Binding and Mode Coupling of Proteins

by

Mert Gür

A Thesis Submitted to the

Graduate School of Engineering

in Partial Fulfillment of the Requirements for

the Degree of

Doctor of Philosophy

in

Computational Science and Engineering

Koç University

October, 2010

Koç Üniversitesi
Fen Bilimleri Enstitüsü

Doktora Tez Sınavı Tutanağı

22 Ekim 2010

Fen Bilimleri Enstitüsü Hesaplamalı Bilimler ve Mühendislik bölümü doktora öğrencilerinden 20070959 numaralı **Mert Gür**'ün sözlü tez jüri sınavı 22 Ekim 2010 tarihinde yapılmış ve adı geçen öğrencinin "Binding and Mode Coupling of Proteins" başlıklı doktora tezi başarılı bulunmuştur.

Tez Jüri Üyeleri:

Prof. Dr. Burak Erman .....

Prof. Dr. İvet Bahar .....

Prof. Dr. Atilla Gürsoy .....

Doç. Dr. Özlem Keskin .....

Yrd. Doç. Dr. Alkan Kabakçıoğlu .....

ABSTRACT

Contribution of intermolecular interactions to the stability of two bound molecules is an important factor in flexible binding problems. In order to understand the the change in thermodynamic properties upon binding and determine the binding sides, two hexa-peptides and their bound complex structures were analyzed. The dynamics of the peptides are obtained via molecular dynamics. In order to extract the thermodynamic properties and determine the binding side, a harmonic model was applied.

The interplay between harmonicity and anharmonicity in proteins was studied in literature, reaching the conclusion that the motion within a local minimum is mainly harmonic and the anharmonic component arises from transitions from one minimum to the other. The harmonic 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.

Mode coupling and anharmonicity in a native fluctuating protein is investigated in modal space by projecting the motion along the eigenvectors of the fluctuation correlation matrix. Molecular dynamics trajectories of Crambin are generated and used to evaluate the terms of the polynomials and to obtain the modal energies. Slowest modes have energies that are below that of the harmonic energy, kT/2 per mode, and a few fast modes have energies significantly larger than the harmonic which is a result of coupling. Detailed analysis of the coupling of these modes to others is presented in terms of the lowest order two mode coupling terms. In addition is was shown that the mode coupling and anharmonicity are important for modeling the multidimensional energy landscape of the protein Crambin. The effect of them on the fluctuational entropy is on the order of a few percent.

In order to understand the structure –binding relationship from a different perspective, the fluctuations and free energy profiles of two very similar proteins, which differ only 2 aminoacids, were investigated; HLA-B51 and HLA-B52. HLA-B51 is related to the Behçet's disease, which is a chronic inflammatory disorder, whereas HLA-B52 is not related to it. The unbinding process of a peptide of sequence YAYDGKDYI, which is one of the catabolic products of HLA-B51 and also known to bind well to HLA-B52, was investigated. Change in the dynamics of 1 helix, residues 60-90, were analyzed. Free energy profiles have shown that unbinding from HLA-B52 results in greater free energy differences than for HLA-B51.

ÖZET

Moleküller arası etkileşimin kararlılık üzerine etkisinin anlaşılması esnek bağlanma probleminde önemli bir faktör oluşturmaktadır. Bağlanmadan kaynaklanan termodinamik değişimlerinin daha iyi anlaşılması ve bağlanma bölgelerinin belirlenebilmesi amacıyla iki hexa-peptit ve onların bağlanmış formu analiz edildi. Moleküler dinamik vasıtasıyla peptitlerin dinamiği elde edildi. Bu bilgiden istenen termodinamik özeliklerin elde edilmesi ve bağlanma bölgelerinin belirlenmesi amacıyla harmonik bir model uygulandı.

Literatürde bir lokal minimum etrafindaki hareketin harmonik olduğu ve anharmonik hareketlerin lokal minimumlar arası geçişlerden kaynaklandığı sonucuna varılmıştır. Büyük hareketleri açıklanabilmesi için harmonik matematiksel formulasyon genişletildi. Dalgalanma olasılık fonksiyonu yüksek derece momentleri kullanarak Hermit serisi açılımı ile oluşturuldu.

Proteinlerin çevre ile enerjik etkileşimi ve emilen enerjinin protein içindeki rezidülere dağılımı protein fonksiyonu açısından çok önem arz eden bir konu. Bu sebepten dolayı modların bağlanması ve anharmonik hareketlerin mod uzayına korelasyon matrisinin eigenvektorlerine yansıtılması yolu ile incelenmiştir. Crambin için moleküler dinamik yörüngeleri elde edildi ve bununla polinomlar oluşturularak mod enerjileri hesaplandı. Yavaş modların enerjisinin harmonik kabulün, mod başına kT/2, altında olduğu görüldü. Hızlı modlardan birkaçının ise harmonik kabulün üzerinde enerjiye sahip oldukları görüldü. Bu sapmaların sebebi çevre ile ve başka bir veya birkaç mod ile enerjik bağlanma olarak açıklanabilinir. Bu enerjik bağlanmalarını detaylı analizi ikili en düşük derecede bağlanmalar cinsiden yapıldı. İlaveten Crambin için modların bağlanması ve anharmonik hareketlerin çok boyutlu enerji yüzeyinin modellenmesinde önemli etkiye sahip olduğu görüldü. Anharmonik hareketlerin ve mod bağlanmasının entropi üzerindeki etkisi sadece yüzde birkaç düzeyinde.

Yapı-bağlanma ilişkisini farklı bir açıdan incelemek amacıyla birbirine çok benzer olan ve sadece 2 amino grubu ile farklılık gösteren HLA-B51 ve HLA-B52 proteinlerinin serbest enerji profilleri ve hareketleri incelendi. HLA-B51 kronik iltihaplı bir hastalık olan Behçet hastalığı ile ilgili bir protein, HLA-B52 ise ilgisiz. YAYDGKDYI sekansına sahip HLA-B51'in katabolik bir ürünü olan ve HLA-B52'ye de iyi bağlandığı bilinen bir peptittin bağlanma prosesi incelendi. 1 heliksindeki dinamik değişimler incelendi. Serbest enerji profilleri HLA-B52 bağlanmasının HLA-51'e göre daha çok enerji açığa çıkardığını gösterdi.

ACKNOWLEDGMENT

I owe my deepest gratitude to my supervisor and mentor, **Prof. Burak Erman,** whose encouragement, guidance and continuous support during my doctoral studies had a profound impact on my academic character. During the past three years he challenged my limits continuously and his perpetual energy and enthusiasm in research has motivated me all along the way. He has always been accessible, patient and willing to help me. I could not wish for a better advisor than him.

Assoc. Prof. Özlem Keskin and **Prof. Attila Gürsoy** deserve special thanks since they have shared their knowledge with me and supported me all through my doctoral studies. Moreover, I am so happy to have had them as thesis committee members. I was delighted to interact with **Prof. Yaman Arkun, Prof. Tekin Dereli, Prof. Ersin Yurtsever, Assoc. Prof. Alper Demir, Assoc. Prof. Alper Tunga Erdoğan, Asst. Prof. Alkan Kabakçıoğlu** and **Asst. Prof. Deniz Yüret.** I thank them for giving me advice and insight on various topics. I offer my deep appreciation to my thesis monitoring committee members, **Prof. Burak Erman**, **Prof. Attila Gürsoy** and **Asst. Prof. Alkan Kabakçıoğlu,** for their constructive criticism and suggestions which guided me throughout my doctoral studies.

I am grateful to **Prof. Ivet Bahar** for her support during my last months of my doctoral studies, being a thesis committee member and offering me a post-doctoral position one year prior to my graduation. This gave me a feeling of security and allowed me to dedicate myself fully to my doctoral work.

I especially thank **Giray Yıllıkçı** and **Gözde Eskici** for their great friendship. **Gözde Eskici** has been always glad to share her biology knowledge whenever it was required. In the past 3 years **Server Levent Yılmaz** at the University of Pittsburgh was always there to answer any of my questions regarding parallel programming and he helped me to optimize my simulations. It gives me great pleasure to acknowledge the support of **Ergun Biçici** and **Mehmet Ali Yatbaz** regarding the software set up on the clusters and helping me out with the computer languages.

In addition, I am so happy that I had such a friendly and cheerful group of fellow students namely **Yasemin Demir, Bekir Yenilmez, Hakan Doğan, Deniz Şanlı, Ceren Tüzmen, Beytullah Özgür, Mehmet Ali Öztürk, Ece Bulut** and the countless other individuals I forgot to mention.

I would like to thank the Koc University and the Vehbi Koç Foundation for their financial support during my doctoral studies.

Last but not least, I am indebted to **my family**. It would have been next to impossible to write this thesis without their love, help and guidance.

TABLE OF CONTENTS

LIST OF FIGURES x

Chapter 1 1

INTRODUCTION 1

Chapter 2 5

HARMONIC FLUCTUATIONS OF TWO PEPTIDES 5

2.1. Introduction 5

2.2. Model and Theory 7

2.2.1. The Harmonic Model 7

2.2.2. Deviation from Harmonicity 15

2.3. Result and Discussions 16

2.3.1. Molecular Dynamics 16

2.3.2. Force Constants Among Two Peptides 17

2.3.3. Binding Energies in the Harmonic Approximation 23

2.3.4. Deviation from Harmonicity 26

2.4. Conclusion 27

Chapter 3 30

QUASI-HARMONIC ANALYSIS OF MODE COUPLING IN FLUCTUATING NATIVE PROTEINS 30

3.1. Introduction 30

3.2. The Model and Simulations 32

3.2.1. Molecular Dynamics Simulations 32

3.2.2. Fluctuations, Correlations and Principal Component Decomposition 33

3.2.3. Tensorial Hermite Series Approximation and Thermodynamic Analysis 35

3.3. The Lowest Order Coupling of Two Modes: 41

3.4. Results 45

3.4.1. Marginal Energies of Modes as a Function of Time and Mode Index 46

3.4.2. Third Order Moments and coupling of Modal Coordinates 51
3.4.3. Time Averaged Third Order Coupling $\langle \Phi_{ij} \rangle$ 54
3.4.4. Third Order Coupling Φ_{ij} as a Function of Time 57
3.4.5. Coupling of Conformations 60
3.5. Discussion 62
3.6. Supplementary Material 63
3.6.1. Coupling of Energies from $\langle \Phi_{ij} \rangle$ Values 66
Chapter 4 71
ANHARMONICITY, MODE-COUPLING AND ENTROPY IN FLUCTUATING NATIVE PROTEIN 71
4.1. Introduction 71
4.2. Modal Expansion and Beyond 71
4.2.1. Hermite Expansion 72
4.2.2. Anharmonicity, Mode-coupling and Entropy in a Fluctuating Native Protein 72
4.3. Anharmonicity vs Mode-coupling 75
4.4. Results 77
4.4.1. Crambin Molecular Dynamics: a Test Ground 77
4.4.2. Entropy Estimation 79
4.4.3. Mode-Coupling Corrections 80
4.4.4. Higher-Order Coupling 80
4.5. Discussion 81
4.6. Conclusion 82
Chapter 5 84
INVESTIGATING THE EFFECT OF THE DIFFERENCE IN THE UNBINDING FREE ENERGY PROFILES OF HLA-B51 AND HLA-B52 ON BEHÇET'S DISEASE 84
5.1. Introduction 84
5.2. Methods 86

5.2.1. Theory 86
5.2.2. MD Simulations 89
5.3. Result and Discussion 90
5.3.1. Change in Dynamics 90
5.3.2. SMD Simulations 96
5.4. Conclusion 103

APPENDIX 104

BIBLIOGRAPHY 106

VITA 113

LIST OF FIGURES

Figure 2-1. Dot product matrix $\left\langle \Delta \boldsymbol{R} \cdot \Delta \boldsymbol{R}^T \right\rangle$. White indicates negative correlations whereas black indicate positive correlations. 18

Figure 2-2: (a) Γ matrix of the bound form. (b) Part $1-2$ of the Γ matrix in the bound form. Black color indicates negative elements (Attractive) whereas white indicates positive elements (Repulsive). 22

Figure 2-3. $\dfrac{\left\langle \Delta \boldsymbol{r}_i^{\,4} \right\rangle}{\left(\left\langle \Delta \boldsymbol{r}_i^{\,2} \right\rangle \right)^2}$ values evaluated using the molecular dynamic trajectory are shown with the solid line and the theoretical values are shown with the dotted line. 27

Figure 3-1. The normalized histograms of the modal coordinates $\Delta \boldsymbol{r}$ for the first 12 slow modes. Filled points are the calculated values and the lines through them are evaluated using equation (14), up to the 17th order terms. 35

Figure 3-2. The RMSD values at 310 K (Black solid line) and at 273.15 K (Grey solid line) 45

Figure 3-3. Energy $\left\langle E_i \right\rangle$ of each mode at 273.15 K relative to the harmonic energy $kT/2$ per mode. Mode index corresponds to an increasing frequency order. Mode 1 is the slowest mode whereas mode 1965 is the fastest mode. 47

Figure 3-4. Energy $\left\langle E_i \right\rangle$ of each mode at 273.15 K relative to the harmonic energy $kT/2$ per mode at different stretches of the trajectory. (a)Stretch 1.25 - 4.25ns, (b) Stretch 5.75-11.25ns, (c)Stretch 8.25 - 11.5ns and (d) Stretch 32ns - 38ns. 49

Figure 3-5. Third order moment $\Delta \mathbf{r}_i^{\,3}$ of mode 310 and 445 at 273.15 K 50

Figure 3-6. The scatter diagram of the largest 500 $\left\langle \Delta \boldsymbol{r}_i \Delta \boldsymbol{r}_j^{\,2} \right\rangle$ terms at 237.15 K. There are of 2346 points due to the multiple presence of one type of coupling. 51

Figure 3-7. Distribution of $\left\langle \Delta r_i \Delta r_j \Delta r_k \right\rangle$ terms at 237.15 K. Rank goes from 1 to 140520. 52

Figure 3-8. Third order coupling terms $\left\langle \Delta \boldsymbol{r}_1^{\,2} \Delta \boldsymbol{r}_j \right\rangle$ and $\left\langle \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_j^{\,2} \right\rangle$ of the first mode to other modes. 53

Figure 3-9. $\left\langle \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_2 \Delta \boldsymbol{r}_k \right\rangle / \left\langle \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_2^{\,2} \right\rangle$ as a function of mode index k. 54

Figure 3-10. (a) Time averaged coupling of energy values $\langle \Phi_{1j} \rangle$ among the first mode and all other modes j (b) among mode 148 and all other modes j, $\langle \Phi_{148\,j} \rangle$ (c) among mode 310 and all other modes j, $\langle \Phi_{310\,j} \rangle$ 56
Figure 3-11. Cumulative coupling of mode i to all other modes. 57
Figure 3-12. (a) Coupling term among mode 1 and 2 (b) Coupling term among mode 2 and 31 (c) Coupling term among mode 148 and 310, all as a function of time. 59
Figure 3-13. (a) Contour plot of $\mathbf{e}_{310}\mathbf{e}_{310}^T$ (b) Contour plot of $\mathbf{e}_{148}\mathbf{e}_{148}^T$ (c) Contour plot of $\mathbf{e}_{148}\mathbf{e}_{310}^T$. 61
Figure 3-14. The normalized histograms of the modal coordinates $\Delta \boldsymbol{r}$ for the first 12 slow modes. Filled points are the calculated values and the lines through them are evaluated using equation (14). 64
Figure 3-15. The scatter diagram of the largest 500 $\langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j \Delta \mathbf{r}_k \rangle$ terms at 310 K. There are of 2332 points due to the multiple presence of one type of coupling. 65
Figure 3-16. Distribution of $\langle \Delta r_i \Delta r_j \Delta r_k \rangle$ terms at 310 K. Rank goes from 1 to 140520. 65
Figure 3-17. Third order coupling terms $\langle \Delta \mathbf{r}_3{}^2 \Delta \mathbf{r}_j \rangle$ and $\langle \Delta \mathbf{r}_3 \Delta \mathbf{r}_j{}^2 \rangle$ among the third mode and all other modes. 66
Figure 3-18. (a) Mean Coupling of energy values $\langle \Phi_{3,j} \rangle$ among the first mode and all other mode j (b) Mean Coupling of energy values $\langle \Phi_{608,j} \rangle$ among the 608$^{\text{th}}$ mode and all other mode j. 67
Figure 3-19. Cumulative coupling of mode i to all other modes. 68
Figure 3-20. (a) Coupling term among mode 3 and 6 as a function of time (b) Coupling term among mode 3 and 608 as a function of time. 69
Figure 3-21. Energy $\langle E_i \rangle$ of each mode at 310 K. 70
Figure 4-1. Graphical representation of $H_4(x)$ tensor elements in two dimensions. 74
Figure 4-2. Time plots of the slowest 1$^{\text{st}}$, 5t$^{\text{h}}$, 10$^{\text{th}}$, and 50$^{\text{th}}$ modes between timesteps 8000-8500. 76
Figure 4-3. A comparison of W_0 and W_1 and W_2 on the normalized histogram of the slowest mode. Note that W_1 and W_2 give the same marginal mode probabilities. 78

Figure 4-4. A comparison of W_1, W_2 and KDE against a scatter plot of the two slowest modes. 80

Figure 5-1. $\left\langle \Delta R_i^2 \right\rangle^{1/2}$ values of the carbon alpha atom of the ligand for the 15ns trajectories where complex is aligned with respect to the ligand (a) and complex is aligned with respect to the chain A of the protein (b). Black dots indicate the fluctuation value for HLA-B52 whereas grey dots indicate the fluctuation values for HLA-B51. 93

Figure 5-2. $\sum_{j=60}^{90} \left\langle \Delta R_{ij}^2 \right\rangle$ values of the carbon alpha atoms of helix-1 (a) $\sum_{j=60}^{90} \left\langle \Delta R_i \Delta R_j \right\rangle$ values of the carbon alpha atoms of helix-1 (b). Black dots indicate overal strength of the correlation values for HLA-B52 whereas grey dots indicate the overal strength of the correlation values values for HLA-B51. 96

Figure 5-3. (a) Reaction coordinates vs force for the 11. SMD simulation (Starting structure selected at 20.892 ns of the conventional molecular dynamics simulation) of HLA-B52 (shown with black line) and for the 11. SMD simulation (Starting structure selected at 13.798ns of the conventional molecular dynamics simulation) simulation of HLA-B51 (Shown with grey line). (b) Reaction coordinates vs average force of all SMD simulation for HLA-B52 (shown with the black solid line) and HLA-B51 (Shown with the grey line). 100

Figure 5-4. PMF of HLA-B52 (Solid black line) and HLA-B51 (Solid grey line) with respect to the reaction coordinate ξ 101

Figure 5-5. Difference between the minimum and maximum work values at each reaction coordinate ξ 102

Chapter 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[1]. In theory, fluctuations are studied at various levels of approximation, ranging from all-atom to coarse-grained scales. Studying the fluctuations of the C_α is a convenient approximation where each successive C_α pair is assumed to be connected by a virtual bond of fixed length and only interactions between residues, represented by their C_α's, are considered. In the present study, both this level of approximation and all atomic level approximation are adopted. 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, 9-11]. 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 addition to work in harmonic fluctuations cited here, anharmonicities of protein fluctuations [12, 13] in the form of nonlinear modes that are localized in certain regions of the protein play important roles in protein function [14, 15]. In this respect, coupling of fast and slow modes resulting in energy flow is the most important process responsible for the protein's function [16, 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 the 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.

In this section, the thermodynamic and statistical basis of fluctuations in native proteins will be given. This basis will be enlarged and applied to physical problems in the following section.

We use the entropy representation for the fundamental relation [18],

$$S = S(U, V, \boldsymbol{R}) \qquad 1\text{-}1$$

Where $S, U, V, \boldsymbol{R}$ are the mean (thermodynamic) values of the entropy, energy, volume, and position vectors of the atoms. Water is not shown explicitly in the fundamental relation. The protein is in diathermal contact with the surrounding water. Similarly, the protein is in contact with a pressure (P) and a force ($\boldsymbol{F}$) reservoir, as a result of which the energy, volume and the positions of residues exhibit fluctuations. Other, not bound proteins, are present in the surroundings but they do not influence the energy levels of the given protein. We call the protein and the surrounding water as an element. The collection of all elements of the system constitutes the ensemble. Statistical mechanics is applicable to a single element. Thermodynamics applies only to an ensemble of the elements. The ensemble of elements with its extensive properties constitutes a macroscopic system [18, 19]. The thermodynamic

variables *S, U, V, **R*** are obtained from the ensemble. For each element, these variables exhibit fluctuations about their native values. The distribution $f\left(\hat{U},\hat{V},\hat{\boldsymbol{R}}\right)$ of the instantaneous extensive variables $\hat{U},\hat{V},\hat{\boldsymbol{R}}$ are given by the relation,

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

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 the Massieu transform of the entropy, which for the specified thermodynamic variables chosen reads as

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

The distribution now takes the explicit form

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

In equation (4), provided that the system remains around the given equilibrium point, i.e., a point on the thermodynamic surface $S=S(U,V,\boldsymbol{R})$, there are no restrictions on the degree of departure of the system, i.e., the magnitude of fluctuations, from the average thermodynamic variables. If the fluctuations are large, the fluctuations may be anharmonic or may induce a jump from one local minimum to another. The applicability of results derived from equation (4) are discussed in detail in the following sections.

The correlation of fluctuations of the ith and jth residues may now be obtained from

$$\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j{}^T \right\rangle = \sum \left(\hat{\boldsymbol{R}}_i - \boldsymbol{R}_i \right) \left(\hat{\boldsymbol{R}}_j - \boldsymbol{R}_j \right)^T \; f\left(\hat{U}, \hat{V}, \hat{\boldsymbol{R}} \right) \qquad 1\text{-}5$$

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

Each of the following chapters is in form of a separate paper. For each of them separate introductions, problem statements and conclusions are provided. This first general introduction chapter aimed to provide the fundamental statistical mechanic background required to understand the following chapters.

Chapter 2

HARMONIC FLUCTUATIONS OF TWO PEPTIDES

2.1. Introduction

Understanding the binding of two molecules is a complex problem that may suitably be simulated by molecular dynamics. In the present study, we use long molecular dynamics trajectories of two bound peptides to extract information from the system. In the interest of simplicity, we adopt a quasiharmonic analysis where we assume that the atoms of the two peptides are connected by linear springs, and the spring constants are obtained from the correlations of fluctuations of the atoms, which are in turn are obtained from the molecular dynamics trajectory.

The idea of obtaining spring constants from fluctuation correlation is not new and was employed by Karplus and collaborators [20, 21] and Lamm and Szabo [22]. The most transparent use of the idea which we adopt in the present work is by Teeter and Case [23].

REACH (Realistic Extension Algorithm via Covariance Hessian) is an elastic network model (ENM) developed by Moristugu and Smith[24], in which the residue interaction spring constants are obtained directly from the atomic-detail variance-covariance matrix calculated using MD simulation [24]. In this way physically-based atomic MD force fields can be projected onto inter-residue spring constants. The REACH spring constants were derived by relating the harmonic-approximated potential energy of the ENM to the Hessian (second-derivative) matrix and then to the variance-covariance matrix[24]. The interactions were divided into 4 classes; virtual 1-2 (between residue i and i+1), virtual 1-3 (between residue i and i+2),virtual 1-4 (between residue i and i+3) and the nonbonded interactions. In REACH, the residue-scale Hessian matrix is calculated using the variance–covariance matrix from the atomistic MD trajectory. Making the harmonic approximation under the equilibrium condition

at constant temperature, T , allows the Hessian matrix to be calculated from the 3N-dimensional variance–covariance matrix [25].

In their work Moristugu and Smith [23] have pointed out that the anharmonicity in atomistic MD simulations may lead to the spring constants deviating from the "ideal" harmonic approximation quantities, but the contribution will likely be small, i.e., this is not the main origin of negative spring constants. Moristugu and Smith [23] have used each 1-ns MD trajectories to calculate the variance-covariance matrix and justified their methodology by pointing out that the time length of 1 ns is long enough to characterize the vibrational component of protein fluctuation, which arises from the harmonic potential, but is not so long that the intramolecular contribution is small compared to the slow, diffusive motion. Since normal modes represent vibrational motions on an effective harmonic potential, the MD time length of 1 ns was concluded to be a suitable choice. Therefore, the MD trajectories were separated into 1-ns long trajectories from each of which the covariance matrix was calculated and then these matrices were averaged in order derive associated spring constants [26] .

On ligand binding, protein dynamics is changed by two effects: the change in conformation of the protein from the unbound to the bound state and the interaction between the protein and the ligand [25]. Moritsugu et al.[25] have focused on the effect of the force field on the vibrational dynamics with a constant protein structure. They have shown that the internal degrees of freedom increases by 6 upon ligand binding which are considered as ligand external motions coupling to protein vibrations.

In their study they have also pointed out that protein vibrations becomes stiffer on ligand binding and that an increase in vibrational entropy on ligand binding arises from the additional six degrees of freedoms. However, with stronger interaction energy, increased entropy on ligand binding would not be observed [25].

The spring constant matrix K is obtained from the expression $K = \left[kT \left\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \right\rangle \right]^{-1}$ where the correlation matrix $\left\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \right\rangle$ of fluctuations of atom positions is obtained from the molecular dynamics trajectory. If, instead of the full trajectory, a specific conformation at a minimum of the energy, U , is found, the more familiar form of the spring constant matrix is obtained from

the Hessian of the system and corresponds to the harmonic or the normal mode analysis. Molecular dynamics show that in biological systems such as two bound peptides in water the system is strongly anharmonic, with frequent transitions from one conformation to another. The next level of approximation over the harmonic is the quasiharmonic that we adopt here. The probability distribution of fluctuations are Gaussian, $f(\Delta \boldsymbol{R}) \propto e^{-\frac{1}{2kT}\Delta \boldsymbol{R}^T \mathrm{K} \Delta \boldsymbol{R}}$, as in the harmonic approximation, but the spring constant matrix K is the fluctuation correlation matrix as discussed above. The name 'quasiharmonic' derives from this difference from the harmonic. The deviation of the full probability distribution function from the quasiharmonic is given in recent work [27, 28].

In this chapter, we adopt a coarse grained model based on the alpha carbon representation. The spring constant matrix of peptide residues yields important information on the binding process and can easily be extended to protein pairs irrespective of their sizes. There are several methods and softwares for determining the protein-protein or protein-ligand binding problem. The stability of binding and the change in the Helmholtz binding energy can be determined by several computational techniques [29-34]. The molecular dynamics scheme gives a realistic estimate of energy and entropy changes in binding. The harmonic and the quasiharmonic approaches are only approximations aimed at (i) simplifying the problem, and (ii) arriving at an analytical treatment. Below, we discuss the various aspects of this approximation, obtain the spring constant matrix for two bound peptides and elaborate on the validity of the approach.

2.2. Model and Theory

2.2.1. The Harmonic Model

The model consists of two interacting native proteins having n_1 and n_2 residues in a thermal reservoir. The coarse grained model is adopted in which the instantaneous positions of the i^{th} alpha carbon of peptide ξ, $(\xi = 1,2)$ is represented by $\hat{R}_i^{(\xi)}$. The mean positions $R_i^{(\xi)} = \left\{ X_i^{(\xi)}, Y_i^{(\xi)}, Z_i^{(\xi)} \right\}$ are assumed fixed, and only the fluctuations

$$\Delta R_i^{(\xi)} = \hat{R}_i^{(\xi)} - R_i^{(\xi)} \qquad 2\text{-}1$$

are of consequence.

The correlation of the equilibrium fluctuations of the ith atom of peptide 1 and the jth atom of peptide 2 is given by the general second moments of the form introduced by equation (5),

$$\left\langle \Delta R_i^{(1)} \Delta R^T{}_j^{(2)} \right\rangle = \sum (\hat{R}_i^{(1)} - R_i^{(1)})(\hat{R}_j^{(2)} - R_j^{(2)})^T f(\hat{\boldsymbol{R}}) \qquad 2\text{-}2$$

where, the superscript T denotes the transpose, and

$$\left\langle \Delta R_i \Delta R_j^T \right\rangle = \begin{bmatrix} \Delta X_i \Delta X_j^T & \Delta X_i \Delta Y_j^T & \Delta X_i \Delta Z_j^T \\ \Delta Y_i \Delta X_j^T & \Delta Y_i \Delta Y_j^T & \Delta Y_i \Delta Z_j^T \\ \Delta Z_i \Delta X_j^T & \Delta Z_i \Delta Y_j^T & \Delta Z_i \Delta Z_j^T \end{bmatrix} \qquad 2\text{-}3$$

The position vector of the ξ^{th} peptide can be defined as

$\hat{R}^{(\xi)} = \left\{ \hat{X}_1^{(\xi)}, \hat{X}_2^{(\xi)}, ..., \hat{X}_{n_1}^{(\xi)}, \hat{Y}_1^{(\xi)}, \hat{Y}_2^{(\xi)}, ..., \hat{Y}_{n_1}^{(\xi)}, \hat{Z}_1^{(\xi)}, \hat{Z}_2^{(\xi)}, ..., \hat{Z}_{n_1}^{(\xi)} \right\}$. Here $X_i^{(\xi)}$ is the mean x position of the ith atom of peptide ξ, with similar definitions holding for the remaining coordinates. Hence, the position vector of both peptides together is defined as $\hat{\boldsymbol{R}} = \begin{bmatrix} \hat{R}^{(1)} \\ \hat{R}^{(2)} \end{bmatrix}$ and in general $\boldsymbol{R} = \begin{bmatrix} R^{(1)} \\ R^{(2)} \end{bmatrix}$. In the remaining sections of the paper, we suspend the superscript Greek letters identifying the peptides, and use the general vector representation for F and R

containing the information for both of the peptides as defined above. Equation (2) can be rewritten in this notation as,

$$\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T} \right\rangle = \sum (\hat{\boldsymbol{R}}_i - \boldsymbol{R}_i)(\hat{\boldsymbol{R}}_j - \boldsymbol{R}_j)^T f(\hat{\boldsymbol{R}}) \qquad 2\text{-}4$$

where the summation is over all allowable states. $f(\hat{\boldsymbol{R}})$ is the probability, that the system has coordinates $\hat{\boldsymbol{R}}$ in contact with the thermal reservoir. The instantaneous variables $\hat{\boldsymbol{R}}$ define the microstates of the complex. More generally, the probability, $f\left(\hat{U},\hat{V},\hat{\boldsymbol{R}}\right)$ of the complex in a thermal and pressure reservoir is given by

$$f\left(\hat{U},\hat{V},\hat{\boldsymbol{R}}\right) = \exp\left\{-k^{-1}S\left[\frac{1}{T},\frac{P}{T},\frac{\boldsymbol{F}}{T}\right] - k^{-1}\left(\frac{1}{T}\hat{U} + \frac{P}{T}\hat{V} - \frac{\boldsymbol{F}}{T}\hat{\boldsymbol{R}}\right)\right\} \qquad 2\text{-}5$$

Here $S\left[\frac{1}{T},\frac{P}{T},\frac{\boldsymbol{F}}{T}\right]$ is the Massieu transform of the entropy and $\frac{1}{T},\frac{P}{T},\frac{\boldsymbol{F}}{T}$ are the entropic intensive parameters whose values are equal to those of the reservoir.

Because of the special form of $f\left(\hat{U},\hat{V},\hat{\boldsymbol{R}}\right)$, equation (4) can be reformulated as

$$\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T} \right\rangle = kT \sum (\hat{\boldsymbol{R}}_i - \boldsymbol{R}_i) \frac{\partial f(\hat{U},\hat{V},\hat{\boldsymbol{R}})}{\partial \boldsymbol{F}_j} \qquad 2\text{-}6$$

where $\boldsymbol{F} = \begin{bmatrix} F^{(1)} \\ F^{(2)} \end{bmatrix}$ and $F^{(\xi)} = \begin{bmatrix} F_x^{(\xi)} \\ F_y^{(\xi)} \\ F_z^{(\xi)} \end{bmatrix}$, $\left(\xi = 1,2\right)$. $F_x^{(\xi)}$ is the force exerted in the x direction on the ξ^{th} peptide. Equation (11) simplifies to

$$\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j{}^T \right\rangle = kT \left(\frac{\partial \boldsymbol{R}_i}{\partial \boldsymbol{F}_j} \right)_{\boldsymbol{T},P,\boldsymbol{F}_{i\neq j}} \qquad 2\text{-}7$$

where the variables to be kept fixed are indicated as subscripts. The variables on the right hand side of equation (7) are the thermodynamic, i.e., average quantities. The reader is referred to Callen [18] for details of the derivation above.

Definition of the mean positions of the C_α's and the harmonic assumption allows us to write the force-deformation relation for the complex as,

$$\boldsymbol{F} = \mathrm{K}\left[\hat{\boldsymbol{R}} \text{-} \boldsymbol{R}^o \right] \qquad 2\text{-}8$$

K can be considered as the symmetric matrix having the harmonic springs as the elements. Hence, the bracket in equation (8) implies that springs are relaxed at the equilibrium positions $\boldsymbol{R}^o$ of the atoms and the energy is zero at those positions.

