluctuations and their correlations with neighbors.

2.3.2 The Anisotropic Network Model

The ANM was introduced by Hinsen [8] in order to couple the fluctuations of residues and domain motions to an external coordinate system. The model rests on the rotationally invariant form of the harmonic potential $U_{ij}(\boldsymbol{R}_{ij})$ given as $U_{ij}(\boldsymbol{R}_{ij}) = k(\boldsymbol{R}_{ij}^{\,0})(|\boldsymbol{R}_{ij}| - |\overline{\boldsymbol{R}}_{ij}|)^2$

where $\boldsymbol{R}_{ij}$ is the instantaneous value of the vector from residue i to j, $\overline{\boldsymbol{R}}_{ij}$ is its reference value, and vertical bars denote the magnitude. It is worth pointing out here that the presence of the magnitude $\left|\boldsymbol{R}_{ij}\right|$ in the pair potential renders the $\boldsymbol{\Gamma}$ matrix of Eq. 2.10 $\left|\boldsymbol{R}_{ij}\right|$ dependent, whereas the $\boldsymbol{\Gamma}$ of the GNM consists of $\left|\boldsymbol{R}_{ij}\right|$ independent constants. The differences between this potential and the GNM potential are further discussed in the Discussion section. Thus, the major interest of the ANM is centered around approximating the $\boldsymbol{\Lambda}$ matrix. In this approximation, the $\boldsymbol{\Lambda}_{ij}$ term given in Eq. 2.15 as $2\frac{\partial \Gamma_{ij}}{\partial R_{ij}^{\,2}}\left(X_j - X_i\right)\left(Y_j - Y_i\right)$ is replaced by $\frac{\gamma}{s_{ij}^{\,2}}\left(X_j - X_i\right)\left(Y_j - Y_i\right)$, where γ is the spring stiffness constant and s_{ij} is the magnitude of the vector between residues i and j. However, the $\boldsymbol{\Gamma}$ term in Eq. 2.16 is omitted in the ANM. Thus, effects coming from topology are not represented in the ANM directly. However, these missing terms, which are identified in the three diagonal block matrices, can indirectly be incorporated into the model by independently adjusting the front factor of the diagonal block. Inasmuch as the $\boldsymbol{\Gamma}$ matrix includes effects from neighbor densities in addition to those that are absent in $\boldsymbol{\Lambda}$, its inclusion into the ANM is expected to improve the agreement with experimental data. The shortcomings of the ANM which are already acknowledged in the literature are perhaps due to the absence of the $\boldsymbol{\Gamma}$ term [47, 48].

There is a basic correspondence between the Hessian matrix $\boldsymbol{H}$ of ANM and $\boldsymbol{\Lambda}$. Considering the $\boldsymbol{H}$ matrix in the block representation (see Appendix A.3.), the correspondence between $\boldsymbol{\Lambda}$ and $\boldsymbol{H}$ would look as follows:

$$H_{ij} = \begin{cases} \dfrac{\gamma \Lambda_{ij}}{2 s_{ij}^{\,2} \dfrac{\partial \Gamma_{ij}}{\partial R_{ij}^{\,2}}} & i \neq j \\ -\sum_{k \neq j} H_{jk} & i = j \end{cases} \tag{2.19}$$

2.4 Relationship of Fluctuation Directions to Protein Geometry

In this and the remaining sections, we use molecular dynamics trajectories of ten proteins (PDB codes: 1BFT, 1BZD, 1CD0, 1CJQ, 1CQK, 1MR8, 1QNZ, 1VFB, 1X2I, AND 1VII) for validating the basic features of the theory presented. Details of the molecular dynamics simulations are given in Appendix A.5. Previously, Micheletti et. al. [41], compared MD results with Gaussian model results. Their in depth analysis of MD trajectories and Gaussian models and their conclusion that quadratic models can efficiently characterize the vibrational motions of proteins near their native state encouraged us to go further in this direction.

When expressed in the block representation, the fluctuation correlation matrix has block diagonal and block off-diagonal components. The diagonal block matrices contain information from both $\boldsymbol{\Gamma}$ that are associated with neighbor density, and $\boldsymbol{\Lambda}$**.** The off-diagonal block matrices contain information on the coupling of fluctuations to an external coordinate system through $\boldsymbol{\Lambda}$. In this section, we discuss the second issue in more detail.

In modal space, the long wavelength fluctuations are those that relate to thermodynamic coordinates of the system [45]. The short wavelength modes are those that identify localized events. Among the thermodynamic coordinates are the volume and shape. On this basis, we postulate that the eigenvectors of the longest wavelength mode of fluctuations lie along the three principal directions of the moment of inertia tensor. This establishes a relationship between fluctuations and protein geometry or shape.

We use the results of the molecular dynamics trajectories to calculate the three principal directions of the fluctuation correlation matrix from the longest wavelength mode and compare these with the principal directions of the moment of inertia tensor.

From the molecular dynamics trajectory of a protein of N residues, the tx3N matrix of position vectors $\boldsymbol{R}$ of C^{α}'s is constructed for t snapshots. The instantaneous mean centered fluctuation, $\Delta\boldsymbol{R}$ is

$$\underset{[t\,x\,3N]}{\Delta\boldsymbol{R}} = \boldsymbol{R} - \boldsymbol{\mu_R} \tag{2.20}$$

where the dimension of the matrix is indicated in square brackets and $\boldsymbol{R}$ is the instantaneous fluctuation matrix composed of $\boldsymbol{R_i^t}$ position sub-matrices for atoms

$$\underset{[t\,x\,3N]}{\boldsymbol{R}} = \begin{bmatrix} R_1^1 & R_2^1 & R_3^1 & \cdots & R_N^1 \\ R_1^2 & R_2^2 & R_3^2 & \cdots & R_N^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ R_1^t & R_2^t & R_3^t & \cdots & R_N^t \end{bmatrix} \qquad \boldsymbol{R_i^t} = \begin{bmatrix} X_i & Y_i & Z_i \end{bmatrix} \tag{2.21}$$

and $\boldsymbol{\mu_R}$ is the mean of the fluctuations over the t snapshots.

We can decompose $\Delta\boldsymbol{R}$ into orthogonal modes of fluctuation using the Singular Value Decomposition, SVD:

$$\underset{[t\,x\,3N]}{\Delta\boldsymbol{R}} = \underset{[t\,x\,t]}{\boldsymbol{U}} \cdot \underset{[t\,x\,3N]}{\boldsymbol{S}} \cdot \underset{[3N\,x\,3N]}{\boldsymbol{V}^{\boldsymbol{T}}} \tag{2.22}$$

where, $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal matrices of the left singular ($\mathbf{u}_i$'s) and right singular vectors ($\mathbf{v}_i$'s), respectively. The v_i vector defines the direction of motion of α-carbon atoms

in the i[th] mode of fluctuation, $\mathbf{u}_i$ bears the time dependency of the directions of motions. The diagonal $\boldsymbol{S}$ matrix contains the singular values s_i for this decomposition which weight the modal fluctuations. The singular values are sorted in decreasing order.

