A COMPUTATIONAL APPROACH FOR ANALYSIS OF COMMUNICATION IN PROTEINS

by

Sinem Özel

B.S., Chemical Engineering, Boğaziçi University, 2005

Submitted to the Institude for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering
Boğaziçi University
2007

ACKNOWLEDGEMENTS

I would like to express my special gratitude to my thesis advisor, Prof. Dr. Pemra Doruker Turgut, for her mentorship, guidance and support. Her encouragement helped me throughout the research and writing of this thesis.

I owe my sincere gratitude to Prof. Dr. Ivet Bahar and Dr. Chakra Chennubhotla, who gave me the opportunity to work with them in the Department of Computational Biology at University of Pittsburgh. Prof. Dr. Ivet Bahar's encouraging and personal guidance; Chakra Chennubhotla's insightful comments and constructive criticisms have been of great value for me.

I would also like to express my appreciation to Prof. Dr. Türkan Haliloğlu for her contribution to enrichment of my knowledge and perspective.

I am much indebted to my colleagues in Polymer Research Center, Özge Kürkçüoğlu, Nigar Kantarcı Çarşıbaşı, Arzu Uyar, Ayşegül Özen, Burcu Aykaç Fas, Nevra Özer, Sertan Cansu and my old collegue Uğur Emekli for their helps and kind attitudes. I also want to thank Başak Işın and Ahmet Bakan for their friendship and support in Pittsburgh.

I acknowledge the financial support of TUBITAK (BIDEB-2210 Fellowship) during my master studies.

The last but not the least, my gratitude is to my family; none of these would be possible without their love and patience. This thesis is dedicated to them.

ABSTRACT

A COMPUTATIONAL APPROACH FOR ANALYSIS OF COMMUNICATION IN PROTEINS

Allostery and communication in proteins are crucial for the regulation of protein functions. The binding of a ligand changes the affinity of the protein at a distal ligand binding site and leads to conformational changes which are important for its function. In other words, the functional motions of the proteins reveal the communication patterns that are inherent to native architecture. The discrete-time, discrete-state Markov process applied on the network of interacting residues determines the potential pathways of signal transduction. This novel method involves the evaluation of two basic quantities: hitting and commute times, which express the information flow between the residue pairs. Furthermore maximum likelihood communication pathways are computed based on the Markov transition probabilities. This methodology is applied to adenylate kinase, triosephosphate isomerase, PDZ signaling protein and Cdc25B. As a result, the key interactions and the communication ability of the residues in the selected networks are ascertained. It is shown that the catalytic residues are located at the minima of the mean commute time curves, which indicates their efficient signal transduction abilities. Mobile residues are found to be slow communicators and the residues in the protein core are illustrated to be efficient communicators of the network. Besides, the residues on the maximum likelihood pathways are, at large, evolutionarily conserved; this issue confirms the functional importance of the residues on the pathways. In summary, the new methodology provides insights into understanding of communication patterns in proteins.

ÖZET

PROTEİNLER İÇİ HABERLEŞMENİN İNCELENMESİ İÇİN HESAPSAL BİR YAKLAŞIM

Allosteri ve protein içi haberleşme protein işlevinin düzenlenmesi için çok önemlidir. Ligantın bağlanması proteinin bağlantı bölgesinden uzak bir bölgenin afinitesini değiştirmekte ve işlevi için önemli konformasyon değişiklikleri ile sonuçlanmaktadır. Bir başka deyişle, proteinin işlevsel hareketleri doğal yapısına özgü olan iletişim modellerini açığa vurmaktadır. Birbirleriyle etkileşen residülerin oluşturduğu ağlara uygulanan kesikli-zaman, kesikli-durum Markov prosesi muhtemel yolakları belirlemektedir. Bu yeni yöntem residü ikilileri arasındaki haber akışını ifade eden iki temel niceliği: vurma zamanı ve gidip-gelme zamanının değerlendirilmesini içermektedir. Ayrıca yüksek olasılıklı iletişim yolakları Markov geçiş olasılıklarına dayandırılarak hesaplanmaktadır. Bu yöntem adenilat kinaz, triosefosfat izomeraz, PDZ sinyal proteini ve Cdc25B'ye uygulanmıştır. Sonuç olarak, seçilen ağ yapıları içerisindeki önemli etkileşmeler ve residülerin haberleşme yetenekleri belirlenmistir. Katalitik residülerin ortalama gidip-gelme zaman eğrilerinin minimum noktalarında bulunduğu gösterilmiştir. Bu durum, katalitik residülerin etken sinyal transdüksiyon yeteneklerinin bir göstergesidir. Hareketli residülerin yavaş haberci oldukları ve proteinin merkezinde bulunan residülerin etken haberci oldukları görülmüştür. Yüksek olasılıklı yolaklar üzerindeki residüler genelde evrimsel olarak korunmuştur, bu sonuç yolaklar üzerindeki bu residülerin işlevsel önemini teyit etmektedir. Özetle, bu yeni yöntem protein içi iletişim modellerinin anlaşılmasına ışık tutmaktadır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

A	Affinity matrix
a_{ij}	Affinity between residue i and j
C(i,j)	Commute time between residue i and j
<C(i)>	Mean commute time for residue i
D	Degree matrix
d_{eff}	Effective/Commute distance
d_i	Local interaction density at residue i
H(j,i)	Hitting time from residue i to j
k_B	Boltzman constant
l	Length of one step
M	Markov transition matrix
m_{ij}	Markov transition probability between residue i and j
N	Number of atom-atom contacts
n	Number of steps
p	Probability distribution
R_i	Position vector of residue i
ΔR_i	Displacement vector for R_i
ΔR_{ij}	Magnitude of the vector ΔR_i-ΔR_j
T	Temperature
u_p	p^{th} eigenvector of Γ matrix
α	Alpha helix
β	Beta strand
γ	Force constant
Γ	Kirchhoff/Connectivity matrix
λ_p	p^{th} eigenvalue of Γ matrix
μ	Mean
π	Stationary distribution
σ	Standard deviation

AK	Adenylate Kinase
AKeco	Adenylate Kinase from Escherichia coli
Ala	Alanine
AMP	Adenosine monophosphate
ANM	Anisotropic Network Model
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ATD	Anisotropic Thermal Diffusion
ATP	Adenosine triphosphate
Cdc25B	Cell division cycle 25 homolog B
Cys	Cysteine
cTIM	Chicken Triosephosphate Isomerase
DHAP	Dihydroxyacetone phosphate
EM	Electron Microscopy
GAP	Glyceraldehyde 3-phosphate
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GNM	Gaussian Network Model
His	Histidine
hTIM	Human Triosephosphate Isomerase
Ile	Isoleucine
KNF	Koshland-Nemethy-Filmer
Leu	Leucine
Lys	Lysine
MC	Monte Carlo
MD	Molecular Dynamics
Met	Methionine
MWC	Monod-Wyman-Changeux
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank

PDZ	Postsynaptic Density-95/Discs large/Zonula occludens-1
PGH	Phosphoglycolohydroxamate
Phe	Phenylalanine
Pro	Proline
RMSD	Root mean square deviation
SCA	Statistical Coupling Analysis
Ser	Serine
std	Standard deviation
Thr	Threonine
TIM	Triosephosphate Isomerase
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

1. INTRODUCTION

Proteins are essential parts of all living organisms participating in cellular processes. Cells can rapidly fine-tune or regulate the activity of a certain key protein using other molecules that bind to and distort it. Allostery ('other space') indicates that the change in one site results in a functional change at a second distal site. Several studies inform on the fundamental role of allosteric communication in achieving the biological functions of proteins. (Changeux and Edelstein, 2005)

The binding of a ligand to a specific region of a protein often triggers large conformational changes in a distant part. (Monod *et al.*, 1965) Thus, it is denoted that dynamics and allosteric processes are almost linked. It has also been postulated that proteins, which are engaged to large conformational changes, are potentially allosteric. (Gunasekaran *et al.*, 2004)

Allostery and communication in proteins for regulation of protein functions and signal transduction have been important topics for more than four decades. (Swain and Gierasch, 2006) In recent studies, there have been significant endeavors to identify the communication pathways in protein structures. It has been revealed that clusters of co-evoluted residues might be important in intramolecular communications. (Lockless and Ranganathan, 1999; Suel et al., 2003). Furthermore, it has been put forth that the propagation of heat, in terms of kinetic energy, might also identify allosteric pathways. (Ota and Agard, 2005) Recently, a systematic computational method has been developed to derive the potential intra- and inter molecular pathways of signal transduction based on the Markov process. (Chennubhotla and Bahar, 2006) This methodology elucidates the potential pathways for allosteric communication favored by the complex molecular architecture.

In the present study, the novel methodology developed and being improved by Chennubhotla and Bahar, is applied to seven proteins from four different families. The structures are modeled as networks of interacting residues and the topology-driven 'information' diffusion is investigated by the Markov process initiated on the networks.

The key residues that are crucial for the communication stochastics are identified and their functional importances are examined. The shortest (maximum likelihood) paths which might be crucial in biological function of the proteins are computed according to the atom-atom contacts between the residues. A comparative investigation has also been carried out to display the changes in communication stochastics upon ligand/substrate binding. Thus, this novel methodology provides insights into the molecular basis of communication mechanisms in biomolecular systems.

2. PROTEIN DYNAMICS AND COMMUNICATION PATTERNS

Proteins are macromolecules that are composed of various combinations of 20 major amino acids. The properties of a protein depend mainly on its shape, which consecutively depends on the linear arrangement of the amino acid that make up the protein. All amino acids have a central carbon atom (Cα) to which a carboxyl group (-COOH), an amino group (-NH_2), a hydrogen and a side chain (-R) are bonded. The side chain of the amino acids can vary in size from just a hydrogen atom (glycine) to a large heterocyclic group (tryphtophan). Depending on the chemical nature of the side chains, amino acids possess different chemical properties (i.e. hydrophobic, polar, etc.).

Table 2.1. Three-letter (Code 1) and one-letter (Code 2) codes of 20 amino acids

Name	**Code 1**	**Code 2**	**Name**	**Code 1**	**Code 2**
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic Acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamine	Gln	Q	Serine	Ser	S
Glutamic Acid	Glu	E	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

2.1. Four Levels of Protein Structure

Structures of proteins are explored according to four major groups: primary, secondary, tertiary and quaternary structures. The arrangement of amino acids along the linear polypeptide chain is the ‘primary’ structure of the protein. It has been pointed out that hydrophobic groups of amino acids are mostly buried in the interior and hydrophilic groups lie on the surface of the protein. The local spatial conformation of the polypeptide’s

backbone, excluding the constituent amino acid's side chain, is defined as the 'secondary' structure. Secondary structure mainly occurs as α-helices or β-strands; on the other hand certain amino acid sequences favor formation of loop regions. The 'tertiary' structure is the folded form of the single protein chain which is the functional form. Finally, the shapes or structures of large assemblies of proteins specify the quaternary structure.

2.2. Structure-Function Relationship

The amino acid sequences specify the three-dimensional structures, which in turn determine how the protein functions: '*sequence* → *structure* → *function*'. Thus, understanding how the three-dimensional structures, which are encoded in the amino acid sequences, is an important step towards understanding the functions of biomolecules. There are several methods with varying resolution for determining the structures of proteins. The tertiary structures which require comprehensive information about the arrangements of the atoms can be determined using X-ray crystallography or NMR spectroscopy. These methods have been developed to obtain three-dimensional models of small protein molecules. The quaternary structure of large proteins or aggregates, i.e. virus particles, ribosome, or gap junctions, can be determined by electron microscopy (EM). (Branden and Tooze, 1999)

2.3. Protein Dynamics

Proteins are not entirely rigid structures; they are composed of a linear sequence of covalently bound amino acids that intrinsically have tolerance for rotations. Each protein folds into a conformation of lowest energy and displays local conformational changes around this so-called 'native' state.

The linkage between structure and function apparently lies in dynamics. Molecular Dynamics (MD) simulations (Verlet, 1967) provide a straightforward approach to elucidate the microscopic details of protein motion as well as their functional significance. However, this method is of limited applicability due to the huge computational effort involved. Monte Carlo (MC) simulations (Cao and Berne, 1990) and knowledge-based algorithms (Sudarsanam *et al.*, 1995; Bahar *et al.*, 1997a; Haliloglu and Bahar, 1998; Deane and

Blundell, 2000) are the other approaches that are also being used in sampling protein conformations.

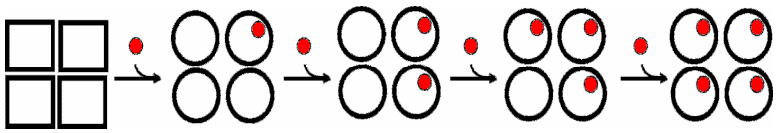
Under native state conditions, protein motions can vary from side-chain conformational changes (local) to collective domain motions (global). Investigating the fluctuation dynamics of proteins around the native state gives insight about the molecular basis and mechanisms of their function. (Bahar and Rader, 2005; Ma 2005) In particular, Normal mode analysis (NMA) and elastic network models (i.e. Gaussian Network Model (GNM), Anisotropic Network Model (ANM), etc.) (Tirion, 1996; Bahar *et al.*, 1997b; Haliloglu *et al.*, 1997; Atilgan *et al.*, 2001) are based on the harmonic approximation of the motion and yield the vibrational modes of a given structure. These modes represent an orthonormal vectorspace and any displacement can be delineated as a linear combination of such modes. Thus knowing the normal modes permits the explicit evaluation of all possible vibrational frequencies in a system. Besides, slow vibrational modes represent the collective motion of the proteins which are associated with their function. (Haliloglu *et al.*, 1997; Kovacks *et al.*, 2004; Lu and Ma, 2005; Chennubhotla *et al.*, 2005; Demirel and Keskin, 2005; Cui and Bahar, 2006) Elastic network models are conceptually simple, hence, are of great interest in exploration of protein dynamics. (McCammon and Harvey, 1994)

2.4. Communication Stochastics in Protein Networks

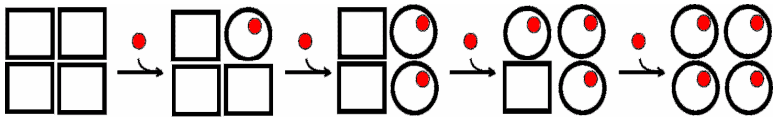
A central goal in structural biology is to elucidate the mechanism of allosteric activation in biomolecular systems. The term ‘allostery’ comes from the Greek words ‘allos’ and ‘stereos’, meaning ‘other’ and ‘space/location’, respectively. This term was introduced by Jacob and Monod in 1961. (Jacob and Monod, 1961) Allostery can be expressed as ‘action at a distance’, because it signifies the coupling of conformational changes between one functional site and another distal one.

Two models have been proposed to depict the allosteric transitions in multimeric proteins; Monod-Wyman-Changeux (MWC) model (Monod *et al.*, 1965) and Koshland-Nemethy-Filmer (KNF) model (Koshland *et al.*, 1966) which are illustrated in Figure 2.1. (Creighton, 1993) The former takes a concerted approach for granted, whereas the latter

appropriates a sequential approach. MWC presumes an all-or-none change in all subunits; however, KWC appropriates the sequential transition of individual subunits.



(a)



(b)

Figure 2.1. MWC and KNF models to explain allosteric transitions

The earliest example of allosteric regulation is the control of oxygen binding to hemoglobin by protons. Hemoglobin binds oxygen in the lungs and sticks very tightly to it. However, once it reaches parts of the body which is short of oxygen, it needs to release the oxygen. This is accomplished by an allosteric interaction between carbon dioxide and oxygen. The carbon dioxide dissolved in the blood produces carbonic acid. The presence of this acid leads to a conformational change in hemoglobin, causing it to release oxygen and take up carbon dioxide. (Hill, 1910)

The term allostery is, presently, used to indicate a broad spectrum of functional interactions. (Krusek, 2004) The interactions are usually induced by substrate binding and followed by the propagation of information transmitting the signals that trigger the functional responses, as it is in hemoglobin. It has been stated that exploring the changes in conformational distribution is important to understand the allosteric effects in proteins. (Gunasekaran *et al.*, 2004) The response to ligand binding can involve all forms of dynamic behavior, ranging from highly correlated low-frequency normal mode vibrations to random local anharmonic motions of individual atoms or groups. However, in some cases communication occurs without conformational change. The basic requirement for long-range communication has been postulated to be existence of atoms dispersed

throughout the protein which, directly or indirectly, experience the presence of ligands. (Cooper and Dryden, 1984)

Allostery is usually associated with multimeric proteins, in other word, monomeric proteins are usually assumed to be nonallosteric. However, there are monomeric proteins in which the modulations of the functional properties are carried out by allosteric effectors. (Ascenzi *et al.*, 2005) Sperm whale myoglobin, human serum albumin, and human-α-thrombin are the examples for these monomeric proteins.

Signal transduction through protein structure is the basis of the allostery and recently, several methodological advances have provided insights into the underlying molecular mechanisms. Ranganathan and coworkers devised a sequence-based-method to identify the networks of amino acid interactions in a protein. (Lockless and Ranganathan, 1999; Suel *et al.*, 2003) This method is based on the assumption that there is statistical covariance during evolution between the residues if the interaction between them is functionally important. It has been exposed that the relevant energetically coupled residues identified from statistical coupling analysis (SCA) forms a network for the propagation of signals in the protein.

A non-equilibrium simulation method, anisotropic thermal diffusion (ATD), is developed by Ota and Agard to follow the transduction of kinetic energy distal from the binding site. (Ota and Agard, 2005) This method also elucidates the physical mechanisms of an intramolecular signal transduction process. Nussinov and coworkers proposed a methodology to highlight the key residues based on the protein topology, they analyzed the robust communication system and determined the nodes, removal of which would lead to significant differences in the system. (del Sol *et al.*, 2006) Another study that evaluates the communication pathway is developed by Atilgan and coworkers. (Atilgan *et al.*, 2007) The weighted-residue-networks are constructed for proteins by using the Miyazawa and Jernigan (Miyazawa and Jernigan, 1996) or the Thomas and Dill (Thomas and Dill, 1996) knowledge-based potentials. Shortest paths in the constructed networks are computed by using the Dijkstra's algorithm (Dijkstra, 1959). It has also been revealed that the shortest path lengths and residue fluctuations are highly correlated.

A new methodology has been developed, by Chennubhotla and Bahar, where a Markov process is defined on atom-based elastic network. It has also been put forward that collective motions of a protein determine the communication patterns that are inherent to the native structure. (Chennubhotla and Bahar, 2006) This methodology gives insights about the topological basis of communication in proteins and design principles for efficient signal transduction. The present work includes the applications of this methodology to example proteins which are stated in the following section. This novel approach is explained in 'Methodology' section.

2.5. Materials

In this study, seven structures from different families (i.e. phosphotransferase, isomerase, hydrolase, peptide recognition) are investigated. Adenylate kinase (AK), triosephosphate isomerase (TIM), PDZ signaling protein and Cdc25B are chosen as the model systems for the analysis of the new methodology. Two structures from PDZ-3 protein and three structures from TIM enzyme are picked out to elucidate the changes in communication ability upon ligand binding. The file names (IDs) of the analyzed structures in Protein Data Bank (PDB) (Berman *et al.*, 2000) are 4AKE for adenylate kinase; 8TIM, 1TPH and 1HTI for triosephosphate isomerase, 1BE9 and 1BFE for PDZ, and 1QB0 for Cdc25B protein.

2.5.1. Adenylate Kinase

Adenylate kinase is an essential enzyme, responsible for recycling AMP in energetically active cells. It catalyzes the transfer of a phosphate group from adenosine triphosphate (ATP) to adenosine monophosphate (AMP). The X-ray structure of the apo-form of the AK from *E.coli* is determined to 2.2 Å resolution. (Muller *et al.*, 1996) The structure (PDB ID: 4AKE) is shown in Figure 2.2, where each domain is displayed in a different color: CORE (blue), AMP binding domain (orange), LID (green). CORE is the stable frame which is mainly preserved during catalysis; on the other hand, the domains AMPbd and LID execute large amplitude motions. AMP binds to AMPbd for the phosphorylation, whereas ATP is bound to the enzyme through the P-loop, GXXGXGK (residues Gly7-Lys13) in the CORE domain.

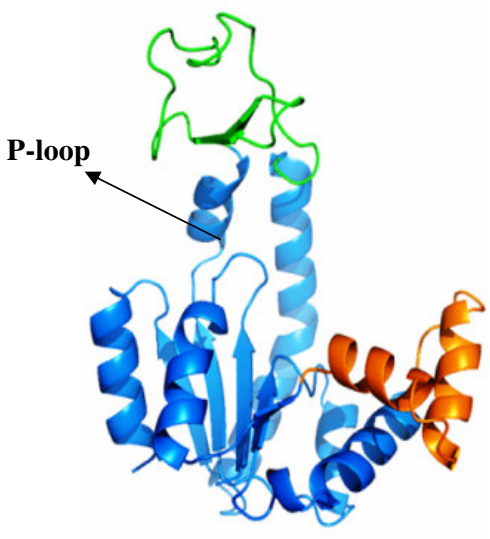


Figure 2.2. The cartoon representation of the free and complex forms of AKeco

2.5.2. Triosephosphate Isomerase

Triosephosphate isomerase (TIM) is an important glycolytic enzyme, catalyzing the interconversion of dihydroxyacetone phosphate (DHAP) into D-glyceraldehyde 3-phosphate (GAP). TIM is a dimer of identical subunits, each with a catalytic center. The X-ray structure of free form of chicken TIM (PDB ID: 8TIM) is determined by Phillips and coworkers at 2.5 Å resolution, the complex form with phosphoglycolohydroxamate (PGH), (PDB ID: 1TPH) is determined by Ringe and coworkers at 1.8 Å resolution. (Zhang *et al.*, 1994) The inhibitor (PGH, shown in Figure 2.3 (a)) is a structural analog of substrate DHAP (Figure 2.3 (b)).

(a) (b)

Figure 2.3. Structures of PGH (a) and DHAP (b)

The free and complex forms of chicken triosephosphate isomerase (cTIM) are shown in Figure 2.4, and are colored in light blue and light pink, respectively. The active

loop (loop 6) has two different conformations: *open* (dark blue), leaving the active site accessible to the solvent; and *closed* (dark pink), preventing the loss of the reaction intermediate during the catalysis. The inhibitor PGH (black) and the active site residues Lys13 (green), His95 (orange) and Glu165 (cyan) are also displayed on the structure.