The spring K in equation (8), written in terms of the coordinates of the C^α 's has a special form

$$\mathrm{K}_{ij} = \begin{cases} -\kappa_{ij} & \text{if } i \neq j \\ \kappa_{ii} = \sum_{j\neq i} \kappa_{ij} \end{cases} \qquad 2\text{-}9$$

Applying equation (8) on equation (7) leads to the correlations of fluctuations

$$\left\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \right\rangle = kT\mathrm{K}^{-1} \qquad 2\text{-}10$$

The energy $\hat{U}(\boldsymbol{\Delta R})$ of the microstate of the complex corresponding to the set $\{\boldsymbol{\Delta R}\}$ follows from the harmonic assumption as

$$\hat{U}(\boldsymbol{\Delta R}) = \frac{1}{2}\boldsymbol{\Delta R}^T \mathrm{K} \boldsymbol{\Delta R} \qquad 2\text{-}11$$

The energy in each interaction on the other hand can be evaluated as

$$E_{i,j} = \frac{1}{2}\mathrm{K}_{ij}\left(\Delta \boldsymbol{R}_i - \Delta \boldsymbol{R}_j\right)^2 \qquad 2\text{-}12$$

Using the energy definition of the microstate $\{\boldsymbol{\Delta R}\}$ provided in Eq.11, the configurational probability density in the Cartesian coordinates is given by

$$f(\boldsymbol{\Delta R}) = \frac{e^{-\frac{1}{2kT}\boldsymbol{\Delta R}^T \mathrm{K} \boldsymbol{\Delta R}}}{\int_{\{\boldsymbol{\Delta R}\}} e^{-\frac{1}{2kT}\boldsymbol{\Delta R}^T \mathrm{K} \boldsymbol{\Delta R}} d\{\boldsymbol{\Delta R}\}} \qquad 2\text{-}13$$

Using equation (10), equation (13) can be reformulated as

$$f(\Delta \boldsymbol{R}) = \frac{e^{-\frac{1}{2}\Delta \boldsymbol{R}^T \left\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \right\rangle^{-1} \Delta \boldsymbol{R}}}{\int_{\{\Delta R\}} e^{-\frac{1}{2}\Delta \boldsymbol{R}^T \left\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \right\rangle^{-1} \Delta \boldsymbol{R}} d\{\Delta \boldsymbol{R}\}} \qquad 2\text{-}14$$

Fluctuation $\Delta \boldsymbol{R}$ in the Cartesian space can be expressed in terms of the fluctuation in mode space as

$$\Delta \boldsymbol{r} = V^T \Delta \boldsymbol{R} \qquad 2\text{-}15$$

Where V is the eigenvector matrix that diagonalizes the covariance matrix $\left\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \right\rangle$

$$\left\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \right\rangle = V \Sigma V^T \qquad 2\text{-}16$$

Using equation (10) the force constant matrix can be formulated as

$$\mathrm{K} = V \Lambda V^T \qquad 2\text{-}17$$

Here Λ is the eigenvalue matrix of K which equals to $\Lambda = kT\Sigma^{-1}$. Since all translational and rotational degrees of freedom are eliminated, $\Delta \boldsymbol{r}$ will consist of 3n-6 nonzero modes; i.e. the system consists of 3n-6 independent internal degrees of freedom. The energy $\hat{U}(\Delta \boldsymbol{R})$ of the microstate of the complex corresponding to the set $\{\Delta \boldsymbol{R}\}$ can be defined in terms of the set $\{\Delta \boldsymbol{r}\}$

$$\hat{U}(\Delta \boldsymbol{R}) = \frac{1}{2} \Delta \boldsymbol{R}^T \mathrm{K} \Delta \boldsymbol{R} = \frac{1}{2} \Delta \boldsymbol{R}^T V \Lambda V^T \Delta \boldsymbol{R} = \frac{1}{2} \Delta \boldsymbol{r}^T \Lambda \Delta \boldsymbol{r} = \frac{1}{2} \sum_{i=1}^{3n-6} \lambda_i \Delta \boldsymbol{r}_i^2 \qquad 2\text{-}18$$

Where $\Delta \boldsymbol{r}_i$ is the modal coordinate of the ith mode and λ_i is the ith eigenvalue of K. Taking the ensemble average of equation (18) and recognizing that $\lambda_i \left\langle \Delta \boldsymbol{r}_i^2 \right\rangle$ equals to kT for every i, the average energy U can be written as

$$U = \frac{1}{2} \sum_{i=1}^{3n-6} kT = \frac{(3n-6)}{2} kT \qquad 2\text{-}19$$

This is the energy of a harmonic system consisting of n particles.

The configurational probability density in modal space can be written for the 3n-6 internal modes,

$$f\left(\Delta \boldsymbol{r}\right) = \frac{e^{-\frac{1}{2kT}\Delta \boldsymbol{r}^T \Lambda \Delta \boldsymbol{r}}}{\int\limits_{\{\Delta \boldsymbol{r}\}} e^{-\frac{1}{2kT}\Delta \boldsymbol{r}^T \Lambda \Delta \boldsymbol{r}} d\left\{\Delta \boldsymbol{r}\right\}} \qquad 2\text{-}20$$

Since all 3n-6 nonzero modes $\Delta \boldsymbol{r}_i$ are independent and harmonic they satisfy a multivariate Gaussian distribution. The denominator of equation (14) is the configurational integral part Z_n of the vibrational partition function. Hence, to evaluate Z_n the denominator of equation (20) is integrated out over all allowable states $\{\Delta r\}$.

$$Z_n = \int\limits_{\{\Delta \boldsymbol{r}\}} e^{-\frac{1}{2kT}\Delta \boldsymbol{r}^T \Lambda \Delta \boldsymbol{r}} d\left\{\Delta \boldsymbol{r}\right\} = \left(2\pi kT\right)^{(3n-6)/2} \left[\det\left(\Lambda^{-1}\right)\right]^{1/2} = \left(2\pi\right)^{(3n-6)/2} \prod_{i=1}^{3n-6} \left(\frac{1}{\lambda_i}\right)^{1/2} \qquad 2\text{-}21$$

where λ_i are the 3n-6 internal eigenvalues of the K matrix. Since free energy makes sense only with respect to reference point, we will assume an arbitrary state whose partition function is in the form of $Z_o = \varepsilon^n$. ε is selected so that entropy won't attain negative values. The excess Helmholtz free energy then follows as

$$A = -kT \ln\left(\frac{Z_n}{Z_o}\right) \qquad 2\text{-}22$$

Where Z_o denotes the partition function of the reference state. Substituting the definition obtained for Z_n in equation (21) into equation (22) we end up with the following result.

$$\begin{aligned} A &= -\frac{3}{2}kT\sum_{1}^{n-6}\left(1-\ln\left(\lambda_i^{-1}\pi e 2kT\right)\right) - kT\ln Z_o \\ &= \frac{3n-6}{2}kT - kT\left(\frac{3}{2}\sum_{1}^{n-6}\ln\left(\lambda_i^{-1}\pi e 2kT\right) + \ln Z_o\right) \end{aligned} \qquad 2\text{-}23$$

Recalling the form of the Helmholtz free energy as

$$A = U - TS \qquad 2\text{-}24$$

And using equation (19) and equation (22), the entropy can be written as

$$\frac{S}{k} = \frac{3}{2}\sum_{1}^{n-6}\ln\left(\lambda_i^{-1}\pi e 2kT\right) + \ln Z_o \qquad 2\text{-}25$$

2.2.2. Deviation from Harmonicity

The even moments $\left\langle x^{2n} \right\rangle$ of the variable x in a Gaussian distribution are given as

$$\left\langle x^{2n} \right\rangle = \frac{\int_{-\infty}^{\infty} x^{2n} \exp(-px^2)dx}{\int_{-\infty}^{\infty} \exp(-px^2)dx} \qquad 2\text{-}26$$

Where the integrals can be evaluated using the following property

$$\int_0^{\infty} x^{2n} \exp(-px^2)dx = \frac{(2n-1)!!}{2\cdot(2p)^n}\sqrt{\frac{\pi}{p}} = \frac{(2n)!}{2\cdot(2p)^n}\sqrt{\frac{\pi}{p}} \qquad 2\text{-}27$$

Hence defining x to be the normalized modal coordinate of the ith mode $x = \Delta \boldsymbol{r}_i$ we end up with the following equations

$$\left\langle \Delta \boldsymbol{r}_i^4 \right\rangle = 2\int_0^{\infty} \Delta \boldsymbol{r}_i^4 \frac{\exp(-0.5\Delta \boldsymbol{r}_i^2)}{Z_n} d\Delta \boldsymbol{r}_i = \frac{1}{Z_n}\int_0^{\infty} \Delta \boldsymbol{r}_i^4 \exp(-p\Delta \boldsymbol{r}_i^2)d\Delta \boldsymbol{r}_i = \frac{3\sqrt{\pi}}{4Z_n\sqrt{0.5^5}}$$

$$\left\langle \Delta \boldsymbol{r}_i^2 \right\rangle = 2\int_0^{\infty} \Delta \boldsymbol{r}_i^2 \frac{\exp(-0.5\Delta \boldsymbol{r}_i^2)}{Z_n} d\Delta \boldsymbol{r}_i = \frac{1}{Z_n}\int_0^{\infty} \Delta \boldsymbol{r}_i^2 \exp(-p\Delta \boldsymbol{r}_i^2)d\Delta \boldsymbol{r}_i = \frac{1\sqrt{\pi}}{2Z_n\sqrt{0.5^3}} \qquad 2\text{-}28$$

Hence the dimensionless correlation among them equals to 3 under harmonic approximation

$$\frac{\left\langle \Delta \boldsymbol{r}_i^{\,4} \right\rangle}{\left(\left\langle \Delta \boldsymbol{r}_i^{\,2} \right\rangle \right)^2} = \frac{\dfrac{3\sqrt{\pi}}{4Z_i\sqrt{0.5^5}}}{\left(\dfrac{1\sqrt{\pi}}{2Z_i\sqrt{0.5^3}} \right)^2} = \frac{3Z_i\sqrt{0.5}}{\sqrt{\pi}} = \frac{3\sqrt{2\pi}\sqrt{0.5}}{\sqrt{\pi}} = 3 \qquad 2\text{-}29$$

2.3. Result and Discussions

2.3.1. Molecular Dynamics

Here the methods of the preceding section are applied to the complex structure of two distinct peptides. In HyperChem7, peptides of sequences ASN1-ASP2-MET3-PHE4-ARG5-LEU6 and LEU7-LEU8-PHE9-MET10-GLN11-HIS12 were constructed. Structures were separately and in complex form geometrically optimized via Polack-Ribiere algorithm with OPLS force field in Hyperchem. These geometrically optimized structures were used as the starting structures prior to the minimization-equilibration cycles in the molecular dynamic simulations.

Simulations were performed in explicit solvent (water) using NAMD 2.6 package with CHARMM27 force field [35]. All simulations were performed at constant temperature (310 K) and pressure (1.01325 bar) in a periodic water box with a 20-Å cushion. Ions were added in order to represent a more typical biological environment. Nonbonded and electrostatic forces were evaluated each time step. In order to keep all degrees of freedom, no rigid bonds were used . To evaluate the non-bonded interaction, cut-off distance was set to 12 Å. The Particle Ewald sum was used as a way of calculating long range forces in the periodic systems. Therefore error introduced by truncation due to the cut-off distance was minimized. Integration time step was set to 1 fs and structure was recorded at 1000-step (1 ps in MD). Prior to MD calculations, all models were subjected to 3 minimization-equilibration cycles. The first ones were applied to relax the water in the first place and the last two ones were applied to find a local minimum of the whole systems energy. All minimization cycles included 20,000 energy minimization steps to relieve close intermolecular contacts and geometric strain. In order to relax the water, the proteins were kept fixed in the first minimization and the 0.1 ns long equilibration cycle. Then the structures were released stepwise by applying harmonic constraining forces to every backbone atom of 1, 0.5 and 0.25 kcal/(mol*Angstrom^2) in magnitude each for 0.05 ns. Finally an additional 0.05ns simulation

were performed without any constraining force. Following the first minimization-equilibration cycle two more minimization were performed, each separated by an 0.1ns equilibration phase. For chain A, calculations were performed using the trajectory part 57.6-61.1 ns after the final minimization-equilibration cycle. For chain B, calculations were performed using the trajectory part 52.8-56.3 ns after the final minimization-equilibration cycle. For the bound form on the other hand the trajectory part 27.5-31 ns after the final minimization-equilibration cycle was used for the calculations. In order to eliminate all the rotational and translational motions, all structures were aligned with respect to the first frame of the production phase using the transformation matrix which shows the best fit for to the C^{α} atoms with the first structure. All transformation matrices were constructed via tcl commands in VMD.

2.3.2. Force Constants Among Two Peptides

For each MD trajectory the 36x36 covariance matrix $\left\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \right\rangle$ was constructed using the recorded snapshots during the production phase. In order to show the fluctuations in a more concise manner the average of the dot products $\left\langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_j \right\rangle = \left\langle \{\Delta x_i, \Delta y_i, \Delta z_i\} \cdot \{\Delta x_j, \Delta y_j, \Delta z_j\} \right\rangle$ were evaluated and shown in figure 1. Part $1-1$ and part $2-2$ give the intramolecular part of the correlations of peptide 1 and peptide 2 respectively. Part $1-2$ and its symmetric mirror image, part $2-1$, provide the intermolecular correlation among the two peptides. White regions indicate the negative values whereas black regions indicate the positive values. Numeric values of the correlation matrix are provided in Appendix.1.

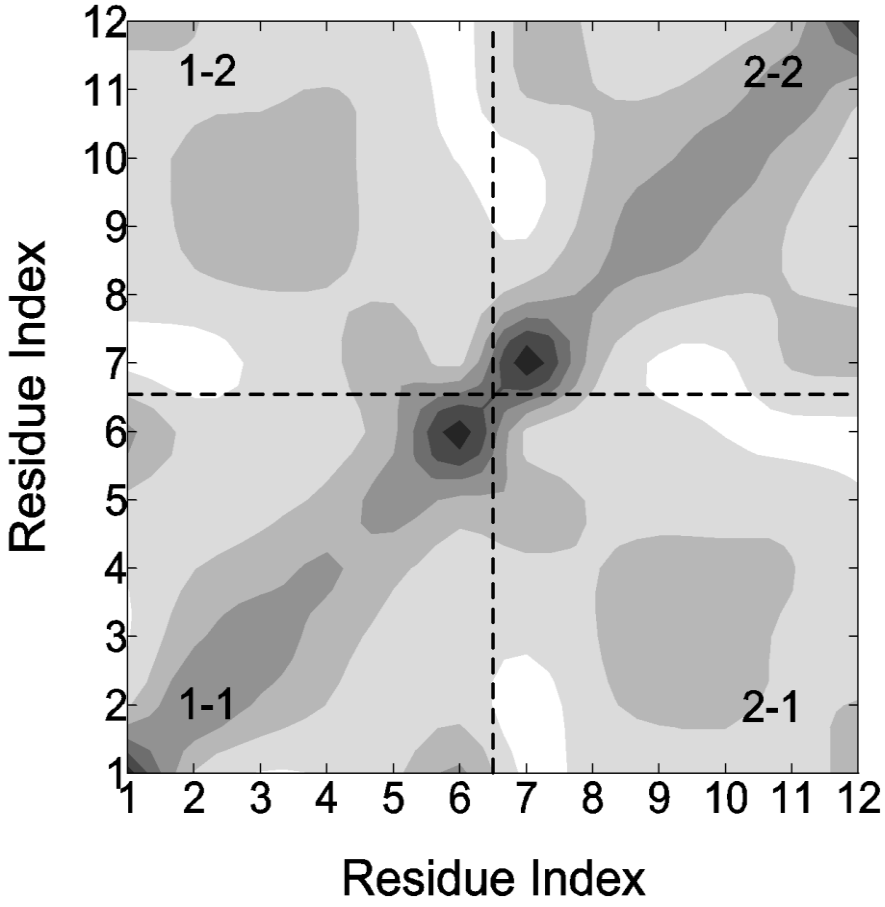


Figure 2-1. Dot product matrix $\left\langle \Delta \boldsymbol{R} \cdot \Delta \boldsymbol{R}^{T} \right\rangle$. White indicates negative correlations whereas black indicate positive correlations.

Using MATLAB's build-in eigenvalues decomposition algorithm the eigenvalues and eigenvectors of the correlation matrix were obtained. Excluding the six zero eigenvalues, the pseudo inverse was formed. By equation (10) the force constant matrix is constructed which has the following form,

$$K = \begin{bmatrix} K_{X,X}^{(1)} & K_{X,Y}^{(1)} & K_{X,Z}^{(1)} & K_{X,X}^{(1,2)} & K_{X,Y}^{(1,2)} & K_{X,Z}^{(1,2)} \\ K_{Y,X}^{(1)} & K_{Y,Y}^{(1)} & K_{Y,Z}^{(1)} & K_{Y,X}^{(1,2)} & K_{Y,Y}^{(1,2)} & K_{Y,Z}^{(1,2)} \\ K_{Z,Z}^{(1)} & K_{Z,Y}^{(1)} & K_{Z,Z}^{(1)} & K_{Z,X}^{(1,2)} & K_{Z,Y}^{(1,2)} & K_{Z,Z}^{(1,2)} \\ K_{X,X}^{(2,1)} & K_{X,Y}^{(2,1)} & K_{X,Z}^{(2,1)} & K_{X,X}^{(2,2)} & K_{X,Y}^{(2,2)} & K_{X,Z}^{(2,2)} \\ K_{Y,X}^{(2,1)} & K_{Y,Y}^{(2,1)} & K_{Y,Z}^{(2,1)} & K_{Y,X}^{(2,2)} & K_{Y,Y}^{(2,2)} & K_{Y,Z}^{(2,2)} \\ K_{Z,X}^{(2,1)} & K_{Z,Y}^{(2,1)} & K_{Z,Z}^{(2,1)} & K_{Z,X}^{(2,2)} & K_{Z,Y}^{(2,2)} & K_{Z,Z}^{(2,2)} \end{bmatrix} \qquad 2\text{-}30$$

As can be seen in equation (30) the force constant matrix is divided into 4 subsections. $K^{(1)}$ and $K^{(2)}$ give the intramolecular correlation of peptide 1 and peptide 2 respectively. $K^{(1,2)}$ and its symmetric mirror image, $K^{(2,1)}$, provide the intermolecular correlation among both peptides. Here $K_{X,X}^{(1,1)}$ for example indicates the force constant among the residues in peptide 1 in the x directions.

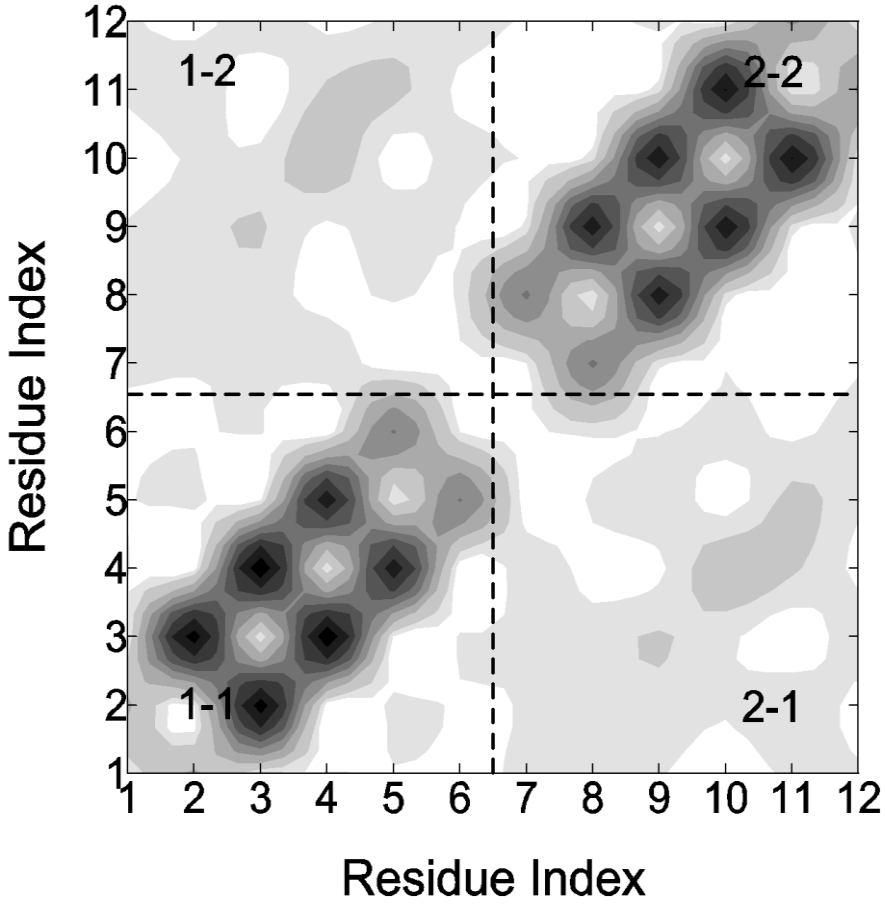
Again to be more concise, simple in presentation and to be in accord with [5, 6] a new force constants matrix is defined in terms of K as follows

$$\Gamma = \begin{bmatrix} K_{X,X}^{(1)} + K_{Y,Y}^{(1)} + K_{Z,Z}^{(1)} & K_{X,X}^{(1,2)} + K_{Y,Y}^{(1,2)} + K_{Z,Z}^{(1,2)} \\ K_{X,X}^{(2,1)} + K_{Y,Y}^{(2,1)} + K_{Z,Z}^{(2,1)} & K_{X,X}^{(2)} + K_{Y,Y}^{(2)} + K_{Z,Z}^{(2)} \end{bmatrix} = \begin{bmatrix} \Gamma^{(1)} & \Gamma^{(1,2)} \\ \Gamma^{(2,1)} & \Gamma^{(2)} \end{bmatrix} \qquad 2\text{-}31$$

The Γ matrix is divided into 4 subsections as it was the case for K The Γ matrix is numerically provided in Appendix 2 and shown in figure 2 (a). Since the diagonal terms are the sum of the off-diagonal terms in each row, they tend to be significantly larger in magnitude then the off-diagonal terms. Therefore, for clarity diagonal terms were set to zero in figure 2 (a) and only off-diagonal terms are shown As can be concluded from equation (9), the negative off-diagonal elements of Γ indicate positive force constants whereas positive

ones indicate negative force constants. Positive force constants κ_{ij} indicate that residue i and residue j tend to stay at their equilibrium distance, $\boldsymbol{norm}\left(\boldsymbol{R}^o{}_i - \boldsymbol{R}^o{}_j\right)$. Any deviation from this distance will cause an increase in energy as indicated in equation (11) which is unfavorable. The equilibrium coordinates were assumed to be the same as the mean coordinates of the atoms $\boldsymbol{R}$ obtained via MD. Negative force constants on the other hand indicate that any deviation from the mean coordinates results in a decrease in energy. Since an energy decrease is favorable, the meaning of a negative force constant is that residue i and j do not prefer to stay at these relative positions, $\boldsymbol{norm}\left(\boldsymbol{R}^o{}_i - \boldsymbol{R}^o{}_j\right)$. Therefore, for the rest of the paper positive force constants will be called attractive force constants and negative force constants will be called repulsive force constants. Hence, large negative off-diagonal elements of Γ show strong attractive interactions. As can be seen in figure 2 (a) attractive force constants for the interactions among residues of the same peptide (intramolecular) attain much larger values than the force constants for the interactions among residues of different peptides (intermolecular). To focus on the intermolecular interactions, part $\Gamma^{(1,2)}$ is shown separately in figure 2 (b).

.


1-2
2-2
1-1
2-1
Residue Index
Residue Index
1
2
3
4
5
6
7
8
9
10
11
12

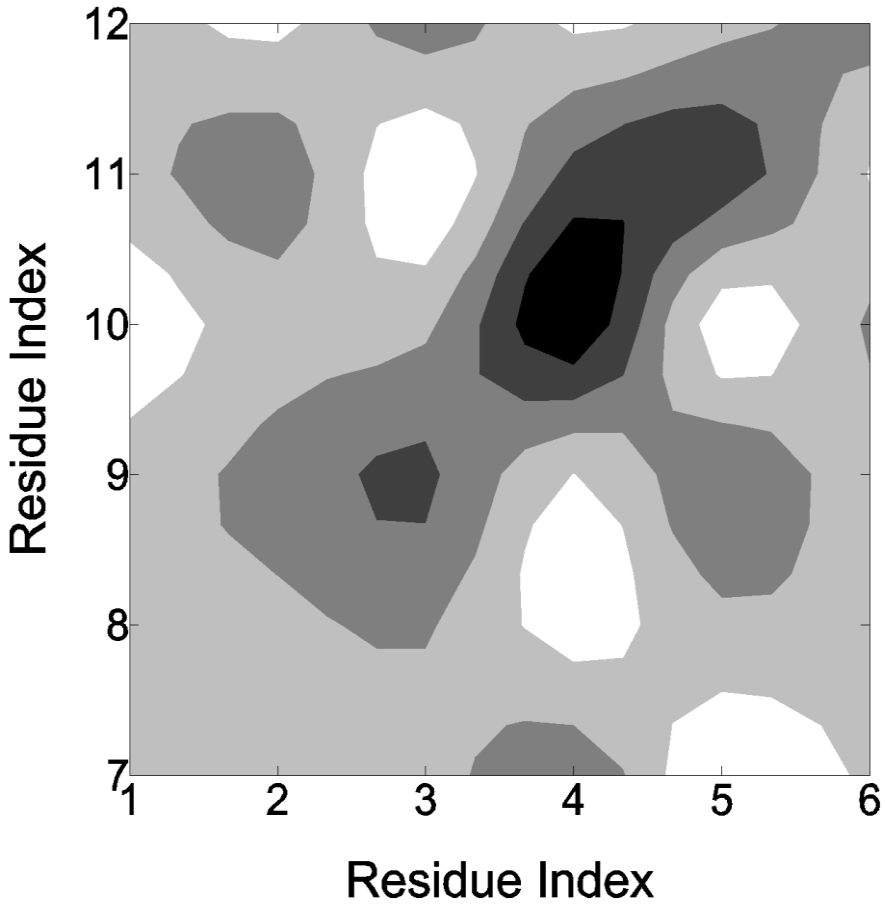


Figure 2-2: **(a)** Γ matrix of the bound form. **(b)** Part $1-2$ of the Γ matrix in the bound form. Black color indicates negative elements (Attractive) whereas white indicates positive elements (Repulsive).

The 6 strongest spring constants are observed in decreasing order among PHE4-MET10, ARG5-GLN11, PHE4- GLN11, MET3-PHE9, PHE4-LEU7 and ARG5-PHE9. Phenylalanine is an aromatic hydrophobic neutral amino acid whereas Methionine is a hydrophobic neutral amino acid. Hence, due to their hydrophobic character, trying to escape the surrounding water molecules, PHE4-MET10 and MET3-PHE9 interact strongly. We interpret the strong attractive spring between ARG5 and GLN11 as a result of an indirect effect, induced by their covalent neighbors, PHE4 and MET10. The strong attraction between PHE4-MET10 results in an induced steric attraction between ARG5 and GLN11. Otherwise, Arginine and Glutamine are both polar hydrophilic residues which can interact with the surrounding water

molecules and a direct strong attraction between them is not expected. Leucine is an aliphatic hydrophobic neutral amino acid. Due to its hydrophobic character, attractive interaction among PHE4-LEU7 is observed. Arginine is a polar hydrophilic positively charged amino acid. The weak interaction between ARG5 and PHE9 may also be seen as an indirect attractive force induced by their hydrophobic neighbors PHE4-MET10 and MET3-PHE9.

2.3.3. Binding Energies in the Harmonic Approximation

In this section, we calculate the change in energy of the system upon binding. The change results from two components: the first is the change in the energy of each peptide upon binding, and the second is the energy stored in the system from the interactions of the two peptides.

For two peptides the force constant matrix K is in the following form.

$$\mathrm{K} = \begin{bmatrix} \mathrm{K}^{(1)} & \mathrm{K}^{(1,2)} \\ \mathrm{K}^{(2,1)} & \mathrm{K}^{(2)} \end{bmatrix} \qquad 2\text{-}32$$

Where $\mathrm{K}^{(1)}$, reflects the interaction among residues of peptide 1 and $\mathrm{K}^{(1,2)}$ reflects the interaction among the residues of peptide 1 and peptide 2.

The energies of of peptides $\xi = 1,2$ were evaluated in their unbound state as

$$U_{unbound}{}^{(\xi)} = \frac{1}{2}\left\langle \Delta R^{(\xi)T} \mathrm{K}^{(\xi)} \Delta R^{(\xi)} \right\rangle = 6kT \qquad 2\text{-}33$$

Here $\mathrm{K}^{(\xi)}$ is evaluated via equation (10) using the covariance matrix which was obtained from the molecular dynamic trajectory of peptide ξ. Using the fluctuations $\Delta R^{(\xi)}$, the

energies corresponding to these coordinates are evaluated as shown in equation (11). Taking the average of these energies over the trajectory the average energy shown in equation (33) were obtained.

In the bound state on the other hand energies are calculated as

$$U_{bound}^{(1)} = \frac{1}{2}\left\langle \Delta R^{(1)T} \mathrm{K}^{(1)} \Delta R^{(1)} \right\rangle = 10.04kT$$

$$U_{bound}^{(2)} = \frac{1}{2}\left\langle \Delta R^{(2)T} \mathrm{K}^{(2)} \Delta R^{(2)} \right\rangle = 10.02kT \qquad \text{2-34}$$

Here the intramolecular force constant part $\mathrm{K}^{(1)}$ of peptide 1 which is shown in the detailed structure of the K matrix in equation (30) is selected. Using the fluctuations $\Delta R^{(1)}$ of peptide 1 in its bound form and the force constant matrix $\mathrm{K}^{(1)}$, the energies corresponding to fluctuation values $\Delta R^{(1)}$ are evaluated. Again average of the energies are evaluated over the trajectory and resulting average energy is shown in equation (34).

The energy stored as direct interaction between the two peptides is calculated as the sum of the two terms:

$$U_{bound}^{(1,2)} = \frac{1}{2}\left\langle \Delta R^{(2)T} \mathrm{K}^{(1,2)} \Delta R^{(1)} \right\rangle = \text{-}2.53kT$$

$$U_{bound}^{(2,1)} = \frac{1}{2}\left\langle \Delta R^{(1)T} \mathrm{K}^{(2,1)} \Delta R^{(1)} \right\rangle = \text{-}2.53kT \qquad \text{2-35}$$

The change in the energy of the system upon binding is

$$\Delta U = \left(U_{bound}{}^{(1)} - U_{unbound}{}^{(1)}\right) + \left(U_{bound}{}^{(2)} - U_{unbound}{}^{(2)}\right) + U_{bound}{}^{(1,2)} + U_{bound}{}^{(2,1)} = 3kT \qquad 2\text{-}36$$

Note that the change in energy, ΔU, is positive. The individual energies of the peptides increase due to binding. This is an expected result since the peptides are forced to tenser conformations instead of staying in their native conformation. The interaction energy between the two peptides are, on the other hand, attractive showing that there is attraction between the two peptides.

In harmonic approximation, the energies of peptide 1, 2, and the bound form are evaluated from equation (19) as: $U_{unbound}{}^{(1)} = 6kT$, $U_{unbound}{}^{(2)} = 6kT$ And $U_{bound}{}^{(1,2)} = 15kT$. $U_{bound}{}^{(1)}$, $U_{bound}{}^{(2)}$ and U_{bound} are the energies of peptide 1, peptide 2 and the bound form respectively.

The difference in energy, ΔU between the bound and unbound states of peptide 1 and 2 is

$$\begin{aligned}\Delta U &= U_{bound} - U_{unbound}{}^{(1)} - U_{unbound}{}^{(2)} \\ &= 15kT - 6kT - 6kT \\ &= 3kT\end{aligned} \qquad 2\text{-}37$$

which is the same as the value calculated by using the force constants, as expected.

The binding free energy $\Delta A_{binding}$ of peptide 1 and peptide 2 is evaluated using the energy formalism proposed by equation (23). For the simulation at $T = 310\text{K}$ the following binding energy emerges,

$$\begin{aligned}\Delta A_{binding} &= A^{b} - A^{(1)} - A^{(2)} \\ &= 12.63kT\end{aligned} \qquad 2\text{-}38$$

Using equation (25) the change in entropy in binding is obtained as $\Delta S_{binding} = -9.63k$.

The increase in energy comes from the fact that, prior to binding 12+12=24 internal degrees of freedom existed whereas subsequent to binding 30 internal degrees of freedom exist.

This increase in the degrees of freedom is caused by the additional 6 relative motions of the peptides with respect to each other which are introduced via binding.

Since peptides are more restricted by binding, the disorder of peptides decreases and therefore the entropy decreases.

However, in biological systems binding occurs inside a solvent. Hence, if the solvent and the peptides are considering together as a closed system, the energy and entropy change would also include the thermodynamic change of the solvent and the solvent-peptide interactions.

It was observed that these peptides stay in their bound form for the entire trajectory. Solvent effects most probably compensate for the thermodynamically unfavorable entropy and energy change observed for the peptides. In addition the unharmonicity present in the system is expected to affect the energy values to some extent.

2.3.4. Deviation from Harmonicity

In order to understand how accurate our harmonic model suits the system, the term $\frac{\left\langle \Delta \boldsymbol{r}_i^{\,4} \right\rangle}{\left(\left\langle \Delta \boldsymbol{r}_i^{\,2} \right\rangle \right)^2}$ is investigated. For each mode its theoretical value was predicted by equation (29) to be 3. Deviations from this value indicate the unharmonicity present in the system. In figure (3) the solid line shows the term $\frac{\left\langle \Delta \boldsymbol{r}_i^{\,4} \right\rangle}{\left(\left\langle \Delta \boldsymbol{r}_i^{\,2} \right\rangle \right)^2}$ which is evaluated using the molecular dynamic trajectory. The dotted line on the other hand shows the theoretical value. As can be seen in the figure the deviation from the theoretical value is %11 for the first mode which is the global (slowest) mode. For the next two slowest modes, mode 2 and 3, deviations are negligible. Deviations seem to dominate at the medium modes. Maximum deviation is around %41 and

observed for mode 21. The fact that deviation from harmonicity is mild at the most global 3 modes indicates that our model acceptable to predict the gross overall behavior the system.

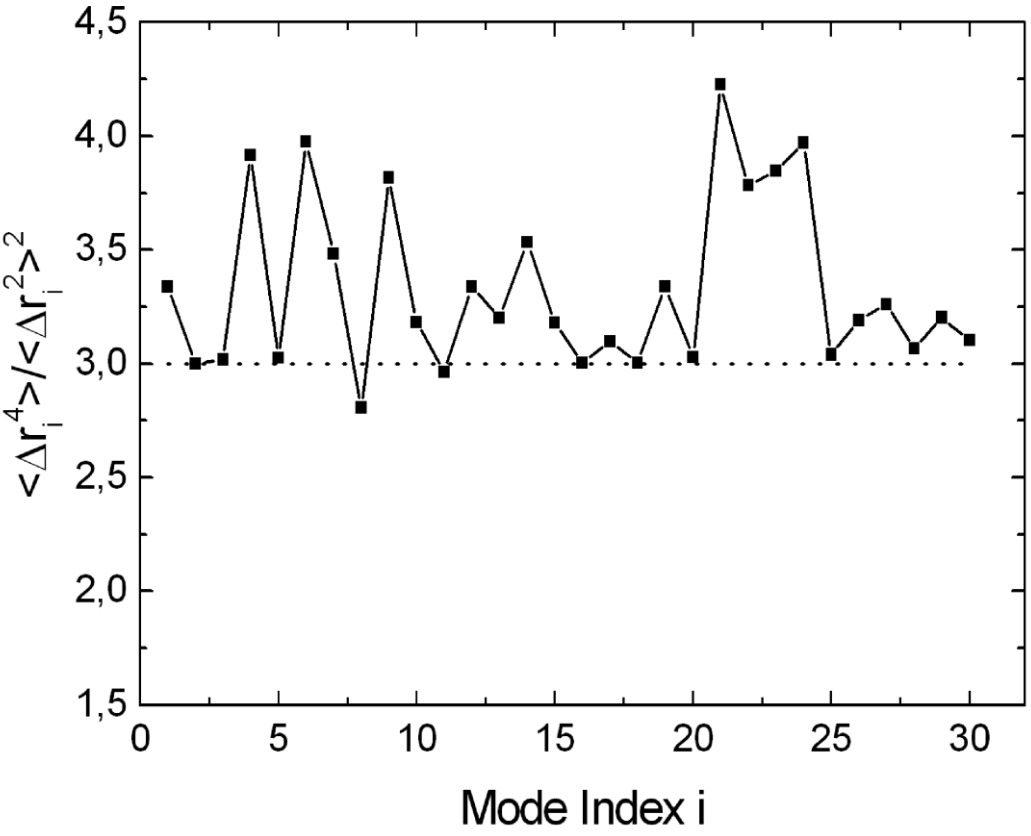


Figure 2-3. $\frac{\left\langle \Delta \boldsymbol{r}_i^{\,4} \right\rangle}{\left(\left\langle \Delta \boldsymbol{r}_i^{\,2} \right\rangle \right)^2}$ values evaluated using the molecular dynamic trajectory are shown with the solid line and the theoretical values are shown with the dotted line.

