Protein intrinsic normal modes are correlated with pre-existing conformer populations

by

Beytullah ÖZGÜR

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

Master of Science
in
Computational Science and Engineering

Koç University
October, 2010

Koc University

Graduate School of Sciences and Engineering

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

Beytullah ÖZGÜR

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

Committee Members:

_____

Assoc. Prof. Özlem Keskin(Advisor)

_____

Prof. Attila Gürsoy (Co-advisor)

_____

Prof. Burak Erman

_____

Assist. Prof. Mehmet Sayar

_____

Assist. Prof. İpek Başdoğan

Date: 07.10.2010

ABSTRACT

Proteins are indispensible components of cellular functions. Although protein structure is determined as a static picture for most of the proteins, the dynamics or the time-dependent behavior of proteins act as main contributors to protein function. In addition to ligand induced structural motions proteins also bear intrinsic motions arising from thermal energy they contain. These ligand-independent motions possess functional importance according to experimental evidences for a large number of proteins and a link between these motions and functional motions are established both in terms of structure and timescale. These intrinsic fluctuations are revealed by low-frequency individual modes of proteins which are determined using a simplified version of normal mode analysis termed as Anisotropic Network Model (ANM). In this study, we apply modal analysis to eleven proteins including enzymes, antibodies and signal proteins. We investigate a kinetic relation between modal analysis and protein motions. For this purpose we employ eigenvalues of Hessian matrix which carry information about the vibrational frequencies of these modes. In ANM studies these eigenvalues are used to determine the low and high frequency modes of protein. Our findings imply a correspondence between eigenvalues and kinetic/thermodynamic properties of protein motions. These intrinsic motions establish a dynamic equilibrium between distinct conformers of the protein and as we have proposed eigenvalues of Hessian matrix correlate well both with the timescales of these motions and thermodynamics of these motions.

ÖZET

Proteinler hücresel fonksiyonların vazgeçilmez elemanlarıdırlar. Çoğu protein yapı olarak durağan bir resme sahip olmasına rağmen, protein dinamiği, bir başka deyişle zamana bağlı protein davranışı protein fonksiyonuna önemli derecede katkıda bulunur. Liganda bağlı, yapısal hareketlerin dışında, proteinler içerdikleri termal enerji sayesinde kendilerine özgü hareketler gerçekleştirebilirler. Liganda bağlı olmadan proteinde yer alan bu yapısal hareketler fonksiyonel olarak önem taşımaktadırlar ve farklı proteinler için yapılan deneysel çalışmalara gore bu hareketler ile fonksiyonel hareketler arasında yapısal ve hareketin süreci bakımından ilişki bulunmaktadır. Bu içsel salınımlar, normal mod analizinin basitleştirilmiş bir versiyonu olan ANM ile belirlenen düşük frekanslı, protein modları tarafından açığa çıkarılabilmektedir. Biz bu çalışmada içerisinde enzim, antikor ve sinyal proteinlerinin bulunduğu 11 farklı proteine mod analizi uyguladık ve mod analizi ile protein hareketleri arasında kinetik olarak herhangi bir bağlantı olup olmadığını irdeledik. Bu amaçla Hessian matriksinden elde ettiğimiz özdeğerleri kullandık. Bu özdeğerler modların titreşim frekanslarına dair bilgi içermekte ve ayrıca ANM çalışmalarında özdeğerler düşük ve yüksek frekanslı modları belirlemektedir. Bizim elde ettiğimiz sonuçlar özdeğerler ile protein hareketlerinin kinetik ve termodinamik özellikleri arasında bir bağlantıyı işaret etmektedir. Önceden de belirtildiği gibi bu içsel hareketler, protein ait farklı yapılar arasında dinamik bir denge oluşturmaktadır ve çalışmamızda önerdiğimiz gibi Hessian matriksinin özdeğerleri, protein hareketlerinin zaman periyodu ve termodinamik özellikleriyle bağdaşmaktadır.

ACKNOWLEDGEMENTS

I would like to thank to my supervisors Prof. Attila Gürsoy and Assoc. Prof. Özlem Keskin for their guidance and patience throughout my graduate study and thesis. I am grateful to take part in COSBI group and appreciate group member's kind interests and patience during my graduate study. I shall also express my gratitude to Koc University for their financial support and Engineering Faculty members for their supports. I owe much to my office mates for their kind interest, patience and friendship not just for my graduate study but for my two years which contributed much to me. Special thanks go to my family and my friends who provided me a comfortable environment during my Master period.

TABLE OF CONTENTS

Protein intrinsic normal modes are correlated with pre-existing conformer populations i

ABSTRACT I

ÖZET II

ACKNOWLEDGEMENTS III

TABLE OF CONTENTS IV

LIST OF FIGURES VII

LIST OF TABLES VII

INTRODUCTION 1

1. LITERATURE REVIEW 5

1.1. Energy Landscape Theory 5

1.2. Conformational Selection Model 7

1.3. Protein Motions 8

1.3.1. Fast-Scale Motions 8

1.3.2. Slow-Scale Motions 9

2. COMPUTATIONAL METHODS 14

2.1. Gaussian Network Model 15

2.2. Anisotropic Network Model 18

2.3. Previous Studies 19

2.4. Available Online Tools 25

2.5. Comparison of EN models 26

2.6. Modal Analysis 27

2.7. Mode Selection and Investigating Relation between Thermodynamics/Kinetics and Mode Contributions....32

3. DATASET PROTEINS....36

3.1. Adenylate Kinase....36

3.2. Monoclonal Antibody: Spe7....37

3.3. Ribonuclease A (RNase)....38

3.4. Dihydrofolate Reductase (DHFR)....39

3.5. Maltose Binding Protein (MBP)....41

3.6. Cytochrome P450 EryK....42

3.7. Triosephosphate Isomerase (TIM)....43

3.8. Thrombin....44

3.9. 34E4 Fab Antibody....45

3.10. Cholesterol Oxidase....46

3.11. NtrC....47

4. RESULTS AND DISCUSSION....48

4.1. B-Factor Comparisons....48

4.2. Modal Analysis....49

4.2.1. Adenylate Kinase (AK)....51
4.2.2. Spe7....52
4.2.3. RNase....53
4.2.4. 34E4 Fab Antibody....54
4.2.5. Maltose Binding Protein (MBP)....55
4.2.6. Thrombin....56
4.2.7. Trisophosphate Isomerase (TIM)....57
4.2.8. Dihydrofolate Reductase (DHFR)....58
4.2.9. Cytochrome P450....59
4.2.10. NtrC....60
4.2.11. Cholesterol Oxidase....60

4.3. Mode Selection....61

CONCLUSION....66

APPENDIX 1 ..... **68**

APPENDIX 2 ..... **69**

APPENDIX 3 ..... **91**

BIBLIOGRAPHY **92**

VITA **99**

List of Figures

Figure 1-1 Protein folding according to unique pathway model (A) and funnel concept (B) [33]. 6
Figure 2-1: Depiction of our thesis study 35

List of Tables

Table 1 The proteins used in this study. 50
Table 2 Kinetic rates / relative populations of dataset proteins 63
Table 3: Selected Modes for proteins with kinetic rates..... 65
Table 4: Selected modes for proteins with relative populations 65

Table A. 1 B-Factor prediction comparisons of dataset proteins..... 68

Table A.2. 1 Mode Graphs and Contribution Spectrum of first 30 modes for AK (open form)..... 69
Table A.2. 2 Mode Graphs and Contribution Spectrum of first 30 modes for AK (closed form)..... 70
Table A.2. 3 Mode Graphs and Contribution Spectrum of first 30 modes for Spe7 (Ab^1 form)..... 71
Table A.2. 4 Mode Graphs and Contribution Spectrum of first 30 modes for Spe7 (Ab^2 form)..... 72
Table A.2. 5 Mode Graphs and Contribution Spectrum of first 30 modes for RNase (open form)..... 73
Table A.2. 6 Mode Graphs and Contribution Spectrum of first 30 modes for RNase (closed form)..... 74
Table A.2. 7 Mode Graphs and Contribution Spectrum of first 30 modes for Fab (open form)..... 75
Table A.2. 8 Mode Graphs and Contribution Spectrum of first 30 modes for Fab (non-binding form)..... 76

Table A.2. 9 Mode Graphs and Contribution Spectrum of first 30 modes for MBP (open form) 77
Table A.2. 10 Mode Graphs and Contribution Spectrum of first 30 modes for MBP (partially closed) 78
Table A.2. 11 Mode Graphs and Contribution Spectrum of first 30 modes for Thrombin (slow form) 79
Table A.2. 12 Mode Graphs and Contribution Spectrum of first 30 modes for Thrombin (fast form) 80
Table A.2. 13 Mode Graphs and Contribution Spectrum of first 30 modes for TIM (open form) 81
Table A.2. 14 Mode Graphs and Contribution Spectrum of first 30 modes for TIM (closed form) 82
Table A.2. 15 Mode Graphs and Contribution Spectrum of first 30 modes for DHFR (NADPH binding form) 83
Table A.2. 16 Mode Graphs and Contribution Spectrum of first 30 modes for DHFR (DHF binding form) 84
Table A.2. 17 Mode Graphs and Contribution Spectrum of first 30 modes for Cytochrome P450 (open form) 85
Table A.2. 18 Mode Graphs and Contribution Spectrum of first 30 modes for Cytochrome P450 (closed form) 86
Table A.2. 19 Mode Graphs and Contribution Spectrum of first 30 modes for NtrC (inactive form) 87
Table A.2. 20 Mode Graphs and Contribution Spectrum of first 30 modes for NtrC (active form) 88
Table A.2. 21 Mode Graphs and Contribution Spectrum of first 30 modes for Cholesterol Oxidase (substrate bound form) 89
Table A.2. 22 Mode Graphs and Contribution Spectrum of first 30 modes for Cholesterol Oxidase (FAD bound form) 90

Table A.3 1 Performance of Random Modes 91

INTRODUCTION

Protein function is highly determined by protein structure and dynamics which are linked in an ultimate level. Architecture of a protein provides itself a rigidity and elasticity that shapes the functional capability of the protein. This ultimate relation between structure and dynamics enable scientists to predict the protein dynamics using solely the protein structure. For this purpose several computational efforts emerge including molecular dynamics and normal mode analysis. In the former one, protein is allowed to follow Newtonian equations of motion in a deterministic way and the protein behavior is monitored through molecular dynamics trajectories [1, 2]. In the latter one, the normal modes of protein motion are extracted through Harmonic motion approximation [3, 4]. Both computational methods help to expand the knowledge about protein dynamics.

The normal mode analysis is simplified by adopting solely C_α atoms of protein instead of all-atoms. These new methods, called Elastic Network Models (ENMs) are inspired both by normal mode analysis and the elasticity theory of random polymer networks [5]. ENMs perform well [6-10] to predict B-factors (also called temperature factors, Debye-Waller factors) which store the mean square fluctuation information of each atom. Further studies about the protein dynamics accomplished by ENMs indicate that the residue fluctuation information stored in global modes correlate well with the conformational transitions of the protein [11-14]. Besides, deformation of the open form along one of the three low frequency modes, in other words global modes transfer the open form to closed form indicating that the conformational transition information is encoded in protein structure.

The previous findings of ENMs actually support preexisting equilibrium model. According to preexisting equilibrium model, both open and closed forms of the protein preexist together in solution and ligand binding acts as a stabilizer for the closed form and eventually, the probability distribution of conformers are altered towards the closed form [15-17]. In a more generalized view, these conformers form an ensemble where

the ensemble dynamics include fluctuations within the same energy level, termed as fast fluctuations [18], and slow scale motions which enable the protein to sample several ensemble conformers [18, 19]. ENM modes deal with both fast and slow scale motions via predicting B-factors (fast motions) and conformational transitions (slow scale motions).

Gaussian Network Model (GNM) and Anisotropic Network Model (ANM) are two types of ENMs. These two methods differ from each other regarding the adopted matrices which store inter-residue potentials in which GNM employs an N-by-N matrix whilst ANM adopts a 3N-by-3N matrix. The fluctuation information supplied by GNM is isotropic (only amplitude) whilst ANM provides both amplitude and direction information for observed fluctuations. The so-called individual modes are constructed using two outcomes of eigenvalue decomposition of the adopted matrix eigenvectors which include mode shapes and eigenvalues which store the frequency information of the corresponding mode. Besides, the contribution of each mode is determined based on the eigenvalue spectrum of the protein in which low frequency modes would contribute much to the overall motion.

As aforementioned open form encodes the closed or closed-like conformers through its thermal fluctuations and these fluctuations can be accessed by low-frequency normal modes of the protein. In literature, a structural relation between the normal modes and conformational transitions between distinct states of proteins has been established through mode analysis of several proteins [12, 13, 20, 21]. In further attempts ANM modes of unligated form of a protein have been found to sample conformational space accessed by different ligand bound forms of the same protein [14]. However, this structural relation between ENM and protein dynamics does not provide any insights about the timelines of protein fluctuations since ENM calculations do not possess any time dependence. There are several evidences provided by NMR experiments [15, 22-25] that protein backbone dynamics drive turnover rates or kinetic rates of conformational transitions highlighting the role of protein dynamics during protein function.

In this thesis, we have combined ANM modal analysis and experimental evidences about protein kinetics/thermodynamics in order to establish a relation between ANM findings and kinetics/thermodynamics of the proteins. As we have stated above ENM is already proven to correlate well with the conformational transitions and we have simply intended to expand success of modal analysis over thermodynamics and kinetics of proteins. For this purpose, we have initially constructed a protein data set that consists of eleven different proteins including enzymes, signal proteins and antibodies. All proteins within the dataset are proven to co-exist as at least two conformers in solution via experimental studies. The thermodynamic/kinetic data regarding transition rates or relative populations of these eleven proteins have been retrieved from NMR or single-molecule FRET studies of these eleven proteins. Therefore, we have a protein dataset comprised of 22 PDB structures (two for each protein) and their respective kinetic or thermodynamic data.

We have performed modal analysis on both structures of eleven proteins using ANM and determined the similarity between first thirty modes and observed experimental displacement via employing two similarity measurements. Mode correlation and overlap scores give amplitude and direction similarities of these two vectors. Once we have calculated similarity scores of individual modes of each conformer, we in further steps determined top three modes considering those similarity scores and eigenvalue spectrum of corresponding conformers.

Modal analysis of eleven proteins have been followed by main emphasize of our study. We have proposed a relation between eigenvalues of Hessian matrix and kinetic or thermodynamic properties of ensemble conformers. For this purpose we have constructed two vectors termed as experimental and computational vectors. Experimental vector included either the ratios of kinetic rates of conformational change which transfers one conformer to other one or this vector included ratios of relative populations of each conformer depending on the data that we have retrieved from literature. Second vector included ratios of eigenvalues of inverse Hessian matrix which simply corresponds to contribution of each mode. Here, we have iterated top three

modes in order to select the significant modes and calculate ratios for computational vector. Thus, the results presented in the last part of Results correspond to Pearson correlation between these experimental and computational vectors. We have found that for both kinetic rates and relative populations, ANM modal analysis produces significant results implying a relation between ANM modal analysis and kinetic and thermodynamic properties of proteins dynamics.

Prior to our modal analysis we have also investigated success of five different ENMs regarding B-factor predictions. We have employed two GNM methods in which one is cutoff-derived (GNM) whilst the other uses a parameter free method (pfGNM). We have also adopted three ANM methods, one is cutoff derived (ANM) and two of them use parameter free methods (pfANM2 and pfANM4). Parameter free methods have been previously found to improve B-factor predictions compared to classic cutoff-derived methods [26]. We have used a dataset comprised of more than a thousand proteins and our prediction results imply similar improvements in B-factor predictions in the case of parameter free methods.

As a future work, the relation between kinetic/thermodynamic data and combined modes can be investigated for a larger dataset. Our results strictly require validation using a larger dataset including different types of proteins.

Chapter 1

1. LITERATURE REVIEW

1.1. Energy Landscape Theory

Protein structure and dynamics constitute a major interest within the biological studies. Functional importance of proteins such as catalysis, regulation and interactions are all directed by protein structure and dynamics. The thermal energy stored in the protein enables the protein itself to explore a large ensemble of conformations around a well defined energy level. Thus, native structure of the protein does not possess a static structure, but represents the average structure of all ensemble structures that are called 'substates'. The conformational ensemble of the protein populates several conformational states, also called substates in differing amounts within a multidimensional energy landscape and the interconversions between these substates determine the behavior of the protein.

Folding funnel hypothesis [27, 28] which has emerged from protein folding studies proposes presence of multiple conformational states and folding pathways that protein can follow in order to reach its native conformation on contrary to the Levinthal's idea of unique folding pathway [29] that is later known as 'Levinthal's paradox' [30]. In a similar manner, folded proteins also exist as multiple states according to study of Frauenfelder et. al for the myoglobin [31]. In this study, they have determined four different binding processes and for the interpretation of the data they have proposed that protein populates four distinct states that are both temperature and concentration dependent. In another study [32], solvent is found as an active participant, equally important as the energy landscape of myoglobin.

The energy landscape theory replaces the Levinthal's paradox for both folding and folded proteins and proposes a new funnel concept instead of the pathway concept. According to funnel landscape model, number of accessible conformations or conformational entropy is represented by the lateral area of energy landscape in any given depth where the depth indicates progress of folding [33]. As the folding advances

number of accessible substates decreases together with degree of freedom and eventually protein reaches its native state. Figure 1-1 shows the folding process according to two different models; unique pathway model (Fig.1A) and funnel concept (Fig 1B). Funnel concept contains multiple pathways that protein can follow and conformational entropy narrows down as the folding advances. N is the native state of the protein. The lateral area at any given depth defines the conformational entropy or number of accessible states [33].

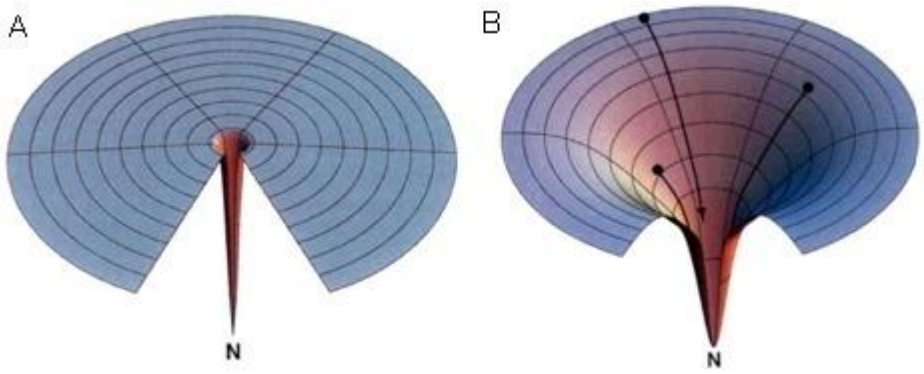


Figure 1-1 Protein folding according to unique pathway model (A) and funnel concept (B) [33].

In terms of energy landscape funnels, the protein behavior is related with the smoothness of this funnel. Bumpy landscapes can lead protein to be stuck in local minima slowing down the folding process. In the case of native state if the landscape exhibits a smooth behavior, thermal fluctuations will lead to small structural changes within the protein. If landscape contains bumpy regions the small fluctuations in energy can lead to large structural changes eventually to "conformational distant relatives" also called as excited states of the protein. The energy barriers separating distinct energy states determine the transition rates between these conformers and also the relative population of each state [33].

To sum up, conformational ensemble of a folded protein spans its energy landscape in a multidimensional manner together with the substates (distinct energy states) and the energy walls separating these substates. The transition rates (kinetics) and relative

populations (thermodynamics) of each state are largely affected by solvent and conditions.

1.2. Conformational Selection Model

Protein folding and function are all directly related with free energy landscape of the protein as aforementioned. The energy landscape theory also has roots in molecular recognition mechanisms. 'Lock and Key' model [34]proposes a recognition mechanism with any structural change neither in protein nor in ligand. This model is followed by Koshland's induced fit model [35] which proposes a ligand based induction within the active site of protein. Both models assume the protein to exist as a single, stable conformation and fail to reflect dynamic nature of protein. Contrary to these two models a new model called "conformational selection model" emerges deriving from energy landscape model. Energy landscape of a protein contains distinct energy states distinguished by energy walls. Each energy state is populated in different amounts in certain conditions and conformational selection model proposes that the substates of corresponding energy landscape plays an important role in molecular recognition and ligand preferably interacts with any of these substates. Ligand interaction alters the population distribution of each substate and leads to a population shift favoring the ligand binding form [36].

This model is supported by several experiments for enzymes [15, 26, 37-41], monoclonal antibodies [16, 42, 43] and regulatory proteins [44]. According to enzyme studies referred above both the open and closed forms of the enzyme are co-populated favoring the open form in the absence of the ligand. Ligand binding simply shifts the relative populations of open and closed forms shifting the equilibrium towards the closed form. These studies mutually conclude indispensible role of pre-existing substates for ligand recognition and binding.

Two antibody examples, Spe7 [16, 43] and 34E4 Fab Antibody [42] support conformational selection model and demonstrate that multiple substates co-exist in solution, interconverting between each other as a result of thermal fluctuations. The

presence of pre-existing substates in solution provide ligand diversity for these antibodies and in the presence of the ligand, the populations are redistributed favoring the ligand binding conformation.

1.3. Protein Motions

Driven by the energy landscape of protein, the timescales and amplitudes of protein motion constitute the protein dynamics. The motions which occur in femtosecond to nanosecond timescales are termed as fast scale motions and the motions which span a timescale of microseconds to milliseconds are termed as slow scale motions. These two types of motion lead to different energetic and structural changes within the protein in terms of energy landscape [45].

1.3.1. Fast-Scale Motions

Fast scale motions are coupled to local, small amplitude fluctuations between close energy states that has small energy barriers around kT [18]. This type of motions with amplitudes of lower than 1.5 Å enables the protein to sample several structurally similar states. Side chain rotations and small amplitude backbone fluctuations are examples of fast scale motions. NMR relaxation experiments [17, 46] allow investigation of fast scale dynamics in terms of bond fluctuations. NMR spectroscopy is also used extensively in order to determine entropic contribution of protein to binding. Calmodulin, which binds a variety of target proteins, demonstrates a good example for the application of NMR to biomolecular binding. Entropic and enthalpic contributions of Ca^{2+}-calmodulin differ widely when binding to different peptides although the binding free energies are similar for target peptides determined by isothermal calorimetry [47-49]. The methyl-order parameters of calmodulin binding to six different target peptides were studied [49] in order to question observed differences in entropic and enthalpic contribution of the protein. In this study, it was shown that sub-nanosecond dynamics of methyl-bearing amino acid side chains exhibit significantly varying motion depending on the nature of the target peptide. The backbone dynamics

on the other hand, are invariant across the complexes [49]. This study concludes that fast side chain dynamics of calmodulin are employed in order to tune its affinity for targets.

X-ray diffraction data supplies information about the average three dimensional structure together with the fast scale fluctuations around this conformation, commonly known as X-ray crystallographic B-factors [50]. The information supplied by X-ray B-factors provides insights about the mean square fluctuations of atomic displacement [51] in an isotropic manner. The directionality information about these fast scale motions can be obtained using higher resolution X-ray technology in sub-angtrom scale [52]. Infrared or fluorescence correlation spectroscopy can also provide insights about local fluctuations [18].

Fast-motions between the ranges of femtoseconds to nanoseconds are within the scope of molecular dynamics (MD), quantum mechanics and molecular mechanics. As stated above the fluctuations within the same energy state can be determined by the experimental methods. However, still an understanding about the exact nature of the motions and the underlying forces driving these motions remain as unknowns. MD simulations cover these questions and help the literature to improve in two ways, first by complementing readily available knowledge about the dynamics and second by inspiring new experimental studies. Since the mid-seventies MD simulations became an integral part of protein dynamics studies through advances in computational power and improvements in accuracy of mathematical models. Now, MD simulations reach up to tens of microseconds scale with a common system of 10^4-10^6 atoms [53].

1.3.2. Slow-Scale Motions

Slower scale motions ranging between microseconds and milliseconds are collective, including large-amplitude motions of secondary structure elements, subunits or domains fluctuating between energetically distinct substates separated by energy barriers of several *kT* [18, 19]. These slow-scale fluctuations enable a protein to resemble relatively more stable states in which inter-conversions between these states

carry importance about the protein function including enzyme catalysis, signal transduction and biomolecular interactions. Consistent with the energy landscape idea these states form the conformational ensemble of the protein with different attributes based on the energy profile such as interconversion rates and relative populations.

NMR spectroscopy and X-ray crystallography are primarily used in order to elucidate the substate structures in atomic resolutions and further information about the kinetics and thermodynamics of protein dynamics can be elucidated using NMR methods. The slow motions as aforementioned in microseconds to seconds scale lead to conformational exchange between kinetically distinct states. In an NMR experiment, these transitions between different magnetic environments lead to a [23, 54] contribution in the transverse relaxation rate (termed as R_2). From this term, three properties of the system can be extracted; chemical shifts between exchanging sites ($\Delta\omega$), transition rates between two states (k_{ex}, kinetics) and relative populations of sites A and B (p_A and p_B). The observed $\Delta\omega$ does not necessarily indicate a structural change, however it refers that the electronic environment around an atom differs between two states [23]. Further information about the protein dynamics depend on the interpretation of these NMR findings.

Cyclophilin A (CypA), which is a cis-trans isomerase, [55] is a good example about NMR spectroscopy studies for conformational change. Eisenmesser et. al. studied CypA dynamics using NMR spectroscopy [22] and this study elucidated a relation between enzyme flexibility and catalysis. The conformational changes taking place during the turnover were found to pre-exist in the free enzyme with frequencies similar to catalytic turnover rates implying the catalysis to be an inner property of the free protein [22]. In the same study authors concluded that CypA dynamics associated with catalysis is a built-in property of the enzyme that is also apparent in the free CypA. For the cytochrome P450 the ensemble dynamics and interconversion rates were determined by crystallographic techniques [26] and within the same study the interconversion rates between corresponding substates have been presented.

NMR relaxation and dispersion experiments provide knowledge also about the relative populations of the ensemble substates and chemical shift differences between corresponding substates. Characterization of catalytic motions of free RNase [56] revealed three substates that pre-exist together in conformational ensemble including free enzyme, enzyme-substrate complex and enzyme-product complex. Such studies established the innate slow scale motions to be relevant to functional motions indicating the effect of the ligand to be relatively small for these motions. The ligand binding acts to stabilize the pre-existing conformer, altering the probability distribution.