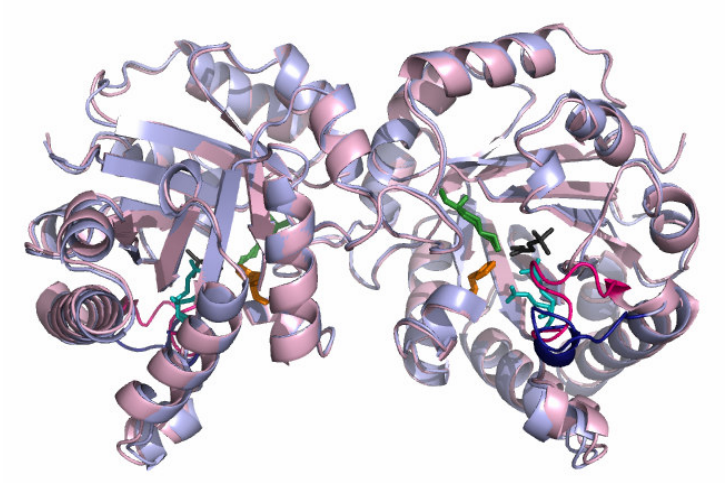


Figure 2.4. Free (light blue) and complex (light pink) forms of cTIM superimposed

Chicken and human TIMs, shown in Figure 2.5, are approximately 90% identical in sequence. The RMS difference between the common C-α atoms of the two structures is 0.69 Å. The flexible loops of the two structures do not have similar conformations. The loops in chicken TIM (8TIM) are open in both subunits. However, the flexible loop of subunit A in human TIM (PDB ID: 1HTI, (Mande *et al.*, 1994)) is open whereas that of subunit B is closed due to the presence of the inhibitor. The human TIM (hTIM) undergoes specific deamidation at asparagines 15 and 71 of each subunit. The binding of the ligand enhances the specific deamidation of Asn71 which is followed by the subsequent deamidation of the Asn15 of the neighboring subunit. The new negative charges at the subunit interface leads to dissociation and unfolding. The deamidation site residues (Asn15 and Asn71) are shown as red sticks in Figure 2.5.

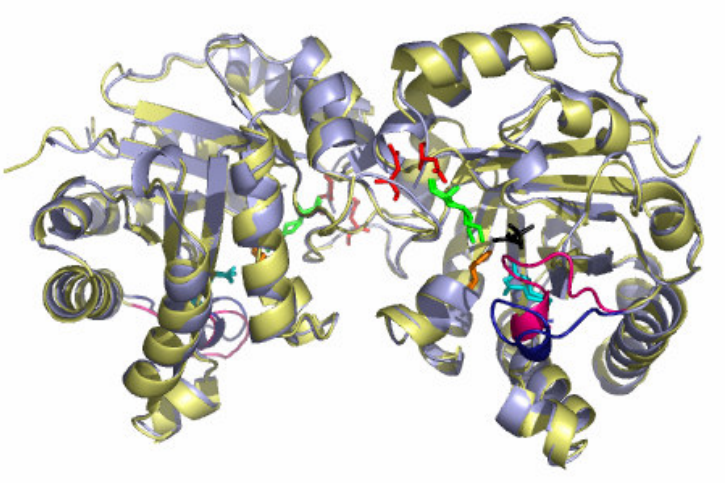


Figure 2.5. The superimposition of hTIM (light yellow) and free form of cTIM (light blue)

2.5.3. PDZ

PDZ (post-synaptic density-95/discs large/zonula occludens-1) domains are found in diverse signaling proteins in bacteria, yeasts, plants, insects and vertebrates. They are protein recognition modules and fundamental in regulating the dynamic organization of the cell. The X-ray structures of free (PDB ID: 1BE9) and complex (PDB ID: 1BFE) forms of the third PDZ domain (PDZ-3) of PSD-95 are determined at 1.82 Å and 2.30 Å resolutions, respectively. (Doyle *et al.*, 1996)

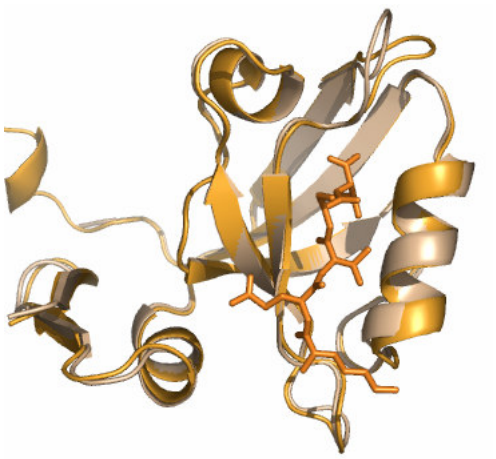


Figure 2.6. PDZ-3 domain proteins, unbounded (bounded) structure is colored in light (dark) orange.

The structure of PDZ consists of two α-helices (αA, Pro346–Ser350; αB, His372–Asn381) and six β-strands (βA–βF). The free and complex forms are superimposed and shown in Figure 2.6. In PDZ ligand binding does not induce large conformational changes of the domain structure. The bound form of PDZ-3 domain is in great interest of the allosteric pathway studies. (Lockless and Ranganathan, 1999; Ota and Agard, 2005)

2.5.4. Cdc25B

Cdc25B (cell division cycle 25 homolog B) is involved in the control of cyclin-dependent kinases (CDK) and progression of cells through the cell control. (Rudolph, 2007) The X-ray structure (PDB ID: 1QB0) is determined by Watenpaugh and coworkers at 1.9 Å resolution. (Reynolds *et al.*, 1999)

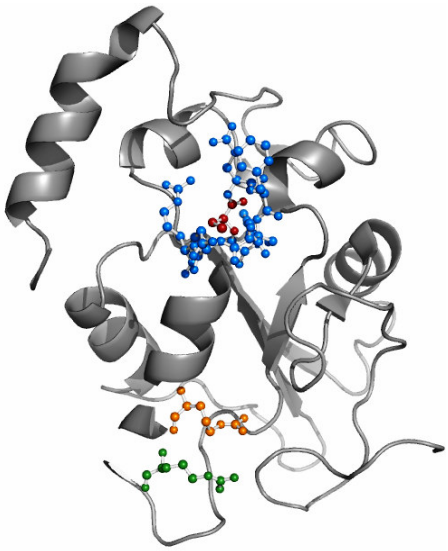


Figure 2.7. The structure of Cdc25B

The structure of Cdc25B is illustrated in Figure 2.7. The catalytic site (His472 through to Arg479), the catalytic residue Cys473 and the two arginine residues (Arg488 & Arg492), which are in the binding pocket and crucial for docking studies, are shown in Figure 2.7. The catalytic site is colored in blue, Cys473 is colored in red, Arg488 and Arg479 are colored in orange and green, respectively. This structure has not been investigated in great detail; in this thesis, main emphasis is on the comparison of the

communication abilities of the two arginines which would give insights into the experimental and computational docking results.

3. METHODOLOGY

A novel method, where the communication stochastics within the proteins is modeled as a discrete-time, discrete-state Markov process of 'information' transfer across the proteins, has been developed. (Chennubhotla and Bahar, 2006) A Markov chain is a random process in which the future states depend only upon the present state, not on any past states.

In this methodology, the three-dimensional structures of proteins are investigated from the network perspective, where the amino acid residues represent the nodes of the network.

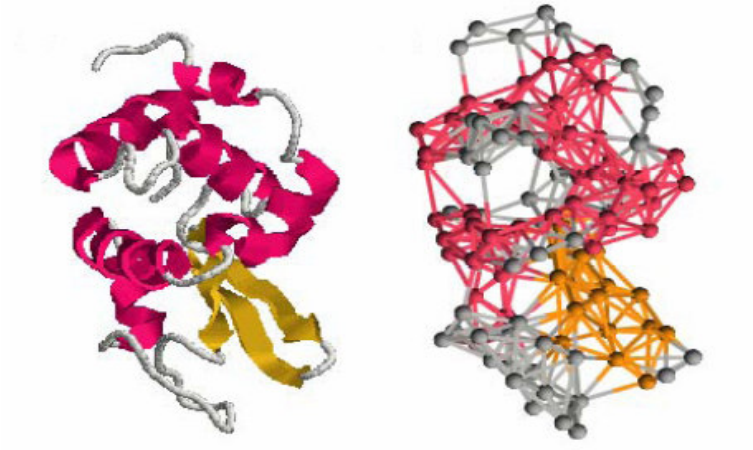


Figure 3.1. Protein Network

3.1. Markov Process of Network Communication

In proteins the topology-driven dynamics underlies function. In particular, these motions also determine communication patterns that are inherent to the native architecture. They are modeled as networks of n nodes, as shown in Figure 3.1 (Cui and Bahar, 2006), here each node represents a given residue v_i. Chennubhotla and Bahar (Chennubhotla and Bahar, 2006) defined the affinity (a_{ij}), which is the interaction strength, between residues v_i and v_j, as:

$$a_{ij} = \frac{N_{ij}}{\sqrt{N_i N_j}} \tag{3.1}$$

where N_{ij} is the total number of atom-atom contacts made between residues v_i and v_j based on a cutoff distance of 4 Å and N_i is the total number of heavy atoms in v_i. d_j is the local interaction density at residue v_j and is defined as:

$$d_j = \sum_{i=1}^{n} a_{ij} \tag{3.2}$$

The communication probabilities between residues are assumed to be the functions of their affinities across the network and thus the conditional probability of transmitting information from residue v_j to residue v_i within *one-step* is defined as:

$$m_{ij} = \frac{a_{ij}}{d_j} \tag{3.3}$$

There is an asymmetry in m. For instance, $m_{jk} \neq m_{kj}$ in the network shown in Figure 3.2.

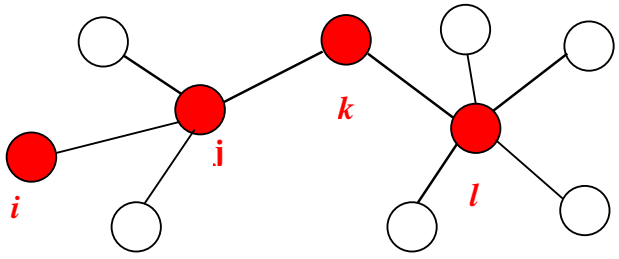


Figure 3.2. Asymmetry in Network

The conditional probability matrix, M=$\{m_{ij}\}$, also called the 'markov transition matrix', is expressed in terms of the affinity matrix (A=$\{a_{ij}\}$) and the degree matrix (D=diag$\{d_{ij}\}$) .

$$M = A D^{-1} \tag{3.4}$$

Markov transition matrix defines the stochastics of a random walk of propagation of information over the network of residues. The probability of initiating the Markov propagation process at node j is p_j^0 and the probability of reaching residue i in one time step is $m_{ij} \times p_j^0$. In matrix notation, the probability of ending up on any of the residues $v = [v_1, v_2, ..., v_n]$ after one step is given by the distribution, $p^1 = Mp^0$, where $p^k = [p_1^k, ..., p_n^k]$. This process can be iterated, so that after β steps is expressed as $p^\beta = M^\beta p^0$. Assume there is a path connecting every pair of residues in the network, as $\beta \to \infty$, the Markov chain p^β approaches a unique stationary distribution, $\pi = [\pi_1, ..., \pi_n]$, the elements of which are given by $\pi_i = d_i / \sum_{k=1}^{n} d_k$. (Norris, 1997; Chennubhotla and Bahar, 2006, 2007 (*submitted*))

3.2. Hitting and Commute Time Concepts

The hitting time and commute time are the basic quantities associated with the flow of information between particular pairs of residues. *Hitting time*, *H(j,i)*, is the expected number of steps it takes to send information from residue v_i to residue v_j, and *commute time*, *C(i,j)*, is defined as:

$$C(i, j) = H(i, j) + H(j, i) = C(j, i) \tag{3.5}$$

Commute time is symmetric by definition. However, there is asymmetry in hitting time, that is to say *H(j,i)* may not be the same as *H(i,j)*. (Chennubhotla and Bahar, 2007 (*submitted*)) By using the recursion equation (Norris, 1997) hitting time is given as:

$$H(n, i) = 1 + \sum_{k=1}^{n-1} H(n, k) \times m_{ki} \tag{3.6}$$

3.3. The Relation between Information Flow and Protein Function

The elastic network constructed to elucidate the communication stochastics can be used to explore the dynamics of proteins around their equilibrium conformations as well. The motions under native state conditions are obtained from the Kirchhoff matrix, Γ, which is defined in terms of the affinity and degree matrices.

$$\Gamma = D - A \tag{3.7}$$

Both Markov matrix and Kirchhoff matrix are expressed in terms of degree and affinity matrices. Thus, the commute and hitting times can also be written in terms of the eigenvalues and eigenvectors of the stiffness matrix that is the natural modes of vibration derived from the Gaussian Network Model. (Bahar *et al.*, 1997b)

$$C(i,j) = \left[\Gamma_{ii}^{-1} + \Gamma_{jj}^{-1} - 2\Gamma_{ij}^{-1}\right] \sum_{k=1}^{n} d_k \tag{3.8}$$

$$H(j,i) = \sum_{k=1}^{n} \left[\Gamma_{ki}^{-1} - \Gamma_{ji}^{-1} - \Gamma_{kj}^{-1} + \Gamma_{jj}^{-1}\right] d_k \tag{3.9}$$

Eigen-mode decomposition of the Kirchhoff matrix gives the natural modes of vibration for the system. For an n node network Γ has a rank of n-1, so it cannot be inverted and its pseudo-inverse is computed after eliminating the contribution of the zero eigen value.

$$\Gamma^{-1} = \sum_{p=2}^{n} \frac{1}{\lambda_p} u_p u_p^T \tag{3.10}$$

The elements of the inverse Kirchhoff matrix provide a measure of the cross-correlation $\langle \Delta R_i . \Delta R_j \rangle$ between the fluctuations ΔR_i and ΔR_j of residues v_i and v_j near mean positions in the equilibrium structure, following the relationship (Bahar *et al.*, 1997b):

$$< \Delta R_i . \Delta R_j >= \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ij} \quad (3.11)$$

where k_B is the Boltzmann constant, and γ is the uniform force constant of the springs that connect the nodes in the GNM description of the biomolecular structure (originally proposed by Tirion (1996) at atomistic scale).

The mean square fluctuations in the distance fluctuations between nodes i and j is:

$$< \Delta R_{ij} . \Delta R_{ij} >= < \Delta R_i . \Delta R_i > + < \Delta R_j . \Delta R_j > - 2 < \Delta R_i . \Delta R_j > \quad (3.12)$$

Equation 3.8 and 3.12 illustrates that commute times are directly proportional to the mean-square fluctuations in inter-residue distances, thus establishing a connection between fluctuation dynamics and communication stochastics.

$$C(i, j) \sim < \Delta R_{ij} . \Delta R_{ij} > \quad (3.13)$$

Equation 3.11 is fundamentally important: it bridges between the informatic-theoretic methodology and statistical mechanical behavior of the network. (Chennubhotla *et al*., *unpublished*)

3.4. Analyzing the sub-system of a Protein Complex

In the comparative investigation the sub-systems of a protein complex might required to be explored. To analyze the dynamics of a sub-system of the protein (Flory, 1976; Brooks *et al.*, 1995; Zheng and Brooks, 2005; Ming and Wall, 2005; Chennubhotla, 2006), the fluctuation vector (Δx) can be rearranged into a relevant portion (Δx_s) and the environment (Δx_e).

$$\Delta x^T = [\Delta x_s^T \quad \Delta x_e^T] \quad (3.14)$$

$$\Gamma = \begin{bmatrix} \Gamma_{ss} & \Gamma_{se} \\ \Gamma_{es} & \Gamma_{ee} \end{bmatrix} = \begin{bmatrix} \Gamma_{ss} & \Gamma_{se} \\ \Gamma_{se}^{T} & \Gamma_{ee} \end{bmatrix} \tag{3.15}$$

$$\Delta x^T \Gamma \Delta x = \Delta x_s^T \Gamma_{ss} \Delta x_s + \Delta x_e^T \Gamma_{ee} \Delta x_e + 2\Delta x_e^T \Gamma_{es} \Delta x_s \tag{3.16}$$

These representations are used to generate a new Kirchhoff matrix for the protein after binding.

$$\Gamma_{new} = \Gamma_{ss} - \Gamma_{se} \Gamma_{ee}^{-1} \Gamma_{es} \tag{3.17}$$

In the present work, this approach is applied when the numbers of residues in two conformations of a protein that is of interest are not equal. (i.e. presence of ligand, differences in PDB structures) This approach elicits the application of comparative investigations.

3.5. Communication Pathways

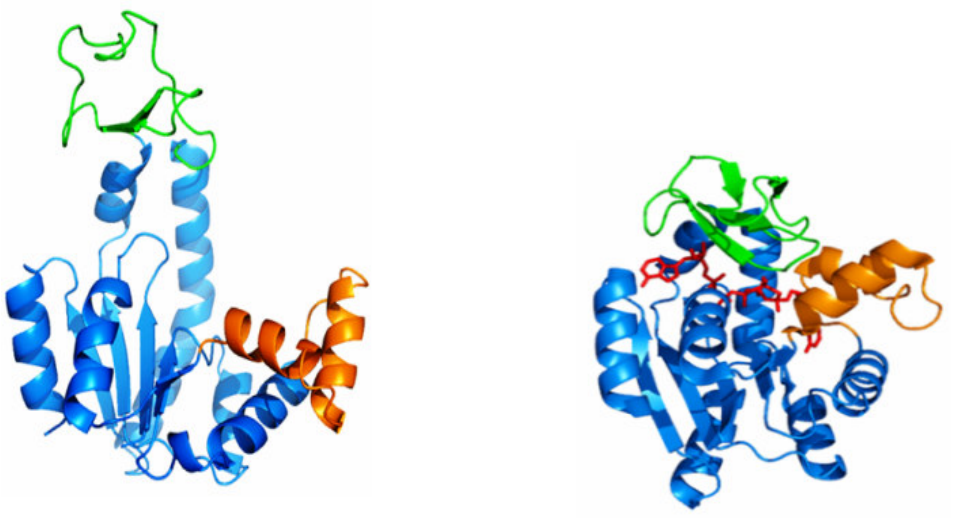
The communication pathways that might be the potential targets for controlling the function are computed by using the 'Dijkstra's shortest path algorithm'. (Dijkstra, 1959) A path from a source vertex to a target vertex is to be 'the shortest path' if its total cost is minimum among all the paths between the source and the target. The algorithm starts with the source; it visits the vertices in order of increasing cost. (Chen, 2003) The pathways between the key residues illustrate the possible allosteric communication in the proteins.

4. RESULTS AND DISCUSSION

In this study, the information flow in the apo-form of E.coli adenylate kinase, triosephosphate isomerase (TIM), PDZ signaling protein and Cdc25B are investigated.

4.1. Adenylate Kinase

Adenylate kinase catalyzes the reaction Mg^{2+}.ATP + AMP → Mg^{2+}.ADP + ADP. (Noda, 1973) Adenylate kinase from *E.coli* (AKeco) has three domains: CORE, AMP binding domain (AMPbd) and LID. CORE is the largest domain including the residues M1-I29, T60-V121, and Q160-G214, AMPbd includes the helix formed by residues S30-V59 and the LID domain includes residues G122-D159. ATP phosphates are bound to enzyme through P-loop GXXGXGK (G7-K13). The structures of the ligand-free enzyme form (PDB ID: 4AKE – Muller *et al.*, 1996) and its complex with the substrate-mimicking inhibitor AP_5A (PDB ID: 1AKE – Muller and Schulz, 1992) are shown in Figure 4.1 (a) and (b), respectively. This figure depicts that the enzyme undergoes large-scale movements upon substrate binding.



(a) (b)

Figure 4.1. (a) Open and (b) Closed form of AKeco

In this study, apo (open) form of the protein (Muller *et al.*, 1996) is analyzed. The distribution of hitting times (a) and commute times (b) in AKeco are shown in Figure 4.2. The upper curves describe the average hitting times (<H(i)>) and average commute times (<C(i)>) (i.e. average over the elements of the corresponding row), respectively. The hitting and commute time values range from low (blue) to high (red). Low hitting/commute times reveals faster communication. The moving domains, AMPbd and LID, are observed to have slow communication propensities. The color-coded maps reveal that LID communicates poorly with both AMPbd and portions of the CORE, however AMPbd is ineffective mostly with respect to the LID. This might suggest that first AMP binds to the structure via the more rigid AMP binding domain. Then a conformational change takes place to accomplish the binding of ATP. This is, possibly, succeeded by the closure of the LID. Thus, analyzing the flow of information gives insight into the possible sequence of events in adenylate kinase catalysis.

The prior experimental and computational findings about the sequence of events in AKeco are contradictory; Deiderichs and coworkers state that the displacement of the AMPbd takes place before the closure of the LID (Schulz *et al.*, 1990), whereas Maragakis and Karplus noted that LID closes initially in the path from open to the close conformation. (Maragakis and Karplus, 2005) The results obtained from the color-coded maps support the experimental findings of Deiderichs and coworkers.

The hitting and commute time distributions are examined to explore and depict the functionally important residues. The important residues of AKeco are ATP binding site (P-loop) and catalytic sites: Lys13, Arg123, Arg156, Asp158, Asp159 and Arg167. The mean versus standard deviation of the commute time distribution is plotted and displayed in Figure 4.3. It is observed that the ATP-binding site residues are at the lower left end of the distribution, indicating that they are ‘fast’ and ‘precise’ communicators. Some of the catalytic sites have moderate communication abilities, but interestingly, the catalytic sites are also fast and precise compared to the other residues *within* their domain. In AKeco, P-loop (ATP-binding site) exhibits the highest allosteric potential (lowest commute time and variance).