2.4. Conclusion

Methodologically the most similar models are the Gaussian Network Model (GNM), Anisotropic Network model (ANM) and the REACH (Realistic Extension Algorithm via

Covariance Hessian) method. All these coarse-grained models are built using the alpha carbon coordinates only. Both the GNM and the ANM assume that residue pairs within a cutoff distance are connected with springs with the same force constant. Since force constants are evaluated posteriori, these models do not predict the magnitude of the fluctuations. In GNM the fluctuations are assumed to be isotropic and Gaussian . The ANM is simply a normal mode analysis applied to an elastic network model [9]. The REACH method on the other hand is an elastic network model in which the residue interaction force constants are evaluated using the variance-covariance matrix obtained from the molecular dynamic (MD simulations) [24]. In REACH, the interactions are divided into 4 classes; virtual 1-2 (between residue i and i+1), virtual 1-3 (between residue i and i+2),virtual 1-4 (between residue i and i+3) and the nonbonded interactions. In [24, 26] wide distributions of force constants for these interactions were obtained, which range from negative values to positive values for 1-3, 1-4 and nonbonded interactions. For a simplified understanding of protein dynamics and for convenient application in coarse grained MD simulation, a constant value model was assumed in REACH for each of the virtual bond interactions and a single exponential decay was assumed for the nonbonded interaction[24]. In other words representative force constants were assumed for each class of the virtual bonds and a representative equation was assumed for the nonbonded interactions. For interactions among residues in different proteins, the same single exponential form which was used for intramolecular nonbonded interaction, was used. Hence the distance dependence of the force constants for nonbonded intramolecular interactions was also applied to the intermolecular interactions. In our work the virtual 1-2 type interactions were found to be around the same order as the ones obtained for myoglobin monomer[24] and for the myoglobin dimer [26]

In our model all force constants were taken as they are. Positive force constants κ_{ij} indicate that residue i and residue j tend to stay at their equilibrium distance, $\boldsymbol{norm}\left(\boldsymbol{R}^{o}_{i} - \boldsymbol{R}^{o}_{j}\right)$. Any deviation from this distance will cause an increase in energy as indicated in equation (11) which is unfavorable. The equilibrium coordinates were assumed to be the same as the mean coordinates of the atoms $\boldsymbol{R}$ obtained via MD. Negative force constants on the other hand indicate that any deviation from the mean coordinates results in a decrease in energy. Since an energy decrease is favorable, the meaning of a negative force constant is that residue i and j

do not prefer to stay at these relative positions, $\boldsymbol{norm}\left(\boldsymbol{R}^{o}{}_{i} - \boldsymbol{R}^{o}{}_{j}\right)$. The aim in our model was to find strong positive force constants, which show important binding sites. The other 3 models on the other hand do not focus on the force constants. They rather use approximate generalized force constant values to investigate the dynamics of the proteins.

Chapter 3

QUASI-HARMONIC ANALYSIS OF MODE COUPLING IN FLUCTUATING NATIVE PROTEINS

3.1. Introduction

The expectation that the fluctuations of residues of a protein should be strongly coupled in order for the protein to perform its function is leading to a growing interest in protein physics [14, 15, 36]. Recently, several papers addressed this issue and gave specific examples of mode coupling: Moritsugu et al [16, 17] were among the first who studied mode coupling due to intramolecular vibrational energy transfer in myoglobin at near zero temperature by molecular dynamics simulations. Calculations were performed by adding a small amount of energy to one mode and monitoring the transfer of this energy into three other modes. The latter were selected according to a Fermi-resonance related argument among frequencies. Sanejouand and collaborators[14, 15] [37] studied the consequences of energy transfer at room temperature by cooling specific residues at the surface of dimeric citrate synthase and observing the transfer of energy to different modes. They explained the localization of energy at specific residues located in the stiff parts of the protein and its transfer to other modes by employing an anharmonic potential. The interplay between the surroundings of the protein and intraprotein dynamics have been also investigated[38, 39]. Of particular relevance to our work is the paper by Moritsugu and Smith in which intra-protein dynamics is investigated at 300 K. The energetic interactions of a protein with its surroundings and the distribution of the energy absorbed by the protein to different residues is the major issue concerning protein function.

Interest in modal decomposition in protein dynamics is not new. Decomposition of trajectories into an essential subspace and a remaining Gaussian subspace has been employed widely [20, 40-45]. In general, the interest has been focused on retaining the essential subspace of anharmonic modes for the analysis of correlated motions and ignoring the remaining Gaussian subspace that consists of the faster modes of motion. The interplay

between harmonicity and anharmonicity in proteins was studied by Hayward et al, by employing molecular dynamics simulations [13], reaching the conclusion that the motion within a local minimum is mainly harmonic and the anharmonic component arises from transitions from one minimum to the other. Although it is generally true that the essential subspace of anharmonic modes are the ones that lead to important couplings between modes, higher modes also play important roles in protein fluctuation dynamics. The findings of Moritsugu et al [16, 17, 44] showing energy transfer between modes 1589, 1860, 5858 out of a total of 7418 modes at zero temperature of myoglobin is an example of the importance of faster modes. In the interest of observing the effects of quasi-harmonic motions, we used the covariance projection technique [44] and analyzed the fluctuations of a small protein, Crambin at 273.15K and 310 K. The principle aim of this paper is to understand how modes in proteins are coupled with each other, whether they persist throughout the full trajectory, how the energy of the protein is distributed to these modes, and what the observable consequences of such coupling are. Our calculations show that, among the 1971 modes of the system, the slowest few modes are strongly coupled among themselves throughout the full 40 ns trajectory. The time-averaged energies of the slowest modes are always below the average energy of the protein. Strong and long lasting couplings are also observed in the faster modes. The time-averaged energies of the faster modes that exhibit strong coupling are always larger than the average energy. The faster modes that exhibit coupling are few in number and they appear and disappear at different stretches of the trajectory. The coupling that we identified between two of the modes suggest a distinct mechanism where the residue fluctuations of one of the modes drive the motions of the other mode. The transfer of energy among different modes, including slow as well as fast modes, and the resulting affinity and conformation changes in a protein is of importance for understanding allostery [36].

In this chapter paper, we present results for Crambin in water at 273.15K. The results for 310 K, which are similar to those for 273.15 K are presented in the end of this chapter as supplementary material.

The plan of this chapter is as follows: In Part 2, we present the model and the simulations. In the results section, we first evaluate the marginal energies of the modes in order to see which modes deviate strongly from the harmonic. Having identified these modes, we then use the

lowest order two mode coupling analysis for their detailed analysis. The paper ends with the discussion section.

3.2. The Model and Simulations

3.2.1. Molecular Dynamics Simulations

Crambin, Protein Data bank code 1EJG.pdb, was selected as an example since it is a relatively small, 46 residue protein and its dynamics is widely studied[41, 46, 47]. All Molecular dynamics simulations were performed for an N,V,T ensemble in explicit solvent (water) using NAMD 2.6[35] package with CHARMM27 force field. Two different simulations were performed at temperatures of 273.15 K and 310 K. The protein was solvated in a waterbox of 12 A cushion and periodic boundary conditions were applied. Ions were added in order to represent a more typical biological environment. Langevin dynamics was used to control the systems temperature and pressure. All atoms were coupled to the heat bath. A time step of 1fs was used. Nonbonded and electrostatic forces were evaluated at each time step. In order to keep all degrees of freedom no rigid bonds were used. Three minimization-equilibration cycles were applied: The first one was applied under N,P,T conditions to relax the water in the first place and the second and third ones were applied under N,V,T conditions to find a local minimum of the whole system's energy[48]. The energy of the initial system was first minimized for 20000 steps. The system was then equilibrated by first keeping the Protein fixed for the first 0.1 ns. At 273.15 K it took 0.01 ns for the volume to converge whereas at 310 K it took 0.02 ns for the volume to converge.. During the remaining time, volume fluctuated around 155500 $Å^3$ at 273.15 K and around 159000 $Å^3$ at 310K. Then, the protein was released stepwise by applying harmonic constraining forces to every backbone atom of 1, 0.5 and 0.25 kcal/(mol*Angstrom^2) in magnitude each for 0.05 ns. Finally, the simulation was performed for an additional 0.05 ns without applying any force. Having finished the first cycle, the second minimization-equilibration cycle was performed, this time the protein was free to move. Again, 20000 steps of minimization were applied and system was equilibrated for 1 ns. After a final 20000 steps of minimization, the protein was equilibrated for an additional 1 ns at both temperatures.

After equilibration, several stretches of the trajectory of different lengths were used for analysis. In the main part of the work, 2.75 ns long trajectories from different stretches of a full 40 ns simulation were used. The first 2.75 ns stretch was taken 10 ns after the final equilibration.

At every 500th time step of the 2.75 ns trajectory, the instantaneous atomic coordinates $\hat{\mathbf{R}}$ of all atoms, the velocities, the pressures and the energies were recorded. In order to eliminate all the rotational and translational motions, all structures were aligned with respect to the structure observed at time instant 1.5 ns of the production phase. Alignments were performed using the transformation matrix which shows the best fit of the backbone atoms. All transformation matrices were constructed via tcl commands in VMD.

The 46 residue Crambin consists of 657 atoms. Hence a set of 1971 modes are obtained. Then the overall rotational and translational motions were eliminated since they are irrelevant for the internal motions [44].Thus, an overall of 1965 non-zero modes were obtained.

3.2.2. Fluctuations, Correlations and Principal Component Decomposition

The fluctuations $\Delta\boldsymbol{R}$ of atoms are defined by $\Delta\boldsymbol{R} = \hat{\boldsymbol{R}} - \bar{\boldsymbol{R}}$, where $\bar{\boldsymbol{R}}$ are the mean atomic coordinates and hence are time independent quantities, which define an average configuration obtained by the protein during the part of the trajectory that we use for calculations.

The covariance matrix $\boldsymbol{C}$ is defined in terms of $\Delta\boldsymbol{R}$ as

$$C = \left\langle \Delta\boldsymbol{R}\Delta\boldsymbol{R}^{T} \right\rangle \qquad 3\text{-}1$$

Here, the angular brackets denote the average, taken over the trajectory. The instantaneous fluctuations are transformed into modal space using principal component decomposition with the covariance matrix as [44, 49],

$$\Delta \boldsymbol{r} = \boldsymbol{C}^{-1/2} \Delta \boldsymbol{R} \qquad 3\text{-}2$$

Where $\Delta \boldsymbol{r}$ are the dimensionless transformed fluctuations in modal space. We let $\boldsymbol{e}$ represent the eigenvector matrix that diagonalizes $\boldsymbol{C}$, and λ represent the eigenvalues. Then, $\left\langle \Delta \boldsymbol{R} \Delta \boldsymbol{R}^T \right\rangle^{-1/2} = diag \lambda^{-1/2} \boldsymbol{e}^T$ and the fluctuations $\Delta \boldsymbol{r}$ are the fluctuations in mode space spanned by the eigenvectors, $\boldsymbol{e}$ [49]. The components of the real trajectory that correspond to a given mode is obtained simply by keeping the eigenvalue of interest, equating all the others to zero, followed by a back transformation of equation (2).

In figure 1 the distributions $W(\Delta r)$ of the modal coordinates $\Delta \boldsymbol{r}$ are shown for the first 12 slowest modes. The range of values taken by $\Delta \boldsymbol{r}$ is divided into twenty-five stations, and the normalized frequency of observing a given $\Delta \boldsymbol{r}$ is shown in the figures by the filled circles. The lines through the points are drawn using a 17^{th} order Hermite series expansion of the fluctuation probability distribution function (See equation (14) below). The distributions of the 1^{st}, 2^{nd}, 3^{rd}, 4^{th}, 8^{th} and 12^{th} modes depart strongly from a harmonic distribution. These modes mainly contain information about the anharmonicities in the system. The overall behavior is that the distributions converge to single peak Gaussians with increasing mode numbers.

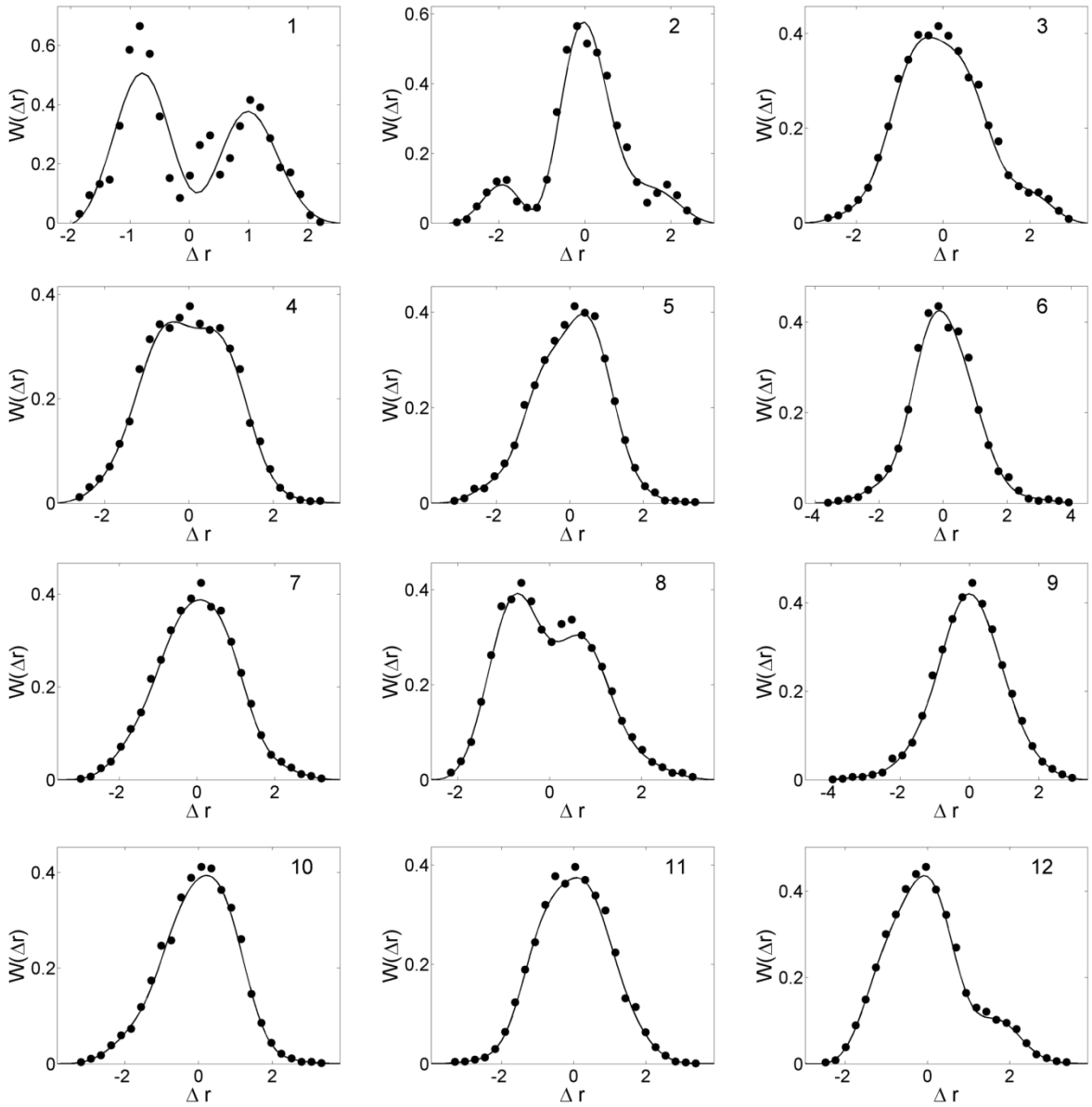


Figure 3-1. The normalized histograms of the modal coordinates Δr for the first 12 slow modes. Filled points are the calculated values and the lines through them are evaluated using equation (14), up to the 17^{th} order terms.

3.2.3. Tensorial Hermite Series Approximation and Thermodynamic Analysis

In order to describe the behavior of the protein in its full generality, we use a 3n dimensional tensorial Hermite series, where n is the number of residues. Originally, a three dimensional Tensorial Hermite series was used by Flory and Yoon to describe the statistical behavior of

short polyethylene chains [50, 51].The method was then generalized and applied to protein fluctuations by Yogurtcu et al. [49] in 3n dimensional space. We adopt the notation of[49-51].

In terms of the actual coordinates $\Delta\boldsymbol{R}$, the probability function $W(\Delta\boldsymbol{R})$ may be approximated by the tensorial Hermite series [49]

$$\boldsymbol{W}(\Delta\boldsymbol{R}) = (2\boldsymbol{\pi})^{-\boldsymbol{N}/2}\left(\det\left\langle\Delta\boldsymbol{R}\Delta\boldsymbol{R}^{\boldsymbol{T}}\right\rangle\right)^{-1/2}\exp\left[-\frac{1}{2}\Delta\boldsymbol{R}^{\boldsymbol{T}}\left\langle\Delta\boldsymbol{R}\Delta\boldsymbol{R}^{\boldsymbol{T}}\right\rangle^{-1}\Delta\boldsymbol{R}\right]\cdot$$

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

Here N is the number of nonzero modes after the elimination of translational and rotational degrees of freedom. The leading term of the distribution function is the Gaussian and the terms in the last square brackets show the deviations from the Gaussian, introduced as corrections in terms of the Hermite polynomials. The expression $\boldsymbol{W}(\Delta\boldsymbol{R})$ is valid for any amplitude of fluctuations. These correction terms become unimportant as the fluctuations become small. The first few polynomials, H_ν, are

$$H_1(\Delta\boldsymbol{R}) = \Delta R_i$$

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

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

$$H_4(\Delta\boldsymbol{R}) = \left(\Delta R^4 - \Delta R^2\delta + \delta^2\right)_{ijkl} \qquad 3\text{-}4$$

$$H_5(\Delta\boldsymbol{R}) = \left(\Delta R^5 - \Delta R^3\delta + \Delta R\delta^2\right)_{ijklm}$$

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

where, δ_{ij} is the Kronecker delta, and $(\Delta R\delta)_{ijk}$ in the expression for H_3 is a shorthand 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 equation (3). For example,

$$\left(\Delta R^2 \delta\right)_{ijkl} = \Delta R_i \Delta R_j \delta_{kl} + \Delta R_i \Delta R_k \delta_{jl} + \Delta R_i \Delta R_l \delta_{kj} + \Delta R_j \Delta R_k \delta_{il} + \Delta R_j \Delta R_l \delta_{ik} + \Delta R_k \Delta R_l \delta_{ij} .$$

The expression given by equation (3) contains all the information on the behavior of the system in its full generality. However, being an infinite series, it is complicated, and being a moment based expansion it may have problems of slow convergence [49-51]. If the higher moments of the system can be calculated easily, such as from a molecular dynamics trajectory, then equation (3) may suitably be used to extract information on the system at different levels of approximation. The deviations from harmonicity are due to two different sources, (i) anharmonicity of pure modes and (ii) coupling of two or more modes. The first is the presence of higher order terms $\Delta \boldsymbol{r}_i^m$ where $m > 2$ of pure modes. The second is the coupling of two modes represented by a nonzero value of the average $\left\langle \Delta \boldsymbol{r}_i^m \Delta \boldsymbol{r}_j^k \right\rangle$. Coupling of three different modes such as $\left\langle \Delta r_i^k \Delta r_j^l \Delta r_k^m \right\rangle$ and coupling of more than three modes also contribute to coupling. However, keeping track of such high order coupling is a formidable task, and we will consider only the lowest order coupling in the present work. It is to be noted that all orders of the moments that appear in $\langle H_v \rangle$ in equation (3) can easily be evaluated from molecular dynamics trajectories. Hermite series are generally known as slow converging series[49-51]. In the present work, we adopt the Hermite series up to the 17th order moments that ensures convergence within a few percent of simulation results. A general scheme and a computer script for easily computing tensorial Hermite polynomials from molecular dynamics simulation trajectories up to any desired order is provided in the next section [28].

In terms of the modal coordinates, $\Delta \boldsymbol{r}$, where $\Delta \boldsymbol{r}$ is the vector with elements $\{\Delta r_1, \Delta r_2 ...., \Delta r_N\}$, the probability function reads as

$$W(\Delta \boldsymbol{r}) = (2\pi)^{-N/2} \exp\left[-\frac{1}{2}\Delta \boldsymbol{r}^T \Delta \boldsymbol{r}\right] \cdot \left[1 + \sum_{\nu=3}^{\infty} (\nu!)^{-1} \langle H_\nu \rangle \cdot H_\nu(\Delta \boldsymbol{r})\right] \quad 3\text{-}5$$

where, the average Hermite polynomials are defined as

$$\langle \boldsymbol{H}_\nu \rangle = \int_{-\infty}^{\infty} \boldsymbol{H}_\nu(\Delta \boldsymbol{r}) W(\Delta \boldsymbol{r}) d\Delta \boldsymbol{r} \quad 3\text{-}6$$

The elements of $\langle \boldsymbol{H}_\nu \rangle$ now contain products of modal coordinates. For example the third order terms are now $\langle \Delta \boldsymbol{r}_i \Delta \boldsymbol{r}_j \Delta \boldsymbol{r}_k \rangle$, and are measures of the extent of mode coupling. The second order modes are decoupled since $\langle \Delta \boldsymbol{r}_i \Delta \boldsymbol{r}_j \rangle = \delta_{ij}$.

equation (5) may be rewritten as

$$W(\Delta \boldsymbol{r}) = C \exp\left[-\frac{1}{2}\Delta \boldsymbol{r}^T \Delta \boldsymbol{r} + \ln\left(1 + \sum_{\nu=3}^{\infty} (\nu!)^{-1} \langle H_\nu \rangle \cdot H_\nu(\Delta \boldsymbol{r})\right)\right] \quad 3\text{-}7$$

where, $C = (2\pi)^{-\frac{N}{2}}$ is the normalization constant. Writing equation (7) as

$$W(\Delta \boldsymbol{r}) = Z^{-1} \exp\left[-\frac{1}{2}\Delta \boldsymbol{r}^T \Delta \boldsymbol{r} + \ln\left(1 + \sum_{\nu=3}^{\infty} (\nu!)^{-1} \langle H_\nu \rangle \cdot H_\nu(\Delta \boldsymbol{r})\right) - \beta E_0\right] \quad 3\text{-}8$$

where $\beta E_0 = -\ln Z - \ln C$, and comparing with the normalized distribution for a T, V, N ensemble

$$f(q) = \frac{e^{-\beta E(q)}}{Z} \qquad 3\text{-}9$$

leads to the energy of the system at a specific microstate having modal coordinates $\Delta \mathbf{r}$ as

$$E(\Delta \boldsymbol{r}) = \frac{1}{\beta}\left[\frac{1}{2}\Delta \boldsymbol{r}^T \Delta \boldsymbol{r} - \ln\left(1 + \sum_{\nu=3}^{\infty}(\nu!)^{-1}\langle H_\nu \rangle \cdot H_\nu(\Delta \boldsymbol{r})\right)\right] + E_0 \qquad 3\text{-}10$$

equation (10) can be considered in two parts: (i) The fluctuation part, which depends on the microstate $\Delta \boldsymbol{r}$ and (ii) the reference energy E_o which depends on the free energy as

$$E_o = F + \frac{1}{2}NkT\ln(2\pi) \qquad 3\text{-}11$$

Using equations (9) and (10), the thermodynamic energy, U, is written as

$$U = \int_{\Delta \boldsymbol{r}} E(\Delta \boldsymbol{r})\, f(\Delta \boldsymbol{r}) d\Delta \boldsymbol{r} = \frac{kT}{2}N - kT\left\langle \ln\left(1 + \sum_{\nu=3}^{\infty}(\nu!)^{-1}\langle H_\nu \rangle \cdot H_\nu(\Delta \boldsymbol{r})\right)\right\rangle + E_0 \qquad 3\text{-}12$$

Substituting $E_o = U - TS + \frac{1}{2}NkT\ln(2\pi)$ into equation (12) leads to the following expression for the entropy

$$S = \frac{k}{2}N\left(1+\ln(2\pi)\right) - k\left\langle \ln\left(1+\sum_{\nu=3}^{\infty}(\nu!)^{-1}\langle H_\nu\rangle \cdot H_\nu(\Delta \boldsymbol{r})\right)\right\rangle \qquad 3\text{-}13$$

Thus correlations, depending on their signs, decrease or increase the entropy with respect to that of the harmonic model. The leading term on the right hand side of equation (12) is the energy of the N modes in the uncoupled case. The second term is the contribution of anharmonicity and coupling. The third term is the reference energy, which depends on the macro state of the system as was defined in equation (11).

The marginal energy of a mode is defined as the energy obtained by considering only one mode, equating all the other modal variables to zero. equation (10) is now written as

$$E_i = E(0,0,..,\Delta \boldsymbol{r}_i,..0,0) = \frac{1}{\beta}\left[\frac{1}{2}\Delta \boldsymbol{r}_i^{\,2} - \ln\left(1+\frac{1}{\nu!}\sum_{\nu=0}^{\infty}\langle H_\nu(\Delta \boldsymbol{r}_i)\rangle \bullet H_\nu(\Delta \boldsymbol{r}_i)\right)\right] + E_0 \qquad 3\text{-}14$$

Here the Hermite series expansion in the parenthesis may be expressed as an ν^{th} order polynomial of the modal coordinate $\Delta \boldsymbol{r}_i$,

$$\frac{1}{\nu!}\sum_{\nu=0}^{\infty}\langle H_\nu(\Delta \boldsymbol{r}_i)\rangle \bullet H_\nu(\Delta \boldsymbol{r}_i) = c_1 + c_2\Delta r_i + c_3\Delta r_i^{\,2} + \ldots\ldots\ldots + c_\nu \Delta r_i^{\,\nu} \qquad 3\text{-}15$$

Here Δr_i^v is the V^{th} power of the ith modal coordinate only and c_v is a function of the various moments of the Hermite polynomials. Since the reference energy E_0 is constant throughout all calculations, for simplicity it will be set to zero in all calculations and figures.

3.3. The Lowest Order Coupling of Two Modes:

The results of MD simulations to be reported below show that the energies of few of the modes deviate strongly from the harmonic resulting from coupling, obtained by the 17th order Hermite series. A mode may be coupled to the surrounding water, or to another mode, or to several other modes simultaneously. We are specifically interested in the coupling of two modes to each other. In order to understand the details of coupling between two modes, we perform a first order approximation keeping the third order moments only, which read as $\left\langle \Delta \boldsymbol{r}_i \Delta \boldsymbol{r}_j^{\,2} \right\rangle$ and $\left\langle \Delta \boldsymbol{r}_i^{\,2} \Delta \boldsymbol{r}_j \right\rangle$. In the lowest order, coupling of three modes, $\left\langle \Delta \boldsymbol{r}_i \Delta \boldsymbol{r}_j \Delta \boldsymbol{r}_k \right\rangle$ where i and j and k are different from each other, may also contribute to coupling, which we also discuss in some detail.

If both of the terms $\left\langle \Delta \boldsymbol{r}_i \Delta \boldsymbol{r}_j^{\,2} \right\rangle$ or $\left\langle \Delta \boldsymbol{r}_i^{\,2} \Delta \boldsymbol{r}_j \right\rangle$ have the same sign, the fluctuations of these two modes are positively correlated. If they have different signs, fluctuations of mode i and j are negatively correlated. In terms of energy, large values of $\left\langle \Delta \boldsymbol{r}_i \Delta \boldsymbol{r}_j^{\,2} \right\rangle$ and $\left\langle \Delta \boldsymbol{r}_i^{\,2} \Delta \boldsymbol{r}_j \right\rangle$ indicate the presence of large energy coupling among them, but the type of coupling, i.e., whether i gains or loses energy while j loses energy and vice versa, cannot be concluded.

Expanding equation (10) up to the third order terms and setting all modes other than i and j to zero, the marginal energy of modes i and j together is obtained as

$$E_{i+j}=\frac{1}{\beta}\Bigg[\ \frac{1}{2}\Delta r_i^{\,2}+\frac{1}{2}\Delta r_j^{\,2}-\ln\Bigg(\ 1+\frac{1}{6}\left\langle\Delta r_i^{3}\right\rangle\Delta r_i\ \left(\Delta r_i^{2}-3\right)+\frac{1}{6}\left\langle\Delta r_j^{3}\right\rangle\Delta r_j\left(\Delta r_j^{2}-3\right)$$

$$+\frac{1}{2}\left\langle\Delta r_i\ \Delta r_j^{2}\right\rangle\left[\Delta r_i\ \Delta r_j^{2}-\Delta r_i\ \right]+\frac{1}{2}\left\langle\Delta r_i^{2}\Delta r_j\right\rangle\left[\Delta r_i^{2}\Delta r_j\ -\Delta r_j\right]\ \Bigg)\ \Bigg] \qquad 3\text{-}16$$

The marginal energy of the ith mode reads as

$$E_i=\frac{1}{\beta}\left[\frac{1}{2}\Delta r_i^{\,2}-\ln\left(1+\frac{1}{6}\left\langle\Delta r_i^{3}\right\rangle\Delta r_i\ \left(\Delta r_i^{2}-3\right)\right)\right] \qquad 3\text{-}17$$

The energy coupling $\Delta E_{ij}\equiv E_{i+j}-E_i-E_j$ among mode i and j at the microstate Δr is obtained from equations (16) and (17) as

$$\Delta E_{ij}=-\frac{1}{\beta}\ln\left[\frac{1+\Theta_i+\Theta_j+\Phi_{ij}}{\left(1+\Theta_i\right)\left(1+\Theta_j\right)}\right] \qquad 3\text{-}18$$

where,

$$\Theta_i=\frac{1}{6}\left\langle\Delta r_i^{3}\right\rangle\Delta r_i\ \left(\Delta r_i^{2}-3\right) \qquad 3\text{-}19$$

is the single mode anharmonicity term, and

$$\Phi_{ij}=\frac{1}{2}\left\langle\Delta r_i\ \Delta r_j^{2}\right\rangle\left(\Delta r_i\ \Delta r_j^{2}-\Delta r_i\ \right)+\frac{1}{2}\left\langle\Delta r_i^{2}\Delta r_j\right\rangle\left(\Delta r_i^{2}\Delta r_j\ -\Delta r_j\right) \qquad 3\text{-}20$$

contains the joint distribution of modes i and j.

We now expand equation (18) up to the second order terms in Θ_i, Θ_j, and Φ_{ij}

$$\Delta E_{ij} = -\frac{1}{\beta}\left[\Phi_{ij} - \frac{1}{2}\Phi_{ij}^{\,2} - \left(\Theta_i + \Theta_j\right)\Phi_{ij} - \Theta_i\Theta_j + ...\right] \qquad 3\text{-}21$$

The function Φ_{ij} is third order in the modal coordinates. Thus, the first term on the right hand side of equation (21) is of third order. The remaining ones are all of sixth order. The energy coupling in the presence of third order term only reads as $\Delta E_{ij} = -\frac{\Phi_{ij}}{\beta}$, or

$$\Delta E_{ij} = -\frac{1}{2\beta}\left[\left\langle \Delta \boldsymbol{r}_i\, \Delta \boldsymbol{r}_j^{\,2}\right\rangle\left(\Delta \boldsymbol{r}_i\, \Delta \boldsymbol{r}_j^{\,2} - \Delta \boldsymbol{r}_i\right) + \frac{1}{2}\left\langle \Delta \boldsymbol{r}_i^{\,2}\Delta \boldsymbol{r}_j\right\rangle\left(\Delta \boldsymbol{r}_i^{\,2}\Delta \boldsymbol{r}_j - \Delta \boldsymbol{r}_j\right)\right] \qquad 3\text{-}22$$

Equation (22) is the lowest order contribution to mode coupling. It is a function of the third order moments $\left\langle \Delta \boldsymbol{r}_i^{\,3}\right\rangle$, $\left\langle \Delta \boldsymbol{r}_j^{\,3}\right\rangle$, $\left\langle \Delta \boldsymbol{r}_i\, \Delta \boldsymbol{r}_j^{\,2}\right\rangle$, $\left\langle \Delta \boldsymbol{r}_i^{\,2}\Delta \boldsymbol{r}_j\right\rangle$, as well as their microstate values $\Delta \boldsymbol{r}_i^{\,3}$, $\Delta \boldsymbol{r}_j^{\,3}$, $\Delta \boldsymbol{r}_i\, \Delta \boldsymbol{r}_j^{\,2}$, $\Delta \boldsymbol{r}_i^{\,2}\Delta \boldsymbol{r}_j$, and of the linear terms $\Delta \boldsymbol{r}_i$ and $\Delta \boldsymbol{r}_j$. At a given time, if $\Phi_{ij} > 0$, then $\Delta E_{ij} < 0$, indicating that the combined energy of modes i and j is less than the sum of the single mode energies. Similarly, when $\Phi_{ij} < 0$, the coupling energy is positive, i.e., $\Delta E_{ij} > 0$. Thermodynamically, $\Delta E_{ij} < 0$ implies that the coupling of modes i and j gives off energy to the surroundings, and $\Delta E_{ij} > 0$ absorbs energy from the surroundings. Thus, the sign of Φ_{ij} can differentiate the case when two modes gain or lose energy simultaneously. Another possibility of coupling is obtained when one mode loses energy and the other one simultaneously gains energy. This type of coupling can not be determined by the sign of Φ_{ij}. One should examine the absolute energies E_i and E_j over the full trajectory for this purpose.