Alternatively, Equation 2.22 can be written as the sum of modal fluctuations:

$$\underset{[t\,x\,3N]}{\Delta \boldsymbol{R}} = \sum_{i=1}^{3N} \underset{[t\,x\,1]}{\boldsymbol{u_i}} \cdot s_i \cdot \underset{[1\,x\,3N]}{\boldsymbol{v_i}^T} \tag{2.23}$$

The longest wavelength mode of fluctuations is then

$$\Delta \boldsymbol{R}^{(1)} = \boldsymbol{u}_1 \cdot s_1 \cdot \boldsymbol{v}_1^T \tag{2.24}$$

where, the superscript 1 in parenthesis identifies the longest wavelength mode. We denote the longest wavelength fluctuation vector for the i[th] frame of the trajectory by $\Delta \boldsymbol{R}^{(1)}(i)$. The elements of the fluctuation covariance matrix, $\boldsymbol{C_F}$, can be calculated

$$\underset{[3\,x\,3]}{\boldsymbol{C_F}} = \frac{1}{t} \sum_i \Delta \boldsymbol{R}^{(1)}(i)^T \cdot \Delta \boldsymbol{R}^{(1)}(i) \tag{2.25}$$

Applying eigenvalue decomposition to this symmetrical covariance matrix yields the eigenvalue and eigenvector matrices, $\boldsymbol{D}$ and $\boldsymbol{Q}$, respectively.

$$\boldsymbol{C_F} = \boldsymbol{Q_F} \cdot \boldsymbol{D_F} \cdot \boldsymbol{Q_F^T} \tag{2.26}$$

In this equation, vectors of the eigenvector matrix $\boldsymbol{Q_F}$ are the principal coordinate axes of the fluctuations. $\boldsymbol{D_F}$ is a diagonal matrix with eigenvalues in descending order. The eigenvalues show the variation of the trajectory along the corresponding eigenvectors.

The moment of inertia tensor of the protein is calculated by multiplication of the atomic coordinate matrix, $\boldsymbol{R_0}$, retrieved from the PDB file, with itself. Here the R_0 matrix is

$$\underset{[N\,x\,3]}{\boldsymbol{R}_0} = \begin{bmatrix} X_1^0 & Y_1^0 & Z_1^0 \\ X_2^0 & Y_2^0 & Z_2^0 \\ \cdots & \cdots & \cdots \\ X_N^0 & Y_N^0 & Z_N^0 \end{bmatrix} \tag{2.27}$$

and the moment of inertia tensor, $\boldsymbol{C_I}$, is

$$\underset{[3\,x\,3]}{\boldsymbol{C_I}} = \boldsymbol{R}_0^T \cdot \boldsymbol{R}_0 \tag{2.28}$$

Applying eigenvalue decomposition to this symmetrical matrix yields the eigenvalue and eigenvector matrices, $\boldsymbol{D_I}$ and $\boldsymbol{Q_I}$, respectively:

$$\boldsymbol{C_I} = \boldsymbol{Q_I} \cdot \boldsymbol{D_I} \cdot \boldsymbol{Q_I^T} \tag{2.29}$$

In this equation, columns of the eigenvector matrix $\boldsymbol{Q_I}$ are the principal directions of the moment of inertia tensor. $\boldsymbol{D_I}$ is a diagonal matrix with eigenvalues in descending order. The eigenvalues show the variation of the shape of the protein along the corresponding eigenvectors. Using the molecular dynamics trajectories for 10 proteins and their

corresponding geometry information, we calculated the angles between the principal directions of the moment of inertia tensor and those of the longest wavelength mode of fluctuations. The results are given in **Figure 2.1**. The filled circles show the angles between the principal moment of inertia direction and the principal direction of the longest wavelength fluctuation that corresponds to the largest eigenvalue of the fluctuation covariance matrix. With the exception of two proteins, the principal axes of fluctuation coincide with the principal geometry axes. The two outliers show a difference of about 20^{o} between the fluctuation and geometry principal axes. The squares and plusses denote results for the remaining two principal directions.

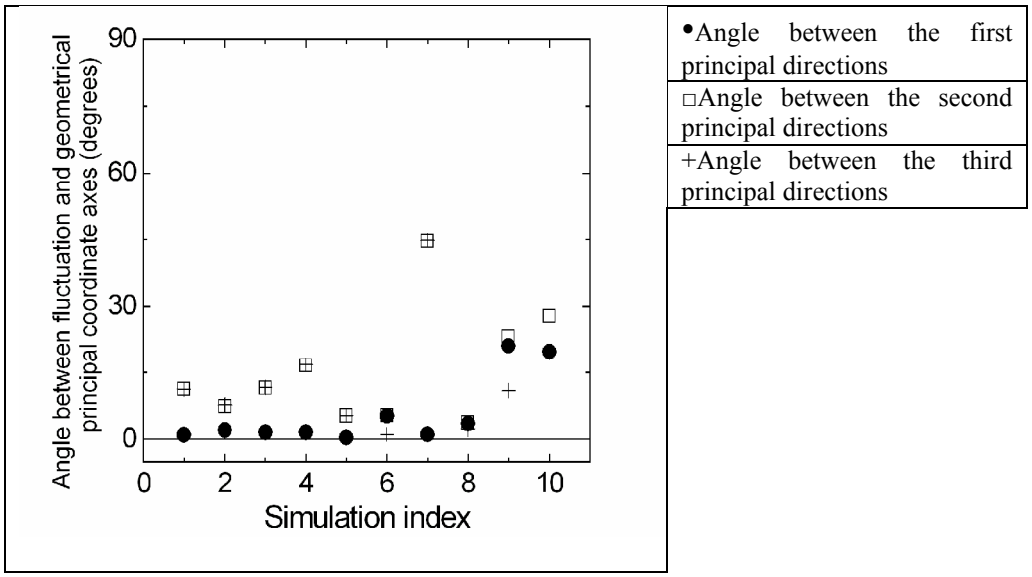


Figure 2.1: Angles between the principal directions of the moment of inertia tensor and of the longest wavelength mode of residue fluctuations.

We also compared the ratios of the eigenvalues for the moment of inertia tensor and those of the 3x3 fluctuation covariance matrix. The results are shown in **Figure 2.2** where

the abscissa is the ratio of an eigenvalue of the covariance matrix to the root sum of squares of the three eigenvalues. The ordinate is the same ratio for the moment of inertia tensor. Only the largest eigenvalues are included in the figure. The points lie approximately on a 45^{o} line which is drawn in the figure to guide the eye. The correlation coefficient of the points relative to the 45^{o} line is 0.85 and the standard deviation is 0.09. There is an outlier point in the figure. If this point is ignored, the correlation coefficient and the standard deviation become 0.97 and 0.07, respectively.

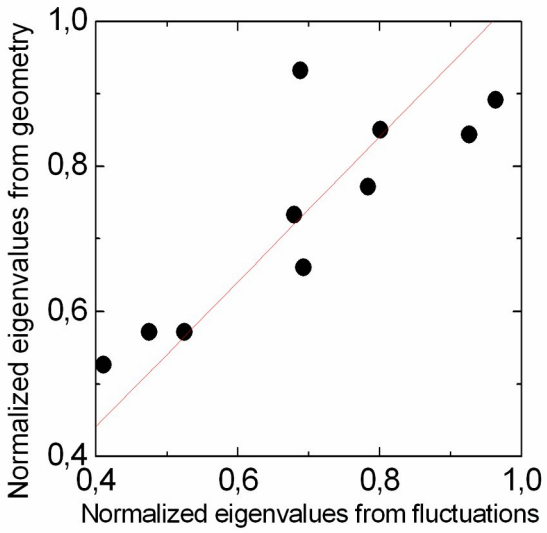


Figure 2.2: Scatter plot of normalized largest of the three eigenvalues of moment of inertia tensor plotted against the corresponding eigenvalue of the longest wavelength fluctuation covariance matrix. The normalization is made by dividing each eigenvalue by the root sum of squares of the three eigenvalues.