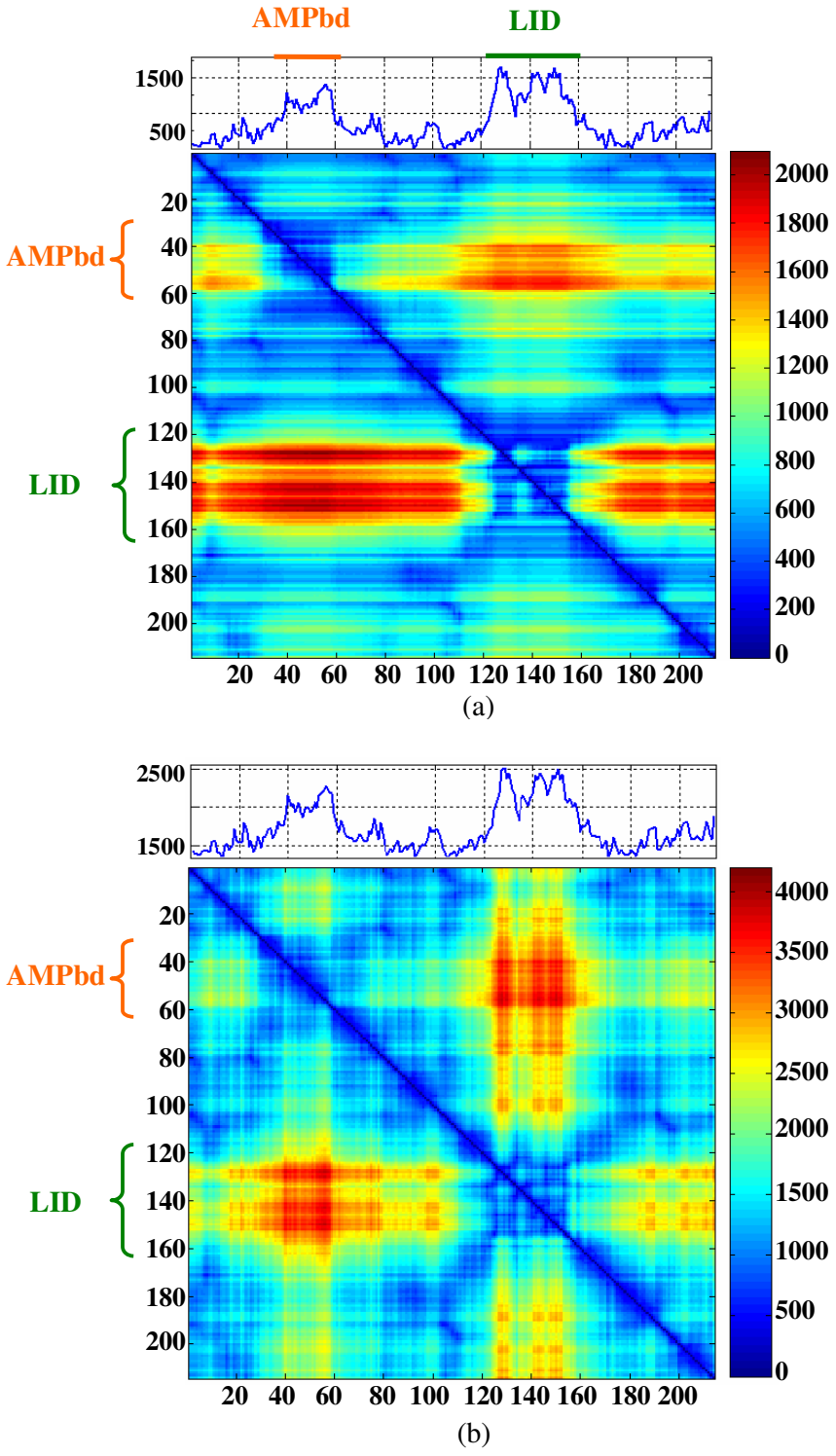


Figure 4.2. Hitting Time (a) and Commute Time (b) Distributions

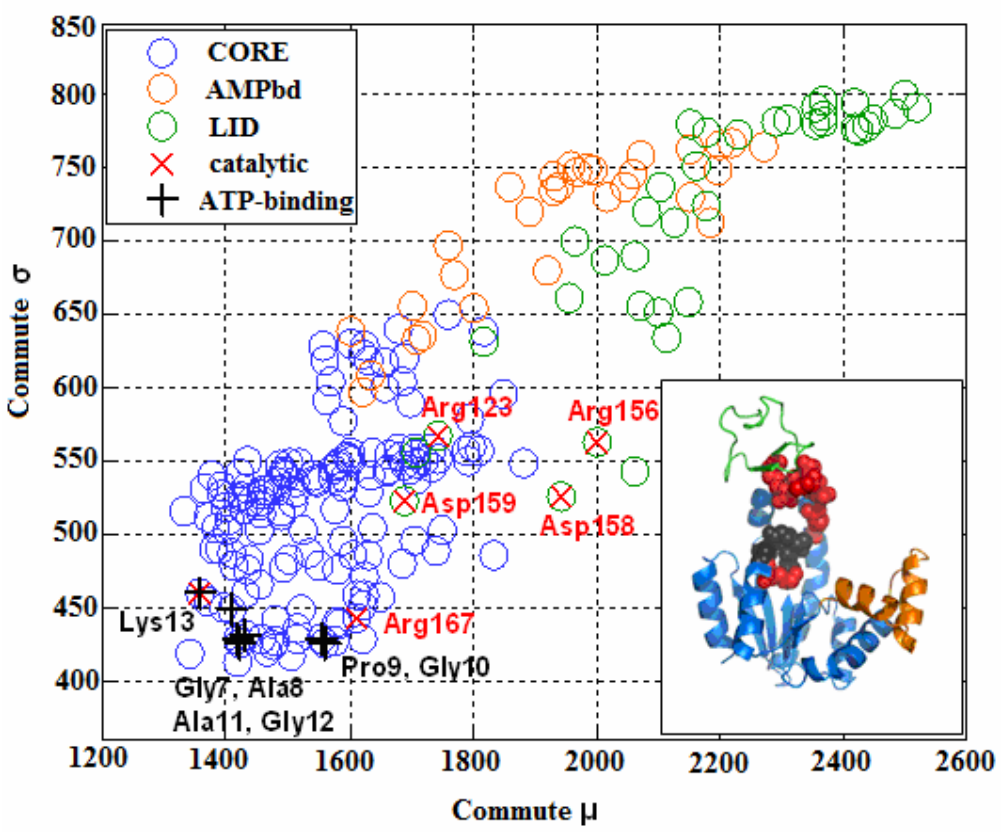


Figure 4.3. Communication abilities of the residues in AKeco

The changes in distances between the pairs of residues belonging to different domains are revealed in Figure 4.4. The largest changes in inter-residue distances occur at the AMPbd-LID residue pairs, succeeded by LID-CORE and then AMPbd-CORE pairs. The ordinate gives the number of residue pairs having the corresponding fluctuation values. The range on abscissa has arbitrary units, the absolute values can be received from FRET (Fluorescence Resonance Energy Transfer) experiments. FRET is a technique for measuring interactions between two nodes of the protein in vivo. The timescale results from FRET measurements might also be used to rescale the arbitrary hitting and commute time values.

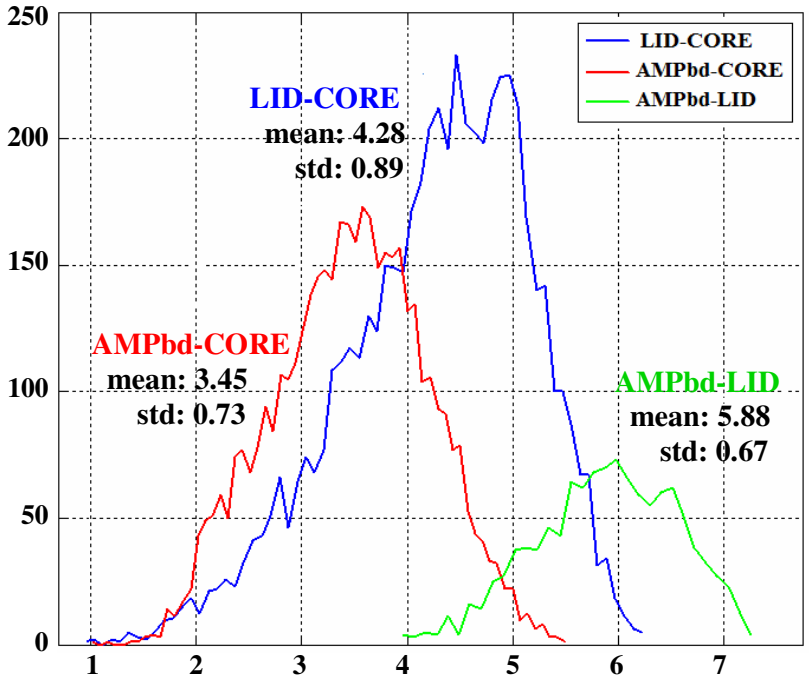


Figure 4.4. Inter-domain inter-residue distance distribution in AKeco

The effect of collective modes on communication stochastics is an important point that is analyzed in the study. Mean commute time (<*C(i)*>) for the first, second, tenth slow modes and the total <*C(i)*> (all modes) as a function of residue index are given in Figure 4.5. Minima in <*C(i)*> refer to residues with fast communication abilities, while peaks are the slow communicators. The correlation values between the mean commute time for all modes (black) and the mean for the selected modes: first (red), second (blue) and tenth (green), are 0.89, 0.42 and 0.09, respectively. This reveals the significant contribution of the slow (global) modes to the communication stochastics.

The fractional contributions of the first and second slow modes to the commute time are given in Figure 4.6 (a) and (b). Panel (a) shows the contribution of the first mode while (b) shows that of the second mode. The structures of AKeco, given in Figure 4.7 (a) and (b) are colored according to the mean commute time values of the first and second global modes, respectively. 'Hot' (red, orange) colors represent the residues with high commute times (slow communication). The slowest GNM mode activates the functional movement of the LID domain whereas the second mode activates the AMP binding

domain. (Temiz *et al.*, 2004) Panel a (b) interprets that first (second) mode designates the communication ability of LID (AMPbd).

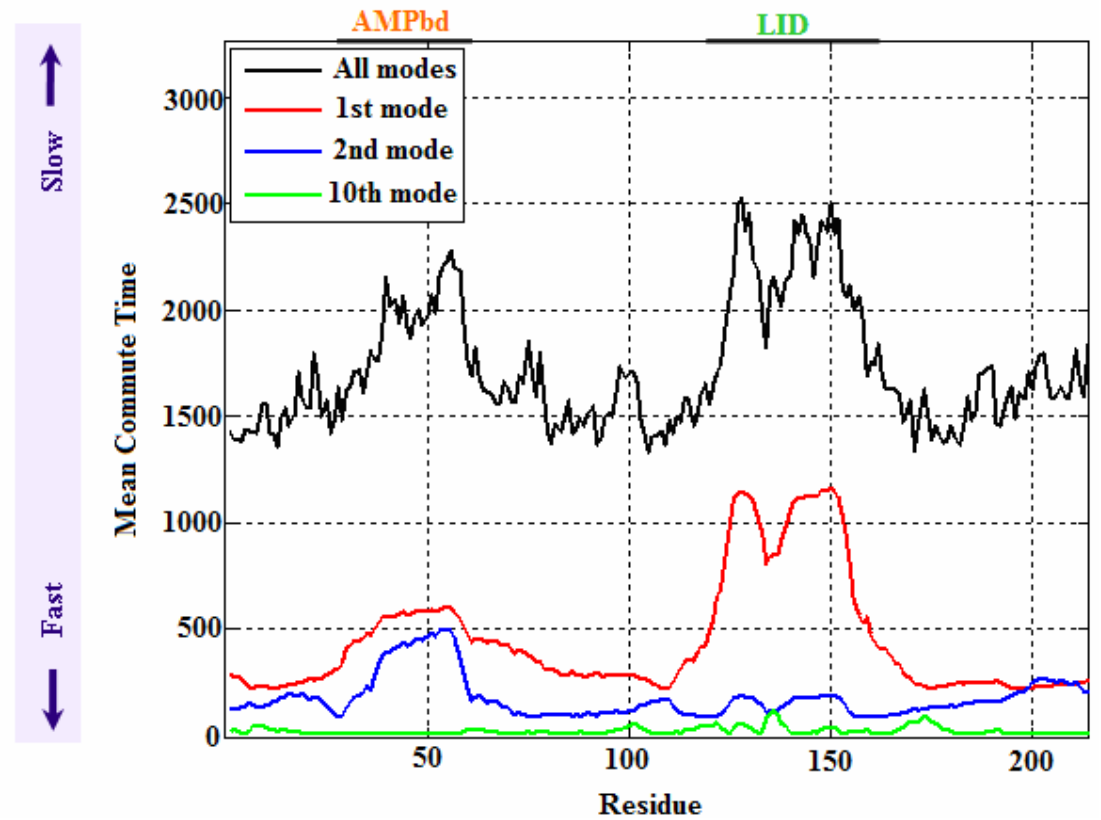


Figure 4.5. Mode contributions to mean commute time

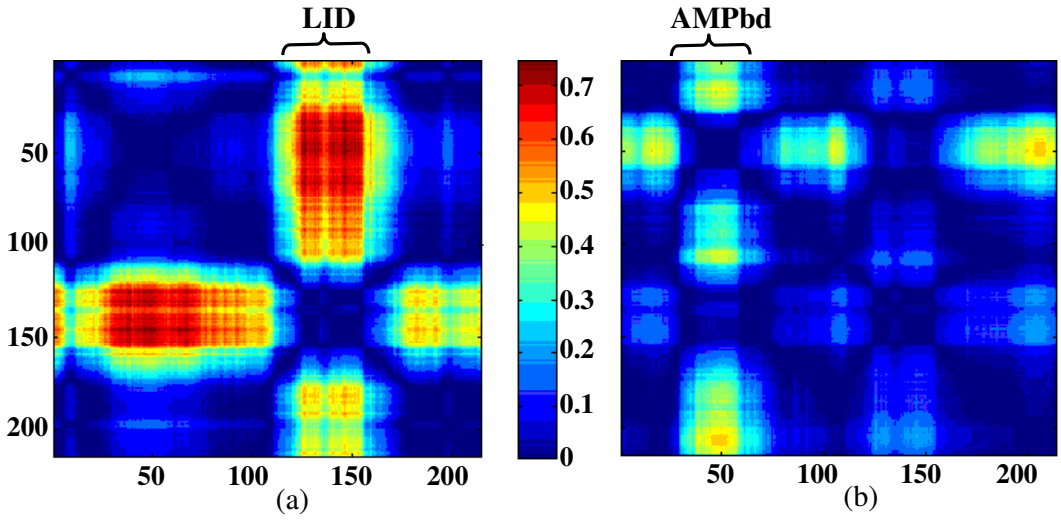


Figure 4.6. Fractional contributions of first (a) and second (b) modes to commute time

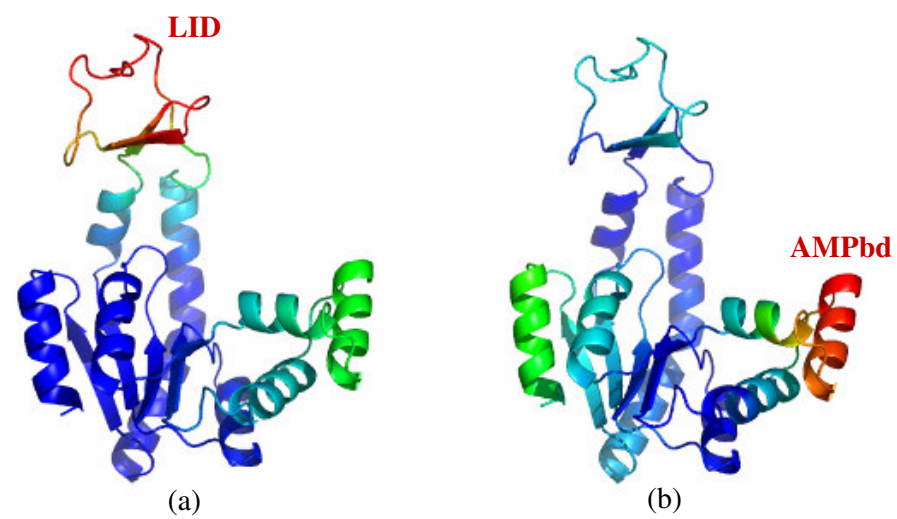


Figure 4.7. Mean commute times in first (a) and second (b) global modes of AKeco

The topology-driven communication in proteins is analyzed with an ideal random walk in three-dimensional space. The mean value of the square of the displacement between two residues is a function of the number of steps it takes to reach the destination times the square of the average length of each step ($d_{eff}^2 = n\, l^2$). It is assumed that the number of steps to reach the destination is half of the commute time ($C(i,j)/2$) (Chennubhotla and Bahar, 2007 (*submitted*)). The average step size for apo-form of adenylate kinase in *E.coli* is calculated to be 3.46 Å. The distribution of the effective distance (calculated from the random walk model) versus the physical distances (calculated from the PDB coordinates) is illustrated in Figure 4.8. The residue pair distances are marked in different colors according to being inter or intra-domain. The physical and effective distances are shorter within the domains. In other words, intra-domain communications are much faster than inter-domain communications. When the residue pairs separated by an identical physical distances of 30 Å are examined, it has been made out that residues in CORE, mostly, prefer to communicate with the residues in the same domain, then with the residues in AMP binding domain, and their communication is least with the residues in LID. Identification of communication ability of the residue pairs and examining their effects on allosteric responses would give insights into better understanding of the molecular basis of signal transduction mechanisms in molecular systems.

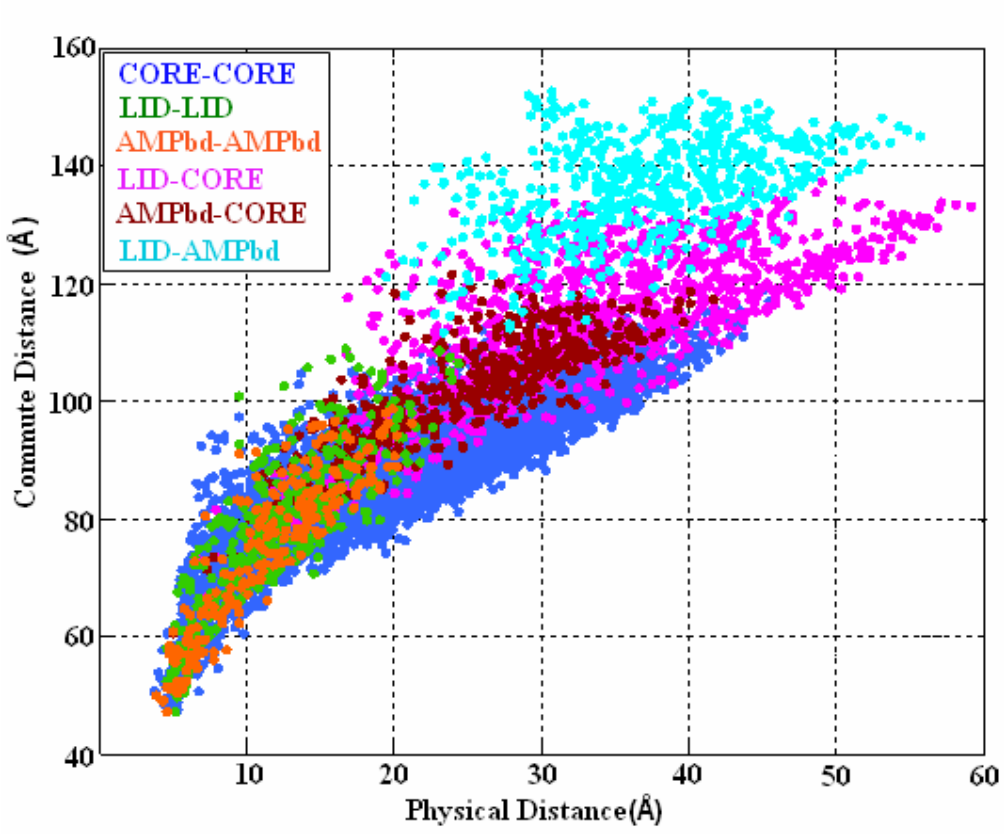
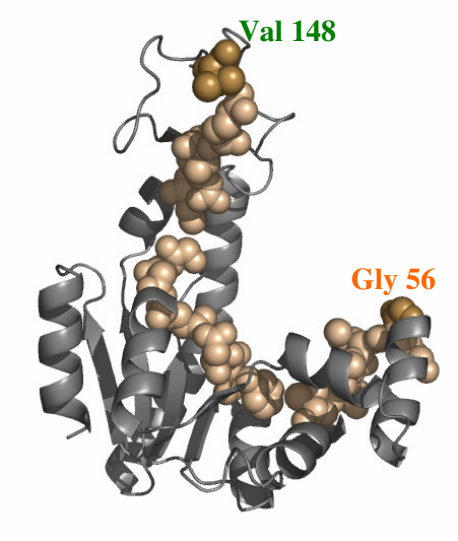


Figure 4.8. Comparison of effective (commute) distances and physical distances in AKeco

The identification of key interactions and communication pathways which might serve as potential targets for controlling functions is one of the crucial goals of the new methodology. The most probable pathways are computed according to the idea that if communication probability between two residues is high, the information would transfer through them. The maximum likelihood pathway between the tip of the residues of the moving domains, Val148 from LID and Gly56 from AMPbd, are illustrated in Figure 4.9. The residues on the pathway are stated, the ones with high conservation scores (evaluated from ConSurf, (Landau *et al*., 2005)) are listed in boldface. In Figure 4.10 the structure of AKeco is colored according to the frequency with which the residues participate in the most probable communication pathways for any pair of residues in the network. The residues with 'hot' colors correspond to the most frequently visited ones are labeled and their side-chains are explicitly shown. These residues form a communication core. It is plausible that residues largely taking part in information propagation should be central in

the network architecture; and they should be on the shortest pathways between the most residue pairs in the network. (del Sol *et al.*, 2006)



148-**131**-132-**123-120-[10-13]-[84-86]-64-[60-58]**-57-**56**

Figure 4.9. Conserved residues and communication pathways

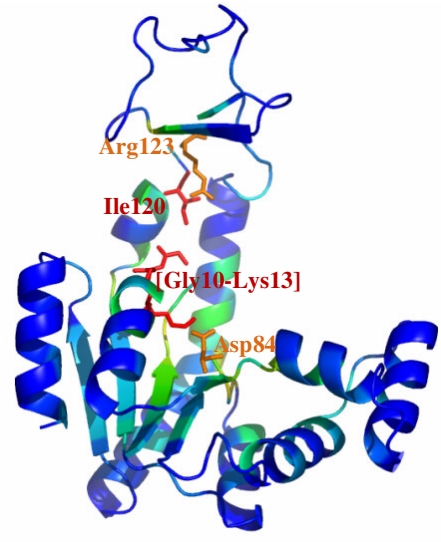


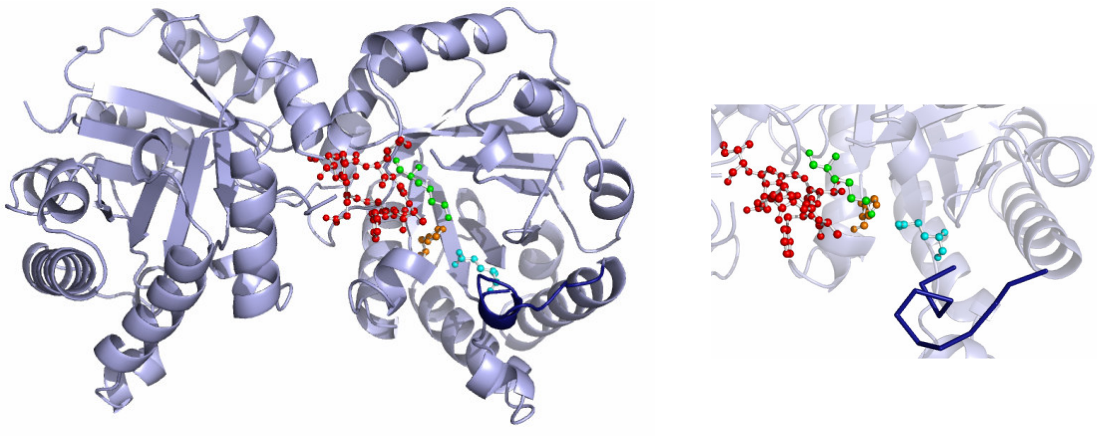
Figure 4.10. Communication core in AKeco

4.2. Triosephosphate Isomerase

Triosephosphate isomerase (TIM) is a dimeric glycolytic enzyme. The dimerization is required for the enzymatic activity (Waley, 1973), suggesting that residues of one subunit are crucial for maintaining the integrity and stability of the active site of the other subunit. The two active centers and the interfacial regions of the subunits are in close contact by the 'interdigitating loop' (residues 71-78) that extends from one subunit to the other near active site pockets. There is no reported allosteric control or cooperativity between the two subunits. To gain further insights into the functional mechanism and the possible role of cooperativity, the flow of information on the enzymes from human and chicken are analyzed.