The average value $\langle \Phi_{ij} \rangle$ of Φ_{ij} reads as

$$\langle \Phi_{ij} \rangle = \frac{1}{2}\left[\left\langle \Delta \boldsymbol{r}_i \, \Delta \boldsymbol{r}_j^2 \right\rangle^2 + \left\langle \Delta \boldsymbol{r}_i^2 \Delta \boldsymbol{r}_j \right\rangle^2 \right] \qquad 3\text{-}23$$

and the energy coupling is

$$\langle \Delta E_{ij} \rangle = -\frac{1}{2\beta}\left[\left\langle \Delta \boldsymbol{r}_i \, \Delta \boldsymbol{r}_j^2 \right\rangle^2 + \left\langle \Delta \boldsymbol{r}_i^2 \Delta \boldsymbol{r}_j \right\rangle^2 \right] \qquad 3\text{-}24$$

Since $\langle \Phi_{ij} \rangle$ is always positive, equation (24) indicates that in third order coupling, $\langle \Delta E_{ij} \rangle \le 0$, i.e., coupling of two modes i and j gives off energy to the surroundings.

The cumulative coupling C_i of a given mode i to all of the other modes is obtained by summing $\langle \Phi_{ij} \rangle$ over all other modes j

$$C_i = \sum_j \langle \Phi_{ij} \rangle = \frac{1}{2} \sum_j \left[\left\langle \Delta \boldsymbol{r}_i \, \Delta \boldsymbol{r}_j^2 \right\rangle^2 + \left\langle \Delta \boldsymbol{r}_i^2 \Delta \boldsymbol{r}_j \right\rangle^2 \right] \qquad 3\text{-}25$$

Comparing equation (24) and (25), we conclude that within the third order approximation the cumulative coupling energy of mode i to the rest of the modes is always negative.

3.4. Results

In order to have a clear picture on the temperature dependence of coupling effects around physiological temperatures, simulations were performed at 273.15 K and at 310 K. Here, we presentresults for 273.15 K. Full data for the 310 K case are presented in the Supplementary section.

The RMSD plots for all non-hydrogen atoms with respect to their average are shown for 310 K and for 273.15 K in figure 2 for a stretch of 2.75 ns of the trajectory. Results indicate that an increase of about 40K approximately doubles the RMSD values.

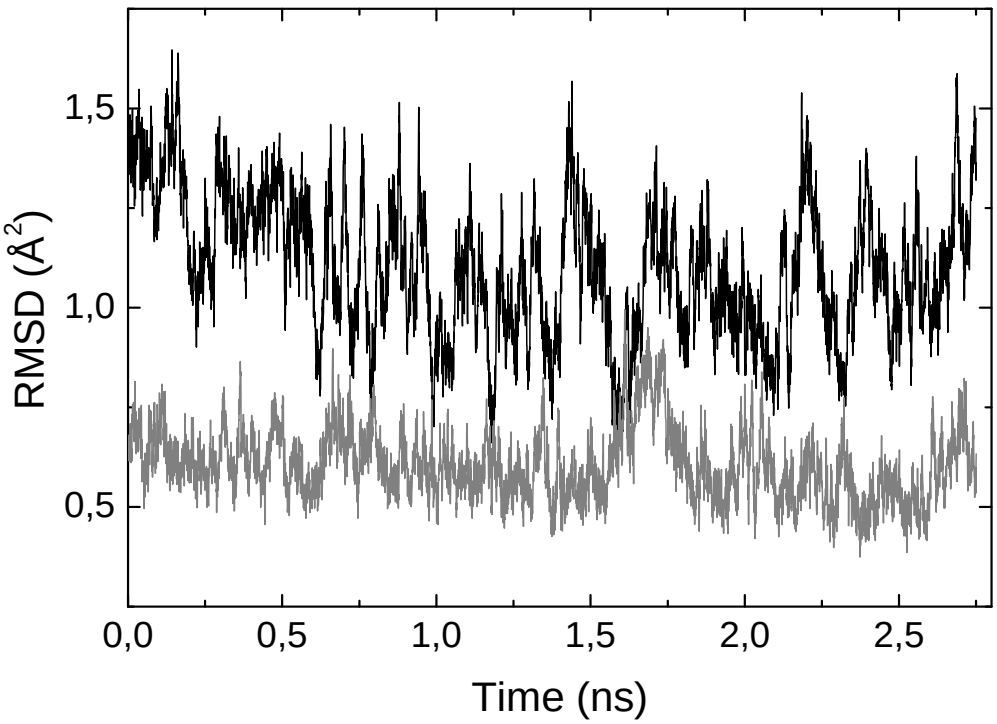


Figure 3-2. The RMSD values at 310 K (Black solid line) and at 273.15 K (Grey solid line)

3.4.1. Marginal Energies of Modes as a Function of Time and Mode Index

The energy of each mode is evaluated using the marginal energy formula via equation (14) up to the $v=17$ th order. As stated before, almost full convergence of the probability function is obtained when 17 terms are taken. The mean marginal energy of each mode is evaluated in the same way by equation (12) and are shown in figure 3 for 273.15 K. E_0 is set to zero, and the results are presented relative to $kT/2$, the harmonic component, as may be seen from equation (12). The figure shows that the majority of the modes have the harmonic energy. However, a few of the slower modes exhibit a significant negative deviation from the harmonic energy. The largest ten negative deviations from the harmonic approximation in decreasing order are for modes 1, 2, 8, 12, 3, 4, 18, 5, 6 and 10. In figure 1 the first six of them can be identified as highly unharmonic. Two of the faster modes, modes 310 and 445, exhibit positive deviation from the harmonic. In the following discussion, the harmonic component of the energy $kT/2$ will be subtracted and the corresponding modes will be named “positive modes” or “negative modes” depending on the sign of the energy.

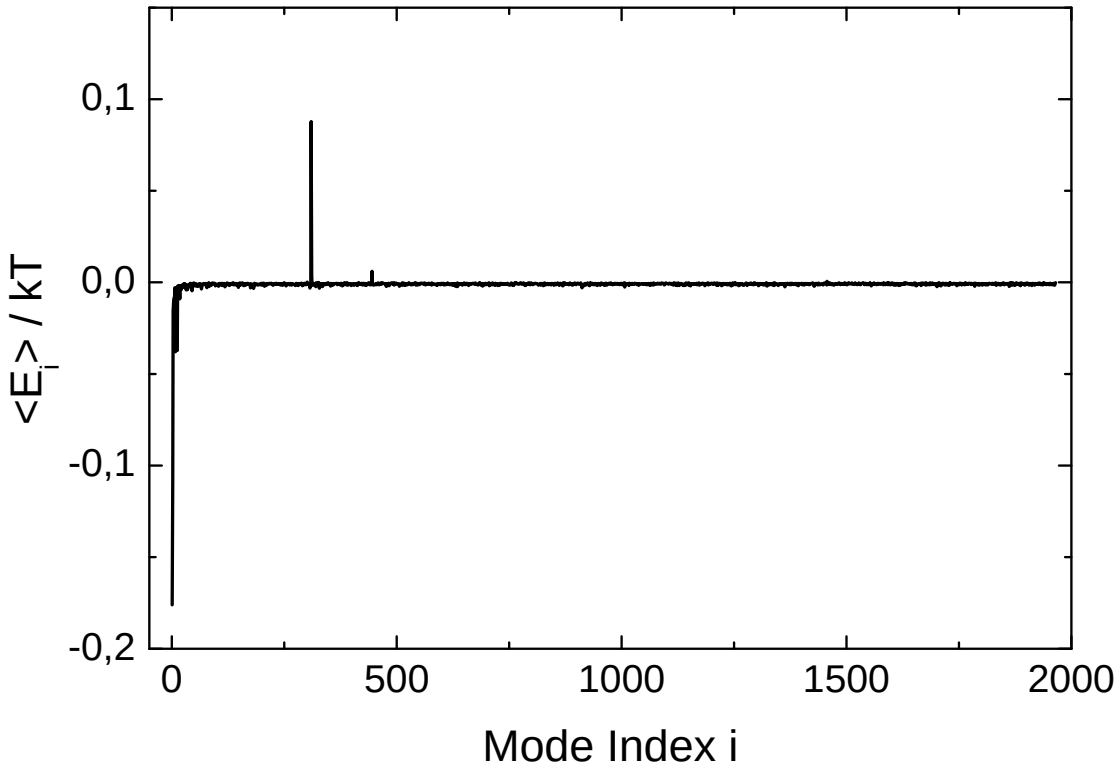
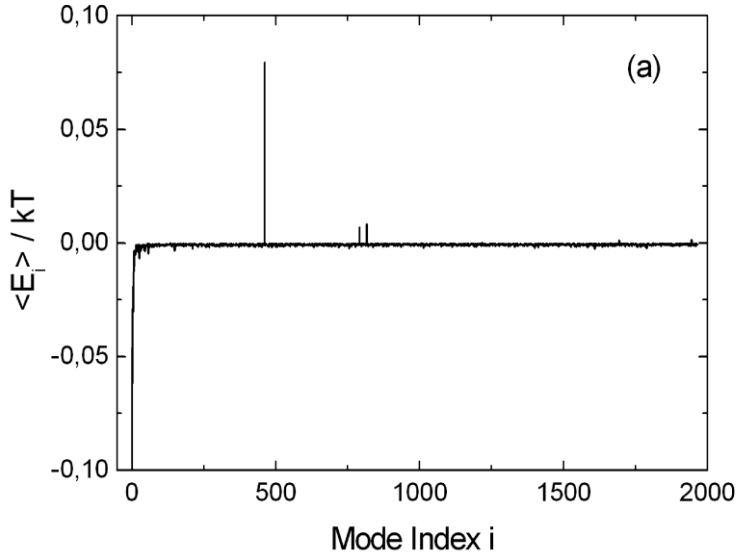
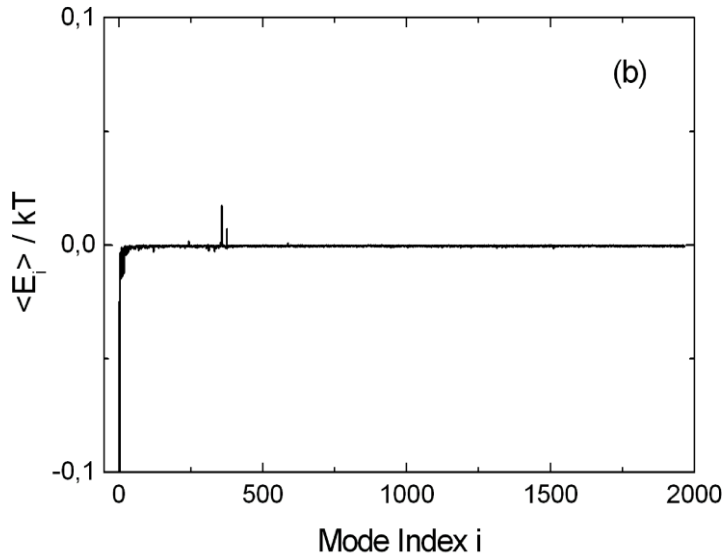


Figure 3-3. Energy $\langle E_i \rangle$ of each mode at 273.15 K relative to the harmonic energy $^{kT}/_{2}$ per mode. Mode index corresponds to an increasing frequency order. Mode 1 is the slowest mode whereas mode 1965 is the fastest mode.


(a)
0,10
0,05
0,00
-0,05
-0,10
<E_i> / kT
0
500
1000
1500
2000
Mode Index i


(b)
0,1
0,0
-0,1
<E_i> / kT
0
500
1000
1500
2000
Mode Index i

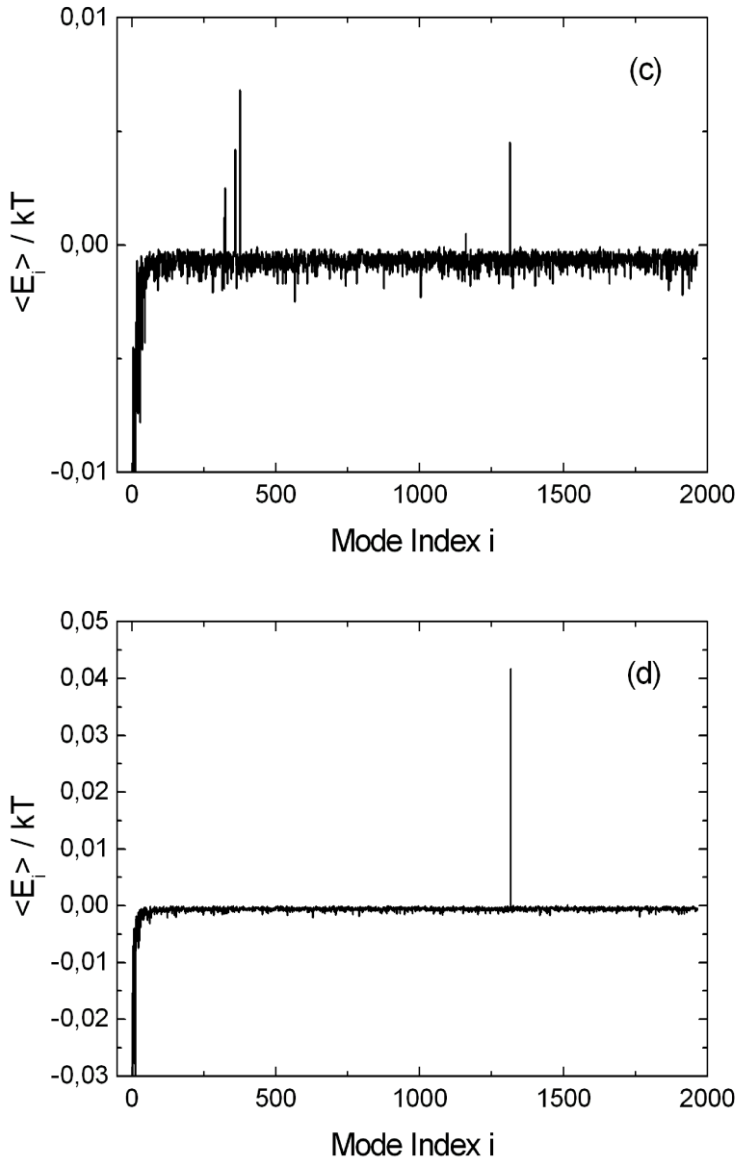


Figure 3-4. Energy $\langle E_i \rangle$ of each mode at 273.15 K relative to the harmonic energy $kT/2$ per mode at different stretches of the trajectory. **(a)** Stretch 1.25 - 4.25ns, **(b)** Stretch 5.75-11.25ns, **(c)** Stretch 8.25 - 11.5ns and **(d)** Stretch 32ns - 38ns.

In figure 4, energies at different time intervals are presented. Each of these stretches was aligned separately. After alignment principal component analysis was performed for each stretch. Using the obtained modes $\Delta \boldsymbol{r}$, the energies $\langle E_i \rangle$ which are depicted in figure 4 were

obtained. One sees from figures 3 and 4 that modes with high positive coupling fall into four regions in the mode spectrum: (i) 310, (ii) 445, (iii) 792 and (iv) 1319. As the trajectory extends to longer and longer times, new modes are not introduced, but the relative magnitudes of the coupled modes vary.

The energy jumps observed in figures 3 and 4 result from strong jumps in modal coordinates. In figure 5, the third order modal coordinates of the modes 310 and 445 are shown as a function of time. Both of these modes make large scale jumps during the simulation. The jumps take place during a short time interval as shown in the insets of the figures. However, a strong jump in the modal coordinate is not a necessary condition for coupling between modes. We will shown below that mode 310 is strongly coupled with mode 148 although the marginal energy of mode 148 does not deviate from the mean energy.

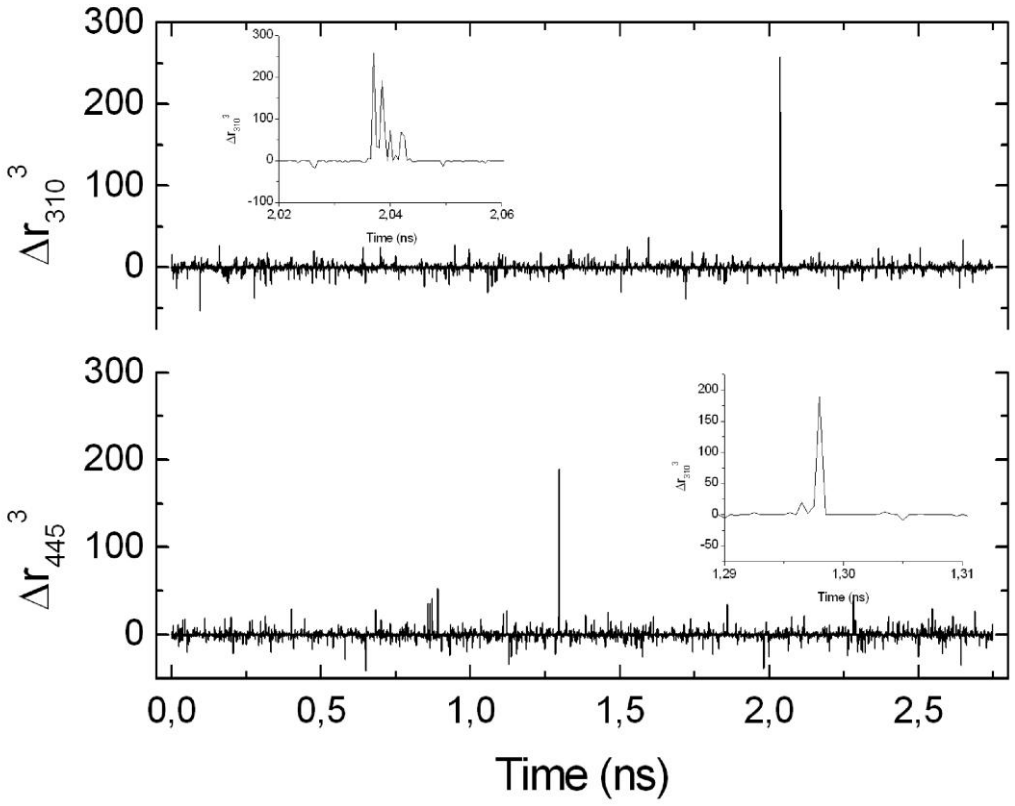


Figure 3-5. Third order moment $\Delta \mathbf{r}_i^3$ of mode 310 and 445 at 273.15 K

3.4.2. Third Order Moments and coupling of Modal Coordinates

The largest 500 $\left\langle \Delta r_i \Delta r_j^{\,2} \right\rangle$ values are shown for 237.15 K in figure 6 in a scatter plot with axes $\boldsymbol{r}_i$, $\boldsymbol{r}_j$, $\boldsymbol{r}_k$, with j=k. The largest 500 $\left\langle \Delta \boldsymbol{r}_i \Delta \boldsymbol{r}_j \Delta \boldsymbol{r}_k \right\rangle$ terms make a total of 2346 points due to the multiple presence of one type of coupling, i.e $\left\langle \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_2^{\,2} \right\rangle$ occurs as $\left\langle \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_2 \Delta \boldsymbol{r}_2 \right\rangle, \left\langle \Delta \boldsymbol{r}_2 \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_2 \right\rangle$ and $\left\langle \Delta \boldsymbol{r}_2 \Delta \boldsymbol{r}_2 \Delta \boldsymbol{r}_1 \right\rangle$. As can be clearly seen the third order couplings among the top 500 are clustered in the slow mode regime. The largest 750000 third order couplings $\left\langle \Delta r_i \Delta r_j \Delta r_k \right\rangle$ were determined. There are a total of 140520 values of $\left\langle \Delta r_i \Delta r_j \Delta r_k \right\rangle$ when multiplicities are removed. In figure 7, we present the absolute values of $\left\langle \Delta r_i \Delta r_j \Delta r_k \right\rangle$ in decreasing order as a function of rank, where the latter goes from 1 to 140520. It is clearly seen from this log-log plot that there is an extended power law region.

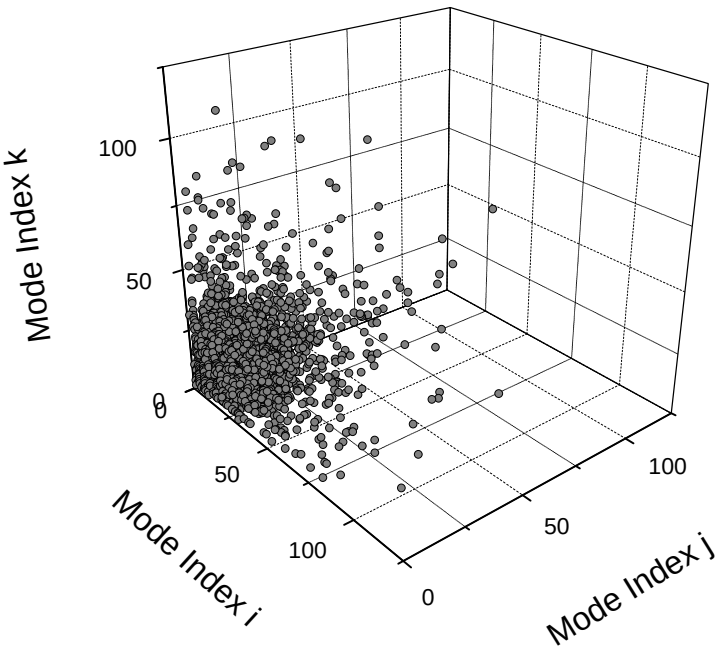


Figure 3-6. The scatter diagram of the largest 500 $\left\langle \Delta \boldsymbol{r}_i \Delta \boldsymbol{r}_j^{\,2} \right\rangle$ terms at 237.15 K. There are of 2346 points due to the multiple presence of one type of coupling.

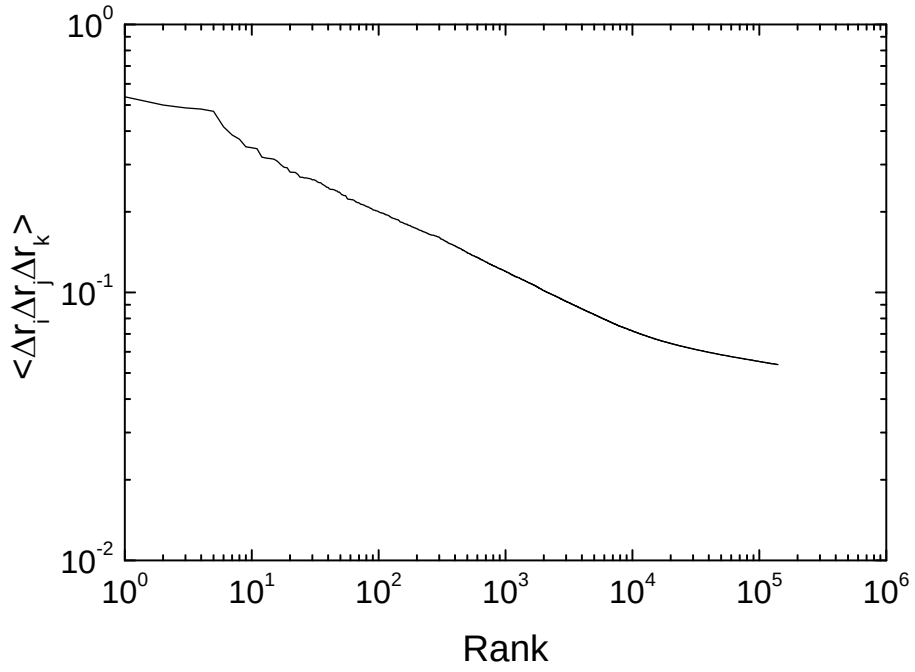


Figure 3-7. Distribution of $\langle \Delta r_i \Delta r_j \Delta r_k \rangle$ terms at 237.15 K. Rank goes from 1 to 140520**.**

Figure 6 shows that third order coupling is confined mostly to slower modes. As an example, we show here the third order coupling of mode 1 with other modes. The third order coupling of mode 1 with all other modes j occurs in two forms; $\langle \Delta \boldsymbol{r}_1^2 \Delta \boldsymbol{r}_j \rangle$ and $\langle \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_j^2 \rangle$. In figure 8, both of these coupling terms are shown as a function of index j. It can be seen that the third order coupling for mode 1 in the form of $\langle \Delta \boldsymbol{r}_1^2 \Delta \boldsymbol{r}_j \rangle$ decreases with increasing mode number. Most important couplings are observed with modes smaller than 200. For the couplings of type $\langle \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_j^2 \rangle$, on the other hand, an increase in the coupling terms with increasing mode numbers is observed after mode 100, which has its peak between modes 200 and 250. Subsequent to mode 300, coupling again decreases up to mode 400. A small but non-negligible coupling of the type $\langle \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_j^2 \rangle$ is present for all modes j as can be seen from the lower panel of figure 8.

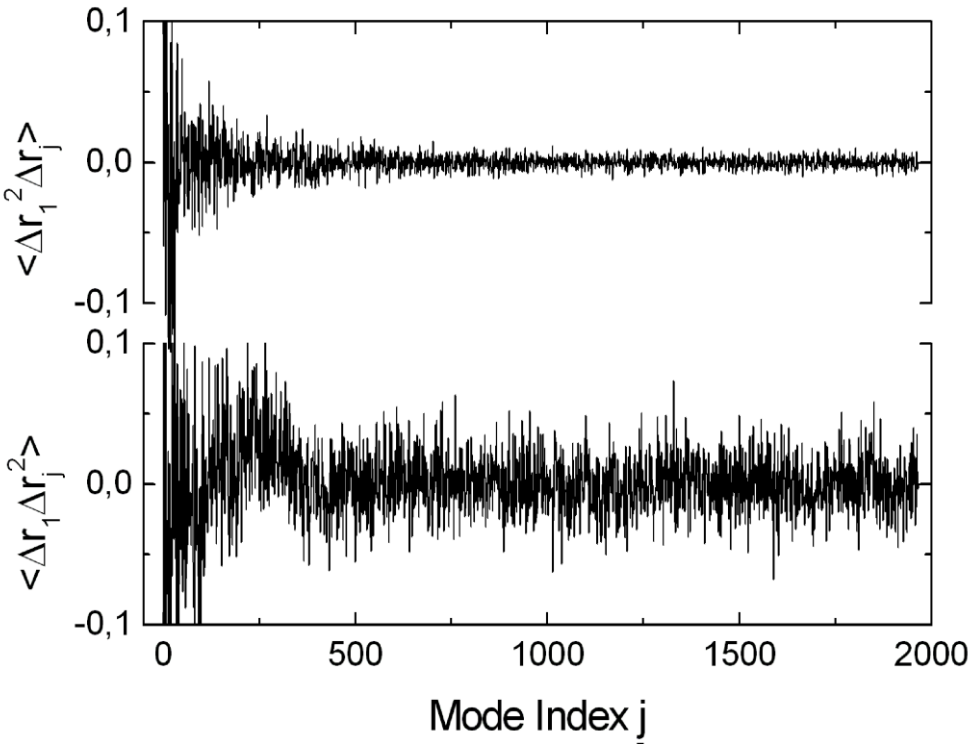


Figure 3-8. Third order coupling terms $\langle \Delta r_1^2 \Delta r_j \rangle$ and $\langle \Delta r_1 \Delta r_j^2 \rangle$ of the first mode to other modes.

In this paper, we have not gone into a detailed analysis of the coupling of three different modes, such as $\langle \Delta r_i \Delta r_j \Delta r_k \rangle$ which is also third order. Our calculations show that mixed three mode terms are generally larger for slow modes. For example one such term is $\langle \Delta r_1 \Delta r_2 \Delta r_8 \rangle$ which is 50% of the correlation $\langle \Delta r_1 \Delta r_2^2 \rangle$. This is shown in more detail in figure 9, where the ratio $\langle \Delta r_1 \Delta r_2 \Delta r_k \rangle / \langle \Delta r_1 \Delta r_2^2 \rangle$ is presented as a function of mode index k.

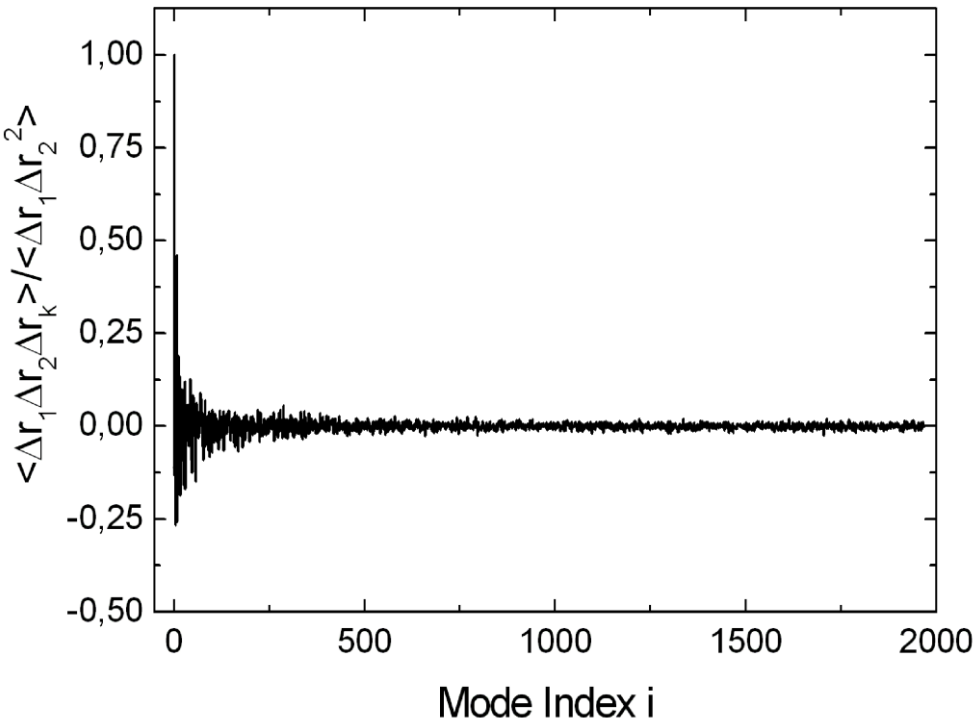
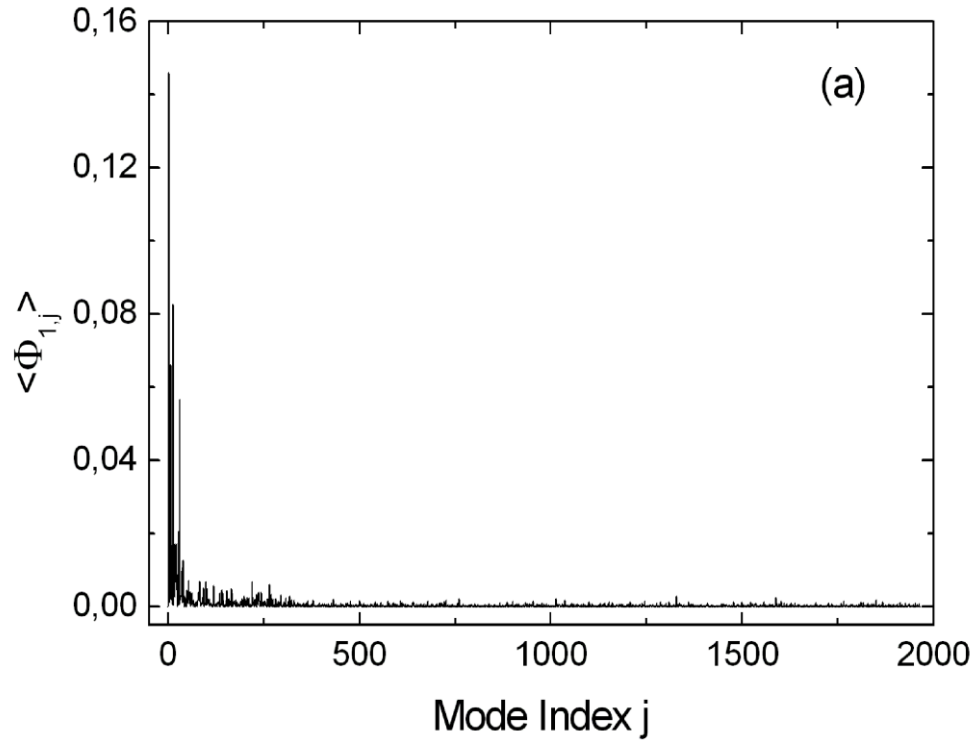
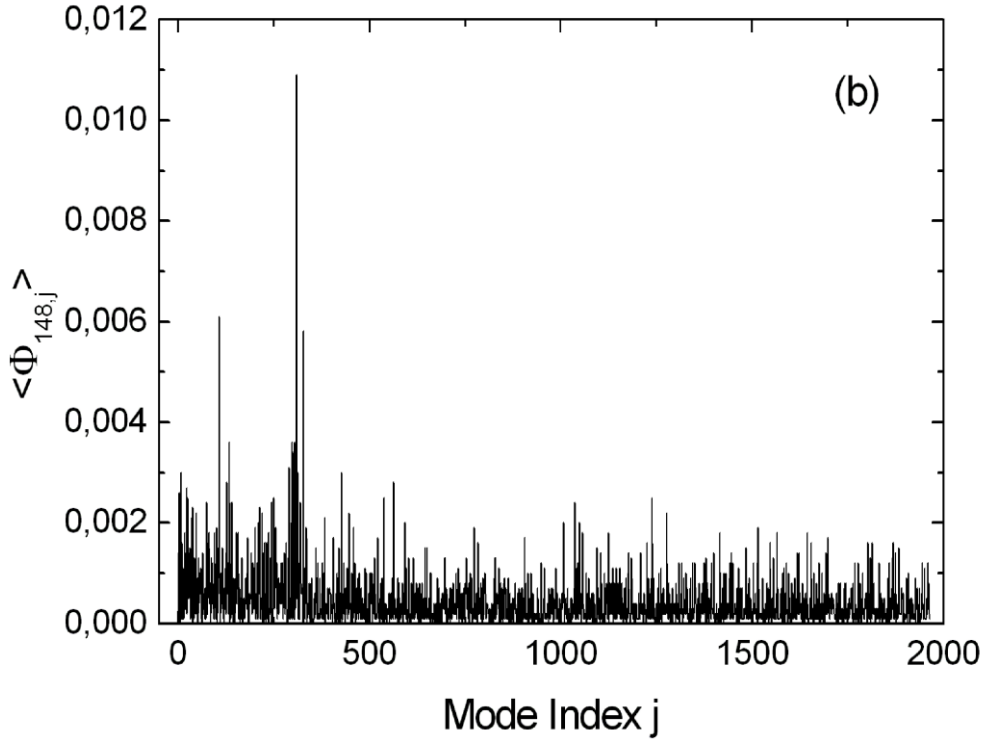


Figure 3-9. $\langle \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_2 \Delta \boldsymbol{r}_k \rangle / \langle \Delta \boldsymbol{r}_1 \Delta \boldsymbol{r}_2{}^2 \rangle$ as a function of mode index k.

