MODAL ANALYSIS OF MYOSIN II AND IDENTIFICATION OF FUNCTIONALLY IMPORTANT SITES

by

Onur Varol

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

Master of Science

in

Computer Science & Engineering

Koç University

June, 2012

Koç University

Graduate School of Sciences and Engineering

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

Onur Varol

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:

Asst. Prof. Dr. Deniz Yuret

Asst. Prof. Dr. Alkan Kabakçıoğlu

Prof. Dr. Burak Erman

Date: _____

To my family

ABSTRACT

Analysis of protein dynamics uses structural and fluctuation based methods. Fluctuation analysis of protein dynamics has proven to be a rewarding venue of research. Mass and spring models are used in previous research commonly. However, fluctuations of this models are based on purely harmonic which has significant gap between the experimental results. Deviations from harmonicity mostly observe in slow, collective modes. Corrections like anharmonic modal decomposition are first step in order to minimize this gap. The contribution of the higher-order corrections is limited because of the interacting modes. Mode-coupling corrections which yield valuable information on means of energy transfer and allostery.

In this work, molecular dynamic results of *Dictyostelium discoideum* myosin II motor domain is used as test ground. Mode fluctuation distributions produced using MD results, fully harmonic models and a model with anharmonic corrections. Tensorial hermite polynomials are used in order to obtain distributions of modal fluctuations. Fluctuations on modal space are transformed back into real space and distribution of residual fluctuations is compared using KL divergence. Analysis results for ligand-bound and free myosin dynamics are used in order to demonstrate that the mode-coupling contributions alone highlight functionally important sites.

ÖZETÇE

Protein dinamiği analizlerinde yapısal ve dalgalanma özellikleri kullanan metodlar uygulanmaktadır. Dalgalanma analizi çalışmalarının sonuçları buradan elde edilecek bilgilerin ümit verici olduğunu göstermektedir. Bundan önceki çalışmalarda kütle ve yay sistemleri sıklıkla kullanılmaktadır. Ancak dalgalanma analizlerinde kullanılan modellerde sistemdeki titreşimlerim tamamen harmonik olmaları nedeniyle deneysel verilerle arasında önemli bir fark oluşmaktadır. Harmonik davranıştan farklılıklar daha çok yavaş ve kollektif modlarda meydana gelmektedir. Anharmonik mod ayrışımı yöntemi bu farkı kapatabilecek düzeltmelerin ilk aşamasıdır. Daha yüksek mertebeden düzeltmeler modların kendi aralarında olan etkileşimlerden dolayı limitlidir. Mode eşlenimi düzeltmeleri ise enerji transferi ve allosterik mekanizmaya yönelik önemli bilgiler içermektedir.

Bu çalışmada, moleküler dinamik sonuçları *Dictyostelium discoideum* myosin II motor yapısını bir test ortamı olarak kabul edip incelenmiştir. Mod dalgalanma dağılımları MD kullanılarak oluşturulup, tamamen harmonik olan sistemlerin üzerine anharmonik düzeltmeler yapılması amacıyla kullanılmıştır. Hermite tensör polinomları ise mod dalgalanma dağılımlarının elde edilmesinde kullanılmıştır. Mod uzayındaki dalgalanmalar daha sonrasında gerçek uzay koordinatlarına dönüştürülüp KL metriği ile karşılaştırılmıştır. Ligand bağlı ve serbest protein durumlarına ait sonuçlar mod çiftlenmi etkisinin tekbaşına proteinin işlevsel olarak önemli kısımlarını gösterdiği görülmüştür.

ACKNOWLEDGMENTS

Scientific studies researches require interdisciplinary studies, which are ongoing simultaneously in different areas. Scientist needs to collaborate with colleagues from different disciplines to exhibit efficient results and to analyse complex systems in nature. I would like to thank my advisor Asst. Prof. Dr. Deniz Yuret and Asst. Prof. Dr. Alkan Kabakçıoğlu for providing me valuable research atmosphere and significant research questions. I would also want to thank Prof. Dr. Burak Erman for sharing his experience and providing substantial knowledge. It was a great pleasure for me to on this thesis under their supervision.

I would also want to thank my friends in Koç University. Nese Aral for her energetic mood and academic chats, my physicist friends Neslihan Oflaz and Duygu Can. Coffee is a miracle beverage while studying. Hüsnü Şensoy joined our group beginning of this year however I felt most close in our office and he became my friend in our coffee breaks. I also want to thank my office friends in AI lab.

I also want to state that I am thankful about being TUBITAK scholar during my master education. Research on molecular dynamic studies needs high computational resources and ITU UYHBM support my project and save valuable time for my research. Computing resources used in this work were provided by the National Center for High Performance Computing of Turkey (UYBHM) under grant number 4001752012.

At last, but not least, I especially want to thank my family who respected my decisions during my all life. I want to thank for my father Cüneyt Varol, my mother Berrin Varol and my sister Beril Varol for their continuous support and understanding during my education. I especially want to thank my fiancé Nur Mustafaoğlu who motivated me all the time.

Finally, I thank to all reputable scientists who prepare the whole infrastructure to contribute for development and accelerate of present scientific studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

GNM	:	Gaussian Network Model
ENM	:	Elastic Network Model
ANM	:	Anisotropic Network Model
C_α	:	Carbon alpha atom
PDB	:	Protein Data Bank
ATP	:	Adenosine triphosphate
ADP	:	Adenosine diphosphate
P_i	:	Inorganic phosphate
NMR	:	Nuclear Magnetic Resonance
MD	:	Molecular dynamic
$\mathbf{\Delta R}$	:	Displacement vectors
Δr	:	Modal space displacement
Γ	:	Covariance matrix
f_0	:	Harmonic model
f_1	:	Model with anharmonic corrections
KL	:	Kullback-Leibler
KDE	:	Kernel density estimation

Chapter 1

INTRODUCTION

Proteins are the essential structures assembled from the amino acids which contains information about functionalities encoded in genes. Proteins performs different task in cell like signalling, transport or structural support. Many proteins are enzymes that catalyse biochemical reactions vital for organism. Proteins perform their tasks by working in coordination with other macromolecules. In allosteric regulation, molecule binding a protein at one site induces conformational change in protein structure. This transform creates a significant changes on distant sites. Challenges in this problem are determination of the active sites and the pathway of energy transfer to the distant sites.

Myosin is one of the allosteric protein which has complex cycle of states. Well accepted model for myosin kinetic cycle is proposed by Lymn and Taylor [41]. Allosteric effect on myosin reveal when hydrolysis of ATP and release of products occurs. Structural changes on lever arm generates force for movement, which is essential for living organisms. Hydrolysis and product release on ligand-binding pocket bring changes on distant sites of the protein. This coordinated mechanism also cause acting binding on the other site of the myosin head. In this thesis, myosin structure is chosen because of its allosteric mechanism and experimental and computational evidence about its functions.

Fluctuation analysis of protein is promising technique in order to obtain valuable insight about functional role of structural elements, prediction for the functional sites and residues. Previous methods also perform successful intuition about protein functionalities. These techniques based on mass-spring system models such as Gaussian (GNM) [2, 23],elastic (ENM) [73], or anisotropic (ANM) [1, 3] network model. These models suggest that interactions between residues are harmonic motions [82] and ligand binding [14, 75] effects are not considered directly. However, molecular dynamics studies reveal that fluctuations of residues and modes are not harmonic. Such deviations from harmonicity are most observed

in slow modes and collective modes.

Modal space analysis contains information of the driving motions of the protein which are also essential for global mechanisms. However allosteric mechanisms cannot be explained using this trend. Allostery can be explained in terms of mode-coupling of residues not only by looking topological properties but also considering mode vibrational synchrony. Energy transfer between modes also studied for hemoglobin [35,45]. Energy transfer pathways also suggested using shortest paths and sequence of conserved residues [67,71]. There are also studies on multiple energy transfer pathways in which energy flows in alternative paths with different quantities [13]. Recent studies on network properties indicates the importance of clique residues [12,42,75]

Growing interest on modal analysis and energy transfer between residues motivate us in order to explain phenomenons like allostery and ligand binding. Anharmonic modal decomposition is the first correction on the fully harmonic system. There are also mode-coupling corrections which yield valuable information about energy transfer during allosteric transition and functionally important sites of protein. Identifying interaction between different modes believed to be the key to understanding allostery and possibly other mechanisms of the proteins inner dynamics. Systematic mathematical analysis of the fluctuation data was introduced in order to distinguish anharmonic and mode-coupling contributions to free energy [34].

In this work, introductory information about protein structures and the structural analysis techniques presented in Chapter 2. In order to understand myosin structures is essential to interpret results and point functionally important sites. Kinetic cycle of the myosin and allosteric mechanisms explained based on Lymn and Taylor [41] in Chapter 3. Allosteric mechanisms and proposed allosteric pathways in the literature given. Experimentally verified residues based on mutation experiment is also extracted from several sources. These experimental results provide information about essential structural mechanisms and functional sites.

Molecular dynamics simulation details and analysis of the trajectory outputs explained in Chapter 4. Configuration of the simulation environment and the parameters that are used during simulations explained. Structural stability is analysed using Ramachandran plots and ligand binding regions are also shown.

In Section 5 our approach in order to measuring residual contribution of mode-coupling and anharmonicty effects explained. Analysis based on mode-coupling contribution of residuals point that mode-coupling contributions alone highlight functional important sites of the myosin.

Chapter 2

PROTEIN STRUCTURE

Proteins evolve in nature in order to perform specific functions. The functional property of proteins depends on their structures and dynamical properties. Three dimensional structure of the proteins may arise different combinations of the amino acids in polypeptide chains. Folding mechanisms of proteins are still unsolved question and researchers tries to improve their knowledge on protein folding. Protein folding remains a problem because there are twenty different amino acids that can be combined countless different ways. However, there are also similar protein structures which has different amino acids sequence. Since three-dimensional structure of the proteins cannot be predicted successfully, these structures divided into small units starting from amino acids.

Protein structures assumed to perform its functional shapes and conformations. First protein structure organization is proposed by Kaj Ulrik Linderstrøm-Lang in 1951 [38]. These organization contains four levels: primary, secondary, tertiary and quaternary structures.

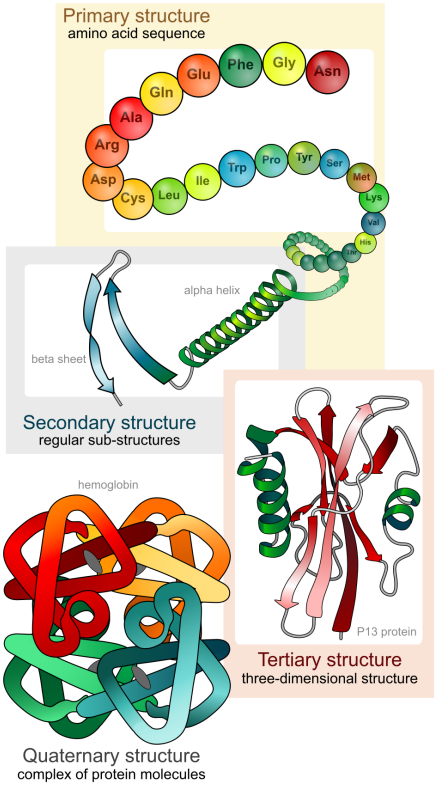


Figure 2.1: Four level of protein structure. [77]

Structural changes and other geometric aspect of protein is important to understand conformational changes. Measuring various positional properties of protein helps to get intuition about conformational changes. Distance between atoms and angles between residues used in various methods.

2.1 Interatomic distances and bond angles

Positions of the atoms are given by (x,y,z) coordinates and distance measured in Cartesian coordinates by Pythagorean calculation. If atom A has coordinates $a = (a_x, a_y, a_z)^T$ and atom B has coordinates $b = (b_x, b_y, b_z)^T$, the distance between A and B is given by

$$d(A,B) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2 + (a_z - b_z)^2} \tag{2.1}$$

Similar to this norm calculation can be formulated as

$$\|a - b\| = \sqrt{(a-b)^T(a-b)} \tag{2.2}$$

Inner product of the normalized vectors u and v can be viewed as cosine of the angle between these vectors. Angle between these vectors can be formulated as

$$\cos\theta = \frac{\langle u, v\rangle}{\|u\|\,\|v\|} \tag{2.3}$$

Considering atoms A and B connected with third atom called C. If vectors u and v defined as $u = a - c$ and $v = b - c$, cosine of the bond angle

$$\cos\theta = \frac{(a-c)^T(b-c)}{\|a-c\|\,\|b-c\|} \tag{2.4}$$

Once these cosine values of this angles are calculated, inverse cosine of this angle gives measurement of degree of bond angle.

2.1.1 Contact Map

A protein contact map represents the Euclidean distance between all pairs of residues. For the two residue i and j contact map element C_{ij} corresponding distance between C_α atoms. Threshold for $C_\alpha - C_\alpha$ distances are selected mostly between 6-12$\mathring{A}$. If distance C_{ij} less than this threshold value contact map representation is equal to 1, otherwise represented as 0.

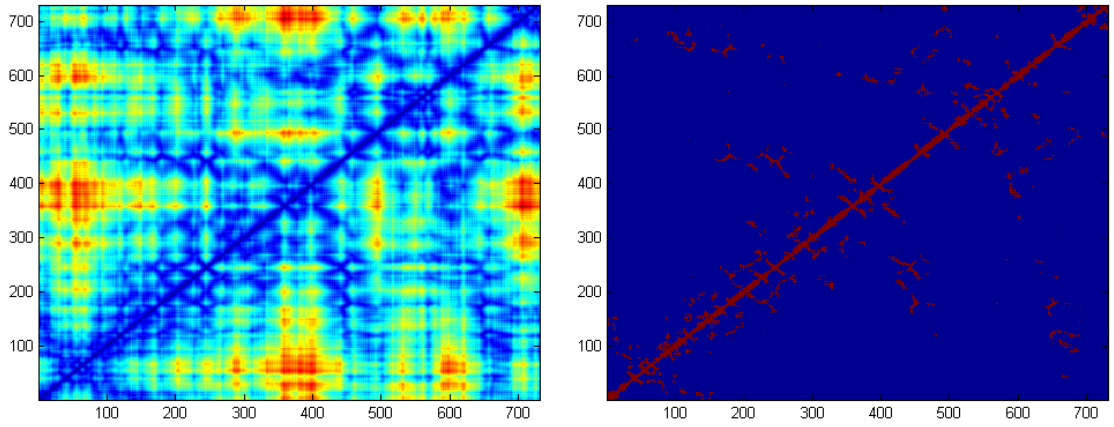


Figure 2.2: Difference of contact maps for binary version on the right and real distances shown on the left.

Contact map provide more reduced representation of a protein structure than it is full 3-D cartesian space coordinates.

2.2 Dihedral Angles

Atoms in protein backbones contains a chain of atoms with same sequence. Distances between this backbone atoms and bonds are almost equal. Consecutive atoms in the protein backbone is $N_{i-1}C_{\alpha_{i-1}}C_{i-1}N_iC_{\alpha_i}C_iN_{i+1}C_{\alpha_{i+1}}C_{i+1}$ also shown in Fig.2.3

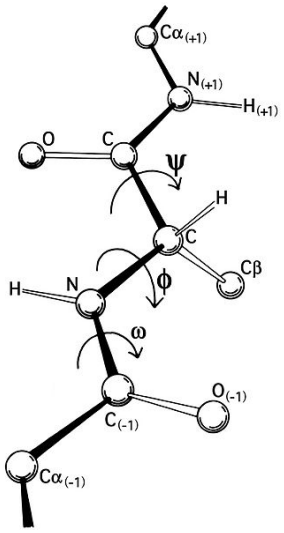


Figure 2.3: Protein backbone dihedral angles phi, psi, and omega [57]

Bond angles are determined by a sequence of three bonded atoms, but a dihedral angle is defined by a sequence of four consecutive bonded atoms. Three consecutive atom represent a plane and angles between these planes represent dihedral angles. Four consecutive atom $C_{i-1}N_iC_{\alpha_i}C_i$ defines two planes that intersect in the line that coincide bond between $N_iC_{\alpha_i}$. Angle between these two plane is referred to as the Phi (ϕ) angle associated with i^{th} residue. Similarly consecutive backbone atoms like $N_iC_{\alpha_i}C_iN_{i+1}$ defines two planes that intersect in the line that coincide bond between $C_{\alpha_i}C_i$ bond. The angle between these two plane is referred as the Psi (ψ) angle associated with i^{th} residue. Another possible angle between intersection of two plane with four consecutive atoms is Omega (ω). This angle defines by backbone atoms sequence $C_{\alpha_i}C_iN_{i+1}C_{\alpha_{i+1}}$.

2.2.1 Ramachandran Plot

Dihedral angles change with the conformations of proteins. However, atoms cannot freely move on backbone and there exist impossible conformations for atomic positions. Because of the collisions, atoms can locate some allowed intervals. An observed combinations of Psi and Phi angles ranging over values $[-\pi, \pi]$ is represented in a 2D plot. This type of scatter plot is called Ramachandran plot [54]. Ramachandran is the first person relate this scatter plot and allowed regions with amino acids types for C_α atoms in protein conformations.

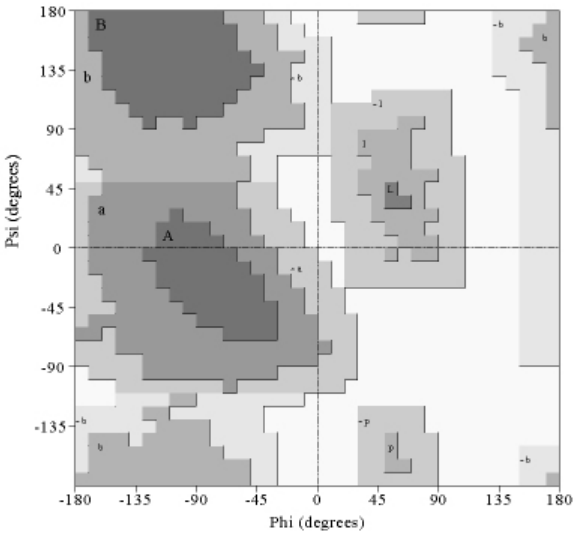


Figure 2.4: Ramachandran plot allowed regions. Different regions given with abbreviations. **A**-Core alpha, **a**-Allowed alpha, **a**-Generous alpha, **B**-Core beta, **b**-Allowed beta, **b**-Generous beta, **L**-Core left-handed alpha, **l**-Allowed left-handed alpha, **l**-Generous left-handed alpha, **p**-Allowed epsilon, **p**-Generous epsilon

Allowed regions of ramachandran plot shown in Fig.2.4. This figure generated by using dihedral angle values of more than 400 proteins and 120000 residues. The area originated from the scattered residues and subdivided into 36^2 regions. Each region is coloured by density of residues in these regions. Dark regions has highest density. Medium coloured regions characterized as allowed regions and lighter regions are called generous regions.

Secondary structure of the proteins can also be predicted using ramachandran plots. Ramachandran plots are also wrapped plots right to left and top to bottom because dihedral angle values are the same at 0° and 360°.

Ramachandran plots also used in machine learning methods to predict secondary structures and stability of the proteins [10, 72]. Amino acids apply probabilistic distribution on psi-phi space. Protein designs based on this statistics to select appropriate amino acid.