2.5 Coupling the GNM to External Coordinates

The Gaussian Network Model is based on the internal coordinates of a protein and information on the relation of residue fluctuation directions to an external coordinate system is lacking. However, the observation that the principal directions of the fluctuation matrix coincide approximately with the principal directions of the moment of inertia tensor may be used to introduce external coordinates to the GNM. Moment of inertia tensor matrix is of size three and its entries are the moment of inertia values for the three coordinate axes, x, y and z. In this section we discuss the possibility of relating GNM results, to external coordinates, approximately and at least in the long wavelength modes. We take the external coordinates to lie along the three principal directions of the moment of inertia tensor of the protein. More specifically, we attempt to obtain the averages $\langle \Delta X_i \Delta X_j \rangle$, $\langle \Delta Y_i \Delta Y_j \rangle$, $\langle \Delta Z_i \Delta Z_j \rangle$, $\langle \Delta X_i \Delta Y_j \rangle$, $\langle \Delta X_i \Delta Z_j \rangle$, $\langle \Delta Y_i \Delta Z_j \rangle$ where the ΔX, ΔY and ΔZ's are now along the principal directions of the moment of inertia tensor.

Our first assumption is that residues fluctuate along preferred directions relative to an external coordinate system. Molecular dynamics simulations indeed show that this assumption is valid to a significant degree. In order to define the anisotropy in fluctuation (including all modes) we calculated the smallest and largest normalized eigenvalues of the 3x3 fluctuation correlation matrix $\boldsymbol{C}_F$ for each residue and obtained the distribution of the ratio of the smallest to the largest eigenvalue for the residues of each protein. If the fluctuations are isotropic, then the eigenvalues will be close to 1/3, each, and the ratio of the smallest to the largest eigenvalue will be around unity. If, on the other hand, the residues fluctuate along preferred directions, this ratio will be close to zero. The first and second lines in **Table 2.1** show the mean and the standard deviation of the ratio for each distribution, respectively. The minimum and maximum ratios of each distribution are presented in the fourth and third rows, respectively. The mean values are much smaller

than unity and the standard deviations are small. Thus, the data shows a significant degree of anisotropy of residue fluctuations. Detailed plots are given in Appendix A.6.

Table 2.1: Anisotropy of fluctuations in MD simulations for C^{α}'s.

	1BFT	**1BZD**	**1CD0**	**1CJQ**	**1CQK**	**1MR8**	**1QNZ**	**1VFB**	**1X2I**	**1VII**	**Average**
Mean	0.34	0.23	0.26	0.22	0.19	0.23	0.27	0.29	0.18	0.08	0.23
Stdev	0.15	0.12	0.14	0.12	0.11	0.15	0.14	0.15	0.12	0.08	0.13
Min	0.09	0.05	0.02	0.05	0.02	0.02	0.02	0.05	0.05	0.01	0.04
Max	0.74	0.56	0.67	0.61	0.55	0.72	0.63	0.85	0.55	0.41	0.63

Below, we calculate the components of the fluctuation correlation matrix from GNM with respect to the principal directions of the moment of inertia tensor and compare with the corresponding results from MD.

The longest wavelength components of the correlations are obtained from the GNM according to the relation

$$\left\langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_j \right\rangle = \left\langle \Delta X_i \Delta X_j \right\rangle + \left\langle \Delta Y_i \Delta Y_j \right\rangle + \left\langle \Delta Z_i \Delta Z_j \right\rangle = \frac{1}{\lambda_1} v_{1i} v_{1j} \tag{2.30}$$

where, $\boldsymbol{\lambda_1}$ is the smallest singular value, and v_{1i} is the i[th] element of the singular direction corresponding to λ_1, and $\boldsymbol{v_{1j}}$ is for the j[th] element. Based on the assumption that residues fluctuate along preferred directions, and that all residues fluctuate in phase in a pure mode allows us to write $\Delta \boldsymbol{X_i} = \left\langle (\Delta \boldsymbol{X_i})^2 \right\rangle^{1/2}$ and $\Delta \boldsymbol{X_i} \Delta \boldsymbol{X_j} = \left\langle \Delta \boldsymbol{X_i} \Delta \boldsymbol{X_j} \right\rangle$, etc., for all i and j. In this notation, Eq. 2.30 is now written as

$$\Delta \boldsymbol{X_i} \Delta \boldsymbol{X_j} + \Delta \boldsymbol{Y_i} \Delta \boldsymbol{Y_j} + \Delta \boldsymbol{Z_i} \Delta \boldsymbol{Z_j} = \left\langle \Delta \boldsymbol{R_i} \cdot \Delta \boldsymbol{R_j} \right\rangle \tag{2.31}$$

The entries on the left hand side of Eq. 2.31 are the unknowns, and the elements of the NxN correlation matrix ones on the right hand side are known from GNM. A higher level of system of equations using linear algebra looks like this:

$$\underset{[N\,x\,3]}{\Delta \boldsymbol{R}} = \begin{bmatrix} \Delta \boldsymbol{X}_1 & \Delta \boldsymbol{Y}_1 & \Delta \boldsymbol{Z}_1 \\ \Delta \boldsymbol{X}_2 & \Delta \boldsymbol{Y}_2 & \Delta \boldsymbol{Z}_2 \\ \cdots & \cdots & \cdots \\ \Delta \boldsymbol{X}_N & \Delta \boldsymbol{Y}_N & \Delta \boldsymbol{Z}_N \end{bmatrix} \qquad \underset{[N\,x\,3][N\,x\,3]}{\Delta \boldsymbol{R}\ \Delta \boldsymbol{R}^T} = \underset{[N\,x\,N]}{\Gamma^{-1}} \tag{2.32}$$

We are searching for a $\Delta \boldsymbol{R}$ matrix that satisfies Eq. 2.32.

The set of equations given by Eq. 2.31 corresponds to $N(N+1)/2$ independent equations. The number of unknowns ΔX_i, ΔY_i, and ΔZ_i are 3N in number. We search for the solution of the 3N unknowns using the $\boldsymbol{N}(\boldsymbol{N}+1)/2$ equations. For $\boldsymbol{N} \geq 6$ the number of equations exceeds the number of unknowns, therefore different solutions are possible. We obtained the solutions for the 3N variables using nonlinear equation solver of Matlab®. This solver applies the Levenberg-Marquardt [49] method to the problem and needs a set of initial values of the 3N variables. Several solutions were obtained with different initial values for the $\Delta \boldsymbol{R}$ matrix of Eq. 2.32. The internal coordinate system for the set of 3N unknowns calculated using a given set of initial conditions are in general different than those calculated using another set of initial conditions.

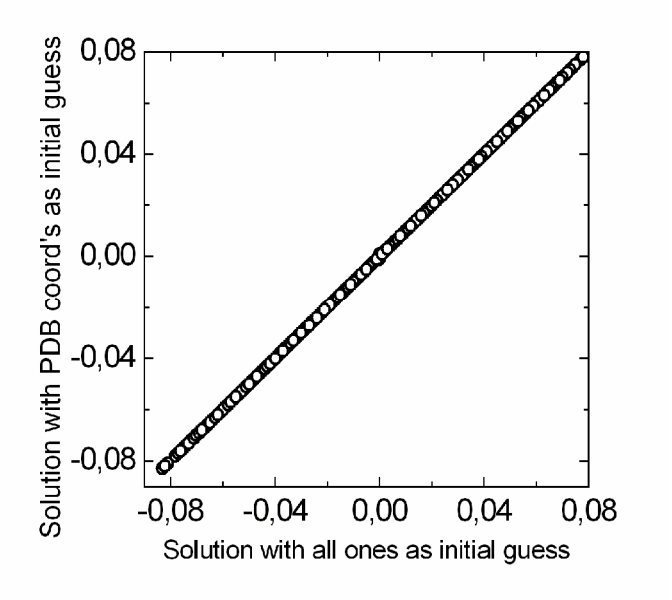


Figure 2.3: Scatter plot of corresponding entries of the correlation matrices obtained with the present method using unity matrix (abscissa) and using PDB coordinates for the Cα's (ordinate) as initial guesses for the nonlinear equation solution applied on the first mode of the correlation matrix of the GNM.