Single molecule FRET (Flourescence Resonance Energy Transfer) is another strong technique to observe different state of the proteins. This technique depends on the energy transfer between the donor and acceptor [57]. The changes in the fluorescence intensity and excited-state lifetimes of both donor and acceptor supply evidences about single-molecule dynamics. Two-state behavior of the proteins can be elucidated in time-resolved FRET by determining the relative distributions in which, each conformational state will demonstrate a different FRET value. FRET also performs well in order to detect conformational transitions [57]. Study of Diez and colleagues [58] about F_0F_1-ATP synthase is an illustrative example about the applicability of single-molecule FRET. This enzyme performs ATP catalysis from ADP and phosphate, in bacteria, mitochondria and chloroplast, through converting electrochemical energy supplied by transmembrane electrochemical proton gradient into chemical energy. The authors employ single-molecule FRET in order to monitor rotations in the γ-subunit of the F_1 part taking place during the ATP synthesis using intact protein complex in liposomes that allows measurements up to hundred of miliseconds. The distances between rotating γ-subunit and static β subunit were determined by attaching two fluorophores to each subunit. The study monitored three distinct distances between two subunits and also capturing rotations between these three states. According to this study, the direction of rotation is reversed during the ATP synthesis compared to ATP hydrolysis. The single-molecule FRET experiments allow studying consecutive and progressive dynamics

which is only possible by employing single-molecule experiments due to synchronization problems.

As aforementioned slow motions reach up to milliseconds exceeding accessibility of conventional MD simulations. However, these motions contain biologically important processes such as protein folding, conformational transitions between distinct states and catalysis. Several simplifications to force fields have been developed including normal mode analysis [4, 59, 60]], elastic network models (will be emphasized in next chapter) [6-9] and the simplified version of MD simulations order to reach microsecond timescales. These simplifications of molecular dynamics include acceleration of system dynamics by employing external forces and the most familiar ones are targeted, steered and accelerated molecular dynamics [53, 60, 61].

The molecular dynamics study about the native state ensemble of ubiquitin [62] exemplifies the ability of molecular dynamics simulations in order to sample slow motions. According to that study, together with the fast dynamics of ubiquitin, slower motions on nanosecond to microsecond timescales occur during the simulations and the observed structural heterogeneity overlaps with the variability observed in different crystal structures of ubiquitin. As we have explained before this behavior of the protein is related with energy landscape theory which enables the protein to populate different substates within an ensemble of conformers together with interconversions between each conformer. In a similar study, Henzler-Wildman et. al. [18] combined both NMR methods and single molecule-FRET experiments together with molecular dynamics simulations and the MD results elucidated that free form of the protein samples the partially closed form in smaller timescales (about nanoseconds) compared to results of NMR and FRET experiments (microseconds to miliseconds).

Molecular dynamics have been also adopted for protein folding studies [63, 64] in microsecond time scales and both studies concluded that the period of protein folding is largely related with the initial structure used in simulations. Selection of force-field is also another contributor to success of simulations since finding accurate force fields that would best suit to the folding process possesses problems [65, 66]. Protein folding is

studied also by adapting energy landscape model [67] through using a simplified model called umbrella sampling [68] in order to determine the equilibrium population distribution of the protein. Initial structures for umbrella sampling are obtained by unfolding simulations and at the end of folding simulations driven by umbrella sampling the density of conformational states are calculated, and subsequently free energy landscape is determined.

Chapter 2

2. COMPUTATIONAL METHODS

In the previous chapter, protein dynamics, together with timescales of motions have been presented in terms of energy landscape theory. Several experimental studies provide evidences for this theory making it applicable to both protein folding and folded proteins. Computational efforts also produce significant results in agreement with experimental results providing insights about the nature of fluctuations. Molecular dynamics simulations perform well for the fast timescale fluctuations of proteins, but require simplifications in order to deal with slower motions which carry functional importance because of computational limitations.

As proposed in previous chapter several coarse-grained methods have been developed in order to eliminate limitations of molecular dynamics simulations. Full atomic, empirical force fields are replaced with coarse-grained models in order to determine functional motions of proteins. Normal mode analysis (NMA) [4, 59, 60] adopts a simplified version of potential or conformational energy of the protein in order to determine the vibrational motions assuming that thermal fluctuations of the protein are harmonic neglecting any anharmonicity in calculations. Bond angles and lengths are fixed only permitting bond rotations thus only dihedral angels are taken into consideration. "The conformational energy is approximated by the multidimensional parabola characterized by second-derivative matrix at the minimum [4]." Eigenvalue decomposition of second-derivative matrix produces normal modes of vibration and their frequencies. Superposition of computed normal modes approximates the dynamics of protein and the calculations presented in paper about the globular protein is in good agreement with B-factors of the same protein obtained by X-ray crystallography [4].

Tirion [11] replaced the detailed parameters which are used to express potential energy by Hookean pairwise potential between two interacting pairs and the potential energy is defined as the sum of pairwise potentials. Tirion adopted a cutoff distance in order to determine number of pairwise interactions which will define the size of

computation. The author further discusses the applicability of this simplified model and concludes that the details neglected in the calculations do not contribute much to the slow vibrations which are collective motion driven by a large number of atoms. So the simplifications perform well for slow motions and reduce the computational cost considerably compared to the cost of molecular dynamics simulations of NMA. The calculations of Tirion including theoretical temperature factors calculated by using lowest 30 modes produce successful results with the temperature factors for G-actin.

2.1. Gaussian Network Model

Inspired by normal mode analysis and elasticity theory of random polymer networks [5, 69, 70] another simplified method has emerged based on the inter-residue potentials of C_α atoms different from normal mode analysis which adopts coarse-grained approach [6, 7]. This method, called Gaussian Network Model (GNM), treats proteins in their native states as three-dimensional elastic networks (EN) in which residues and bonds are represented as nodes and springs respectively. Residue fluctuations follow Gaussian distribution and are determined by the interaction potentials between residue itself and closely located neighbor residues. As in normal mode analysis, GNM also assumes residue fluctuations to follow harmonic behavior.

Inter-residue potentials for interacting pairs are independent from the residue types and are determined based solely on the inter-residue distances. The distances are calculated using atomic coordinates of C_α supplied in X-ray crystals or NMR structures. The Kirchoff (Γ) matrix, which is a symmetric N-by-N dimensional matrix, includes information about the interaction potential of each residue pair and is determined as in Eq. 2.1:

$$\Gamma_{ij} = \begin{cases} -\gamma, if\ i \neq j\ and\ R_{ij} \leq r_c \\ 0, if\ i \neq j\ and\ R_{ij} \geq r_c \\ -\sum_{i \neq j} \Gamma_{ij}, if\ i = j \end{cases} \quad (Eq.\ 2.1)$$

In Eq. 2.1 Γ_{ij} is the interaction potential between i^{th} and j^{th} residues, γ is the empirical normalization constant, counterpart of the single-parameter in the Hookean potential of Tirion's work [11], R_{ij} is the distance between i^{th} and j^{th} residues and r_c is the cutoff value which determines interacting residue pairs. If the inter-distance between two residue pair is below defined cutoff r_c, then these residues are connected with a spring and the interaction potential for these residues are stored in Γ as $-\gamma$. If the residues are located far from the cutoff value (r_c), no interaction potential is assigned between them. The diagonal elements of the Kirchoff matrix store the total number of bonded and nonbonded interactions for corresponding residue indices.

The potential energy (V) of the protein associated with fluctuations of C_α atoms can be computed using Kirchoff matrix (Γ) and N dimensional fluctuation vector ($\Delta\mathbf{R}$) whose elements are the fluctuation vector of each residue as presented in Eq. 2.2:

$$V = \Delta \boldsymbol{R}^T \Gamma \Delta \boldsymbol{R} \; (Eq. 2.2)$$

The configurational partition function of a network with N nodes (Z_N) which defines the statistical attributes of the system in terms of thermodynamics can be expressed as in Eq. 2.3 with an analogy to the studies about the random Gaussian networks [5, 69, 70]:

$$Z_N = K\, e^{-\Delta \boldsymbol{R}^T \Gamma \Delta \boldsymbol{R}} \; (Eq. 2.3)$$

The cross-correlations between residue fluctuations of ith and jth residues can be written as in Eq. 2.4 and consequently takes form as in Eq. 2.5:

$$< \Delta R_i \cdot \Delta R_j > = \left(\frac{1}{Z_N}\right) \int \left(\Delta R_i \cdot \Delta R_j\right) e^{-V/kT}\, d\{\Delta R\} \; (Eq. 2.4)$$

$$< \Delta R_i \cdot \Delta R_j > = \frac{3kT}{\gamma} \, [\Gamma^{-1}]_{ij} \; (Eq. 2.5)$$

Here, k is the Boltzmann constant and T is the temperature. Γ^{-1} is the inverse of the Kirchoff matrix. If the Γ^{-1} is written in terms of eigenvectors and eigenvalues of Γ the cross-correlations take form as shown below in Eq. 2.6:

$$< \Delta R_i \cdot \Delta R_j > = \sum_k [\Delta R_i \cdot \Delta R_j]_k = \frac{3kT}{\gamma} \sum_k [\lambda_k^{-1} \boldsymbol{u}_k \boldsymbol{u}_k^T]_{ij} \; (Eq. 2.6)$$

The equation tells us that the total cross-correlations can be expressed as superposition of fluctuations distributed to each individual mode of peptide (termed as k here). An elastic network with N residues contains N-1 nonzero eigenvalues and the same number of modes. λ_k is the eigenvalue of k^{th} mode and u_k is the corresponding eigenvector. Eq. 2.6 signifies the computational efficiency of GNM which requires only eigenvalue decomposition of Γ matrix and the matrix depends solely on the C_α atom positions of residues extracted from X-ray crystal or NMR structures.

From Eq.6 one can also find mean square fluctuations of each residue using Γ matrix. The fluctuation amplitudes determined in GNM are isotropic, containing no direction information. As in the work of Tirion [11], the GNM findings are validated using Debye-Waller factors which are calculated using Eq.2.7:

$$B_k = 8\pi^2 \frac{\langle \Delta R_k \cdot \Delta R_k \rangle}{3} \quad (Eq. 2.7)$$

Here B_k is the B-factor of k^{th} atom and ΔR_k is the fluctuation amplitude of k^{th} atom. The results between theoretical findings and experimental B-factors exhibit good agreement indicating that GNM performs well in order to determine the fluctuation pattern of the protein. In this model only coordinates of C_α atoms are used thus any interactions involving sidechain atoms are ignored within the GNM calculations. Thus, GNM fail to predict the motion successfully within the regions where the stability is provided by interactions involving sidechain atoms.

The cutoff value adopted in GNM is 7.0 Å and includes all neighbors of the corresponding residue within the first coordination shell. Another empirical parameter γ is the normalization factor determined by normalizing theoretical mean-square fluctuations with respect to the experimental ones and depends on the secondary structure content and the size of the protein [6]. The proteins with a large content of secondary structures exhibit larger γ values. The adopted cutoff value is also important in order to determine γ value. Adoption of larger cutoff values will increase the interaction strength subsequently increasing constraints over atomic fluctuations. Increase in constraints over atomic fluctuations lead to increase in γ value. This value

does not affect the distribution of the fluctuations of each residue, but rescales the amplitude of fluctuations uniformly. As aforementioned, the value of force constant γ does not exhibit differences for different amino acid pairs.

Despite the simplicity of the model, GNM still performs well and exhibits good agreement with experimental findings. The main reason behind success of the GNM is that vibrational motions are largely dominated by collective motions in the native state of the protein and these low-frequency, collective motions contribute much more than the other according to their eigenvalues (Eq. 2.6).

2.2. Anisotropic Network Model

The GNM model provides fluctuation information in an isotropic manner without any further direction information. However, together with the amplitude information, directions of these modes are equally important. Anisotropic Network Model (ANM) has emerged as another elastic network model in order to determine atomic fluctuations with both amplitude and direction information.

ANM [8, 9] also depends on the theory of elastic networks like GNM. The residues are still represented as nodes using atomic coordinates of C_α atoms connected by springs. Like GNM, ANM also adopts a single-parameter force constant γ for all residue pairs with no discrimination done between different residue types. ANM employs Hessian matrix that is counterpart of the Kirchoff matrix used in GNM. Different from GNM Hessian matrix is N-by-N matrix composed of super elements of 3-by-3 matrices. The interaction potential between residue pairs are defined as second derivatives of potential energy with respect to X, Y and Z components of residue pairs as presented in Eq. 2.8:

$$\boldsymbol{H_{ij}} = \begin{bmatrix} \partial^2 V/\partial X_i \partial X_j & \partial^2 V/\partial X_i \partial Y_j & \partial^2 V/\partial X_i \partial Z_j \\ \partial^2 V/\partial Y_i \partial X_j & \partial^2 V/\partial Y_i \partial Y_j & \partial^2 V/\partial Y_i \partial Z_j \\ \partial^2 V/\partial Z_i \partial X_j & \partial^2 V/\partial Z_i \partial Y_j & \partial^2 V/\partial Z_i \partial Z_j \end{bmatrix} \quad (Eq.\,2.8)$$

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1N} \\ H_{21} & H_{22} & \dots & H_{2N} \\ \vdots & & \ddots & \vdots \\ H_{1N} & & & H_{1N} \end{bmatrix} \quad (Eq.\ 2.9)$$

Hessian matrix is a 3N-by-3N matrix and eigendecomposition of this matrix yields 3N-6 non-zero eigenvalues and same number of eigenvectors. In ANM, the cross-correlations between residue pairs is defined as in Eq. 2.10 as modified version of Eq. 2.6

$$< \Delta R_i \cdot \Delta R_j > = \sum_k [\Delta R_i \cdot \Delta R_j]_k = \frac{3kT}{\gamma} \sum_k tr([\lambda_k^{-1} \boldsymbol{u}_k \boldsymbol{u}_k^T])_{ij} \quad (Eq.\ 2.10)$$

Here, k defines the indices of modes and Hessian matrix contains 3N-6 non-zero individual modes and the cross-correlations of each residue pair is determined by taking the trace of 3-by-3 matrix that is trace of H^{-1}_{ij}. The eigenvectors in ANM are comprised of N-dimensional elements with each element containing 3-dimensional vector in which fluctuation information in three Cartesian coordinates are determined.

The adopted cutoff value in ANM is relatively larger than one adopted in GNM. Studies [9] show that adoption of a cutoff about 7.0 Å leads to calculation of more than six zero eigenvalues not as expected. Also, small cutoff values favor large fluctuations in one direction for particular residues. In order to remove these inconveniences ANM adopts relatively larger cutoff values and a spring constant of 1.0 ± 0.5 kcal/(mol.A^2) is determined for cutoff values between 12-15 Å.

Both elastic network models vibrational frequency (ω_i) of each motion can be calculated as $\omega_i = (\gamma\lambda_i)^{1/2}$. Like in GNM, the atomic fluctuations determined in ANM are largely dominated by low-frequency, collective motions. EN models do not require any minimization or equilibration since it depends on the pure harmonic motions of the protein that are extracted using eigenvalue decomposition making it an analytical tool.

2.3. Previous Studies

Elastic Network Models immediately have become a useful technique to study protein dynamics due to their simplicity and no requirement for molecular dynamics

simulations. GNM and its anisotropic counterpart ANM have initially been adopted for studying correlations between experimental B-factors (also called Debye-Waller factors) and mean square atomic fluctuations determined by EN models [6-9]. Doruker et. al. [7] investigated efficiency of GNM, ANM and MD over experimental findings of α-amylase inhibitor. According to this study dispersion of modes within the slow, intermediate and fast scales differ for each method such that slow modes of MD contribute most to the determined fluctuations. ANM becomes intermediate at this scale between MD and GNM however, matches GNM within the intermediate scale. At the fast regimes, the relative importance of GNM and MD is inverted and contribution of fast modes diminishes for MD simulations logarithmically where contributions for ANM and GNM fast modes decrease linearly.

The same study also concludes that GNM follows a smooth pattern for mode shapes, where others, ANM and MD exhibits relatively sharper mode shapes. GNM fails to determine high mobility regions, on the other hand ANM and MD performs this task accurately. The correlated and anti-correlated residue fluctuations are also determined for each mode using EN models [7] together with the collectivity of each mode. In another study [9] authors employed ANM to investigate dynamics of RBP protein and characterized the highly coupled motions of two loop regions which they propose to have functional roles for both gating of cavity and recognition, consistent with the MD simulations of the same protein.

Large conformational changes between open and closed forms of 20 proteins have been studied using normal mode analysis [20]. Low-frequency collective modes of proteins have been adopted in order to define conformational change observed between open and closed forms of the protein. These transitions have essential roles for protein function such as catalysis, regulation and binding. According to this study, one of the low-frequency normal modes of the open form performs well to define the transitions between open and closed form. In this, study the authors employ overlap which determines the similarity between directions of normal modes and corresponding experimental displacement. Success of the normal mode analysis depends on the

collectivity of the experimental displacement according to the study of Tama and Sanejouand [20] and they conclude that the motions with a collectivity larger than 0.18 can be defined by one of the three lowest-frequency modes with an overlap value higher than 0.50. These implications further suggest that distinct forms of the protein can be encoded by the normal modes of the open form and protein sequence is evolved such that favoring a single normal mode.

Further studies of EN models [12, 13, 71] emphasize over the low-frequency collective modes determined by ANM and GNM. Bahar and Tobi [12] elucidated that C_α displacements observed between unbound and bound forms of four different protein pairs which exhibit substantial conformational changes upon ligand binding can be explained using one of the lowest three individual modes which are also called as global modes. These findings of Bahar and Tobi provide evidence for the pre-existing equilibrium model which adopts energy-landscape theory of proteins favoring presence of distinct conformers. Keskin [13] studied antibody diversity in terms of pre-existing equilibrium model using ANM modal analysis. She investigated flexibility of two antibodies using normal modes of the corresponding proteins in their unbound states. The results of this study imply the importance of the intrinsic flexibility of the antibodies for recognition and binding process of antibodies and ANM performs well in order to extract these functional fluctuations through normal modes of the unbound form.

Another attempt for studying large conformational transitions was performed by Sternberg et. al [21] for a large benchmark of proteins which exhibit different scale of conformational changes using EN models. They investigated the effect of several parameters like amplitude of observed conformational change, collectivity and length of protein over success of normal modes. The authors determined the first 20 low frequency modes for each protein and revealed that the proteins with experimentally known as undergoing large conformational changes produce eigenvalues with lower frequencies corresponding to more rigid ones. They have found that protein size can also predict extent of conformational change less reliably than mode frequency. The

authors' results imply that the unbound form of the protein sample bound-like conformations through their low-frequency modes and normal modes can be used efficiently in order to predict the extent of conformational change. Still, bound conformations differ slightly from predicted structures due to final arrangements induced by ligand. This study also concentrates over the importance of normal modes for the docking studies by predicting candidate target conformational subspaces.

Normal mode analysis stands as a powerful analytical tool for investigation of functional protein motions as can be seen in above described studies providing evidences for the pre-existing equilibrium model. Bakan and Bahar [14] expanded normal mode studies of conformational changes to the conformational spaces that protein occupy in their bound forms and investigated the role of global normal modes for the corresponding motions. The authors have adopted three proteins p38 MAP kinase, HIV-1 reverse transcriptase and cyclin-dependent kinase including their unbound forms and the inhibitors binding to them, totally involving 292 structures. They have determined conformational space accessed upon ligand binding for each of three proteins and extracted dominant functional motions through principal component analysis of experimental data. They have also analyzed subsets of protein-inhibitor interactions in order to understand dominant motions based on the type of the inhibitor and binding sites. The study emphasizes mainly over the relation between first two principal components of functional variations and normal modes of unbound form of each protein. The results are promising and imply that there are individual modes that correlate well with these two principal components and generally the corresponding individual mode is among the collective modes of unbound form. The authors conclude that the collective normal modes determined by ANM are thermodynamically favorable in terms of entropic considerations and the close correspondence between normal modes and experimentally determined functional variations imply the entropic effects as a driving force for the observed conformational transitions during ligand binding. ANM also acts as predictive to form representative ensemble of conformers by determining global modes.

Elastic network models depend on a pre-defined cutoff value and interacting pairs are defined using this cutoff distance. Jernigan et. al. [10] investigates the presence of this cutoff value and proposes a new model which discards the requirement for a pre-defined cutoff value. As aforementioned the matrices containing inter-residue potentials are classically constructed using some empirical cutoff values which considers only the interactions within the range of the cutoff-value and ignores the out of-range interactions. However, definition of a cutoff value is not a straightforward process and depends on the architecture of the protein, showing differences from protein to protein. Also, the inter-residue potential is determined independent from the distances between interacting residue pairs. Jernigan and his collegues propose a new method called parameter-free model in order to assign inter-residue potentials to each residue pair. According to this model, all residue pairs are accepted as interacting and the inter-residue potential between this residues are determined using inter-distances between these pairs independent from a cutoff value. In such way, the authors take also the long range interactions into considerations within calculations of the EN models. This model is inspired from recent studies of Lin et. al [72] which defines a weighted contact number model for prediction of B-factors differing from previous contact number-based models by discarding the cutoff value, and instead weighting the interactions in a distance dependent manner. The same theory is adopted in EN models by Jernigan and his group and the interaction potential between i^{th} and j^{th} residue is determined as in Eq. 2.11:

$$W(i,j) = \frac{1}{d(i,j)^2} \; Eq.2.11$$

Here, W is the interaction potential between i^{th} and j^{th} residues and d(i,j) is the distance between two residues. So, the determination of Kirchoff and Hessian matrices take the form as in Eq. 2.12 and Eq. 2.13 respectively:

$$\Gamma_{ij}^{pf} = \begin{cases} d(i,j)^{-2} & if\ i \neq j \\ -\sum_{i,j \neq i} \Gamma_{ij} & if\ i = j \end{cases} \quad Eq.2.12$$

$$H_{ij}^{pf} = H_{ij} d(i,j)^{-2} \; Eq. 2.13$$

Here, H_{ij} defines the each super-element of Hessian matrix.

The authors in the same study compare the success of B-factor predictions for each EN model including cutoff employed and parameter free GNM and ANM. Results imply a significant improvement in the case of parameter free GNM and ANM. They also conclude that for larger conformational changes the power of parameter-free models can be increased.

In a recent work, Liu et. al [19] compared the accuracy of molecular dynamics and GNM over experimental data using a designed sugar-binding peptide. The mean-square residue deviations of the protein derived from NMR and B-factors of x-ray crystallography have been compared with GNM and MD predictions. The comparison results exhibit that GNM yielded a better correlation with experimental data compared to MD predictions. Individual modes determined by GNM or molecular dynamics exhibit similarities and differences in terms of contributions to fluctuation. These modes extracted by each technique are comparable and the low-frequency modes of MD simulations contribute much to the observed motion whereas in GNM the contribution of higher modes to the fluctuation is higher compared to MD. Thus, the modes calculated by each technique differ mainly in terms of high frequency modes. Moreover, the GNM performed better agreement with NMR findings compared to X-ray crystallography. The mean square deviations obtained from an ensemble of conformations rather than considering only the calculations of B-factors using just a single crystal structure is preferable for GNM that improves the agreement observed between NMR and GNM findings. Also, the B-factors calculated by X-ray crystallography can be influenced by crystal packing effects due to the interactions between the crystal molecule and its neighbors subsequently affecting equilibrium motions.

2.4. Available Online Tools

The normal mode analysis and EN models presented here are also available as online tools developed by different groups. MoViES server developed by Chen and colleagues [73], determine vibrational normal modes by using AMBER molecular mechanics force fields in terms of a self-consistent harmonic approximation method. NOMAD-Ref [74], another online tool that performs normal mode analysis in a full-atomic manner and this tool enables the computed normal modes to be graphically accessed and utilized by different programs for several purposes such as refinement of structures against experimental data, calculating overlaps between eigenvectors and eigenvalues of different conformations of the same molecule. ProMode database [75] performs full-atom NMA calculations using dihedral angels of the structure instead of Cartesian coordinates differing from previous tools. ProMode is also capable of representing properties of normal modes graphically in more detail. WEBnm [76] computes slowest non-zero normal modes in a residue level with the animations of corresponding modes presented in the web-browser together with calculation of normalized squared atomic displacements and vector field representation. AD-ENM [http://enm.lobos.nih.gov] developed by Zheng and his colleagues [77] is a simplified ENM tool and deal with low-frequency spectrum of normal modes calculating the contribution of each slowest normal mode to conformational change between different conformers of same protein, determining the deformation of the whole structure according to a given conformation for a subset of residues. AD-ENM also enables detection of hot-spots through calculating fluctuations of a given functional site by means of normal modes. ElNemo [78] determines slowest 100 normal modes at the residue level without any size limits due to its 'rotation-translation-block' (RTB) approximation method. Parameters and visualizations such as residue mean-square fluctuations, degree of collectivity and correation between predicted and experimental B-factors are presented. If the protein has two available structures, ElNemo identifies the normal modes that contribute most to the conformational transition. iGNM database

[79] performs GNM calculations for the protein structures deposited in PDB database and the results are available for users with the graphical analysis such as cross-correlations between residues and fluctuations according to each mode enabled. oGNM server [80] is developed by the same group for online GNM calculations for a broad range of cutoff value (7.3-15Å). ANM calculations are also provided as an online tool [81] with flexible options that enable user to control model parameters and the output formats. This ANM server presents collective modes as interactive animations with the computed parameters available for download.

2.5. Comparison of EN models

In the first part of our study we have compared accuracy of five different EN models over prediction of experimental B-factors. Two of these models are GNMs and the remaining three methods are ANMs. One of GNM methods compute theoretical B-factors using a predefined cutoff value of 7.3 Å. The other GNM method adopts a parameter free approach and as aforementioned weights the interactions based on the inter-distances between interacting residue pairs. In our calculations we have employed a power of 2 as done in the work of Jernigan and colleagues [10]. Two out of three ANMs constructed in our study depend on parameter free calculations and differ only in the defined powers that weight the strength of interactions. These methods are called as pfANM2 and pfANM4 with respective powers of two and four. One of ANMs depends on the classical calculations for Hessian matrix with a pre-defined cutoff value of 13.0 Å.

The proteins used for accuracy comparisons are retrieved from PDB database using PDB-REPRDB [82] which determines representative protein chains according to user-defined input parameters. We have retrieved 1171 proteins restricting the X-ray resolution of these structures below 2.0 Å and with a length shorter than 300 residues. All of these proteins contained experimentally determined B-factors with no missing residues. We have accomplished the Pearson correlation calculations using a MATLAB script that determines the theoretical B-factors through utilizing all normal modes of the

protein computed for each five different EN models including two GNM and three ANM calculations.

2.6. Modal Analysis

This part constitutes the main focus of our study. Here, we had a postulate and performed our further calculations in order to test this hypothesis. According to pre-existing equilibrium model the two conformers (substates) of a protein are at equilibrium with forward and reverse rate constants as k_1 and k_{-1}. If transition free energies of two substates are ΔG_1 and ΔG_2, then

$$k_1 = A_1 \exp\left(-\frac{\Delta G_1}{kT}\right) \; and \; k_{-1} = A_2 \exp\left(-\frac{\Delta G_2}{kT}\right) \quad (Eq. 2.14)$$

where A1 and A2 are pre-exponential factors and R and T are the gas constant and absolute temperature.