2.3 Amino acids and Primary Sequence

The primary structures of a protein is exact sequence of amino acids. All of the 20 amino acids have in common a central carbon atom (C_α) to which attached a hydrogen atom, an amino group (NH_2) and carboxyl group (COOH). Classification of the amino acids are

shown in Fig.2.5. Their names abbreviated with both three letter and one letter are shown. Sequence of genetic codes and hydrophobicity are also shown.

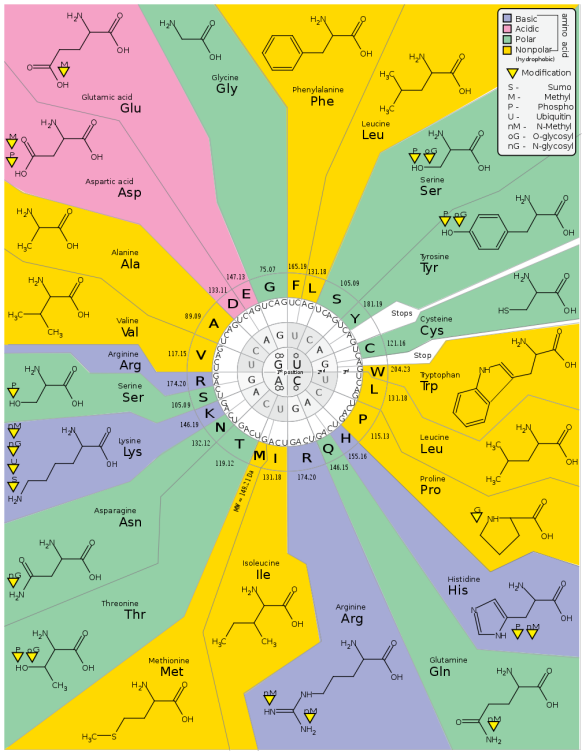


Figure 2.5: Amino acids are shown in table according to their structure, hydrophobicity, genetic codes and additional groups. Three letter codes and one letter abbreviations are also shown in this figure. [21]

Sequence of amino acids connected by peptide bonds generates a main chain or in other words backbone chain. Amino acids joined end to end during protein synthesis. Peptide bonds are formed while carboxyl group of one amino acid combines with the amino group of the next.

2.4 Secondary Structure

The secondary structure is assigned based on hydrogen bonding patterns. There exist three common secondary structures in proteins: α helices, β sheets and turns or bends. Some of the proteins cannot be assigned any of these structures and they are called random coils.

2.4.1 Alpha Helices

Helices are repetitive structure of α-carbons. Dihedral angles of these structures are same for each residue along the helix. Helices can be defined number of residue per helical turn (n), the rise per helical residue (d) and rise per full turn (p). Sign of n value corresponds to turn of helix.

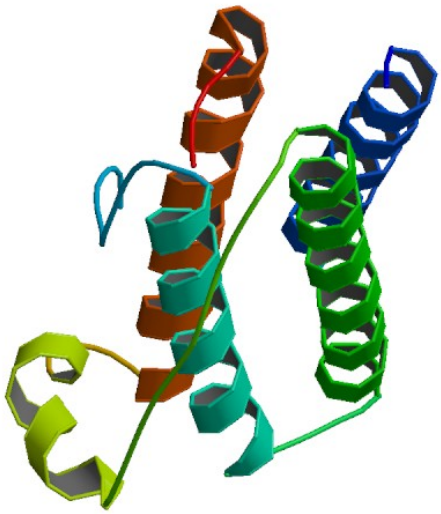


Figure 2.6: Human obesity protein Leptin contains 4 α-helices (PDB:1AX8) [81].

2.4.2 Beta Strand

Similar to alpha helices, beta strands are also secondary structures formed by hydrogen bondings. In this structure consecutive amino acids interact with another distant consecutive amino acids. In this distant interaction hydrogen bonds contribute to the structure and there can be two different orientation of the beta sheets: parallel and antiparallel.

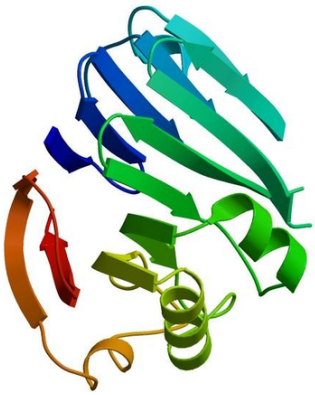


Figure 2.7: Structure of PDB:1VJN contains beta sheets

2.4.3 Loops

Another category of secondary structure is the loop. Loops are consecutive amino acids that have no hydrogen interaction with other parts of the protein. Proteins containing loops have stability problems and their conformation can stay stable by disulphide bridges.

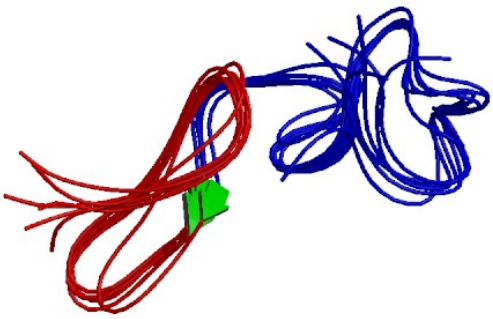


Figure 2.8: Structure of Flavoridin (PDB:1FVL)

Chapter 3

MYOSIN

Creatures can adapt themselves to dynamic environments and change their position and orientation in order to access foods or avoiding from the possible threats. Thus, movement is one of the most important property of living organisms. Various evolutionarily adaptation techniques are utilized by creatures for the necessity of accessing food. Even plants can change their position for their benefits and animals has muscular or more basic forms of the force generating systems.

There exist three major class of molecular motor proteins: myosin, dynein and kinosin [76]. Myosin is one of the molecule which is responsible from the movement. Myosin uses chemical energy to perform motion. Hydrolyzation of ATP produces energy to act stroke motion of myosin. Myosin plays essential roles in biological activities including muscle contraction, cell movements, membrane transport and several essential signalling pathways.

3.1 Myosin Kinetic Cycle

Kinetic model for myosin is first proposed in 1971 by Lymn and Taylor [41]. Mechanism of this process is shown Fig.3.1 as follow: ATP rapidly binds Actin-Myosin complex on ATPase site, then myosin hydrolyses ATP in the absence of Actin. The binding of ATP induces conformational change of myosin head and makes it dissociate from actin. After that, ADP and phosphate are released and then recombines with actin. This transition produces conformational change leading the stroke movement, which is called recovery stroke. ADP and inorganic phosphate are formed and remain bonded to the head. This conformational change produce recovery stroke. When the phosphate is released, conformational change in lever arm occurs, which is called power stroke. Following state ADP dissociates and chemical and mechanical cycle continues.

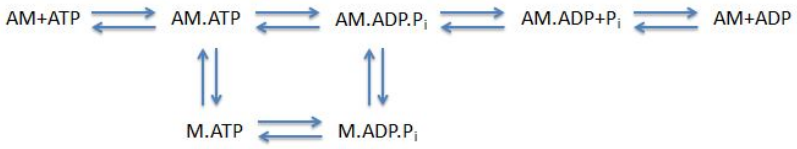


Figure 3.1: Lymn-Taylor Cycle. In this representation A correspond to Actin, M for myosin. Other ligands are ADP and ATP are also shown in different states and hydrolization also produce inorganic phosphate (P_i).

3.2 Myosin Structure and Dynamics

There are 35 classes of myosin known, 13 of them appears in human body [69]. Myosin molecule may have different structures distinguished by structural and functional features. There exist one and two headed myosin molecules. Myosin I, Myosin III and Myosin XV have one head. However, Myosin II and Myosin V have two heads with tails with different length. In this work, Myosin II molecule is studied, which is also called conventional myosin type and responsible for muscle contraction. PDB structures explained in this section specific for *Dictyostelium discoideum* myosin II motor domain.

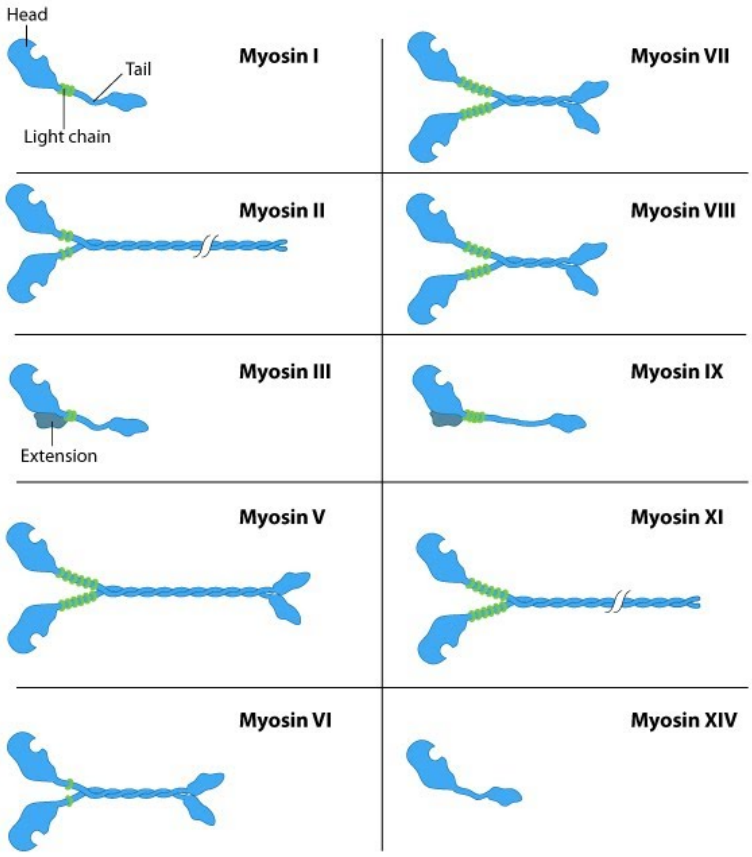


Figure 3.2: Myosin family of motor proteins. Most of the myosin proteins have two head domain. Light chains are also part of the myosin connecting head domain and tail [31]

Myosin is composed of two chain: heavy and light chains. Myosin heavy chain consist of three domains, which are head domain, hinge region and tail domain. Head domain responsible from binding actin filament and ATP molecule. Hinge region binds to myosin light chains and functions as a lever arm. Heavy chain varies for different myosin structures and fulfil different mechanisms.

Myosin II has a short lever arm with two light chains. Myosin tails can also be stacked together to form a large complex. Each light chain weights about 20 kDa and head has a molecular weight of 95 kDa. The globular head, which forms the actin binding site and ATPase site, has three different regions. Secondary structure of head contains alpha-helices which holds the head region together. Light chain wrapped around the heavy chain and stabilized by a hydrophobic residues and also has the site of phosphorylation by myosin

light chain kinase. Heavy chain has the actin binding site and also ligand binding site. This area centered around seven-stranded beta sheet which places the actin [55].

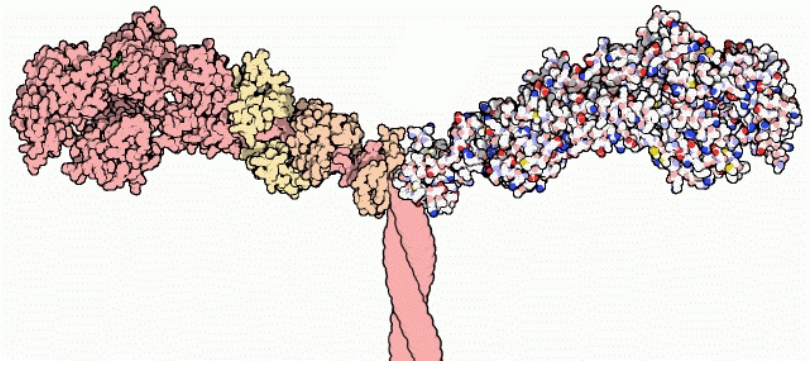


Figure 3.3: Part of the myosin structure, atoms in the heavy chain are coloured red on the left-hand side, and atoms in the light chains are coloured orange and yellow.

All myosin structures consist of four subdomains these are N terminal subdomain, upper 50 kDa subdomain (U50), lower 50 kDa subdomain (L50) and converter. These subdomain are connected by four ellastic joints which are called relay helix, strut, SH1 helix and switch II of ligand binding site [27, 82].

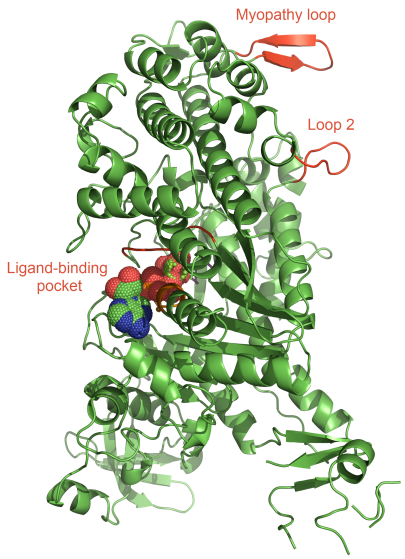


Figure 3.4: Structural domains of Myosin II protein with PDB code 1VOM.

Structural components of Myosin II exist in similar locations on different PDB files. These components are Switches, P-loop relay helix and loop, converter domain and Strut. Aminoacid types and alignments may change, but structure index couples are: switch I (233-238), switch II (454-459), p-loop (179-186), relay helix and relay loop (466-518), converter domain (710-747) and strut (590-593) [80]. In this work different states of Myosin ,all in common contains converter domain and ligand binding sites, corresponding to the head structure of Myosin.

3.3 Allosteric Mechanism

Conformational transitions have an important role for protein function. These conformational transitions occur via allosteric messaging through pathway of residues. Allosteric communication describes the mechanism of conformational changes via signal detection in one side of protein and influences to the distant sites.

Most known example of the allosteric mechanism occurs in Hemoglobin protein. Oxygen binding generates conformational changes between two equilibrium states of the Hemoglobin. These two states, tense and relaxed, characterized by their affinities for oxygen [33].

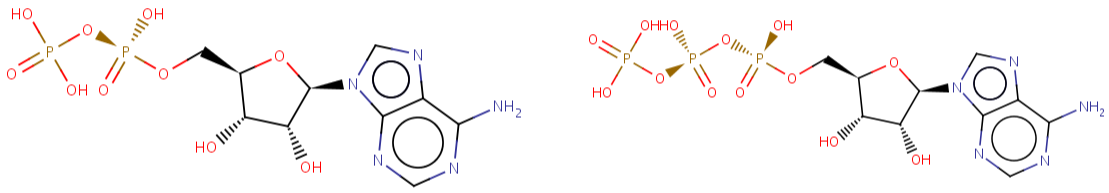
Allosteric mechanisms are first studied as MWC [44] and KNF [37] models. These models try to explain allosteric mechanism by using transition between known end states. Bounded states are less likely to interact with ligands. These two models differ in their assumption about interactions and states of the protein. There are also alternative approaches that claim multiple pathways for allostery. These multiple pathways exist between perturbation and the substrate binding sites [13].

MWC model claims that there exists different interconvertible states in the absence of regulators. Proteins change their conformations from one state to another by binding ligand to a site that is different from the active site [44].

In KNF model, allosteric change in one subunit does not necessarily change other subunits. Ligand binding by induced fit changes conformation, but this change does not propagated to other subunits, adjacent subunits affinity to substrate is changes [37].

ATP hydrolysis is an important reaction in which energy stored on chemical bond of phosphate releases to produce work. ATP hydrolyses to ADP and inorganic phosphate and energy stored in bonds are released. Chemical representations of ADP and ATP shown in

Fig.3.5.



(a) ADP (b) ATP

Figure 3.5: Essential macromolecules used in kinetic cycle of myosin.

3.4 Myosin States

Kinetic cycle states of the myosin known from the dynamical models and experimental techniques. There are appropriate structures known for corresponding states. Protein Data Bank (PDB) structures for these proteins are contains information about different states of the myosin. These structural characteristics either obtained X-Ray Crystallography or NMR (Nuclear Magnetic Resonance) methods.

In this work three major conformation states of the Myosin II are examined. These structures are classified by structural properties like their ligand binding states and sub-domain positions relatively each other in kinetic cycle. These conformations are pre-stroke which is also called transition state, near-rigor and post-rigor states.

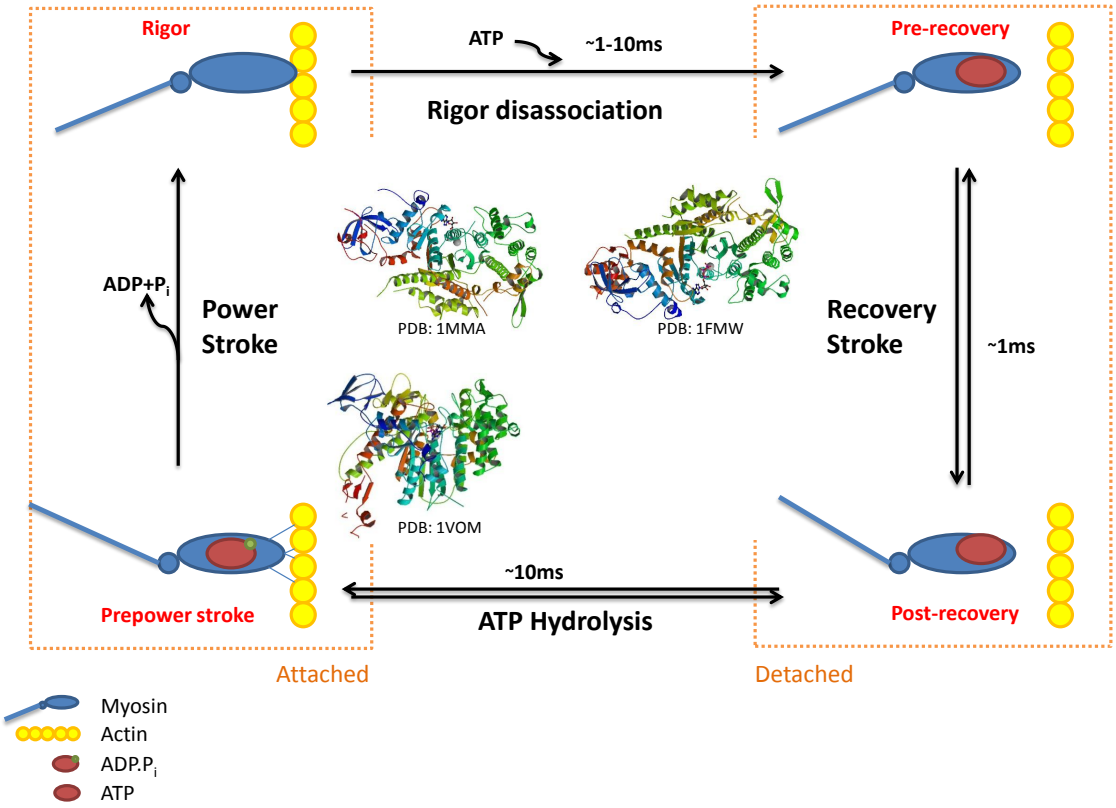


Figure 3.6: Transitions between myosin states. Transition properties and conformations of states are shown in figure. Two of these conformations are in attached to actin and other states detached. PDB structure of these states also shown in middle with given PDB codes. PDB structure 1VOM is in pre-powerstroke stage. ADP and inorganic phosphate release leads to rigor state. Between the rigor stage and pro-power stroke there exist near rigor conformation 1MMA. ATP bound to actin myosin complex cause disassociation from actin and this state is called pre-recovery and corresponding PDB is 1FMW.