3.4.3. Time Averaged Third Order Coupling $\langle \Phi_{ij} \rangle$

The form of equation (23) shows that the values of $\langle \Phi_{ij} \rangle$ are positive. The largest coupling terms $\langle \Phi_{ij} \rangle$ are among the slow mode pairs, among modes 1-31. The magnitude of coupling decreases with increasing mode number. However, among the faster modes, a strong coupling between modes 148 and 310 exists. We compare the relative magnitude of the couplings of these two modes with the couplings of the slowest mode in figure 10. In figure 10 (a), the coupling energy $\langle \Phi_{ij} \rangle$ between mode i = 1 and all other modes is shown. Strong couplings are mainly observed among mode 1 and other slow modes. This plot is essentially the sum of the two parts given by figure 8. In figure 10 (b) and (c), the coupling of modes i=148 and 310 with all other modes is presented. The strong coupling between modes 148 and 310 is evident from an inspection of figures (b) and (c).


(a)
0,16
0,12
0,08
0,04
0,00
<Φ1,j>
0
500
1000
1500
2000
Mode Index j


(b)
0,012
0,010
0,008
0,006
0,004
0,002
0,000
<Φ148,j>
0
500
1000
1500
2000
Mode Index j

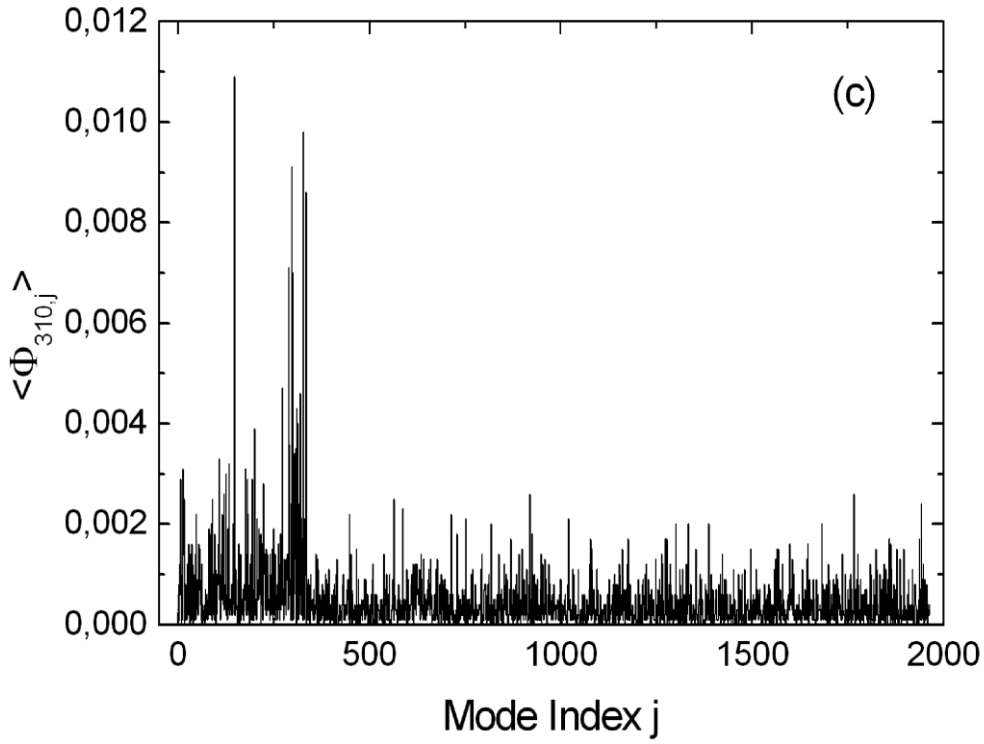


Figure 3-10. **(a)** Time averaged coupling of energy values $\langle \Phi_{1j} \rangle$ among the first mode and all other modes j **(b)** among mode 148 and all other modes j, $\langle \Phi_{148j} \rangle$ (c) among mode 310 and all other modes j, $\langle \Phi_{310j} \rangle$.

The second panel of figure 10 shows that mode 148 couples strongly with mode 310, and the third panel shows that mode 310 couples strongly with mode 148. Weaker coupling of these two modes to other modes also exist as may be seen from the figures.

In figure 11, we present the cumulative correlation C_i of a mode i with all other modes, defined as $C_i = \sum_j \langle \Phi_{ij} \rangle$.

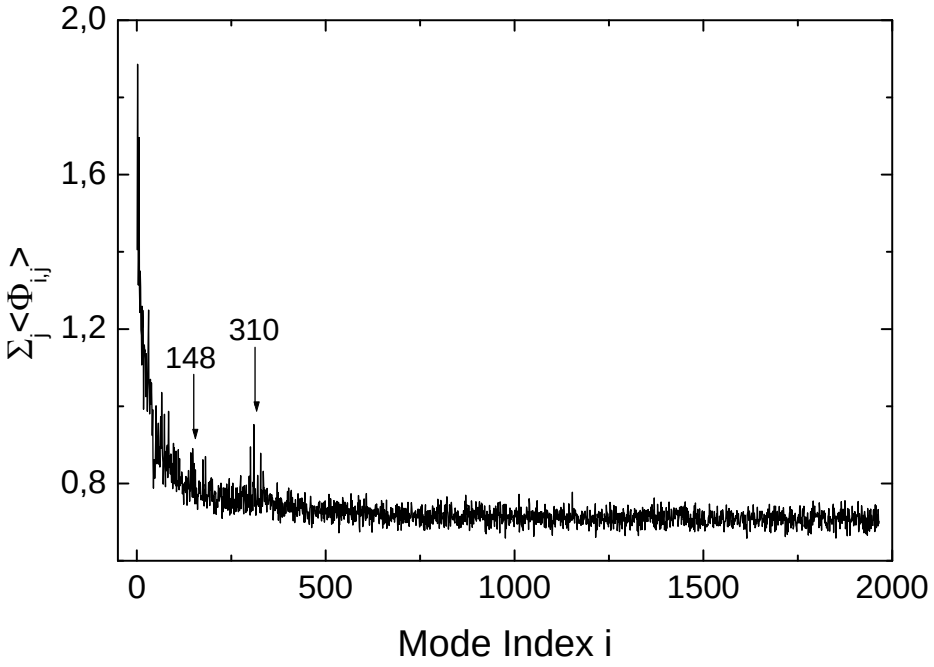


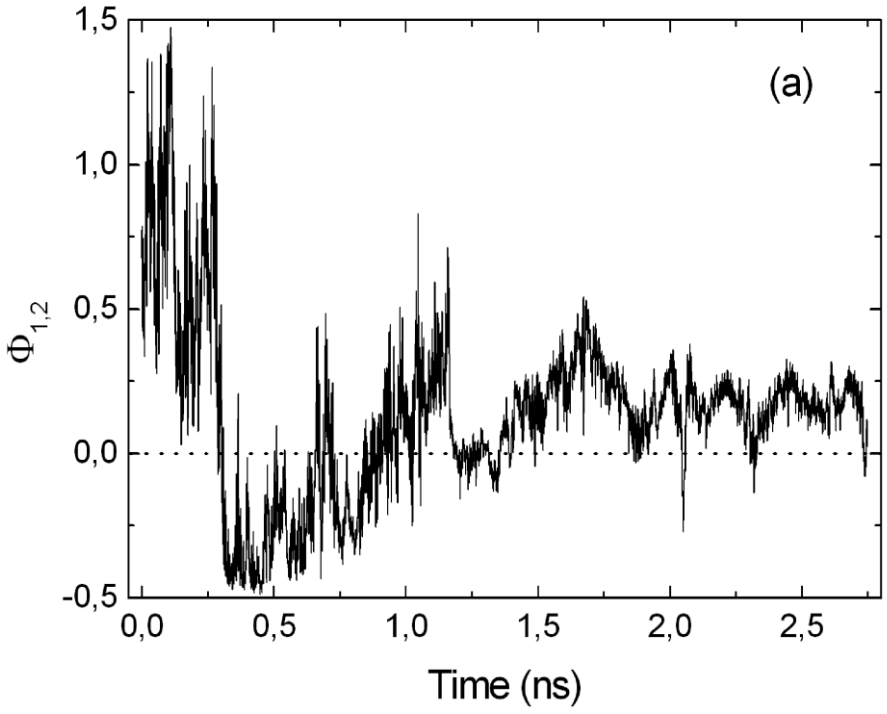
Figure 3-11. Cumulative coupling of mode i to all other modes.

In this figure, the peak at mode 310 is visible. More precisely, the peak consists of several closely spaced spikes around mode 310. The second group of modes around 148 may also be seen in the figure. Modes 149 and 310 are strongly coupled as will be discussed in more detail below.

3.4.4. Third Order Coupling Φ_{ij} as a Function of Time

There may be instances when the values of $\langle \Phi_{ij} \rangle$ become small due to the cancellation of positive and negative values along the trajectory when the average is taken. For this reason, we present the non averaged values of Φ_{ij} as a function of time. Again, the slower modes exhibit pronounced coupling with small modes. In figure 12 (a), for example, the coupling term $\Phi_{1,2}$ among mode 1 and 2 is shown. Slow variation of positive and negative couplings

along the trajectory is apparent. Overall, the coupling is positive when the full trajectory is considered. Coupling shows positive characteristic in the first 0.3 ns. After 0.3 ns coupling switches between positive and negative values. In figure 12 (b) the coupling term $\Phi_{2,31}$ is depicted. These two modes are selected because their coupling is strong compared to other slow modes. Coupling is strong for the first 0.29 ns and the period 0.86-1.17 ns. Other than these periods, an additional peak is observed at 0.660 ns. In figure 12 (c) the coupling term $\Phi_{148,310}$ is shown. An outstanding peak is observed at time instant 2.0375 ns. The maximum magnitudes of $\Phi_{1,2}$ and $\Phi_{2,31}$ are 1.5 and 3.5, respectively, whereas that of $\Phi_{148,310}$ is 18 as may be seen from the comparison of figures 12 (a), (b) and (c). The coupling among modes 1-2 and 2-31 is continuous and generally larger than the coupling among 148-310, except for the peaks. The strong coupling among modes 148 and 310 in the period 2.0375- 2.0450 ns is shown in the inset of figure 12 (c). Additional smaller peaks outside these periods are observed, mainly at 0.9975 ns and 0.4125 ns.



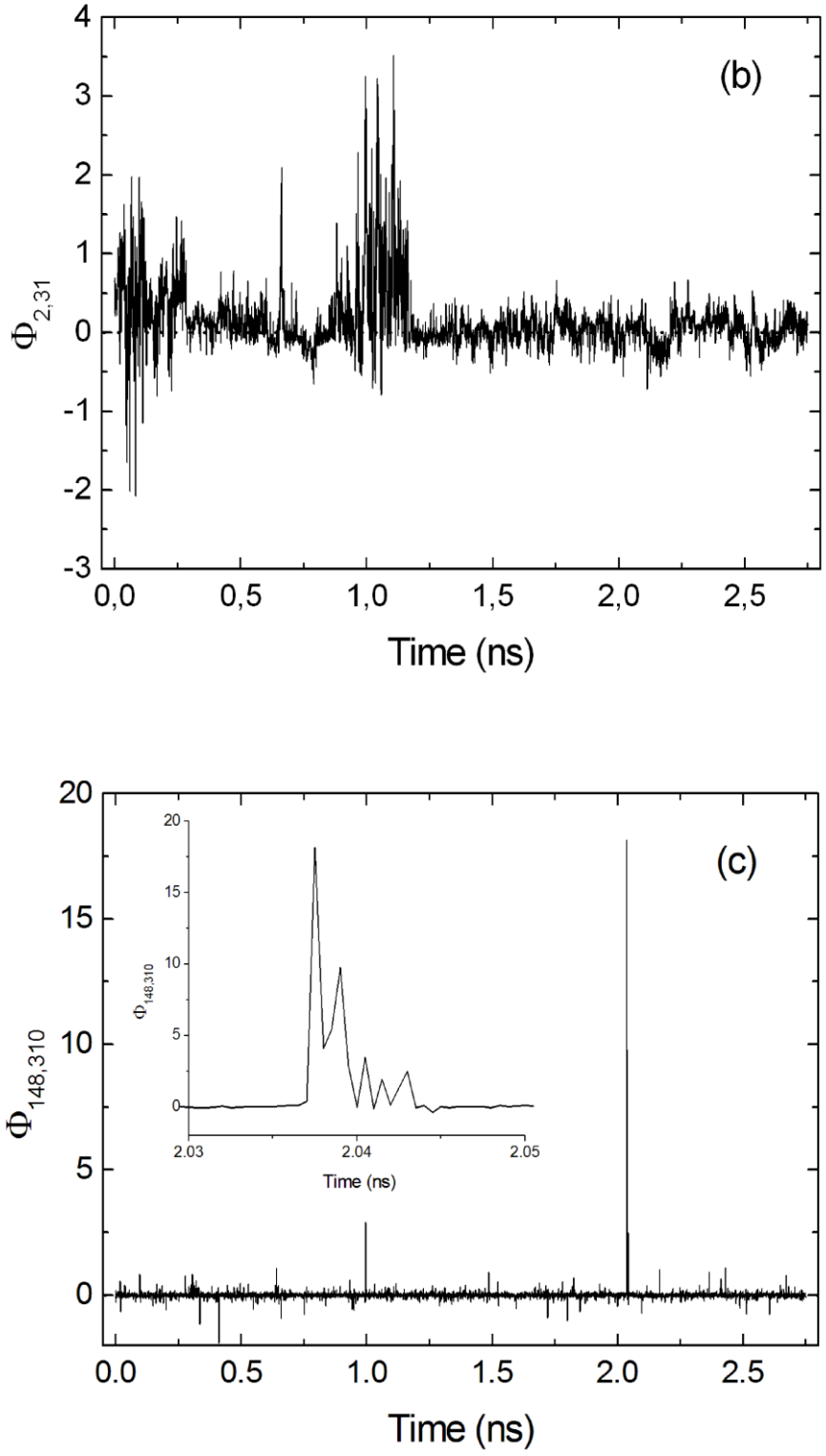


Figure 3-12. (a) Coupling term among mode 1 and 2 **(b)** Coupling term among mode 2 and 31 **(c)** Coupling term among mode 148 and 310, all as a function of time.

3.4.5. Coupling of Conformations

The couplings discussed above are all in modal space. Here, we investigate the effects of the stated couplings on the three dimensional conformations of the protein. The dot product of the eigenvectors of modes i and j gives the correlation of the positional fluctuations of these modes since,

$$\frac{\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T} \right\rangle}{\sqrt{\lambda_i \lambda_j}} = \boldsymbol{e}_i \boldsymbol{e}_j^{T} \qquad \text{3-26}$$

Here $\Delta \boldsymbol{R}_i$ is the fluctuation vector obtained by the back transformation from the modal space by keeping only the i^{th} mode. For simplicity, only the carbon alpha atoms will be considered.

In figure 13 (a) the correlations of the residues in mode 310 $\mathbf{e}_{310}\mathbf{e}_{310}^{T}$ are shown. The most active residues in this mode are 3, 4, 31 and 36. Especially residue 31 is correlated to a wide range of other residues. The 5 largest negatively correlated residue pairs of mode 310 are CYS4-SER6, CYS3-SER6, CYS4-GLY31, ARG17-GLY31 and CYS3- ILE33.

CYS3 and CYS32 are connected together by two parallel hydrogen bonds, resulting in a strong interaction. Both residues GLY31 and ILE33, which are negatively correlated to CYS3, are neighbor residues of CYS32. The constraint imposed by the CYS3-CYS32 bridge against the anticorrelated fluctuations of both GLY31 and ILE33 is expected to store energy into the system, which is observed as the positive deviation of the energy of mode 310 from the harmonic in figure 3.

In figure 13 (b) the correlations of the residues in mode 148 $\mathbf{e}_{148}\mathbf{e}_{148}^{T}$ are shown.

The most active residue in this mode is GLY 37 and it is anticorrelated with PRO5, LEU18, and ALA45.

In figure 13 (c) the correlations of the residues in mode 148 and mode 310 $\mathbf{e}_{148}\mathbf{e}_{310}^{T}$ are shown. GLY 37 in mode 148 is the key residue that is correlated with a large number of residues fluctuating in mode 310 such as GLY37- ARG17, GLY37- VAL8, GLY37-ILE33.

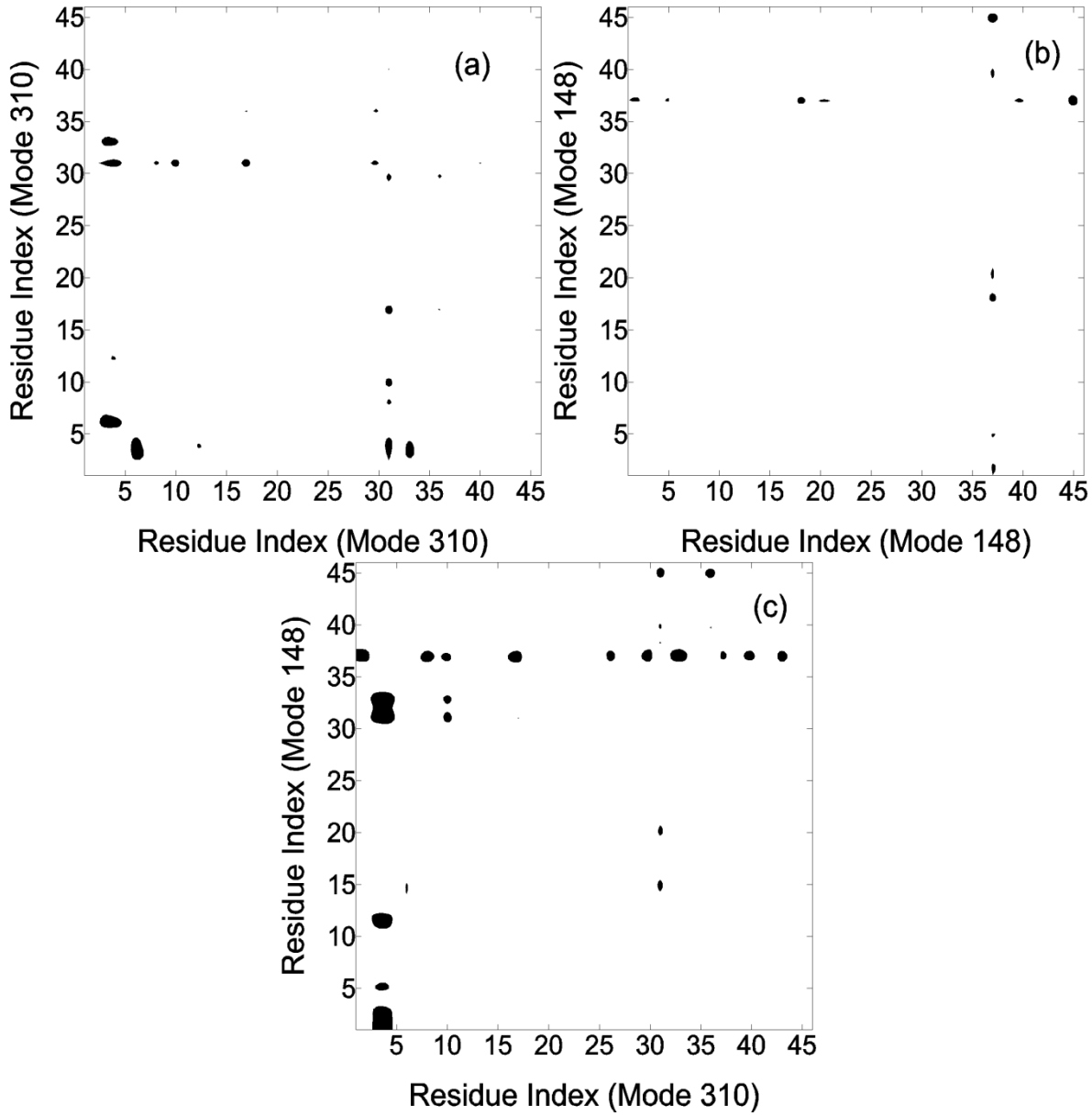


Figure 3-13. (a) Contour plot of $\mathbf{e}_{310}\mathbf{e}_{310}^{T}$ **(b)** Contour plot of $\mathbf{e}_{148}\mathbf{e}_{148}^{T}$ **(c)** Contour plot of $\mathbf{e}_{148}\mathbf{e}_{310}^{T}$.

3.5. Discussion

A direct method of identifying mode coupling from molecular dynamics trajectories would be to project the trajectory onto the eigenvectors obtained from principle component analysis, and calculate the energies for each mode. If a mode i and j are dependent, then, $E(i+j) \neq E(i)+E(j)$, where $E(i+j)$ is the energy calculated in the presence of the two modes. Softwares such as NAMD can calculate the energy of the system when the trajectory is given. Although this approach would lead to the energies of the modes for small fluctuations, such as that obtained at very low temperatures, it becomes unreliable when applied to proteins around physiological temperatures where fluctuations are dominant. For this reason, in the present paper, we take an alternate route where we search for signs of coupling by expanding the probability function of residue fluctuations into tensorial Hermite series expressed in principal components. This approach is valid for all fluctuation levels, and therefore acts as a good method of identifying coupling of all orders in modes at physiological temperatures. For clarity and brevity, we concentrate mostly on the third order coupling terms.

If the system were harmonic, the total energy would be 3NkT/2. This energy would be distributed to the 3N modes equally, with an energy of kT/2 per mode. Our calculations show that the slowest modes have energies below the harmonic. Slow modes exhibit coupling with other slow modes. Few fast modes have energies above the harmonic energy. It is these modes that exhibit strong coupling. When the third order coupling term is plotted as a function of time, we see that the coupling of fast modes is about an order of magnitude stronger than that of the slow modes.

One of the modes, mode 310, results from the presence of CYS-CYS bridge that restricts the fluctuations of the protein significantly. In mechanical terms, such a restriction would increase the strain energy of the protein, and this excess energy should transfer to another mode. In this case, this is achieved by coupling to mode 148 via the residue GLY37. The present approach makes it possible to identify the important modes and the residues that take part in these modes, and to estimate their contributions to the behavior of the native protein.

It is interesting to note that the coupling of modes is not active continuously throughout the full trajectory. The coupling of two modes is ephemeral; it persists for several picoseconds, and then disappears, but reappears at a later time during the trajectory. However, it is to be noted that the coupling involves only the same few modes throughout the trajectory.

3.6. Supplementary Material

This data contains results for s 2.75ns long stretch of the trajectory obtained at 310K.

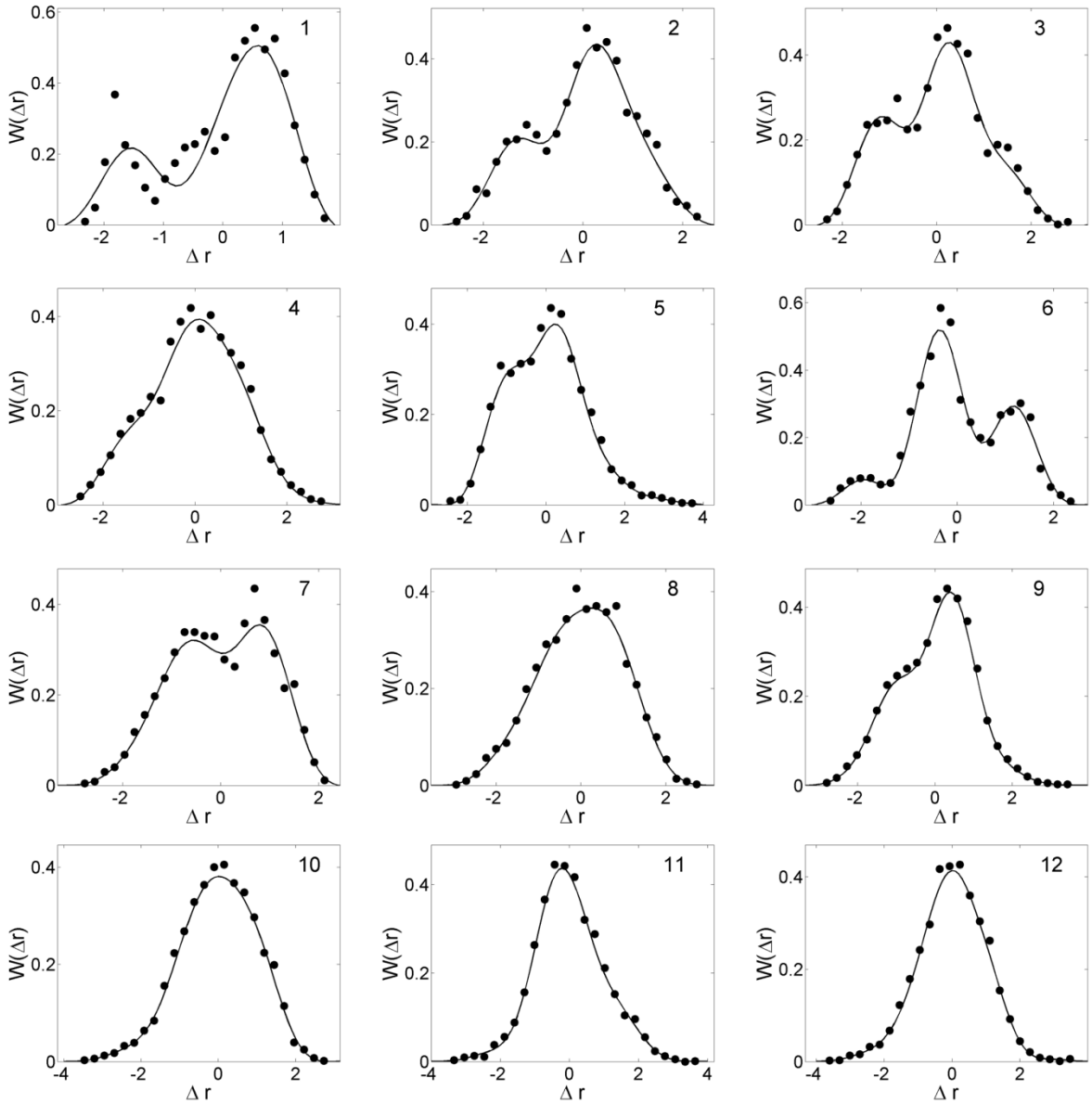


Figure 3-14. The normalized histograms of the modal coordinates $\Delta \boldsymbol{r}$ for the first 12 slow modes. Filled points are the calculated values and the lines through them are evaluated using equation (14).

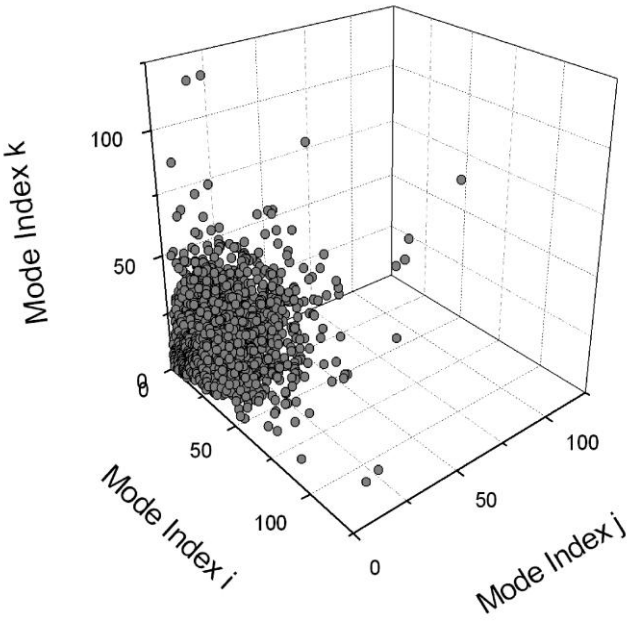


Figure 3-15. The scatter diagram of the largest 500 $\langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j \Delta \mathbf{r}_k \rangle$ terms at 310 K. There are of 2332 points due to the multiple presence of one type of coupling**.**

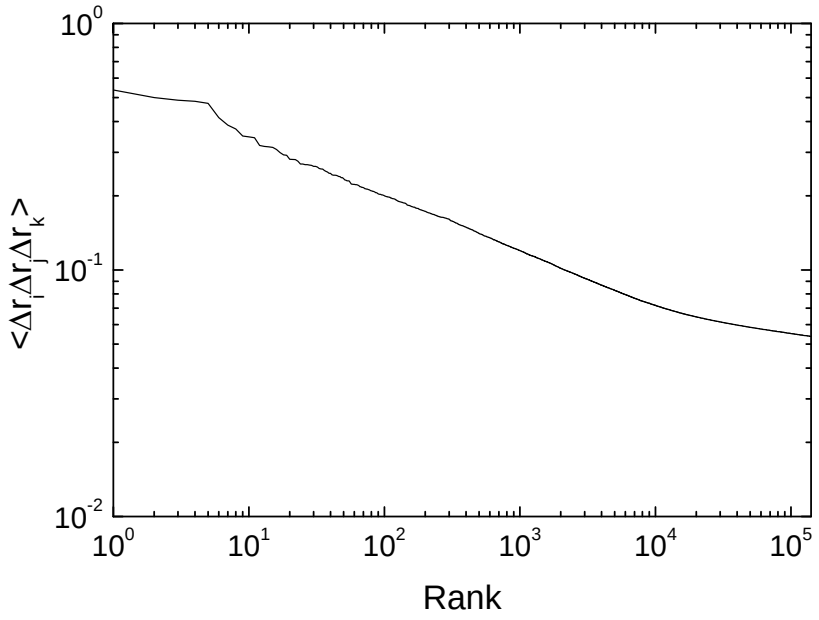


Figure 3-16. Distribution of $\langle \Delta r_i \Delta r_j \Delta r_k \rangle$ terms at 310 K. Rank goes from 1 to 140520.

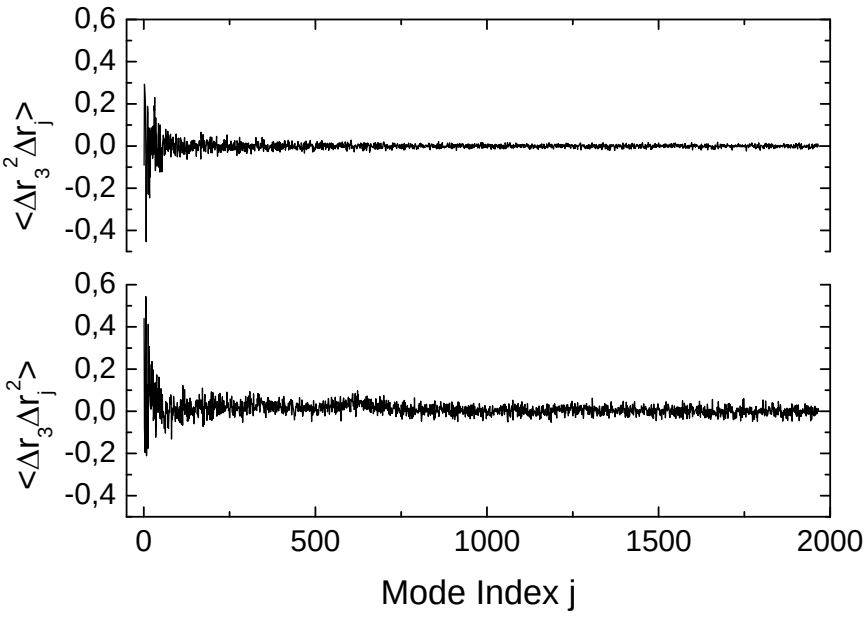


Figure 3-17. Third order coupling terms $\left\langle \Delta \mathbf{r}_3{}^2 \Delta \mathbf{r}_j \right\rangle$ and $\left\langle \Delta \mathbf{r}_3 \Delta \mathbf{r}_j{}^2 \right\rangle$ among the third mode and all other modes.

3.6.1. Coupling of Energies from $\langle \Phi_{ij} \rangle$ Values

The values of $\langle \Phi_{ij} \rangle$ are all positive. The largest average 10 coupling terms $\langle \Phi_{ij} \rangle$ are observed for the mode pairs 3- 6, 4-2, 10-5, 7-3, 1-6, 7- 6, 3-13, 25-13, 3-1 and 16-5.

In figure 18 (a), the energy coupling $\langle \Phi_{ij} \rangle$ between mode i = 3 and all other modes is shown. Strong couplings are mainly observed among mode 3 and other slow modes. In figure 5 (b) the same is shown for modes i=608.

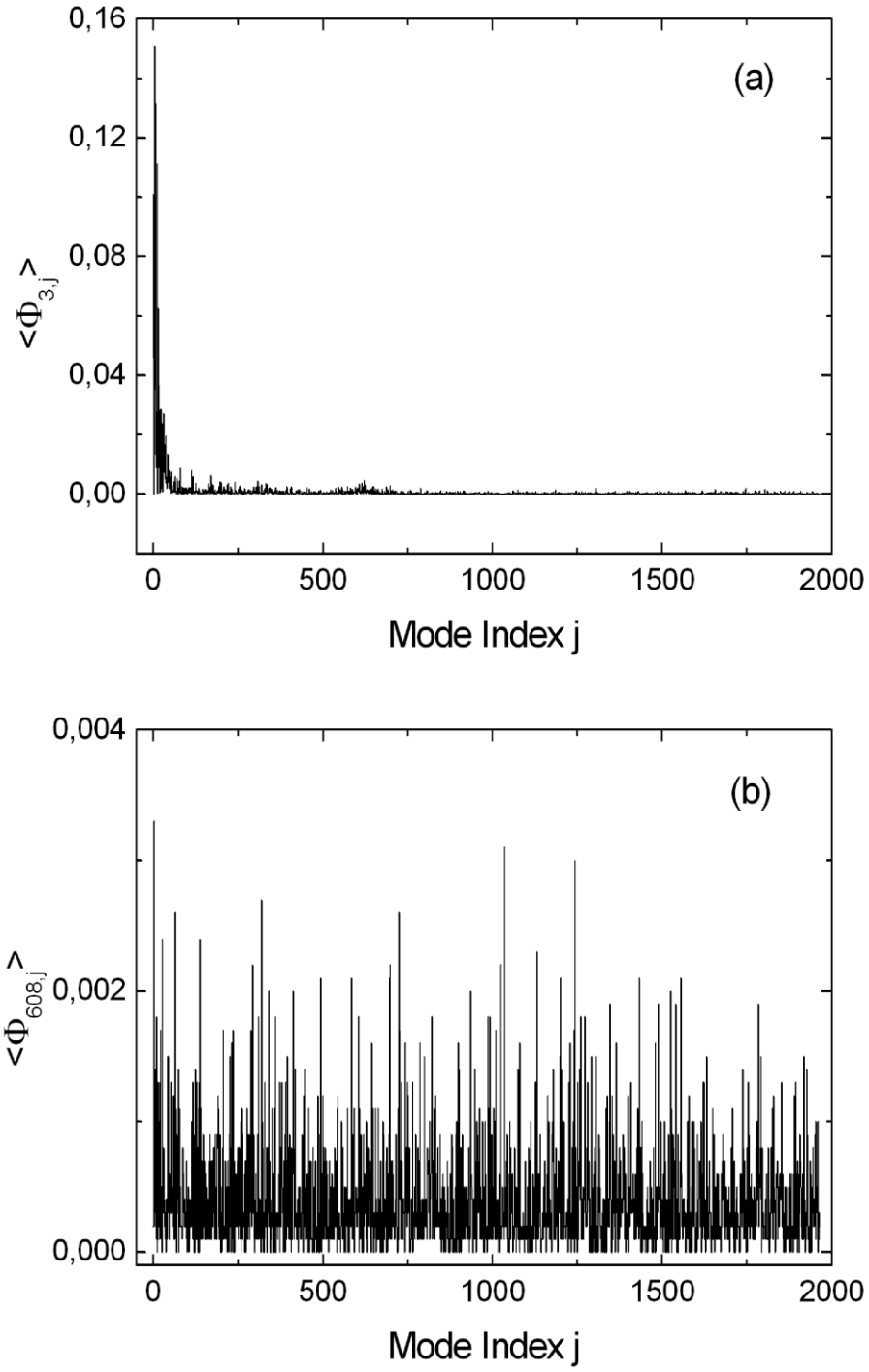


Figure 3-18. **(a)** Mean Coupling of energy values $\langle \Phi_{3,j} \rangle$ among the first mode and all other mode j **(b)** Mean Coupling of energy values $\langle \Phi_{608,j} \rangle$ among the 608^{th} mode and all other mode j.