4.2.1 Free Form of cTIM

Analyzing the communication abilities of free form (PDB ID: 8TIM) gives provides better understanding of the communication patterns in the enzyme. The side chains of the catalytic residues: Lys13 (green), His95 (orange), Glu165 (cyan); the active loop (dark blue) and the interdigitating loop (red) are shown in Figure 4.11. Four residues on interdigitating loop (G72, A73, F74, T75) are in close contact with the active sites of the adjacent subunits, these residues were supposed to form part of the active sites. (Banner *et al*., 1975)



(a) (b)

Figure 4.11. Free Form of cTIM

The color-coded maps in Figure 4.12 reveal the hitting and commute times for every pair of residues in free cTIM. The numbering scheme corresponds to TIM sequence from 4 to 248 for monomer A and from 304 to 548 for monomer B. The interdigitating loops (residues 71-78 and 371-378) have low hitting and commute times with the adjacent subunit residues, illustrating that this loop is an effective communicator. There are also other good communicators in the structure which are illustrated in Figure 4.13 (b) to be positioned in the core of the protein.

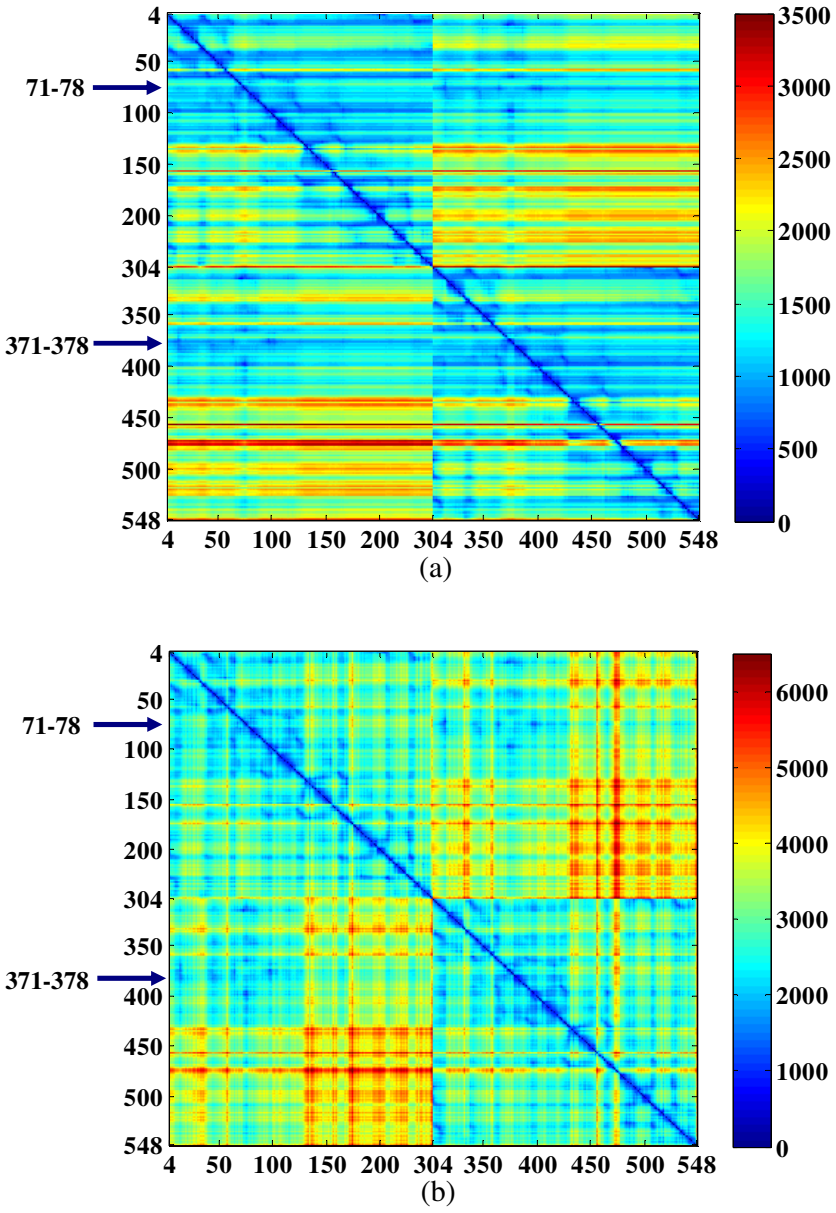
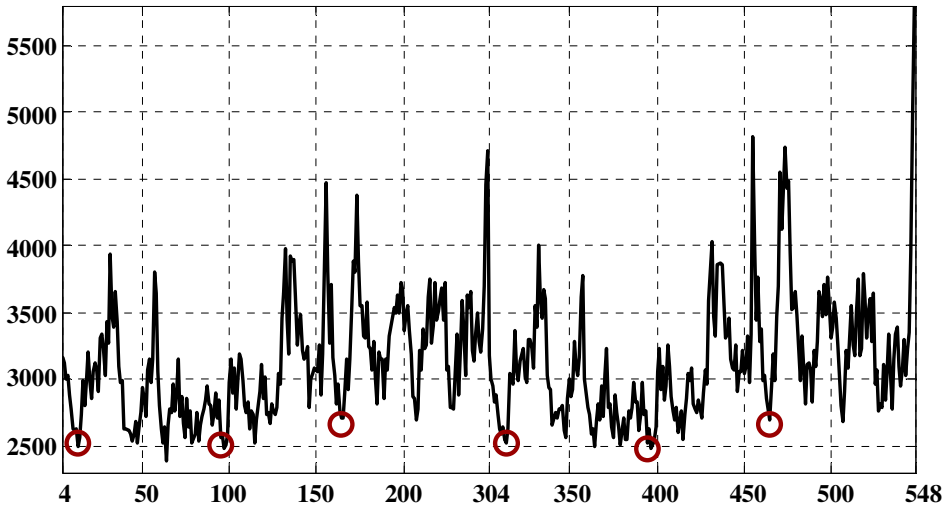
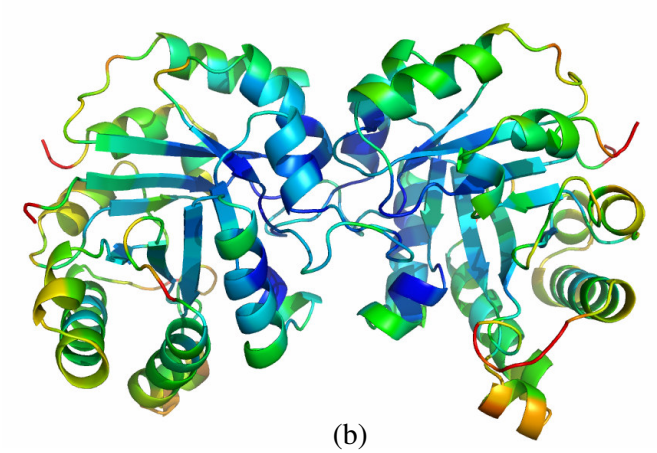


Figure 4.12. Hitting Time and Commute Time Distributions

Figure 4.13 (a) shows the mean commute times (<C(i)>) for each residue in free TIM, residues at minima have very efficient communication abilities. The catalytic residues Asn11, Lys13, His95 and Glu165 in both subunits are marked with red circles, these residues are at the minima of the mean commute time distribution. In Figure 4.13 (b), the X-ray structure of free TIM is colored according to the mean commute time values of the residues. 'Hot' colors (red, orange) represent the residues with high commute times. The residues close to the dimer interface have low commute times (in blue color), indicating that they are fast communicators.



(a)



(b)

Figure 4.13. Mean Commute Times of residues in free form of cTIM

Analyzing both the mean and standard deviation of commute times for each residue represents the communication ability of the residues in the network. The active-site residues and the residues supposed to play a part in the inter-subunit contact (M14, N15, S45, I46, L48, Q53, Q64, N65, Y67, V69, G72-I78, P80, M82, I83, D85, I86, I92, E97, R98, E104, and K112) (Banner *et al.*, 1975) have low mean commute times and the standard deviations are relatively low, thus they have 'fast and precise' communication abilities. Figure 4.14 shows the communication abilities of the residues in the free form of chicken TIM. The dots represent the residues in the network. They are grouped according to the mean value and standard deviation of their commute times. The X-ray structure colored according to the cluster of the color in the plot is shown in the inset. The residues in blue and cyan are in the subunit interface and are fast and precise communicators. There is a gradual increase in communication ability of the residues as they are closer to the subunits interface. This elucidates the importance of these residues in communication among the subunits.

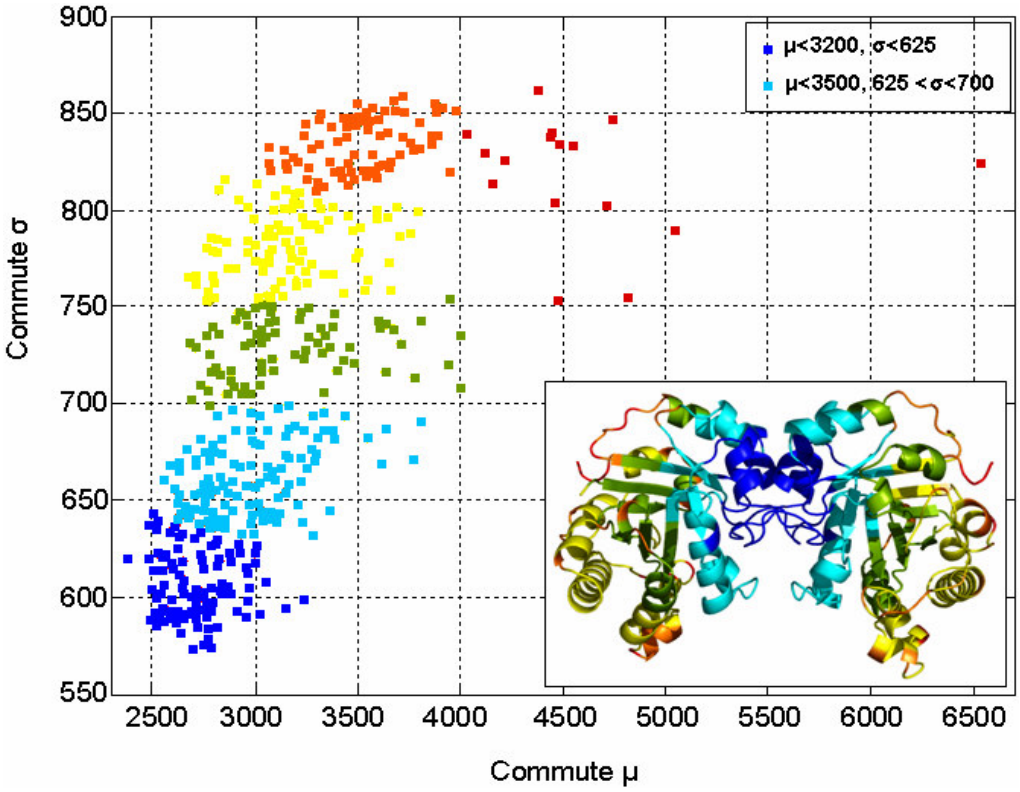
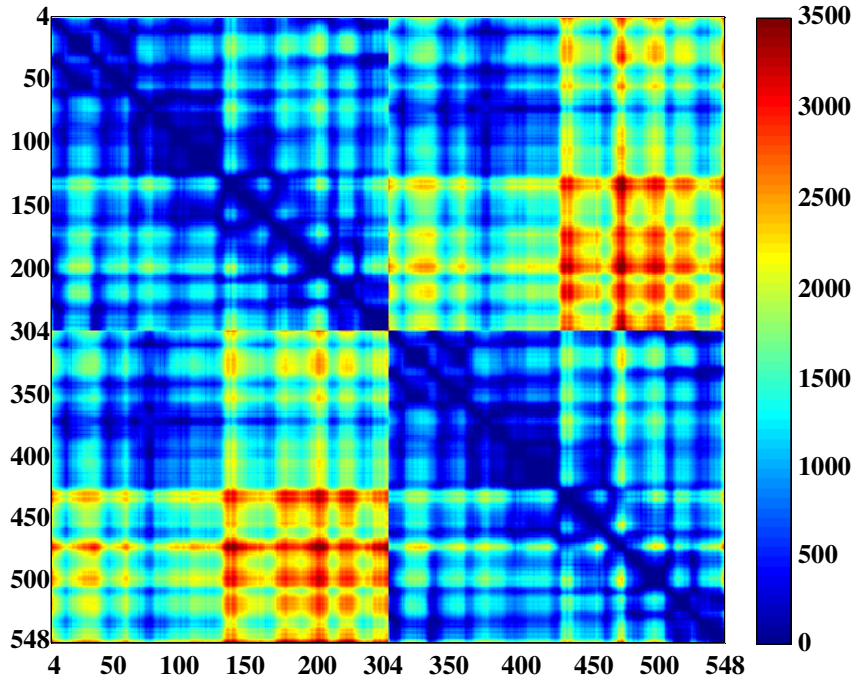


Figure 4.14 Communication abilities of the residues in free TIM

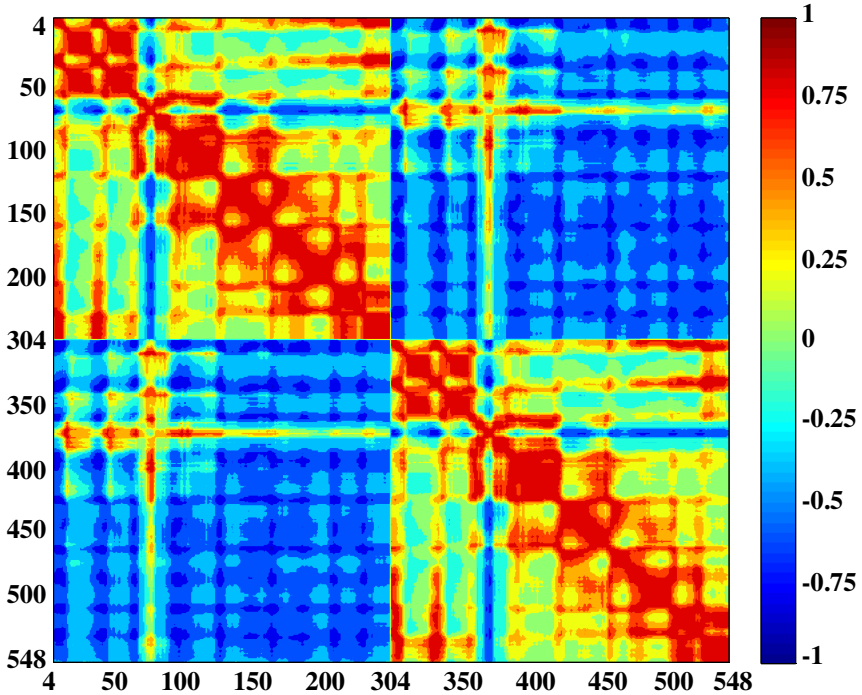
For triosephosphate isomerase, there is no reported evidence of allostery or cooperativity among the subunits that is important for its function. The experimental study carried out by Schnackerz and Gracy (Schnackerz and Gracy, 1991) indicates that the inactivity of the catalytic sites of one subunit does not affect the activity of the other subunit. However, it has also been found (Sun *et al.*, 1992a and 1992b) that the active catalysis by one monomer requires the support of the other subunit. Then, it has been put forth that catalytic residues get the correct orientation by the dimerization. (Borchert *et al.*, 1994) Besides, a computational study (Kurkcuoglu *et al.*, 2006) on mixed coarse grained ANM on TIM indicates that the opening/closure of the flexible loop (loop 6) is driven by the inherent collective motions of the enzyme in the absence of the ligand. Thus, the dynamics of the enzyme is of great interest.

Equation 3.13 shows that commute time is related to the intrinsic motions of a protein. The term ($-2\langle \Delta R_i . \Delta R_j \rangle$) in inter-residue fluctuations is positive when the fluctuations of a pair of residues are anti-correlated, it means the commute time is higher than the case where they are correlated. Figure 4.15 (a) and (b) illustrate the cumulative commute times and the cross-correlations of the slowest ten GNM modes in free cTIM, respectively. The positively correlated pairs have lower commute times, indicating that it is easy to send information between these residue pairs. The closer examination of the intersubunit commute times and cross correlations reveals that residues in dimer interface are positively correlated with some of the residues in the adjacent subunit. The interdigitating loop (71-78, 371-378) is positively correlated with 'all' the residues in the adjacent subunit and it has low commute times, which elucidates its effective communication ability with the other subunit.

The cross-correlation maps might be used to identify the probable domains in the structure. Here, TIM can be separated into five domains; the residues from 4 to 69 form *domain 1*, 70 to 81 form *domain 2*, 82 to 120 form *domain 3*, 121 to 230 form *domain 4* and 231 to 248 form *domain 5*. The structure shown in Figure 4.16 is colored according to the stated domains. Domain 1 is colored in pink, domain 2 in green, domain 3 in yellow, domain 4 in orange and domain 5 is colored in blue. The boundary for the domains are not so distinct, however they are consistent with the cross-correlation values.



(a)



(b)

Figure 4.15. Cumulative Commute Times (a) and Cross-Correlations (b) of the slowest ten modes in free form of cTIM

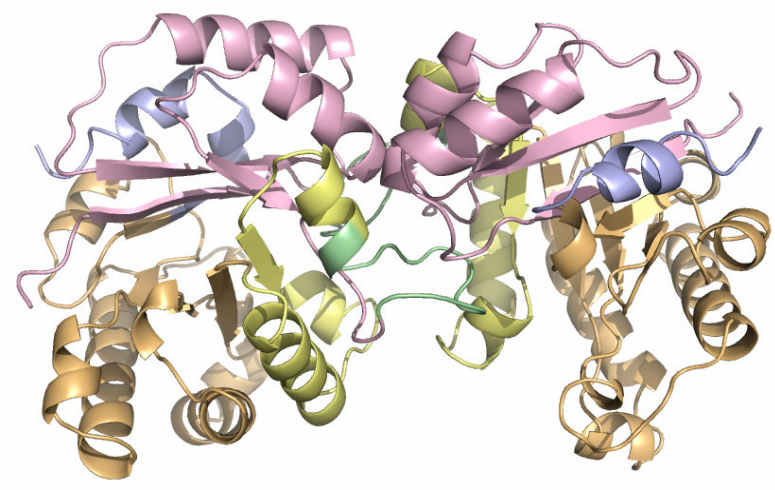


Figure 4.16. Domains in cTIM

The dynamic behavior of the networks has been investigated in a recent study by Vishveshwara and coworkers (Ghosh *et al.*, 2007), the changes in the nodes that play important roles in the stability of the network structure has been examined by molecular dynamics simulations (MD). During an MD simulation, the changes in the side-chain orientations might lead to changes in communication patterns in the protein networks. In most of the previous MD studies performed on TIM the main focus was on the active-site pocket and the motion of the remaining parts of the protein was restricted. Although the enzyme is active as a dimer, most of the simulations were performed only for the monomer. (Massi *et al.*, 2006)

In this study, the snap-shots of the MD simulations (Cansu *et al.*, *unpublished*) performed on the dimer of the free form of cTIM (8TIM) are analyzed to investigate the communication patterns in TIM. The full-atom, explicit solvent simulation is performed at 300 K and the duration of the simulation is 60 ns. When the open and closed X-ray structures of cTIM are compared, the RMSD between the superimposed structures is found to be 0.62 Å, but the flexible loop (loop 6) moves about 7 Å and closes over the ligand. Thus, the snapshots at 4 ns, 12 ns and 38 ns at which the flexible loops have various conformations are chosen to be used in the analysis. Figure 4.17 shows the chosen MD trajectory snapshots. The X-ray structure (gray), snapshots at 4 ns (blue), 12 ns (red) and at 38 ns (green) are superimposed to show the motion of the enzyme in the MD simulation.

The overall RMS differences between the X-ray structure and the snapshots at 4 ns, 12 ns and 38 ns are 2.309 Å, 1.933 Å and 1.731 Å, respectively. Besides, the distances between the tip of the loops (Thr172) of the X-ray structure and the snapshots are 4.68 Å, 4.62 Å and 3.76 Å, respectively.