Let's say, the relative populations of these substates, depicted from experiments, are P_1 and P_2. The energy barrier between the two states (ΔG_2-ΔG_1) will be proportional to the relative probability (and thus the population) of each state. We hypothesize that these pre-existing substates can be studied by modal analysis and there exist some modes that will drive conformer 1 to conformer 2 (say i^{th} mode of conformer 1), and the associated weight of this mode is w_i, and conformer 2 to conformer 1 (say j^{th} mode of conformer 2), and the associated weight of this mode is w_j; then there exist a relation between the weights of the modes and the relative populations of these conformers:

$$\frac{P_1}{P_2} = \frac{k_{-1}}{k_1} = \frac{w_j}{w_i} \quad (Eq. 2.15)$$

In order to investigate our hypothesis in the second part of our study we have employed pfANM4 to extract normal modes of pre-existing conformers of each protein. We have retrieved eleven different proteins (mentioned in Chapter 3) that each exhibits a conformational change between two conformers which are generally termed as open and closed forms. The experimental displacement vector for each protein stores the

isotropic displacement of each residue between the open and closed forms calculated after superposition of these structures. As explained in several studies in this chapter individual modes resemble the observed experimental displacement between different conformers of the same protein. The main idea behind the superposition that we have employed is to maximize the performance of individual modes over the mobile regions of the protein.

The experimental data about the thermodynamics/kinetics of the protein elucidate presence of at least two conformers for each one of the eleven proteins that we have adopted in our modal analysis calculations. EN model calculations depend on a crystal structure which stores the atomic coordinates of C_α atoms. However, low stability and fast conversion to majorly populated conformers stand as a problem for the crystallization of minorly populated conformers. For some of our proteins we encounter these difficulties since they do not have any corresponding structure in PDB database. Thus, in order to overcome this drawback we have adopted PDB structures that are structurally similar to experimentally proposed conformers for Cholesterol oxidase, DHFR and Fab Antibody. Experimental data about the cholesterol oxidase [24] proposes an interconversion between the (E-FAD.S) and E-$FADH_2$ form. In the former conformer substrate is bound to enzyme together with FAD and in the latter one only $FADH_2$ is bound to enzyme. We have adopted the PDB structure (1coy) that includes the cholesterol oxidase and a steroid substrate bound together with FAD to enzyme. However, for the $FADH_2$ bound form, the exact structure was not deposited in the PDB database thus, we have adopted PDB structure (3cox) that includes enzyme complexed with FAD. For DHFR [37] experimental data describes an interconversion between two free forms of the enzyme which are capable to bind only NADPH or only H_2F. We have adopted only NADPH (1rx1) and only H_2F (1rx7) bound forms of the enzyme in order to resemble unbound forms of the enzyme. Fab Antibody has two forms, one non-binding form and the other one ligand binding form (3cfj) of the antibody. Structure of non-binding form of the antibody is not deposited in PDB database thus, we have

adopted the hapten bound closed form of the enzyme (1y0l) in order to represent non-binding closed form of the antibody.

Except DHFR, Fab Antibody and cholesterol oxidase, remaining proteins including Adenylate Kinase (4ake, 1ake), RNAse (7rsa, 1u1b), TIM (1ypi, 7tim), NtrC (1dc7, 1dc8) and Cytochrome P450 (EryK; 2wio, 2jjn) include both open and closed structures deposited in PDB database. Monoclonal antibody Spe7 includes two distinct free conformers that are determined by X-ray crystallography. Maltose Binding Protein (MBP; 2v93) includes open and partially closed forms of the protein that are determined using paramagnetic NMR. For Thrombin we have retrieved Na^+ free (1sgi) and Na^+ bound form (1sg8) of the enzyme. Our eleven dataset proteins for modal analysis have been presented in Table 1 and Table 2 in Chapter 4 together with the PDB ids for each conformer, the retrieved chain and information about the experimental data used for further investigations.

The PDB structures have been pre-processed before employing modal analysis because of the presence of missing residues. Two conformers of each protein need to have equal residue length. Thus, if one of two structures contains missing residues, these residues have to be discarded from the second conformer.

One of the PDB structures of EryK starts from 19^{th} residue (2WIO), the other (2JJN) starts from 17^{th} residue. Thus the residues 17 and 18 have been excluded from 2JJN. 4^{th} residue of 1COY, 148^{th} residue of 1SG8 (chain E) and residues 232, 233 have been removed from 1Y0L (chain H) for ANM calculations. For MBP both open and closed structure coordinates have been derived from the same structure (PDB id: 2V93) which contains atomic coordinates of both open and closed structures of MBP.

Experimental displacement between two conformers has been determined by calculating the displacement of each residue between two different conformers. We have superimposed Adenylate Kinase, 34E4 Fab Antibody and NtrC using the secondary structures that is common in both substates. The secondary structure information is extracted from PDB file itself which presents residues with helical or β-

sheet content. Two structures for EryK have been superimposed using conserved regions of the protein that spans helices D, L, I and E [26]. Spe7 antibody contains three loop regions that are responsible for antigen binding thus we have adopted remaining parts of the protein which comprises regions between residues Q1-S25, W36-R50, K67-R98 and W110-S120 for superimposition of two isomers of Spe7. RNase interacts with the substrate through B1 (loop between residues D14-N24) and B2 (loop residues between A64-N71) subsites so that the enzyme structures were superimposed using the residues except B1 and B2 subsites [83]. Active site of the TIM is comprised of loop 6 region and overall motion of the protein is restricted to the motion of this loop region which contains the residues between V167 and T177 [39]. Structural matching of two conformers of the TIM has been accomplished using the regions except loop 6 region. Superimposition of the MBP structures has been done by using the whole residues of the Maltose Binding Protein. Similar to the behavior of TIM, the major conformational changes of DHFR also includes motion of an active site loop region (residues A9 to L24) which is named as Met^{20}. This loop is closed when the enzyme is bound to substrate [25]. We have considered flexibility of this loop region and superimposed two DHFR structures using the remaining residues. For Thrombin we have performed superposition using all residues of the protein. For cholesterol oxidase we have adopted SCOP domain information and superimposed two structures using residues R4-W318 and G451-I506 which constitute cholesterol oxidase of GMC family.

The collectivity (K) of the experimental displacement for each protein has also been calculated. This value defines the collectivity of the motion according to amount of participating residues within the observed motion. The motion which spans a large number of residues as participants will have a higher collectivity than the local motions which span a major motion restricted to a small region. Collectivity of the experimental displacement is determined using Eq. 2.16 [20] where K is the collectivity, N is the length of the experimental displacement vector, α is the normalization factor that makes $\sum_1^N \alpha \Delta R_i^2$ equal to 1, ΔR_i^2 is the mean square displacement of the ith residue.

$$K = \frac{1}{N}\exp\left(-\sum_{i=1}^{N} \alpha\Delta R_i^2 log\, \alpha\Delta R_i^2\right) \quad (Eq.2.16)$$

We have analyzed the correlation between individual modes and experimental displacement in both isotropic and anisotropic manner. For this purpose we have adopted two parameters, one is the mode correlation (*C*) and the other is the Overlap (*I*). Mode correlation (*C*) determines the isotropic amplitude similarity between experimental displacement and individual normal mode vector that stores residue fluctuations. The experimental displacement and normal modes initially calculate the mean square fluctuations of each residue thus we have calculated the square roots of both experimental displacement and normal mode vectors. The Pearson correlation (r) is calculated as in Eq. 2.17 where M_i is the ith individual mode, ΔR_i is the isotropic experimental change vector and N is the number of nodes.

$$r = \frac{\sum M_i \Delta R_i - {\sum M_i \sum \Delta R_i}/{N}}{\sqrt{\left(\sum M_i^2 - {(\sum M_i)^2}/{N}\right)}\sqrt{\left(\sum \Delta R_i^2 - {(\sum \Delta R_i)^2}/{N}\right)}} \quad (Eq.2.17)$$

Overlap (*I*) on the other hand calculates the direction similarity between the experimental displacement and normal modes. The eigenvalues calculated by ANM include fluctuation information in three dimensions. The anisotropic experimental vector is determined by calculating the displacements of each residue in all three dimensions between the two structures of the same protein. Once these two vectors have been determined the overlap is calculated as in Eq. 2.18:

$$I_j = \frac{\left|\sum_1^{3N} \mathrm{u}_{ij} \,.\, \Delta \mathrm{r_i}\right|}{\left[\sum_1^{3N} \mathrm{u}_{ij}{}^2 \sum_1^{3N} \Delta r_i{}^2\right]^{1/2}} \quad (Eq.2.18)$$

Here, u_j is the j^{th} eigenvector, composed of 3N elements in which each residue is represented as fluctuations in x,y and z directions respectively. Δr is the experimental displacement of the protein containing displacements in the x,y and z directions in the

same manner with u_j and has the same dimensions as u_j. I_j is the overlap between the directions of the j^{th} eigenvector and experimental displacement.

We have employed pfANM4 calculations to both conformers of all eleven proteins and determined the amplitude and direction similarity between experimental displacement and individual modes through mode correlation and overlap calculations.

The individual mode predictions have been validated through random mode analysis. As we have stated above we had two displacement vectors the first one is N dimensional (isotropic displacement of each entry) and the second one is 3N dimensional (containing displacement of each residue in x, y, z coordinates). We have adopted these vectors for mode correlation and overlap calculations respectively. In random mode analysis part, we have investigated success of randomly created modes in terms of mode correlation and overlap scores for each protein. For this purpose, we have produced two sets of random vectors which contain N-dimensional and 3N-dimensional vectors. Each entry of N-dimensional random vector has been assigned with a random value between a range of minimum and maximum amplitudes of experimental displacement. In the case of 3N-dimensional vector, three N-dimensional vectors have been created separately and combined in further steps. Here the aim is to define the amplitude ranges of each entry according to each three directional components (displacements in x, y, z) which differ according to their maximum and minimum values. For each protein we have produced 5000 random mode vectors using a uniform MATLAB function ‘rand’.

2.7. Mode Selection and Investigating Relation between Thermodynamics/Kinetics and Mode Contributions

Our study is followed by the analysis of eigenvalues of individual modes which exhibit considerable similarity with the observed experimental displacement. For this purpose we have selected top three modes based on the mode correlation and overlap

scores together with the visual inspection of individual modes and termed these modes as 'significant modes'.

Eigenvalues of these modes have been processed with the formula given below that gives the normalized contribution of each mode with respect to the whole spectrum of eigenvalues to obtain mode weights (also called as weighted eigenvalues):

$$w_i = \frac{1/\lambda_i}{\sum_1^{3n-6} 1/\lambda_i} \quad (Eq. 2.19)$$

Here W_i is the mode weight of the i^{th} mode and λ_i is the raw eigenvalue of the i^{th} mode.

We have constructed two vectors in order to investigate the relation between the mode weight and kinetic rates/relative populations called computational (**C**) and experimental (**E**) vectors. These two vectors, **C** and **E**, contain ratios of mode weights and ratios of kinetic rates/relative populations of the proteins respectively. Each entry of the **C** represents computational findings of each protein and is determined with following formula:

$$C_J = \frac{w_{j,m2}}{w_{i,m1}} \quad (Eq. 2.20)$$

Here $w_{i,m1}$ and $w_{j,2m}$ are the mode weights of the first and second conformers of the m^{th} protein respectively which were determined by Eq. 2.19.

Based on the type of the experimental data, whether it contains relative populations of conformers or the kinetic conversion rates between these states we construct our **E** using following formulas:

$$E_j = \frac{p_{1j}}{p_{2j}} \quad (Eq. 2.21)$$

By above formula we construct our experimental vector for proteins with population data. Here p_{1j} and p_{2j} are the relative populations of the first and second conformer respectively.

For the proteins with kinetic rates we determine each entry of **E** by following formula:

$$E_j = \frac{1/r_{1j}}{1/r_{2j}} \ (Eq.\ 2.22)$$

Following construction of **C** and **E** for dataset proteins with kinetic rates or populations, we have calculated the similarity between **C** and **E** using Pearson correlation.

We have provided a depiction of our study in Figure 2-1. We have constructed three different experimental vectors, one containing ratios calculated using kinetic rates, one containing ratios of relative populations and the last one containing combined data of ratios of kinetic rates and relative populations. First two of these vectors contained six proteins and the last one contained whole dataset of eleven proteins where for Spe7 (both kinetic rates and relative populations present) we have adopted ratios determined via kinetic rates. In the same manner computational vector has been constructed three times each of them including different proteins with the last one containing whole dataset proteins.

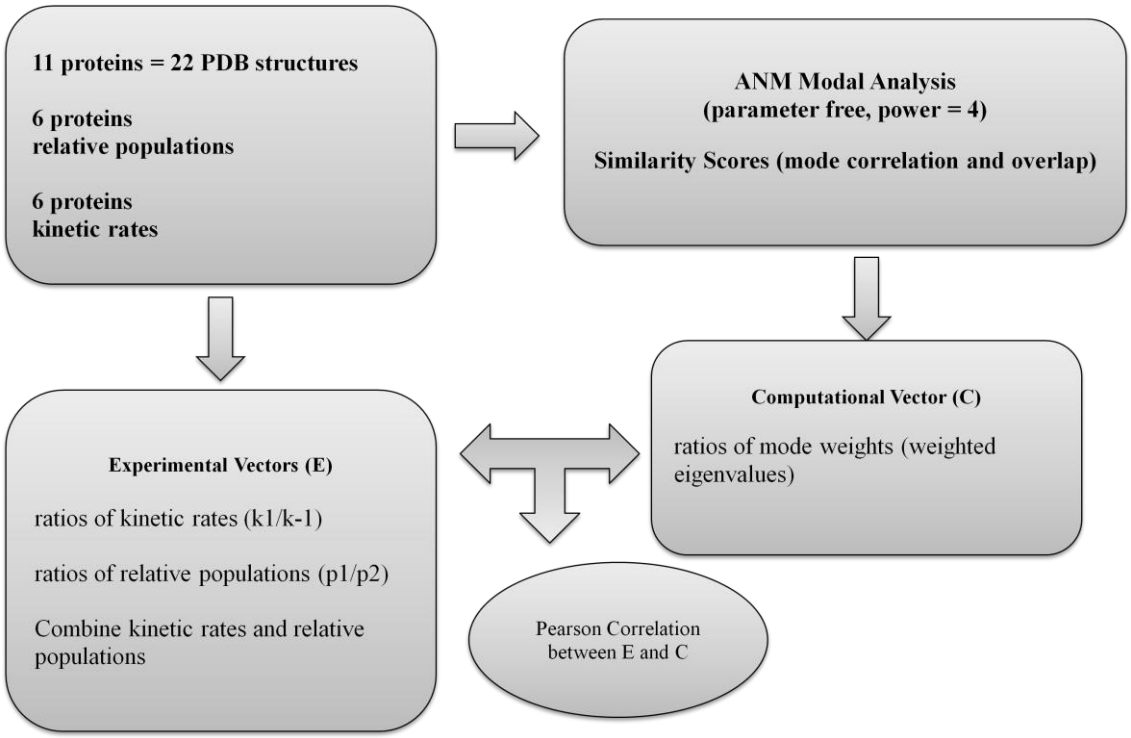


Figure 2-1: Depiction of our thesis study

Chapter 3

3. DATASET PROTEINS

In this chapter, the proteins we have used for modal analysis are presented together with their structural and kinetic properties. Consistent with the aim of our study, these proteins exhibit behaviors that comprise pre-existing equilibrium model and perform conformational changes in different scales from loop motions to collective domain motions. This chapter is dedicated to present brief knowledge about structural properties and dynamics of the adopted proteins.

3.1. Adenylate Kinase

Adenylate Kinase (AK) performs reversible catalysis of ATP and AMP into ADP via transferring phosphoryl group in solvent excluded environment. The protein interconverts between open and closed conformation directed by the motions of two ligand binding domain, the ATP lid and AMP binding domain. The substrate binding leads to closure of both the ATP lid and the AMP binding regions over the core region. The solvent is excluded from the active site by the closure providing a hydrophobic environment for phosphoryl transfer between ATP and AMP. Once the catalysis is accomplished the enzyme releases its products and converts into the open form [23, 40]. The crystal structures of AK from several organisms have been determined both in free and in complex forms [84].

Both experimental and computational efforts contribute much to the knowledge about the structure, dynamics and catalysis together with behavior of the AK enzyme. NMR study of Kern and colleagues [15] about the opening and closing dynamics of AK enzyme in two organisms, one mesophilic and the other thermophilic, elucidate the catalysis of the enzyme. Authors initially compared the catalytic activities of the enzyme in these two organisms as a function of temperature and found that the optimum temperatures that enzyme exhibits maximal activities differ considerably and in the 20

ºC activity of the thermophilic enzyme is reduced significantly compared to activity of mesophilic enzyme. In further steps authors concentrated over the rate-limiting steps for catalysis of enzyme and determined the opening and closing rates of the enzyme for both organisms. The findings of this study indicate that reduced activity of the thermophilic enzyme is related with lid-opening rate of the thermophilic enzyme.

Henzler-Wildman et. al. [85] have adopted the kinetic rates of opening and closing and found that these slow motions of the enzyme are facilitated by faster, in pico-to nano second timescale, local atomic fluctuations of the enzyme that are located in the hinge regions of the enzyme. Computational study of Lu and Wang [86] have determined the pathway of transition between open and closed conformations of the enzyme together with transition state ensembles and this computational study also calculated similar kinetic rates with those of Kern et. al [15]. In another computational attempt [87], the fluctuations of the open enzyme have been studied using principal component analysis and varimax rotation analysis that employs determined principal components to a certain subset of residues. The results of this study inplies that the open form encodes the holo-like conformations through fluctuations of a certain subset of residues.

3.2. Monoclonal Antibody: Spe7

The behavior of Spe7 is a good example about the antibody diversity. This antibody is shown to follow a pre-existing equilibrium model co-existing in solution as two distinct conformations, dubbed as $Ab^1$and Ab^2 and the structures of these conformations in their free and complexed forms with four different antigens are determined using X-ray crystallography [43]. In three of four hapten complexes (haptens are DNP-Ser, Az or Fur) Spe7 is determined in another conformation which is termed as Ab^3. When complexed with the fourth protein antigen (Trx-Shear) Spe7 exhibits another conformation termed as Ab^4. The architecture of binding sites of Spe7 is different between these four distinct conformations. These four distinct conformations

distinguish from each other mainly by the orientations of H3 and L3 loops. Having different backbone orientations, these four conformers also exhibit distinct side-chain rotamers. Apparent conformational diversity of Spe7 enables recognition and binding of a diversity of ligands including haptens and antigens.

In a further study of the same group [16] authors emphasize over the isomerization process of the antibody. This study has determined the kinetic rates for the distinct steps of antigen binding including initial binding and induced fit together with the interconversions between the free conformers Ab^1 and Ab^2. Existence of two conformers as free forms enables the "kinetic discrimination" more than a thousand fold from a diversity of ligands which are highly related and also share similar affinities. In further stages of recognition only specific ligands provide a "locking in" which helps the antibody to convert into another distinct form Ab^3. The structural similarities of L3 loop and architecture of similar binding pocket between Ab^2 and Ab^3 imply that the binding occurs initially by interconversions between pre-existing conformers Ab^1 and Ab^2 which are populated as 80% and 20% respectively and in the presence of the ligand antibody-ligand interactions lead to another distinct conformer of Spe7 [16, 43].

3.3. Ribonuclease A (RNase)

RNase performs the breakdown of RNA in two sequential steps where breakage of diester bonds occur initially and this step is followed by hydrolyzing the intermediate to pyrimidine 3-phosphate. The enzyme is relatively small, composed of 124 amino acids and the stability of the enzyme is provided by four disulfide bonds [56, 88]. Binding site of the enzyme comprises highly specific interactions between enzyme and ligand and is located within a deep cavity. The active site of the enzyme indicates a high flexibility especially for the residues involving H119, K7, K41 and K66. The deep cleft of the enzyme that accommodates active site residues is surrounded by two halves of the hinge that directs opening and closing of the enzyme [88]. Two subsites of the enzyme termed as B1 and B2 are responsible for binding of nitrogen containing groups of the ligand

and three other subsites called P0, P1 and P2 are responsible for the binding of phosphate groups. Important residues for nitrogen containing group binding of B1 and B2 are D83, T45 and N71, E111 respectively. K66 for P0 subsite, Q11, K41 and D121 together with highly conserved residues H12 and H119 for P1 subsite and K7, R10 for P2 subsite are other catalytically important residues [83].

NMR studies for the motions of RNase [56] have revealed three substates that pre-exist together in conformational ensemble including free enzyme (E), enzyme-substrate complex (ES) and enzyme-product (EP) complex. The backbone alterations observed during conformational transition between ligand-free and enzyme-substrate complex involve two loop regions as loop 4 and loop 1. Loop 4 is responsible for purine specificity and loop 1 is located 20 Å from the active site. In all E, ES and EP complexes exchange parameters of NMR experiments do not exhibit considerable changes implying a ligand-independent nature for enzyme motions. Similar activation barriers for transitions between open and closed forms are observed for both E and ES complex. The amount of major conformation of the enzyme which depends on both the presence of the ligand and progress of catalysis differ between 93%-%95 with the other forms of the enzyme populated in minor amounts. According to these studies the functional enzymes motions are readily available even in the absence of the ligand and the ligand acts as a stabilizer of ES complex.

3.4. Dihydrofolate Reductase (DHFR)

Dihydrofolate reductase catalyzes the reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF) by utilizing NADPH for the hydride transfer. DHFR is an important enzyme for purine and thymidylate metabolism that are required for cell growth and proliferation and the enzyme is the only source for the production of THF [89]. Several crystal structures for the DHFR are deposit in RCSB and previously Wright and his colleagues crystallized the intermediates (E:NADPH, E:NADPH:DHF,

E:$NADP^+$:THF and E:NADPH:THF) during the catalysis of the enzyme together with the ligands [25].

The DHFR enzyme of *E. coli* is composed of a central eight stranded β-sheet and four flanking α-helices and consists of two structural subdomains that are divided by the active site of the enzyme. The adenosine binding domain accommodates the adenosine region of the cofactor and the second domain called major domain is relatively larger compared to the adenosine binding domain dominated by three loops, termed as Met^{20} (residues 9-24), F-G (residues 116-132) and G-H (residues 142-150) that are facing the ligand binding region of the domain. In the presence of the ligand the hinge bending motions occur in about K38 and V88 moving the adenosine binding domain over major domain [89] leading to closure of active site cleft.

Crystal structures of DHFR complexes imply that conformation of the Met^{20} loop is determined by the presence of ligands and cofactors in corresponding binding sites. In the presence of only the substrate, the loop adopts an occluded conformation and in the presence of adenosine moiety of the cofactor the Met^{20} loop packs against the nicotinamide ring of the cofactor providing closure of active site cleft and protecting it from solvent. The closure of Met^{20} loop is required for the catalysis in which the cofactor and the substrate are positioned in close proximity within the active site cleft [. The closed conformation of the Met^{20} loop is observed in the complexes E:NADPH and E:NADPH:DHF, on the other hand the occluded form is adopted in the E:$NADP^+$:THF, E:THF and E:NADPH:THF [25, 89]. The characterization of intermediate states of the enzyme during catalysis together with the kinetic rates of the hydride transfer and conformational transitions [25, 90] indicate the importance of the energy landscape of the protein in order to drive the enzyme through its catalytically competent conformations along a energetically favorable kinetic path.

Another study of the same enzyme [37] accomplished using single-molecule measurements elucidated an interesting interaction mechanism between enzyme-substrate and enzyme-cofactor. According to this study, two distinct free-forms of the

enzyme is present in solution each binding NADPH or DHF with these structures interconverting between each other.

3.5. Maltose Binding Protein (MBP)

Maltose Binding Protein (MBP) belongs to the bacterial periplasmic binding protein family [91]. This family of proteins is responsible for transmembrane transport or chemotaxis where the ligand binding induces a large conformational transition that subsequently facilitates the signal transduction through interactions with the proteins responsible for chemotaxis and transport proteins.

MBP is essential for maltose uptake and consists of two domains termed as amino-terminal and carboxy-terminal domains, NTDs and CTDs respectively. Two domains of the protein are linked via a short helix and two-stranded β-sheet [41]. X-ray crystallography studies of the protein elucidate the structure of two domains that shares similar secondary structure topologies containing a central β pleated sheet with two or three parallel α helices flanking on both sides [92]. The protein cleft accommodates the maltose binding site and in the binding site several polar and aromatic groups are located providing formation of extensive interactions between substrate and the cleft. The maltose is buried in the cleft in the closed form inaccessible by the solvent and in the absence of the substrate MBP adopts an open conformation [92].The transition between the open form and the closed form of the protein involves a hinge-bending motion about the linker region resulting with a rigid body domain reorientation [93, 94].

Conformational dynamics of the MBP has been studied using paramagnetic NMR and results show that in the apo state of the protein a minor conformation (populated ~%5) is also populated together with the major conformation (populated ~%95) that is open form of the protein [41]. Paramagnetic NMR performs well in order to reveal minor populated species in the solution even the amount is as low as 0.5-1%. There is a rapid transition between major and minor species of the protein according to the paramagnetic relaxation enhancement data. The relative of orientations of two domains, NTD and CTD differ from the open or closed states of the protein but indicates

presence of another structure which is partially closed state of the protein [41]. Findings imply the presence of a long-postulated dynamic equilibrium between open and partially closed apo states of the protein. The transition between these species is fast within the timescales of nano-to-microseconds. The presence of the substrate redistributes the populations of open and partially closed states of the protein favoring the latter one. Pre-existence of apo and ligand binding conformers even in the absence of the ligand facilitates the transition to the holo structure.

3.6. Cytochrome P450 EryK

Cyctochrome P450 proteins comprise one of the largest families of enzyme proteins and are diversely found in all known organisms. Although amino acid sequences of P450 proteins differ largely with a similarity less than 20% in some cases, the structural fold has been conserved throughout evolution. These heme-binding proteins perform diverse functions primarily responsible for oxygen chemistry such as biosynthesis of hormones and structural components of living organisms [95]. Although EryK exhibits a large structural conservation, it also recognizes a wide range of substrates [95]. Due to its importance in the oxygen chemistry and drug metabolism, several drug discovery studies targeting P450 proteins have been accomplished [96, 97].

A recent study of Savino et. al. [26] emphasizes over the structural and biophysical attributes of EryK through analysis of three crystallized conformers of the protein where two of them are in free form and one is the ligand complexed conformer. The protein core is highly conserved and is formed by a four-helix bundle, three parallel helices, termed as D, L and I and an antiparallel one termed as E. The heme group is positioned between the helices I and L and is bound to the C353, highly conserved fifth ligand of the heme iron. The high substrate specificity is provided by six ‘substrate recognition sites’ (SRS) that occupies the active site of the protein. Depending on the salt concentration there are two ligand-free structures of this protein crystallized, as open and closed and the differences between these two conformers involve the these SRS.