The largest coupling term $\langle \Phi_{3,j} \rangle$ is observed among the mode 3 and 6 .The largest coupling term $\langle \Phi_{608,j} \rangle$ is observed among the mode 3 and 608.

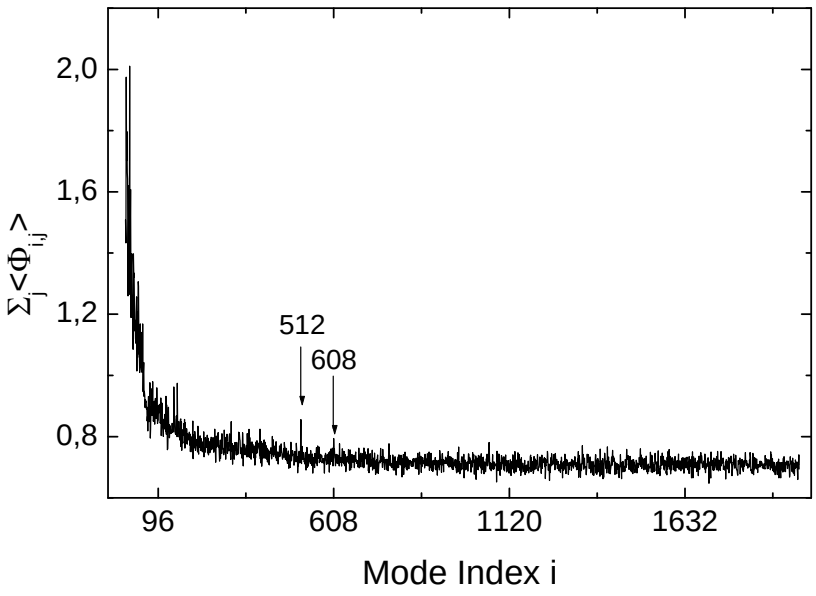


Figure 3-19. Cumulative coupling of mode i to all other modes.

In figure 20 (a) the third order coupling function $\Phi_{3,6}$ is shown over time. In figure 20 (b) the third order coupling function $\Phi_{3,608}$ is shown over time.

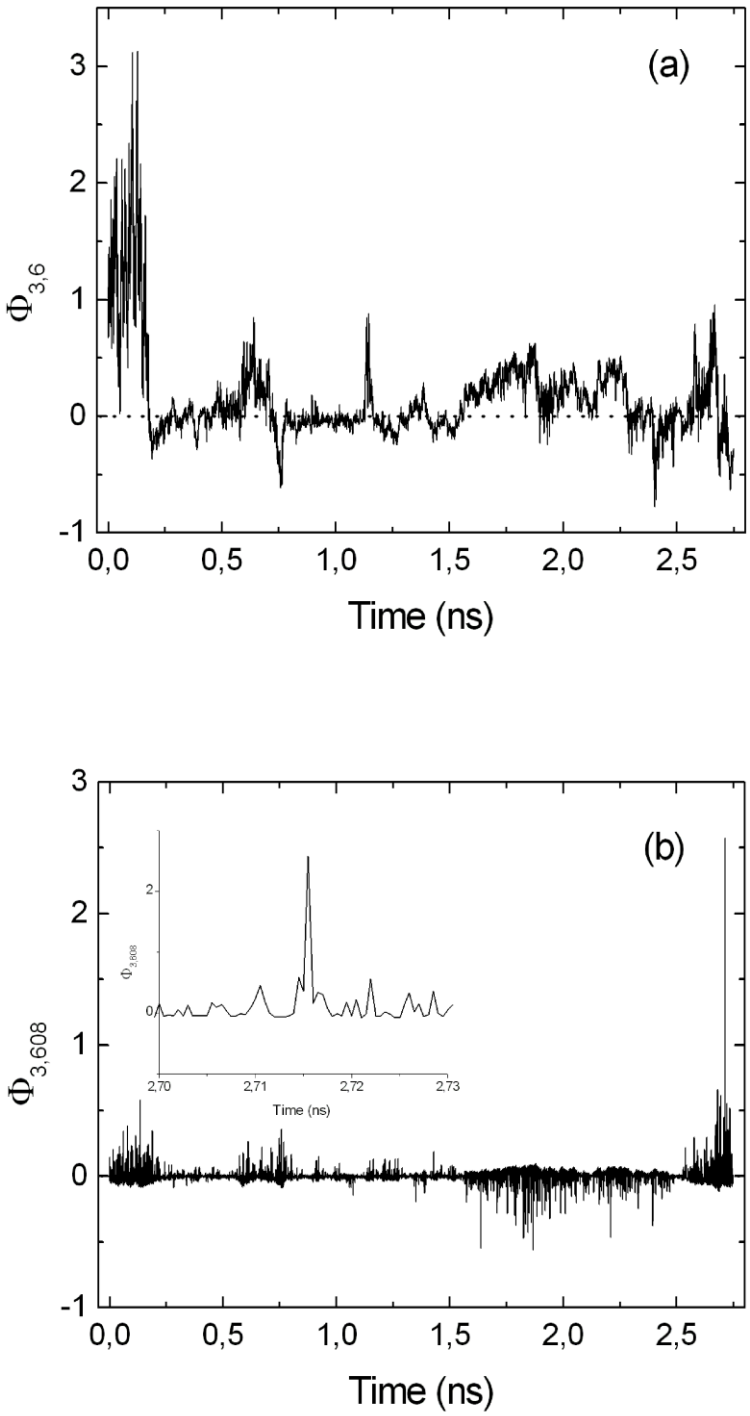


Figure 3-20. (a) Coupling term among mode 3 and 6 as a function of time **(b)** Coupling term among mode 3 and 608 as a function of time.

The energy of each mode is evaluated using the marginal energy formula via equation (14) up to the $v=17$th order and are shown for all modes in figure 21. The largest 2 positive peaks are at 608 and 512. The ten largest negative peaks are located at modes 1, 6, 3, 7, 2, 13, 5, 16, 9 and 4 with respective values of 0.3258, 0.4177, 0.4558, 0.4563, 0.4574, 0.4706, 0.4746 and 0.4748.

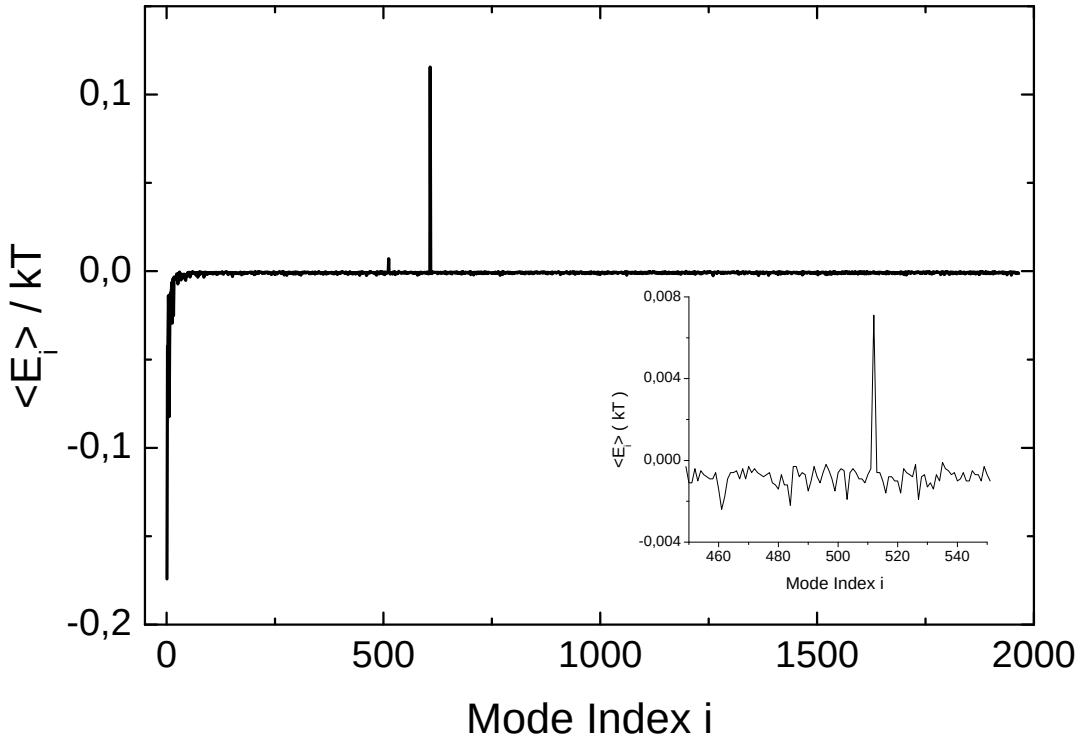


Figure 3-21. Energy $\langle E_i \rangle$ of each mode at 310 K.

Chapter 4

ANHARMONICITY, MODE-COUPLING AND ENTROPY IN FLUCTUATING NATIVE PROTEIN

4.1. Introduction

Several studies have pointed out that for the protein to perform its function, fluctuations of proteins have to be coupled [14, 15, 36]. The protein dynamics were investigated extensively under harmonic conditions in the literature [8, 9, 13, 20, 46, 52] and mode coupling were studied under harmonic conditions[16]. However, it is nowaday general knowledge that the slow modes are highly anharmonic [13, 49, 52, 53] . For this reason, the modal space were decomposed into an anharmonic essential subspace and a remaining harmonic Gaussian subspace [20, 40-45]. It was pointed out by Moritsugu et al.[16] and Leitner et al. [54] that coupling among modes is essential for the protein dynamics, such as energy transfer inside the protein. In this chapter the contributions of the anharmonicty and the mode couplings to the thermodynamic properties are investigated.

4.2. Modal Expansion and Beyond

Sampling the time evolution of a protein by using molecular dynamics reveals a multivariate probability distribution function (pdf) $f(\boldsymbol{\Delta R})$ for the deviations of atoms (assume there are N of them) from their mean positions, i.e. $\boldsymbol{\Delta R}_i = \hat{\boldsymbol{R}}_i - \boldsymbol{R}_i$, i = 1, . . . , 3N. We here adopt a coarse-grained representation of this pdf where only C_α atoms are considered, so that N is also the number of residues. Accordingly, $\boldsymbol{R}$ i are the mean C_α coordinates corresponding to the average configuration of the protein during the part of the trajectory that is used for the calculations.

4.2.1. Hermite Expansion

Since the deviations from this free energy minimum are expected to be harmonic for sufficiently small amplitudes, Hermite polynomials - which are orthogonal *wrt* a Gaussian weight function - constitute a natural basis for representing $W(\boldsymbol{\Delta R})$. First, following, Ref. [49, 55], we perform the transformation

$$\boldsymbol{\Delta r} = \left\langle \boldsymbol{\Delta R}\boldsymbol{\Delta R}^T \right\rangle^{-1/2} \boldsymbol{\Delta R} \qquad 4\text{-}1$$

This diagonalizes the covariance matrix $\boldsymbol{\Gamma} = \left\langle \boldsymbol{\Delta R}\boldsymbol{\Delta R}^T \right\rangle$ ($\langle\cdot\rangle$ represents averaging over the trajectory) and would give the normal modes of the protein if fluctuations were harmonic. Otherwise, the distribution function for $\boldsymbol{\Delta r}$ in its most general form, can be expressed as [50]

$$W(\Delta \boldsymbol{r}) = (2\pi)^{-3n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{3n} \Delta \boldsymbol{r}_i^{\,2} \right] \cdot \left[1 + \sum_{\nu=3}^{\infty} C_\nu \cdot H_\nu(\Delta \boldsymbol{r}) \right] \qquad 4\text{-}2$$

where $C_\nu = (\nu!)^{-1} \langle H_\nu \rangle$ (constant) and H_ν (derived below) are tensors of rank ν, and the dot product refers to $\sum_{ij\ldots k} C_\nu^{\,ij\ldots k} H_\nu^{\,ij\ldots k}$. The fluctuations $\boldsymbol{\Delta r}$ in this mode space spanned by the eigenvectors of $\boldsymbol{\Gamma}$ are meanless, i.e., $\langle \boldsymbol{\Delta r}_i \rangle = 0$, and decoupled at the lowest (second) order, i.e.,

$$\left\langle \boldsymbol{\Delta r}_i^{\,T} \boldsymbol{\Delta r}_j \right\rangle = \boldsymbol{\delta}_{ij} \qquad 4\text{-}3$$

4.2.2. Anharmonicity, Mode-coupling and Entropy in a Fluctuating Native Protein

The modes are numbered in the order of the corresponding nonzero eigenvalues of $\boldsymbol{\Gamma}$,

i.e., smaller ranking modes are also the slowest ones. A purely harmonic model is given by $C_\nu = 0, \forall \nu$. Note that, the atomic fluctuations corresponding to a given mode can easily be calculated by setting to zero all the eigenvalues except the one of interest, followed by a back transformation of equation (1).
Tensor Hermite polynomials can be obtained by successive differentiation using Rodrigues' formula:

$$H_\nu^{ij...k}(\Delta \boldsymbol{r}) = \frac{(-1)^\nu}{g(\Delta \boldsymbol{r})} \nabla^{ij...k} g(\Delta \boldsymbol{r}) \qquad 4\text{-}4$$

Above, $g(x) = (2\pi)^{3n/2} \exp\left[-x^2/2\right]$ is the multidimensional Gaussian distribution and $\nabla^{ij...k} = \nabla^i \nabla^j .. \nabla^k$ is the gradient tensor with $\nabla^i = \partial / \partial x_i$. The tensor coefficients that appear in $W(\Delta \boldsymbol{r})$ follow from the orthogonality relation as

$$C_\nu = \frac{1}{\nu!} \int_{-\infty}^{\infty} H_\nu(\Delta \boldsymbol{r}) W(\Delta \boldsymbol{r}) d\Delta \boldsymbol{r} = \frac{\langle H_\nu(x) \rangle}{\nu!} \qquad 4\text{-}5$$

Therefore, the problem reduces to obtaining the expectation values of the polynomial tensor elements for the system. At the lowest nonvanishing order they read

$$\begin{aligned} H_3^{111}(x) &= x_1^3 - 3x_1 \\ H_3^{112}(x) &= x_1^2 x_2 - x_2 = H_3^{121}(x) = H_3^{211}(x) \\ H_3^{123}(x) &= x_1 x_2 x_3 = H_3^{213}(x) = ... \end{aligned} \qquad 4\text{-}6$$

Higher order tensor elements can be calculated using a diagrammatic technique. A graphical representation of H_4 in one and two dimensions is given in figure 1.

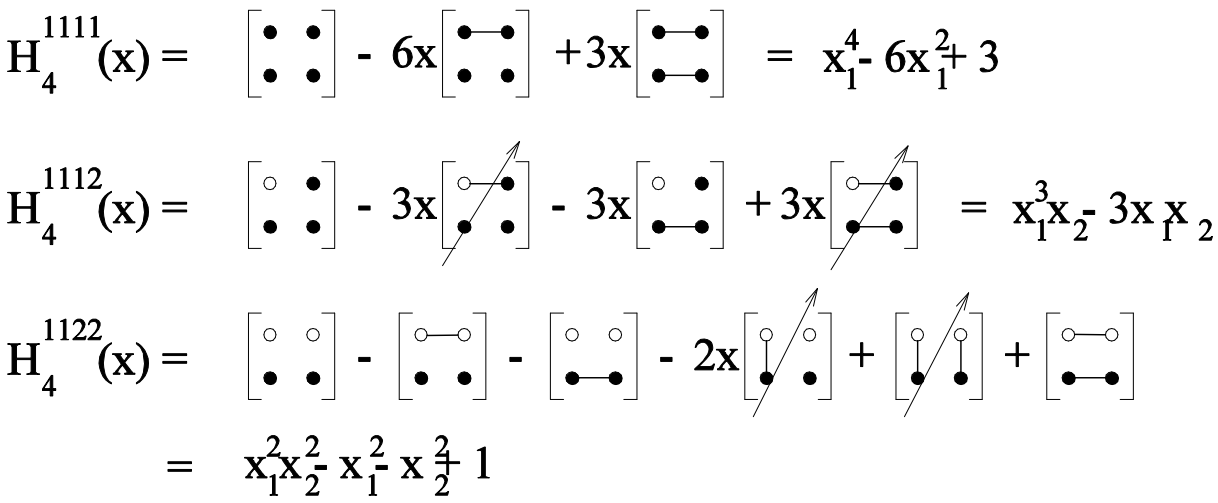


Figure 4-1. Graphical representation of $H_4(x)$ tensor elements in two dimensions.

Terms that vanish by virtue of equation (3) are crossed.

The inclusion of mode-coupling necessitates consideration of mixed indices (nondiagonal tensor elements). Here, we focus on the coupling between mode pairs and ignore threesome and higher order mixing, i.e., we consider only the bi-polynomials $H_\nu^{i_1 i_2 \ldots i_\nu}(\Delta \boldsymbol{r}_k, \Delta \boldsymbol{r}_l)$ with $i_m \in \{k,l\}, k,l = 1,2,.....,3n$. At first sight, estimating the contribution of mode-coupling even at this lowest level appears to be a formidable task, because the number of distinct expectation values to be extracted from the data grows combinatorially. We show below that, the factorization property of the off-diagonal tensor elements and the orthogonality of the modes at the second order bring a significant reduction in complexity, which we exploit to investigate the impact of anharmonicity and mode-coupling separately on the protein dynamics.

We first observe that, the value of a tensor element $H_\nu^{i_1 i_2 \ldots i_\nu}(\Delta \boldsymbol{r})$ does not depend on the order of the indices due to the commutativity of the gradient operator, $\nabla_k \nabla_l - \nabla_l \nabla_k = 0$. Therefore,

$$H_\nu^{i_1 i_2 \ldots i_\nu}(\Delta \boldsymbol{r}) = H_\nu^{p}(\Delta \boldsymbol{r}_k, \Delta \boldsymbol{r}_l)$$

where p is the number of indices equal to k (and the remaining $\nu = p$ indices are equal to l). The fact that the covariance matrix in the normal basis is diagonal further implies that

$$H_\nu^p(\Delta \boldsymbol{r}) = H_p(\Delta \boldsymbol{r}_1) \times H_{\nu-p}(\Delta \boldsymbol{r}_2) \qquad 4\text{-}7$$

as is also evident from the Rodrigues's formula in equation (4).

4.3. Anharmonicity vs Mode-coupling

Combining equation (5) and equation (7), the Hermite expansion in equation (2) can be cast into the
following form:

$$W(\Delta \boldsymbol{r}) = (2\pi)^{-3n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{3n} \Delta \boldsymbol{r}_i^{\,2}\right] \cdot \left[1 + \sum_i \sum_{\nu=3}^{\infty} \frac{1}{\nu!} \left\langle H_\nu(\Delta \boldsymbol{r}_i)\right\rangle \cdot H_\nu(\Delta \boldsymbol{r}_i) + \sum_i \sum_{\nu=3}^{\infty} \frac{1}{\nu!} \sum_{p=1}^{\nu-1} \binom{\nu}{p} \left\langle H_p(\Delta \boldsymbol{r}_i) H_{\nu-p}(\Delta \boldsymbol{r}_j)\right\rangle H_p(\Delta \boldsymbol{r}_i) H_{\nu-p}(\Delta \boldsymbol{r}_j) + \sum_{i \neq j \neq k} \ldots\right] \qquad 4\text{-}8$$

The first term in equation (8) corresponds to a purely harmonic model given by the Gaussian probability distribution

$$W_o(\Delta \boldsymbol{r}) = (2\pi)^{-3n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{3n} \Delta \boldsymbol{r}_i^{\,2}\right] \qquad 4\text{-}9$$

This is the starting point for most of the past studies on protein fluctuations [9].
The next term in equation (9) is appreciable when the fluctuations are anharmonic, but gives no information about mode-coupling. In fact, the most general mode-amplitude distribution of an anharmonic model composed of decoupled modes is

$$W_1(\Delta \boldsymbol{r}) = (2\pi)^{-3n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{3n}\Delta \boldsymbol{r}_i^{\,2}\right] \cdot \prod_i \left[1 + \sum_{\nu=3}^{\infty} \frac{1}{\nu!} \left\langle H_\nu(\Delta \boldsymbol{r}_i)\right\rangle \cdot H_\nu(\Delta \boldsymbol{r}_i)\right] \qquad 4\text{-}10$$

The approximation to the true distribution given in equation (10) is named W_1 in order to remind the reader that it qualitatively improves on the Gaussian approximation W_o of equation (9). The difference between the full pdf given in equation (8) and the approximation W_1
is the mode-coupling corrections such as

$$\left\langle H_p(\Delta \boldsymbol{r}_i) \times H_{\nu-p}(\Delta \boldsymbol{r}_j)\right\rangle - \left\langle H_p(\Delta \boldsymbol{r}_i)\right\rangle \times \left\langle H_{\nu-p}(\Delta \boldsymbol{r}_j)\right\rangle \neq 0$$

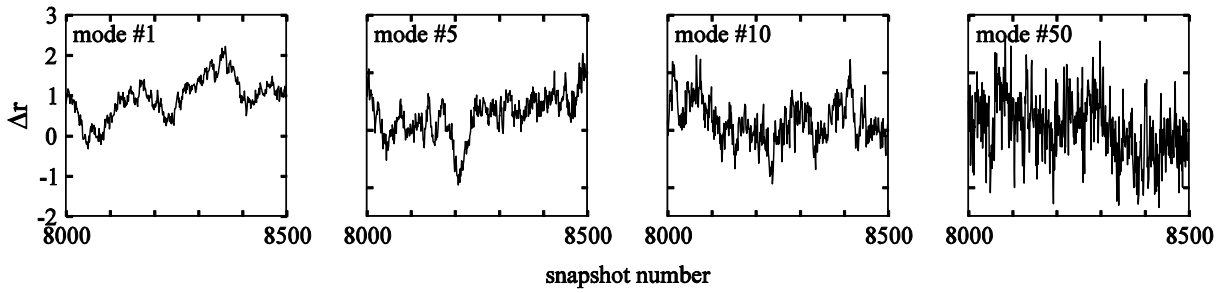


Figure 4-2. Time plots of the slowest 1st, 5th, 10th, and 50th modes between timesteps 8000-8500.

and higher order cumulants. Note that, marginal distributions are transparent to such corrections

$$W(\Delta \boldsymbol{r}_i) \equiv \int_0^{\infty} d\Delta \boldsymbol{r}_j W(\Delta \boldsymbol{r}) \qquad 4\text{-}11$$

as a merit of the orthogonality relation in equation (5). Therefore, even if the marginal distributions are reproduced to good accuracy, the multidimensional free-energy landscape of the protein may still be very different from that implied by a model based on equation (10). We demonstrate below that this is the case for the protein Crambin. To this end, we improve the approximation in equation (10) one step further and approximate $W(\Delta \boldsymbol{r})$ by

$$W_2(\Delta \boldsymbol{r}) \equiv W(\Delta \boldsymbol{r}) - \sum_{i \neq j \neq k} \cdots \qquad 4\text{-}12$$

, i.e., the part of the Hermite expansion spelled out in equation (9) which takes into account the mode-coupling corrections at the lowest order they appear while ignoring cubic and higher-order terms. The magnitude of the neglected terms can be estimated through other means presented at the end of our paper.

4.4. Results

4.4.1. Crambin Molecular Dynamics: a Test Ground

Crambin (Protein Data bank code 1EJG.pdb) was selected as a test ground since it is a relatively small protein and its dynamics is widely studied [41, 46, 47]. The 46 residue Crambin consists of 657 atoms. Taking only the alpha carbons into account a set of 138 modes were obtained, six of which have a zero eigenvalue corresponding to the translation of the center of mass and three global rotational degrees of freedom around it.
All molecular dynamics simulations were performed for an N,P,T ensemble in explicit solvent (water) at 310 K using NAMD 2.5 package with CHARMM27 force field. The protein was solvated in a waterbox of 15Å cushion and periodic boundary conditions were applied. Ions were added in order to represent a more typical biological environment. Langevin dynamics was used to control the system's temperature and pressure. All atoms were coupled to the heat bath of temperature of 310 K. A time step of 1fs was used. Nonbonded and electrostatic forces were evaluated at each time step.

In order to keep all degrees of freedom no rigid bonds were used. All structures were translated so that their centers of mass are positioned at the origin and rotated to yield the best mass weighted RMSD agreement with the initial structure [49].

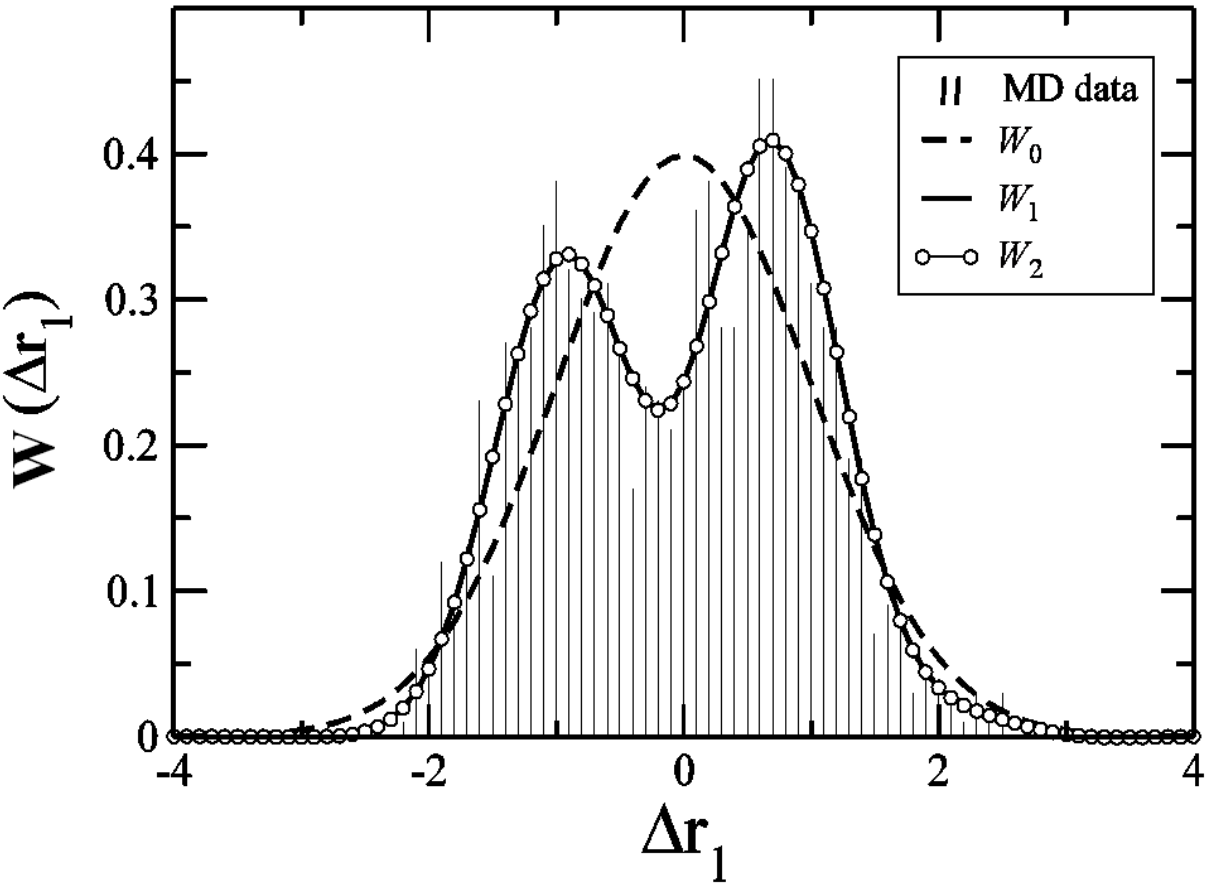


Figure 4-3. A comparison of W_0 and W_1 and W_2 on the normalized histogram of the slowest mode. Note that W_1 and W_2 give the same marginal mode probabilities.

The full dataset consists of 8967 snapshots of 132 modes taken at 0.1 ps intervals. To prevent overfitting, every 9^{th} snapshot (a total of 996) was reserved as the test set and the rest (7971 snapshots) were used as the training set. Figure 2 gives the time plots of a few of the sample modes. The average values of the Hermite polynomials given in equation (5) are calculated by

averaging over 7971 snapshots of the fluctuation trajectory. A maximum rank of 17 for the Hermite polynomial tensors was observed to be sufficientfor the convergence of our results.

4.4.2. Entropy Estimation

In order to compare the quality of the three models W_0 and W_1 and W_2 (given respectively by Eqs. (9, 10, and 12)), we use the average log likelihood of the snapshots in the test data, given the parameters optimized for the training data. equation (13) defines the average log likelihood of the data. $\Delta \boldsymbol{r}^{(i)}$ denotes the ith snapshot and M is the number of snapshots.

$$\begin{aligned} \left\langle \log W(\Delta \boldsymbol{r}) \right\rangle &= \int W(\Delta \boldsymbol{r}) \log W(\Delta \boldsymbol{r}) \, d\Delta \boldsymbol{r} \\ &\approx \frac{1}{M} \sum_{i=1}^{M} \log W\left(\Delta \boldsymbol{r}^{(i)}\right) \end{aligned} \qquad \text{4-13}$$

From the definition of the log likelihood given in equation (13) it immediately follows that the entropy, S, can be estimated as $S/k_B = -\left\langle \log W(\Delta \boldsymbol{r}) \right\rangle$ where k_B is the Boltzmann constant. The average log likelihood based on W_o is -187.5 per snapshot, corresponding to 115.5 kcal/mol contribution to the free energy. The latter is obtained as $TS = -RT\left\langle \log W(\Delta \boldsymbol{r}) \right\rangle$, R=1.986 cal/mol , T = 310 K.

Figure (3) compares the W_o and W_1 distributions for the slowest mode, obtained from the test data. The anharmonicity of the dynamics is clear and is well represented by W_1. The free energy equivalent of the W_1 entropy is 115.1 kcal/mol which is only 0.4% less than that of W_o.

4.4.3. Mode-Coupling Corrections

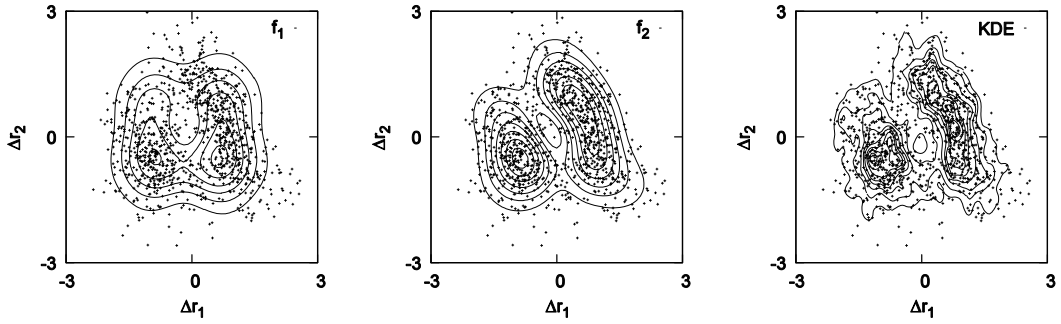


Figure 4-4. A comparison of W_1, W_2 and KDE against a scatter plot of the two slowest modes.

The W_2 approximation in equation (8) introduces pairwise interactions between modes. Note that, the correction due to the mode-coupling terms included in W_2 is invisible at the level of the marginal distibutions, such as in figure (3). Therefore we compare in the first two panels of figure 4 the contour plots of W_1 and W_2 distributions with the scatter plot of the two slowest modes. The free energy landscape of the protein is captured visibly better by W_2 when compared to W_1. The contribution at the level of W_2 to free energy equivalent of entropy is 110.9 kcal/mol. This results in a correction of ≈ -4.6 kcal/mol to the free energy *solely due to mode-coupling.*

4.4.4. Higher-Order Coupling

To quantify the effects of phenomena beyond pairwise interactions, we further used non-parametric kernel density estimation (KDE) procedure to model the probability density of the training set. In this approach, one fits the data using second-order Gaussian kernels with fixed bandwidths optimized by likelihood cross validation. We used the “npudensbw” and “npudens” procedures from the “np” package for the “R” statistical computing environment[56].

In figure 4, the results of KDE are presented in the third panel. These results are representative of all orders of contributions to anharmonicity and mode coupling and may be used to obtain a reasonable upper bound on the impact of mode-coupling on protein energetics. The free energy equivalent of the entropy calculated by the KDE is 108.4 kcal/mol. This shows that the total reduction in entropy due to anharmonicities and coupling relative to Gaussian is 7.1 kcal/mol, or 6.1% for Crambin. In spite of the fact that the simulation time is too short to faithfully sample the free-energy landscape, the result above is found to be stable under a ten-fold variation in the size of the analysed data set.

4.5. Discussion

Anharmonicity of protein motions has been addressed in several earlier papers by means of principal component analysis [23, 52, 57]. Hayward et al.[52] used a quasiharmonic approximation to the anharmonic energy surface having multiple minima, where the fluctuation distribution function was treated as a multivariable Gaussian, with the variables being the normal modes of the molecular dynamics trajectory. If hopping between different minima is suppressed, then the quasiharmonic approximation reduces to normal mode analysis. In the present paper, the quasiharmonic approximation would obtain if the sum in the brackets in equation (2) were set to zero, subject to the conditions given by Eqs. (1&3). The higher-order moments in equation (2) include terms necessary to go beyond the quasiharmonic approximation. Maisuradze and Leitner [57] applied the principal component analysis to a tetrapeptide and analyzed the free energy surface using only the largest two eigenvectors obtained from the dihedral angle space. Their analysis contains the coupling effects between the first two eigenvectors. In the present work we also expressed the free energy surface keeping mode pairs, but contrary to the the treatment of Ref.[57] all pairs are included. Ref.[57] also discussed the problems with sampling convergence based on the first two mode analysis. In our case, the relative entropy of non-harmonic effects measured over all of the modes appears stable under varying data size.