In myosin kinetic cycle shown in Fig.3.6, four conformational state describe Actin-Myosin interaction and ligand binding states. Two of the states are attached states where myosin head interact with actin molecule. In rigor state myosin molecule interacts with actin until ATP molecule binds myosin head. Rigor disassociation takes 1-10 millisecond and myosin detached from actin. Major structural changes occur in the orientation of the lever arm. In post-rigor or pre-recovery state myosin and ATP complex creates recovery stroke and reach post-recovery state, where ATP hydrolysis starts. Hydrolysis of ATP into ADP and inorganic phosphate takes about 10 millisecond. In pre-power stroke state myosin head at-

tached with actin and power-stroke occur when ADP and inorganic phosphate disassociate. Power stroke leads system back to the rigor state.

Myosin kinetic cycle generates repetitive motion. During this cycle of the motion structural changes occur in every step. Transition between steps can also be explained these structural changes and points functionally important sites of Myosin. Table 3.1 summarize some of the important structure and transitions between states. Mechanism of myosin domains and important transitions explained to draw more clear picture. Each state has its own functional sites and transition of these sites leads overall dynamics for given state.

Table 3.1: Structural characteristics and conformational states of studied Myosin proteins. Data in the table partially used from [80]

State	**PDB Code**	**Ligand**	**Switch I**	**Switch II**	**P-loop**	**Lever Arm**	**Strut**
			233-238	454-459	179-186	710-747	590-593
Post-rigor	1FMW	ATP, Mg^{+2}	Closed	Open	Down	Down	Open
Pre-powerstroke	1VOM	ADP, Mg^{+2}	Closed	Closed	Down	Up	Open
Near-rigor	1Q5G	None	Open	Closed	Up	Down	Closed
	1MMA	ADP, Mg^{+2}	Closed	Open	Up	Down	Open

Structural changes of myosin cannot be monitored during kinetic cycle. Because of this, transition between different states are estimated by experimental results and theoretic analysis. Initial and final states during transitions are known from crystallography results. Analysis of transition between kinetic cycle stages of myosin based on existing crystal structures. Pre-power stroke state (PDB:1VOM) is weakly bound state that occurs after hydrolysis [65]. Near-rigor state (PDB:1MMA) is proposed as a weakly bound to ligand that occurs shortly after detaching from actin [22]. Following the 1MMA state inorganic phosphate and ADP released and this will generate rapid binding to acting. This state is called rigor state. Rigor state is followed by ATP binding and system transforms into pre-recovery state (PDB:1FMW). After pre-recovery state recovery stroke generated and myosin reaches post-recovery stage, which is followed by ATP hydrolysis.

Classification of the states can be achieved by different ways. Ligand bound-unbound states is one of the classification. In the bound conformation catalytic site is active and switch I and II bound with the ligand. Another way of classifying stages is open/close states of switch I and II of the ligand-binding pocket.

Transition between pre-power stroke and rigor state explained by ligand-binding pocket

structures. Different models point same structures in different order, but role of switch I, switch II and P-loop is essential for cleft opening and ADP release. The major differance between C/C and C/O states corresponding to 1VOM and 1MMA states comparison with respect to the switch II position is in the movement of the switch II away from the ligand. During this transition very little change on Switch I is observed. This change on switch I linked to change on other structures including P-loop, N terminal and converter domain [83].

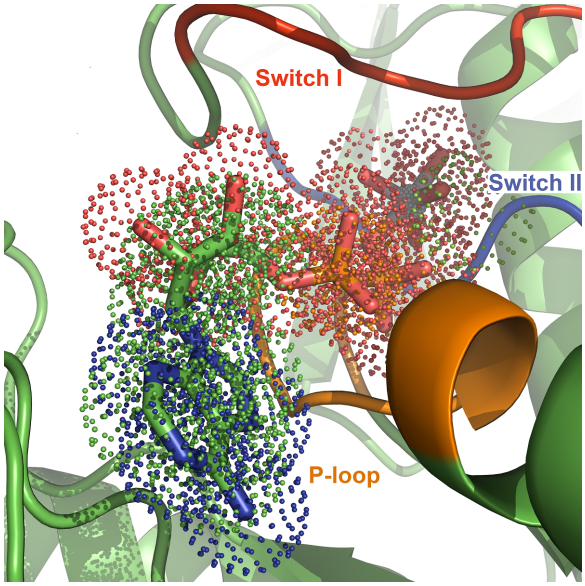


Figure 3.7: Ligand binding domain of Myosin (PDB:1VOM)

Inorganic phosphate release disrupts the hydrogen bond between switch II and γ-phosphate, which allow switch II to open, leading ADP release. Switch II contains conserved Gly-457 which has hydrogen bond with γ-phosphate in the closed from. Loss of inorganic phosphate broke this bond in the open state and leads ADP release and generate swing motion of the lever arm. Switch II change also moves relay helix and this cause a significant transition on lever arm during the force generation and formation of a tightly bound state of the rigor conformation [43, 56].

Release of the inorganic phosphate and ADP lead myosin head to bind actin with high affinity. Cleft closure changes switch I and ligand-binding pocket opens. Myopath loops has hydrophobic residues, which is essential for the actin binding. Deletion and mutation experiments resulted lose of functionality myosin for acting binding. Loops in myosin II join

large domains of the protein. The actin binding sites contain loop 2 structure which has an important role in regulating ligand release and actin affinity [47]. Actin binding creates rapid conformational changes and loop 2 rearranges its conformation [68]. Surface loops importance experimentally verified and interaction between loop 2 and negatively charged part of the actin [66]. Loop 2 cooperatively perform actin binding with the other surface loops and most likely to create actin binding interface that function triggering phosphate release [69]. A bent at the distal end of the long helix between residues 411-440 drives myopathy loop to bind to the surface of actin. Actin binding twists a large beta-sheet that forms myosin motor domain backbone. This twist changes relationship between switch I, swith II and P-loop in order to release of the ADP and P_i permanently [17, 43].

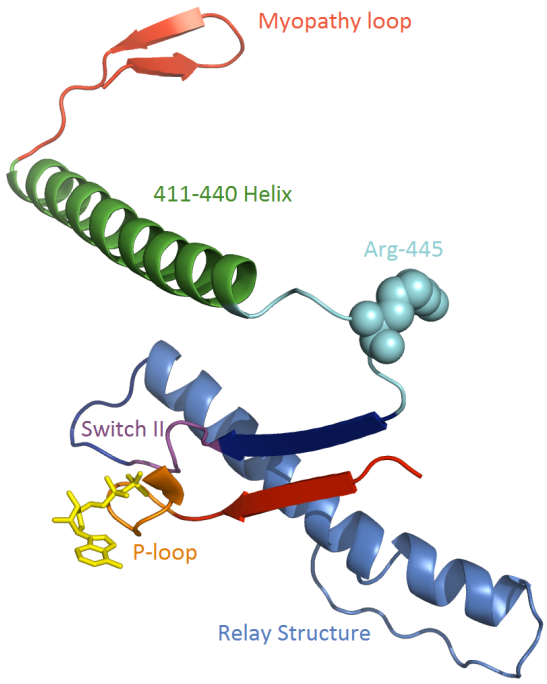


Figure 3.8: Important structural element effective ADP and P_i release

Conformational change between ATP and ADP bound states start with presence of the single phosphate group [76]. Detection of phosphate group is evident for the existence of γ-phosphate sensor. ATP hydrolysis requires interaction between switch II and γ-phosphate

that lead to closure of the γ-phosphate pocket, which prevents inorganic phosphate release after hydrolysis [29, 56]. Switch II is also coupled with both relay and SH1 helices during recovery stroke. Because of this coupling see-saw like motion rotation of the converter domain controlled. Localized motion around ligand-binding pocket, in particular, switch II loop observed stepwise displacement. Coupling with relay helix and SH1 helix generates rotation on the converter domain [36]. These transitions leads switch II loop closure and protein structure form in post-recovery stage.

3.5 Energy Pathway

Allosteric communication pathways are significant to understand biological significance of protein dynamics. In order to understand this pathways there are several methods suggested in literature. These methods are using evolutionarily constraint residues [71], shortest paths on graph networks [13, 71], mode analysis [14, 64, 70], hinge residues [82] and Molecular Dynamic simulations [80]. Experimental information of allostery commonly determined by personal structural intuitions based on available biochemical and functional data.

Previous results for energy pathways are organized in this chapter and our results will reference these results later on. Allosteric communication of Myosin II supposed to start from ATP binding site to the beginning of the lever arm. Different states on kinetic cycle of Myosin has different energy transfer pathways. Flexibility of switch I in near-rigor state observed and U50 subdomain rotates while U50-L50 cleft closes slightly.

Clue about communication pathway of residues are obtained by Zheng and Brooks [82]. In this work normal mode analysis and hinge residue detection used in order to determine essential residues on Myosin energy pathways.

Highly connected and correlated hinge residues are computed and sorted. These residues are distributed over several domain and structures. These residues and domains are organized and shown in Table 3.2. Information in the table is generated by DynDom [25] software and explained in work of Zheng and Brooks [82].

Table 3.2: Domain of correlations and hinge residues of the Myosin structure (PDB:1VOM). Residues marked by asteriks(*) are essential residues in domain correlation.

Domains of correlation	**Subdomains and Joints**	**Hinge Residues**
Ligand-binding site	Switch I Switch II P-loop	233*, 231, 232* 461*, 463 182*
Actin-binding site	Strut	573*, 574*, 575*, 586*, 587*, 589, 590*, 591, 629
Converter	-	742*
Relay Helix	-	481, 483*, 484, 485, 486, 487*, 488*, 489, 490*, 491, 493*, 494, 499, 500, 501, 502, 503, 504, 505, 506*, 508, 509*, 510*
SH1 Helix	-	674, 675*, 680*, 682*, 683*, 686, 687, 688, 689*, 691, 692, 693*, 695

In study of Zheng and Brooks [82], residues with high connectivity of myosin (PDB:1VOM) listed in descending order. Some of these residues are: 693, 121, 689, 485, 505, 490, 508, 494, 692, 742, 231, 695, 591, 502, 739, 688, 506, 120, 691, 662, 503, 747, 660, 488, 272, 509, 234, 483, 683, 493, 420, 97, 487, 182, 486, 499, 500, 587, 484, 590, 489, 504, 629, 674, 266, 686, 233, 573, 432, 501, 651, 259, 122, 98, 687, 680, 574, 682, 158, 194, 262, 481, 491, 650, 675, 745, 586, 589, 216, 463, 510, 461 and 232. Structural representation of these residues are shown in Fig.3.9.

Figure 3.9: Identified hinge residues in 1VOM shown in figure. Color blue and red corresponds to low and high connectivity of correlation values [82].

Another approach to determine allosteric communication is suggested in work of Tang and Liao [71]. In this study, pathway of residues that mediate allosteric communication predicted using only a combination of distance constraints and evolutionary data. They claim that allosteric communication starts from ATP catalytic site and destination is identified as a region corresponding to the beginning of the lever arm.

Surface residues are rated with high path efficiency metric, which is calculated by distance between start and end points divided by the sum of path conservation weights. Shortest path calculation is carried on pre-stroke myosin II proteins and important residues are listed with high path efficiency metric. Most of these residuals are located at the beginning of the lever arm. These residuals are 492, 496, 498, 500, 504, 693, 695, 738, 739, 741, 742 and 743 for 1VOM. Residues 492, 496 and 498 are part of the relay helix. Residue 742 selected as path endpoint because it had the highest score. In Fig. 3.10, top 25 high scoring residues of 1VOM shown. A green star indicates the origin of the signal propagation site.

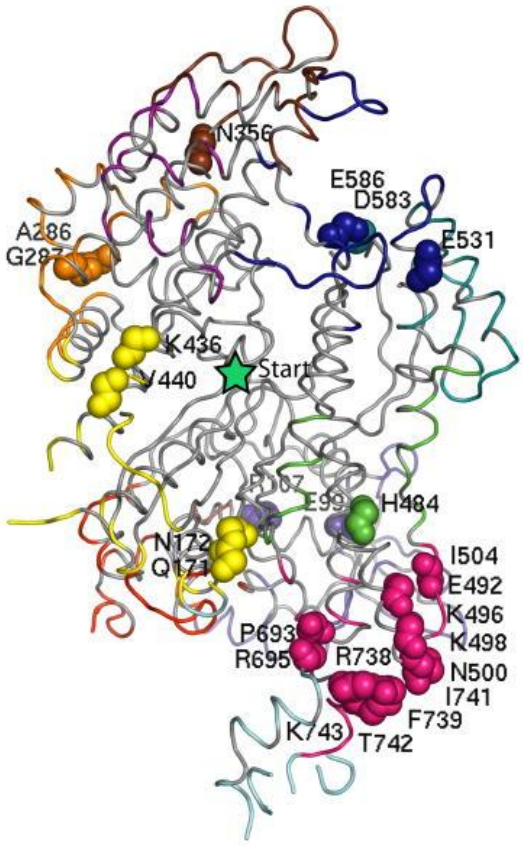


Figure 3.10: Pathway of surface residues with high potential of allosteric coupling. Essential residuals with top 25 score is shown and green star indicates origin of the allosteric messaging in the ATPase site [71].

Allosteric pathways are different for each state of the Myosin II, because of the ligand binding sites, interaction and conformations. Shortest paths in pre-stroke structure (1VOM) have multiple alternative paths before residue 482, which also supports multiple allosteric path hypothesis [13]. One option is to jump switch II residue 457 and then followed by residue 458. The other option is to traverse residue 181 in the P-loop followed by residue 458.

Allosteric path also calculated for port-rigor configuration (1MMA), which starts in the P-loop residue 186 and propagates through switch I residues 237 and 238, switch II residues 471 and 475, SH1 and SH2 residues between 670-700 and arrives residue 742.

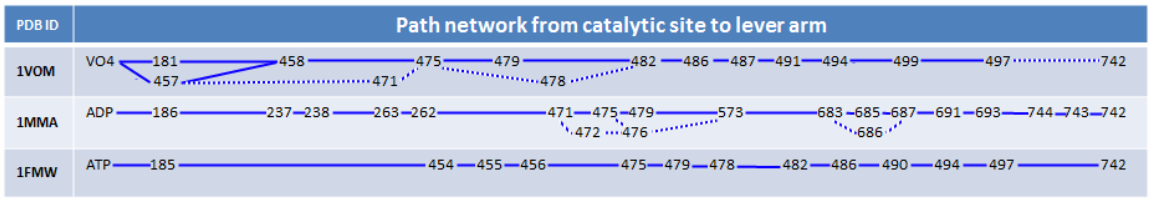


Figure 3.11: Pathway of residues resulted in the work of Tang and Liao for different conformations of Myosin.

3.6 Experimentally Verified Results

Coordination between the ligand binding domain and the converter domain of myosin is essential for its motor activities. Communication between these two distant region is open problem and there are different techniques suggested in literature. Experimental evidence of these finding also suggests two different hypothesis. One is through the interaction between the relay helix and the converter [15]. Experimental finding about this hypothesis test disruption of a hydrophobic linkage between I499 in the relay helix and converter region [62]. Other hypothesis suggest interaction between the relay helix and the SH1-SH2 helix [28]. There are also other experimental studies show significant interactions explain in this section.

Coupling between chemical reactions and conformational changes is crucial. Experimental techniques are the most informative way to determine functional importance of residues for the protein dynamics. However, these techniques are time-consuming and expensive. Mutations on the genes resulted different amino acid sequences for the protein [60]. These changes may effect proteins dynamics significantly. Experimental information are used in literature to validate computational results acquired [42].

Table 3.3: Mutation sites experimentally tested and list of observed residues which is essential for protein dynamic. In this table mutation experiment results for *Dictyostelium discoideum* myosin II motor domain (PDB:1VOM) is collected.

Mutation sites	**Reference**
Glu-459, Asn-464, Glu-746, Asn-475, Cys-470, Phe-481, Asn-472, Tyr-473	[59]
Glu-467, Gly-624, Gly-680, Gly-691, Tyr-494, Trp-501, Glu-586, Gly-740	[51]
Asp-454, Gly-457, Phe-458, Glu-459	[61]
Phe-487, Phe-506, Trp-501	[74]
Glu-531, Pro-536, Arg-562	[18]
Ans-233, Ser-237, Arg-238	[63]
Asp-403, Val-405	[48]
Gly-680, Gly-691	[50]
Phe-482, Gly-680	[32]
Ile-499, Phe-692	[62]
Ser-236	[17]

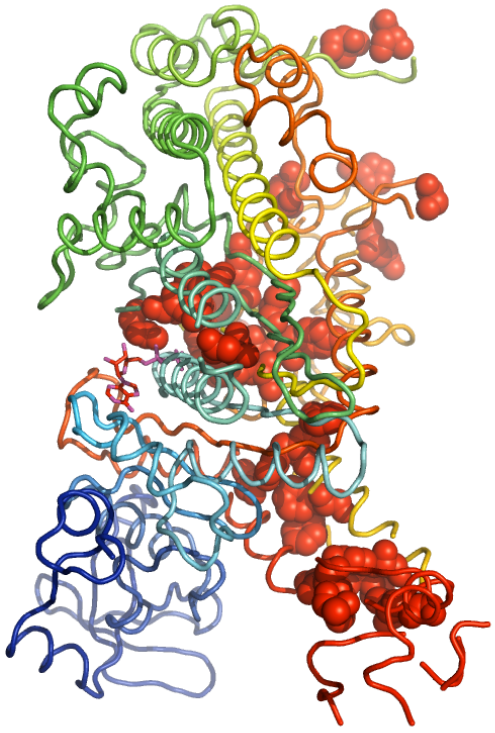


Figure 3.12: Essential residues extracted from experimental analysis are represented as spheres.

Chapter 4

SIMULATIONS

In field of protein dynamics, molecular dynamics simulations are essential to capture vibrations within protein and conformational changes. In order to obtain this information, molecular dynamic (MD) simulations made for Myosin II on different states of kinetic cycle.

Side chain motions of Myosin play essential role to understand dynamics of protein. Examination of the imaging techniques do not reveal any path related to the allosteric mechanisms.

All molecular dynamics simulations used in this work were performed using NAMD 2.7 software package [52] with CHARMM27 force field [8] in explicit solvent (water) at 310 K. Simulations were performed using Langevin dynamics to control temperature and pressure for an N,P,T ensemble. Waterbox with cushion 15$\mathring{A}$ and periodic boundary conditions were applied. Timesteps of the simulations were selected small to capture motion of the protein both in slow and fast frequencies. A time step 1 fs was used and both non-bonded and electrostatic forces were evaluated and no rigid bonds were used. Trajectories of motions are captured every 50 fs. Trajectory files are large to display using VMD [30] and working with because of the memory needs. In order to overcome this problem Carma software [19] used to reduce trajectory file sizes by removing atom other than backbone atoms.

4.0.1 1FMW

Structure of Myosin II in post-rigor state is captured by X-ray crystallography methods. PDB code of this molecule is 1FMW [5]. In this state Myosin head is detached from the actin and followed by pre-recovery stroke. Hydrolises of MgATP converts energy into direct movement.