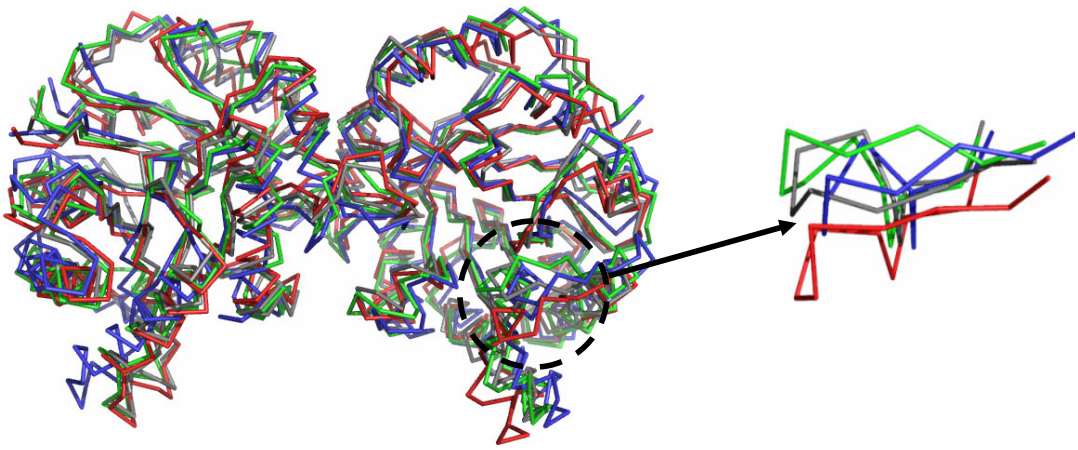


Figure 4.17. MD trajectory snapshots: 4 ns (blue), 12 ns (red), 38 ns (green) and the X-ray structure of free cTIM (gray)

The commute time maps of the molecular dynamics trajectory snapshots are illustrated in Figure 4.18. Four maps are given on the same scale, the first one (X-ray structure) differs from the others in color-code; this is probably due to being crystallized. In the other structures the solvent affects the structure explicitly and thus the contact maps differ. The energy minimization and equilibration steps are the relaxation period, it might be better to have these relaxation periods prior to analyzing the intramolecular pathways in a structure. Thereby, the effects of unnecessary crystal contacts would be removed from the system. It is conceivable that the biological systems perform their functions in the solution, thus investigating the communication propensities subsequent to the relaxation period might lead to more realistic results.

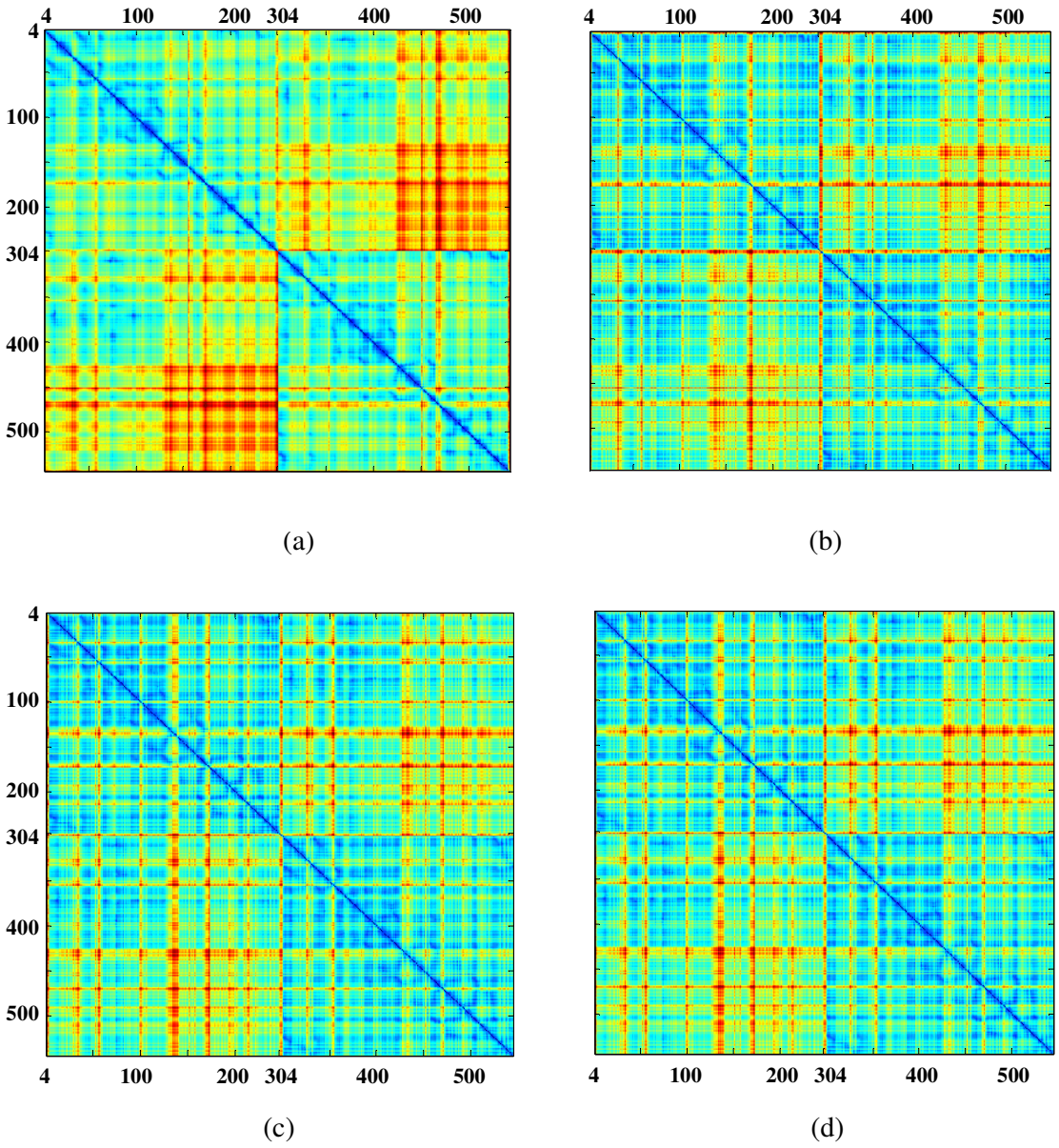


Figure 4.18. Commute Time maps of X-ray structure (a) MD snapshots at 4 ns (b) 12 ns(c) 38 ns (d)

Figure 4.19 displays the structure of cTIM where the eight α/β units are differentially colored. The units are numbered 1-8 and colored in red, magenta, orange, yellow, light green, green, cyan and blue, respectively.

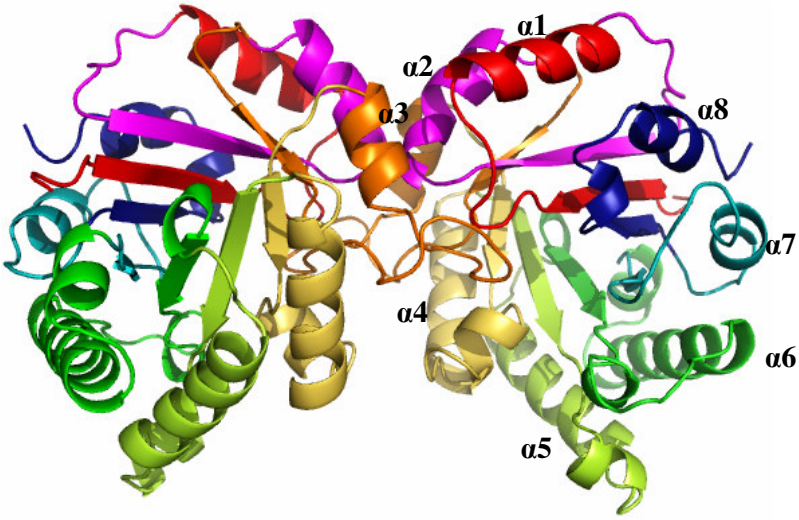


Figure 4.19. *α/β* units in cTIM

It is hard to identify the changes in communications in MD simulation by comparing the maps illustrated in Figure 4.18. Thus, difference matrices for commute times are built. Figure 4.20 (a), (b) and (c) show the difference between the commute times of snapshots at 4 ns & 12 ns, 4 ns & 38 ns and 12 ns & 38 ns, respectively. Negative values on Figure 4.20 (a) indicate that commute times between the corresponding residue pairs in snapshot at 4 ns are higher than snapshot at 12 ns. Similarly, negative values on Figure 4.20 (b) indicate that commute times in snapshot at 4 ns are higher than at 38 ns. During MD simulations the collective motion of the protein lead to differences in commute times of the residues, on loop2 (i.e. 31-34, '+'), *α*2 (i.e. 48-53, '*'), loop4 & *α*4 (i.e. 93-96 & 102-106, '•'), loop5 (i.e. 130-135, '▪') and loop6 (166-176, '▲') and loop7 (209-211, '♦'). Thus, most of the changes in commute times occur in the residues on loops and the residues near the subunit interface.

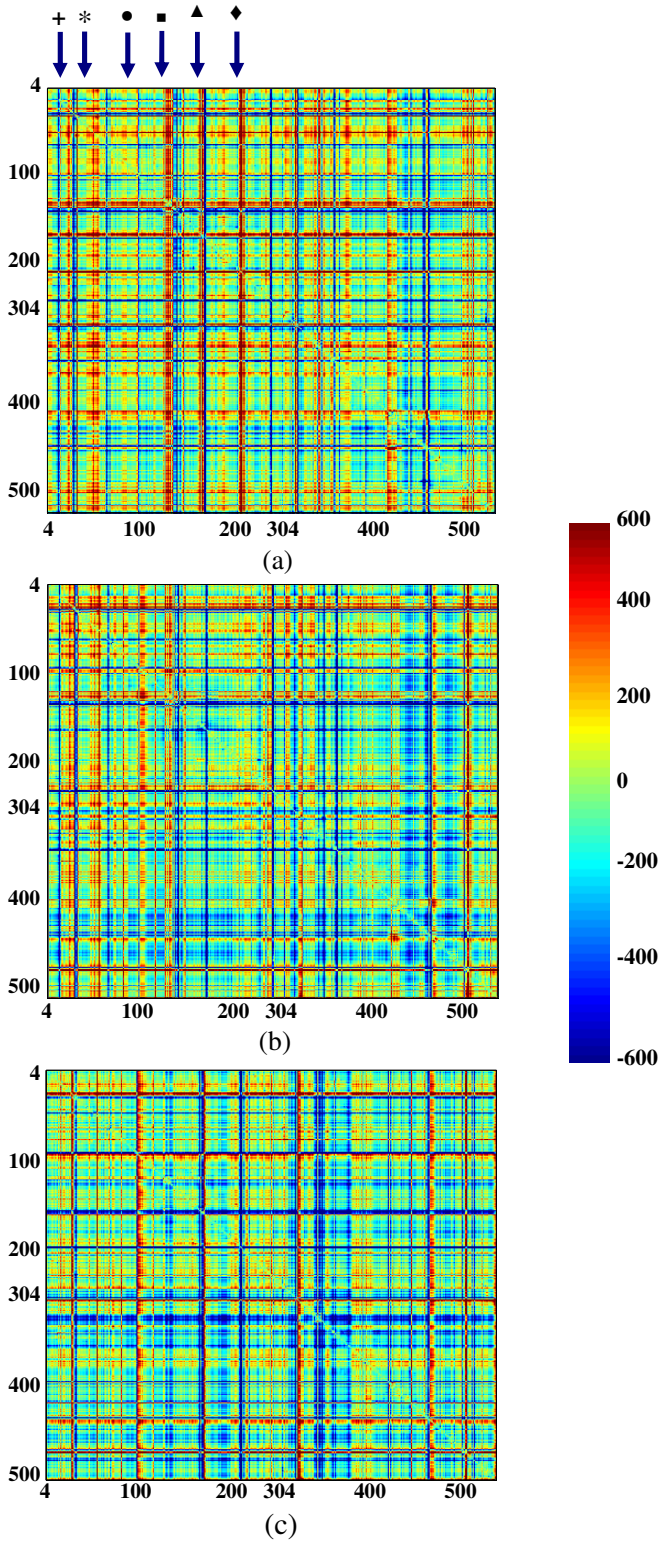


Figure 4.20. Changes (differences) in commute times of MD simulation snapshots between 4ns & 12ns (a), 4ns & 38ns (b), 12ns & 38ns (c)

The maximum likelihood pathways between the tips of the flexible loops (Thr172) in the adjacent subunit are analyzed to illustrate the changes in communication patterns in an MD simulation. Table 4.1 states the residues on the pathways where the conserved residues are listed in boldface.

Table 4.1. Residues on maximum likelihood pathways between Thr172 and Thr472

X-ray	**172-169-167-96-98-65**-66-67-74-**73-397-396-467-470-472**
4 ns	**172-131-130**-100-101-102-**367-366-393-426-465-469-472**
12 ns	**172-171-131-130**-100-101-102-**404-399-396-470-471-472**
38 ns	**172-171-167**-96-97-374-367-366-**393-394-395-465-469-472**

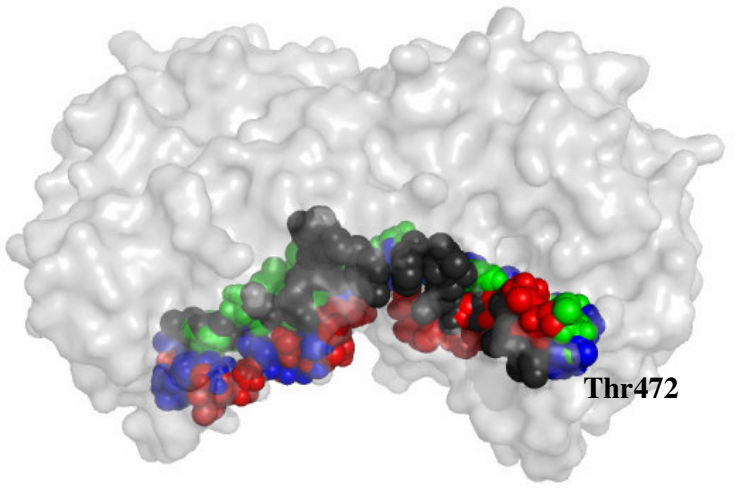


Figure 4.21. Maximum Likelihood pathways between Thr172 and Thr472

The residues on the pathways are illustrated on structure in Figure 4.21. The paths in snapshot at 4 ns, 12 ns, 38 ns and X-ray structure are shown with blue, red, green and black spheres, respectively. The side-chain orientations, thus the communication and the maximum likelihood pathways in the network change during MD simulations.

The path between the tips of the flexible loops (Thr172) is also computed by the AlloPathFinder[1] package. The underlying assumption in AlloPathFinder (Tang *et al.*, *unpublished*) is that residues participating in allosteric communication should be fairly

[1] The package is downloadable from 'https://simtk.org/home/allopathfinder'

conserved and that they should be close in space (r-cut: 3.25 Å). The path computed by the program is shown in Figure 4.22. The source and the destination are colored in green, whereas the remaining residues on the pathway are colored in orange. The path passes through the following residues (the conserved residues are written in boldface): **172**-**168**-**166**-**165**-**96**-**98**-**77**-78-66-**112**-**104**-**398**-**397**-**396**-**465**-**466**-**470**-**472**. The same pathway computed by the new methodology on X-ray structure of cTIM is shown in Figure 4.21 (black spheres). The comparison of pathways depicts that both pathways are composed of conserved residues. However, the path of the new methodology is shorter than the path computed by the server. Thus, it may be concluded that, the pathways evaluated from Dijktra's shortest path algorithm is more probable.

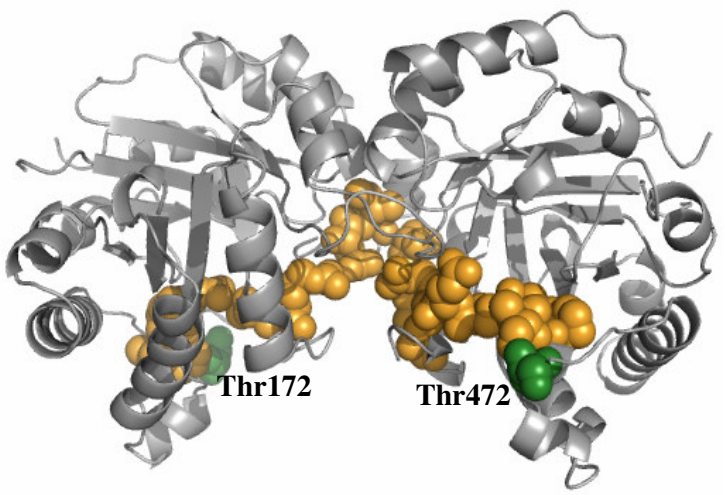


Figure 4.22. Pathway between the tips of the flexible loops of the adjacent subunits

4.2.2 Complex Form of cTIM

Analyzing the complex form of the enzyme and comparing it with the free form elucidates the changes in the enzyme upon ligand binding. Furthermore, the communication of the residues with the inhibitor, PGH, clarifies the communication patterns of the enzyme. The catalytic residues Lys 13 (green), His95 (orange), Glu165 (cyan), the active loop (dark pink), the interdigitating loop (red) and the bound PGH (black) are shown in Figure 4.23.

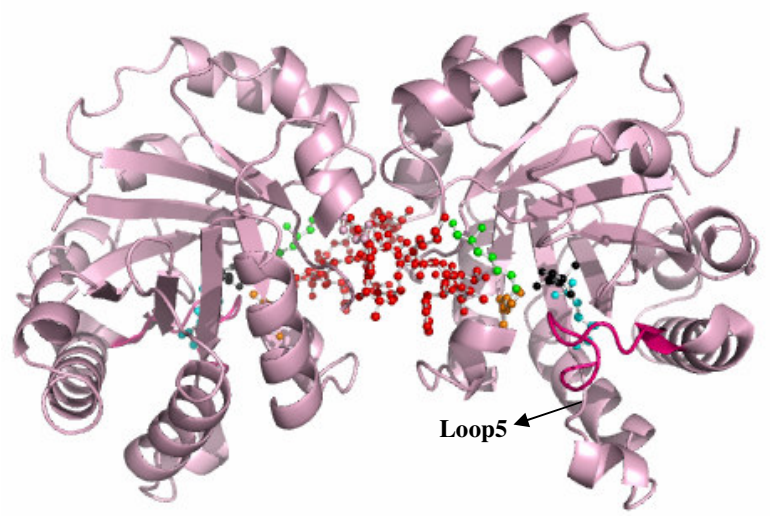


Figure 4.23. Complex Form of cTIM (1TPH)

The commute time distribution of the complex form of the enzyme, shown in Figure 4.24, reveals that the residues in loop5 (Lys130-Glu135) communicates poorly with the other residues in the network. This might be due to the changes in loop5 as a result of the closure of the active loop (loop6). (Wierenga *et al.*, 1992) Asp152 is another residue that is a poor communicator. This is, probably, due to the side-chain orientation, the side-chain points out into solution and hence it does not have so many contacts to communicate in the protein network. On the other hand, Ser211, Gly232 and Gly233 are good communicators of the network. These residues are known to have hydrogen bonds with PGH and the corresponding hydrogen bonds stabilize and fix the substrate molecule in a reactive conformation. (Zhang *et al.*, 1994)

The difference in commute times of complex and free forms of cTIM is given in Figure 4.25 to illustrate the changes upon the ligand. The negative (positive) values indicate that $C_{complex}$ is lower (higher) than C_{free}. The presence of the inhibitor (PGH) changes the contact map and the communication in the network. The flexible loop becomes a better communicator in the bound form. In the open structure, residues 171, 173, 175, and 176 make no hydrogen bonds to the rest of the protein, whereas in the closed structure only residue 175 makes no hydrogen bond. (Joseph *et al.*, 1990) The difference in the interactions changes the communication pattern in the enzyme. Besides the flexible loop,

there is a distinct change in some residues between Asp225 and Phe240. The residues in this region are known to interact with the substrate.

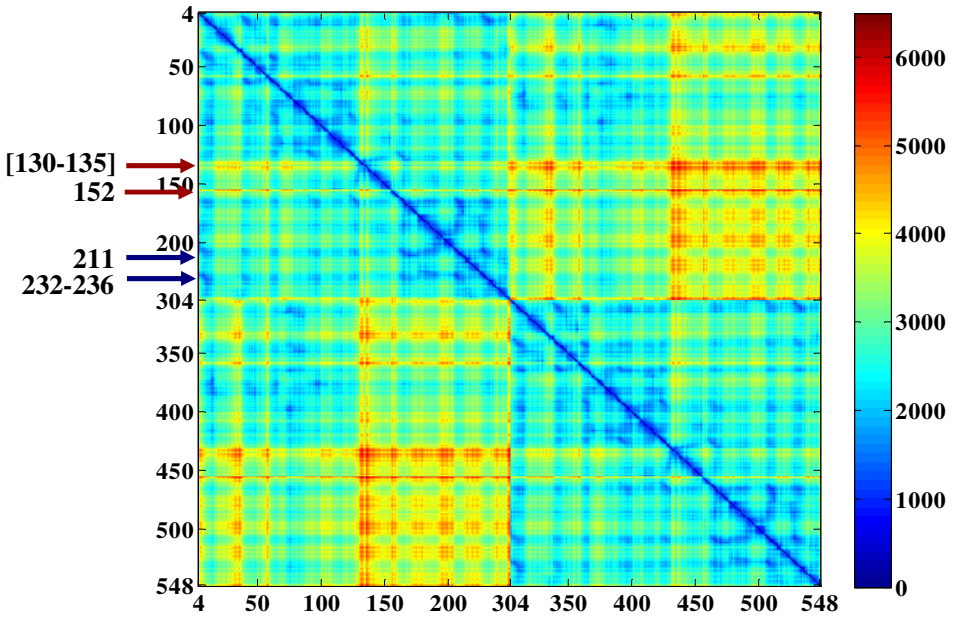


Figure 4.24. Commute time distribution in cTIM-PGH complex

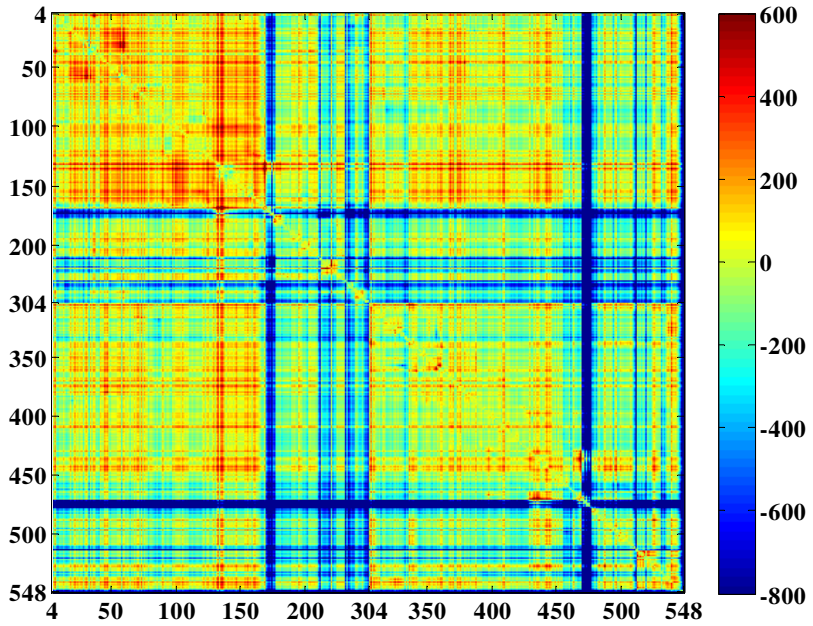


Figure 4.25. Changes in commute times upon ligand binding

Mean commute times of the residues in free and complex forms of cTIM are shown in Figure 4.26. In general, the commute times of residues in the complex form are slightly lower than the free form. There is a significant change in the mean commute times for the flexible loops.