Each solution corresponds to a different internal reference frame. However, all different solutions become identical when expressed in the principal coordinates of the 3x3 correlation matrix by simple rotation as follows:

$$\underset{[Nx3]}{\Delta \boldsymbol{R}_R} = \underset{[Nx3]}{\Delta \boldsymbol{R}} \cdot \underset{[3x3]}{\boldsymbol{Q}_F}$$

where Q_F is the principal direction matrix of Eq. 2.26 and ΔR_R is the rotated version of solution for residue fluctuation direction.

In **Figure 2.3**, we show this identity for two solutions for one of which all the entries of the initial Nx3 $\Delta \boldsymbol{R}$ matrix were taken as unity and for the other, the PDB coordinates were assigned as the initial set for the 3N variables. Once the components of the fluctuation vectors are obtained in the principal coordinate system, we form the 3Nx3N correlation matrix. In **Figure 2.3**, the abscissa and the ordinate of a circle show the values of the correlation matrix obtained from the first and second solutions. The solutions are identical up to the third decimal point. Same is true for the other modes.

Having evaluated the elements of the fluctuation correlation matrix $\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T} \right\rangle$ by the numerical solver described above using GNM results for the longest wavelength, we now compare with the corresponding values obtained from MD trajectories. We note that the elements $\left\langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_j \right\rangle$ obtained for the GNM are in good agreement with the corresponding values from MD if the spring constant in the $\boldsymbol{\Gamma}$ matrix of the GNM is taken as 0.2.

As noted before by Doruker et. al. [42], there is a general correspondence between the eigenmodes of the GNM and MD simulations. In **Figure 2.4** we compare the values of the correlation matrix obtained for 1CQK by MD and GNM. A good agreement was obtained when the first mode of MD simulations (ordinate values) and the first three modes of GNM were compared as may be seen from **Figure 2.4**.

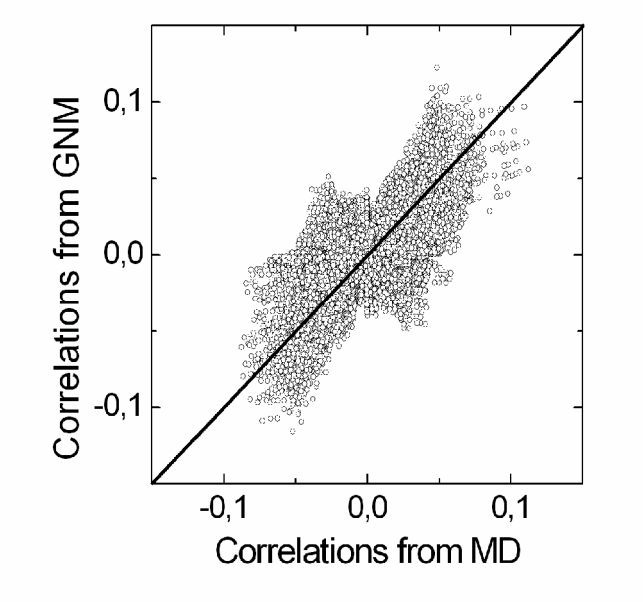


Figure 2.4: Scatter plot of corresponding entries of the correlation matrices obtained with the present method and from the 6 ns. MD simulation of the MAK33 Antibody (PDBID: 1CQK)

In Figure 2.5, we compare the contour plots for the anticorrelations obtained from the first three modes of the GNM (**Figure 2.5-a**) and from the first mode of MD simulations (**Figure 2.5-b**). A comparison of **Figures 2.5-a** and **2.5-b** shows the close resemblance of the results from MD and GNM.

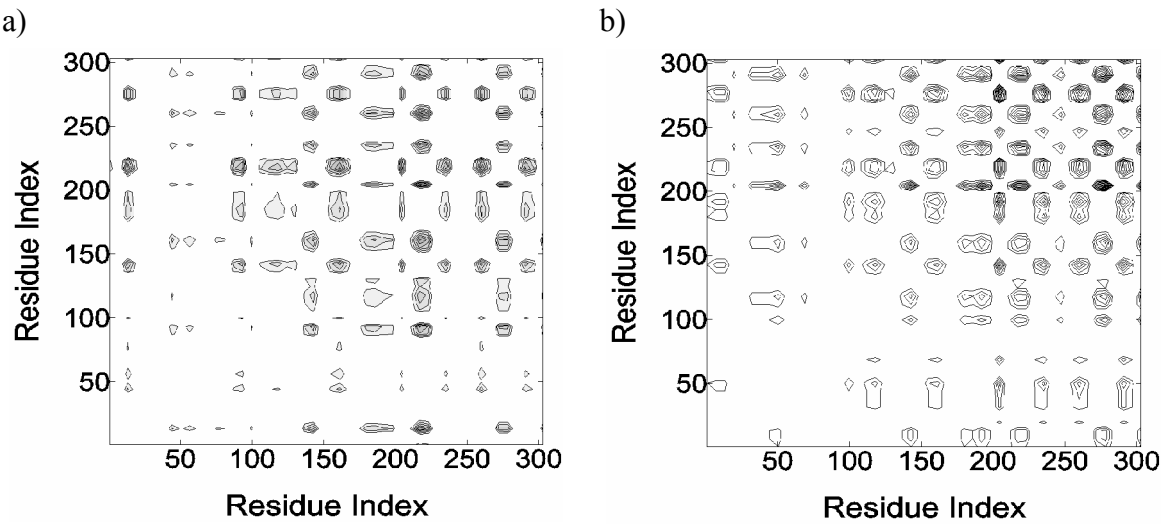


Figure 2.5: Contour plots for the anticorrelations in residue fluctuations obtained from the first three modes of the GNM (a) and from the first mode of MD simulations (b). The values in the upper right-hand block matrices in the plots express the correlation values in the dominant principal coordinate axis.

Chapter 3

CONCLUSIONS AND DISCUSSIONS

Statistical thermodynamics of fluctuations, represented by the Onsager relation given by Eq. 2.6 gives a general format for analyzing fluctuations in native proteins in a constant temperature-pressure bath. The right-hand side of Eq. 2.6 depends only on the mean positions of the residues and the forces that are functions of these mean positions. The relationship of mean positions to forces is obtained from thermodynamics by the spatial gradient of the governing potential. In the present thesis, we used the general Mie potential as an example to illustrate the various derivations for the models. It is to be noted that the choice of a translationally and rotationally invariant potential function is satisfactory for this purpose, and the Mie potential is one of them. The gradient of a general potential results in Eq. 2.12 that contains two matrix components, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$. In the most rigorous thermodynamic treatment, both of these matrices should contribute to the correlations as indicated in the present work. The GNM, being an isotropic approximation does not contain the $\boldsymbol{\Lambda}$ term. The ANM, on the other hand has been formulated with the $\boldsymbol{\Lambda}$ term only. It should be noted that these are approximate models, and the agreement of their results with a wide range of experimental data is remarkable. Recently, the approximation implicit in the GNM with the $\boldsymbol{\Lambda}$ term missing has been criticized by Thorpe on the grounds that this leads to loss of rotational invariance [50]. However, the GNM potential is non-metric and it is constructed by counting the number of neighbors of a given residue, only [48]. The same is true for the ANM. Thus, the absence of the $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ terms in these models is inconsequential as long as the models lead to consistent results. The present thesis is only an attempt to indicate the complete picture if the potentials were taken in their full