The rearrangements taking place in these sites especially in the helices F, G and the B-C loop at high salt concentrations prevents access to the active site of the protein, thus this form is called as the closed form of the enzyme. The conformational transitions between the open and closed forms of the ligand-free enzyme continuously provide a dynamic equilibrium for the enzyme. The enzyme was also crystallized as a complex (ErD-EryK). The structure of this complex again reveals another closed form of the enzyme where formation of specific interactions between active site and the substrate contribute mainly to the substrate specificity. The substrate is positioned on the active site followed by closure of BC loops and the helices F and G. Flexibility of the active site provide specificity in the substrate recognition [26].

3.7. Triosephosphate Isomerase (TIM)

TIM performs the catalysis of dihydroxy-acetone phosphate (DHAP) to glyceraldehide 3-phosphate in a reversible manner within the fifth step of glycolysis. The enzyme is composed of two identical subunits and each subunit has an $(\alpha/\beta)_8$ fold also known as TIM-barrel fold [39, 56, 98].

The catalysis of TIM is performed by opening and closing motion of a well-characterized Ω-loop also called as loop 6 located in the active site of TIM. This loop comprises the residues between V167 and T177. In the open form of the loop the active site of the enzyme is accessible by both solvent and substrate and in the closed form it prevents loss of reaction intermediate [56, 98] and the stability of loop 6 is stabilized by the hydrogen bonds formed between this loop and the neighbor loop 7 [56]. The overall motion of TIM is restricted to this loop 6 region together with the conformational transitions observed in two other sites, the residue K84 that is located on helix C and the residues of the helix G. The opening closing motions of loop 6 seems to be rate-limiting step during the catalysis [39].

Motion of the loop 6 is dictated by the hinge residues located in the both sides of the loop and the loop moves as a rigid body with the loop residues exhibiting a synchronous

motion between open and closed conformations [56]. These two functional states of the loop 6 is found to co-exist in the solution independent from the presence of the ligand supported by both NMR studies and computational studies [39].

3.8. Thrombin

Thrombin is an important protein in the coagulation cascade. The protein has several functions within the coagulation cascade such as acting as a procoagulant through performing cleavage of fibrinogen to fibrin and activation of several coagulation factors and stimulation of platelet aggregation [99]. After a maturation period of thrombin from membrane bound prothrombin, it becomes soluble and is released into the blood vessel lumen. The mature thrombin (α-Thrombin) consists of two chains linked with a disulfide bridge termed as A and B-chains which differ in amino acid length and composition. The A-chain is comprised of 36 residues and B-chain is comprised of 259 residues and B-chain is organized in two adjacent β-barrels. The junction of these barrels accommodates active site residues S195, H57 and D102 and active site cleft lies across both barrels as a deep and narrow groove. The charge distribution of thrombin surface is an important feature which provides highly positively and highly negatively charged fields and these fields bind different substrates [99].

Thrombin exhibits an allosteric behavior and requires Na^+ binding as a modulator. There are two well characterized allosteric states of the thrombin termed as fast and slow forms depending on the absence or presence of the ion [38, 99]. Studies present that Na^+ binding triggers a conformational transition in thrombin involving the specificity sites of the enzyme and 148-loop from a more closed and flexible conformation (slow form of the enzyme) to a more open and rigid form which is termed as the fast form [100]. These findings are consistent with the previous findings [101] that proposes enhanced thrombin activity upon Na^+ binding. Study of Bah et. al [38] emphasizes over the kinetics of Na^+ binding to thrombin. According to this study ion binds to thrombin in two step corresponding to slow and fast steps in which former one involves the slow transition between the free enzyme and another form of the enzyme

that is unable to bind Na^+. Fast step indicates the formation of enzyme-ion complex. These three conformers, non-binding, free and complex forms of the thrombin pre-exist together in solution populated in different amounts that are supplied in the same study.

3.9. 34E4 Fab Antibody

Antibodies exhibit a large diversity for antigens and our protein dataset contain two different antibodies. First of them is Spe7 as mentioned previously in this chapter and second one is 34E4 Fab Antibody. Antibody 34E4 catalyzes the conversion of benzisoxazoles to salicylonitriles [102]. Conformational transitions are also present in the antibody 34E4 similar to behavior of Spe7. There are two free forms of the enzyme in equilibrium and interconverting between each other and only one of these two forms is eligible to bind substrate. Thus, these transitions between two forms affect the catalytic efficiency of the antibody and are prominent. Comparison between the crystal structures of unligated and hapten bound form of the antibody elucidates the structural changes taking place in the catalytic antibody upon binding [42, 102]. According to these studies the hapten binding does not induce any major movements within the CDR (complementary determining regions loops L1-L3 and H1-H3of antibody) loops of the antibody or does not alter the relative disposition of V_L and V_H domains. The most important distinction between the unligated and hapten bound forms of the antibody is the orientation of Trp^{L91} indole ring. The transition between the different orientations of the indole ring alters the architecture of binding site through extending it from a shallow indentation to a deep cavity. In the absence of the ligand, in other words in the closed state, the indole ring occupies the binding pocket and thereby prevents access. When the indole ring moves out of the binding site, the deep cavity becomes eligible to accommodate substrate. Two conformations of the ligand, both the substrate binding and non-binding forms of the enzyme co-exist together in the solution favoring the non-binding form in a 3:1 ratio. Debler et. al. [42] concludes that the presence of these two

conformers of the antibody limit the catalysis of the antibody due to presence of a slow transition step between non-binding and ligand-binding form.

3.10. Cholesterol Oxidase

Cholesterol Oxidase is a FAD-dependent enzyme that undergoes oxidation and reduction reactions catalyzing isomerization of steroids [103]. Catalysis involves two sequential steps: firstly oxidation of cholesterol to cholesterone and reduction of FAD to $FADH_2$ followed by reduction of oxygen to H_2O_2 and oxidation of $FADH_2$ to FAD [24, 104]. According to its crystal form the enzyme comprises two domains: a FAD-binding domain and a steroid binding domain. The active site of the protein is located between these two domains where the six-stranded antiparallel beta-sheet of the steroid binding domain forms the roof of the active-site cavity. The floor of the enzyme is occupied by the flavin ring system of the FAD that is tightly bound to the center of the protein and is surrounded by hydrophobic binding pocket for cholesterol [103].

Functional motions of the enzyme have been studied using single-molecule experiments involving cholesterol oxidase as a case study [24, 104]. These studies emphasize over the behavior of the enzyme that is proposed to follow a Michaelis-Menten (MM) model. This model assumes the enzyme can be found in two distinct states and the enzyme activity is related with independent stochastic transitions between these states. This model neglects the effect of protein chain and assumes it as only a solid support for catalytic center. However, studies of Lerch et. al. [24] revises this simple model based on the fluorescence correlation spectroscopy (FCS) data and concludes that the conformational motions of the enzyme have prominent roles during the turnover cycle and these slow fluctuations of the enzyme modulates the transition rates exhibiting differences between several subsequent turnover cycles. Several methods for the behavior of the enzyme have been presented in the same study [24] in order to interpret FCS data and the kinetic rates according to these models have been determined using experimental data.

3.11. NtrC

Nitrogen Regulatory Protein C (NtrC) is a signal transduction protein belonging to 'two-component system' signaling family. These proteins mutually switch between two states termed as inactive and active states and the activation of protein can be promoted by several factors including other proteins, ligands and covalent modifications. In the case of NtrC, the allosteric activation is provided by phosphorylation of D54 [44].

NMR experiments elucidate a pre-existing equilibrium of both inactive and active states in the absence of phosphorylation. Inactive conformer dominates the solution but inactive state is also populated in minor amounts (differing between 2-10 %) [44]. In the case of phosphorylation the stability of active conformer is increased and the population distribution is shifted towards the active state together with inactive forms in minor populations.

The protein consists of five α-helices and five-stranded parallel β-sheet [44]. The MD simulations of the protein elucidate large movements in helix 4 and loop β3-α3 [105] during activation. The loop β3-α3 accommodates phosphorylation site and upon binding the region termed as 3445 face including secondary structures α3, β4, α4, β5 moves away from the phosphorylation site with a twist motion. T82 of α4 is another important residue and it also exhibits large rearrangements upon phosphorylation and eventually a hydrogen bond between the active site and T82 forms. This hydrogen bond stabilizes the active conformation and that explains the low stability of active conformation in the unphosphorylated state [105].

Chapter 4

4. RESULTS and DISCUSSION

In this part we present our results which include B-factor comparisons and modal analysis of eleven dataset proteins. Second part of this chapter comprises the main focus of this study in which we have supplemented mode analysis results for each conformer of eleven proteins.

4.1. B-Factor Comparisons

First part of our studies includes comparison of five different EN models based on the Pearson correlations between the experimental and theoretical data. B-Factors, also known as temperature factors or Debye-Waller factors define mean square atomic displacements. The theoretical B-factors determined by EN models have been found to produce significant correlations with experimental counterparts. Here we have analyzed performance of five different EN models in order to predict B-factors including two GNMs with (GNM) and without cutoff values (pfGNM) and three ANMs including one cutoff-dependent (ANM) and two cutoff independent ANMs (pfANM2 and pfANM4) (*see B-Factor Comparisons Chapter 2*).

The comparison results of a total of 1171 proteins imply that cutoff-independent methods called as parameter free methods perform better than cutoff dependent EN models in order to predict experimental B-factors. Among the GNM calculations 32.5% of all proteins produced better correlations with cutoff dependent GNM but remaining 67.5% gave the best correlations with pfGNM. Success of cutoff-independent methods was even increased if we consider only Anisotropic Network Models. Parameter free methods of ANM that adopts a power of two and four respectively for pfANM2 and pfANM4 predicted significantly better results than cutoff dependent ANM methods for 87.4% of all proteins. The ANM predictions of these proteins that are included in 87.4% are improved by an average of 45% in the case of parameter free ANMs. Parameter free GNM improved cutoff-dependent GNM predictions by an average of 17.5%.

If we consider comparison of all these five methods with making no distinction between GNM and ANM findings, we observe that pfGNM predictions outperformed other four methods for about 36% percent of proteins. Second method was the pfANM2 and third method was GNM followed by pfANM4 and ANM as fourth and fifth methods respectively. Consistent with previous findings [106] we conclude that theoretical B-factors predicted by GNM perform better results compared to predictions done by ANM.

Selection of an optimum ANM technique for our further ANM calculations was one of the main concerns in these B-factor prediction comparisons. Thus, we have also included our dataset proteins for B-factor prediction comparisons and according to these calculations supplemented in Appendix 1. pfANM4 produced slightly higher correlations amongst all three ANM methods for our dataset proteins. Thus, we have adopted a parameter free method with a power of four for our further ANM calculations.

4.2. Modal Analysis

Normal mode analysis has been used to study the collective motions of proteins extensively as presented in Chapter 2. Some of these normal modes of several proteins were found to be strongly correlated with the large amplitude conformational changes of these proteins observed upon ligand or protein binding.

Advances in biophysical techniques allow investigation of protein dynamics in detail together with the timescales of the protein motion. These studies elucidate time-dependent behavior of proteins enabling scientists to study the so-called energy landscape of the protein together with kinetic and thermodynamic properties of the landscape. Native state structures of proteins called ensemble conformers have been determined using NMR, or X-ray crystallography, together with the relative populations of these conformers and amplitudes of interconversions taking place between protein conformers. Such studies involving different proteins accumulate in literature supplying

Protein Name	# of resid.	pdb I	pdb II	RMSD	K
Adenylate Kinase	214	4ake(A)	1ake (A)	5.55	0.40
Spe7	121	1oaq (H)	1ocw (H)	1.23	0.32
DHFR	159	1rx1 (A)	1rx7 (A)	0.50	0.08
34E4 Fab Antibody	224	3cfj (H)	1y0l (H)	1.79	0.17
Cholesterol Oxidase	500	1coy (A)	3cox (A)	0.24	0.24
EryK Cytochrome	411	2jjn (A)	2wio (A)	2.63	0.42
RNase	124	7rsa (A)	1u1b (A)	0.65	0.34
TIM	247	1ypi (A)	7tim (A)	0.46	0.07
Maltose Binding P.	366	2v93o	2v93c	4.97	0.62
NtrC	121	1dc7(A)	1dc8 (A)	1.78	0.46
Thrombin	250	1sgi (E)	1sg8 (E)	0.41	0.25

Table 1 The proteins used in this study.

findings about the thermodynamics (relative populations of conformers) and kinetics (timescales of interconversions between different conformers) of the protein in detail. As aforementioned normal mode analysis produce significant results correlating with the behavior of the protein resembling conformational changes between conformers of the protein.

In this study we have retrieved thermodynamic/kinetic properties of eleven proteins together with their PDB structures. As mentioned in Chapter 3 these proteins mutually populate distinct conformers in solution consistent with the pre-existing equilibrium model. Table 4-1 lists the dataset proteins, together with the PDB structure of each conformer. Because we have employed modal analysis to all protein conformers, we have determined conformational change taking place between given PDB structures of the protein. Root mean square displacement (rmsd) of this change and the collectivity of the protein motion is also listed in Table 1 for each protein. Superposition details for each protein are supplemented in methods part in Chapter 2. Listed residue lengths for

each protein are obtained after discarding the missing residues from both PDB structures which subsequently assures two PDB structures to have the same number of residues.

We have constructed our Hessian matrix (*mentioned in Chapter 2*) using a parameter-free method to determine the interaction potentials between different node pairs. In parameter free method the interaction potential between each pair is considered independent from a cutoff value and is determined based on the inter-distances of each residue pair. In this case the potentials between distant pairs have been taken into account which was neglected in the classic cutoff derived methods due to presence of a cutoff value.

For each conformer we have extracted lowest non-zero, thirty modes and calculate similarity between the experimental displacement and individual mode. For this purpose we have employed two similarity scores termed as mode correlation and overlap. Mode correlation simply determines the Pearson correlation between the individual mode and experimental displacement in an isotropic manner. Overlap, on the other hand defines the directional similarity between mode and conformational transition (*see Methods*).

In further parts we have presented our modal analysis results for each conformer together with supplementing mode correlations, overlaps and mode graphs of top three modes as supplementary information in Appendix 2.

4.2.1. Adenylate Kinase (AK)

AK performs opening and closing motions during its catalytic cycle through two mobile regions, ATPlid and AMP binding site (AMPbd), to provide a suitable environment for the catalysis[22, 40, 84]. ATPlid spans residues between G122 and D159 and AMPbd comprises a region between T31 and V59. These two regions perform hinge motions during opening and closing and remaining parts of the protein do not exhibit high mobility compared to ATPlid and AMPbd regions [84]. Collectivity

of the AK motion between open and closed forms is calculated as 0.40 which define a relatively high collectivity.

Modal analysis of the open form indicates importance of low-frequency modes especially first ten modes. These ten modes resemble the motion of ATPlid and AMPbd well according to mode correlations they produce, however if overlaps are considered two modes outperform others significantly which are second and fifth modes. Eighth modes follow these modes according to overlap calculations. Although first mode produces the best mode correlation (0.84), overlap (0.13) of the same mode is lower than previously mentioned three modes. As the mode indices increase mode correlations and overlaps diminish due to the collective nature of the motion. This sharp decrease is also reflected within the eigenvalue spectrum of the protein where the contribution of each mode declines after the first ten modes.

Individual modes of closed form produced relatively lower similarities considering the individual modes of open form. The best mode correlation and overlap have been calculated by first mode as 0.48 and 0.54 respectively. This mode is followed by the second and fourth modes according to similarity calculations. The motion of the open form can be defined using the first ten modes; however, higher frequency modes of closed form are also important to define its motion completely. According to these results we can state that both open and closed forms have intrinsic tendencies to resemble each other, but the energy state of the closed form is lower than the open form thus, it encounters difficulties in order to resemble a higher energy state (open form). This is the main reason behind the low similarity calculations of closed form.

4.2.2. Spe7

This antibody exists as two free conformers, Ab^1 and Ab^2 (pre-existing equilibrium), within the solution and the population distribution of each conformer depends on the type of ligand. In the presence of proper ligand this antibody can populate up to four

distinct conformers (Ab^3 and Ab^4, induced fit) [16, 43]. Here we present modal analysis of only free conformers of the antibody (Ab^1 and Ab^2).

RMSD between two structures is 1.23 and the collectivity of the motion is calculated as 0.32. The orientation of loop H3 and residues W33 and R50 comprise main differences and the experimental displacement between these two conformers exceed 8 Å within these regions. Except these regions the experimental displacement for most of the residues has lower amplitudes below 2 Å. Amplitude of H3 loop displacement is best defined by seventh mode with a mode correlation of 0.74 and an overlap of 0.27. However, best overlap (0.35) is calculated with the second mode that has a mode correlation of 0.58 that is the second best mode correlation. Sixth mode of Ab^1 follows second and seventh modes with a mode correlation of 0.51 and an overlap of 0.19.

Modal analysis of second conformer defines similar calculations for Ab^2 also in which fifth mode defines the motion of H3 loop with a best mode correlation of 0.70 and an overlap of 0.28. However, the best overlap is obtained by second mode as 0.34 (mode correlation 0.49). Third mode follows similarity calculations of second and fifth mode with a mode correlation of 0.3 and an overlap of 0.25. These two conformers are energetically more close to each other than open and closed forms thus individual modes perform similar correlation and overlap calculations. The eigenvalue distributions of these two conformers are highly similar to each other indicating that the mode contributions of two conformers follow a similar pattern.

4.2.3. RNase

RNase performs hydrolysis of RNA molecule into NTP in two sequential steps [83, 88]. The active site of the enzyme exists in open or closed forms and the experimental displacement between these two forms is increased within the loop 1 (residues 14-24) and loop 4 (residues 64-71) regions [83]. RMSD between two structures is 0.65 and the collectivity of the motion is calculated as 0.34. Modal analysis of open form indicates importance of first three modes which is enough to resemble experimental displacement

and after these modes the similarity calculations diminish significantly. First mode of the open form has the best mode correlation and overlap as 0.69 and 0.62 respectively and the similarity can even increase more if we remove the terminal residues. Second mode follows first mode with a mode correlation of 0.64 and overlap of 0.40. Third mode is also significant with a mode correlation of 0.59 and an overlap of 0.36.

Individual modes of closed form also exhibit similar behavior in which first three modes distinguish from the rest based on their similarity scores. The mode correlations are calculated as 0.61, 0.49 and 0.50 and the overlaps are calculated as 0.49, 0.48 and 0.47 for first, second and third modes respectively. Actually similarity scores can both be improved in both structures with the removal of terminal residues. Eigenvalue distributions of both conformers share a similar pattern in which lower frequency modes simply dominates the motion which is reflected by high collectivity of the motion for the enzyme.

RNase discriminates from the AK if we consider similarity scores of closed states. As presented above closed form of AK encounters difficulties because of high RMSD between two structures but lower RMSD between two structures of RNase enable the closed form to sample open form better. Eigenvalue distributions also differ between two proteins. Two structures of RNase share a similar pattern for eigenvalue distribution, however the eigenvalue distribution of open and closed structures of AK differ significantly.

4.2.4. 34E4 Fab Antibody

34E4 Fab Antibody is another antibody adopted in our database and akin to Spe7 based on their behaviors. Like Spe7, 34E4 also follow a conformational diversity due to presence of pre-existing conformers. One of these conformers is unable to bind and stays in closed form and another conformer binds ligand with an open architecture for binding site. As presented in Chapter 3 antibody do not possess major structural changes between free and bound forms except a ring in L chain [42, 102]. Here, we

perform modal analysis of non-binding (adopted bound form of the antibody from RSCB in order to resemble non-binding form PDB: 1Y0L) and free forms of the antibody. RMSD between two structures is 1.79 and collectivity of the motion is 0.17. The experimental displacement of the antibody has a peak around the residues 140 and 148 with amplitude of about 10 Å within a loop region located between two β-sheet structures. Except this region, most of the residues have maximum amplitude of 1.5 Å.

Both open and closed forms of the antibody perform best similarity scores using first, second and sixth modes. According to modal analysis results these modes differ from the others significantly based on their similarity scores. The mode correlations of open form are calculated as 0.53, 0.65 and 0.61 for first, second and sixth modes respectively. The first mode gave an overlap of 0.40; where second and sixth modes share an overlap of 0.34. The best mode correlation after these modes have been calculated as 0.50 (overlap 0.02, mode 10) and the best overlap has been calculated as 0.16 (mode correlation 0.34, mode 11). Non-binding closed form produced mode correlations as 0.44, 0.59 and 0.62 and overlaps 0.45, 0.29 and 0.33 for first, second and sixth modes respectively. Mode correlations of that form still produce significant scores for some other modes including third and fourth, however overlap scores of these modes are fairly lower than these three modes. According to mode contributions of Fab antibody lowest frequency modes of both conformers fairly dominate observed motion.

4.2.5. Maltose Binding Protein (MBP)

MBP is a periplasmic binding protein composed of two domains, N-Terminus and C-Terminus domains respectively. A central helix links two domains [41]. We have performed our ANM calculations over free and partially closed forms of the protein. RMSD between two forms is 4.97 and the collectivity of the motion is calculated as 0.62. The motion of NTD comprises fluctuations with amplitudes between 2 Å and 10 Å followed by a more rigid linker and in CTD fluctuation amplitudes again increase up to 8 Å. Analysis of individual modes of open form indicated that the overlap scores immediately diminish after fourth mode although still we can find higher frequency

modes that has mode correlation scores as high as first four modes. These four modes produced mode correlations as 0.39, 0.27, 0.32 and 0.33 with overlap scores of 0.46, 0.61, 0.31 and 0.25 respectively.

According to similarity scores of the partially closed form the motion is largely dominated by first three modes because overlap scores diminish immediately after these modes similar to the open form. Mode correlations are similar for the first three modes (0.35, 0.31 and 0.3 respectively), but the overlap score of the second mode has the highest overlap score of 0.67 followed by first and third modes which share a overlap of 0.36. Individual modes of both structures resemble mobile regions well but fail to resemble the rigid structure of linker region.

Eigenvalue spectrums of two structures are akin to each other in which the low frequency modes become insufficient to determine whole displacement and motion is not dominated by a unique mode instead it is distributed amongst a large spectrum of eigenvalues implying importance of higher modes.

4.2.6. Thrombin

Thrombin which is an important protein in coagulation cascade possess Na^+ dependent behavior and pre-exist as a slow (free thrombin) and fast form (Na^+ bound) in solution [38]. The rmsd between these two conformers is 0.41 and the collectivity of the experimental displacement is 0.25. Because of low rmsd, the experimental displacement of thrombin upon Na^+ binding is limited to amplitudes lower than 0.5 Å for most of the proteins.

Similarity scores of individual modes for free form indicate importance of four modes including first, second, fourth and fifth modes. Mode correlations of these modes are calculated as 0.47, 0.50, 0.50 and 0.63 respectively. Overlap scores are 0.20, 0.33, 0.27 and 0.22 respectively. The discrimination of these four modes among the rest is caused overlap scores. While there are modes that produce similar mode correlations with these four modes, the overlap scores of the rest fail to produce similar scores.

The distinction by the overlap scores is also present in the fast form of the protein. Although there are several modes that produce significant mode correlations there are only five modes that produce significant overlap scores. These modes are first, second, fourth, fifth and eighth with overlaps 0.20, 0.27, 0.23, 0.22 and 0.25 respectively. Mode correlations of the same modes are 0.40, 0.42, 0.39, 0.54 and 0.64 respectively.

According to mode contributions both structures share a similar pattern indicating importance of higher frequency modes for determined overall motion. The decline in mode contributions is not sharp as in Adenylate Kinase or RNase.

4.2.7. Trisophosphate Isomerase (TIM)

The catalytic activity of TIM comprises an opening and closing motion of a well defined loop between open and closed states. This loop called as loop 6 consists of the residues between V167 and T177. Except the motion of this loop, the protein does not possess any major changes between open and closed forms [56]. This case is reflected in the collectivity of the enzyme motion which is calculated as 0.07, with an rmsd of 0.46.

The low collectivity of the enzyme distinguishes TIM from other proteins presented until now. Lowest frequency modes perform well to represent collective motions thus in order to represent local motions modes with higher frequencies need to be considered. If we investigate the individual modes of the open structure of enzyme, we can see the failure of lowest three modes to resemble the local motion. However for the higher modes like sixth, seventh or even twentieth mode, the similarity scores are improved considerably. For example best mode correlation is produced by twentieth mode as 0.69 (overlap: 0.22). This mode is followed by sixth mode which has a mode correlation of 0.68 with an overlap score of 0.37 which is the best overlap score produced amongst the first thirty modes.

The similarity calculations decrease significantly in the case of closed form. Highest mode correlation (0.43) and overlap (0.22) is produced by the 26^{th} mode. According to

overlap scores this mode is followed by seventh and fourteenth modes with an overlap of 0.20 and 0.14 respectively (mode correlations: 0.12 and 0.27 respectively).

Eigenvalue contributions of open and closed forms mutually indicate that the overall enzyme motion cannot be defined by the collective motions solely and require further higher frequency modes.

4.2.8. Dihydrofolate Reductase (DHFR)

DHFR performs reduction of DHF to THF. According to literature DHFR exists as two distinct conformers within the solution with each capable of binding DHF or NADPH. Akin to behavior of TIM, DHFR also exhibits a loop motion (termed as Met^{20} loop) [25, 89]. Collectivity of the motion is 0.08 and the rmsd between two structures is calculated as 0.50.

Modal analysis results of the NADPH binding form of the enzyme indicates that to represent observed experimental displacement the structure requires higher modes. The best mode correlation (0.48) and overlap (0.27) scores were produced by thirteenth mode and this mode was followed by twelfth mode which produced a mode correlation of 0.33 and an overlap of 0.20.

DHF binding form of the enzyme produced its best similarity scores through 24^{th} mode with a mode correlation of 0.46 and an overlap of 0.24. This mode was followed by thirteenth and twenty fourth modes which produced mode correlations of 0.43 and 0.40 and overlaps of 0.19 and 0.13 respectively. The failure of low frequency modes are related with low collectivity of the enzyme similar to similarity scores of TIM, DHFR also favored higher modes to define loop motion.

Contribution spectrum indicates that there is not a unique mode that defines the whole motion but the distribution pattern favors a large spectrum of modes in order to define the whole motion.

4.2.9. Cytochrome P450

For Cyctochrome P450 EryK literature [26] proposes that binding of protein to the ligand (ErD) includes an interconversion between ligand-binding (open) and non-binding (closed) forms of the enzyme. This protein plays an important role in the oxidative metabolism of endogenous compounds such as antibiotics and steroids. The rmsd between open and closed forms of the enzyme is calculated as 2.63 and the collectivity of the experimental displacement is 0.42.