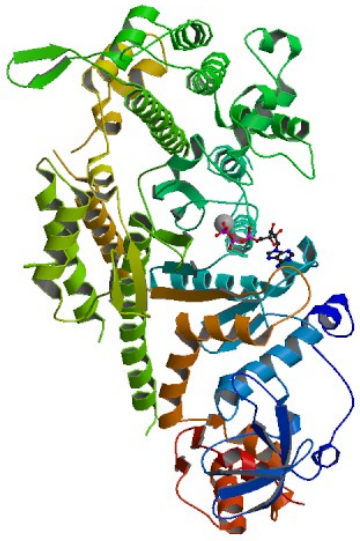


Figure 4.1: 1FMW: Crystal structure of the MgATP complex for the motor domain of Myosin II

Stability of 1FMW is also high. All residuals placed in allowed regions in ramachandran plots and 95.6 percentage of all residuals located in favoured regions and there is no outliers exists [11].

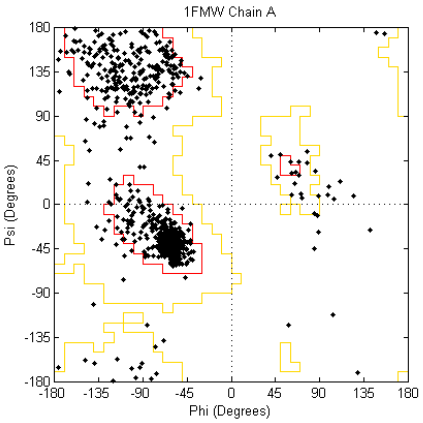


Figure 4.2: Ramachandran plot of the 1FMW protein

Interactions between protein residues and Ligands shown in Fig.4.3. These interaction also reveal active sites and binding regions. Interaction with Mg^{+2} is ionic and there are also

hydrogen bonds exists with Asn127, Tyr135, Ser181, Gly182, Lys185 and Asn233. There are also van-der-waals bond exists in residues Ile115, Pro128, Lys130, Ala183, Gly184, Thr186, Glu187, Ser236 and Ser237.

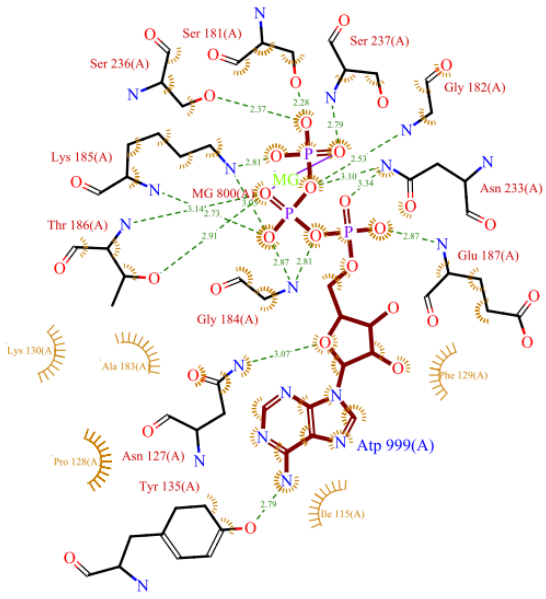


Figure 4.3: Interactions between 1FMW and Ligands. Interaction on residues shown in yellow color represents van-der-waals bonds and red ones represents hydrogen bonds [78].

This PDB file contains protein with 738 residue. In this state of Myosin II ligands also placed in PDB file, which are ATP and Mg^{+2}. Simulation of 1FMW takes 2.15 ns and 43000 frame available in trajectory file. Last 30000 frame are considered as equilibrium and only those motions are used in this study. RMSD of 1FMW is shown in Fig.4.4.

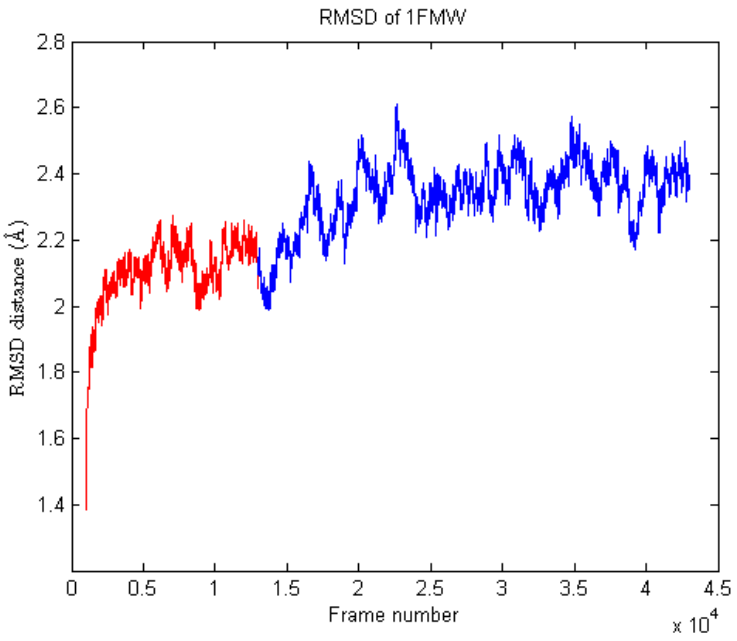


Figure 4.4: RMSD of the 1FMW protein

4.0.2 1VOM

Pre-power stroke state of the Myosin II is represented by protein structure with PDB code 1VOM [65]. In pre-power stroke conformation of Myosin motor domain which contain MgADP complex. In this state myosin weakly bound to actin occurs after hydrolysis. Furthermore, vanadate structure (O_4V) in 1VOM is a model for the transition state of ATP hydrolysis and used in order to provide appropriate crystallization.

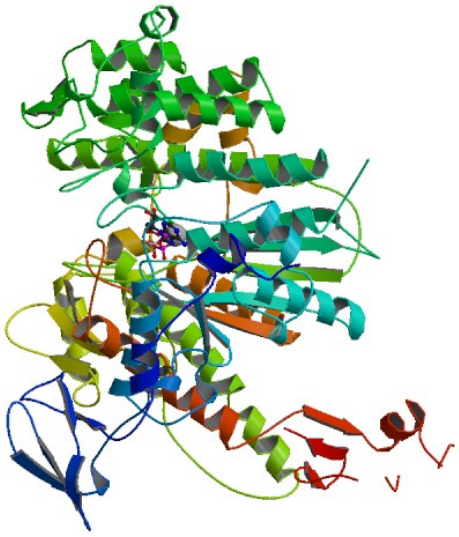


Figure 4.5: 1VOM: Crystal structure of the MgADP complex for the motor domain of Myosin II

Stability of 1VOM structure is calculated using MolProbity servers [11]. There exist one outlier residue Asp733. However all other residues are located in allowed regions and 96.8 percent of residues are located in favoured regions.

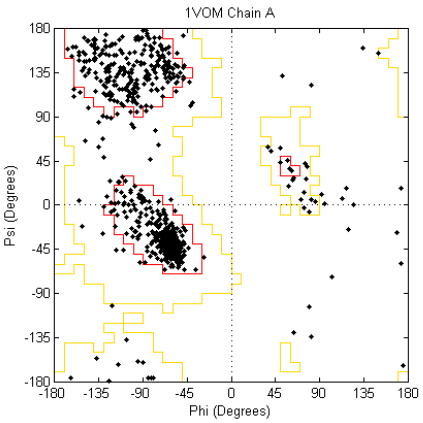


Figure 4.6: Ramachandran plot of the 1VOM protein

Interactions between protein residues and Ligands shown in Fig.4.7. These interaction also reveal active sites and binding regions. Interaction with O_4V is covalent and Mg^{+2} is ionic. There are also hydrogen bonds exists with Asn127, Tyr135, Gly182, Gly184, Lys185

and Asn233. There are also van-der-waals bond exists in residues Pro128, Phe129, Lys130, Ala183, Thr186.

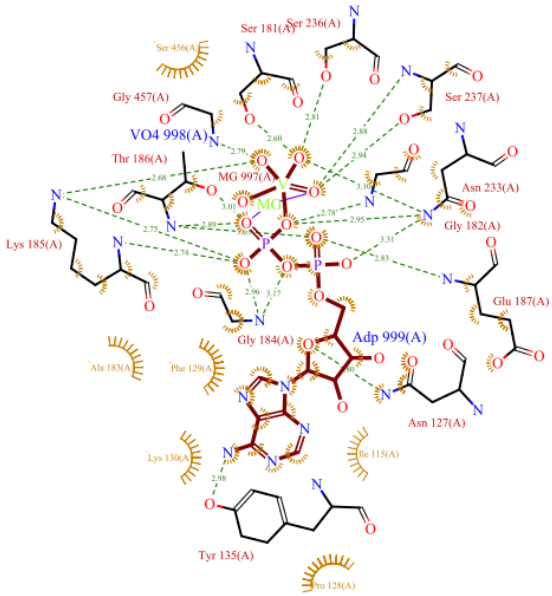


Figure 4.7: Interactions between 1VOM and Ligands. Interaction on residues shown in yellow color represents van-der-waals bonds and red ones represents hydrogen bonds [78].

This PDB file contains protein with 730 residue. In this state of Myosin II ligands also placed in PDB file, which are ADP and Mg^{+2}. Simulation of 1VOM takes 2.15 ns and 43000 frame available in trajectory file. Last 30000 frame are considered as equilibrium and only those motions are used in this study. RMSD of 1VOM is shown in Fig.4.8.

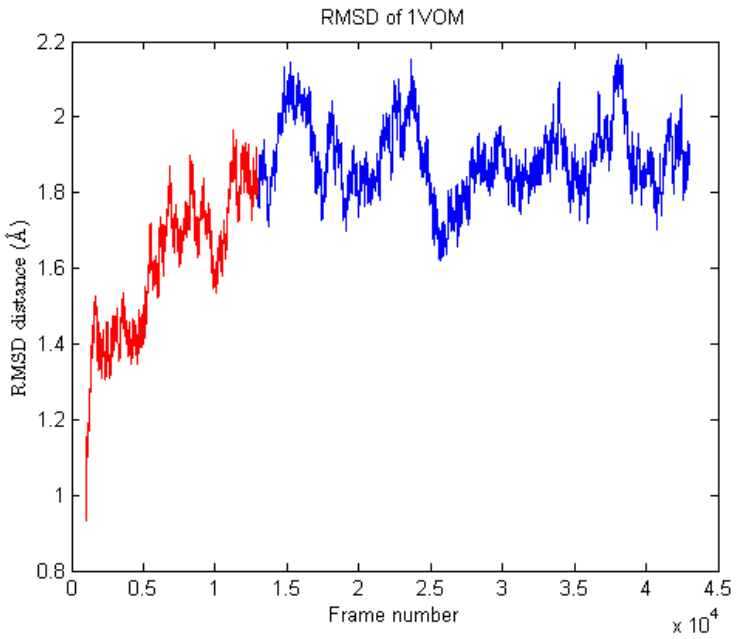


Figure 4.8: RMSD of the 1VOM protein

1VOM longer simulation

In this simulation 23.65 ns simulation made with 5000fs intervals. This simulation continue using snaphot after 2.1 ns simulation. There exist 4730 frame exist and last 4000 frame considered as equilibrium and only those frames used in this study.

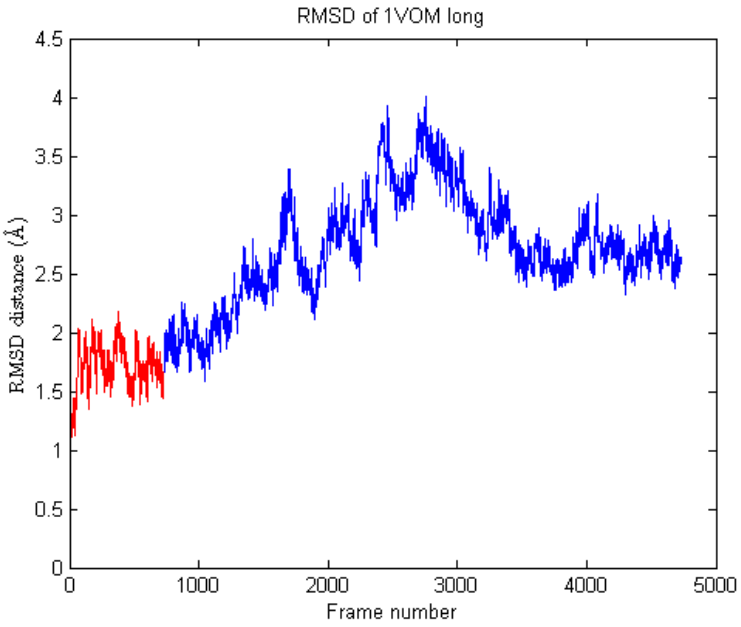


Figure 4.9: RMSD of the 20 ns 1VOM simulation

1VOM without ligand

In this special case, effect of the ligand on protein dynamics are seek. PDB structure of 1VOM is used in this simulations without using Ligands. ADP and magnesium ion are removed from the PDB file. This PDB file contains protein structure same as 1VOM. Simulation of 1VOM without ligand takes 2.1 ns and 42000 frame available in trajectory file. Last 30000 frame are considered as equilibrium and only those motions are used in this study. RMSD of 1VOM without ligand is shown in Fig.4.10.

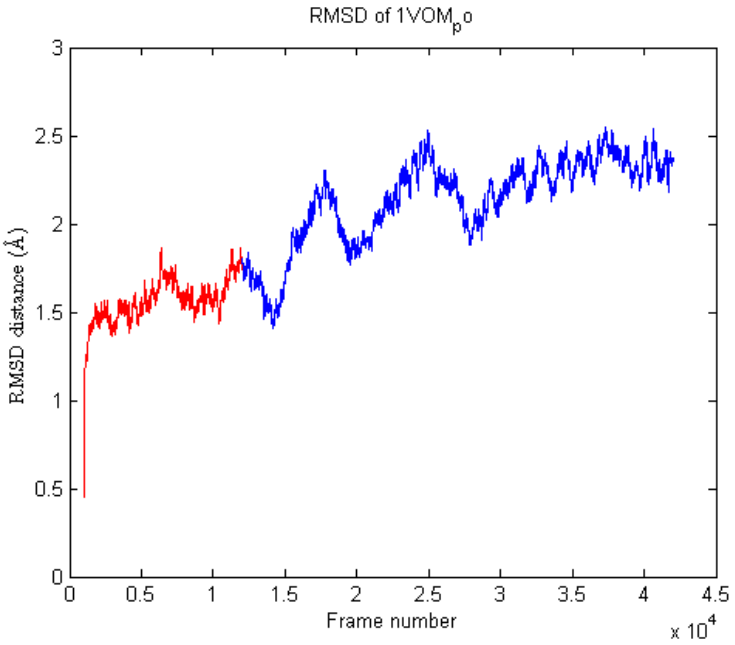


Figure 4.10: RMSD of the 1VOM protein when ligands are removed from the system

In this structure system is not in equilibrium state. Absence of ligands drive motion of the protein to different state.

4.0.3 1MMA

Post-rigor state of the Myosin II is represented by protein structure with PDB code 1MMA [22]. Conformation of myosin in 1MMA structure contains Mg^{+2} and ADP as ligand which are generated after ATP hydrolysis. During power stroke ADP and inorganic phosphate leave and reached attached rigor state.

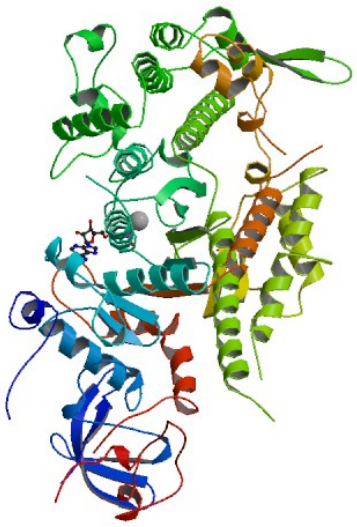


Figure 4.11: 1MMA: Crystal structure of the MgADP complex for the motor domain of Myosin II in post-rigor state

Stability of 1MMA structure is calculated using MolProbity servers [11]. There exist 16 outliers these residues are Asp69, Asp86, Ala363, Thr373, Asn410, Phe535, Pro536, Thr688, Gly691, Phe692, Pro693, Asn694, Tyr698, Glu735, Lys743 and Arg753. However 97.7 percent of residues are located in allowed regions and 89.3 percent of residues are located in favoured regions.

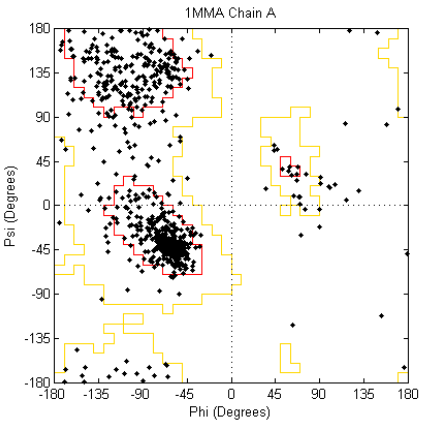


Figure 4.12: Ramachandran plot of the 1MMA protein

Interactions between protein residues and Ligands shown in Fig.4.13. These interaction

also reveal active sites and binding regions. Interaction with Mg^{+2} is ionic and Glu187 is electrostatic. There are also hydrogen bonds exists with Asn127, Tyr135, Gly182, Gly184, Lys185 and Asn233. There are also van-der-waals bond exists in residues Pro128, Phe129, Lys130, Ala183.

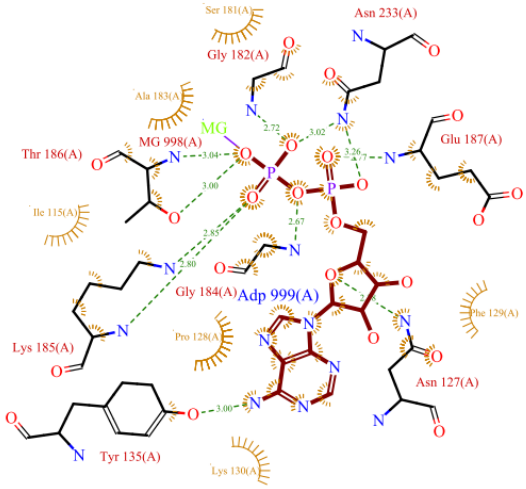


Figure 4.13: Interactions between 1MMA and Ligands. Interaction on residues shown in yellow color represents van-der-waals bonds and red ones represents hydrogen bonds [78].

This PDB file contains protein with 719 residue. In this state of Myosin II ligands also placed in PDB file, which are ADP and Mg^{+2}. Simulation of 1MMA takes 0.7 ns and 14500 frame available in trajectory file. Last 10000 frame are considered as equilibrium and only those motions are used in this study. RMSD of 1MMA is shown in Fig.4.14.