generality. In order to have a deeper understanding of the contributions of the $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ matrices, we discuss the problem in some more detail here. For the interest of transparency, we adopt the widely used Lennard-Jones potential obtained by taking m = 12 and n = 6 in the Mie potential. In this case, the total energy E of the system is obtained from Eq. A.2.1 as

$$E = \frac{1}{2}\sum_{j>i} {E_{ij}}^{0}\left[\left(\frac{\overline{R}_{ij}}{R_{ij}}\right)^{12} - 2\left(\frac{\overline{R}_{ij}}{R_{ij}}\right)^{6}\right] \tag{3.1}$$

where, $\overline{R}_{ij}$ is now the distance between residues i and j for which the potential is a minimum. Expanding Eq. 3.1 in Taylor's series and keeping the first two terms leads to the Gaussian approximation [42]

$$E = -\frac{1}{2}\sum_{j>i} {E_{ij}}^{0} + \frac{1}{2}\sum_{j>i}\left(\frac{36{E_{ij}}^{0}}{{\overline{R}_{ij}}^{2}}\right)\left(R_{ij} - \overline{R}_{ij}\right)^{2} \tag{3.2}$$

Here, R_{ij} and $\overline{R}_{ij}$ are the magnitudes, as appropriately indicated by Hinsen [8]. Linearization of Eq. 3.2 for small displacements leads to the following expression which is used in the literature for the ANM:

$$\boldsymbol{E} = -\frac{1}{2}\sum_{\boldsymbol{j>i}} {\boldsymbol{E_{ij}}}^{0} + \frac{1}{2}\sum_{\boldsymbol{j>i}}\left(\frac{\boldsymbol{36}{\boldsymbol{E_{ij}}}^{0}}{{\overline{\boldsymbol{R}}_{\boldsymbol{ij}}}^{2}}\right)\left[\left(\Delta\boldsymbol{R}_i - \Delta\boldsymbol{R}_j\right)\cdot\left(\overline{\boldsymbol{R}}_i - \overline{\boldsymbol{R}}_j\right)\right]^{2} \tag{3.3}$$

The steps leading to the term in the square brackets in Eq. 3.3, starting from Eq. 3.2 are outlined in Appendix A.4. The energy expression given by Eq. 3.3 is invariant under an infinitesimal rigid body rotation of the system about an axis.

The GNM replaces the harmonic term of Eq. 3.2 by $(\boldsymbol{R_{ij}} - \overline{\boldsymbol{R}}_{ij})^2$ which may be written in the equivalent form as $(\Delta\boldsymbol{R_i} - \Delta\boldsymbol{R_j})^2$. This leads to the GNM expression:

$$E = E_0 + \frac{1}{2}\boldsymbol{\Delta R^T}\ \boldsymbol{\Gamma}\ \boldsymbol{\Delta R} \tag{3.4}$$

where $\boldsymbol{\Gamma}$ is now a matrix whose elements are independent of $\boldsymbol{R_{ij}}$.

The term in the parentheses $(\boldsymbol{R_{ij}} - \overline{\boldsymbol{R}}_{ij})^2$ in the GNM expression is the difference between the vector $\boldsymbol{R_{ij}}$ from residue i to j at any time and its average value $\overline{\boldsymbol{R}}_{ij}$. There is no contribution to the energy if the vector $\boldsymbol{R_{ij}}$ lies along $\overline{\boldsymbol{R}}_{ij}$, and if their magnitudes are equal. However, it is also possible that $\boldsymbol{R_{ij}}$ may exhibit a pure rotation without its magnitude being changed. In this case, $(\boldsymbol{R} - \overline{\boldsymbol{R}}_{ij})^2$ will not be zero and pure rotations will be contributing to the energy. We note that the fluctuations of residues have a strong tendency to lie along a fixed direction, as implied by the MD data presented in Table I and by the more detailed calculations based on the MD trajectories for the ten proteins. In this case, the error introduced to the Gaussian formulation by pure rotational contributions is small, which justifies the adequacy of the Gaussian model as shown by a wide body of literature.

For the first time we show, using MD trajectories for 10 different proteins, that there is a strong direct relation between protein geometry and anisotropy of fluctuations. This observation, if verified by further examples, is expected to have far-reaching consequences

in the understanding of protein structure-function relations among which we can count protein design and docking studies.

Finally, using the geometry-fluctuation relation concept, we showed that the GNM can predict the anisotropy of fluctuations, at least in the few longest wavelength modes. Obviously, extending the GNM to predict anisotropy of fluctuations is only an attempt for understanding the fundamental features of the model and its limits. Otherwise, the ANM, properly formulated with both the $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ terms, is most suitable for predictions of anisotropic fluctuations.

APPENDIX

A.1. Onsager Relation

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

$$\frac{\partial f}{\partial\left(\boldsymbol{F}/T\right)} = k^{-1}\left(-R - \frac{\partial}{\partial \boldsymbol{F}} S\left[\frac{1}{T},\frac{P}{T},\frac{\boldsymbol{F}}{T},\frac{\mu}{T}\right]\right) f = -k^{-1}\left(\hat{\boldsymbol{R}} - \boldsymbol{R}\right) f_{U,V,N} \tag{A.1.1}$$

where, the subscripts of *f* on the right-hand side indicate that the system is in contact with a reservoir corresponding to *U, V, N*. Equation A.1.1 is valid irrespective of system size and is therefore suitable for the study of a single protein. Following Callen [45] and using A.1.1 in the definition for the correlation of fluctuations, Eq. 2.5, we obtain

$$\left\langle \Delta\hat{\boldsymbol{R}}_i \Delta\hat{\boldsymbol{R}}_j{}^T \right\rangle = -kT \sum\left(\hat{\boldsymbol{R}}_j - \hat{\boldsymbol{R}}_i\right)\frac{\partial f}{\partial F_j} = -kT\frac{\partial \boldsymbol{R}_i}{\partial \boldsymbol{F}_j} \tag{A.1-2}$$

The intermediate steps leading to the right-hand side of Eq. A.1.2 are given in Callen, p 427 [45].

A.2. Derivation of Matrix Entries Using Mie Potential

The intermolecular potential that forms the basis of the formulation has to be a scalar in R_{ij}'s and must be translation and rotation independent. In this Appendix, we give a special form for this potential, the Mie potential, so that the general theory presented in the text can be checked by means of this example potential.

The Mie potential is written in its full generality as

$$E_{ij} = E_{ij}^{\,0}\left[\left(\frac{n}{m}\right)^{\frac{m}{m-n}} - \left(\frac{n}{m}\right)^{\frac{n}{m-n}}\right]\left[\left(\frac{R_{ij}^{\,c}}{R_{ij}}\right)^{m} - \left(\frac{R_{ij}^{\,c}}{R_{ij}}\right)^{n}\right] \qquad \text{(A.2.1)}$$

where $E_{ij}^{\,0}$ is the interaction energy of the residues i and j, R_{ij} is the distance between them, conventionally taken as the distance between α-carbons, and $R_{ij}^{\,c}$ is the distance when the two residues are in contact. For $E_{ij}^{\,0}<0$ and the special case of n = -2, and (m/n)$\rightarrow 0^{-}$, the term in the first squared brackets in Eq. A.2.1 goes to unity and the Mie potential reduces to the harmonic potential.