Maisuradze et al. [58] investigated the folding and unfolding of the B domain of staphylococcal protein by a coarse-grained principal component analysis and showed that while a one or two dimensional free energy landscape is sufficient for describing folding and unfolding, it may fail in describing the stability of the native state. Their work considers the folding/unfolding of a protein, and therefore deviates from the present work that focuses on fluctuations about the native state. Nevertheless, their observations are relevant, and the need to go to higher dimensions is pertinent. Similarly, Altis et al. showed that a five dimensional landscape in dihedral space is necessary to properly characterize the free energy landscape [23].

The main difference of the present work from these past studies is that, by means of a Hermite expansion we take into account all of the principal components, in contrast with the restricted numbers considered in [23, 57, 58]. The constraint on the Hermite series expansion is twofold: maximum tensor rank and the order of coupling. A cutoff on the maximum rank can be made as large as computationally possible, since their evaluation is straightforward. The order of coupling (two-body, three-body, etc) is more subtle. In the present paper, we treated the first- and second-order couplings only. Including the higher-order terms is a computational challange. Instead, we supplemented our results by an alternative nonparametric "kernel density" estimation method that in effect considers all orders, but does not allow a separation of elastic, anharmonic and mode-coupling contributions to the free energy. Our results show that the difference between the Hermite representation at the second-order and the KDE do not show marked differences from each other.

4.6. Conclusion

The probability distributions of residue fluctutations obtained by the Hermite series expansion and the KDE give consistent measures of the fluctuational entropy of the protein Crambin in its native state. The Gaussian approximation W_o gives a value of $TS = 115.5$ kcal/mol. Introduction of anharmonicities in the absence of mode coupling reduces the entropy to $TS =$ 115.1 kcal/mol. Inclusion of second order mode coupling further reduces the entropy to $TS =$

110.9 kcal/mol. The KDE, which takes all orders of correlations and mode coupling into account yields an entropy of $TS =$ 108.4 kcal/mol.

In conclusion, we find that the effect of anharmonicity and correlations on the entropy is about 6% as determined from the difference between the Gaussian and the KDE approximations. The Hermite expansion of the configurational p.d.f. suggests that the correlations (i.e., mode-coupling) account for a large portion of this correction. In a strict sense, 6% is a lower bound on the best estimate one can obtain from the given data - although, presumably not too far from it in view of the nonparametric nature of the KDE method. Finally, in spite of their relatively small weight, we see that both corrections to the Gaussian approximation strongly modify the shape of the free-energy landscape of the protein in the vicinity of its native state. Most of the contribution due to anharmonicity is from the modes corresponding to the largest eigenvalues of the covariance matrix (i.e., slowest modes), yet a quantitive evaluation of the contribution to the fluctuation entropy in terms of the eigenvalues is not obvious. It is nevertheless worth stressing that, unlike in a harmonic system, the different fluctuational modes contribute unequally to the free energy of a native protein.

Chapter 5

INVESTIGATING THE EFFECT OF THE DIFFERENCE IN THE UNBINDING FREE ENERGY PROFILES OF HLA-B51 AND HLA-B52 ON BEHÇET'S DISEASE

5.1. Introduction

Behçet's disease is a type of chronic inflammatory disorder with a substantial genetic background, which is chiefly featured by perennial oral aphthous ulceration, genital ulceration, skin lesions and uveitis [59]. In the literature, the role of the immunological abnormalities on the mechanism by which the Behçet' disease is caused has been contended for a long time [59-62]. These immunological abnormalities are presumably caused by microbial pathogens in genetically delicate individuals [59]. Comas et.al.[63] have shown by using mitochondrial studies that the disease might have a genetic basis. It was observed that Behçet's disease is seen mainly at the regions around the ancient silk road and that it overlaps with a elevated frequency of HLA-B51 in the healthy population [59, 64-66]. The most powerful proof underpinning the involvement of genetic factors in the pathogenesis of Behçet's disease is its relation with HLA-B51 [64]. HLA-B5, of which HLA-B51 is a serotype, is a human gene that encodes a protein that plays a critical role in the immune system. However, if HLA-B51 has an immediate effect in the development the of Behçet's disease or if the mentioned correlation implies linkage disequilibrium with a susceptibility gene for Behçet's disease which is stationed close to HLA-B locus, has been studied in the literature extensively [59]. It was shown using stratification analyses, genotypic differentiation and allelic association in various ethnic clusters that HLA-B51 is presenting the most powerful relation with Behçet 's disease [59, 67-69]. Moreover, HLA-B sequence specifying HLA-B*51 is proposed to be the main responsible genetic factor in Behçet's disease [70, 71]. The B*51 allele, HLA-B*51001, is probably the strongest related B*51 allele to the Behçet's disease since it is observed in higher levels in Behçet's patients compared to healthy individuals [64, 71].

Assuming that HLA-B*5101 association with the mechanisms of Behçet's disease is correct, an minor differences in peptide sequences between the HLA molecules should be able to provide detailed information about the disease etiology [71, 72]. HLA-B51 selects peptides with eight or nine amino acids and with a hydrophobic C-terminus [71, 73, 74]. Lemmel et al. [71] have performed a mass spectrometric analysis of the 22 HLA-B*5101 ligands which were obtained by extraction and identification from cells which include HLA-B5101 molecules. As a result of the comparison among these peptides it was found that the amino acids at positions 2 and 9 are anchor residues. These anchor residues were identified to be Alanine, Proline and Glycine at position 2 and to be Isoleucine, Phenylalanine and Valine at position 9[71].

Since the peptide sequence, YAYDGKDYI, which is one of the catabolic products of HLA-B51, includes these particular anchor amino acids; it is selected as the ligand in this study. The protein structure of HLA-B*5101 was extracted from the pdb file 1E27.

-B5 split antigen HLA-B52 is not related with Behçet's disease and differs only 2 amino acids at the α 1 helix from HLA-B51[59]. Asparagine and phenylalanine in HLA-B51 (positioned at 63 and 67 of the α 1 helix), which are located at the B pocket of the antigen binding groove, are exchanged with glutamic acid and serine in HLA-B52 [59, 66, 75] . It was shown by molecular typing of HLA-B51 molecules that the presence of certain Behçet's disease related peptides with their distinctive F and B pockets features may be one of the causative factors of Behçet's disease [74]. Moreover, it was suggested the motif of the peptides that are able to bind to the HLA molecules may be altered by the modifications in the B pocket [71, 74]. The structure of HLA-B52 was obtained by the modification of the HLA-B*5101 structure at positions 63 and 67.

In order to gain detailed information about the relationship between HLA-B51 and Behçet's disease we compared binding of the peptide, YAYDGKDYI, to the B pocket of HLA-B51 and of HLA-B52.

Unbinding is simulated in NAMD via steered molecular dynamics. Force profiles, Potential of mean force (PMF) profiles are constructed and free energy differences are evaluated. The

purpose of this analysis is to identify the strength, characteristics and difference among the unbinding process for HLA-B51 and HLA-B512.

5.2. Methods

5.2.1. Theory

In this work the unbinding of the ligand from the proteins are performed with a finite velocity. Because of this finite velocity, friction is present in the system and entropy generation due to friction occurs. Hence, the process becomes a non-equilibrium (irreversible) process. In addition, for an T,V,N ensemble the temperature of the system in reality does not stay exactly constant during the molecular dynamics simulations. It fluctuates slightly around the temperature of the bath. Therefore, energy transfer over a finite temperature difference may occurs which might be another source of irreversibility.

Some of work that is applied on the system during the unbinding process will be lost to these irreversibilities. Hence, the change in free energy of the system will be lower than the work applied on the system. The second law of thermodynamic states that the free energy difference between the initial and final state can not be larger than the average work performed on the system during this process.

$$\Delta A = A_{final} - A_{initial} \leq \langle W \rangle \qquad 5\text{-}1$$

equation (1) provides an upper limit to the free energy differenc. The Jarzynski's equality states that the following equality holds regardless of the speed of the process [29, 76]

$$e^{-\beta \Delta A} = \left\langle e^{-\beta W} \right\rangle \qquad 5\text{-}2$$

Jarzynski's equality is a relation between equilibrium free energy differences and work done through nonequilibrium processes[77]. The major difficulty of the Jarzynski'equality is that its average is dominated by small work values that are observed only rarely. Therefore, if only

a small number of steered molecular dynamic simulations are performed, the velocity should be small enough to permit such small work values. Moreover, in the literature this difficulty was overcome to some extend by applying the cumulant expansion [29, 78, 79] .

$$\log\left\langle e^{-\beta W}\right\rangle = -\beta\langle W\rangle + \frac{\beta^2}{2}(\left\langle W^2\right\rangle - \langle W\rangle^2) - \frac{\beta^3}{3!}(\left\langle W^3\right\rangle - 3\langle W\rangle^2\langle W\rangle + 2\langle W\rangle^3) + .. \qquad 5\text{-}3$$

Using the cumulant expansion two kind of error are involved; Systematic error due to the truncation of higher order terms and statistical error due to insufficient sampling [77]. For a finite number of trajectories the statistical error is larger than the systematic error. Therefore, as [77] have pointed out approximate formulas may give better results because lower order cumulants are estimated with smaller statistical error.

The system is described as a classical mechanical system of N particles which is described by a molecular dynamics simulation at constant temperature T and volume V. A microstate of the system is defined by the 3N dimensional position vector $\mathbf{R}$ and momentum vector $\mathbf{p}$. In SMD an external harmonic guiding potential is applied so that by adjusting λ the system is guided along the reaction coordinate ξ as [80],

$$h_\lambda(\mathbf{R}) = \frac{k}{2}(\xi(\mathbf{R}) - \lambda)^2 \qquad 5\text{-}4$$

For both HLA-B52 and HLA-B51 the reaction coordinate ξ was selected as the distance among the center of mass of THR97 in chain A and the center of mass of GLY5 in chain C . The modified Hamiltonian is defined as

$$H_\lambda(\mathbf{R},\mathbf{p}) = H_o(\mathbf{R},\mathbf{p}) + h_\lambda(\mathbf{R}) \qquad 5\text{-}5$$

Where $H_o(\mathbf{R},\mathbf{p})$ is the Hamiltonian of the original system. The average $\langle W\rangle$ is taken over the ensemble of trajectories whose initial states $\mathbf{R}_o,\mathbf{p}_o$ are sampled from the canonical ensemble

corresponding to the Hamiltonian $H_o(\mathbf{R}_o,\mathbf{p}_o)$ [77] . For that purpose structures of the N,V,T simulation, each 10 ps away from each other will be used as starting structures for Steered Molecular Dynamics.

Constant velocity SMD simulations were performed, meaning that λ is changed with a constant velocity as [80],

$$\lambda(t) = \lambda(0) + vt \qquad 5\text{-}6$$

Here t is time and $\lambda(t)$ is the λ parameter value at time t of the simulation. The external work is evaluated as,

$$W_{\lambda(0)\to\lambda(t)} = \int_{\lambda(0)}^{\lambda(t)} F d\lambda \qquad 5\text{-}7$$

Where F is the force applied on the system which is evaluated as $F = \frac{dh_\lambda(\mathbf{R})}{d\lambda} = -k\left[\xi(\mathbf{R}_{t'}) - \lambda(0) - vt'\right]$ and $d\lambda = vdt$. In order to obtain the force F in the direction of pulling, the dot product of the force $\vec{f}$ applied on the SMD atoms in the simulation and direction of pulling $\vec{n}$, which is the reaction coordinate, need to be calculated $F = \vec{f} \bullet \vec{n}$ [48].

The Jarzynski's equality provides the methodology to evaluate the free energy differences $A(\lambda(t)) - A(\lambda(0))$ using the work values $W_{\lambda(0)\to\lambda(t)}$. However $\xi(\mathbf{R}_t)$ fluctuates among trajectories, meaning that at the same $\lambda(t)$ value, different reaction coordinates $\xi(\mathbf{R}_t)$ may be observed. Hence, to calculate the PMF $\Phi(\xi)$ at ξ, $W_{\lambda(0)\to\lambda(t)}$ values at different time t but being at the same reaction coordinate ξ have to be combined. When the spring constant k of the guiding potential is sufficiently large so that the reaction coordinate follows the constraint center λ closely, the following stiff-spring approximation emerges [77]

$$A(\lambda) \approx \Phi(\lambda) \qquad 5\text{-}8$$

Hence, the PMF $\Phi(\lambda)$ will be evaluated by the Jarzynski's equality using the work values $W_{\lambda(0)\to\lambda(t)}$.

Due to the external potential applied to the SMD atoms the conformation of the peptide will be lightly biased. Therefore the final states will not be in equilibrium state. However, to relax these final states no external work is required. Therefore, Jarzynski's equality can be stated in terms of transformations between equilibrium states [80]

5.2.2. MD Simulations

All Molecular dynamics simulations were performed for an N,V,T ensemble in explicit solvent (water) using NAMD 2.6 package with CHARMM27 force field. Simulations were performed at 310 K temperature and 1 bar pressure. The reaction coordinates were aligned with the positive x axis. The proteins were then solvated in a waterbox of 40 **Å** cushion in the positive x direction and 10 **Å** cushions in the other directions. Periodic boundary conditions were applied. Ions were added in order to represent a more typical biological environment. Langevin. dynamics was used to control the systems temperature and pressure. All atoms were coupled to the heat bath. A time step of 1fs was used. Nonbonded and electrostatic forces were evaluated each time step. In order to keep all degrees of freedom no rigid bonds were used. Three minimization-equilibration cycles were applied: The first one was applied under N,P,T conditions to relax the water in the first place and the second and third ones were applied under N,V,T conditions to find a local minimum of the whole system's energy[35]. The energy of the initial system was first minimized for 20000 steps. Water was then equilibrated by keeping the Protein fixed for 0.25 ns. Then, the protein was released and an additional 20000 step minimization was performed this time under N,V,T condition. Then the system was equilibrated for 0.5ns. After a final 20000 steps of minimization, the final conventional molecular dynamic simulation (CMD) was performed. At every 500^{th} time step of final conventional molecular dynamic simulation, the instantaneous atomic coordinates **R** of all atoms, the pressures and the energies were

recorded. For both HLA-B52 and HLA-B51, 15ns long trajectories at time 10-25ns of the CMD were used for the calculation. For HLA-B51 the initial states $\mathbf{R}_o, \mathbf{p}_o$ of the SMD simulations were selected starting at 20.782 ns with 10 ps space between them. For HLA-B51 the initial states $\mathbf{R}_o, \mathbf{p}_o$ of the SMD trajectories were selected starting at 13.688 ns with 10ps space between them.

5.3. Result and Discussion

5.3.1. Change in Dynamics

As was mentioned before HLA-B51 differs only 2 aminoacids from HLA-B52; Asparagine and phenylalanine at positions 63 and 67 of the 1 helix of the HLA-B51 molecule are replaced with glutamic acid and serine at the same positions in the HLA-B52 [81]. In order to understand the effect of this modification on the dynamics of the bound forms of HLA-B51 and HLA-B52, the fluctuations of the ligand are investigated.

In order to eliminate translational and rotational degrees of freedom, the structures in each trajectory were aligned with respect to the structure at the middle of the production simulation (time 7.5ns). Alignments were performed using the transformation matrix which shows the best fit of the backbone atoms. These alignments were carried out in two different ways; 1- The protein-ligand complex was aligned with respect to the ligand. 1-The protein-ligand complex was aligned with respect to the chain A of the protein, which is the chain with the binding pocket.

For type 1 alignment, only the vibrational fluctuation of the ligand is of consequence. In the second type of alignment, since the rotational and translational degrees of freedom of chain A are eliminated, the relative fluctuations of the ligand with respect to the chain A and the ligand's vibrational fluctuations are of consequence.

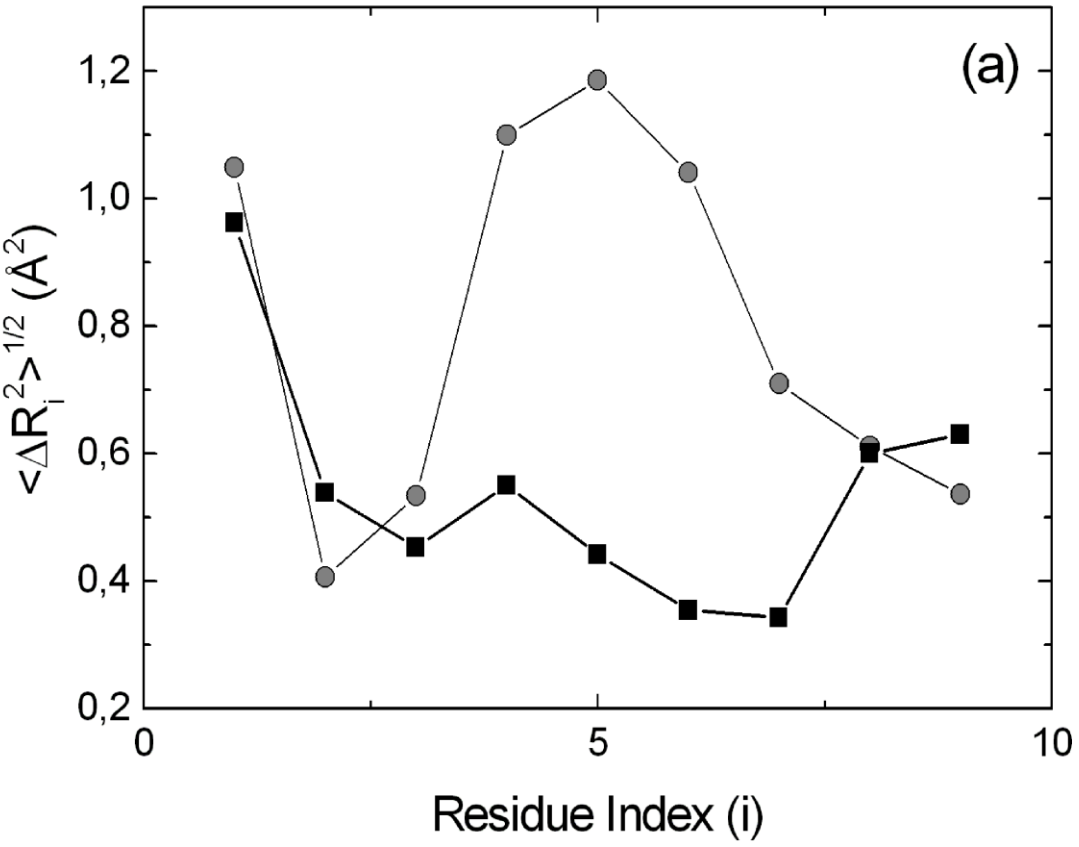
The square root of the mean fluctuation of each carbon alpha atom will be used as a numeric indication of the fluctuations as,

$$\left\langle \Delta R_i^2 \right\rangle^{1/2} = \left\langle \left(R_i - \bar{R}_i \right)^2 \right\rangle^{1/2} \qquad 5\text{-}9$$

Here $\bar{R}_i$ is the mean atomic coordinate of the ith carbon alpha atom and R_i is its instantaneous coordinate. In figure 1 (a-b) the $\left\langle \Delta R_i^2 \right\rangle^{1/2}$ values of are provided for the 15ns long trajectories. Black dots indicate the fluctuation value for HLA-B52 whereas grey dots indicate the fluctuation values for HLA-B51. In figure 1(a) type 1 alignments (with respect to the ligand) were performed. In figure 1 (b) on the other hand type 2 alignments (with respect to chain A) were performed.

In figure 1 (a) it is observed that the vibrational fluctuation of the ligand differ significantly among HLA-B52 and HLA-B51. On the other hand, in figure 1(b) it is observed that when aligned with respect chain A, the ligand seems to be on the overall be more mobile in HLA-B51

It can be concluded that for this ligand the HLA-B51 binding pocket is sloppier than that of HLA-B52.


(a)
1,2
1,0
0,8
0,6
0,4
0,2
0
5
10
$<\Delta R_i^2>^{1/2}$ (Å^2)
Residue Index (i)

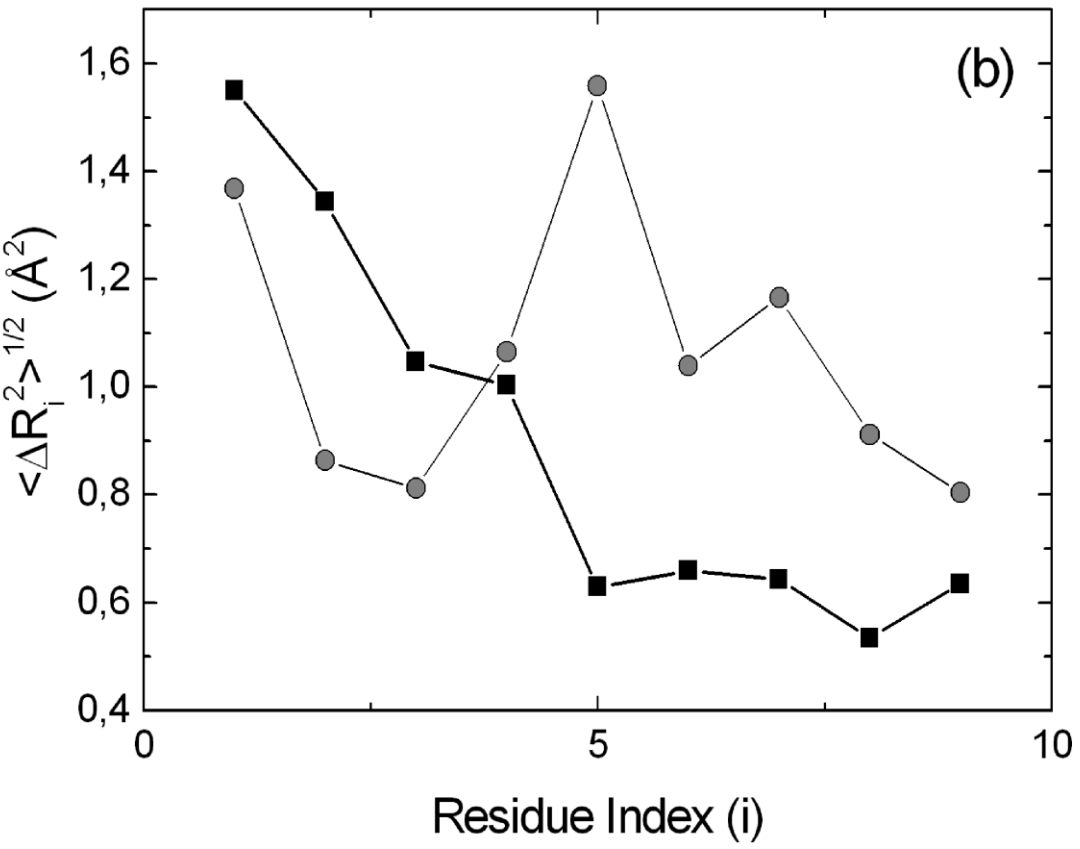


Figure 5-1. $\left\langle \Delta R_i^{\,2} \right\rangle^{1/2}$ values of the carbon alpha atom of the ligand for the 15ns trajectories where complex is aligned with respect to the ligand **(a)** and complex is aligned with respect to the chain A of the protein **(b).** Black dots indicate the fluctuation value for HLA-B52 whereas grey dots indicate the fluctuation values for HLA-B51.

Since the difference among HLA-B51 and HLA-B52 are only the aminoacids at positions 63 and 67 of the 1 helix, the change in the fluctuations of this helix are investigated. For that purpose the fluctuations among each residue pair i and j are evaluated using the alpha carbon coordinates as follows

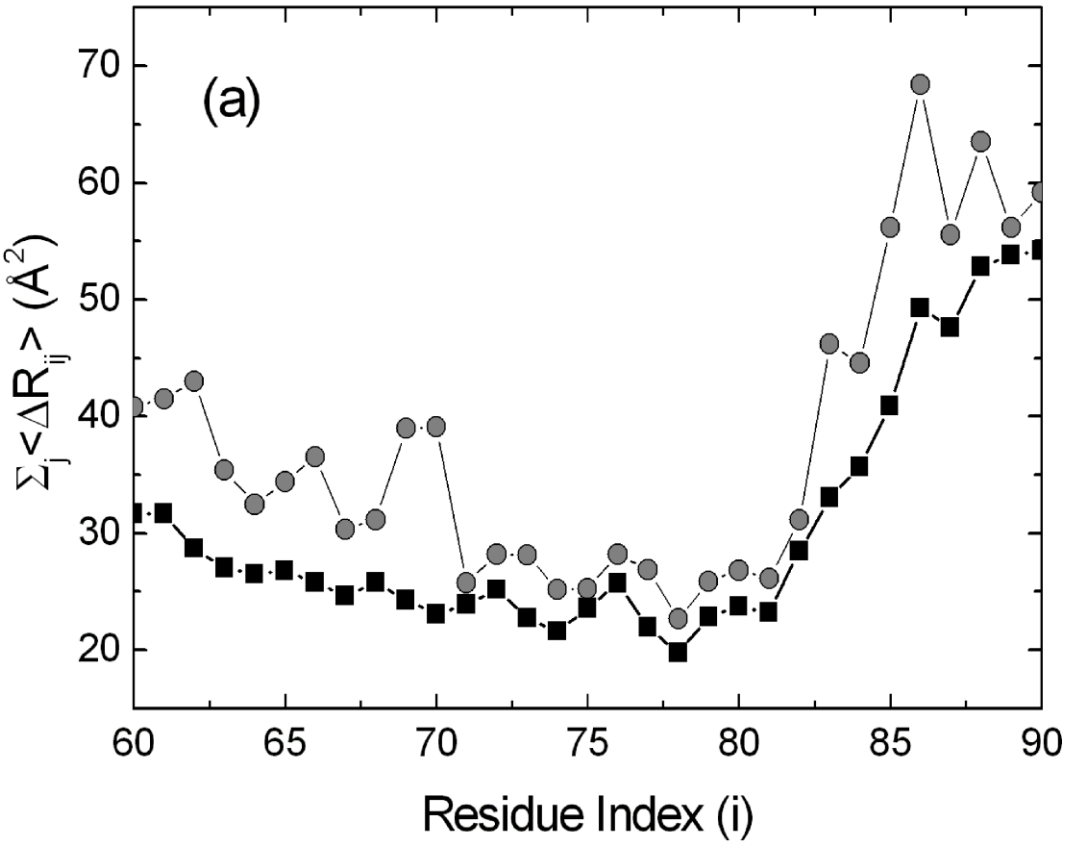
$$\left\langle \Delta R_{ij}^{\,2} \right\rangle = \left\langle \left(R_{ij} - \bar{R}_{ij} \right)^2 \right\rangle \qquad 5\text{-}10$$

Where $R_{ij} = R_i - R_j$ is the distance among residue i residue j of chain A. In order to indicate the overall strength of the correlation of each ith residue with all others, summation of $\left\langle \Delta R_{ij}^{\,2} \right\rangle$ over all other residues at helix 1 (residues 60-90) are performed as $\sum_{j=60}^{90} \left\langle \Delta R_{ij}^{\,2} \right\rangle$ and shown in figure 2 (a).

The same procedure is repeated by constructing the covariance matrix $\left\langle \mathbf{\Delta R \Delta R^T} \right\rangle$ for the helix and performing the summation as $\sum_{j=60}^{90} \left\langle \Delta R_i \Delta R_j \right\rangle$. $\sum_{j=60}^{90} \left\langle \Delta R_i \Delta R_j \right\rangle$ values are depicted in figure 2 (b) . Black dots indicate the overall strength of the correlation values for HLA-B52 whereas grey dots indicate the overall strength of the correlation values for HLA-B51.

It has to be noted that alignments prior to the evaluations were performed with respect to the backbone atoms of the whole protein-ligand complex.

As can be seen in both figures the fluctuations among the residues of helix-1 for HLA-B51 are significantly larger than that for HLA-B52.


(a)
70
60
50
40
30
20
60
65
70
75
80
85
90
Residue Index (i)
$\Sigma_j <\Delta R_{ij}>$ (Å^2)

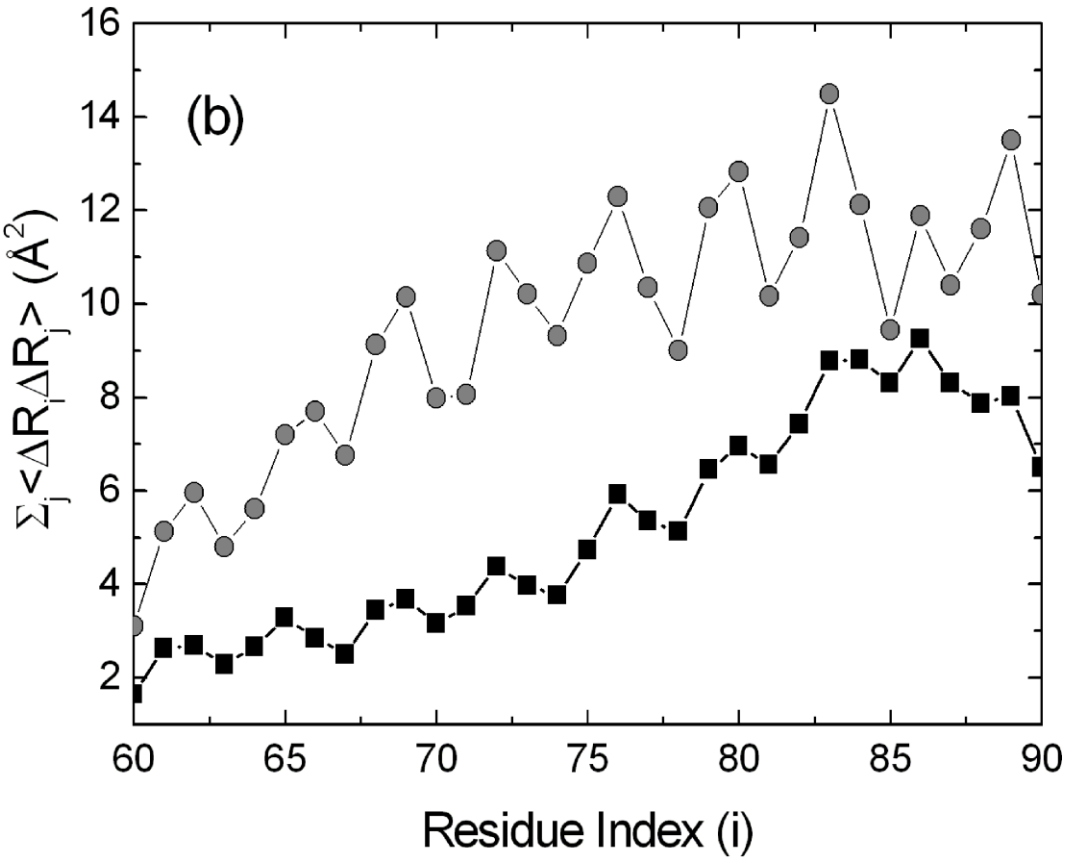


Figure 5-2. $\sum_{j=60}^{90} \left\langle \Delta R_{ij}{}^2 \right\rangle$ values of the carbon alpha atoms of helix-1 **(a)** $\sum_{j=60}^{90} \left\langle \Delta R_i \Delta R_j \right\rangle$ values of the carbon alpha atoms of helix-1 **(b).** Black dots indicate overal strength of the correlation values for HLA-B52 whereas grey dots indicate the overal strength of the correlation values values for HLA-B51**.**

5.3.2. SMD Simulations

Park et al. [77] have shown that for the unfolding of helical Deca-alanine in vacuum a pulling speed of $0.1\ \text{Å}/ns$ resulted in an reversible process. In our work a pulling speed of $10\ \text{Å}/ns$ which is 100 times higher than this reversible pulling speed were applied. Pulling speeds and spring constants of similar studies are as follows; $20\ \text{Å}/ns$ (time step is 2fs) and spring constant of $5\ kcal/mol\text{Å}^2$ for Acetylcholine unbinding from the alpha7 nicotinic acetylcholine

receptor ligand binding domain [82] . Zhang et.al. [82] have concluded that a pulling velocity of $20\ \text{Å}/ns$ were sufficiently slow to produce good statistics from the number of trajectories they have used .

The initial structures for the SMD trajectories were selected from 25 snapshots each taken with 10ps interval among them. Vashisth &Abrams [83] have randomly sampled their starting structures from the ensemble of configurations in the final ~100ps of equilibration simulation. Zhang et.al. [82] have used the final state of the equilibration as a restart point for further SMD study. Park et. al. [77] have selected the initial coordinates of the SMD simulation from an ensemble generated by a 1ns equilibrium simulation.