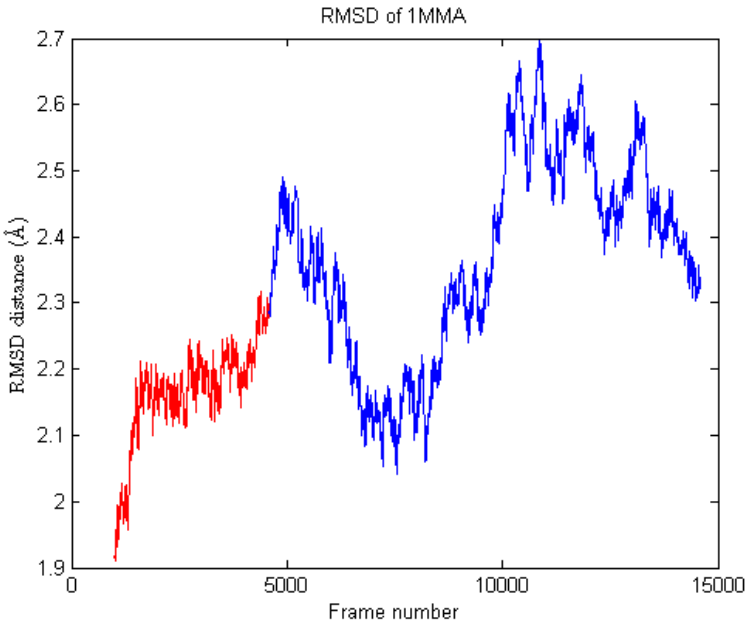


Figure 4.14: RMSD of the 1MMA protein

4.0.4 1Q5G

Conformatonal model for near-rigor state of Myosin II is theoretically suggested [56]. Despite of theoretical nature, structural properties are studied and simulations also support information about this conformation. This structure offered as a state between 1VOM and 1FMW.

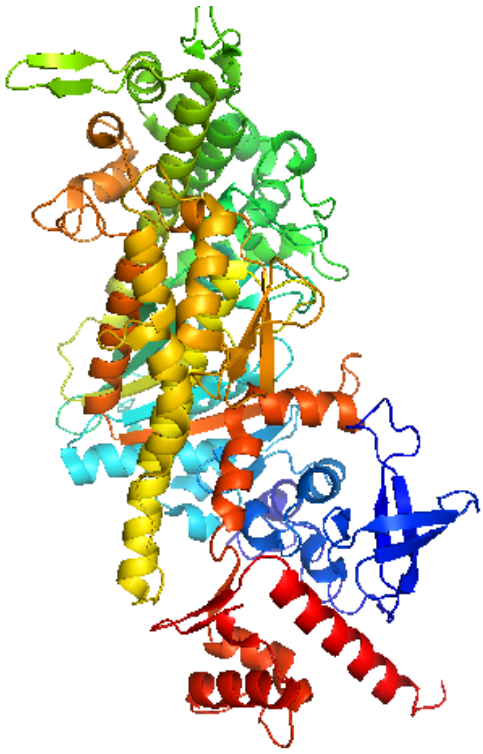


Figure 4.15: 1Q5G: Theoretically produced structure of Myosin II in near-rigor state

Stability of 1Q5G structure is calculated using MolProbity servers [11]. There exist 8 outliers these residues are Asp23, Phe25, Thr28, Gly67, Ala205, Arg620, Ala621 and Lys622. However 99 percent of residues are located in allowed regions and 95.8 percent of residues are located in favoured regions.

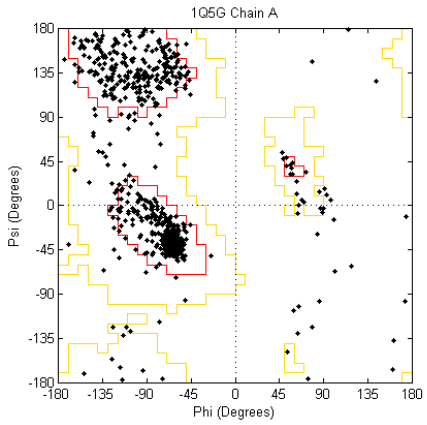


Figure 4.16: Ramachandran plot of the 1Q5G protein

This PDB file contains protein with 770 residue. In this state of Myosin II theoretically suggested. Simulation of 1Q5G takes 1.23 ns and 24500 frame available in trajectory file. Last 20000 frame are considered as equilibrium and only those motions are used in this study. RMSD of 1Q5G is shown in Fig.4.17.

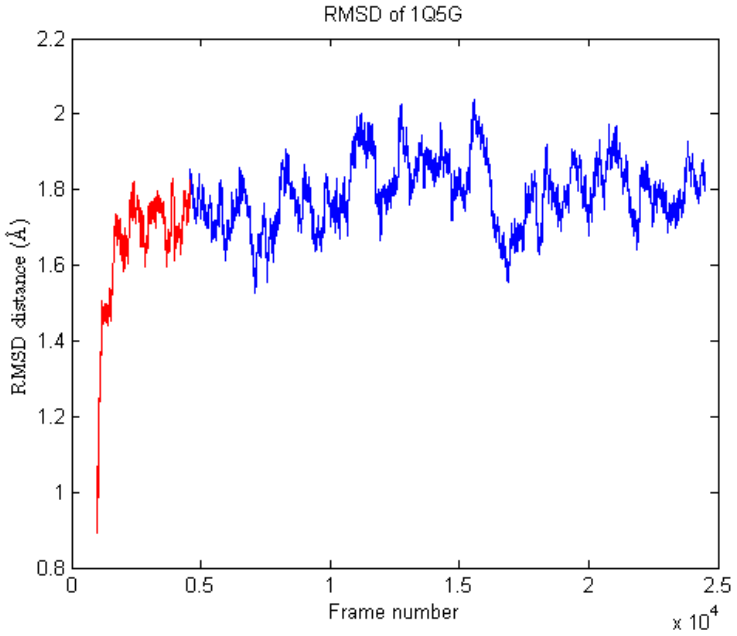


Figure 4.17: RMSD of the 1Q5G protein

Chapter 5

DETECTION OF FUNCTIONALLY IMPORTANT SITES

Fluctuations of the residuals reveal essential information about structural properties and functional sites of protein. Protein dynamics analysis based on fluctuations of residues are promising method and have growing interest in protein physics [24, 53]. Proteins simply modelled as mass and spring systems in elastic network models (ENMs) which has independent harmonic modes of fluctuations [2]. Protein dynamics more complex and believed that anharmonic and mode-coupling interaction relevant for the information tranfer through protein which cannot be captured by harmonic models [40, 46].

Modal decomposition of modes reveal relations between modes and functions. Slow modes known as mostly anharmonic and fast modes contributes energy exchange [26, 79]. Vibrational energy transfer through mode-coupling studied for specific modes before [45]. Systematic analysis for energy transfer between modes are still open question. Mode-coupling occurs between two modes for a short period of time and one mode driver the other mode and creates conformational changes important for understanding allostery [24].

In this paper, we present a systematic way of representing mode-coupling and anharmonic contributions of residues. Hermite expansion of modal fluctuations is used in order to generate anharmonic corrections on purely harmonic model. These distributions are compared with molecular dynamic results of residue fluctuations. Result shows that mode-coupling effect dominant on functionally important residues.

5.1 Hermite Expansion

Molecular dynamics trajectory contains coordinates of the atoms and these atoms deviates from their mean coordinates within time. Fluctuations of residue coordinates defines a multivariate probability distribution function $f(\Delta\mathbf{R})$. Deviation from mean position represents as $\Delta R_i = R_i - R_i^0$, $i = 1, ..., 3N$. In coarse grained representation N corresponds to number of alpha carbons. R_i, R_{i+N} and R_{i+2N} correspond to displacements of residue i on x,y,z

coordinated. Transformation from space coordinates to modal coordinates applied using covariance matrix $\Gamma = \left\langle \Delta\mathbf{R}\Delta\mathbf{R}^T \right\rangle$.

$$\Delta\mathbf{r} = \left\langle \Delta\mathbf{R}\Delta\mathbf{R}^T \right\rangle^{-1/2} \Delta\mathbf{R} \tag{5.1}$$

Tensorial expansion for the probability distribution function represented as orthogonal basis with hermite polynomials [34]. Probability distributions of residue fluctuations represented as

$$\begin{aligned} f(\mathbf{\Delta r}) = \frac{1}{\sqrt{(2\pi)^N}} e^{-\sum_i \Delta r_i^2/2} [1 + \sum_i \sum_{\nu=3}^{\infty} \frac{1}{\nu!} \left\langle H_\nu(\Delta r_i) \right\rangle H_\nu(\Delta r_i) \\ + \sum_{i \neq j} \sum_{\nu=3}^{\infty} \frac{1}{\nu!} \sum_{p=1}^{\nu-1} \binom{\nu}{p} \left\langle H_p(\Delta r_i) H_{\nu-p}(\Delta r_j) \right\rangle H_p(\Delta r_i) H_{\nu-p}(\Delta r_j) + \sum_{i \neq j \neq k} \cdots] \end{aligned} \tag{5.2}$$

where $\langle . \rangle$ denotes time average evaluated over molecular dynamics data. The first term in Eq. 5.2 corresponds to a harmonic contribution term given by the Gaussian probability distribution and called f_0 in our study. This harmonic model is the basis of the many protein fluctuation models [1, 2, 79].

$$f_0(\mathbf{\Delta r}) = \frac{1}{\sqrt{(2\pi)^N}} e^{-\sum_i \Delta r_i^2/2} \tag{5.3}$$

Anharmonic model which is also called f_1 model represented by contribution of decoupled modes to f_0 harmonic model. Second term of this function is corresponds to anharmonic terms and gives no information about mode couplings. The general form of the mode-amplitude distributions composed of using harmonic effects and decoupled modes are

$$\begin{aligned} f_1(\mathbf{\Delta r}) = \frac{1}{\sqrt{(2\pi)^N}} e^{-\sum_i \Delta r_i^2/2} \\ \times \prod_i [1 + \sum_i \sum_{\nu=3}^{\infty} \frac{1}{\nu!} \left\langle H_\nu(\Delta r_i) \right\rangle H_\nu(\Delta r_i)] \end{aligned} \tag{5.4}$$

The difference between the MD results pdf and the approximations f_1 model is the

mode-coupling which is given equations Eq.5.2 and Eq.5.4. Hermite polynomials degree important in this case and empirical observation shows that 32 is appropriate degree and used in our further analysis. Effect of the hermite degree can be seen in Fig.5.1. Further corrections can be applied to model f_1 in order to cover mode-coupling corrections at the lowest order they appear while ignoring cubic and other higher order terms [34].

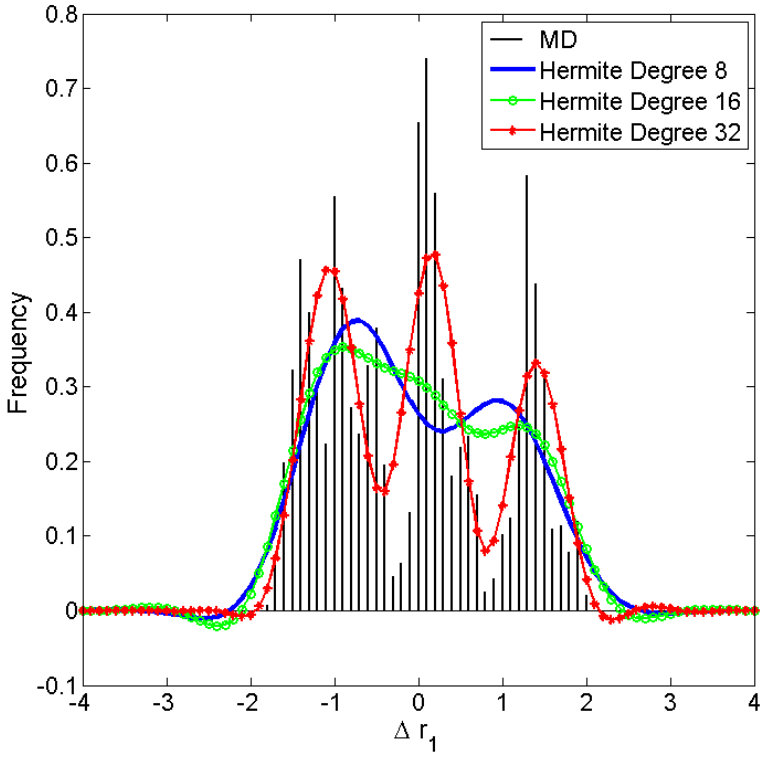


Figure 5.1: Different hermite degree orders compared with MD data.

5.2 Mode-coupling and anharmonicity effects

The probability distributions of residue fluctuations obtained by the Hermite series expansion. Purely harmonic model (f_0) and model with the anharmonic corrections (f_1) is compared with the molecular dynamic results. Distributions of displacement on real space are compared using KL distance. MD-f_1 difference reveals mode-coupling contributions and differ from f_1-f_0 difference on the functionally important sites.

It is important to emphasize that distribution compared using KL metric is the distribu-

tion of the fluctuations in real space. Modal space coordinates transformed into real space coordinates using inverse gamma matrix that is used to transform MD trajectory data into modal space as explained in Appendix A. This production does not affect properties of the distributions. Linear combinations of the values remain applying same distribution.

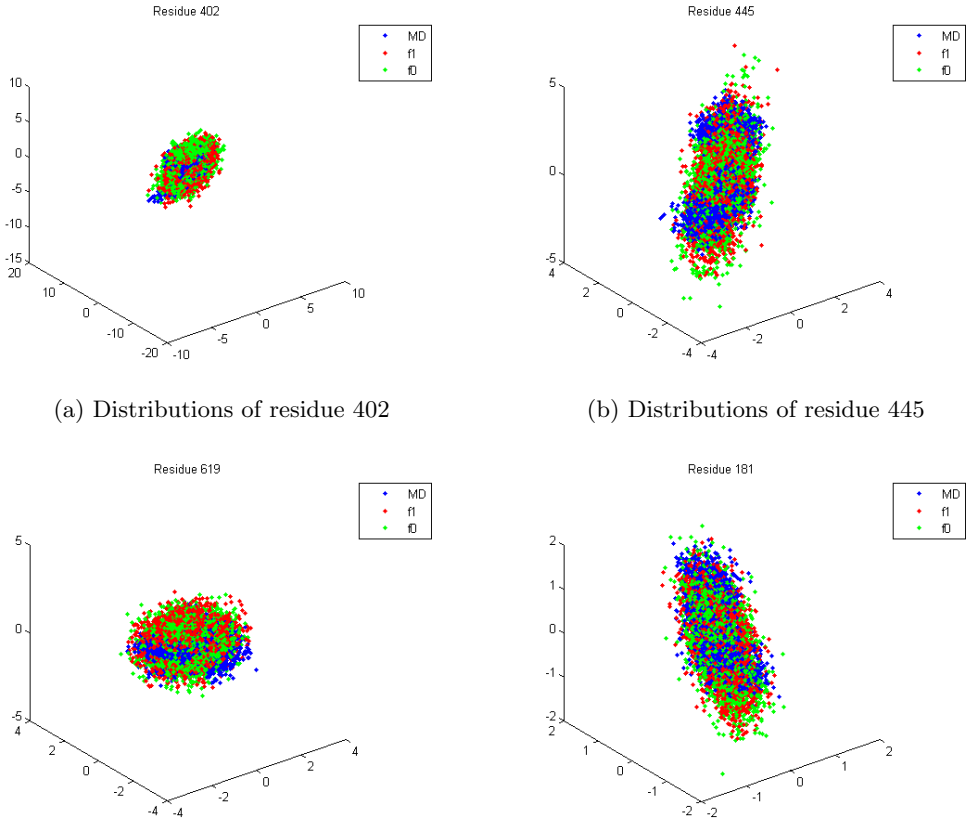
Our process starts with the production of the f_0 model samples. These set contains samples taken space of the modal coordinates. Each sample taken from this set with elements selected from normal distribution.

Samples taken from f_1 distribution by using rejection sampling [7]. Rejection sampling based on selecting random samples from uniform distribution or other distributions with known density function. Samples within curve distributed uniformly after rejecting samples outside of the curve. In our study random samples taken from normal distribution. Height of this gaussian curve update is the sample point from unknown distribution remain upper from the gaussian curve. Gaussian curve selected in order to cover all sample points below.

5.3 Distributional differences

Difference between distributions are measured by using Kullback-Leibler (KL) divergence [39]. KL divergence is a non-symmetric measure for comparing two probability distributions. Information gain or relative entropy for given distributions P represents real distribution data and Q represents model distribution. For probability distributions of discrete random variables P and Q, KL divergence are expressed as

$$D_{KL}\left(P \parallel Q\right) = \sum_{i} P(i) ln \frac{P(i)}{Q(i)} \tag{5.5}$$



(a) Distributions of residue 402

(b) Distributions of residue 445

(c) Distributions of residue 619

(d) Distributions of residue 181

Figure 5.2: Distributional differences of residues which has large KL divergence between MD and f_1 distributions. Sample points from f_0 and f_1 distributions and MD data are plotted for x,y,z displacements ΔR_i for each frame or sample configuration.

Samples from f_0, f_1 and MD samples for each set in $\mathbf{\Delta R}$ are considered as random variables from unknown probability density functions. Estimation of probability density functions are possible using kernel density estimation (KDE) [49, 58]. In statistics, KDE is known to be a non-parametric estimation technique for probability density function (pdf) of a random variable. KDE is useful for making inference about population using finite number of data sample. Samples from distribution f with unknown distribution are used to calculate estimated probability distribution $\hat{f}$ as

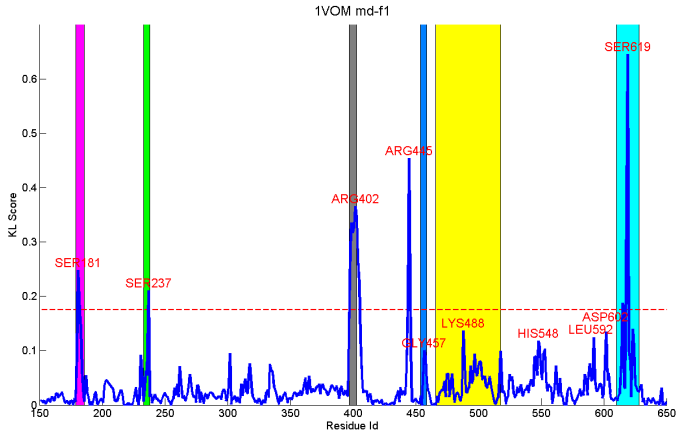
$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{5.6}$$

where K is the kernel and h is the smoothing parameter which is also called bandwidth.

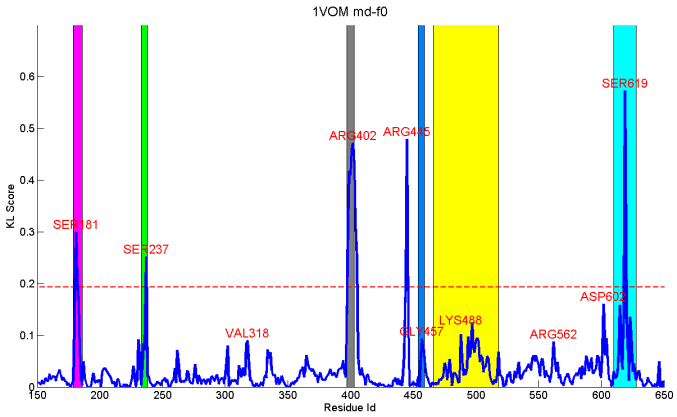
Kernel function selected as Gaussian function in order to use practical estimation of the bandwidth [6].