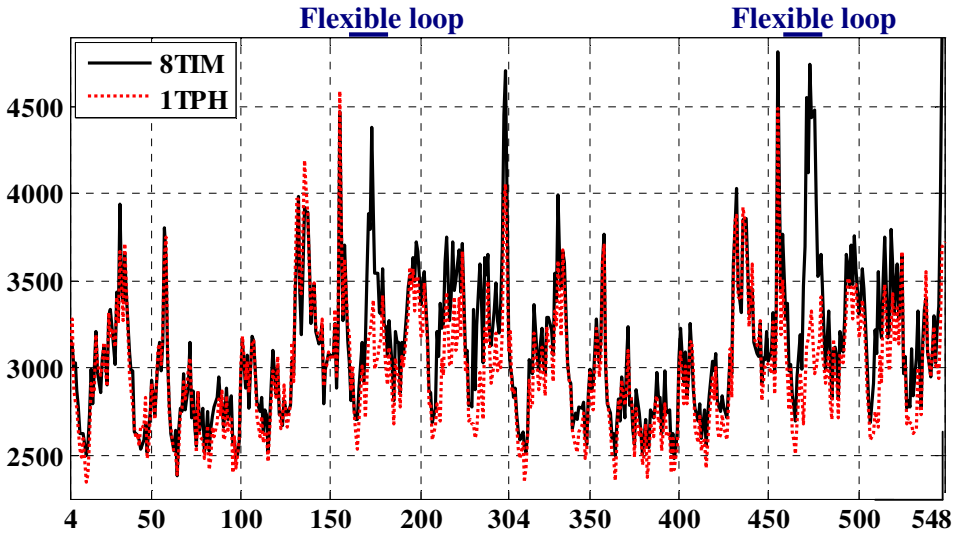
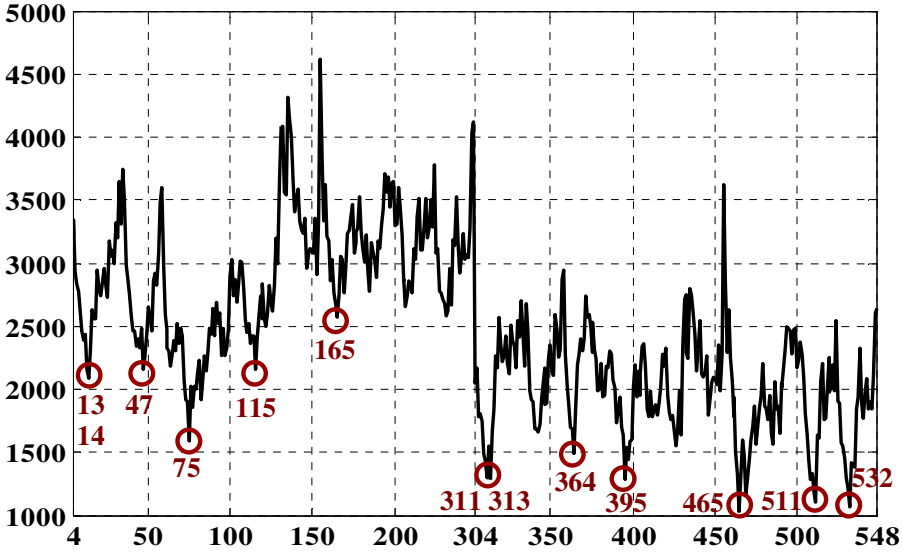
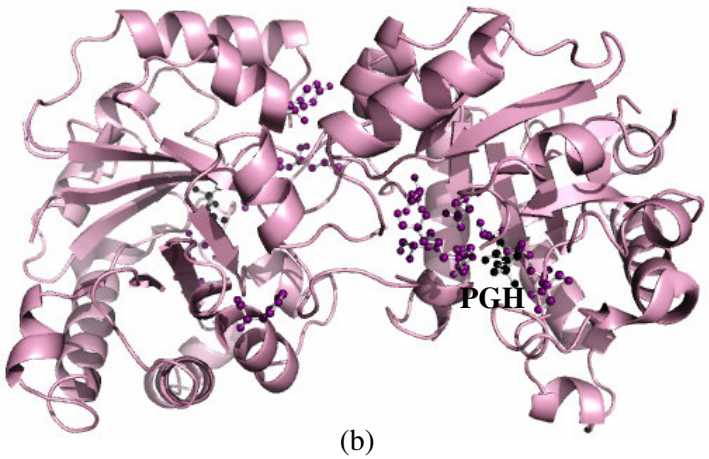


Figure 4.26. Mean Commute Times of the free and complex forms of cTIM

Previous plots depict that the presence of the inhibitor affects the communication stochastics of cTIM. The interactions of the residues in the network with the PGH bound to subunit B are illustrated in Figure 4.27 (a). The commute time between PGH and the tip of the interdigitating loop of the adjacent subunit (Thr75) is less than the commute times between PGH and many residues in the same subunit (subunit B). The residues marked with red circles are at the minima, thus have low commute times with PGH. The residues at the minima of subunit A are the residues that are catalytically important (Lys13, Glu165) or taking part in intersubunit interactions. (Met14, Tyr47, Thr75 and His115) Marked residues have high conservation score, hence are supposed to be functionally important. These residues are shown on the structure in Figure 4.27 (b).



(a)

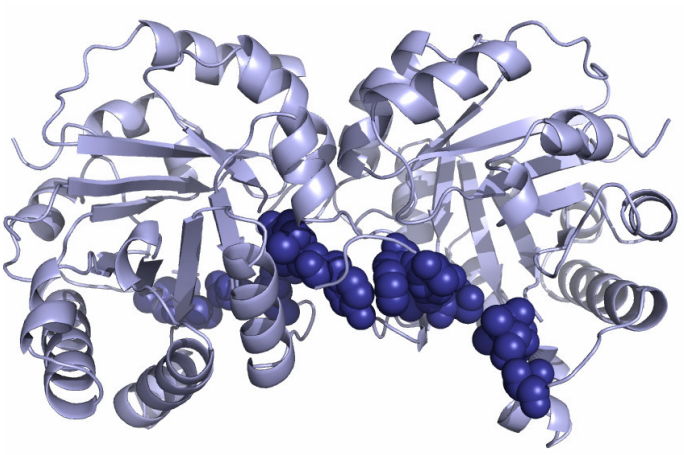


(b)

Figure 4.27. Commute times of residues with the PGH (bound to subunit B) (a) Interacting residues shown on the structure (b)

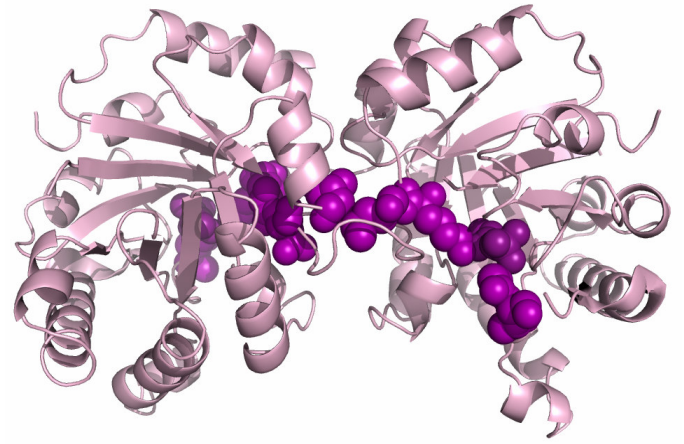
In a recent study carried out by Daily and Gray, the allosteric and non-allosteric proteins are examined and compared according to the rearrangement in residue-residue contacts upon the local conformational differences between the two structures of the proteins. (Daily and Gray, 2007) They analyzed cTIM as an example for the non-allosteric proteins. However, they have found out that like the allosteric proteins, triosephosphate isomerase presents motion which is small relative to the size of the enzyme. Thus, the key

interactions and communication pathways in cTIM are analyzed to determine the potential targets for controlling the function. Figure 4.28 shows the maximum likelihood pathways between the tips of the flexible loops in the free (blue) and complex (pink) forms of the enzyme, respectively. The residues on these pathways are highly conserved, suggesting that they are functionally important.



172 – 169 – 167 – 96 – 98 – 65 – 66 – 67 – 74 – 73 – *397* – *396*– *467* – *470* – *472*

(9) (9) (9) (9) (9) (9) (3) (5) (6) (9) (9) (9) (9) (9) (9)

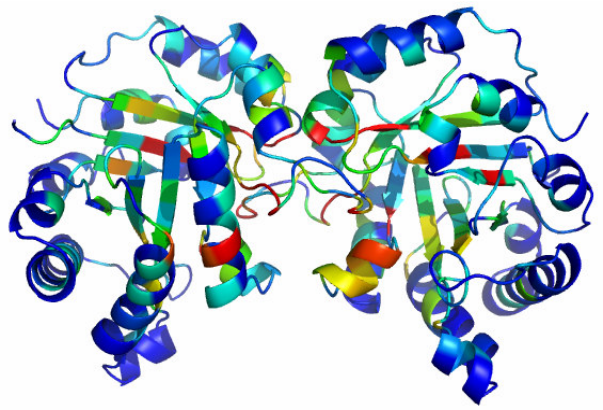


172 – 171 – **PGH** – 13 – 64 – 65 – 75 – *314* – *313*– ***PGH*** – *471* – *472*

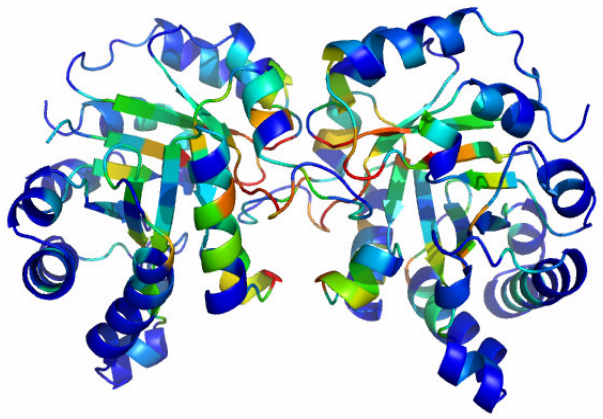
(9) (9) (9) (9) (9) (9) (8) (9) (9) (9)

Figure 4.28. Conserved residues on communication pathways

In Figure 4.29, the cartoon representations for 8TIM and 1TPH are colored according to the frequency with which the residues participate in the most probable communication pathways for any pair of residues in the network. The residues with 'hot' colors (red, orange) correspond to the most frequently visited ones. The binding of the ligand leads to changes in the most visited residues. However, in both cases approximately 90% of the 'hot' colored residues are highly conserved.



(a)



(b)

Figure 4.29. Most frequently visited nodes in (a) 8TIM and (b) 1TPH

The residue pairs which are frequently visited for the paths between the subunits in 8TIM are: 73-397 (97-373), 45-345, 11-375 (75-311). Among the stated residues, all except Ser45 are strictly conserved. There are hydrogen bonds between 45-345 and 11-375 (75-311). For position 45, the interaction between the residue in adjacent subunits differs

according to the type of the aminoacid. For instance, in typosonomal TIM there is phenylalanine in position 45 that is a part of the hydrophobic pocket in the subunit interface. Ser45-Ser45 residue pair is also one of the mosly visited pair in the paths between the subunits for the complex form of cTIM.

4.2.3. Human TIM

Although the hTIM and cTIM are almost identical in both sequence and structure, the presence of the ligand in one subunit, the conformational difference in the flexible loops and also the presence of deamidation sites in hTIM cause some differences in their communication patterns. The methodology is also applied to hTIM to compare the topology-driven information diffusion in cTIM and hTIM.

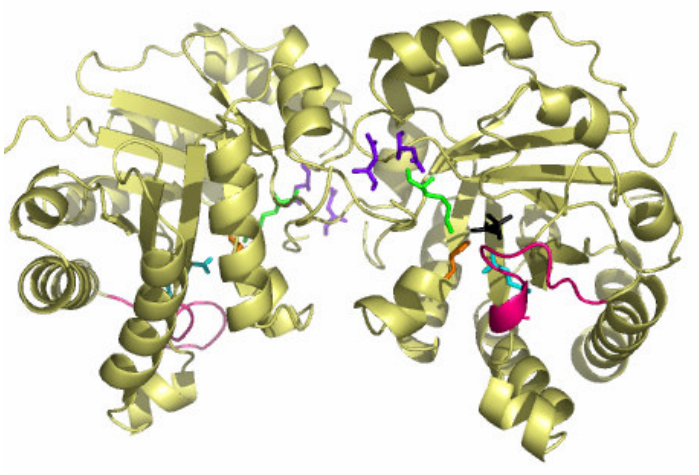


Figure 4.30. Human TIM, deamidation sites (Asn15 and Asn71) are colored in purple.

The following figure shows the commute time distribution for human TIM. This color-coded map looks so similar to that for cTIM; however there are some minor differences. The mean commute time values of residues in the network of cTIM (PDB ID: 8TIM) and hTIM ((PDB ID: 1HTI – Mande *et al.*, 1994)) are represented in Figure 4.32.The correlation between the mean commute time distribution of cTIM and hTIM is 0.882. Not surprisingly, the correlation in subunit A is higher between cTIM and hTIM. This is due to the presence of the inhibitor in subunit B in 1HTI. It has been presented by Haliloglu and coworkers that the binding of substrate only on one subunit provides space at the subunit interface which allows the side-chains of asparagine residues to adopt the

proper conformation for the deamidation reaction. (Konuklar *et al.*, 2006) The new negative charges at the subunit interface of hTIM lead to dissociation and unfolding. The aminoacid types at position 71 are different in cTIM and hTIM. There is Asn in hTIM whereas it is Lys in cTIM. The mean commute time distributions alter at the deamidation sites. These residues (Asn15 and Asn71) possess lower commute times in hTIM network; that is to say, they are faster communicators.

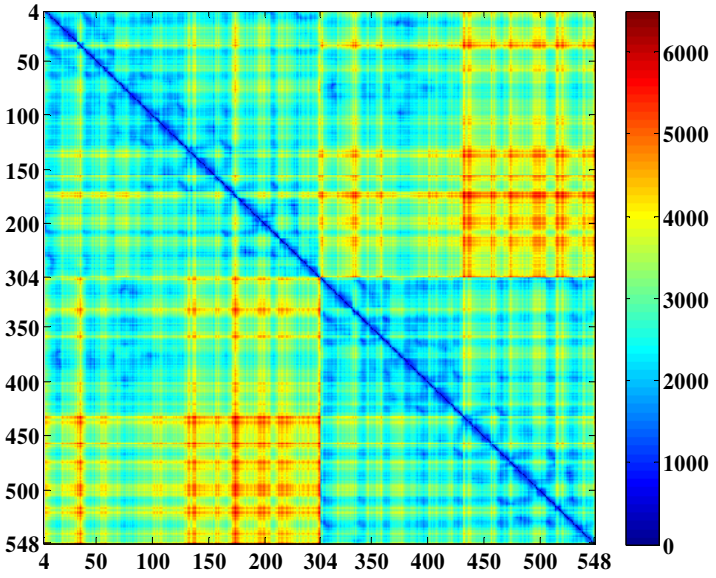


Figure 4.31. Commute Time Map for hTIM

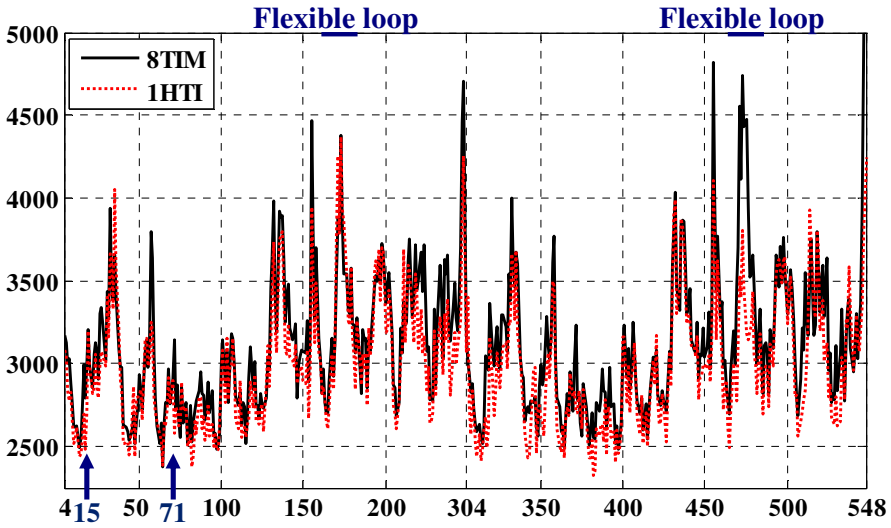


Figure 4.32. Mean Commute Times of cTIM and hTIM

Figure 4.33 shows how frequently the residues are visited in the shortest (most probable) paths. Like in the cTIM structures, the catalytic residues and the residues in contact with these residues (i.e. Met14, His95) are frequently visited, so represented in 'hot' colors. The importance of Met14 and Arg98 in the stability of the dimeric hTIM has been pointed out in a study by Goraj and coworkers. (Mainfroid *et al.*, 1996) These residues are in contact with the catalytic residues Lys13 and His95, respectively. Thus, the mutations in these residues might affect the interactions and thus preventing the correct positioning of the catalytic residues in the active site.

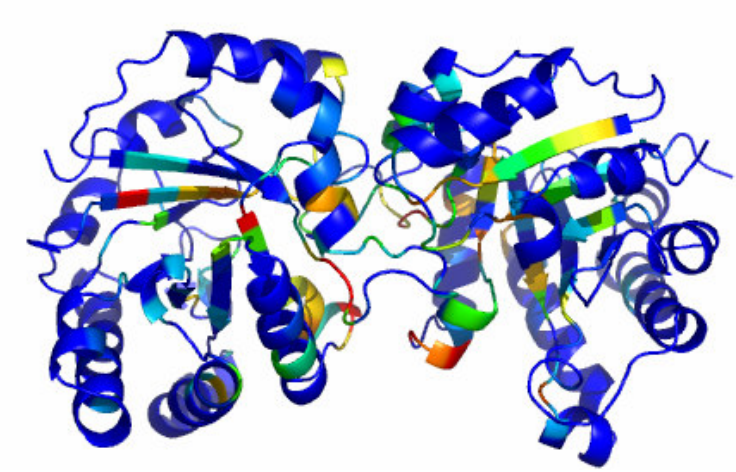


Figure 4.33. Frequently visited nodes in 1HTI

4.3. PDZ-3 Domain

PDZ domains are fundamental in regulating the dynamic organization of the cell. They are protein-protein interaction domains with approximately 100 amino acid residues. A variety of molecules can be docked with the PDZ domains and these bindings might be allosterically regulated. This is probably the reason why this small protein is frequently being analyzed in the studies about the intra-molecular signaling between distant sites within a protein. In PDB, there are both the bound and unbound X-ray structures for the third PDZ domain. (PDZ-3) The missing atoms are added to the structure according to the topology of the residues by using AMBER. The communication patterns in the new structures (without missing atoms) are examined in this study.

Figure 4.34 shows the superimposed structures of third PDZ domain of PSD-95, the unbounded and bounded forms are colored in light and dark orange, respectively. The bound peptide is shown in stick representation and the residues forming the hydrophobic pocket (Leu323, Phe325, Ile327 and Leu379) in the binding groove are shown with spheres. The PDZ-3 structure is composed of two α-helices (αA: Pro346-Ser350 & αB: His372-Asn381) and six β-strands. The RMS difference between the bounded and unbounded forms is roughly 0.6 Å. PDZ-3 is an example of the proteins where the communication occurs with minimal structural change. (Fuentes *et al.*, 2004) The peptide binding induces displacements, mostly, in loop1 and αB.

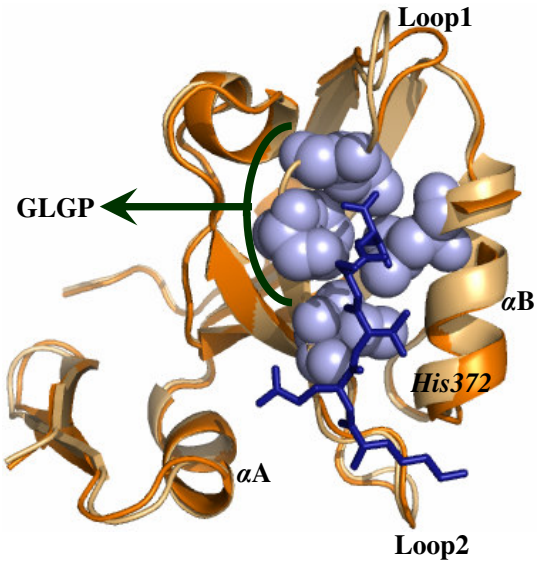


Figure 4.34. PDZ-3 domain unbounded and bound structures

The presence of the highly conserved 'GXGX' motif in the binding groove is one of the general features of the PDZ domains; this binding groove particularly targets peptides with hydrophobic C-terminal. The key residue that is crucial for ligand specificity is His372. (van Ham and Hendriks, 2003) Mackinnon and coworkers stated that this residue forms hydrogen bonds with the ligand. (Doyle *et al.*, 1996) His372 is labeled in Figure 4.34.