To generate the PMF, Zhang et.al. [82] used 22 to 45 separate SMD runs for each trajectory. Park et. al.[77], on the other hand have used 100 trajectories for the helix coil transition of deca-alanine in vacuum and grouped them into ten blocks of ten trajectories. It was shown for the unfolding of helical deca-alanine, that for both pulling velocities $10\ \text{Å}/ns$ and $100\ \text{Å}/ns$ the second order cumulant expansion gave better result than the Jarzynski's equality[77]. The finite-sampling estimate of a non-linear average is biased[77]. Therefore instead of using the second order cumulant expansion directly, the unbiased estimate introduced by [77] will be used as

$$\log\left\langle e^{-\beta W}\right\rangle = \frac{1}{\beta}\left\{\frac{1}{M}\sum_{i=1}^{M} W_i - \frac{\beta}{2}\frac{M}{M-1}\left[\frac{1}{M}\sum_{i=1}^{M} W_i^2 - \left(\frac{1}{M}\sum_{i=1}^{M} W_i\right)^2\right]\right\} \qquad 5\text{-}11$$

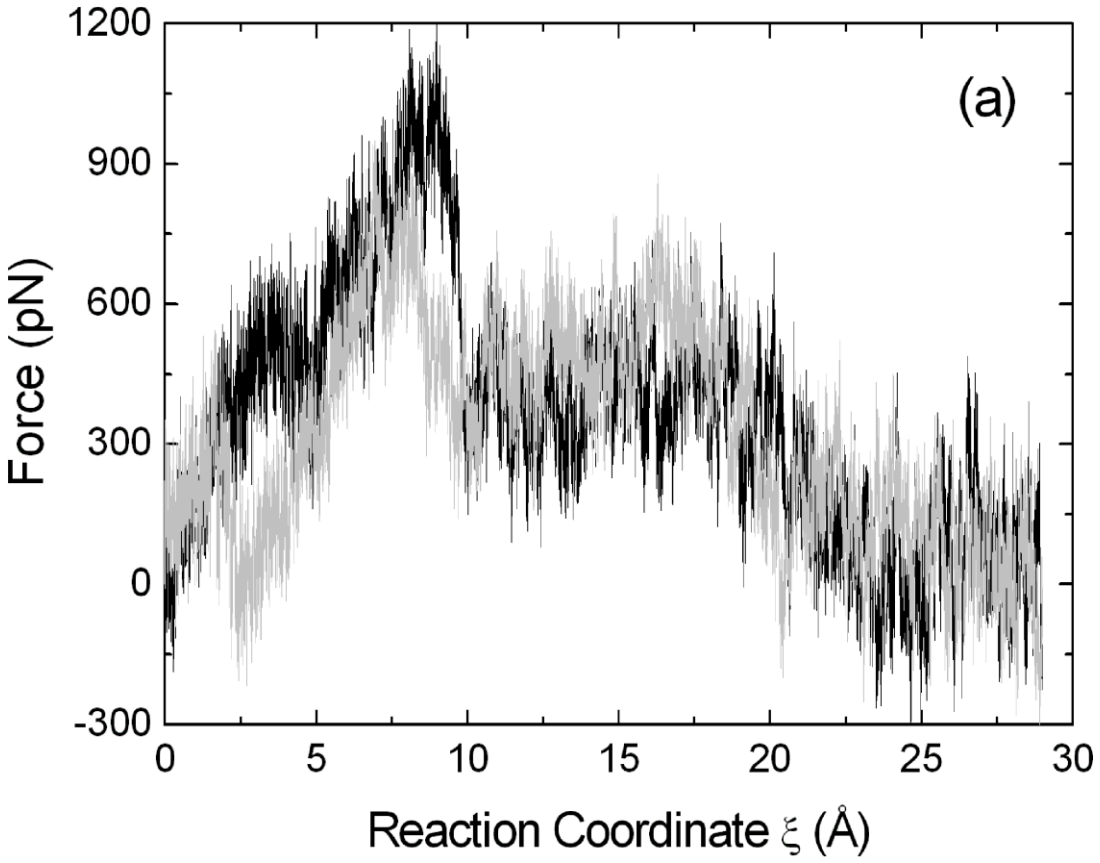
Here M is the total number of trajectories and W_i is the work obtained from the i^{th} trajectory.

In figure 3 (a) the force profiles for the 11. SMD simulations are provided for HLA-B51 and HLA- HLA-B52. The forces shown in figure 3 (a) fluctuate to both positive and negative values which indicate that the thermal fluctuations of the peptides are larger than the perturbation from the pulling force. For this condition the unbinding process was concluded to be near equilibrium [82]. The reasoning behind that statement maybe better understood by the statement provided previously by [84] which says that smaller velocities generally result

in longer times and smaller forces required to induce unbinding because of the higher probability for the ligand to overcome the lowered energy barrier due to thermal fluctuations. In figure 3 (b) the average force profiles of HLA-B52 (black line) and HLA-B51 (grey line) with respect to the reaction coordinate ξ for the SMD simulation are provided as,

$$\bar{W}(\xi) = \frac{\sum_{i=1}^{M} W_i(\xi)}{M} \qquad 5\text{-}12$$

Here the sum is over a total number of M SMD simulations. Based on the magnitude of the force peaks for the unbinding pathways of HLA-B51 and HLA-B52 it can be concluded that unbinding of HLA-B51 is smoother than for HLA-B52. Hence, indicating that HLA-B52 is stronger bound. After 23 **Å** force values fluctuate around a constant value. This related small work values are due to the friction force generated by the finite pulling velocity. These work value may be more eminent then other works due to the larger size of the peptide that is moved. Therefore, free energy profiles after $\xi = 23$**Å** are expected to increase at a constant rate.


(a)
Force (pN)
1200
900
600
300
0
-300
0
5
10
15
20
25
30
Reaction Coordinate ξ (Å)

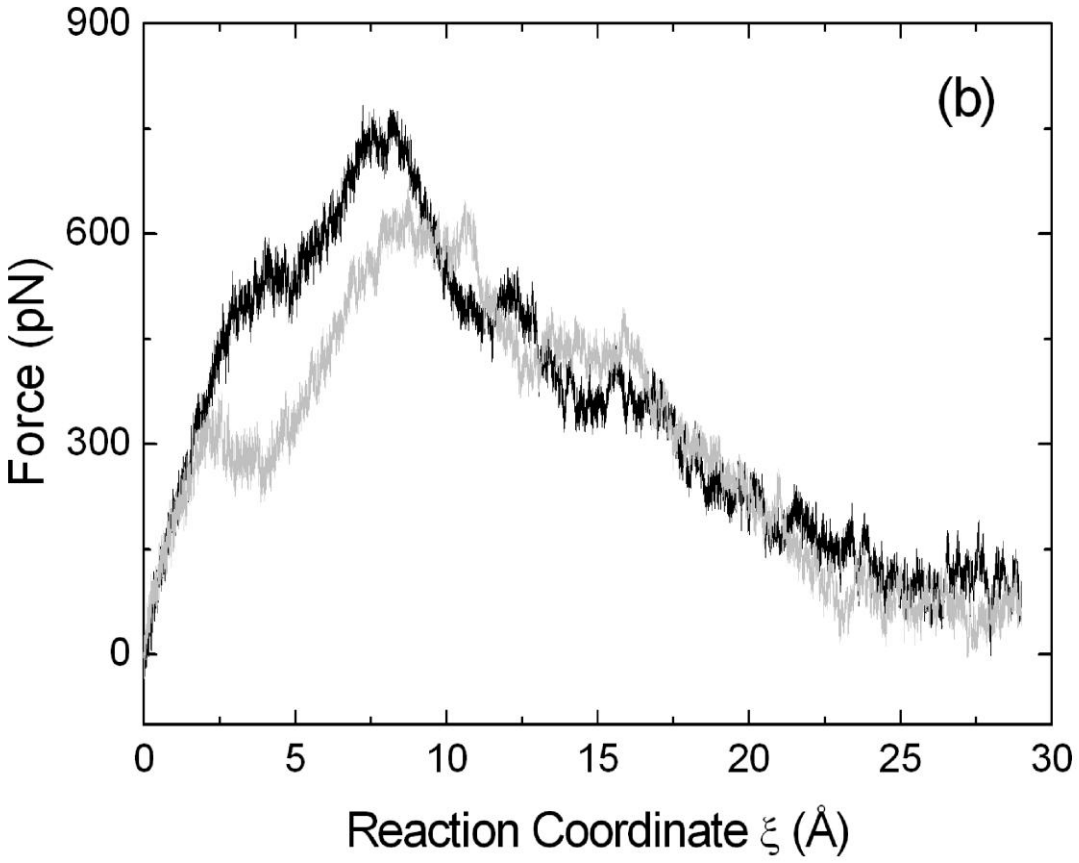


Figure 5-3. (a) Reaction coordinates vs force for the 11. SMD simulation (Starting structure selected at 20.892 ns of the conventional molecular dynamics simulation) of HLA-B52 (shown with black line) and for the 11. SMD simulation (Starting structure selected at 13.798ns of the conventional molecular dynamics simulation) simulation of HLA-B51 (Shown with grey line). **(b)** Reaction coordinates vs average force of all SMD simulation for HLA-B52 (shown with the black solid line) and HLA-B51 (Shown with the grey line).

In figure 4 the potential of mean forces (PMF) evaluated using equation (11) are shown for the reaction coordinate ξ. The black line is the potential of mean force of HLA-B52 whereas the grey line is the PMF of HLA-B51. The PMF of HLA-B51 clearly follows a lower curve than HLA-B52, indicating that unbinding of HLA-B51 is easier than HLA-B52. This result is in total agreement with the previous findings stated above.

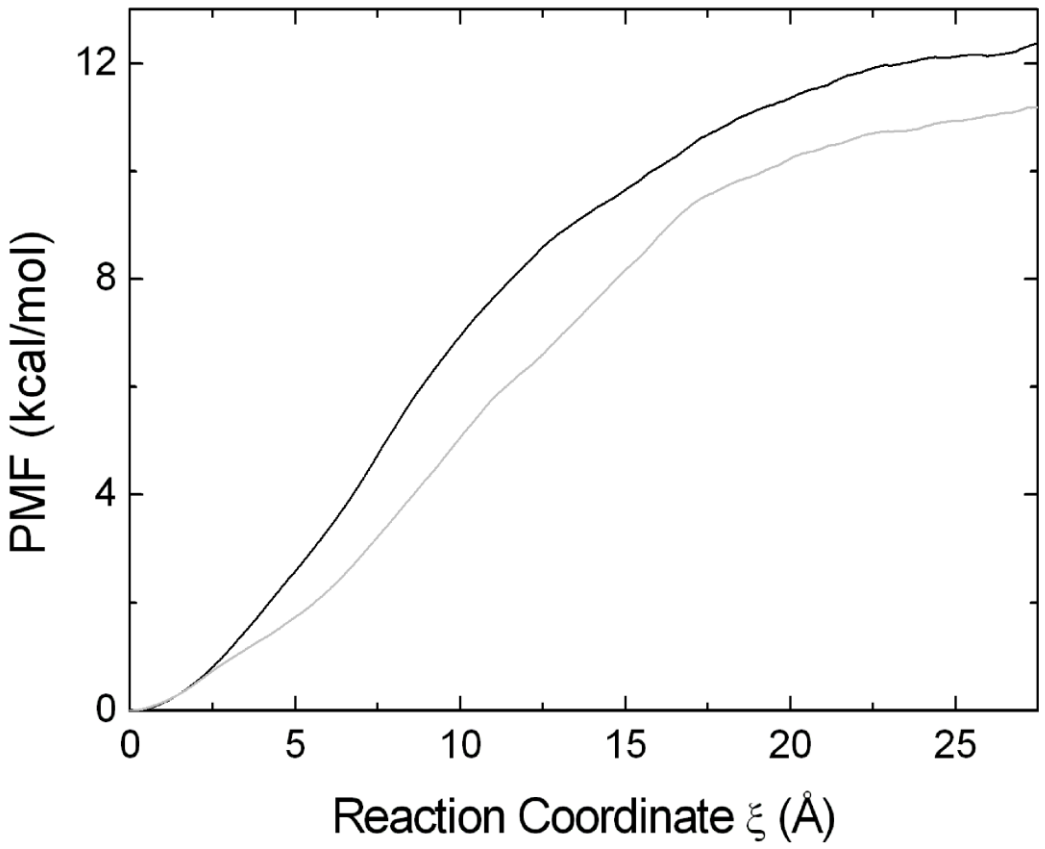


Figure 5-4. PMF of HLA-B52 (Solid black line) and HLA-B51 (Solid grey line) with respect to the reaction coordinate ξ

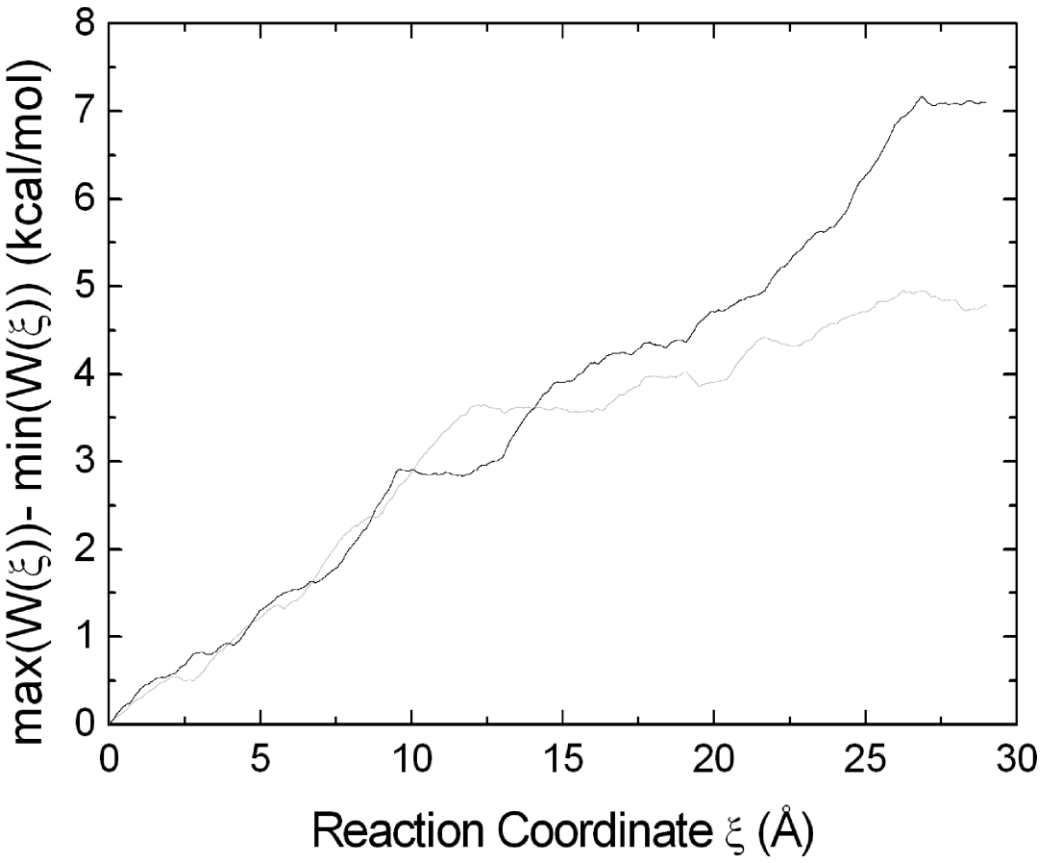


Figure 5-5. Difference between the minimum and maximum work values at each reaction coordinate ξ

Direct use of Jarzynski's equality requires the deviation of work distribution within a few kT [82]. In figure 5 the difference between the minimum and maximum work values at each coordinates are shown. As can be seen the standard deviation grows as the peptide unbinds. It is important to note that equation (11) favor small work values. Therefore, the PMF is closer to the lower boundary of the work values. A large work deviation means that there are cases for which the work required is much larger than the PMF. For HLA-B52 the work deviation is much larger than that of HLA- HLA-B51 at the end. Therefore, in addition to the larger PMF there are also much larger work values.

Although SMD provides an approach to investigate the binding/unbinding events, yet the pulling direction of the force in SMD is chosen randomly or by guessing on the basis of structural information, which makes that the force applied to the ligand in such chosen directions may not move it along a favorable pathway [85].

5.4. Conclusion

The PMF and force profiles show that binding of peptide YAYDGKDYI to HLA-B51 is clearly more floppy than binding of the same peptide to HLA-B52. HLA-B51 differs only 2 aminoacids from the other -B5 split antigen HLA-B52, which is not related with Behçet's disease. Asparagine and phenylalanine at positions 63 and 67 of the 1 helix of the HLA-B51 molecule are replaced with glutamic acid and serine at the same positions in the HLA-B52. These modifications seem to increase the fluctuations at helix 1 and hence make the binding side floppier. This decrease in binding affinity may the reason of the pathogenic mechanism HLA-B51. Further study has to be conducted using other peptides in order to understand if this observed behavior is of general characteristic.

APPENDIX

A.1

$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle$ Å^2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	0.68	-0.01	-0.16	-0.13	-0.04	0.18	-0.19	-0.11	-0.15	-0.07	-0.10	0.10
2	-0.01	0.28	0.12	0.00	-0.11	-0.05	-0.24	-0.05	0.05	0.03	-0.02	0.01
3	-0.16	0.12	0.23	0.08	-0.06	-0.09	-0.12	-0.01	0.09	0.04	-0.02	-0.11
4	-0.13	0.00	0.08	0.17	0.04	-0.08	-0.04	-0.01	0.05	0.04	0.01	-0.12
5	-0.04	-0.11	-0.06	0.04	0.22	0.11	0.14	-0.02	-0.05	-0.07	-0.05	-0.13
6	0.18	-0.05	-0.09	-0.08	0.11	0.69	-0.04	-0.09	-0.09	-0.16	-0.19	-0.19
7	-0.19	-0.24	-0.12	-0.04	0.14	-0.04	0.70	0.07	-0.18	-0.17	0.02	0.04
8	-0.11	-0.05	-0.01	-0.01	-0.02	-0.09	0.07	0.16	0.07	0.00	0.00	-0.02
9	-0.15	0.05	0.09	0.05	-0.05	-0.09	-0.18	0.07	0.25	0.12	-0.01	-0.14
10	-0.07	0.03	0.04	0.04	-0.07	-0.16	-0.17	0.00	0.12	0.23	0.08	-0.09
11	-0.10	-0.02	-0.02	0.01	-0.05	-0.19	0.02	0.00	-0.01	0.08	0.24	0.05
12	0.10	0.01	-0.11	-0.12	-0.13	-0.19	0.04	-0.02	-0.14	-0.09	0.05	0.60

A.2

Γ $kJ/mol \cdot Å^2$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	31.0	-33.1	1.2	-0.8	0.6	1.7	-1.0	-1.5	-0.8	6.1	-1.9	-1.6
2	-33.1	206.2	-161.5	5.5	-3.7	1.1	-0.5	-2.7	-6.1	-1.3	-7.4	3.6
3	1.2	-161.5	349.2	-168.6	1.2	-0.9	-1.3	-4.7	-11.1	-2.6	6.8	-7.5
4	-0.8	5.5	-168.6	333.1	-145.0	8.2	-9.1	4.5	2.1	-20.9	-12.0	3.1
5	0.6	-3.7	1.2	-145.0	232.1	-72.1	7.9	-2.3	-8.7	5.6	-13.6	-2.0
6	1.7	1.1	-0.9	8.2	-72.1	71.8	0.6	-1.3	0.5	-5.0	2.3	-7.1
7	-1.0	-0.5	-1.3	-9.1	7.9	0.6	69.4	-73.7	-2.6	0.5	8.0	1.7
8	-1.5	-2.7	-4.7	4.5	-2.3	-1.3	-73.7	228.4	-144.0	-5.8	0.2	2.9
9	-0.8	-6.1	-11.1	2.1	-8.7	0.5	-2.6	-144.0	325.1	-147.5	-5.0	-1.9
10	6.1	-1.3	-2.6	-20.9	5.6	-5.0	0.5	-5.8	-147.5	321.5	-151.3	0.8
11	-1.9	-7.4	6.8	-12.0	-13.6	2.3	8.0	0.2	-5.0	-151.3	227.8	-53.9
12	-1.6	3.6	-7.5	3.1	-2.0	-7.1	1.7	2.9	-1.9	0.8	-53.9	62.0

BIBLIOGRAPHY

1. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. 28(1): p. 235-242.
2. Tirion, M.M., *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis.* Physical Review Letters, 1996. 77(9): p. 1905-1908.
3. Bahar, I., A.R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential.* Folding & Design, 1997. 2(3): p. 173-181.
4. Haliloglu, T., I. Bahar, and B. Erman, *Gaussian dynamics of folded proteins.* Physical Review Letters, 1997. 79(16): p. 3090-3093.
5. Flory, P.J., M. Gordon, and N.G. McCrum, *Statistical Thermodynamics of Random Networks [and Discussion]*. Proc. R. Soc. Lond. A, 1976. 351(1666): p. 351-380.
6. Kloczkowski, A., J.E. Mark, and B. Erman, *Chain Dimensions and Fluctuations in Random Elastomeric Networks.I. Phantom Gaussian Networks in the Undeformed State.* Macromolecules, 1989. 22(3): p. 1423-1432.
7. Hinsen, K., *Analysis of domain motions by approximate normal mode calculations.* Proteins-Structure Function and Genetics, 1998. 33(3): p. 417-429.
8. Atilgan, A.R., et al., *Anisotropy of fluctuation dynamics of proteins with an elastic network model.* Biophysical Journal, 2001. 80(1): p. 505-515.
9. Cui, Q. and I. Bahar, *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*. 2006: Chapman Hall/GRC.
10. Bahar, I., et al., *Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability.* Physical Review Letters, 1998. 80(12): p. 2733-2736.
11. Haliloglu, T., E. Seyrek, and B. Erman, *Prediction of binding sites in receptor-ligand complexes with the Gaussian Network Model.* Physical Review Letters, 2008. 100(22): p. -.
12. Levy, R.M., D. Perahia, and M. Karplus, *Molecular-Dynamics of an Alpha-Helical Polypeptide - Temperature-Dependence and Deviation from Harmonic Behavior.* Proceedings of the National Academy of Sciences of the United States of America-Physical Sciences, 1982. 79(4): p. 1346-1350.

13. Hayward, S., A. Kitao, and N. Go, *Harmonicity and Anharmonicity in Protein Dynamics - a Normal-Mode Analysis and Principal Component Analysis.* Proteins-Structure Function and Genetics, 1995. 23(2): p. 177-186.
14. Juanico, B., et al., *Discrete breathers in nonlinear network models of proteins.* Physical Review Letters, 2007. 99(23): p. -.
15. Piazza, F. and Y.H. Sanejouand, *Discrete breathers in protein structures.* Physical Biology, 2008. 5(2): p. -.
16. Moritsugu, K., O. Miyashita, and A. Kidera, *Vibrational energy transfer in a protein molecule.* Physical Review Letters, 2000. 85(18): p. 3970-3973.
17. Moritsugu, K., O. Miyashita, and A. Kidera, *Temperature dependence of vibrational energy transfer in a protein molecule.* Journal of Physical Chemistry B, 2003. 107(14): p. 3309-3317.
18. Callen, H.B., *Thermodynamic and an Introduction to Thermostatistics*. Second Edition ed. 1985: John Wiley & Sons.
19. Hill, T.L., *Thermodynamics of Small Systems*. July 1994: Dover Publications.
20. Karplus, M. and J.N. Kushick, *Method for Estimating the Configurational Entropy of Macromolecules.* Macromolecules, 1981. 14(2): p. 325-332.
21. Levy, R.M., et al., *Evaluation of the Configurational Entropy for Proteins - Application to Molecular-Dynamics Simulations of an Alpha-Helix.* Macromolecules, 1984. 17(7): p. 1370-1374.
22. Lamm, G. and A. Szabo, *Langevin Modes of Macromolecules.* Journal of Chemical Physics, 1986. 85(12): p. 7334-7348.
23. Altis, A., et al., *Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis.* Journal of Chemical Physics, 2008. 128(24): p. -.
24. Moritsugu, K. and J.C. Smith, *Coarse-grained Biomolecular simulation with REACH: Realistic extension algorithm via covariance hessian.* Biophysical Journal, 2007. 93(10): p. 3460-3469.
25. Moritsugu, K. and J.C. Smith, *REACH: A program for coarse-grained biomolecular simulation.* Computer Physics Communications, 2009. 180(7): p. 1188-1195.
26. Moritsugu, K., V. Kurkal-Siebert, and J.C. Smith, *REACH Coarse-Grained Normal Mode Analysis of Protein Dimer Interaction Dynamics.* Biophysical Journal, 2009. 97(4): p. 1158-1167.
27. Gur, M. and B. Erman, *Mode Coupling in Proteins.* Physical Biology, 2010. in print.

28. Kabakcioglu, A., et al., Physical Biology, 2010. in print.
29. Jarzynski, C., *Nonequilibrium equality for free energy differences.* Physical Review Letters, 1997. 78(14): p. 2690-2693.
30. Gullingsrud, J.R., R. Braun, and K.J. Schulten, *Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations.* Biophysical Journal, 1999. 76(1): p. A200-a200.
31. Kumar, S., et al., *The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules.1. The Method.* Journal of Computational Chemistry, 1992. 13(8): p. 1011-1021.
32. Henin, J. and C. Chipot, *Overcoming free energy barriers using unconstrained molecular dynamics simulations.* Journal of Chemical Physics, 2004. 121(7): p. 2904-2914.
33. den Otter, W.K., *Thermodynamic integration of the free energy along a reaction coordinate in Cartesian coordinates.* Journal of Chemical Physics, 2000. 112(17): p. 7283-7292.
34. Wang, C.X., et al., *Thermodynamic Integration Calculations of Binding Free-Energy Difference for Gly-169 Mutation in Subtilisin Bpn'.* Proteins-Structure Function and Genetics, 1993. 15(1): p. 5-9.
35. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. 26(16): p. 1781-1802.
36. Hawkins, R.J. and T.C.B. McLeish, *Dynamic allostery of protein alpha helical coiled-coils.* Journal of the Royal Society Interface, 2006. 3(6): p. 125-138.
37. Piazza, F. and Y.H. Sanejouand, *Long-range energy transfer in proteins.* Physical Biology, 2009. 6(4): p. -.
38. Moritsugu, K. and J.C. Smith, *Langevin model of the temperature and hydration dependence of protein vibrational dynamics.* Journal of Physical Chemistry B, 2005. 109(24): p. 12182-12194.
39. Reat, V., et al., *Solvent dependence of dynamic transitions in protein solutions.* Proceedings of the National Academy of Sciences of the United States of America, 2000. 97(18): p. 9961-9966.
40. Horiuchi, T. and N. Go, *Projection of Monte-Carlo and Molecular-Dynamics Trajectories onto the Normal Mode Axes - Human Lysozyme.* Proteins-Structure Function and Genetics, 1991. 10(2): p. 106-116.

41. Teeter, M.M. and D.A. Case, *Harmonic and Quasiharmonic Descriptions of Crambin.* Journal of Physical Chemistry, 1990. 94(21): p. 8091-8097.
42. Perahia, D., R.M. Levy, and M. Karplus, *Motions of an Alpha-Helical Polypeptide - Comparison of Molecular and Harmonic Dynamics.* Biopolymers, 1990. 29(4-5): p. 645-677.
43. Dinola, A., H.J.C. Berendsen, and O. Edholm, *Free-Energy Determination of Polypeptide Conformations Generated by Molecular-Dynamics.* Macromolecules, 1984. 17(10): p. 2044-2050.
44. Amadei, A., A.B.M. Linssen, and H.J.C. Berendsen, *Essential Dynamics of Proteins.* Proteins-Structure Function and Genetics, 1993. 17(4): p. 412-425.
45. Ichiye, T. and M. Karplus, *Collective Motions in Proteins - a Covariance Analysis of Atomic Fluctuations in Molecular-Dynamics and Normal Mode Simulations.* Proteins-Structure Function and Genetics, 1991. 11(3): p. 205-217.
46. Levitt, M., C. Sander, and P.S. Stern, *Protein Normal-Mode Dynamics - Trypsin-Inhibitor, Crambin, Ribonuclease and Lysozyme.* Journal of Molecular Biology, 1985. 181(3): p. 423-447.
47. Lange, O.F. and H. Grubmuller, *Can principal components yield a dimension reduced description of protein dynamics on long time scales?* Journal of Physical Chemistry B, 2006. 110(45): p. 22842-22852.
48. Isgro, T., et al. *NAMD Tutorial*. 2003 [cited.
49. Yogurtcu, O.N., M. Gur, and B. Erman, *Statistical thermodynamics of residue fluctuations in native proteins.* Journal of Chemical Physics, 2009. 130(9): p. -.
50. Flory, P.J. and D.Y. Yoon, *Moments and distribution functions for polymer chains of finite length. I Theory.* Journal of Chemical Physics, 1974a. 61: p. 5358-65.
51. Flory, P.J. and D.Y. Yoon, *Moments of PE.* J. Chem. Phys., 1974b. 61: p. 5358.
52. Hayward, S., A. Kitao, and N. Go, *Harmonic and Anharmonic Aspects in the Dynamics of Bpti - a Normal-Mode Analysis and Principal Component Analysis.* Protein Science, 1994. 3(6): p. 936-943.
53. Pontiggia, F., et al., *Anharmonicity and self-similarity of the free energy landscape of protein G.* Physical Review Letters, 2007. 98(4): p. -.
54. Leitner, D.M., *Energy flow in proteins.* Annual Review of Physical Chemistry, 2008. 59: p. 233-259.
55. Garcia, A.E., *Large-Amplitude Nonlinear Motions in Proteins.* Physical Review Letters, 1992. 68(17): p. 2696-2699.

56. Hayfield, T. and J.S. Racine, *Nonparametric econometrics: The np package*. Journal of Statistical Software, 2008. 27(5): p. 1-32.
57. Maisuradze, G.G. and D.M. Leitner, *Free energy landscape of a biomolecule in dihedral principal component space: Sampling convergence and correspondence between structures and minima.* Proteins-Structure Function and Bioinformatics, 2007. 67(3): p. 569-578.
58. Maisuradze, G.G., A. Liwo, and H.A. Scheraga, *How Adequate are One- and Two-Dimensional Free Energy Landscapes for Protein Folding Dynamics?* Physical Review Letters, 2009. 102(23): p. -.
59. Gul, A., *Behcet's disease: An update on the pathogenesis.* Clinical and Experimental Rheumatology, 2001. 19(5): p. S6-S12.
60. Lehner, T., *Immunopathogenesis of Behcet's disease.* Ann Med Interne, 1999(150): p. 483-7.
61. Lehner, T., *The role of heat shock protein, microbial and autoimmune agents in the aetiology of Behcet's disease.* Int Rev Immunol, 1997(4): p. 21-32.
62. Sakane, T., N. Suzuki, and H. Nagafuchi, *Etiopathology of Behcet's disease: immunological aspects*. Yonsei Med J, 1997(38): p. 350-8.
63. Comas, D., et al., *Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations.* American Journal of Human Genetics, 1998. 63(6): p. 1824-1838.
64. Ohno, S., et al., *Close Association of Hla-Bw51 with Behcets-Disease.* Archives of Ophthalmology, 1982. 100(9): p. 1455-1458.
65. Verity, D.H., et al., *Behcet's disease, the Silk Road and HLA-B51: historical and geographical perspectives.* Tissue Antigens, 1999. 54(3): p. 213-220.
66. Mizuki, N., H. Inoko, and S. Ohno, *Pathogenic generesponsible for the predisposition to Behçet's disase.* Intern Rev Immunol, 1997. 14: p. 33-48.
67. Mizuki, N., et al., *Association analysis between the MIC-A and HLA-B alleles in Japanese patients with Behçet's disease.* Arthritis Rheum, 1999.(42(9)): p. 1961-6.
68. Wallace, G.R., et al., *MIC-A allele profiles and HLA class I associations in Behcet's disease.* Immunogenetics, 1999. 49(7-8): p. 613-7.
69. Mizuki, N., et al., *Localization of the pathogenic gene of Behcet's disease by microsatellite analysis of three different populations.* Invest Ophthalmol Vis Sci, 2000. 41(12): p. 3702-8.

70. Sano, K., et al., *The absence of disease-specific polymorphisms within the HLA-B51 gene that is the susceptible locus for Behcet's disease.* Tissue Antigens, 2001. 58(2): p. 77-82.

71. Lemmel, C., H. Rammensee, and S. Stevanovic, *Peptide Motif of HLA-B*5101 and the Linkage,* in *Immunology of Behçet's Disease,* M. Zierhut and S. Ohno, Editors. 2003, Swets & Zeitlinger: Amsterdam. p. 127-137.

72. Gul, A., *Immunology of Behcet's Disease*, M. Zierhut and S. Ohno, Editors. 2003, Swets & Zeitlinger: Lisse. p. 73-79.

73. Sakaguchi, T., et al., *Binding of 8-mer to 11-mer peptides carrying the anchor residues to slow assembling HLA class I molecules (HLA-B*5101).* Immunogenetics, 1997. 45(4): p. 259-265.

74. Gul, A. and S. Ohno, *Genetics of Behçet's Disease*, in *Behçet's Syndrome*. 2010, Springer New York. p. 265-276.

75. Falk, K., et al., *Peptide Motifs of Hla-B51, Hla-B52 and Hla-B78 Molecules, and Implications for Behcets-Disease.* International Immunology, 1995. 7(2): p. 223-228.

76. Jarzynski, C., *Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach.* Physical Review E, 1997. 56(5): p. 5018-5035.

77. Park, S., et al., *Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality.* Journal of Chemical Physics, 2003. 119(6): p. 3559-3566.

78. Jensen, M.O., et al., *Energetics of glycerol conduction through aquaglyceroporin GlpF.* Proceedings of the National Academy of Sciences of the United States of America, 2002. 99(10): p. 6731-6736.

79. Hummer, G., *Fast-growth thermodynamic integration: Error and efficiency analysis.* Journal of Chemical Physics, 2001. 114(17): p. 7330-7337.

80. Park, S. and K. Schulten, *Calculating potentials of mean force from steered molecular dynamics simulations.* Journal of Chemical Physics, 2004. 120(13): p. 5946-5961.

81. Stewart-Jones, G.B.E., et al., *Crystal structures and KIR3DL1 recognition of three immunodominant viral peptides complexed to HLA-B*2705.* European Journal of Immunology, 2005. 35(2): p. 341-351.

82. Zhang, D.Q., J. Gullingsrud, and J.A. McCammon, *Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain (vol 128, pg 3019, 2006).* Journal of the American Chemical Society, 2006. 128(13): p. 4493-4493.

83. Vashisth, H. and C.F. Abrams, *Ligand Escape Pathways and (Un)Binding Free Energy Calculations for the Hexameric Insulin-Phenol Complex*. Biophysical Journal, 2008. 95(9): p. 4193-4204.

84. Kosztin, D., S. Izrailev, and K. Schulten, *Unbinding of retinoic acid from its receptor studied by steered molecular dynamics*. Biophysical Journal, 1999. 76(1): p. 188-197.

85. Yang, L.J., et al., *Steered Molecular Dynamics Simulations Reveal the Likelier Dissociation Pathway of Imatinib from Its Targeting Kinases c-Kit and Abl*. Plos One, 2009. 4(12): p. -.

VITA

Mert Gür was born in Hamburg, Germany, on May 23, 1983. He received the B.Sc. Degree in Mechanical Engineering from Middle East Technical University (METU) in 2006. In 2006-2007 he finished one year of M.Sc studies in the department of mechanical engineering at Koc University. Starting his PhD at the same university in 2007, he received the Ph.D. degree from Koç University in Computational Science and Engineering in 2010. From September 2006 to September 2010 he worked as teaching and research assistant at Koç University.

His main research interest was theoretical and computational molecular-scale biophysics problems. He investigated the statistical thermodynamics of protein fluctuations, mode coupling in proteins, free energy pathways in protein binding-unbinding problems and unharmonicity in protein fluctuations. In order to find solutions to the selected problems in their research fields they applied analytical and theoretical models as well as molecular dynamic simulations

He has published several articles in prestigious journals such as Journal of Chemical Physics and Physical Biology.

He will continue his academic career as a Postdoctoral Associate in the Department of Computational & Systems Biology, School of Medicine, University of Pittsburgh, where he will focus on dynamics of membrane proteins.