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5} \approx 1.06\hat{\sigma}n^{-1/5} \tag{5.7}$$

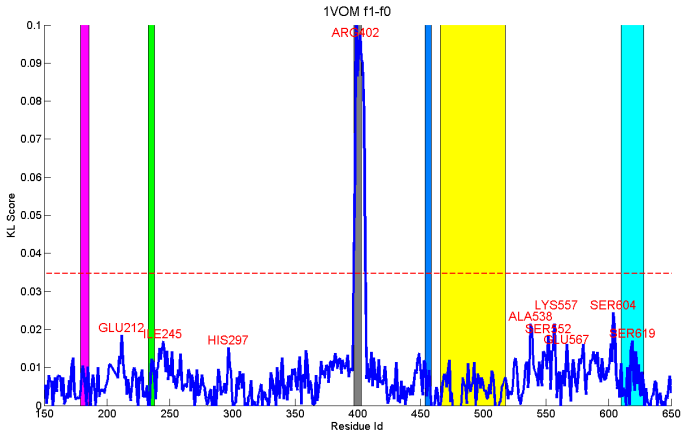
0 where $\hat{\sigma}$ is the standard deviation of the n sample points. Sample distributions on real space shown in Fig 5.3. Comparison between this distributions measured by KL-divergence for each axis. Overall KL score is generated by taking $L1$ distance in other words summing distances of each axis.



(a) KL divergence between MD-f_1 in other words residual mode-coupling contributions



(b) KL divergence between MD-f_0 contains both anharmonic and mode-coupling contributions



(c) KL divergence between f_1-f_0 which contains anharmonic effect

Figure 5.3: KL divergences between different distributions. Yellow regions point functionally important sites p-loop, switch I, myopathy loop, switch II, relay helix and loop 2

5.4 Functional sites of Myosin

Mode-coupling contribution on protein dynamics revealed from the differences between distributions MD and f_1. Significant differences point functionally important sites such as ligand binding site and structures essential for kinetic cycle of the myosin. Differences between two KL score reveals essential sites produced by mode-coupling. Functional importance of residues emerged as peaks in MD-f_1 distribution difference explains using biological background about myosin in Chapter 3

KL score results for different distribution pairs shown in Fig.5.3. In this figure difference between MD and f_1 distributions shown mode-coupling contribution for each residue. Anharmonicity contributions are quiet different and less than mode-coupling effect. Difference between this scores for pair of distributions are 10 times greater in MD-f_1 plot. Difference of mode-coupling contribution and anharmonicity shown in Fig 5.4 Significant information obtain from the mode-coupling effects.

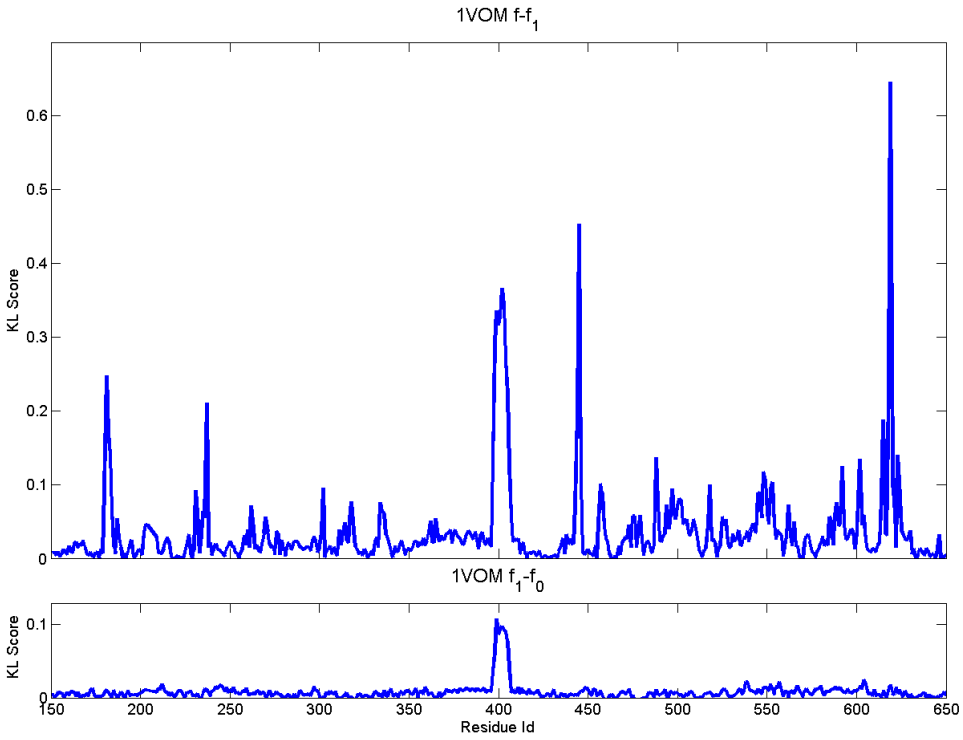
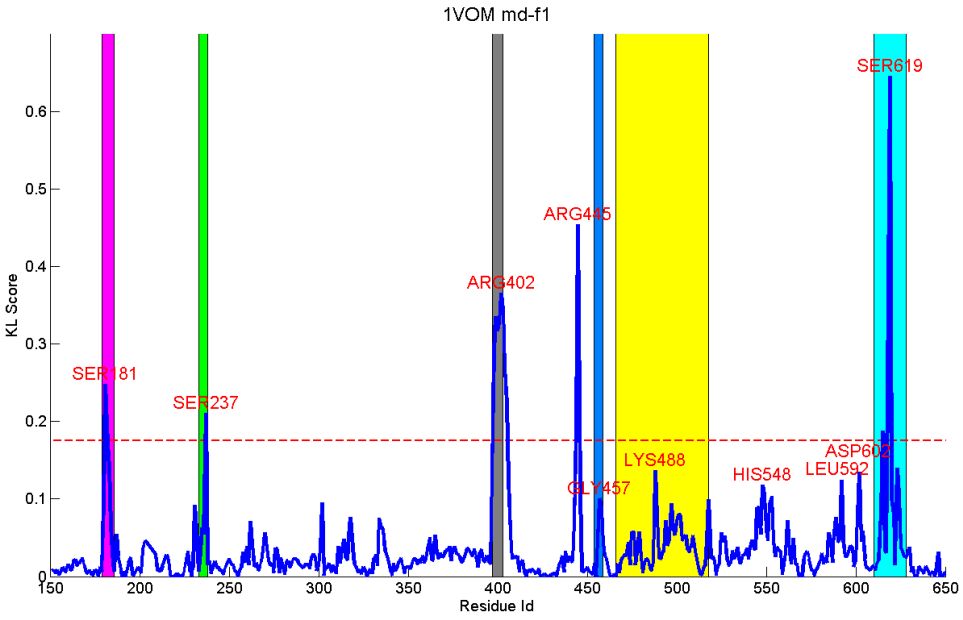
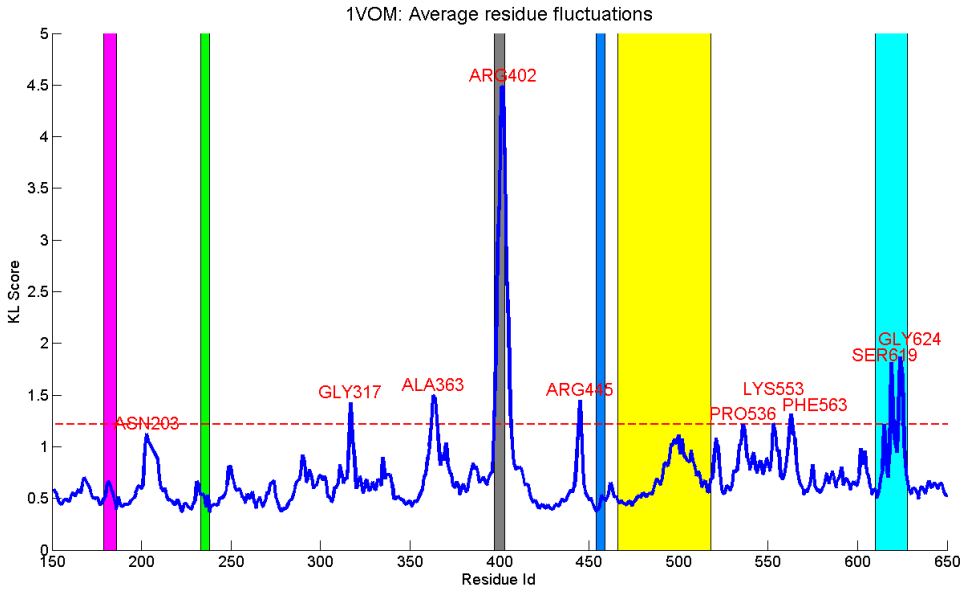


Figure 5.4: Difference between KL scores corresponding to mode-coupling and anharmonicity contribution.

Results of the fluctuations also considered in order to eliminate one possible explanation of our findings. Fluctuations contain information about the residual energy, but our result provides mode-coupling contributions. Fluctuations also shown in Fig. 5.5b peaks are commonly corresponding to loops with large displacements.



(a) KL divergence between MD-f_1 in other words residual mode-coupling contributions

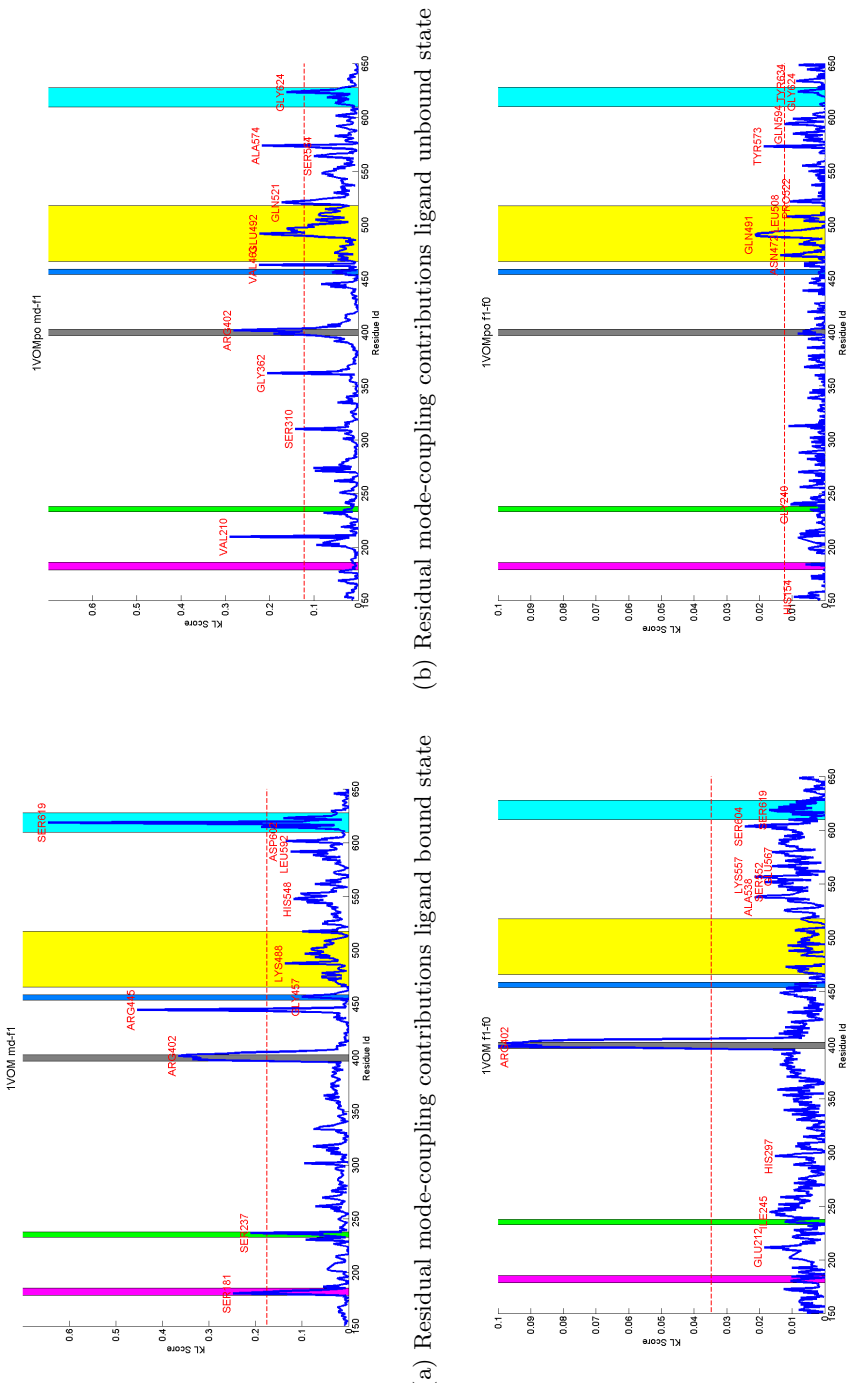


(b) Residual fluctuations of MD data

Figure 5.5: Mode-coupling contributions compared with fluctuations of the residues.

Ligand effect shown in 1VOM simulations, where switch I, switch II and p-loop structures both active. Simulation of the same protein structure without ligand also investigated.

Comparison of the ligand bound and unbound states shown in Fig.5.6. Significant differences between these figures occur in mode-coupling contribution Fig.5.6a and Fig.5.6b. Ligand binding event related with the ligand binding pocket which composed of switch I, switch II and p-loop [47]. All peaks around ligand binding sites are reduced on unbounded state, thought extra peaks emerged when compress in bound state. Ligand binding can also be considered as compressing some region in order to reach desired structure. In allostery this mechanism still an open question. However it is believed that proper ligand binding event suppress activation of some residues in order to activate other sites which is active in mechanism [9, 35, 45].



(a) Residual mode-coupling contributions ligand bound state (b) Residual mode-coupling contributions ligand unbound state

(c) Residual anharmonicity contributions ligand bound state (d) Residual anharmonicity contributions ligand unbound state

Figure 5.6: Mode-coupling and anharmonicity contributions compared for ligang bound and unbound states. Important structures shown with color codes p-loop (magenta), switch I(green), myopathy loop(gray), switch II(blue), relay structure(yellow), loop 2(cyan)

Simulation of the Myosin structure 1VOM prepared with 500fs interval of frames for 2ns length. Longer simulation with larger intervals also prepared which has 5ps intervals and 20ns length as explained in section 4. Comparison of these results for mode-coupling and anharmonicity effects shown in Fig.5.7.

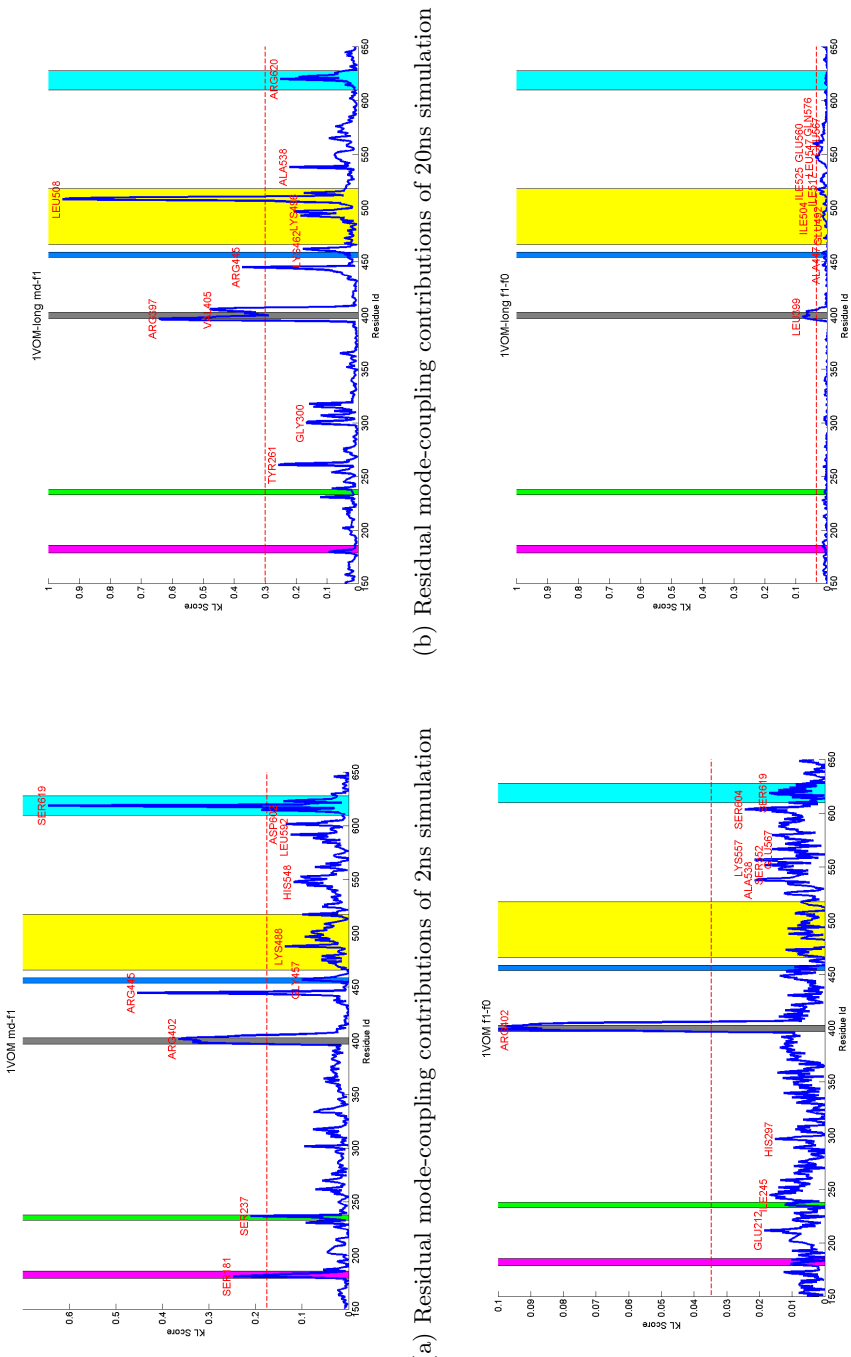
Longer simulation started to have a characteristics of transition between pre-power stroke (1VOM) stage to near-rigor (1MMA) state. During this transition structures like central ligand binding site, beta-sheet and relay helix changes in order to release ADP and inorganic phosphate and rotate lever arm. The major difference between 1VOM and 1MMA states is switch II position in which the movement of the switch II away from the ligand resulted very little change on Switch I. This change on switch I linked to change on other structures including P-loop, N terminal and converter domain [83]. In our KL score comparison in figures 5.7a and 5.7b, value corresponding to switch I region reduces below 3σ line. However, KL scores around switch I still have significant difference.

Release of the inorganic phosphate and ADP lead myosin head to bind actin with high affinity. Cleft closure changes switch I and ligand-binding pocket opens. Myopath loops has hydrophobic residues which is essential for the actin binding. Deletion and mutation experiments resulted lose of functionality myosin for acting binding. Loops in myosin II join large domains of the protein. The actin binding sites contains loop 2 structure which has an important role in regulating ligand release and actin affinity [47]. Actin binding creates rapid conformational changes and loop 2 rearranges its conformation [68]. Surface loops importance experimentally verified and interaction between loop 2 and negatively charged part of the actin [66]. Loop 2 cooperatively perform actin binding with the other surface loops and most likely to create actin binding interface that function triggering phosphate release [69].