As presented in Methods first nineteen residues of EryK was not considered within the ANM calculations. Literature implies presence of both core and mobile regions within EryK. Motion of helix F (comprising residues W165 and Q173) has the maximum amplitude. Together with helix F, BC loop also possesses a major displacement between two conformers. Modal analysis of open form indicates importance of second, ninth and twelfth modes according to their similarity scores. Mode correlations for these modes are 0.31, 0.42 and 0.53 and overlap scores are calculated as 0.19, 0.22 and 0.27 respectively.

For the closed form of the protein, overlap scores are declined below 0.10 for most of the proteins. Thus, we have neglected overlap scores for this conformer of the protein and just dealt with the mode correlations. According to these results modes 3, 13 and 18 produced best mode correlations as 0.37, 0.39 and 0.61 respectively.

Although the collectivity of motion indicates a collective motion for the protein, local displacements of BC loop and helix F force the protein to seek for higher modes reducing the similarity scores of lower modes.

According to mode contributions higher modes still preserve their importance for both substates, however for the closed form of the enzyme the mode contributions diminish faster than the open form which implies that eigenvalues spectrum of open form favors higher modes significantly higher than non-binding form.

4.2.10. NtrC

NtrC is a signaling protein switching between inactive and active states during signal transduction driven by phosphorylation. According to literature, independent from phosphorylation of the protein it can also populate active form together with inactive form [44]. The structures of both conformers have been determined by NMR [44] and the rmsd between these structures have been calculated as 1.78 with a collectivity of 0.46.

Phosphorylation site together with helix α4 exhibited highest displacement after superposition and third mode of the inactive form has both the highest correlation and the highest overlap scores as 0.42 and 0.24 respectively. Fifth and seventh modes followed this mode with mode correlations as 0.23 and 0.35 and overlaps as 0.23 and 0.16 respectively.

Inactive form of the protein produces slightly lower similarity scores with the best mode found to be the second one based on the overlap scores. The mode correlation and overlap of this mode has been calculated as 0.29 and 0.21 respectively. Ninth and twelfth modes have also found as significant with a mode correlation of 0.22 and overlap score of 0.16 for ninth mode and these scores were determined as 0.35 and 0.19 for twelfth mode respectively.

Similar to the behavior of Cytochrome P450 eigenvalue contributions of NtrC also diminish faster for the ligand-bound form. Inactive form of the protein favors contribution of higher modes much more compared to active form.

4.2.11. Cholesterol Oxidase

FAD dependent enzyme cholesterol oxidase performs isomerization of steroids. For this purpose cholesterol oxidase utilizes FAD during oxidation reduction reactions [103]. We have adopted steroid bound (PDB: 1COY) and FAD bound conformers of the enzyme (PDB: 3COX) and performed ANM analysis to both of them. RMSD between two structures have been calculated as 0.24 with a collectivity of 0.24.

The cholesterol oxidase is a large two domain protein comprising 500 residues. The experimental displacement was mostly restricted below 0.4 Å except a region that exceeds 1.5 Å. The modal analysis produced fairly low similarity scores below 0.20 for mode correlation and with a highest overlap of 0.20 for androgen bound form of the enzyme. First mode has given a mode correlation of 0.18 with an overlap 0.14. Best overlap score was calculated as 0.20 with the third mode (mode correlation: 0.1).

The second conformer, only FAD bound form of the enzyme has produced similar results with a highest correlation of 0.18 (overlap 0.14) obtained by first mode. The best overlap has been obtained by sixth mode as 0.21 (mode correlation 0.07). As can be observed from these results individual modes fail to perform well for a limited conformational change. Instead a more collective and large-amplitude displacements can be well predicted by individual modes.

Eigenvalue spectrum of both structures of protein highlights the importance of higher modes since low frequency modes do not define the experimental displacement completely.

4.3. Mode Selection

In previous pages we have analyzed modal analysis results of each conformer. The results elucidate several dependencies of similarity scores including rmsd, collectivity and length of the protein. For example a high rmsd between the open and closed conformers of the Adenylate Kinase produces differences between similarity scores of open and closed forms in which individual modes of closed form encounter difficulties to define the experimental displacement. In the reverse case, when the rmsd is low as in the case of cholesterol oxidase, success of individual modes is limited for both similarity scores. The size of cholesterol oxidase is another factor affecting the success of individual modes together with the low rmsd.

Modal analysis requires a certain level of collectivity for the conformational transition [20]. In the case of low collectivity, for example loop motions of TIM and

DHFR, success of individual modes are declined significantly. Global modes which are the first three modes of these proteins fail to define these local motions; instead higher modes perform better for limited motions.

We have also investigated the performance of randomly created mode vectors for corresponding conformational transitions to check if the success of individual modes can be repeated by random modes or they actually carry information about the fluctuation pattern. The highest three mode correlations and overlap scores obtained from random modes are presented in Appendix 3. As we have stated previously there are several modes within our dataset that produces significant similarity results. Comparing these ANM produced modes with randomly created mode vectors supply evidence about the importance of the corresponding modes. According to the mode correlations, most of the selected modes perform higher similarities except NtrC, Cholesterol oxidase and TIM. For these proteins random modes give similar mode correlations with the selected ones. However, for the selected modes of these proteins given in Appendix 2 the overlap scores of each structure still performs fairly better than the random modes.

In further parts we have utilized mode analysis results for selection of the significant modes. Our results imply that this process is not straightforward for most of the proteins in which several modes exhibit considerable similarities with the experimental displacement. According to our results, for most of the proteins, there are three modes that produce significant similarities mainly based on their overlap scores. Thus, we have selected top three modes for our further analysis.

Main concentration of this study is focused in the last part. Here, we propose that in addition to structurally resembling ensemble conformers modes also carry intrinsic relations with thermodynamics and kinetics of the protein through the mode weights. As we have presented in first chapter, the energy landscape theory includes several energy minimas, in other words substates, distinguished by energy barriers of several kT.

Protein Name	PDB I	PDB II	kinetic rates (s^{-1})	population
Adenylate Kinase	open	closed	1374/286	No population data
Spe7	Ab[1]	Ab[2]	17/58	80 % /20 %
DHFR	NADPH binding	DHF binding	0.034/0.068	No population data
34E4 Fab Antibody	Ligand-binding	Non-binding	1.55/0.51	No population data
Cholesterol Oxidase	1coy (A)	3cox (A)	9.0/1.03	No population data
EryK Cytochrome	Non-binding	Binding form	1.31/0.51	No population data
RNase	Open	Closed	No kinetic data	93-95% / 5-7%
TIM	Open	Closed	No kinetic data	4 % / 96 %
Maltose Binding P.	Open	Partially closed	No kinetic data	95% / 5%
NtrC	inactive	active	No kinetic data	90-98% 2-10%
Thrombin	slow form	fast form	No kinetic data	40% / 60%

Table 2 Kinetic rates / relative populations of dataset proteins

Protein populates these distinct substates according to environmental conditions and height of these energy barriers.

The interconversions between distinct substates provide a dynamic nature for the protein. ANM is readily found to retrieve intrinsic tendency of different substate structures which permits them to resemble each other. Thus, we have investigated applicability of modal analysis for the thermodynamics/kinetics of conformational ensemble which comprises relative populations and interconversion rates respectively. For this purpose we have employed contribution of each mode that is determined by using solely eigenvalues. In Table 2 we have supplemented experimental data about the kinetic rates or relative populations of proteins conformers, except Spe7 which includes

both kinetic conversion rates and relative populations. These properties are directly retrieved from experimental studies. As presented in Table 4-2 six out of eleven proteins including Spe7 contained experimental data about the kinetic rates and six others together with Spe7 contained population data. The kinetic rates and relative populations for each protein have been supplemented sequentially for corresponding conformers.

We have initially considered kinetic rates and relative populations separately. In this case we have six proteins containing kinetic rates and six others containing relative populations. For each dataset we have constructed computational and experimental vectors as told in Chapter 2 by utilizing eigenvalues of the selected modes. After construction of these two vectors for each dataset we have calculated the Pearson correlation between computational findings and experimental data. For the dataset with kinetic rates we have calculated a correlation of 0.97 with the selected modes presented in Table 4.3. For the population dataset we have calculated a correlation of 0.95 with the selected modes presented in Table 4.4. For each protein selected modes have been presented together with the selected mode of second conformer given in brackets. The values given in X column indicated experimental ratios and the values in Y indicated computational ratios.

If we repeat this calculation with the modes which produce best mode correlations, the similarity becomes 0.65 and 0.27 for kinetic rates and relative populations respectively. In the case of the best overlapped modes, we obtain a result of 0.06 and 0.5 for kinetic rates and relative populations respectively. Still, these modes presented in table 3 are significant ones according to their overlap and mode correlation scores with the most of them produced whether best overlap or best mode correlation.

In the case of complete dataset with the kinetic rates and populations data used together we have obtained a calculation of 0.94 with the same modes used in Table 3 and 4 for 22 substates. In these calculations we adopted kinetic rates of Spe7 which contained both data. In all three cases including set of kinetic rates, set of population

and set of both we have calculated significant correlations between the computational findings and the experimental data.

Table 3. Relation between Kinetic Rates and mode weights					
Protein	**PDB I**	**PDB II**	**Modes**	**X**	**Y**
Adenylate Kinase	4ake	1ake	2(1)	0.208	0.45
Spe7	1oaq	1ocw	7(5)	3.45	1.21
DHFR	1rx1	1rx7	13(24)	2	0.70
Fab Antibody	3cfj	1y0l	2(6	0.32	0.33
Cholesterol Oxidase	1coy	3cox	1(6)	0.11	0.41
EryK	2wio	2jjn	12(13)	2.56	0.875
Correlation					**0.97**

Table 3: Selected Modes for proteins with kinetic rates

Table 4. Relation between relative populations and mode weights					
Protein	**PDB I**	**PDB II**	**Modes**	**X**	**Y**
Spe7	1oaq	1ocw	7(5)	4.0	1.21
RNase	7rsa	1u1b	3(1)	15.7	2.29
TIM	1ypi	7tim	6(7)	0.04	0.8
NtrC	1dc7	1dc8	3(2)	15.0	2.21
MBP	2v93o	2v93c	3(1)	19.0	1.95
Thrombin	1sgi	1sg8	5(8)	0.67	0.77
Correlation					**0.95**

Table 4: Selected modes for proteins with relative populations

Chapter 5

CONCLUSION

Proteins in their native states possess innate flexibility which enables them to sample a space of numerous conformers. Existence of proteins as not a static structure instead as distinct conformers in their native states has so many implications in protein function and ligand binding. Kinetic conversion rates of pathway intermediates are compatible with turnover kinetics of several enzymes implying enzyme dynamics to have direct roles during catalysis. Ligand recognition and binding is also related with preexisting conformers where ligand specifically selects appropriate conformer. This model helps to understand antigen diversity observed in antibodies.

Together with experimental studies, computational efforts also improve knowledge about protein dynamics. The innate flexibility of proteins can be extracted through global modes which are determined by coarse-grained EN models. EN models which have a history in protein dynamics for more than a decade produce significant results for both temperature factor predictions and conformational transitions.

ANM and GNM have been mostly adopted in protein dynamics to predict structural changes providing information about the fluctuation amplitudes; however had no implications for the timeline of these fluctuations. Here, we have aimed to reproduce success of EN models for fluctuation timelines and thermodynamics of conformers. The dataset we have adopted here includes antibodies, enzymes and signaling proteins. Modal analysis of eleven proteins or 22 different PDB structures indicated that together with collectivity, modal analysis scores depend also on protein length and root mean square deviation between distinct states. As in the case of DHFR and TIM which perform loop motions the displacement is restricted a small region indicated with a low rmsd and collectivity. In these cases, higher modes of the protein become more important compared to high frequency modes of proteins with collective motions.

Our findings indicate a relation between kinetics/thermodynamics of ensemble conformers and eigenvalues of Hessian matrix. These eigenvalues are known to

determine the effect of each individual mode within the overall motion. According to the results of this study eigenvalues do not only carry information about the mode contribution, but also have importance about the kinetic and thermodynamic properties of the system. This study highlights the importance of protein dynamics in which together with the encoded nature of ensemble conformers, kinetic and thermodynamic properties of these conformers are also encoded in protein structure. EN models access this encoded information through individual modes.

Our hypothesis is supported by the results of eleven proteins but require further enlargement within the dataset via addition of new proteins in order to be validated. Since we combine several distinct information including different PDB structures (conformers) of the same protein and the kinetic or thermodynamic data about these conformers, our study does largely depend on accumulation of these findings. Besides, the proteins in their open forms resemble closed-like structures or vice versa. However, the exact structures of closed-like or open-like structures are not known, thus, instead we adopt open or closed structures in order to represent indicated structures neglecting final ligand-induced rearrangements in binding site. Thus, together with addition of new proteins our results can be improved via using exact closed-like or open-like structures in the presence of structures for corresponding conformers.

As a future work our study can be extended to combined modes which improve similarity scores. The link between eigenvalues and kinetic rates or eigenvalues and relative populations can be investigated using results of combined modes.

APPENDIX 1

B-Factor Prediction Comparisons of Dataset Proteins

B-Factor prediction comparisons of our nine dataset proteins are given in following table. Two of dataset proteins have been determined using NMR methods thus, B-factors were not presented in PDB file.

PDB	Chain	Length	GNM	ANM	pfGNM	pfANM2	pfANM4
1AKE	A	214	0.48	0.55	0.58	0.55	0.59
4AKE	A	214	0.73	0.79	0.83	0.77	0.80
1OCW	H	120	0.24	0.22	0.21	0.25	0.29
1OAQ	H	120	0.62	0.56	0.73	0.68	0.61
1RX1	A	159	0.50	0.36	0.53	0.42	0.49
1RX7	A	159	0.47	0.44	0.49	0.45	0.49
1Y0L	A	216	0.62	0.61	0.62	0.57	0.60
3CFJ	A	215	0.62	0.58	0.62	0.49	0.51
1COY	A	501	0.53	0.52	0.60	0.54	0.58
3COX	A	500	0.70	0.71	0.77	0.73	0.77
2JJN	A	395	0.51	0.14	0.34	0.20	0.39
2WIO	A	393	0.52	0.45	0.57	0.54	0.59
7RSA	A	124	0.64	0.42	0.66	0.66	0.65
1U1B	A	124	0.33	0.11	0.40	0.36	0.39
1YPI	A	247	0.52	0.23	0.54	0.50	0.54
7TIM	A	247	0.51	0.21	0.52	0.53	0.52
1SGI	E	250	0.70	0.55	0.76	0.75	0.70
1SG8	E	251	0.53	0.50	0.64	0.63	0.59

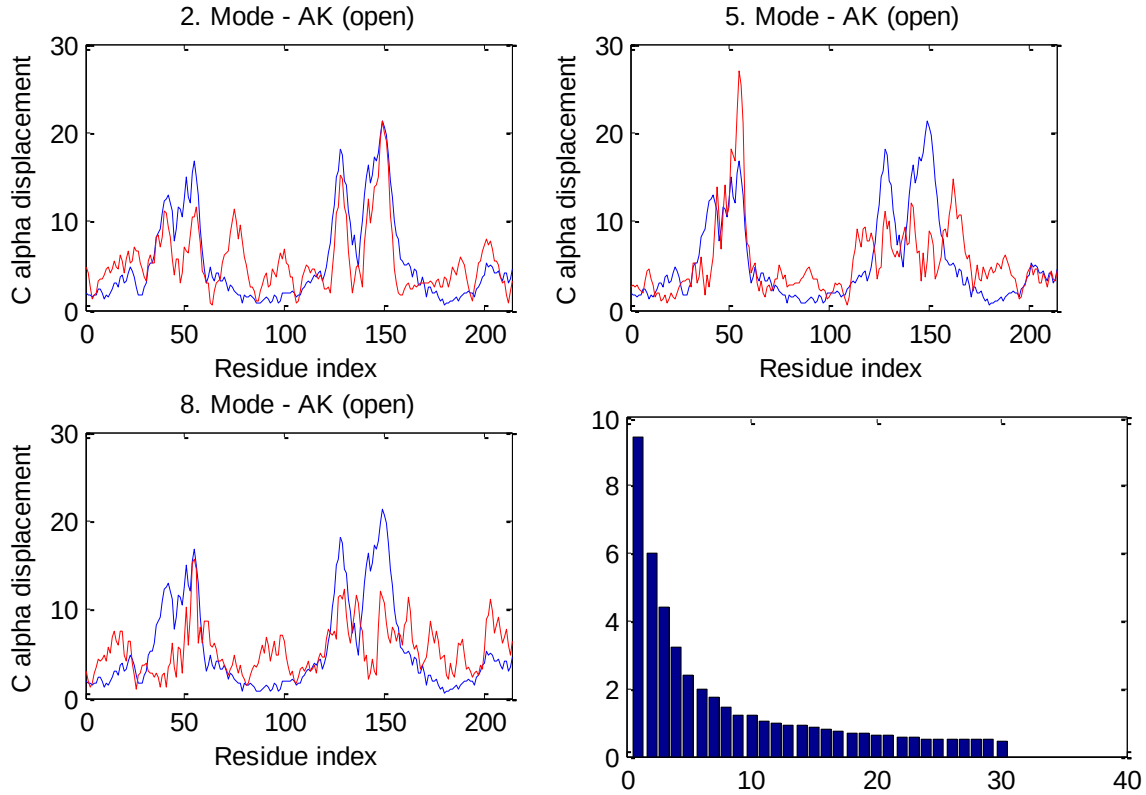
Table A. 1 B-Factor prediction comparisons of dataset proteins

APPENDIX 2

Mode Graphs and Contribution Spectrums

In this part, graphs of significant modes and contribution spectrum of first thirty modes are supplemented together with residue fluctuations calculated between two structures. In those graphs the residue fluctuations between open and closed form is represented by blue color and red color stand for corresponding mode fluctuations. The mode fluctuations are weighted such that the areas under both lines become equal. Contributions of each mode are given in percentages in contribution spectrums.

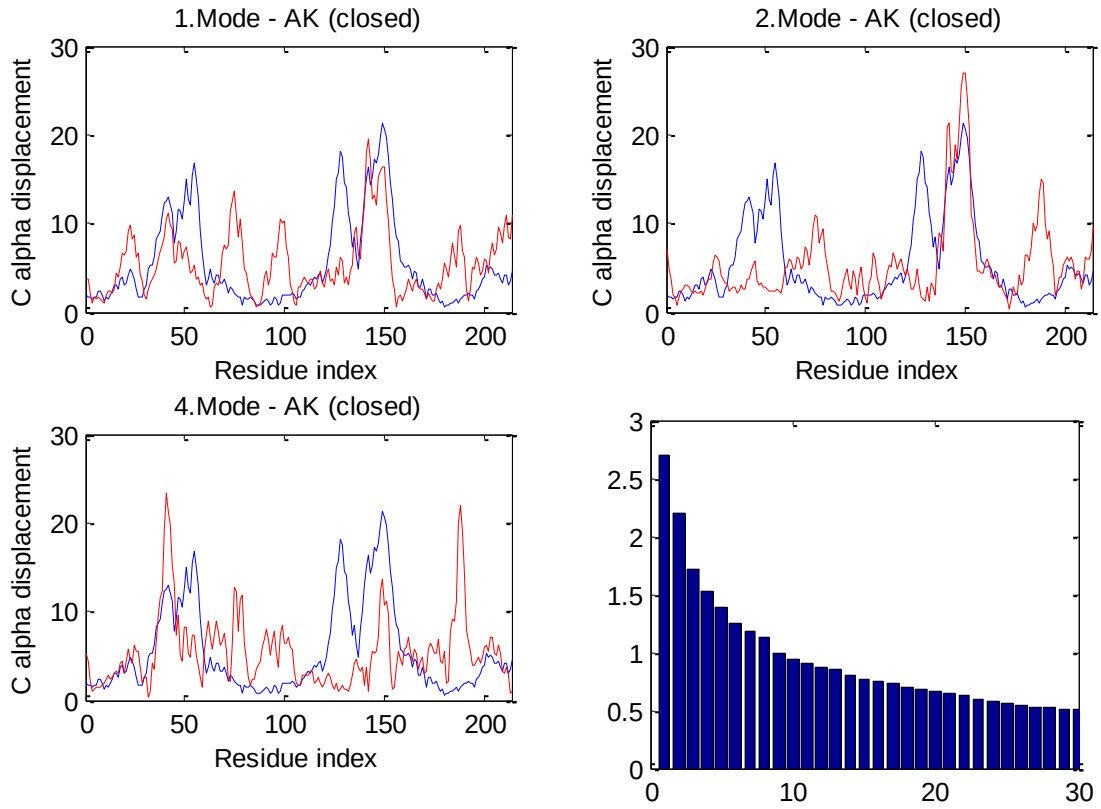
Adenylate Kinase (open form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
2	0.77	0.69	2,93E-05	0,094
5	0.54	0.36	1,15E-04	0,024
8	0.47	0.21	1,88E-04	0,015

Table A.2. 1 Mode Graphs and Contribution Spectrum of first 30 modes for AK (open form)

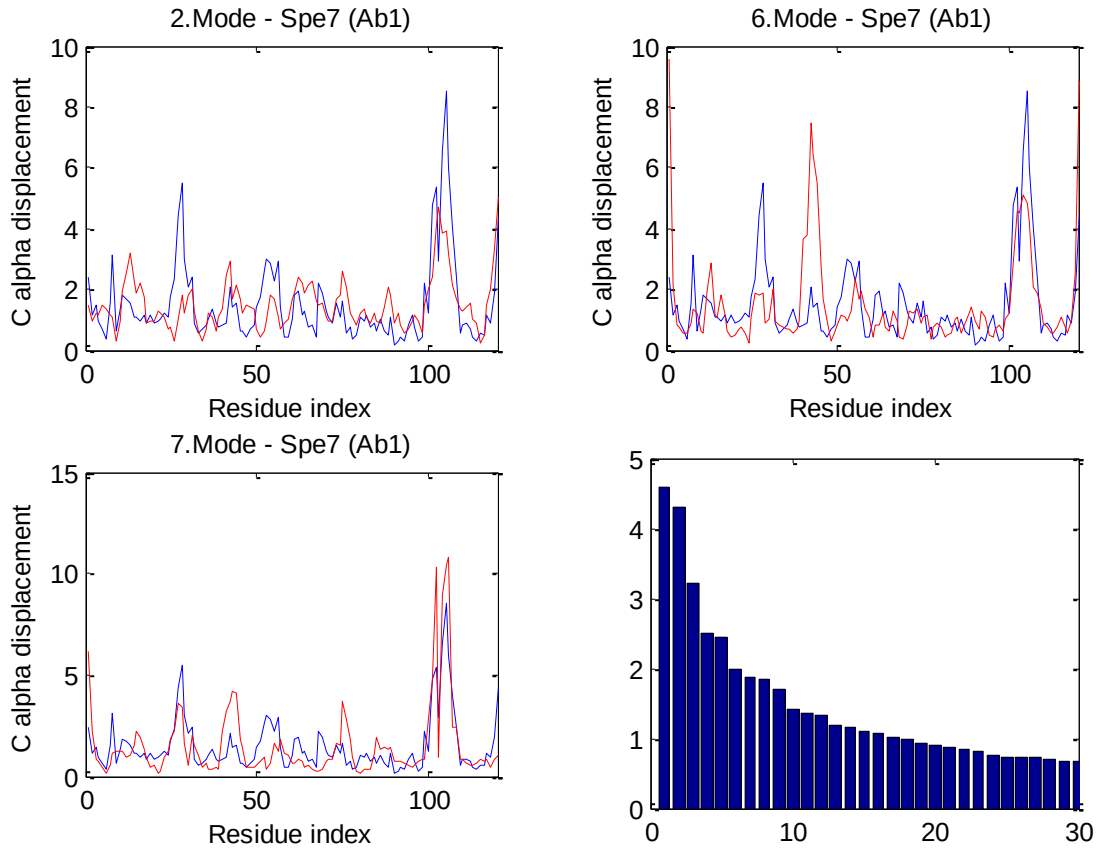
Adenylate Kinase (closed form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
1	0,48	0,54	1,48E-04	0,027
2	0,47	0,21	1,83E-04	0,022
4	0,12	0,29	2,64E-04	0,015

Table A.2. 2 Mode Graphs and Contribution Spectrum of first 30 modes for AK (closed form)

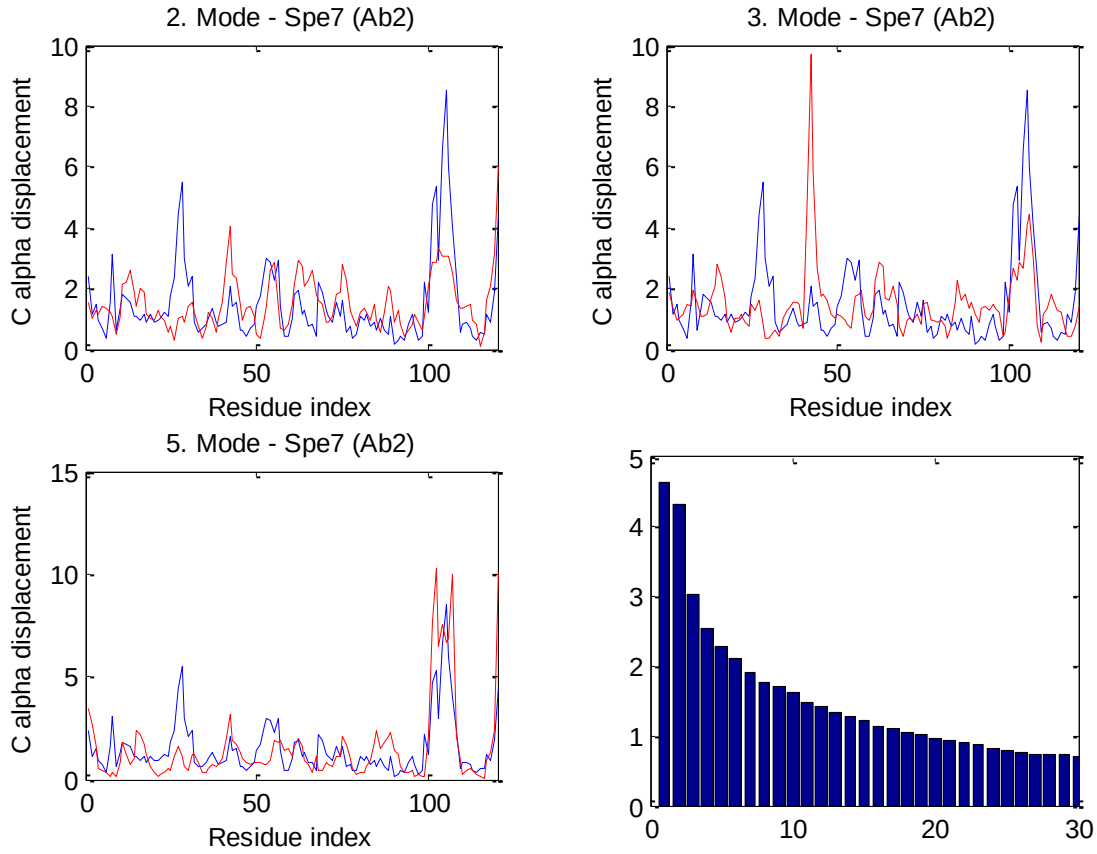
Spe7 (Ab[1] form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
2	0,58	0,35	1,30E-04	0,043
6	0,51	0,19	2,90E-04	0,02
7	0,74	0,27	3,00E-04	0,019