The hitting (a) and commute time (b) distributions for the unbounded (PDB ID: 1BFE – Doyle *et al.*, 1996) and the bounded (PDB ID: 1BE9 – Doyle *et al.*, 1996) forms of PDZ-3 are shown in Figure 4.35 and 4.36 respectively.

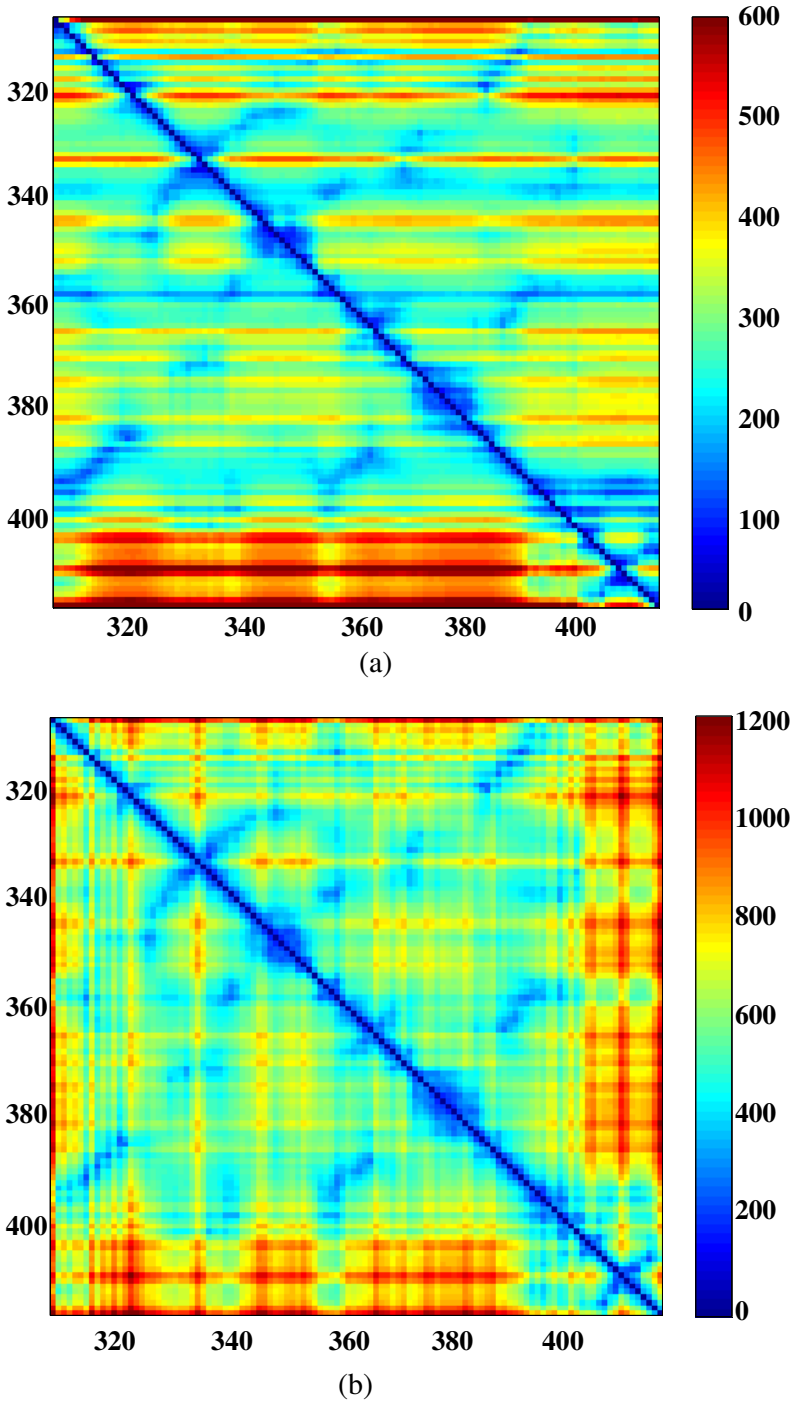


Figure 4.35. Hitting and Commute Time Maps for unbounded PDZ-3

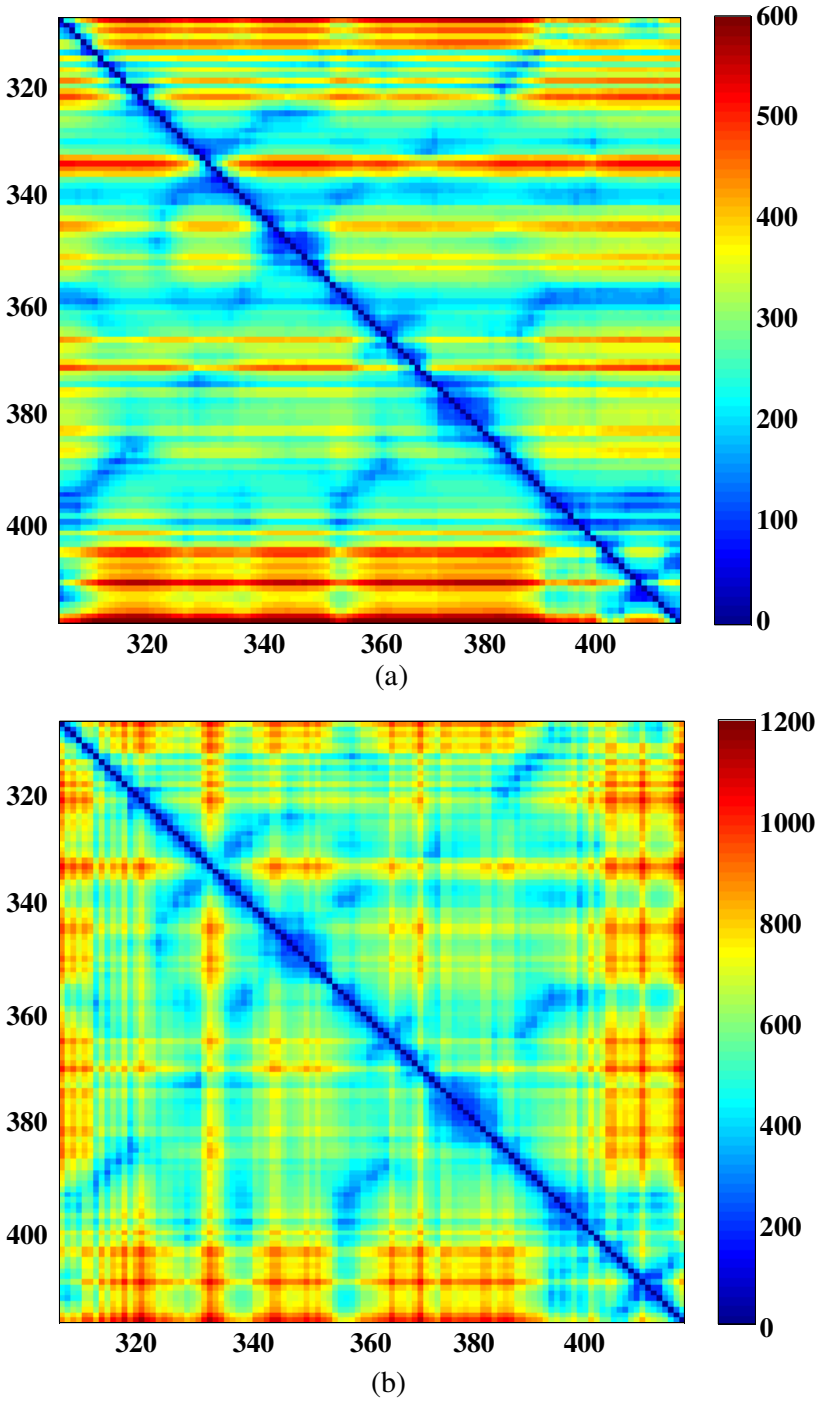


Figure 4.36. Hitting and Commute Time Maps for bounded form of PDZ-3

When the unbounded and bounded are compared, there are insignificant changes in the communication propensity of the residues. The difference between the commute times of each residue pairs between the bounded and unbounded forms is shown in Figure 4.37.

This figure exposes that the communication ability of Glu310, Glu331-Glu334 and Asn369 decrease; whereas the communication ability of Arg313, His372, Ser409 and the chain ends increase significantly. The residues with decreasing (increasing) communication propensity are colored in red (blue) and shown in Figure 4.38. The changes, except Ser409, are typically on the residues close to the bound peptide. Ser409 is close to the chain ends, this is probably the reason of the change in communication ability, thus the changes in these residues can be ignored as it was done in the NMA study on PDZ (De Los Rios *et al.*, 2005).

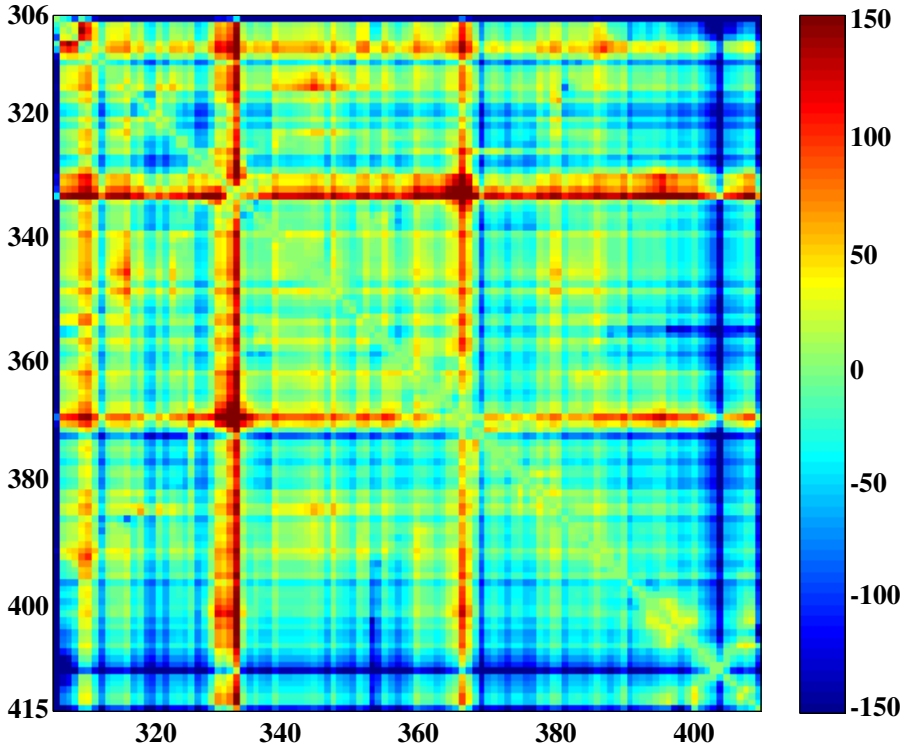


Figure 4.37. Changes in Commute Times in PDZ upon peptide binding

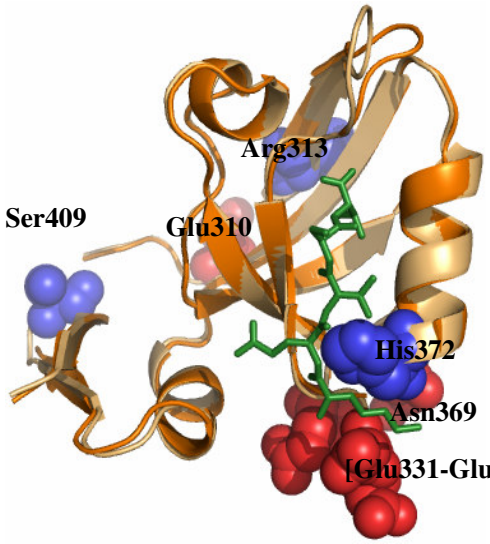


Figure 4.38. Residues with considerable changes in communication upon peptide binding

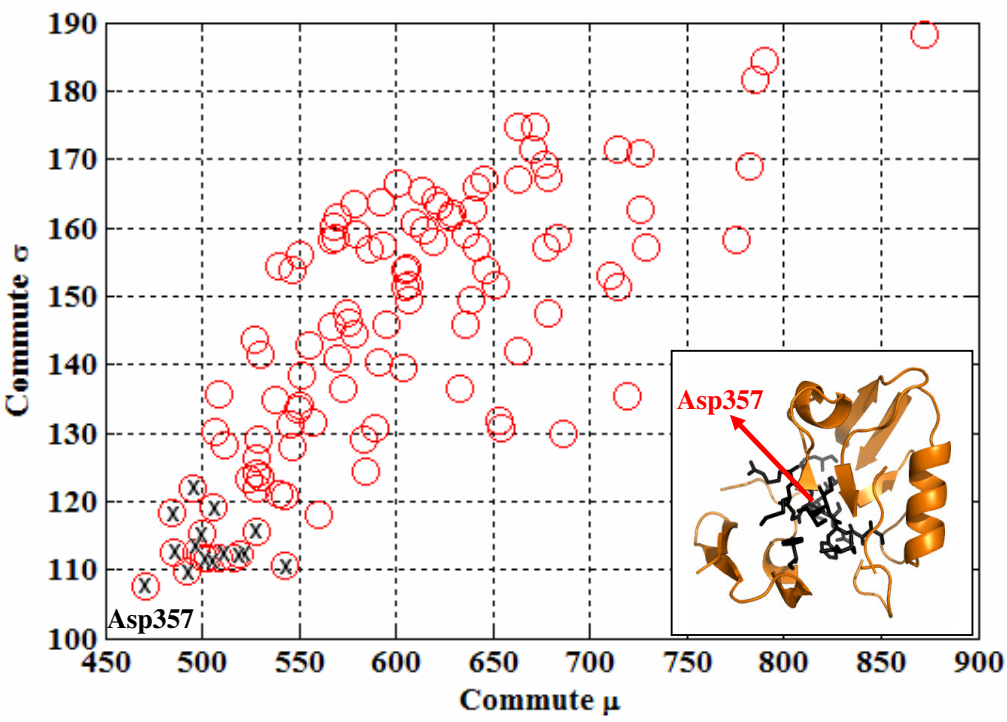


Figure 4.39. Communication Ability of residues in PDZ-3

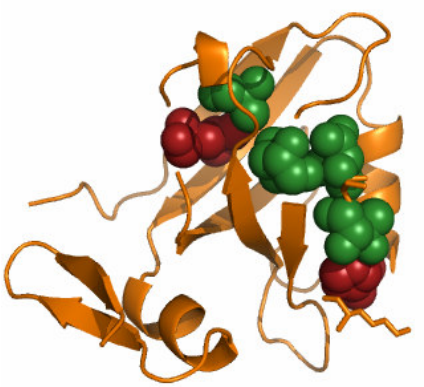
Figure 4.39 reveals the distribution of the communication ability of the residues in PDZ. The 'fast and precise' communicators forming the core of the protein are colored in

black. The residue on the lower left of the distribution is Asp357. Although it has not been pointed out to be a key residue, it has a high conservation score.

The methods about intramolecular signaling pathways, SCA (statistical coupling analysis) (Lockless and Ranganathan, 1999) and ATD (anisotropic thermal diffusion) (Ota and Agard, 2005), have also been applied to the PDZ domain family proteins. In the stated studies His372 is picked out as the source. Despite being on the opposite site of the protein Leu353 is coupled with His372, and in both studies Leu353 is chosen as the destination node. This residue pair puts forth a potential allosteric pathway, the functional importance of which is confirmed by the experimental mutagenesis. The path in Lockless and Ranganathan's study (a), Ota and Agard's study (b) and the new methodology (c) are displayed in Figure 4.40, where the conservation scores (evaluated from ConSurf, (Landau *et al.*, 2005)) of the residues are listed in parenthesis. The residue numbers of the bound peptide are designated from -4 to 0, as Val0, Ser-1, Thr-2, Gln-3 and Lys-4.

The paths from SCA and the new methodology pass through the peptide, but the path from ATD method does not. Furthermore, the paths from ATD and SCA pass through Phe325 which is, as a result, supposed to be a key residue taking part in mediating the signal propagation for ligand binding. The pathway computed by the new methodology does not pass through Phe325; instead it passes through Ser339 and Ile338 which have higher conservation scores than Phe325.

Figure 4.40 also depicts that different methodologies lead to different pathways. The communication path evaluated for PDZ family by SCA represents the residues shared by majority of PDZ domains. However, individual PDZ domains might have different key residues and, hence, additional pathways. ATD and the new methodology reveal the communication paths in the third PDZ domain, actually in PDB entry 1BE9 according to its topology.

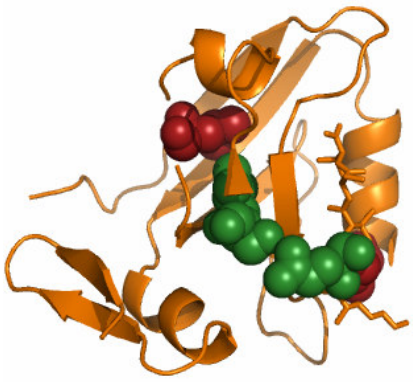


His372 (9) **–** ***Thr-2*** **–** ***Val0*** **–** ***Phe325*** (7) **–** ***Ala347*** (9) **–** ***Leu353*** (9)

(a)

His372 (9) **–** ***Ile327*** (9) **–** ***Phe325*** (7) **–** ***Ile341*** (8) **–** ***Leu353*** (9)

(b)



His372 (9) **–** ***Gln-3*** **–** ***Ser339*** (9) **–** ***Ile338*** (8) **–** ***Leu353*** (9)

(c)

Figure 4.40. Paths between His372 and Leu353

4.4. Cdc25B

Cdc25B (cell division cycle 25 homolog B) is a phosphatase that is involved in the start of the cell cycle. It contains the active site motif H-C-5X-R, where H is a highly conserved histidine residue, C is the catalytic cysteine, 5X motif forms a loop wherein the five amide nitrogens form hydrogen bond with the phosphate of the protein substrate, and R is the conserved arginine in the active site.

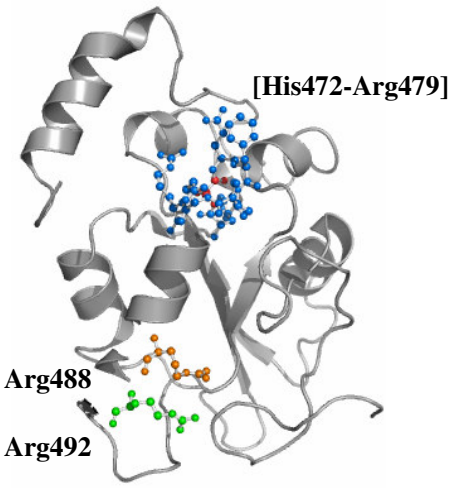


Figure 4.41. X-ray structure of Cdc25B

In this study the structure determined by Watenpaugh and coworkers (PDB ID: 1QB0) is investigated. (Reynolds *et al.*, 1999) The X-ray structure is shown in Figure 4.41, where the catalytic site is colored in blue and the catalytic cysteine in red; the two arginines: Arg488 and Arg492, which are the remote hot spots, are colored in orange and green, respectively.

An improper activation of Cdc25B is a general feature of cancer and tumor development. Therefore, the inhibition of the phosphatase might be used to generate novel anti-cancer agents. The computational and experimental docking studies are applied to achieve this target. The hot spot arginines (Arg488 and Arg492) are supposed to be crucial for the docking studies. (Rudolph, 2007) Bahar and coworkers have performed docking studies on Cdc25B and noticed that the inhibitors, mostly, bind to Arg488. (Bakan, 2007)

The new methodology is applied to Cdc25B to determine and compare the communication abilities of these arginine residues and to elucidate the validity of the method.

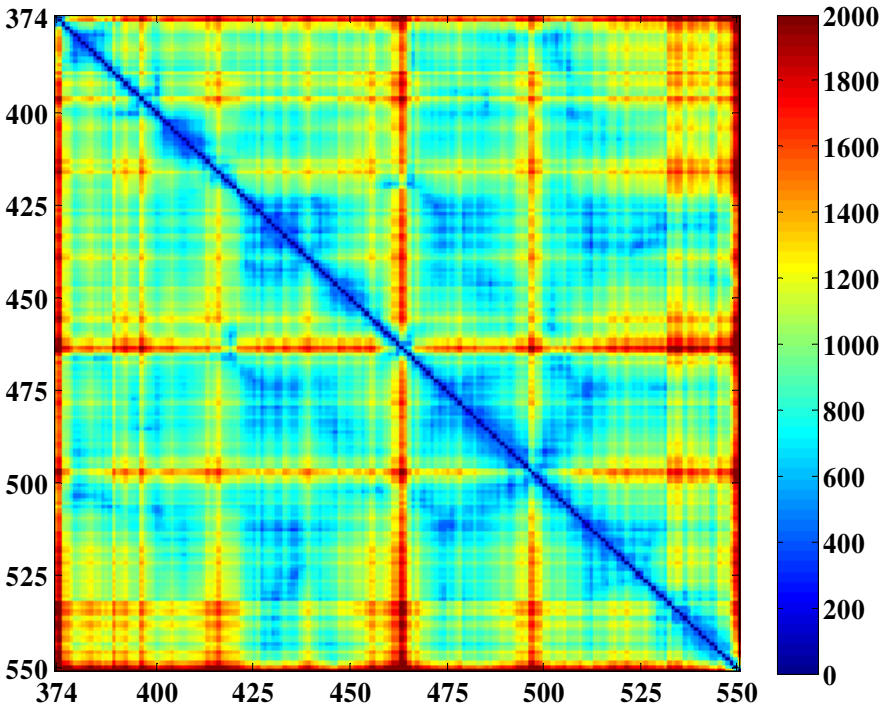


Figure 4.42. Commute Time Map for Cdc25B

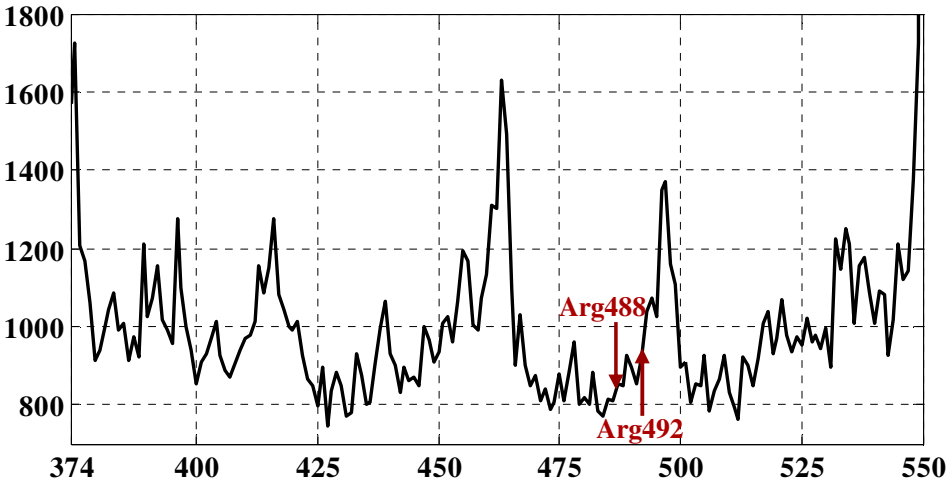


Figure 4.43. Mean Commute Time Distribution for Cdc25B

Figure 4.42 shows the commute time distribution in the network constructed for Cdc25B and Figure 4.43 shows the mean commute time values for each residue. The mean commute time of Arg488 is lower than that of Arg492. That is to say, Arg488 is a faster communicator compared to Arg492. This plot demonstrates the difference barely, thus the standard deviation versus mean commute time plot is prepared for Cdc25B and displayed in Figure 4.44.