A bent at the distal end of the long helix between residues 411-440 drives myopathy loop to bind to the surface of actin. Structure that drive myopathy loop starts with the structure corresponding one of our peak region Arg-445. Actin binding twists a large beta-sheet that forms myosin motor domain backbone. This central beta-sheet starts from switch I and ends residue 261. It also interacts with other beta-sheets in parallel starts residue 446 and ends at 462 which is beginning of relay structure. These beta-sheets are interacts with hydrogen bonds. Twist of the beta-sheets rotates all parallel sheets together because of the hydrogen

bonds. This twist changes relationship between switch I, swith II and P-loop in order to release of the ADP and P_i permanently [17, 43].

During this transition relay structure performs a special see-saw like motion. Peaks on relay structure in Fig.5.7b corresponds to starting (Lys-462) and ending (Lys-496) residues of relay helix. Relay helix structure slides on an axis residue at each side drives that motion. During this movement relay helix changes position in nanometer scale and relay loop moves significantly in order to support movement of relay helix [43]. One of this relay loop residue is emerged in our results in long simulation on Leu-508.



(a) Residual mode-coupling contributions of 2ns simulation

(b) Residual mode-coupling contributions of 20ns simulation

(c) Residual anharmonicity contributions of 2ns simulation

(d) Residual anharmonicity contributions of 20ns simulation

Figure 5.7: Mode-coupling and anharmonicity contributions compared for simulations with different length and interval length. Important structures shown with color codes p-loop (magenta), switch I(green), myopathy loop(gray), switch II(blue), relay structure(yellow), loop 2(cyan)

Comparison of the emerged peak values for different graphs visualized in the form of peak residues if the value of the peak is greater than the 3σ of the values. In Fig.5.8 peak values shown with color codes. Fluctuation or anharmonic contributions success rate is not close to mode-coupling results. Longer simulation and simulation without ligand also high precision compared to the anharmonicity and fluctuation results.

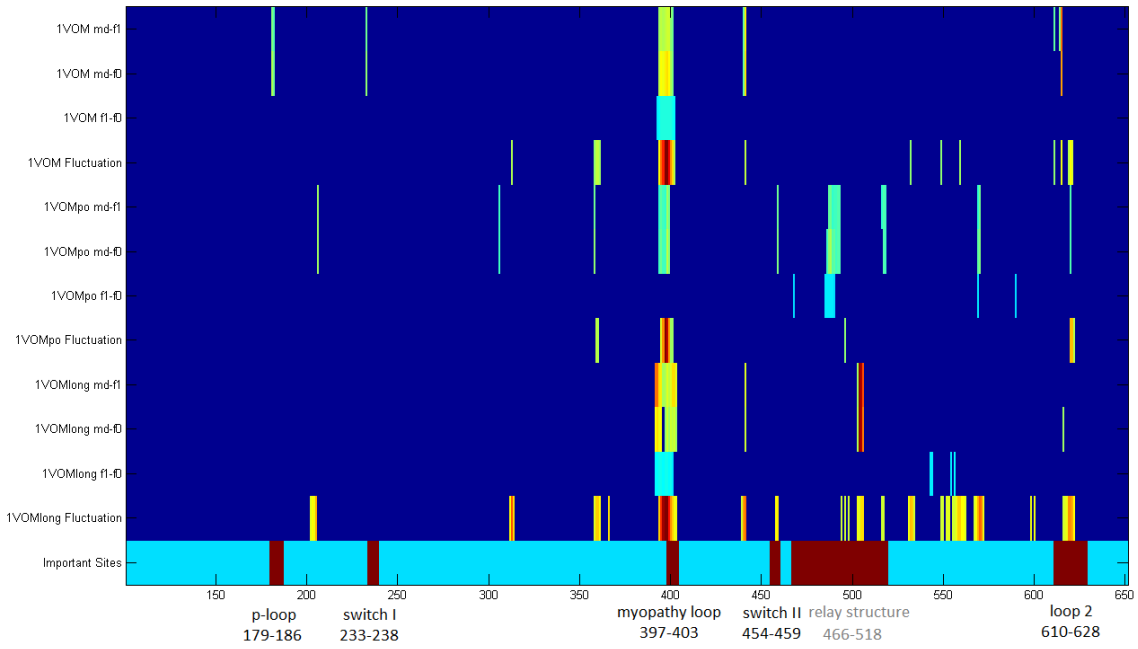


Figure 5.8: Residual contributions and fluctuation values. Values greater than three standart deviation which is statistically different from other residues are represented with color codes. Last line also correspond to the region where residues are part of the funvtionally important sites.

This figure summerized in a table to emphasize predictive power of mode-coupling contribution.

Table 5.1: Precision of KL scores for different MD simulation results

KL Score	Simulation		
	1VOM Precision	1VOMpo Precision	1VOMlong Precision
MD-f_1	$4/5 = 0.80$	$5/9 = 0.55$	$3/4 = 0.75$
f_1-f_0	$1/1 = 1.00$	$2/4 = 0.50$	$1/3 = 0.33$
Fluctuation	$4/10 = 0.4$	$2/4 = 0.5$	$6/16 = 0.38$

Chapter 6

CONCLUSION

In this thesis, we analyse modal space representations and fluctuations of the proteins and obtained information about the functionally important sites. Residual contribution of mode-coupling and anharmonicity systematically represented in this work.

Previous research approaches based on structural information obtained from crystallography studies. Dynamic models, however, uses harmonic spring-mass system in order to describe dynamics. This assumption is used in several research and still recognized as successful. In our study hermite expansion and expectations used in order to make corrections on basic harmonic model. Anharmonicity and mode-coupling effects taken into account in this approach. Free energy contribution of mode couples are known to be a successful tool [34]. Based on this modal correction technique fluctuations of residues for different models are compared.

Residual contribution of anharmonicity and mode-coupling effects are computed by measuring differences between distribution that our model generate and molecular dynamics data. Distributions are generated starting from fully harmonic model in which distribution of fluctuations are Gaussian. Random samples taken from this mode space composed of harmonic modes. Using these distributions and the hermite corrections for anharmonicity more realistic model for capturing anharmonic interactions are generated. Rejection sampling used in order to select samples randomly from new modal space which contains modes with anharmonic correction. These modal space samples tranforms into real space residue fluctuations and distribution of these fluctuations compared using KL divergence. KL divergence for every space coordinate is measured and one measure finally assigned for each residue.

We shown that, mode-coupling effects point functionally important residues with high precision. Myosin protein used as a test ground for this approach because of existence of experimental and theoretical studied. Results for pre-power stroke state of myosin

(PDB:1VOM) point significant residues span residues from switch I and p-loop in ligand-binding site and myopathy loop and loop 2 which are known to be essential on actin-binding. Molecular dynamics studies for 2ns and 20ns simulations and ligand extracted simulations both show significant predictably results using mode-coupling contributions.

Our result and previous results based on modal analysis shows promising results for different aspects of protein research. Allostery can be explained in terms of mode-couplings of residues and drug design techniques can be improved not only by searching topological matchings but also vibrational synchronization of the ligand and protein binding domain. Interactions based on vibrations enlighten energy transfer between residues and coordinative motion of the residues [35,45]. Traverse of the energy is assumed to be on conserved residues or shortest paths [16]. However there are also results for multiple energy pathways in which energy flows in different path with different quantity [13]. Interaction of the residues also form in different pattern in order to maximize these energy flow. Recent studies on network properties indicates the importance of clique residues [12, 75].

Other important development on this study can be achieved on understanding allosteric mechanism. Perturbation studies on molecular dynamics tries to understand excite on one site how to react from distance sites. General model can be produced based on modal space. It is known that major movements are achieved by slow modes contribution. However, energy transfer between slow modes and fast modes observed to generate major differences [9]. Suppression of modes can be used to control structural changes and estimation of these changes will be valuable tool in order to understand allostery and produce functional molecules.

Appendix A

GAUSSIAN NETWORK MODEL

Gaussian network model (GNM) is one of the topology based model. On this model contact maps used to determine adjacent residues and fluctuations of residues are determined Cartesian coordinates of the protein structure. In this model each C_α atom is modelled as an atom connected with springs to other adjacent atoms [2, 23].

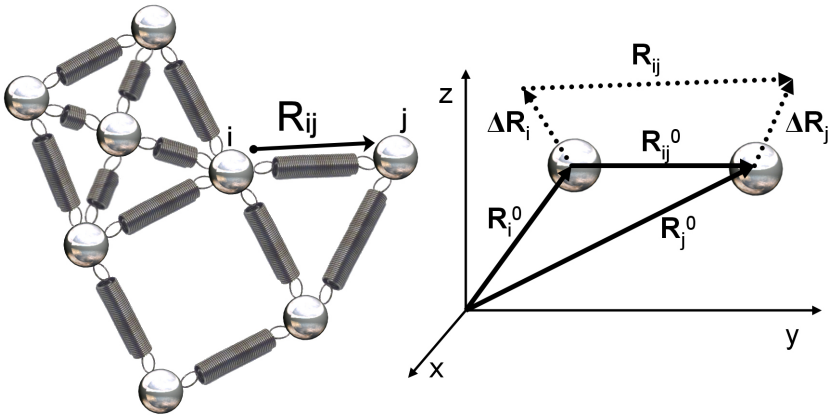


Figure A.1: Gaussian Network Model representation with mass-spring system [4].

A coarse grained GNM analysis based an harmonic oscillator potential energies. The position of the i^{th} residue is denoted by $\mathbf{R_i}$. Distance between residues i and j is denoted by R_{ij}, which is euclidean distance of two vector $R_{ij} = |\mathbf{R_i} - \mathbf{R_j}|$. Contact map of the proteins are 2D binary representations of the distances between residue pairs. Cut-off distance, r_{cutoff} for the residue pairs are commonly selected between 6-12$\mathring{A}$.

Displacement of the residues are detemined using positions at equilibrium of the atoms. For the residue i and j, position vectors are distance between this two residue is represented as $\mathbf{R_i^0}$, $\mathbf{R_j^0}$ and R_{ij}^0. Fluctuations vectors of residue i is given by $\mathbf{\Delta R} = \mathbf{R_i} - \mathbf{R_i^0}$.

Kirchhoff matrix of inter-residue contacts are given by $\mathbf{\Gamma}$ defined as followed in Equation A.1.

$$\Gamma_{ij} = \begin{cases} -1 & if\ i \neq j,\ R_{ij} \leq r_{cutoff} \\ 0 & if\ i \neq j,\ R_{ij} > r_{cutoff} \\ -\sum_{j,j\neq i}^{N} \Gamma_{ij} & if\ i = j \end{cases} \tag{A.1}$$

Molecular Dynamics simulations of proteins reveal fluctuation distributions of each residual from their mean coordinates. This fluctuations are transformed into modal space using principle component fluctuations with the covariance matrix **C**. Covariance matrix is generated as stated below

$$C = \left\langle \Delta\mathbf{R}\Delta\mathbf{R}^T \right\rangle \tag{A.2}$$

Using covariance matrix any fluctuation vectors $\Delta\mathbf{R}$ transform into modal space. Fluctuations in model space $\Delta\mathbf{r}$ denoted as

$$\Delta\mathbf{r} = C^{-\frac{1}{2}}\Delta\mathbf{R} \tag{A.3}$$

Component of the displacements in simulations can be easily decomposed into contributions of different modes by applying eigen decomposition of C matrix. Eigenvalues of C denoted as λ and eigenvector matrix that diagonalizes C represented as $\mathbf{e}$.

$$\left\langle \Delta\mathbf{R}\Delta\mathbf{R}^T \right\rangle^{-\frac{1}{2}} = diag\lambda^{-1/2}\mathbf{e}^T \tag{A.4}$$

The contribution of mode i can be easily generated by changing eigenvalues of all modes other than i to zero in eigenvalue vector λ. Modal coordinate values change within time and slow modes has long interval between two maxima as seen in Fig. A.2.

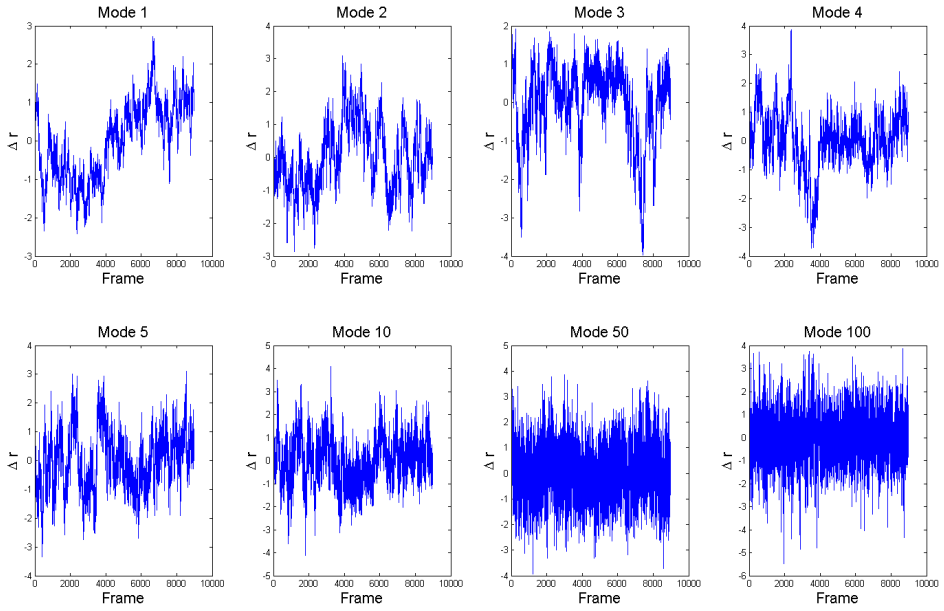


Figure A.2: Time plots of the slowest five modes and 10^{th},50^{th},100^{th} modes.

Probability distribution of modal coordinates tend to be a Gaussian distribution in theory. However simulation results show that PDFs of slow modes mostly non-gaussian and anharmonic interactions has effect in this picture. Overall behaviour of this PDFs converges a gaussian distributions with increasing mode numbers. First 10 modes modal coordinate distributions are shown in Fig. A.2

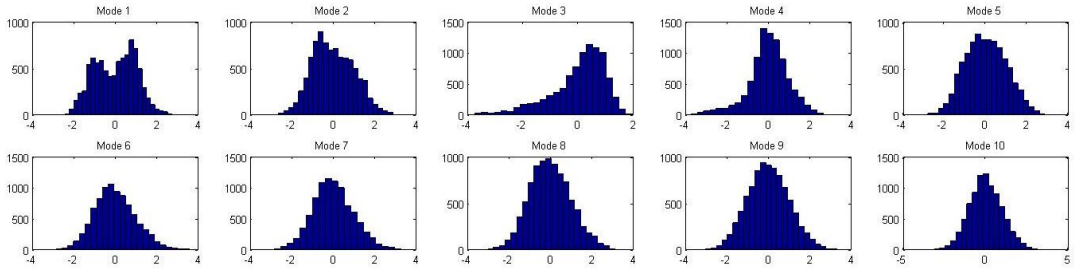


Figure A.3: Histogram of modal coordinates for first 10 modes of 1EJG.

Appendix B

RESIDUAL ENERGIES

Contribution of the fluctuation energy transferred between adjacent residues, which initially comes from interaction energies like electrostatic interactions, hydrogen and covalent bondings. Each residue contribute this energy transfer different portion. In GNM, it assumed that harmonic potentials are related with fluctuations of residues. Fluctuation of the energy ΔU_{ij} is related with the residue position fluctuations $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_i$. Thus, $\Delta U_{ij} = \gamma(\Delta\mathbf{R}_i - \Delta\mathbf{R}_j)^2$, where γ corresponding to the spring constant.

Molecules undergo many internal vibrational between residues. Potential energy stored in this internal energy tend to increase proteins heat capacity. Contributions of residue coupling to heat capacity means interaction between this pair of residue. Energy correlation for residues i and k stated as

$$\langle \Delta U_i \Delta U_k \rangle = \frac{1}{4}\gamma^2 \sum_j \sum_l C_{ij} C_{kl} \left\langle (\Delta\mathbf{R}_i - \Delta\mathbf{R}_j)^2 (\Delta\mathbf{R}_k - \Delta\mathbf{R}_l)^2 \right\rangle \tag{B.1}$$

Correlation given by Eq.B.1 proportional to contributions of residue i and k. Sum over index k of this equation leads to energetic interaction of residue i with all other residues.

$$\Delta U_i = \sum_k \langle \Delta U_i \Delta U_k \rangle \tag{B.2}$$

Correlation between residues and energy contribution of each residue shown in Fig.B.1. Residues with high coupling and energy contribution is essential for dynamics. Residues with high energy contributions appear along a path with one end located at surface and the other located at ligand binding site [75].

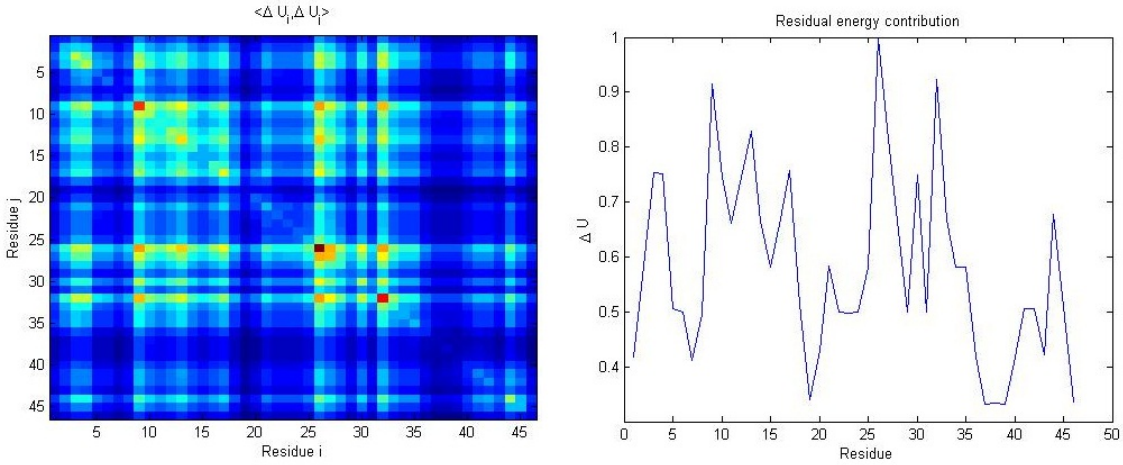


Figure B.1: Correlation between residue pairs shown on the right figure. Energy contributions of each residue shown on the left and energy peaks represents essential residues.

Appendix C

EVOLUTIONARILY CONSERVED RESIDUES

Conserved aminoacid sequences mostly required for the basic functions of the proteins. Sequence similarities are used as an evidence for the biological importance. Most highly conserved residues are observed as a active binding sites of enzymes, and the binding site of the proteins. Energy pathways of the allosteric proteins are known to have a residual path between ligand binding site and surface residue. This residual path contain evolutionarily conserved residues. Previous studies on myosin II shows that allosteric path hypothesis mostly contains evolutionary conserved residues [71].