Table A.2. 3 Mode Graphs and Contribution Spectrum of first 30 modes for Spe7 (Ab[1] form)

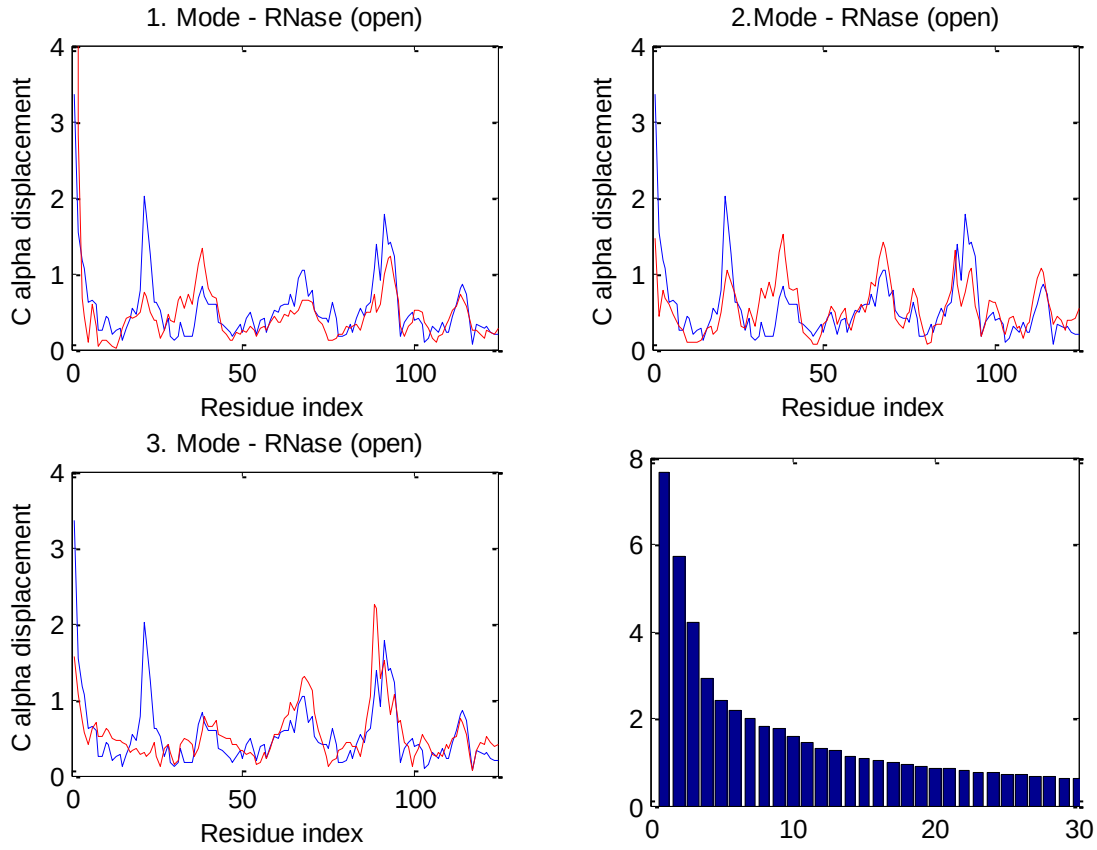
Spe7 (Ab^2 form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
2	0,49	0,34	1,30E-04	0,043
3	0,3	0,25	1,90E-04	0,03
5	0,7	0,28	2,50E-04	0,022

Table A.2. 4 Mode Graphs and Contribution Spectrum of first 30 modes for Spe7 (Ab^2 form)

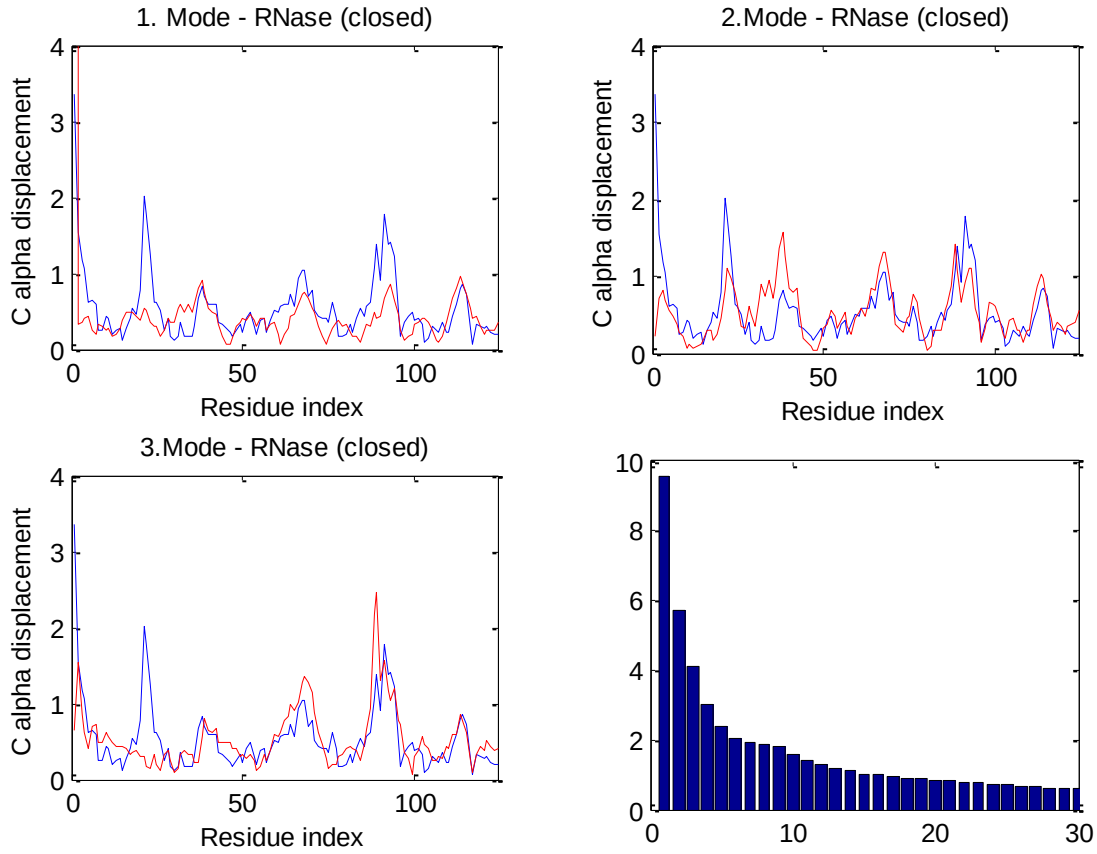
RNase (open form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
1	0,69	0,62	6,56E-05	0,076
2	0,64	0,4	8,81E-05	0,057
3	0,59	0,36	1,20E-04	0,041

Table A.2. 5 Mode Graphs and Contribution Spectrum of first 30 modes for RNase (open form)

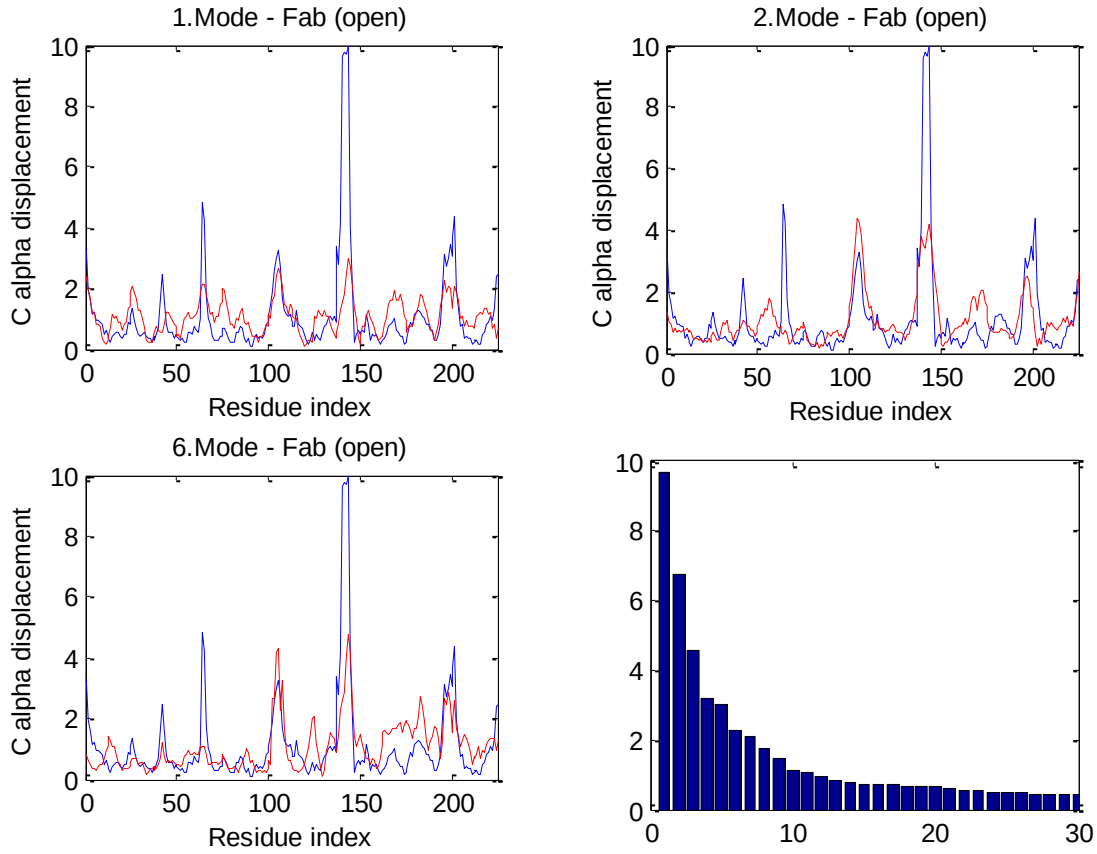
RNase (closed form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
1	0,61	0,49	5,10E-05	0,096
2	0,49	0,48	8,55E-05	0,057
3	0,5	0,47	1,20E-04	0,041

Table A.2. 6 Mode Graphs and Contribution Spectrum of first 30 modes for RNase (closed form)

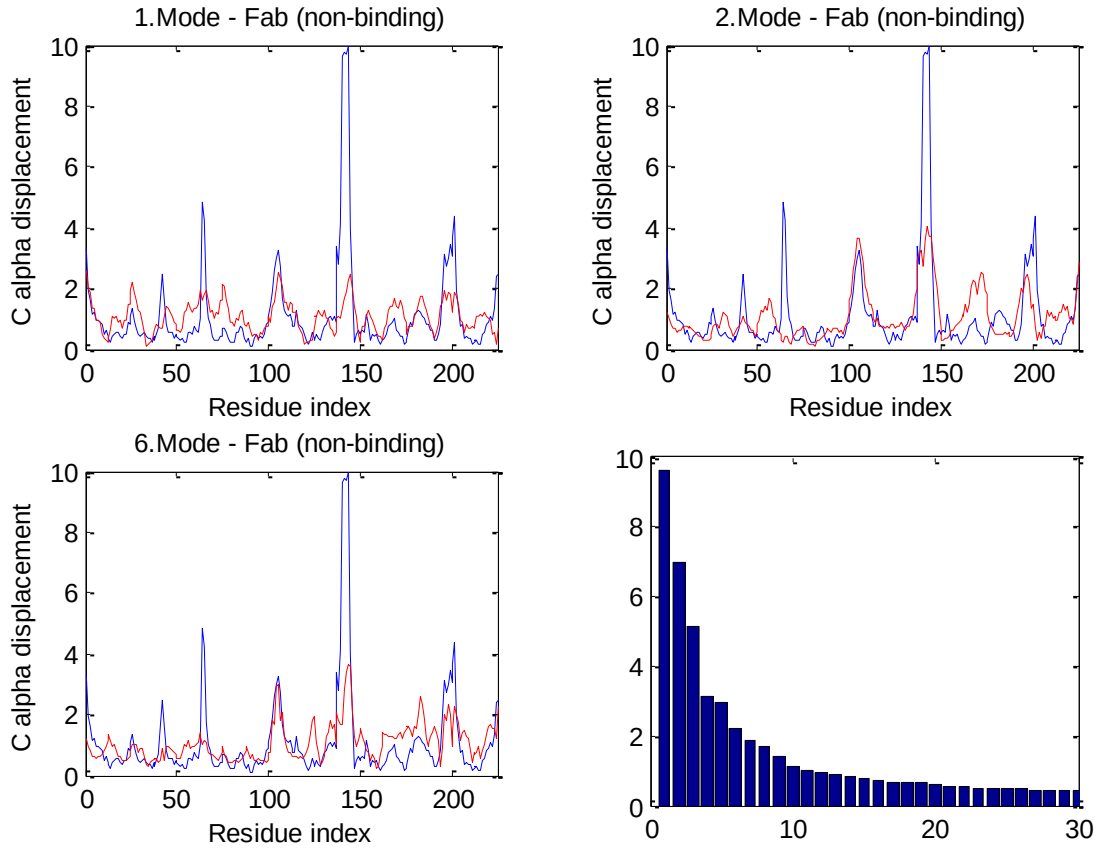
Fab Antibody (open form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
1	0,53	0,4	2,35E-05	0,097
2	0,65	0,34	3,38E-05	0,067
6	0,61	0,34	1,00E-04	0,023

Table A.2. 7 Mode Graphs and Contribution Spectrum of first 30 modes for Fab (open form)

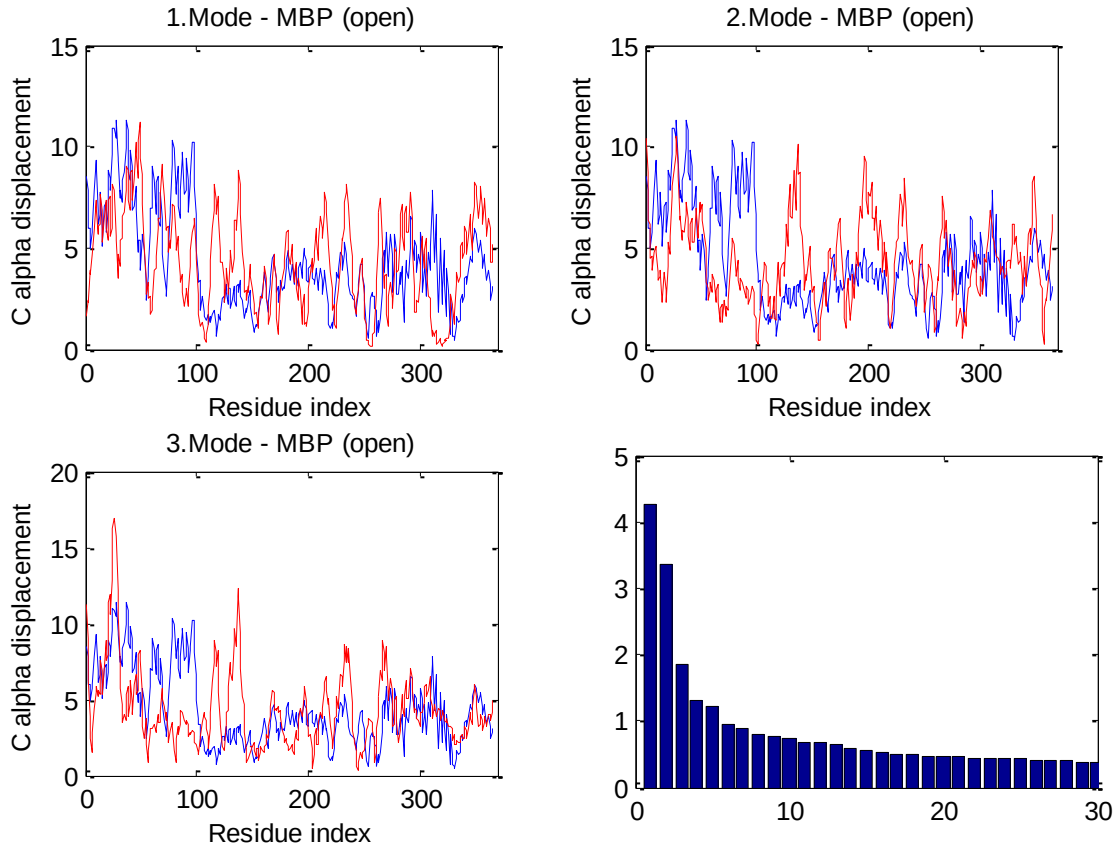
Fab Antibody (closed form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
1	0,44	0,45	2,49E-05	0,096
2	0,59	0,29	3,44E-05	0,07
6	0,62	0,33	1,00E-04	0,022

Table A.2. 8 Mode Graphs and Contribution Spectrum of first 30 modes for Fab (non-binding form)

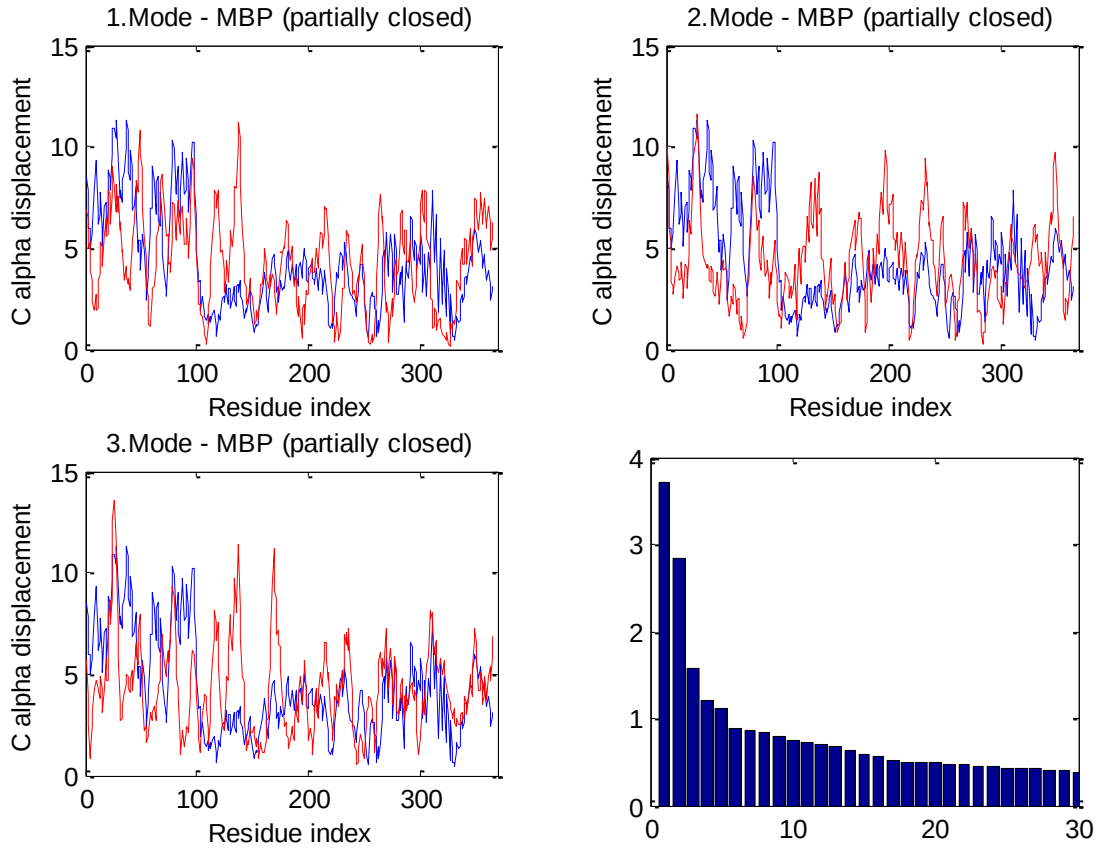
Maltose Binding Protein (open form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
1	0,39	0,46	5,92E-05	0,043
2	0,27	0,61	7,46E-05	0,034
3	0,32	0,31	1,35E-04	0,019

Table A.2. 9 Mode Graphs and Contribution Spectrum of first 30 modes for MBP (open form)

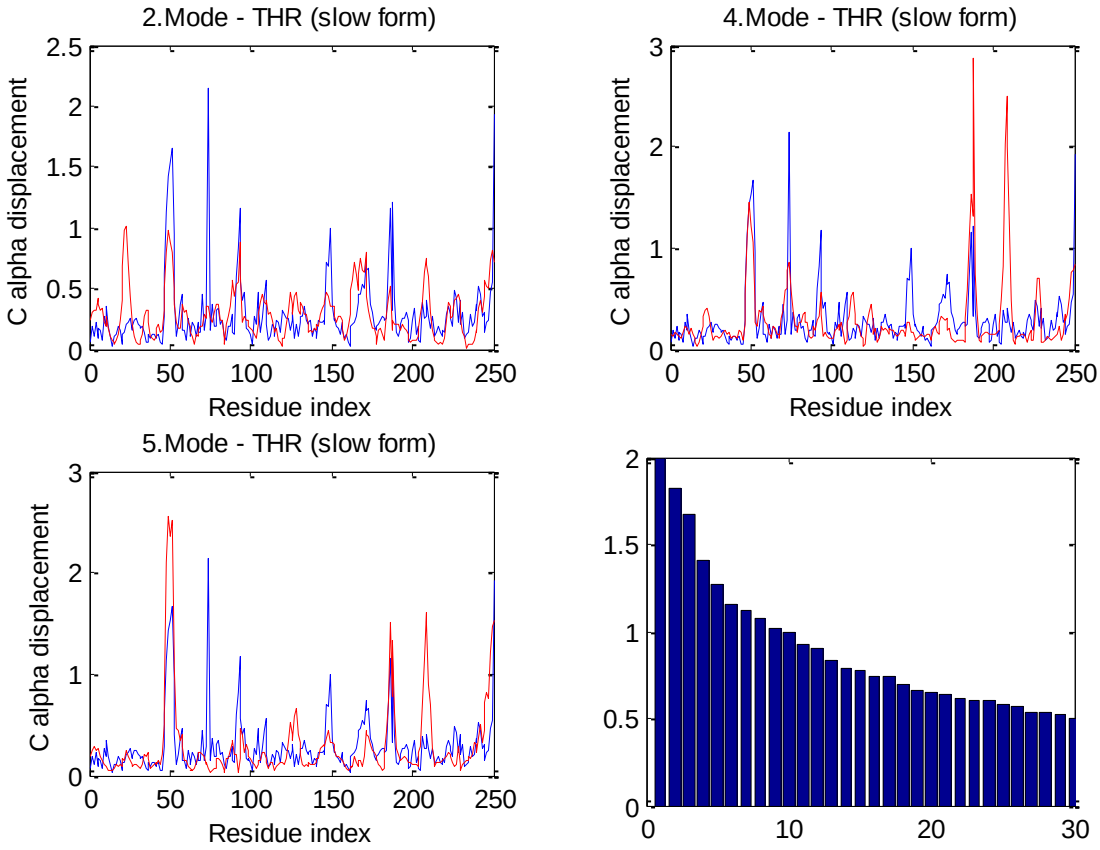
Maltose Binding Protein (partially closed form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
1	0,35	0,36	6,68E-05	0,037
2	0,31	0,67	8,69E-05	0,028
3	0,3	0,36	1,57E-04	0,016

Table A.2. 10 Mode Graphs and Contribution Spectrum of first 30 modes for MBP (partially closed)

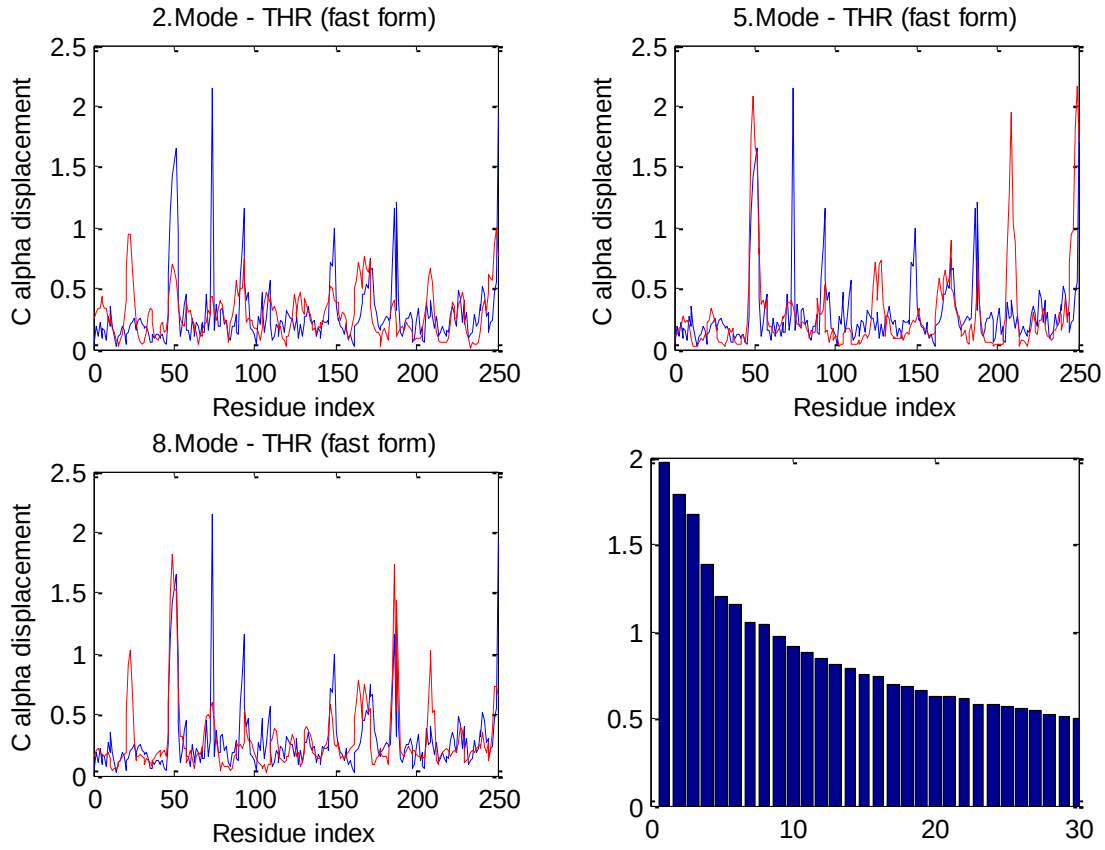
Thrombin (slow form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
2	0,5	0,33	1,96E-04	0,018
4	0,5	0,27	2,50E-04	0,014
5	0,63	0,22	2,80E-04	0,013