The gradient is obtained by the chain rule,

$$\left(\frac{\partial E_{ij}}{\partial\left(R_{ij}\right)^2}\right)\frac{\partial\left(R_{ij}\right)^2}{\partial \boldsymbol{R}_j} = \left(\frac{\partial E_{ij}}{\partial\left(R_{ij}\right)^2}\right)\frac{\partial\left(\boldsymbol{R}_i\cdot\boldsymbol{R}_i - 2\boldsymbol{R}_i\cdot\boldsymbol{R}_j + \boldsymbol{R}_j\cdot\boldsymbol{R}_j\right)}{\partial \boldsymbol{R}_j} = \left(\frac{\partial E_{ij}}{\partial\left(R_{ij}\right)^2}\right)\left(\boldsymbol{R}_j - \boldsymbol{R}_i\right) \qquad \text{(A.2.2)}$$

Using the Mie potential for Eq. 2.7, using Eq. A.2.2 in the differentiation and rearranging the terms into a matrix form, we obtain Eqs. 2.10 where $\boldsymbol{\Gamma}$ is a NxN matrix the components of which are functions of $R_{ij}^{\,2}$. They are obtained from the Mie potential as follows:

$$\Gamma_{ij} = \begin{cases} 2E_{ij}^{\,0} \dfrac{\left[\left(\frac{n}{m}\right)^{\frac{m}{m-n}} - \left(\frac{n}{m}\right)^{\frac{n}{m-n}} \right]}{\left(R_{ij}^{\,c}\right)^2} \left[n \left(\dfrac{R_{ij}^{\,c\,2}}{R_{ij}^{\,2}} \right)^{\frac{n+2}{2}} - m \left(\dfrac{R_{ij}^{\,c\,2}}{R_{ij}^{\,2}} \right)^{\frac{m+2}{2}} \right] & i \neq j \\ -\sum_{k \neq j} \Gamma_{jk} & i = j \end{cases} \qquad \text{(A.2.3)}$$

The gradient of $\boldsymbol{\Gamma}$ is needed for obtaining the $\boldsymbol{\Lambda}$ matrix defined by $\Lambda_{ik} = \Gamma_{ij,k} R_j$. In order to perform this differentiation, we first take the first column of the $\boldsymbol{\Gamma}$ matrix, take its gradient, which gives a vector and then dot this with $\boldsymbol{R}$ and obtain the first column of $\boldsymbol{\Lambda}$. Applying the same operation to the remaining columns of Γ leads to the 3Nx3N $\boldsymbol{\Lambda}$ matrix. Rearranging the terms leads to Eqs. 2.14 and 2.15. The derivative $\partial \Gamma_{ij} / \partial R_{ij}^2$ that appears in these equations is given as follows for the Mie potential:

$$\partial \Gamma_{ij} / \partial R_{ij}^{\,2} = E_{ij}^{\,0} \frac{\left[\left(\frac{n}{m}\right)^{\frac{m}{m-n}} - \left(\frac{n}{m}\right)^{\frac{n}{m-n}} \right]}{\left(R_{ij}^{\,c}\right)^4} \left[-n(n+2) \left(\frac{R_{ij}^{\,c\,2}}{R_{ij}^{\,2}} \right)^{\frac{n+4}{2}} + m(m+2) \left(\frac{R_{ij}^{\,c\,2}}{R_{ij}^{\,2}} \right)^{\frac{m+4}{2}} \right] \qquad \text{(A.2.4)}$$

A.3 Permutation of Correlation Matrix Entries

Before we discuss the general features of Eq. 2.10, we point out that there are two different representations of the matrices $\boldsymbol{\Gamma}$ 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. 2.10 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 $\Delta\boldsymbol{R} = col\left[\Delta X_1, \Delta X_2, ..., \Delta X_N, \Delta Y_1, \Delta Y_2, ..., \Delta Y_N, \Delta Z_1, \Delta Z_2, ..., \Delta Z_N\right]$. In other elastic network models the standard MD representation is used according to which, $\Delta\boldsymbol{R}^t = col\left[\Delta X_1, \Delta Y_1, \Delta Z_1, \Delta X_2, \Delta Y_2, \Delta Z_2, ..., \Delta X_N, \Delta Y_N, \Delta Z_N\right]$. The correlation matrix $\boldsymbol{C}$ is accordingly written either as $\mathbf{C} = <\Delta R \Delta R^T>$ or $\mathbf{C^t} = <\Delta Rt \Delta R^{tT}>$. Both $\boldsymbol{C}$ and $\boldsymbol{C^t}$ are of order 3Nx3N, where N is the number of residues. The passage from one to the other is made by $\boldsymbol{C = TC^tT^T}$ where, $\boldsymbol{T}$ is a 3Nx3N permutation matrix formed as:

$$T_{ij} = \begin{cases} 1, & for\ i = 1,2,...,3N \quad and \quad j = 3((i-1) mod N) + \left\lfloor \dfrac{i-1}{N} \right\rfloor + 1 \\ 0, & otherwise \end{cases} \tag{A.3.1}$$

In the block representation the matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ are partitioned into submatrices as

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}(XX) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Gamma}(YY) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Gamma}(ZZ) \end{bmatrix} \quad \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}(\boldsymbol{XX}) & \boldsymbol{\Lambda}(\boldsymbol{XY}) & \boldsymbol{\Lambda}(\boldsymbol{XZ}) \\ - & \boldsymbol{\Lambda}(\boldsymbol{YY}) & \boldsymbol{\Lambda}(\boldsymbol{YZ}) \\ - & - & \boldsymbol{\Lambda}(\boldsymbol{ZZ}) \end{bmatrix} \tag{A.3.2}$$

where, each submatrix is NxN. The second submatrix $\boldsymbol{\Lambda}(XY)$ for example, has the mixed products $\Delta X_i \Delta Y_j$.

A.4 Linearzation of Hinsen's Potential

Starting with the term in the parenthesis, $(R_{ij} - \overline{R}_{ij})^2$, in Eq. 3.1 and expanding leads to:

$$R_{ij}{}^2 - 2R_{ij} \cdot \overline{R}_{ij} + \overline{R}_{ij}{}^2 \qquad \text{(A.4.1)}$$

We approximate (A.4.1) with respect to R_{ij} by writing the second term as $2R_{ij} \cdot \overline{R}_{ij} \approx 2\overline{R}_{ij}{}^2$ to obtain $(R_{ij} - \overline{R}_{ij})^2 \approx R_{ij}^2 - \overline{R}_{ij}^2$. Linearizing the term R_{ij}^2 as $R_{ij} \cdot \overline{R}_{ij}$ leads to:

$$\left(R_{ij} - \overline{R}_{ij}\right)^2 \approx R_{ij}\overline{R}_{ij} - \overline{R}_{ij}{}^2 = \left(\boldsymbol{R}_i - \boldsymbol{R}_j\right) \cdot \left(\overline{\boldsymbol{R}}_i - \overline{\boldsymbol{R}}_j\right) - \overline{R}_{ij}{}^2 \qquad \text{(A.4.2)}$$