Conservation scores of each residue generated by ConSurfDB webserver [20]. Conservation scores are sorted between 1 and 9. In other words, sorted from least conserved to most conserved. There are also some residues denoted with conservation score zero, which corresponds to insufficient data.

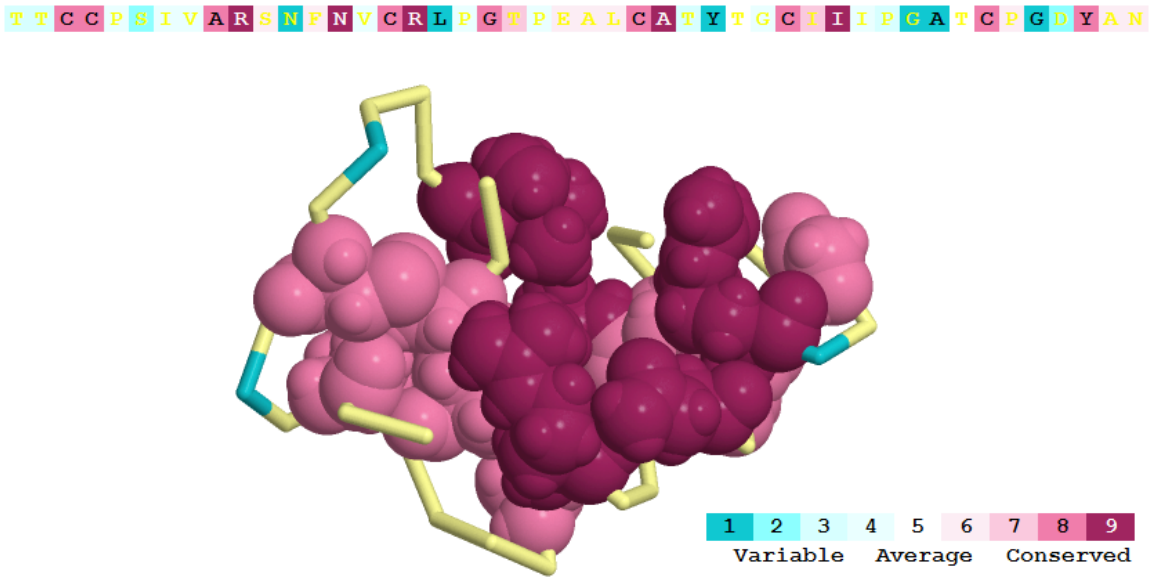


Figure C.1: Conservation score of 1EJG and sequence of residues coloured by their conservation scores. Residues with conservation score higher than 7 represented as ball models other represented as sticks.

BIBLIOGRAPHY

[1] A. Atilgan. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*, 80(1):505–515, January 2001.

[2] Ivet Bahar, Ali R. Atilgan, and Burak Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, 2(3):173–181, June 1997.

[3] Ivet Bahar, Timothy R. Lezon, Lee-Wei W. Yang, and Eran Eyal. Global dynamics of proteins: bridging between structure and function. *Annual review of biophysics*, 39(1):23–42, June 2010.

[4] Ahmet Bakan. Gaussian network model. $http://en.wikipedia.org/wiki/File:MassSpringNetwork.jpg$.

[5] Cary B. Bauer, Hazel M. Holden, James B. Thoden, Robert Smith, and Ivan Rayment. X-ray Structures of the Apo and MgATP-bound States ofDictyostelium discoideum Myosin Motor Domain. *Journal of Biological Chemistry*, 275(49):38494–38499, December 2000.

[6] Bernard. *Density Estimation for Statistics and Data Analysis (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)*. Chapman and Hall/CRC, 1 edition, April 1986.

[7] C.M. Bishop. *Pattern Recognition And Machine Learning*. Information Science and Statistics. Springer, 2006.

[8] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis,

J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The Biomolecular Simulation Program. *JOURNAL OF COMPUTATIONAL CHEMISTRY*, 30(10, Sp. Iss. SI):1545–1614, JUL 30 2009.

[9] M. Cecchini, A. Houdusse, and M. Karplus. Allosteric Communication in Myosin V: From Small Conformational Changes to Large Directed Movements. *PLoS Comput Biol*, 4(8):e1000129+, August 2008.

[10] S. M. Cowell, Y. S. Lee, J. P. Cain, and V. J. Hruby. Exploring Ramachandran and chi space: conformationally constrained amino acids and peptides in the design of bioactive polypeptide ligands. *Current medicinal chemistry*, 11(21):2785–2798, November 2004.

[11] Ian W. Davis, Andrew Leaver-Fay, Vincent B. Chen, Jeremy N. Block, Gary J. Kapral, Xueyi Wang, Laura W. Murray, W. Bryan Arendall, Jack Snoeyink, Jane S. Richardson, and David C. Richardson. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research*, 35(Web Server issue):W375–W383, July 2007.

[12] Dhruba Deb, Saraswathi Vishveshwara, and Smitha Vishveshwara. Understanding Protein Structure from a Percolation Perspective. *Biophys J*, 97(6):1787–1794, September 2009.

[13] Antonio del Sol, Chung-Jung Tsai, Buyong Ma, and Ruth Nussinov. The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways. *Structure*, 17(8):1042–1050, August 2009.

[14] Matthias Dietzen, Elena Zotenko, Andreas Hildebrandt, and Thomas Lengauer. On the Applicability of Elastic Network Normal Modes in Small-Molecule Docking. *J. Chem. Inf. Model.*, February 2012.

[15] Roberto Dominguez, Yelena Freyzon, Kathleen M. Trybus, and Carolyn Cohen. Crystal Structure of a Vertebrate Smooth Muscle Myosin Motor Domain and Its Complex with the Essential Light Chain. *Cell*, 94(5):559–571, September 1998.

[16] Stefan Fischer, Kenneth W. Olsen, Kwangho Nam, and Martin Karplus. Unsuspected pathway of the allosteric transition in hemoglobin. *Proceedings of the National Academy of Sciences*, 108(14):5608–5613, April 2011.

[17] Jeremiah J. Frye, Vadim A. Klenchin, Clive R. Bagshaw, and Ivan Rayment. Insights into the Importance of Hydrogen Bonding in the γ-Phosphate Binding Pocket of Myosin: Structural and Functional Studies of Serine 236,. *Biochemistry*, 49(23):4897–4907, May 2010.

[18] Kim C. Giese and James A. Spudich. Phenotypically Selected Mutations in Myosin's Actin Binding Domain Demonstrate Intermolecular Contacts Important for Motor Function†. *Biochemistry*, 36(28):8465–8473, July 1997.

[19] Nicholas M. Glykos. Software news and updates carma: A molecular dynamics analysis program. *J. Comput. Chem.*, 27(14):1765–1768, 2006.

[20] Ofir Goldenberg, Elana Erez, Guy Nimrod, and Nir Ben-Tal. The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Research*, 37(suppl 1):D323–D327, January 2009.

[21] Kosi Gramatikoff. A (corrected) chart showing the relationship between codons and amino acids. $http://en.wikipedia.org/wiki/File:GeneticCode21.svg$.

[22] Andrew M. Gulick, Cary B. Bauer, James B. Thoden, and Ivan Rayment. X-ray Structures of the MgADP, MgATPγS, and MgAMPPNP Complexes of the Dictyostelium discoideum Myosin Motor Domain†,‡. *Biochemistry*, 36(39):11619–11628, September 1997.

[23] Turkan Haliloglu, Ivet Bahar, and Burak Erman. Gaussian Dynamics of Folded Proteins. *Physical Review Letters*, 79(16):3090–3093, October 1997.

[24] Rhoda J. Hawkins and Tom C. McLeish. Coupling of global and local vibrational modes in dynamic allostery of proteins. *Biophysical journal*, 91(6):2055–2062, September 2006.

[25] S. Hayward and H. J. Berendsen. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins*, 30(2):144–154, February 1998.

[26] S. Hayward, A. Kitao, and N. Go. Harmonic and anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis. *Protein science : a publication of the Protein Society*, 3(6):936–943, June 1994.

[27] A. Houdusse, A. G. Szent-Györgyi, and C. Cohen. Three conformational states of scallop myosin S1. *Proceedings of the National Academy of Sciences*, 97(21):11238–11243, October 2000.

[28] Anne Houdusse, Vassilios N. Kalabokis, Daniel Himmel, Andrew G. Szent-Györgyi, and Carolyn Cohen. Atomic Structure of Scallop Myosin Subfragment S1 Complexed with MgADP. *Cell*, 97(4):459–470, May 1999.

[29] Anne Houdusse and H. Lee Sweeney. Myosin motors: missing structures and hidden springs. *Current Opinion in Structural Biology*, 11(2):182–194, April 2001.

[30] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[31] MBI Info. Myosin. $https://sites.google.com/a/mechanobio.info/mbinfo/Home/glossary-of-terms/mechano-glossary--m/mechano-glossary-myosin$.

[32] Kohji Ito, Taro Q. P. Uyeda, Yoshikazu Suzuki, Kazuo Sutoh, and Keiichi Yamamoto. Requirement of Domain-Domain Interaction for Conformational Change and Functional ATP Hydrolysis in Myosin. *Journal of Biological Chemistry*, 278(33):31049–31057, August 2003.

[33] V. Jayaraman, K. R. Rodgers, I. Mukerji, and T. G. Spiro. Hemoglobin allostery: resonance Raman spectroscopy of kinetic intermediates. *Science*, 269(5232):1843–1848, September 1995.

[34] A. Kabakçıoğlu, D. Yuret, M. Gür, and B. Erman. IOPscience - Anharmonicity, mode-coupling and entropy in a fluctuating native protein. *Physical Biology*, 7(4):046005+, December 2010.

[35] Maja Kobus, Phuong H. Nguyen, and Gerhard Stock. Coherent vibrational energy transfer along a peptide helix. *The Journal of Chemical Physics*, 134(12):124518+, 2011.

[36] Sampath Koppole, Jeremy C. Smith, and Stefan Fischer. The structural coupling between ATPase activation and recovery stroke in the myosin II motor. *Structure (London, England : 1993)*, 15(7):825–837, July 2007.

[37] D. E. Koshland, G. Némethy, and D. Filmer. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5(1):365–385, January 1966.

[38] Linderstrom-Lang KU. *Proteins and Enzymes*. Lane Medical Lectures, Stanford University Publications, University Series, Medical Sciences, vol. 6, Stanford University, 1952.

[39] S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[40] David M. Leitner. Energy Flow in Proteins. *Annual Review of Physical Chemistry*, 59(1):233–259, 2008.

[41] R. W. Lymn and E. W. Taylor. Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry*, 10(25):4617–4624, December 1971.

[42] Cheng-Wei Ma, Zhi-Long Xiu, and An-Ping Zeng. Discovery of Intramolecular Signal Transduction Network Based on a New Protein Dynamics Model of Energy Dissipation. *PLoS ONE*, 7(2):e31529+, February 2012.

[43] András Málnási-Csizmadia, Jane L. Dickens, Wei Zeng, and Clive R. Bagshaw. Switch movements and the myosin crossbridge stroke. *Journal of Muscle Research and Cell Motility*, 26(1):31–37, February 2005.

[44] J. MONOD, J. WYMAN, and J. P. CHANGEUX. On the nature of allosteric transitions: A plausible model. *Journal of molecular biology*, 12:88–118, May 1965.

[45] Kei Moritsugu, Osamu Miyashita, and Akinori Kidera. Vibrational Energy Transfer in a Protein Molecule. *Physical Review Letters*, 85(18):3970–3973, October 2000.

[46] Kei Moritsugu, Osamu Miyashita, and Akinori Kidera. Temperature Dependence of Vibrational Energy Transfer in a Protein Molecule. *J. Phys. Chem. B*, 107(14):3309–3317, March 2003.

[47] C.T. Murphy and J.A. Spudich. Variable surface loops and myosin activity: accessories to a motor. *Journal of muscle research and cell motility*, 21(2):139–151, 2000.

[48] Hirofumi Onishi and Manuel F. Morales. A closer look at energy transduction in muscle. *Proceedings of the National Academy of Sciences*, 104(31):12714–12719, July 2007.

[49] E. Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

[50] B. Patterson, K. M. Ruppel, Y. Wu, and J. A. Spudich. Cold-sensitive mutants G680V and G691C of Dictyostelium myosin II confer dramatically different biochemical defects. *The Journal of biological chemistry*, 272(44):27612–27617, October 1997.

[51] B. Patterson and J. A. Spudich. Cold-sensitive mutations of Dictyostelium myosin heavy chain highlight functional domains of the myosin motor. *Genetics*, 143(2):801–810, June 1996.

[52] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, 26(16):1781–1802, December 2005.

[53] Francesco Piazza and Yves-Henri Sanejouand. Long-range energy transfer in proteins. *Physical Biology*, 6(4):046014+, November 2009.

[54] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, July 1963.

[55] I. Rayment, W. R. Rypniewski, K. Schmidt-Base, R. Smith, D. R. Tomchick, M. M. Benning, D. A. Winkelmann, G. Wesenberg, and H. M. Holden. Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science*, 261(5117):50–58, July 1993.

[56] Thomas F. Reubold, Susanne Eschenburg, Andreas Becker, F. Jon Kull, and Dietmar J. Manstein. A structural model for actin-induced nucleotide release in myosin. *Nature structural biology*, 10(10):826–830, October 2003.

[57] Jane Shelby Richardson. Protein backbone dihedral angles phi, psi, and omega. *http* : $//en.wikipedia.org/wiki/File : Protein - backbone - PhiPsiOmega - drawing.jpg$.

[58] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pages 832–837, 1956.

[59] K. M. Ruppel and J. A. Spudich. Structure-function studies of the myosin motor domain: importance of the 50-kDa cleft. *Molecular biology of the cell*, 7(7):1123–1136, July 1996.

[60] Maniatis T. Sambrook J., Fritsch E. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York, USA, 1989.

[61] N. Sasaki, T. Shimada, and K. Sutoh. Mutational analysis of the switch II loop of Dictyostelium myosin II. *The Journal of biological chemistry*, 273(32):20334–20340, August 1998.

[62] Naoya Sasaki, Reiko Ohkura, and Kazuo Sutoh. Dictyostelium Myosin II Mutations That Uncouple the Converter Swing and ATP Hydrolysis Cycle†. *Biochemistry*, 42(1):90–95, December 2002.

[63] T. Shimada, N. Sasaki, R. Ohkura, and K. Sutoh. Alanine scanning mutagenesis of the switch I region in the ATPase site of Dictyostelium discoideum myosin II. *Biochemistry*, 36(46):14037–14043, November 1997.

[64] Lars Skjaerven, Aurora Martinez, and Nathalie Reuter. Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins*, 79(1):232–243, 2011.

[65] Clyde A. Smith and Ivan Rayment. X-ray Structure of the Magnesium(II)·ADP·Vanadate Complex of the Dictyostelium discoideum Myosin Motor Domain to 1.9 Å Resolution†,‡. *Biochemistry*, 35(17):5404–5417, January 1996.

[66] J. A. Spudich. How molecular motors work. *Nature*, 372(6506):515–518, December 1994.

[67] Gurol M. Suel, Steve W. Lockless, Mark A. Wall, and Rama Ranganathan. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Structural & Molecular Biology*, 10(1):59–69, December 2002.

[68] M. Sun, M.B. Rose, S.K. Ananthanarayanan, D.J. Jacobs, and C.M. Yengo. Characterization of the pre-force-generation state in the actomyosin cross-bridge cycle. *Proceedings of the National Academy of Sciences*, 105(25):8631, 2008.

[69] H. Lee Sweeney and Anne Houdusse. Structural and functional insights into the Myosin motor mechanism. *Annual review of biophysics*, 39:539–557, June 2010.

[70] F. Tama and Y. H. Sanejouand. Conformational change of proteins arising from normal mode calculations. *Protein engineering*, 14(1):1–6, January 2001.

[71] Susan Tang, Jung-Chi C. Liao, Alexander R. Dunn, Russ B. Altman, James A. Spudich, and Jeanette P. Schmidt. Predicting allosteric communication in myosin via a pathway of conserved residues. *Journal of molecular biology*, 373(5):1361–1373, November 2007.

[72] Daniel Ting, Guoli Wang, Maxim Shapovalov, Rajib Mitra, Michael I. Jordan, and Roland L. Dunbrack. Neighbor-Dependent Ramachandran Probability Distributions

of Amino Acids Developed from a Hierarchical Dirichlet Process Model. *PLoS Comput Biol*, 6(4):e1000763+, April 2010.

[73] Monique M. Tirion. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters*, 77(9):1905–1908, August 1996.

[74] Georgios Tsiavaliaris, Setsuko Fujita-Becker, Renu Batra, Dmitrii I. Levitsky, F. Jon Kull, Michael A. Geeves, and Dietmar J. Manstein. Mutations in the relay loop region result in dominant-negative inhibition of myosin II function in Dictyostelium. *EMBO reports*, 3(11):1099–1105, November 2002.

[75] Ceren Tuzmen and Burak Erman. Identification of Ligand Binding Sites of Proteins Using the Gaussian Network Model. *PLoS ONE*, 6(1):e16474+, January 2011.

[76] Ronald D. Vale and Ronald A. Milligan. The Way Things Move: Looking Under the Hood of Molecular Motor Proteins. *Science*, 288(5463):88–95, April 2000.

[77] Mariana Ruiz Villarreal. Main protein structures levels. *http* : $//en.wikipedia.org/wiki/File : Main - protein - structure - levels - en.svg$.

[78] A. C. Wallace, R. A. Laskowski, and J. M. Thornton. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering*, 8(2):127–134, February 1995.

[79] O.N. Yogurtcu, M. Gur, and B. Erman. Statistical thermodynamics of residue fluctuations in native proteins. *J. Chem. Phys.*, 130(9), 2009.

[80] Haibo Yu, Liang Ma, Yang Yang, and Qiang Cui. Mechanochemical Coupling in the Myosin Motor Domain. II. Analysis of Critical Residues. *PLoS Comput Biol*, 3(2):e23+, February 2007.

[81] Faming Zhang, Margret B. Basinski, John M. Beals, Stephen L. Briggs, Lisa M. Churgay, David K. Clawson, Richard D. DiMarchi, Thomas C. Furman, John E. Hale, Hansen M. Hsiung, Brigitte E. Schoner, Dennis P. Smith, Xing Y. Zhang, Jean-Pierre

Wery, and Richard W. Schevitz. Crystal structure of the obese protein Ieptin-E100. *Nature*, 387(6629):206–209, May 1997.

[82] Wenjun Zheng and Bernard Brooks. Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. *Journal of molecular biology*, 346(3):745–759, February 2005.

[83] Wenjun Zheng and D. Thirumalai. Coupling between normal modes drives protein conformational dynamics: illustrations using allosteric transitions in myosin II. *Biophysical journal*, 96(6):2128–2137, March 2009.

VITA

Onur Varol was born in Izmir, Turkey on July 12, 1988. After graduating from Karsiyaka Anatolian High School in 2006, he got admitted Istanbul Technical University where he received his B.Sc degree in Electronics and Physics Engineering degrees respectively in 2010 and 2012. After graduating from Electronics Engineering program he enrolled M.Sc program Computer Science and Engineering at Koc University in 2010.

He enrolled PhD program in Informatics specialized in Complex System track at Indiana University, Bloomington IN in 2012.