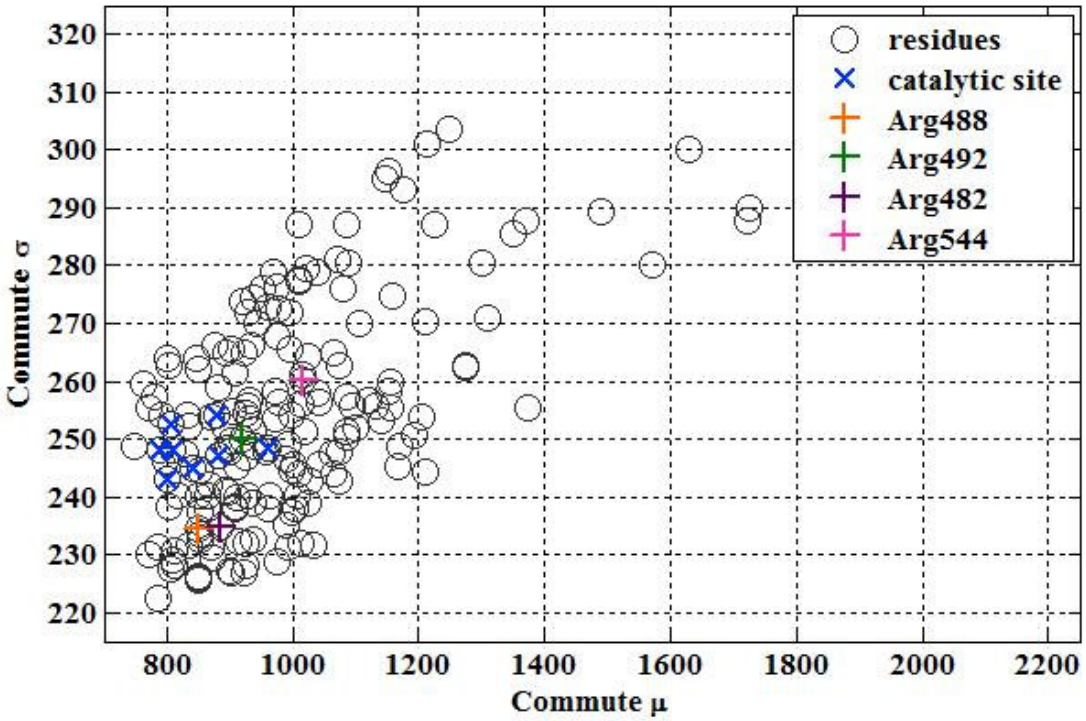


Figure 4.44. Communication Ability of residues in Cdc25B

In this plot the catalytic site residues, two hot spot arginines and another two arginines which are proposed to be important for reversible enzyme inhibition in the docking study of Novellino and coworkers. (Lavecchia *et al.*, 2006) Arg488 has the best communication ability among the stated residues. Thus, this methodology supports the results obtained from experimental and computational docking studies.

5. CONCLUSIONS

Allosteric communication is a fundamental process of signal propagation from one site of a protein to the functionally important distant site. Although it has been studied for decades the basis of allostery or intra-molecular communication is still an intricate subject. The present study demonstrates the application of a new methodology that proposes a Markovian process of information diffusion over the network of residues. (Chennubhotla and Bahar, 2006) It is stated that the topology-driven dynamics underlies function and the motions determine communication patterns that are inherent to the native architecture. The connection between the fluctuation dynamics predicted by physics-based elastic network models and information theory-based Markov stochastics is established by this novel approach. Joining these two disciplines is carried out in identification of critical interactions and their functional effects.

Several observations arise from analyzing the flow of information in the specific examples (Adenylate kinase, TIM, PDZ and Cdc25B). It has been inferred that hitting and commute times are information theoretic concepts which have physical origins, i.e. they are related to the intrinsic motions of a protein as determined by their inter-residue contact topology. The residues at the minima of the mean hitting or commute time curves have very efficient signal transduction abilities. It has also been pointed out that pairs of residues subject to smaller fluctuations have more efficient communication, and vice versa. The communication abilities of the residues differ according to their position in the structure; residues in the protein core are fast (in terms of mean value) and precise (in terms of standard deviation) communicators. The catalytic residues are illustrated to be the efficient communicators of the network. The modes of elastic network models are examined closely to depict the connections between the communication and the fluctuations, the global modes contribute significantly to the communication stochastics. The determination of the shortest paths between residue pairs is one of the important results of the methodology. The shortest paths computed in the study can be characterized as the maximum likelihood pathways because they are evaluated according to the transition probabilities between the residue pairs. The residues on the communication pathways are, at large, evolutionarily conserved; this confirms their functional importance in the network.

This methodology might as well be useful for exploring the communication in multi-molecular interaction networks. Identification of key residues and examining their effects in allosteric responses will be a major step towards gaining a better understanding of the basis of signal transduction mechanisms.

REFERENCES

Ascenzi, P., A. Bocedi, A. Bolli, M. Fasano, S. Notari and F. Polticelli, 2005, "Allosteric Modulation of Monomeric Proteins", *Biochemistry and Molecular Biology Education*, Vol. 33, pp. 169-176.

Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, 2001, "Anisotropic of fluctuation dynamics of in proteins with an elastic network model", *Biophysical Journal*, Vol. 80, pp. 505-515.

Atilgan, A. R., D. Turgut, C. Atilgan, 2007, "Screened non-bonded interactions in native proteins manipulate optimal paths for robust residue communications", *Biophysical Journal*, Vol. 92, pp. 3052-3062.

Bahar, I. and A. J. Rader, 2005, "Coarse-grained normal mode analysis in modeling dynamics of biomolecular complexes", *Current Opinion in Structural Biology*, Vol. 13, pp. 373-380.

Bahar I., B. Erman, T. Haliloglu and R. L. Jernigan, 1997a, "Efficient Characterization of Collective Motions and Interresidue Correlations in Proteins by Low-Resolution Simulations", *Biochemistry*, Vol. 36, pp. 13512-13523.

Bahar, I., A. R. Atilgan and B. Erman, 1997b, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential", Folding and Design, Vol. 2, pp. 173-181.

Bakan, A., 2006, *Personal Communication*

Banner, D. W., A. C. Bloomer, G. A. Petsko, D. C. Phillips, C. I. Pogson, I. A. Wilson, P.H. Corran, A. J. Furth, J. D. Milman, R. E. Offord, J. D. Priddle and S.G. Waley, 1975, "Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution: using amino acid sequence data", *Nature*, Vol. 255, pp. 609-614.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, 2000, "The Protein Data Bank", *Nucleic Acids Research*, Vol. 28, pp. 235-242.

Borchert, T. V., R. Abagyan, R. Jaenicke and R. K. Wierenga, 1994, "Design, creation, and characterization of a stable, monomeric triosephosphate isomerase", *Proceedings of the National Academy of Sciences*, Vol. 91, pp. 1515-1518.

Branden C. and Tooze J., 1999, *Introduction to Protein Structure*, Garland Publishing, Inc., New York.

Brooks, B. R., D. Janezic and M. Karplus, 1995, "Harmonic Analysis of Large Systems", *Journal of Computational Chemistry*, Vol. 16, pp. 1522-1542.

Cansu, S. and P. Doruker, 2007, *unpublished.*

Cao, J. and B. J. Berne, 1990, "Monte Carlo methods for accelerating barrier crossing: Anti-force-bias and variable step algorithms", *Journal of Chemical Physics*, Vol. 93, pp. 1980-1985.

Changeux, J. P. and S. J. Edelstein, 2005, "Allosteric Mechanisms of Signal Transduction", *Science*, Vol. 308, pp. 1424-1428.

Chen, J., 2003, "Dijkstra's Shortest Path Algorithm", *Formulized Mathematics*, Vol. 11, pp. 237-247.

Chennubhotla, C. and I. Bahar, 2006, "Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES", *Molecular Systems Biology*, Article no. 36.

Chennubhotla C. and I. Bahar, 2007, "Markov Propagation of Signals in Proteins and its Relation to Equilibrium Fluctuations", *submitted to Public Library of Science Computational Biology*.

Chennubhotla, C., A. J. Rader, L. Yang, I. Bahar, 2005, "Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies", *Physical Biology*, Vol. 2, pp. 173-180.

Chennubhotla, C., S. Ozel, P. Doruker and I. Bahar, 2007, "Topology-Driven Information Diffusion and Its Relation to Protein Function", *unpublished*.

Chennubhotla, C., 2006, *Personal Communication*.

Cooper A. and D. T. F. Dryden, 1984, "Allostery without conformational change", *European Biophysics Journal*, Vol. 11, pp. 103-109.

Creighton, T. E., 1993, *Proteins: Structures and Molecular Properties*, W. H. Freeman and Company, New York.

Cui, Q. and I. Bahar, 2006, *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, CRC Press, Taylor & Francis Group, Florida.

Daily, M. D. and J. J. Gray, 2007, "Local Motions in a Benchmark of Allosteric Proteins", *Proteins*, Vol. 67, pp. 385-399.

De Los Rios, P., F. Cecconi, A. Pretre, G. Dietler, O., Michielin, F. Piazza and B. Junico, 2005, "Functional Dynamics of PDZ Binding Domains: A Normal- Mode Analysis", *Biophysical Journal*, Vol. 89, pp. 14-21.

Deane, C. M. and T. L. Blundell, 2000, "A novel exhaustive search algorithm for predicting the conformation of polypeptide segments in proteins", *Proteins*, Vol. 40, pp. 135-144.

del Sol, A., H. Fujihashi, D. Amoros and R. Nussinov, 2006, "Residues crucial for maintaining short paths in network communication mediate signaling in proteins", *Molecular Systems Biology*, Article no. 19.

Demirel, M. C. and O. Keskin, 2005, "Protein Interactions and Fluctuations in a Proteomic Network using an Elastic Network Model", *Journal of Biomolecular Structure & Dynamics*, Vol. 22, pp. 381-386.

Dijkstra, E. W., 1959, "A note on two problems in connexion with graphs", *Numerische Mathematik*, Vol. 1, pp. 269-271.

Doyle, D.A., A. Lee, J. Lewis, E. Kim, M. Sheng and R. MacKinnon, 1996, "Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ", *Cell*, Vol. 85, pp. 1067-1076.

Flory, P. J, 1976 "Statistical Thermodynamics of random networks", *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 351, pp. 351-378

Fuentes, E. J., J.D. Channing and A. L. Lee, 2004, "Ligand-dependent Dynamics and Intramolecular Signaling in a PDZ Domain", *Journal of Molecular Biology*, Vol. 235, pp. 1105-1115.

Ghosh, A., K. V. Brinda and S. Vishveshwara, 2007, "Dynamics of Lysozyme Structure Network: Probing the Process of Unfolding", *Biophysical Journal*, Vol. 92, pp. 2523-2535.

Gunasekaran, K., B. Ma and R. Nussinov, 2004, "Is Allostery an Intrinsic Property of All Dynamic Proteins?", *Proteins*, Vol. 57, pp. 433-443.

Haliloglu, T., I. Bahar and B. Erman, 1997, "Gaussian Dynamics of Proteins", *Physical Review Letters*, Vol. 79, pp. 3090-3093.

Haliloglu T. and I. Bahar, 1998, "Coarse-Grained Simulations of Conformational Dynamics of Proteins: Application to Apomyoglobin", *Proteins*, Vol. 31, pp. 271-281.

Hill, A. V., 1910, "The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves", *The Journal of Physiology*, Vol. 40, pp. 4-8.

Jacob, F. and J. Monod, 1961, "Genetic regulatory mechanisms in the synthesis of proteins", *Journal of Molecular Biology*, Vol. 3, pp. 318-356.

Joseph, D., G. A. Petsko and M. Karplus, 1990, "Anatomy of a conformational change: Hinged lid motion of the triosephosphate isomerase loop", *Science*, Vol. 249, pp. 1425-1428.

Konuklar, F. A. S., V. Aviyente and T. Haliloglu, 2006, "Coupling of Sstructural Fluctuations to Deamidation Reaction in Triosephosphate Isomerase by Gaussian Network Model", *Proteins*, Vol. 62, pp. 715-727.

Koshland, D. E., G. Nemethy and D. Filmer, 1966, "Comparison of experimental binding data and theoretical models in proteins containing subunits", *Biochemistry*, Vol. 5, pp. 365-385.

Kovacks, J. A., P. Chacon and R. Abagyan, 2004, "Predictions of protein flexibility: first order measures", *Proteins*, Vol. 56, pp. 661-668.

Krusek, J., 2004, "Allostery and Cooperativity in the Interaction of Drugs with Ionic Channel Receptors", *Physiological Research*, Vol. 53, pp. 569-579.

Kurkcuoglu, O., R. L. Jernigan and P. Doruker, 2006, "Loop Motions of Triosephosphate Isomerase Observed with Elastic Networks", *Biochemistry*, Vol. 45, pp. 1173-1182.

Landau, M., I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko and N. Ben-Tal, 2005, "*ConSurf 2005*: the projection of evolutionary conservation scores of residues on protein structures", *Nucleic Acids Research*, Vol. 33, pp. 299-302.

Lavecchia, A,, S. Cosconati, V. Limongelli and E. Novellino, 2006, "Modeling of Cdc25B Dual Specifity Protein Phosphatase Inhibitors: Docking of Ligands and Enzymatic Inhibition Mechanism", *ChemMedChem*, Vol. 1, pp. 540-550.

Lockless, S. W. and R. Ranganathan, 1999, "Evolutionarily Conserved Pathways and Energetic Connectivity in Protein Families", *Science*, Vol. 286, pp. 295-299.

Lu, M. and J. Ma, 2005, "The Role of Shape in Determining Molecular Motions", *Biophysical Journal*, Vol. 89, pp. 2395-2401.

Ma, J., 2005, "Usefulness and limitations of normal mode analysis in molecular dynamics of biomolecular complexes", *Current Opinion in Structural Biology*, Vol. 13, pp. 373-380.

Mainfroid, V., P. Terpstra, M. Beauregard, J. Frere, S. C. Mande, W. G. J. Hol, J. A. Martial and K. Goraj, 1996, "Three hTIM Mutants that Provide New Insights on why TIM is a Dimer", *Journal of Molecular Biology*, Vol. 257, pp. 441-456.

Mande, S. C., V. Mainfroid, K. H. Kalk, K. Goraj, J. Martial and W. G. J. Hol, 1994, "Crystal structure of recombinant human triosephosphate isomerase at 2.8 Å resolution. Triosephosphate isomerase-related human genetic disorders and comparison with typanosomal enzyme", *Protein Science*, Vol. 3, pp. 810-821.

Maragakis, P. and M. Karplus, 2005, "Large Amplitude Conformational Change in Proteins Explored with a Plastic Network Model: Adenylate Kinase", *Journal of Molecular Biology*, Vol. 352, pp. 807-822.

Massi, F., C. Wang and A. G. Palmer, III, 2006, "Solution NMR and Computer Simulation Studies of Active Site Loop Motion in Triosephosphate Isomerase", *Biochemistry*, Vol. 35, pp. 10787-10794.

McCammon J. A. and S. C. Harvey, 1994, *Dynamics of proteins and nucleic acids*, Cambridge University Press, Cambridge.

Ming, D., M. E. Wall, 2005, "Allostery in a Coarse-Grained Model of Protein Dynamics", *Physical Review Letters*, Vol. 95, Article no. 198103.

Miyazawa, S. and R. L. Jernigan, 1996, "Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading", *Journal of Molecular Biology*, Vol. 256, pp. 623-644.

Monod, J., J. Wyman, J. P. Changeux, 1965, "On nature of allosteric transitions: a plausible model", *Journal of Molecular Biology*, Vol. 12, pp. 88-118.

Muller, C. W. and G. E. Schulz, 1992, "Structure of Complex Between Adenylate Kinase from Escherichia coli and the Inhibitor Ap_5A Refined at 1.9 Å Resolution: A Model for a Catalytic Transition State", *Journal of Molecular Biology*, Vol. 224, pp. 159-177.

Muller, C. W., G. J. Schlauderer, J. Reinstein and G. E. Schulz, 1996, "Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding", *Structure*, Vol. 4, pp. 147-156.

Noda, L. H., 1973, *In the Enzymes*, Vol. 8, pp 279-305, Academic Press, New York.

Norris, J. R., 1997, *Markov Chain*, Cambridge University Press, Cambridge.

Ota, N. and D. A. Agard, 2005, "Intramolecular Signaling Pathways Revealed by Modeling Anisotropic Thermal Diffusion", *Journal of Molecular Biology*, Vol. 351, pp. 345-354.

Reynolds, R. A., A. W. Yem, C. L. Wolfe, M. R. Deibel Jr., C. G. Chidester and K. D. Watenpaugh, 1999, "Crystal structure of the catalytic subunit of Cdc25B required for G2/M phase transition of the cell cycle", *Journal of Molecular Biology*, Vol. 293, pp. 559-568.

Rudolph, J., 2007, "Inhibiting transient protein–protein interactions: lessons from the Cdc25 protein tyrosine phosphatases", *Nature Reviews- Cancer*, Vol. 7, pp. 202-211.

Schnackerz, K. D. and R. W. Gracy, 1991, "Probing the catalytic sites of triosephosphate isomerase by ^{31}P-NMR with reversibly and irreversibly binding substrate analogues", *European Journal of Biochemistry*, Vol. 199, pp. 231-238

Schulz, G., C. W. Muller and K. Deiderichs, 1990, "Induced-fit movements in adenylate kinases", *Journal of Molecular Biology*, Vol. 213, pp. 627-630.

Sudarsanam, S., R. F. DuBose, C. J. March and S. Srinivasan, 1995, "Modeling protein loops using a {phi}(i+1), {psi}(i) dimer database", *Protein Science*, Vol. 4, pp. 1412-1420

Suel, G. M., S. W. Lockless, M. A. Wall and R. Ranganathan, 2003, "Evolutionarily conserved networks of residues mediate allosteric communication in proteins", *Nature Structural Biology*, Vol. 10, pp. 59-70.

Sun, A. Q., K. U. Yuksel, G. S. J. Rao and R. W. Gracy, 1992a, "Effects of active-site modification and reversible dissociation on the secondary structure of triosephosphate isomerase", *Archives of Biochemistry and Biophysics*, Vol. 295, pp. 421-428.

Sun, A. Q., K. U. Yuksel and R. W. Gracy, 1992b, "Interactions between the Catalytic Centers and Subunit Interface of Triosephosphate Isomerase Probed by Refolding, Active Site Modification, and Subunit Exchange", *The Journal of Biological Chemistry*, Vol. 267, pp. 20168-20174.

Swain, J. F. and L. M. Gierasch, 2006, "The changing landscape of protein allostery", *Current Opinion in Structural Biology*, Vol. 16, pp. 102-108.

Tang S., J. Schmidt and J. Liao, 2007, *unpublished*.

Temiz, N. A., E. Meirovitch and I. Bahar, 2004, "Escherichia coli Adenylate Kinase Dynamics: Comparison of Elastic Network Model Modes with Mode-Coupling ^{15}N-NMR Relaxation Data.", *Proteins*, Vol. 57,. Pp. 468-480.

Thomas, P. D. and K. A. Dill, 1996, “An Iterative Method for Extracting Energy-like Quantities from Protein Structures”, *Proceedings of the National Academy of Sciences*, Vol. 93, pp. 11623-11633.

Tirion, M.M., 1996, “Large amplitude elastic motions in proteins from a single-parameter, atomistic analysis”, *Physical Review Letters*, Vol. 77, pp. 1905-1908.

van Ham, M. and W. Hendriks, 2003, “PDZ domains–glue and guide”, *Molecular Biology Reports*, Vol. 30, pp. 69-82.

Verlet, L., 1967, “Computer ‘Experiments’ on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”, *Physical Review*, Vol. 159, pp. 98-103.

Waley, S. G., 1973, “Refolding of Triose Phosphate Isomerase”, *Biochemical Journal*, Vol. 135, pp. 165-172.

Wierenga, R.K., T.V. Borcher and M.E.M Noble, 1992, "Crystallographic binding studies with triosephosphate isomerases: conformational changes induced by substrate and substrate- analogues", *Federation of European Biochemical Societies (FEBS)*, Vol. 307, pp. 34-39.

Zhang, Z., S. Sugio, E. A. Komives, K. D. Liu, J. R. Knowles, G. A. Petsko and D. Ringe, 1994, “Crystal Structure of Recombinant Chicken Triosephosphate Isomerase-Phosphoglycolohydroxamate Complex at 1.8 Å Resolution’, *Biochemistry*, Vol. 33, pp. 2830-2837.

Zheng, W. and B. R. Brooks, 2005, “Probing the Local Dynamics of Nucleotide-Binding Pocket Coupled to the Global Dynamics: Myosin versus Kinesis”, *Biophysical Journal*, Vol. 89, pp. 167-178.