After a series of rearrangements, we get:

$$\left(\boldsymbol{R}_i - \boldsymbol{R}_j\right) \cdot \left(\overline{\boldsymbol{R}}_i - \overline{\boldsymbol{R}}_j\right) - \overline{R}_{ij}{}^2 = \boldsymbol{R}_i \cdot \overline{\boldsymbol{R}}_i - \boldsymbol{R}_i \cdot \overline{\boldsymbol{R}}_j - \boldsymbol{R}_j \cdot \overline{\boldsymbol{R}}_i + \boldsymbol{R}_j \cdot \overline{\boldsymbol{R}}_j - \overline{R}_{ij}{}^2$$

$$= \boldsymbol{R}_i \cdot \overline{\boldsymbol{R}}_i - \boldsymbol{R}_i \cdot \overline{\boldsymbol{R}}_j - \boldsymbol{R}_j \cdot \overline{\boldsymbol{R}}_i + \boldsymbol{R}_j \cdot \overline{\boldsymbol{R}}_j - \left(\overline{\boldsymbol{R}}_i - \overline{\boldsymbol{R}}_j\right)^2$$

$$= \boldsymbol{R}_i \cdot \overline{\boldsymbol{R}}_i - \boldsymbol{R}_i \cdot \overline{\boldsymbol{R}}_j - \boldsymbol{R}_j \cdot \overline{\boldsymbol{R}}_i + \overline{\boldsymbol{R}}_j \cdot \boldsymbol{R}_j - \overline{\boldsymbol{R}}_i{}^2 + 2\overline{\boldsymbol{R}}_j \cdot \overline{\boldsymbol{R}}_i - \overline{\boldsymbol{R}}_j{}^2$$

$$= \left[\left(\boldsymbol{R}_i - \overline{\boldsymbol{R}}_i - \boldsymbol{R}_j + \overline{\boldsymbol{R}}_j\right) \cdot \left(\overline{\boldsymbol{R}}_i - \overline{\boldsymbol{R}}_j\right)\right]^2$$

$$= \left[\left(\Delta\boldsymbol{R}_i - \Delta\boldsymbol{R}_j\right) \cdot \left(\overline{\boldsymbol{R}}_i - \overline{\boldsymbol{R}}_j\right)\right]^2 \qquad \text{(A.4.3)}$$

This is the linearized form used in Hinsen's work and in all subsequent applications of ANM. It is valid for small displacements because it rests on the linearization assumption used in its derivation. This form of the energy is invariant if an infinitesimal rigid body rotation is applied.

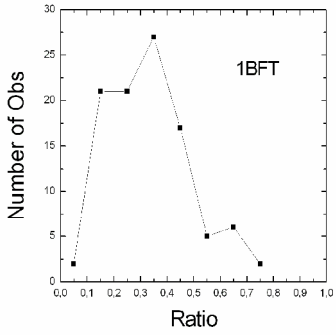
A.5. MD Simulation Configuration

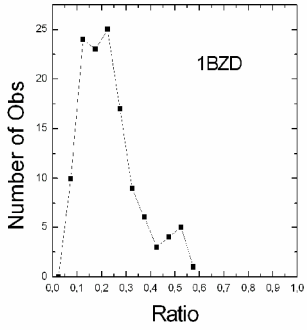
The same approach as Yogurtcu et. al. [44] is used. The MD simulations were performed with NAMD [51] with CHARMM27 [52] force-field parameters for 6 ns. The protein complexes were solvated using the TIP3 water molecules in the VMD [53] package. Particle-mesh Ewald [54] was applied in the simulations. We used the VMD Autoionize, adding sodium and chlorine ions, to neutralize the system. NVT ensemble and periodic boundary conditions with a rectangular box were applied in the simulations. The temperature in the simulations was kept constant at 300 K by using Langevin dynamics. Initial equilibration was done for 10,000 steps, followed by 6-40 ns runs. The time step was 2 fs. The first 1-ns runs were further discarded to assure that the data collected are after equilibration. Trajectories were sampled at 40-ps intervals. The simulations were carried out in a Linux-based cluster from a Racksaver cluster and each node has two 3.06 GHz Intel Pentium Xeon processors and Beowulf Cluster with nodes having Intel Pentium 4 2.4 GHz processors.

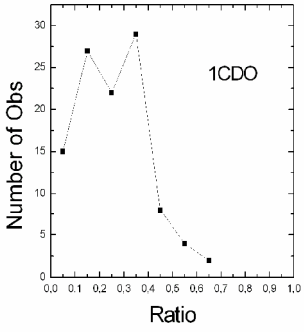
A.6. Distributions of Anisotropy Values

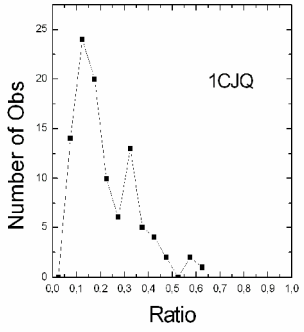
The distributions of the anisotropy values for the ten proteins are given in **Figure A.6.1**. The abscissa in the ten figures below is the ratio of the eigenvalues as described in the thesis, and the ordinate is the frequency of observation. The proteins are identified in each subfigure. All of the figures above show that the peaks are significantly shifted to smaller

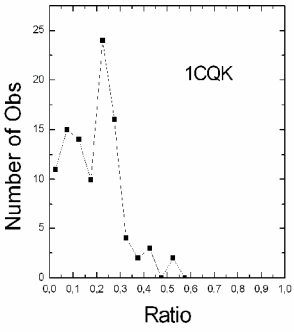
values that is the signature of anisotropy. The values given in **Table 2.1** are taken from these graphs.

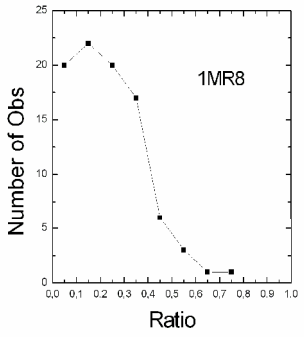












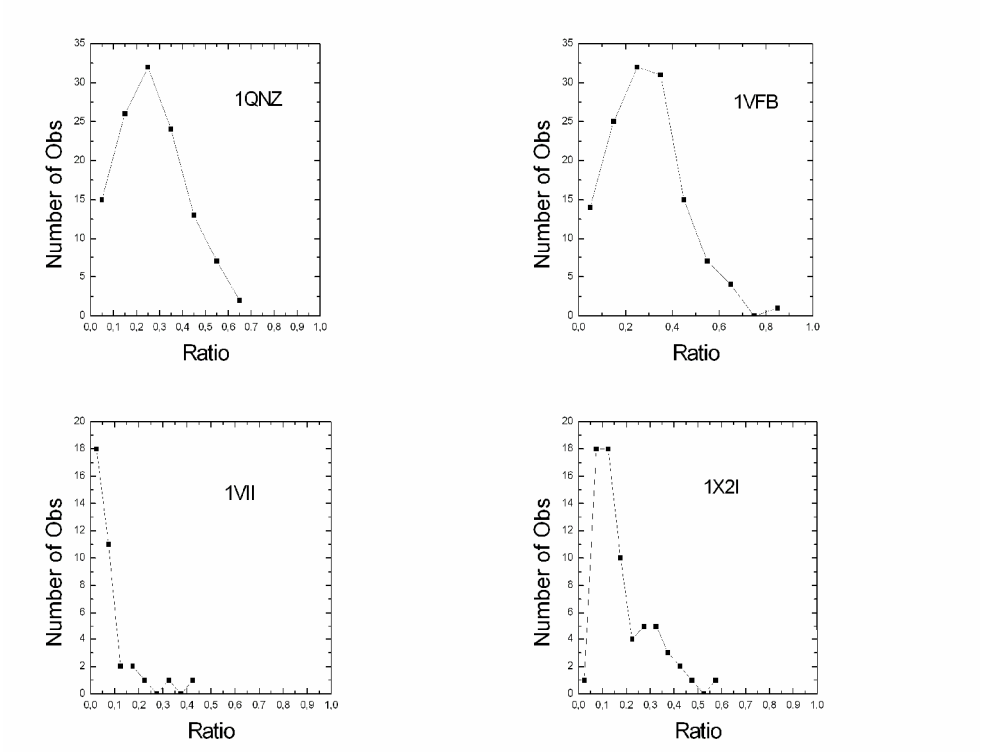


Figure A.6.1: The distributions of the fluctuation anisotropy values for the ten proteins