Table A.2. 11 Mode Graphs and Contribution Spectrum of first 30 modes for Thrombin (slow form)

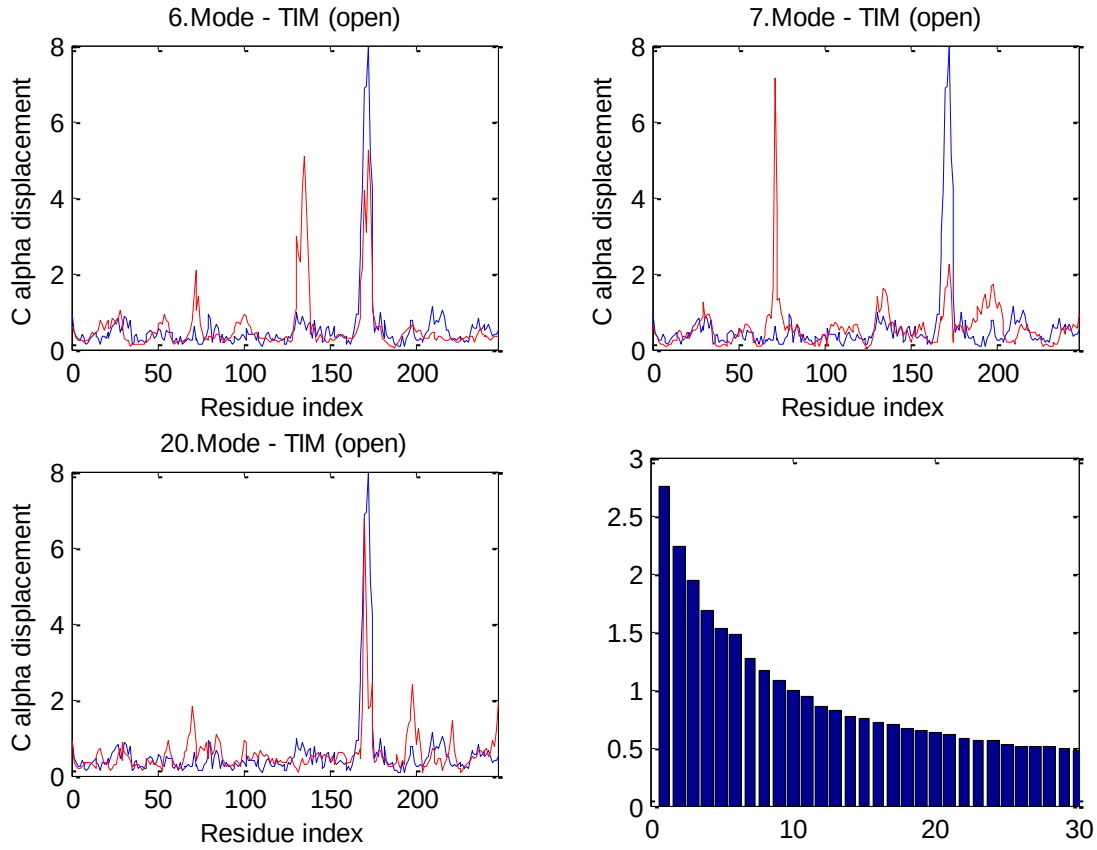
Thrombin (fast form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
2	0,42	0,27	2,00E-04	0,018
5	0,54	0,22	2,99E-04	0,012
8	0,64	0,25	3,45E-04	0,01

Table A.2. 12 Mode Graphs and Contribution Spectrum of first 30 modes for Thrombin (fast form)

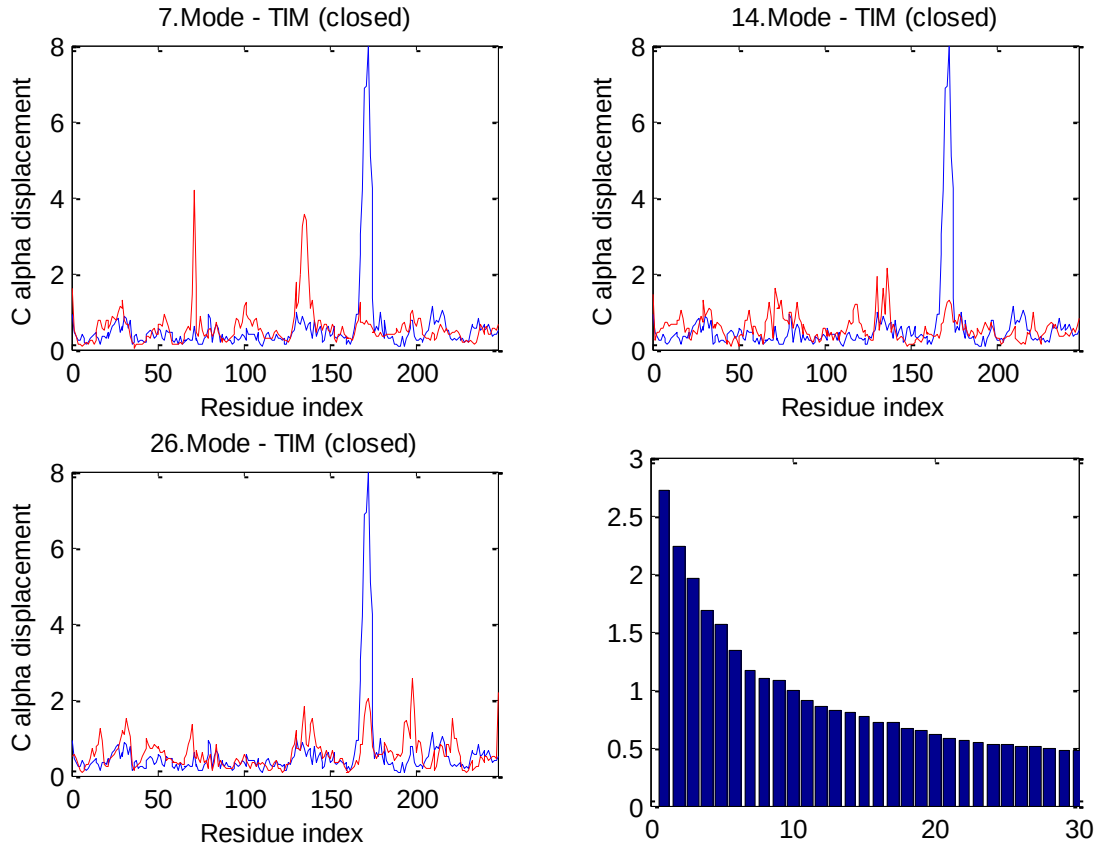
Triosephasphate Isomerase (open form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
6	0,68	0,37	2,40E-04	0,015
7	0,27	0,3	2,80E-04	0,013
20	0,69	0,22	5,60E-04	0,006

Table A.2. 13 Mode Graphs and Contribution Spectrum of first 30 modes for TIM (open form)

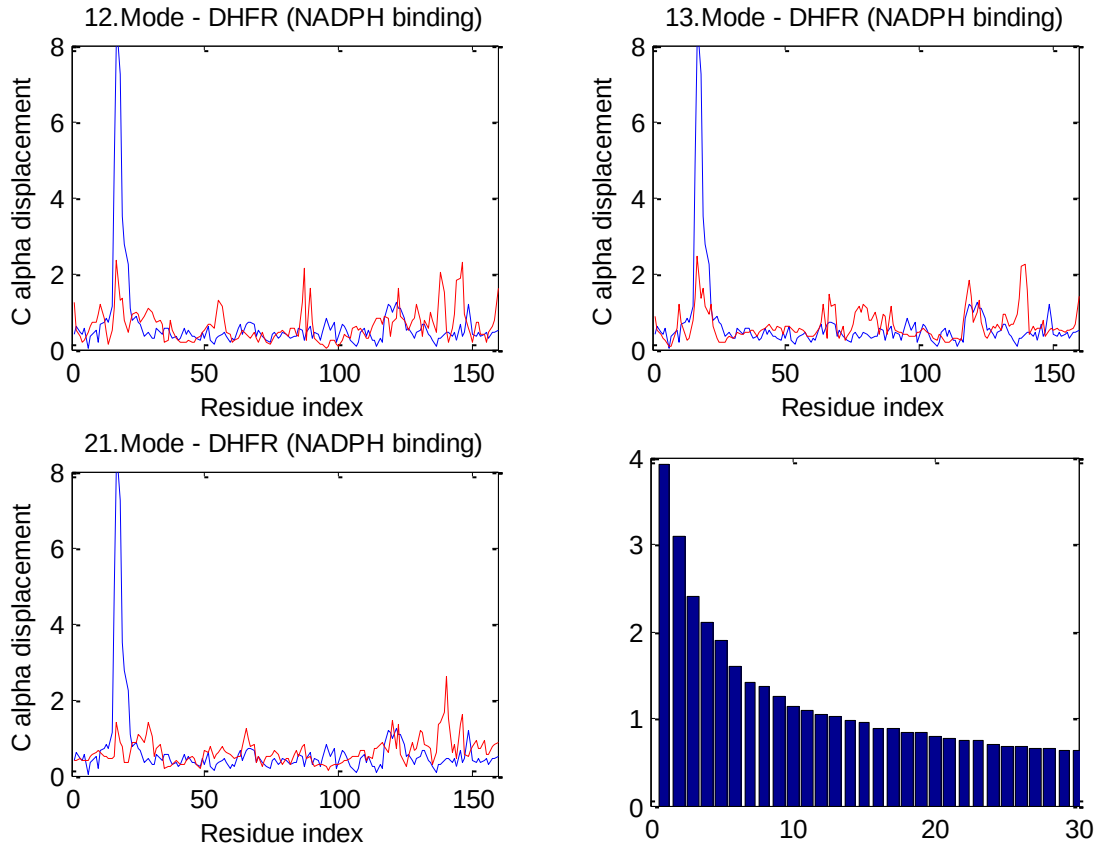
Triosephasphate Isomerase (closed form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
7	0,12	0,2	3,20E-04	0,012
14	0,27	0,14	4,60E-04	0,008
26	0,43	0,22	7,20E-04	0,005

Table A.2. 14 Mode Graphs and Contribution Spectrum of first 30 modes for TIM (closed form)

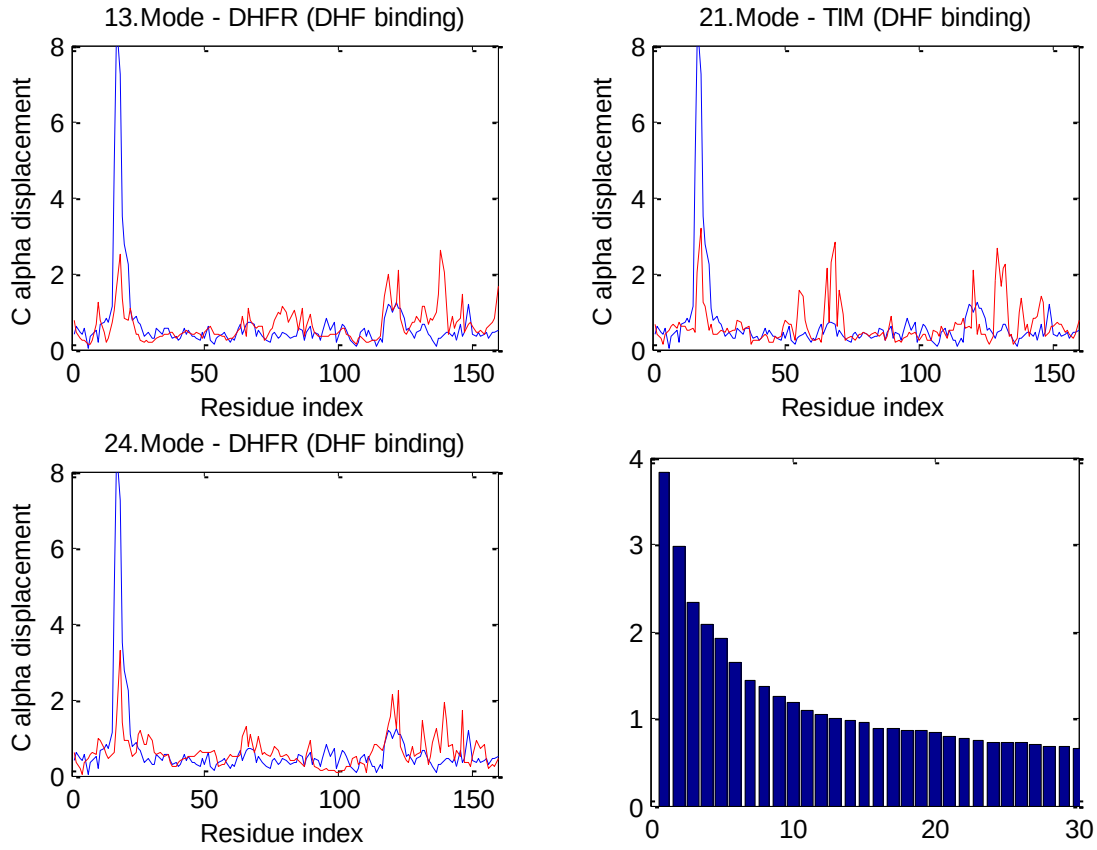
DHFR (NADPH binding form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
12	0,33	0,2	4,27E-04	0,01
13	0,48	0,27	4,36E-04	0,01
21	0,15	0,16	5,76E-04	0,008

Table A.2. 15 Mode Graphs and Contribution Spectrum of first 30 modes for DHFR (NADPH binding form)

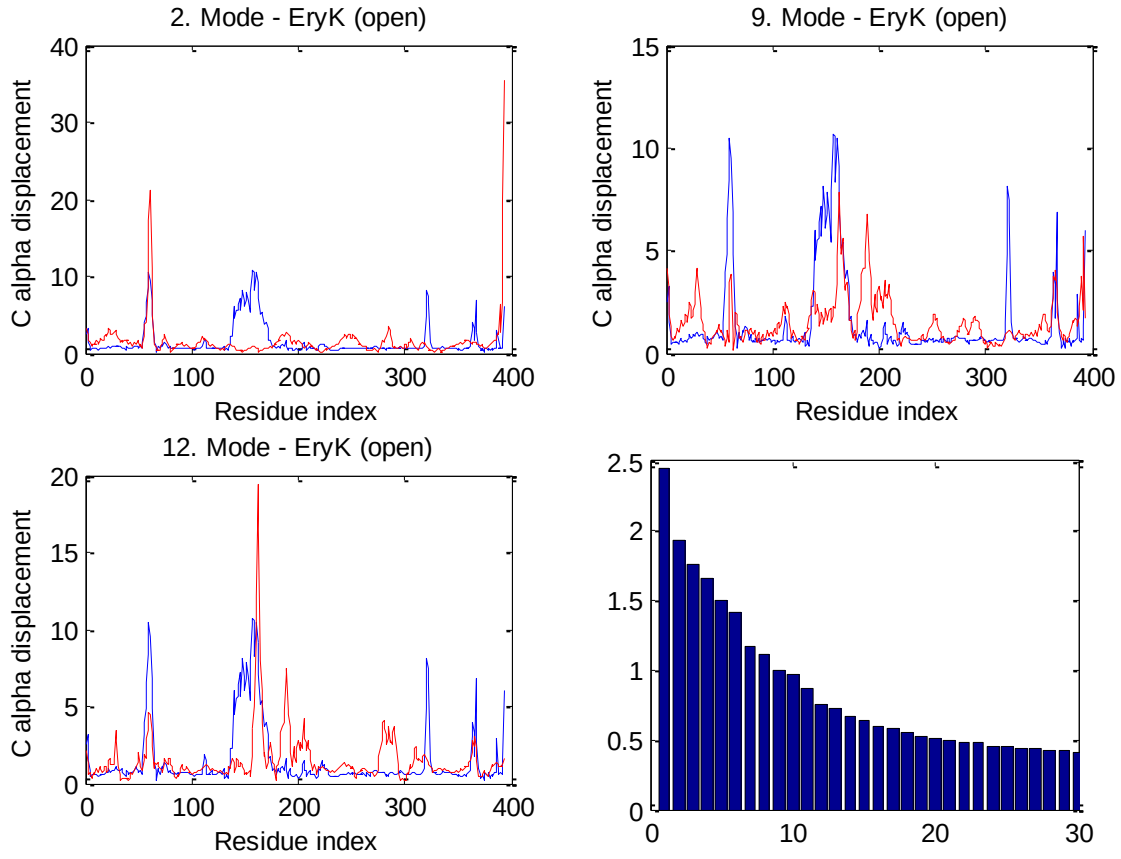
DHFR (DHF binding form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
13	0,43	0,19	4,33E-04	0,01
21	0,4	0,13	5,51E-04	0,008
24	0,46	0,24	5,95E-04	0,007

Table A.2. 16 Mode Graphs and Contribution Spectrum of first 30 modes for DHFR (DHF binding form)

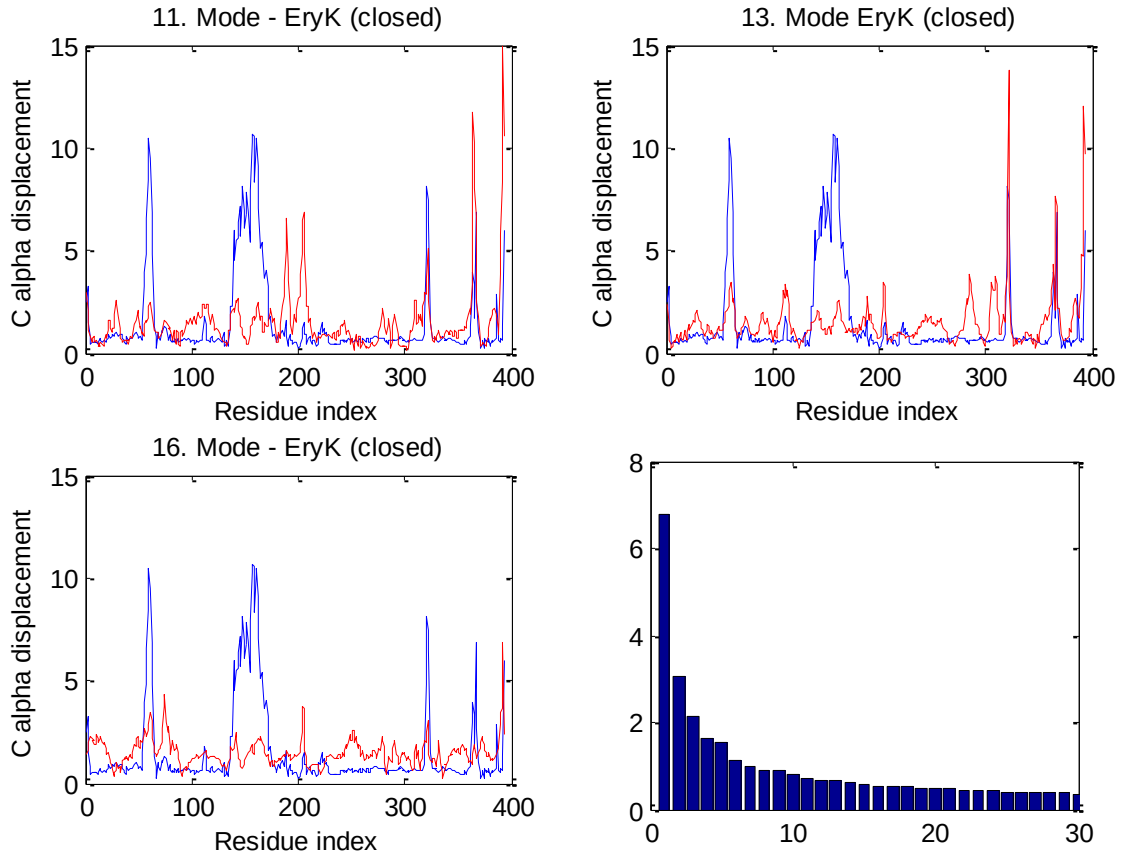
EryK cytochrome P450 (open form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
2	0.23	0.23	1,11E-04	0,019
9	0.36	0.27	2,13E-04	0,01
12	0.47	0.32	2,81E-04	0,0075

Table A.2. 17 Mode Graphs and Contribution Spectrum of first 30 modes for Cytochrome P450 (open form)

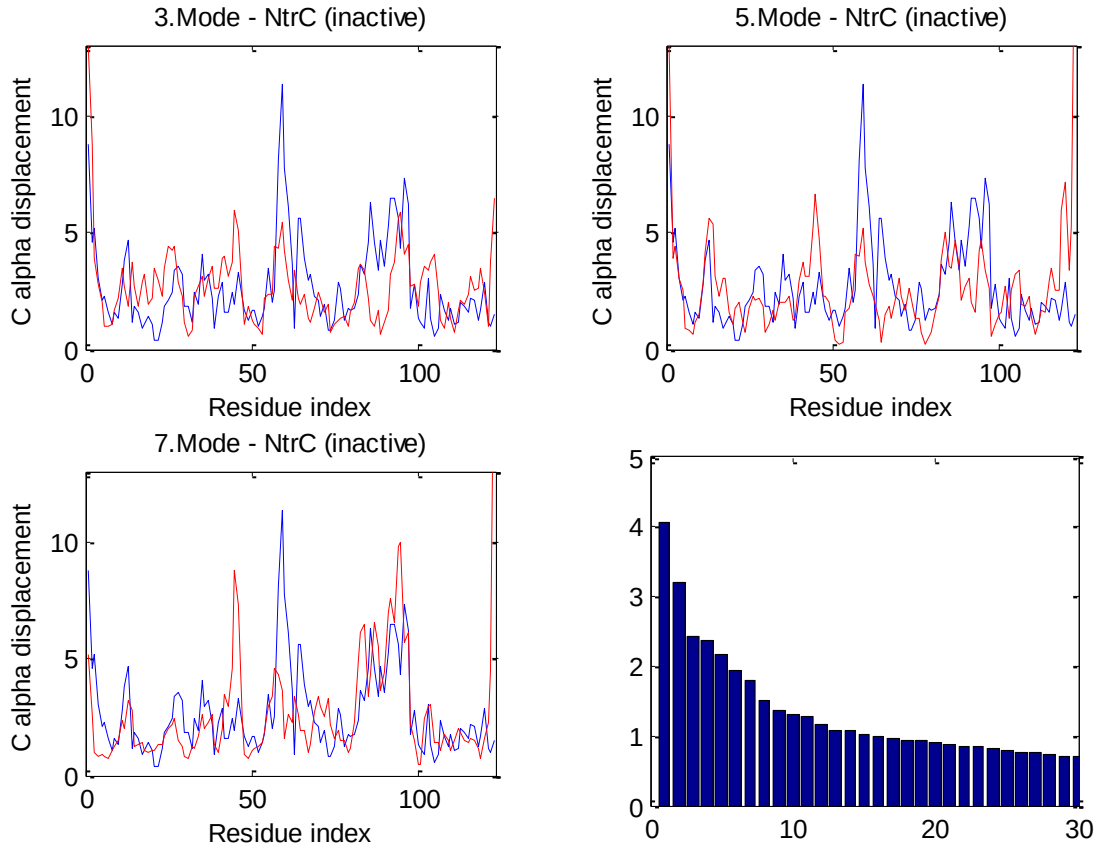
EryK cytochrome P450 (closed, non-binding form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
11	0,21	0,08	2,8E-04	0,007
13	0,30	0,10	3,04E-04	0,0066
16	0,22	0,12	3,69E-04	0,005

Table A.2. 18 Mode Graphs and Contribution Spectrum of first 30 modes for Cytochrome P450 (closed form)

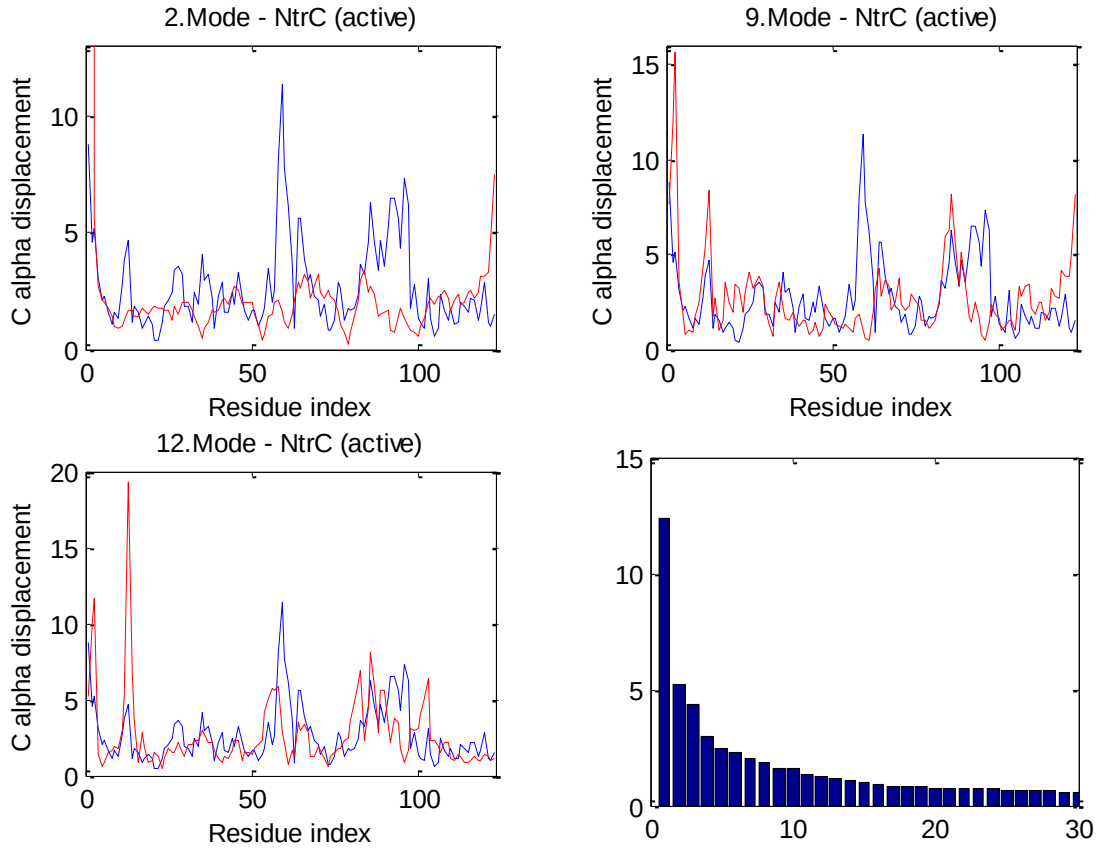
NtrC (inactive form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
3	0,42	0,24	2,80E-04	0,024
5	0,23	0,16	3,10E-04	0,022
7	0,35	0,13	3,70E-04	0,018

Table A.2. 19 Mode Graphs and Contribution Spectrum of first 30 modes for NtrC (inactive form)

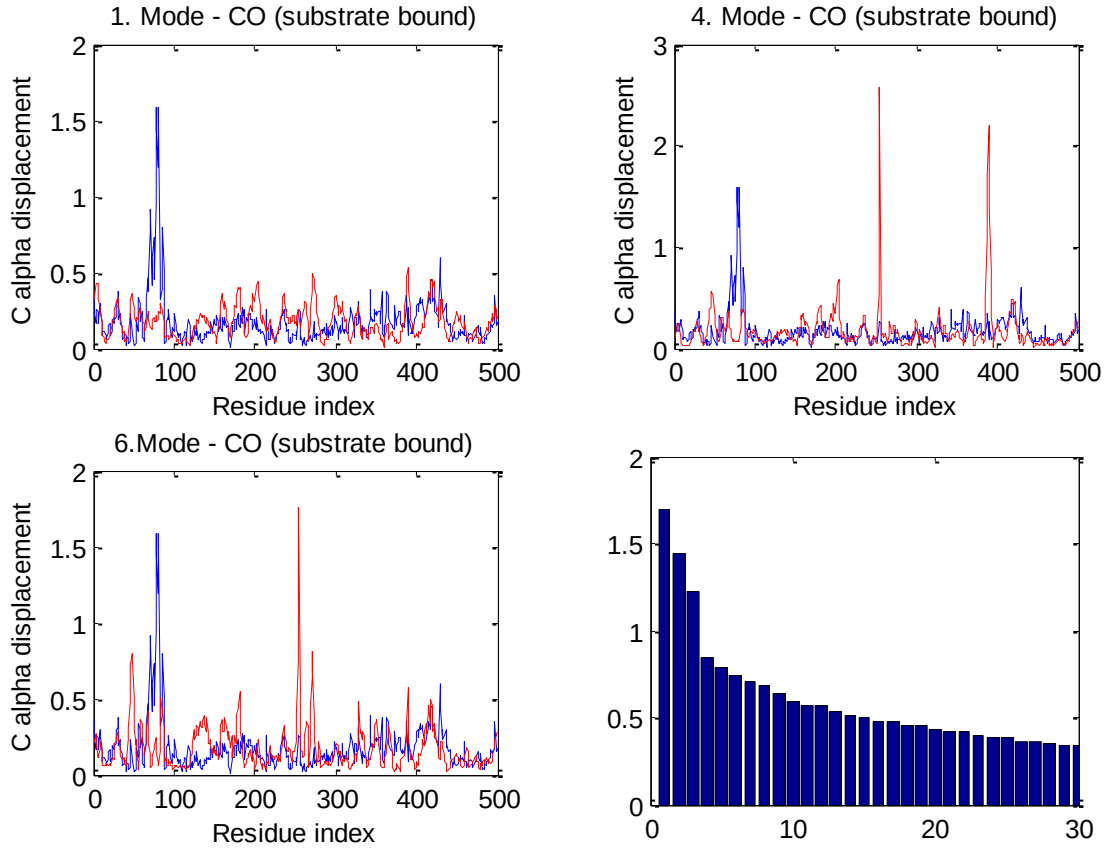
NtrC (active form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
2	0,29	0,21	9,46E-05	0,052
9	0,22	0,16	3,20E-04	0,016
12	0,35	0,19	4,00E-04	0,012

Table A.2. 20 Mode Graphs and Contribution Spectrum of first 30 modes for NtrC (active form)

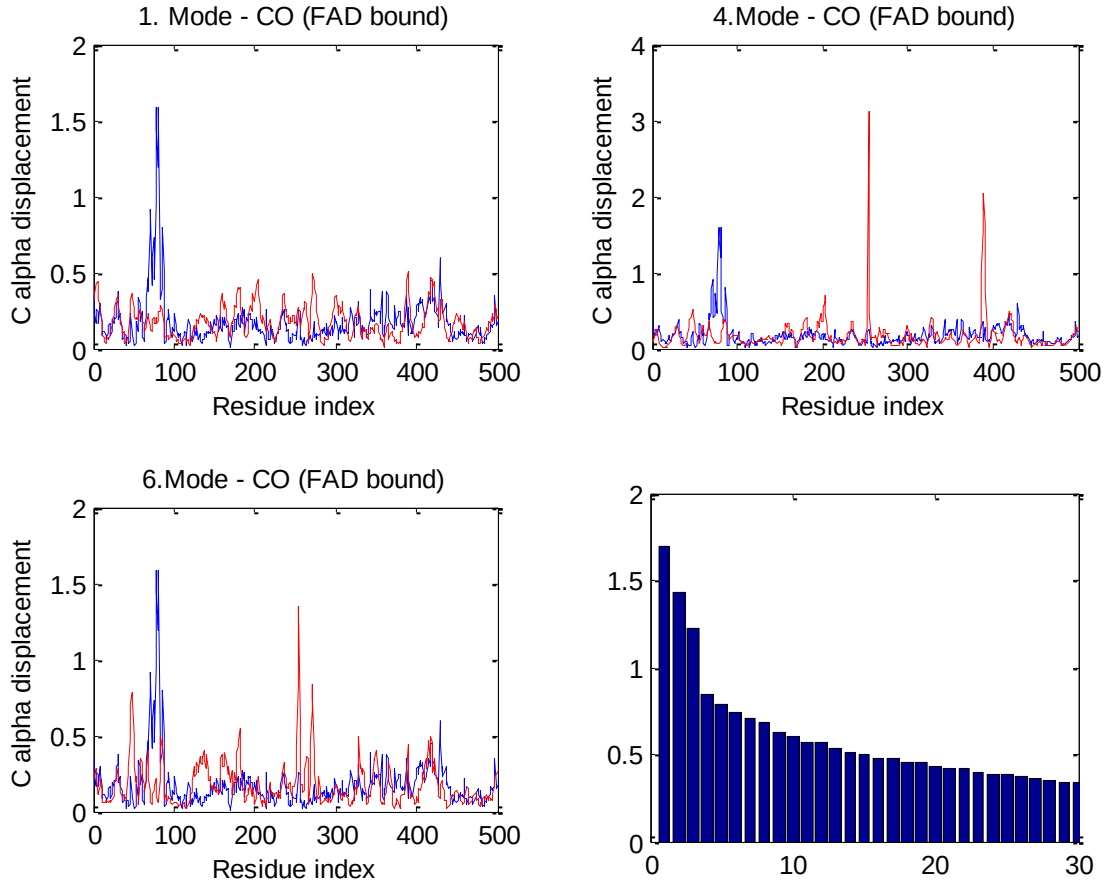
Cholesterol Oxidase (substrate bound form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
1	0,18	0,14	1,23E-04	0,017
4	0,11	0,18	2,48E-04	0,008
6	0,09	0,2	2,82E-04	0,007

Table A.2. 21 Mode Graphs and Contribution Spectrum of first 30 modes for Cholesterol Oxidase (substrate bound form)

Cholesterol Oxidase (FAD bound form)



Mode Indice	Correlation	Overlap	Eigenvalue	W.Eigenvalue
1	0,18	0,14	1,24E-04	0,017
4	0,11	0,17	2,49E-04	0,008
6	0,07	0,21	2,83E-04	0,007

Table A.2. 22 Mode Graphs and Contribution Spectrum of first 30 modes for Cholesterol Oxidase (FAD bound form)

APPENDIX 3

Random Mode Results

Similarity calculations of random mode vectors have been supplemented here. For each protein, the highest three mode correlation and overlap scores are given in Table A.3.1.

Table 2. Performance of Random Modes

Protein	Mode Correlation			Overlap		
Adenylate Kinase	0.295	0.283	0.237	0.130	0.127	0.123
Spe7	0.296	0.293	0.288	0.262	0.167	0.163
TIM	0.202	0.193	0.188	0.094	0.091	0.089
RNase	0.350	0.312	0.293	0.192	0.182	0.164
NtrC	0.327	0.325	0.311	0.179	0.171	0.154
DHFR	0.249	0.235	0.227	0.105	0.100	0.099
MBP	0.189	0.186	0.169	0.140	0.116	0.102
Thrombin	0.223	0.216	0.201	0.117	0.116	0.115
Cholesterol Oxidase	0.151	0.145	0.141	0.077	0.075	0.073
FAB Antibody	0.231	0.216	0.215	0.119	0.098	0.098
EryK P450 Cytochrome	0.184	0.174	0.158	0.107	0.104	0.102

Table A.3 1 Performance of Random Modes

Bibliography

1. Scheraga, H.A., M. Khalili, and A. Liwo, *Protein-folding dynamics: overview of molecular simulation techniques.* Annu Rev Phys Chem, 2007. **58**: p. 57-83.
2. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nat Struct Biol, 2002. **9**(9): p. 646-52.
3. Brooks, B. and M. Karplus, *Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor.* Proc Natl Acad Sci U S A, 1983. **80**(21): p. 6571-5.
4. Go, N., T. Noguti, and T. Nishikawa, *Dynamics of a small globular protein in terms of low-frequency vibrational modes.* Proc Natl Acad Sci U S A, 1983. **80**(12): p. 3696-700.
5. Flory, P.G., M; McCrum, N, *Statistical Thermodynamics of Random Networks [and Discussion].* Proceedings of the Royal Society A: Mathemical, Physical and Engineering Sciences, 1976. **351**(1666): p. 351-380.
6. Bahar, I., A.R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential.* Fold Des, 1997. **2**(3): p. 173-81.
7. Doruker, P., A.R. Atilgan, and I. Bahar, *Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor.* Proteins, 2000. **40**(3): p. 512-24.
8. Haliloglu, T.B., I & Jernigan, RL, *Collective dynamics of class I MHC/peptide complexes and binding to T cell receptors.* Biophys J, 2000. **78**(34A).
9. Atilgan, A.R., et al., *Anisotropy of fluctuation dynamics of proteins with an elastic network model.* Biophys J, 2001. **80**(1): p. 505-15.
10. Yang, L., G. Song, and R.L. Jernigan, *Protein elastic network models and the ranges of cooperativity.* Proc Natl Acad Sci U S A, 2009. **106**(30): p. 12347-52.
11. Tirion, M.M., *Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis.* Phys Rev Lett, 1996. **77**(9): p. 1905-1908.
12. Tobi, D. and I. Bahar, *Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state.* Proc Natl Acad Sci U S A, 2005. **102**(52): p. 18908-13.
13. Keskin, O., *Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies.* BMC Struct Biol, 2007. **7**: p. 31.
14. Bakan, A. and I. Bahar, *The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding.* Proc Natl Acad Sci U S A, 2009. **106**(34): p. 14349-54.
15. Wolf-Watz, M., et al., *Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair.* Nat Struct Mol Biol, 2004. **11**(10): p. 945-9.

16. James, L.C. and D.S. Tawfik, *Structure and kinetics of a transient antibody binding intermediate reveal a kinetic discrimination mechanism in antigen recognition.* Proc Natl Acad Sci U S A, 2005. **102**(36): p. 12730-5.
17. Kay, L.E., *NMR studies of protein structure and dynamics.* J Magn Reson, 2005. **173**(2): p. 193-207.
18. Henzler-Wildman, K.A., et al., *Intrinsic motions along an enzymatic reaction trajectory.* Nature, 2007. **450**(7171): p. 838-44.
19. Liu, L., et al., *A comparative analysis of the equilibrium dynamics of a designed protein inferred from NMR, X-ray, and computations.* Proteins, 2009. **77**(4): p. 927-39.
20. Tama, F. and Y.H. Sanejouand, *Conformational change of proteins arising from normal mode calculations.* Protein Eng, 2001. **14**(1): p. 1-6.
21. Dobbins, S.E., V.I. Lesk, and M.J. Sternberg, *Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking.* Proc Natl Acad Sci U S A, 2008. **105**(30): p. 10390-5.
22. Eisenmesser, E.Z., et al., *Intrinsic dynamics of an enzyme underlies catalysis.* Nature, 2005. **438**(7064): p. 117-21.
23. Kern, D., E.Z. Eisenmesser, and M. Wolf-Watz, *Enzyme dynamics during catalysis measured by NMR spectroscopy.* Methods Enzymol, 2005. **394**: p. 507-24.
24. Lerch, H.P., R. Rigler, and A.S. Mikhailov, *Functional conformational motions in the turnover cycle of cholesterol oxidase.* Proc Natl Acad Sci U S A, 2005. **102**(31): p. 10807-12.
25. Boehr, D.D., et al., *The dynamic energy landscape of dihydrofolate reductase catalysis.* Science, 2006. **313**(5793): p. 1638-42.
26. Savino, C., et al., *Investigating the structural plasticity of a cytochrome P450: three-dimensional structures of P450 EryK and binding to its physiological substrate.* J Biol Chem, 2009. **284**(42): p. 29170-9.
27. Leopold, P.E., M. Montal, and J.N. Onuchic, *Protein folding funnels: a kinetic approach to the sequence-structure relationship.* Proc Natl Acad Sci U S A, 1992. **89**(18): p. 8721-5.
28. Wolynes, P.G., *Recent successes of the energy landscape theory of protein folding and function.* Q Rev Biophys, 2005. **38**(4): p. 405-10.
29. Levinthal, C., *Are there pathways for protein folding?* J. Chim. Phys., 1968. **65**: p. 44-45.
30. Dill, K.A., *Theory for the folding and stability of globular proteins.* Biochemistry, 1985. **24**(6): p. 1501-9.
31. Austin, R.H., Beeson, K W, Eisenstein, L, Frauenfelder, H, Gunsalus, I C, *Dynamics of ligand binding to myoglobin.* Biochemistry, 1975. **14**(24): p. 5355-5373.
32. Fenimore, P.W., Frauenfelder, H, McMahon, B H, Parak, F G, *Slaving: solvent fluctuations dominate protein dynamics and functions.* Proceedings of the

National Academy of Sciences of the United States of America, 2002. **99**(25): p. 16047-16051.
33. Dill, K.A. and H.S. Chan, *From Levinthal to pathways to funnels.* Nat Struct Biol, 1997. **4**(1): p. 10-9.
34. Fischer, E., *Einfluss der Configuration auf die Wirkung der Enzyme.* Berichte der deutschen chemischen Gesellschaft, 1894. **27**(3): p. 2985-2993.
35. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis.* Proc Natl Acad Sci U S A, 1958. **44**(2): p. 98-104.
36. Boehr, D.D., R. Nussinov, and P.E. Wright, *The role of dynamic conformational ensembles in biomolecular recognition.* Nat Chem Biol, 2009. **5**(11): p. 789-96.
37. Zhang, Z., et al., *Single-molecule and transient kinetics investigation of the interaction of dihydrofolate reductase with NADPH and dihydrofolate.* Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2764-9.
38. Bah, A., et al., *Rapid kinetics of Na+ binding to thrombin.* J Biol Chem, 2006. **281**(52): p. 40049-56.
39. Massi, F., C. Wang, and A.G. Palmer, 3rd, *Solution NMR and computer simulation studies of active site loop motion in triosephosphate isomerase.* Biochemistry, 2006. **45**(36): p. 10787-94.
40. Aden, J. and M. Wolf-Watz, *NMR identification of transient complexes critical to adenylate kinase catalysis.* J Am Chem Soc, 2007. **129**(45): p. 14003-12.
41. Tang, C., C.D. Schwieters, and G.M. Clore, *Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR.* Nature, 2007. **449**(7165): p. 1078-82.
42. Debler, E.W., et al., *Conformational isomerism can limit antibody catalysis.* J Biol Chem, 2008. **283**(24): p. 16554-60.
43. James, L.C. and D.S. Tawfik, *Conformational diversity and protein evolution--a 60-year-old hypothesis revisited.* Trends Biochem Sci, 2003. **28**(7): p. 361-8.
44. Volkman, B.F., et al., *Two-state allosteric behavior in a single-domain signaling protein.* Science, 2001. **291**(5512): p. 2429-33.
45. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins.* Nature, 2007. **450**(7172): p. 964-72.
46. Palmer, A.G., 3rd, *NMR characterization of the dynamics of biomacromolecules.* Chem Rev, 2004. **104**(8): p. 3623-40.
47. Wintrode, P.L. and P.L. Privalov, *Energetics of target peptide recognition by calmodulin: a calorimetric study.* J Mol Biol, 1997. **266**(5): p. 1050-62.
48. Brokx, R.D., et al., *Energetics of target peptide binding by calmodulin reveals different modes of binding.* J Biol Chem, 2001. **276**(17): p. 14083-91.
49. Frederick, K.K., et al., *Conformational entropy in molecular recognition by proteins.* Nature, 2007. **448**(7151): p. 325-9.
50. Lumry, R., *Protein substructures and folded stability.* Biophys Chem, 2002. **101-102**: p. 81-92.

51. Frauenfelder, H.G., A. Petsko; Tsernoglou, Demetrius, *Temperature-dependent X-ray diffraction as a probe of protein structural dynamics.* Nature, 1979. **280**: p. 558-563.
52. Merritt, E.A., *Expanding the model: anisotropic displacement parameters in protein structure refinement.* Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 6): p. 1109-17.
53. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins.* Chem Rev, 2006. **106**(5): p. 1589-615.
54. Mulder, F.A., et al., *Studying excited states of proteins by NMR spectroscopy.* Nat Struct Biol, 2001. **8**(11): p. 932-5.
55. Trzesniak, D. and W.F. van Gunsteren, *Catalytic mechanism of cyclophilin as observed in molecular dynamics simulations: pathway prediction and reconciliation of X-ray crystallographic and NMR solution data.* Protein Sci, 2006. **15**(11): p. 2544-51.
56. Loria, J.P., R.B. Berlow, and E.D. Watt, *Characterization of enzyme motions by solution NMR relaxation dispersion.* Acc Chem Res, 2008. **41**(2): p. 214-21.
57. Ha, T., *Single-molecule fluorescence resonance energy transfer.* Methods, 2001. **25**(1): p. 78-86.
58. Diez, M., et al., *Proton-powered subunit rotation in single membrane-bound F0F1-ATP synthase.* Nat Struct Mol Biol, 2004. **11**(2): p. 135-41.
59. Karplus, M. and J.N. Kushick, *Method for estimating the configurational entropy of macromolecules.* Macromolecules, 1981. **14**(2): p. 325-332.
60. Ma, J. and M. Karplus, *Ligand-induced conformational changes in ras p21: a normal mode and energy minimization analysis.* J Mol Biol, 1997. **274**(1): p. 114-31.
61. Schlitter, J., et al., *Targeted molecular dynamics simulation of conformational change - application to the T-R transition in insulin.* Mol.Simul., 1993. **10**: p. 291-308.
62. Lange, O.F., et al., *Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution.* Science, 2008. **320**(5882): p. 1471-5.
63. Ensign, D.L., P.M. Kasson, and V.S. Pande, *Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece.* J Mol Biol, 2007. **374**(3): p. 806-16.
64. Freddolino, P.L., et al., *Ten-microsecond molecular dynamics simulation of a fast-folding WW domain.* Biophys J, 2008. **94**(10): p. L75-7.
65. Best, R.B., N.V. Buchete, and G. Hummer, *Are current molecular dynamics force fields too helical?* Biophys J, 2008. **95**(1): p. L07-9.
66. Freddolino, P.L., et al., *Force field bias in protein folding simulations.* Biophys J, 2009. **96**(9): p. 3772-80.
67. Shea, J.E. and C.L. Brooks, 3rd, *From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding.* Annu Rev Phys Chem, 2001. **52**: p. 499-535.

68. Torrie, G.M. and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling.* Journal of Computational Physics, 1977. **23**(2): p. 187-199.
69. Pearson, D.S., *Scattered intensity from a chain in a rubber network.* Macromolecules, 1977. **10**(3): p. 696-701.
70. Kloczkowski, A., J. Mark, and B. Erman, *Chain dimensions and fluctuations in random elastomeric networks. 1. Phantom Gaussian networks in the undeformed state.* Macromolecules, 1989. **22**(3): p. 1423–1432.
71. Yang, L., G. Song, and R.L. Jernigan, *How well can we understand large-scale protein motions using normal modes of elastic network models?* Biophys J, 2007. **93**(3): p. 920-9.
72. Lin, C.P., et al., *Deriving protein dynamical properties from weighted protein contact number.* Proteins, 2008. **72**(3): p. 929-35.
73. Cao, Z.W., et al., *MoViES: molecular vibrations evaluation server for analysis of fluctuational dynamics of proteins and nucleic acids.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W679-85.
74. Lindahl, E., et al., *NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W52-6.
75. Wako, H., M. Kato, and S. Endo, *ProMode: a database of normal mode analyses on protein molecules with a full-atom model.* Bioinformatics, 2004. **20**(13): p. 2035-43.
76. Hollup, S.M., G. Salensminde, and N. Reuter, *WEBnm@: a web application for normal mode analyses of proteins.* BMC Bioinformatics, 2005. **6**: p. 52.
77. Zheng, W. and S. Doniach, *A comparative study of motor-protein motions by using a simple elastic-network model.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13253-8.
78. Suhre, K. and Y.H. Sanejouand, *ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W610-4.
79. Yang, L.W., et al., *iGNM: a database of protein functional motions based on Gaussian Network Model.* Bioinformatics, 2005. **21**(13): p. 2978-87.
80. Yang, L.W., et al., *oGNM: online computation of structural dynamics using the Gaussian Network Model.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W24-31.
81. Eyal, E., L.W. Yang, and I. Bahar, *Anisotropic network model: systematic evaluation and a new web interface.* Bioinformatics, 2006. **22**(21): p. 2619-27.
82. Noguchi, T. and Y. Akiyama, *PDB-REPRDB: a database of representative protein chains from the Protein Data Bank (PDB) in 2003.* Nucleic Acids Res, 2003. **31**(1): p. 492-3.
83. Beach, H., et al., *Conservation of mus-ms enzyme motions in the apo- and substrate-mimicked state.* J Am Chem Soc, 2005. **127**(25): p. 9167-76.

84. Vonrhein, C., G.J. Schlauderer, and G.E. Schulz, *Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases.* Structure, 1995. **3**(5): p. 483-90.
85. Henzler-Wildman, K.A., et al., *A hierarchy of timescales in protein dynamics is linked to enzyme catalysis.* Nature, 2007. **450**(7171): p. 913-6.
86. Lu, Q. and J. Wang, *Single molecule conformational dynamics of adenylate kinase: energy landscape, structural correlations, and transition state ensembles.* J Am Chem Soc, 2008. **130**(14): p. 4772-83.
87. Cukier, R.I., *Apo adenylate kinase encodes its holo form: a principal component and varimax analysis.* J Phys Chem B, 2009. **113**(6): p. 1662-72.
88. Hammes, G.G., *Multiple conformational changes in enzyme catalysis.* Biochemistry, 2002. **41**(26): p. 8221-8.
89. Schnell, J.R., H.J. Dyson, and P.E. Wright, *Structure, dynamics, and catalytic function of dihydrofolate reductase.* Annu Rev Biophys Biomol Struct, 2004. **33**: p. 119-40.
90. Fierke, C.A., K.A. Johnson, and S.J. Benkovic, *Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from Escherichia coli.* Biochemistry, 1987. **26**(13): p. 4085-92.
91. Tam, R. and M.H. Saier, Jr., *Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria.* Microbiol Rev, 1993. **57**(2): p. 320-46.
92. Spurlino, J.C., G.Y. Lu, and F.A. Quiocho, *The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis.* J Biol Chem, 1991. **266**(8): p. 5202-19.
93. Sharff, A.J., et al., *Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis.* Biochemistry, 1992. **31**(44): p. 10657-63.
94. Quiocho, F.A. and J.C. Spurlino, *Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor.* Structure, 1997. **5**(8): p. 997-1015.
95. Werck-Reichhart, D. and R. Feyereisen, *Cytochromes P450: a success story.* Genome Biol, 2000. **1**(6): p. REVIEWS3003.
96. Lewis, D.F., S. Modi, and M. Dickins, *Structure-activity relationship for human cytochrome P450 substrates and inhibitors.* Drug Metab Rev, 2002. **34**(1-2): p. 69-82.
97. Lamb, D.C., et al., *Cytochromes P450 and drug discovery.* Curr Opin Biotechnol, 2007. **18**(6): p. 504-12.
98. Xiang, J., J.Y. Jung, and N.S. Sampson, *Entropy effects on protein hinges: the reaction catalyzed by triosephosphate isomerase.* Biochemistry, 2004. **43**(36): p. 11436-45.
99. Bode, W., *Structure and interaction modes of thrombin.* Blood Cells Mol Dis, 2006. **36**(2): p. 122-30.

100. De Filippis, V., et al., *Effect of Na+ binding on the conformation, stability and molecular recognition properties of thrombin.* Biochem J, 2005. **390**(Pt 2): p. 485-92.
101. Di Cera, E., Q.D. Dang, and A. Y.M., *Molecular mechanisms of thrombin function.* Cell Mol Life Sci., 1997. **53**(9): p. 701-730.
102. Debler, E.W., et al., *Structural origins of efficient proton abstraction from carbon by a catalytic antibody.* Proc Natl Acad Sci U S A, 2005. **102**(14): p. 4984-9.
103. Vrielink, A., L.F. Lloyd, and D.M. Blow, *Crystal structure of cholesterol oxidase from Brevibacterium sterolicum refined at 1.8 A resolution.* J Mol Biol, 1991. **219**(3): p. 533-54.
104. Lu, H.P., L. Xun, and X.S. Xie, *Single-molecule enzymatic dynamics.* Science, 1998. **282**(5395): p. 1877-82.
105. Hu, X. and Y. Wang, *Molecular dynamic simulations of the N-terminal receiver domain of NtrC reveal intrinsic conformational flexibility in the inactive state.* J Biomol Struct Dyn, 2006. **23**(5): p. 509-18.
106. Kundu, S., et al., *Dynamics of proteins in crystals: comparison of experiment with simple models.* Biophys J, 2002. **83**(2): p. 723-32.

VITA

Beytullah Özgür was born in Razgrad (Bulgaria) in 1986. His family later moved to Turkey and he had been to Kagithane Anadolu Lisesi for his high school education. In 2004 he had started his undergraduate study in İTÜ in the department of Molecular Biology and Genetics. He had graduated from İTÜ in 2008 and has worked at Koc University as a research and teaching assistant during 2008-2010. He pursued his MSc. career on Computational Sciences and Engineering.

contact e-mail: bozgur@ku.edu.tr