BIBLIOGRAPHY

1. US NIH, National Human Genome Research Institute

2. Berman, H. M., Westbrook, J., Feng, Z., et al. The Protein Data Bank 2000

3. Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis 1996

4. Bahar, I., Atilgan, A. R. and Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential 1997

5. Haliloglu, T., Bahar, I. and Erman, B. Gaussian Dynamics of Folded Proteins 1997

6. Flory, PJ, Gordon, M. and McCrum, NG. Statistical Thermodynamics of Random Networks [and Discussion] 1976

7. Kloczkowski, A., Mark, JE and Erman, B. Chain dimensions and fluctuations in random elastomeric networks. 1. Phantom Gaussian networks in the undeformed state 1989

8. Hinsen, K. Analysis of domain motions by approximate normal mode calculations 1998

9. Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O. and Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model 2001

10. Delarue, M., Sanejouand, Y. H. Simplified Normal Mode Analysis of Conformational Transitions in DNA-dependent Polymerases: the Elastic Network Model 2002

11. Keskin, O., Bahar, I., Flatow, D., Covell, D. G. and Jernigan, R. L. Molecular mechanisms of chaperonin GroEL-GroES function 2002

12. Zheng, W., Brooks, B. R. and Hummer, G. Protein conformational transitions explored by mixed elastic network models 2007

13. Yang, L., Song, G. and Jernigan, R. L. How well can we understand large-scale protein motions using normal modes of elastic network models? 2007

14. Cui, Q., Bahar, I. Normal Mode Analysis: Theory And Applications to Biological And Chemical Systems 2006

15. Demirel, MC, Atilgan, AR, Jernigan, RL, Erman, B. and Bahar, I. Identification of kinetically hot residues in proteins 1998

16. Jeong, J. I., Jang, Y. and Kim, M. K. A connection rule for alpha-carbon coarse-grained elastic network models using chemical bond information 2006

17. Leherte, L., Vercauteren, D. P. Collective motions in protein structures: Applications of elastic network models built from electron density distributions 2008

18. Song, G., Jernigan, R. L. vGNM: a better model for understanding the dynamics of proteins in crystals 2007

19. Lavery, R., Sacquin-Mora, S. Protein mechanics: a route from structure to function 2007

20. Marsella, L. Modeling Truncated Hemoglobin vibrational dynamics 2006

21. Kundu, S., Sorensen, D. C. and Phillips, G. N.,Jr. Automatic domain decomposition of proteins by a Gaussian Network Model 2004

22. Eom, K., Baek, S. C., Ahn, J. H. and Na, S. Coarse-graining of protein structures for the normal mode studies 2007

23. Doruker, P., Jernigan, R. L. and Bahar, I. Dynamics of large proteins through hierarchical levels of coarse-grained structures 2002

24. Haliloglu, T., Bahar, I. Structure-based analysis of protein dynamics: comparison of theoretical results for hen lysozyme with X-ray diffraction and NMR relaxation data 1999

25. Bahar, I., Atilgan, A. R., Demirel, M. C. and Erman, B. Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability 1998

26. Bahar, I., Erman, B., Jernigan, R. L., Atilgan, A. R. and Covell, D. G. Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function 1999

27. Eyal, E., Chennubhotla, C., Yang, L. W. and Bahar, I. Anisotropic fluctuations of amino acids in protein structures: insights from X-ray crystallography and elastic network models 2007

28. Yang, L. W., Eyal, E., Chennubhotla, C., Jee, J., Gronenborn, A. M. and Bahar, I. Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions 2007

29. Eyal, E., Yang, L. W. and Bahar, I. Anisotropic network model: systematic evaluation and a new web interface 2006

30. Yang, L. W., Rader, A. J., Liu, X., et al. oGNM: online computation of structural dynamics using the Gaussian Network Model 2006

31. Chennubhotla, C., Bahar, I. Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES 2006

32. Yang, L. W., Liu, X., Jursa, C. J., et al. i GNM: a database of protein functional motions based on Gaussian Network Model 2005

33. Wang, Y., Rader, AJ, Bahar, I. and Jernigan, R. L. Global ribosome motions revealed with elastic network model 2004

34. Rader, AJ, Bahar, I. Folding core predictions from network models of proteins 2004

35. Xu, C., Tobi, D. and Bahar, I. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin TR2 transition 2003

36. Erman, B., Dill, K. Gaussian model of protein folding 2000

37. Erkip, A., Erman, B., Seok, C. and Dill, K. Parameter optimization for the Gaussian model of protein folding 2002

38. Liao, J. L., Beratan, D. N. How does protein architecture facilitate the transduction of ATP chemical-bond energy into mechanical work? The cases of nitrogenase and ATP binding-cassette proteins 2004

39. Tama, F., Brooks, C. L.,3rd. The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus 2002

40. Bahar, I., Rader, A. J. Coarse-grained normal mode analysis in structural biology 2005

41. Micheletti, C., Carloni, P. and Maritan, A. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models 2004

42. Doruker, P., Atilgan, A. R. and Bahar, I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor 2000

43. Erman, B. The gaussian network model: precise prediction of residue fluctuations and application to binding problems 2006

44. Yogurtcu, O. N., Erdemli, S. B., Nussinov, R., Turkay, M. and Keskin, O. Restricted Mobility of Conserved Residues in Protein-Protein Interfaces in Molecular Simulations 2008

45. H. B. Callen, Thermodynamics and an introduction to thermostatistics, Second ed. (Wiley, 1985).

46. T. L. Hill, Thermodynamics of small systems. (Dover, New York, 1994).

47. Kundu, S., Melton, J. S., Sorensen, D. C. and Phillips, G. N. Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models 2002

48. Bahar, I. On the theoretical foundations of the Gaussian network model and its applications to proteins 2007

49. Mor, J. J. The Levenberg-Marquardt algorithm: implementation and theory 1977

50. Thorpe, M. Comment on elastic network models and proteins Phys 2007

51. Kale, L., Skeel, R., Bhandarkar, M., et al. NAMD2: Greater scalability for parallel molecular dynamics 1999

52. MacKerell, A. D.,Jr, Banavali, N. and Foloppe, N. Development and current status of the CHARMM force field for nucleic acids 2000

53. Humphrey, W., Dalke, A. and Schulten, K. VMD: visual molecular dynamics 1996

54. Darden, T., York, D. and Pedersen, L. Particle mesh Ewald: An N.log (N) method for Ewald sums in large systems 1993

Vita

Osman N. Yoğurtçu was born in 1983, Istanbul, Turkey. He had been to Vefa Anadolu and Kültür Fen Lisesi for high school education. He received his BS degree in 2006, in Electrical & Electronics Engineering at Koç University. He collaborated on various projects with professors B. Erman, E. Erzin, A. Gürsoy, M. Türkay, Ö. Keskin, R. Nussinov, N. Ben-Tal and T. Haliloglu. At Koç University, he pursued MSc degree in Computational Sciences and Engineering during 2006-2008.

He currently lives in Istanbul, Turkey, and will join the department of mechanical engineering at Johns Hopkins University, Maryland, USA to complete his PhD degree.