ANALYSIS OF LIGAND BINDING EFFECT ON ENZYME DYNAMICS USING ELASTIC NETWORK MODELS

by

Hüseyin Doğa Fındık

B.S., Chemical Engineering, Boğaziçi University, 2012

Submitted to the Institute 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

2015

ACKNOWLEDGEMENTS

I would like to express my gratitude to my thesis advisor Prof. Pemra Doruker for her great knowledge, for her patience, and leading me throughout my work. I feel very thankful for being able to prepare this thesis with her leadership.

I also want to thank Prof. Türkan Haliloğlu and Prof. Demet Akten Akdoğan for reading my thesis.

Another point that I want to mention my thanks for is the graduate scholarship of TÜBİTAK that supported my project with the number 113M237.

None of this work was possible without Zeynep Kürkçüoğlu's support. I especially want to thank her for kindly sharing her data with me and helping me whenever I needed. I would like to express my gratitudes and thanks to Mehmet Tarık Can, Yasemin Yeşiltepe, and all the other members of the PRC, making this environment a family. Canan Dedeoğlu and Yüksel Yıldırım showed their sincere friendship during my studies.

My family believed in all my works and supported me in this sacrificing road whenever I needed. I would like to express my last gratitude to them for raising me to this point and appreciating me and my thesis.

I want to dedicate this work to endless human endeavor to acquire knowledge.

ABSTRACT

ANALYSIS OF LIGAND BINDING EFFECT ON ENZYME DYNAMICS USING ELASTIC NETWORK MODELS

The objective of this thesis was to investigate the effect of ligand binding on global vibrational modes of an enzyme, triosphosphate isomerase (TIM). For this aim, mixed coarse-grained elastic network model ENM (MCG_ENM) was used to observe the frequency shifts in the most collective modes due to the presence of the ligand. In MCG_ENM, the ligand is taken in atomistic detail and the protein is in low-resolution. First analysis was performed on previous blind docking results, which comprise a total of 1496 poses of five benzothiazoles, three of which are inhibitors of catalytic activity, docked on three different TIM conformers. Several different binding sites/poses were detected including the tunnel region, catalytic site and others. The inhibitors bound to the previously identified tunnel region could be differentiated based on their impact on global vibrational modes of the enzyme. Later, six independent MD runs (three apo and three complex with inhibitor bt10) were analyzed by performing MCG_ENM on a total of 54,000 MD snapshots of TIM. Through this computationally efficient technique, altered collective modes and positive shifts in eigenvalues were detected in the complex runs due to the constraining effect of inhibitor binding at the tunnel region. Lastly, a new computational technique was introduced for scanning protein side chains in terms of their constraining effect on ENM modes. In this methodology, the ligand binding effect in the vicinity of a specific residue is mimicked by adding extra nodes to its side chain atoms. ENM-based scanning of TIM also pinpointed to the tunnel region as a key binding site that can alter global dynamics of the enzyme. Scanning of other enzymes indicated marked constraining effect around the ligand positions observed in crystal structures.

ÖZET

ENZİM DİNAMİĞİ ÜZERİNE LİGAND ETKİSİNİN ELASTİK AĞ MODELLERİYLE İNCELENMESİ

Bu tezin amacı, triofosfat izomeraz (TIM) adlı enzime ligand bağlanmasının enzimin global titreşimsel modları üzerindeki etkilerini araştırmaktır. Bu amaçla, sisteme ligand eklenmesinin en kollektif modlarda meydana getirdiği frekans sapmalarını incelemek üzere, karışık kaba ölçekli elastik ağyapı modeli (MCG_ENM) kullanılmıştır. İlk analizler hâlihazırda literatürde bulunan, TIM yapısına beş adet benzotiazol molekülünün yerleştirildiği toplam 1496 docking pozuna uygulanmıştır. Tünel bölgesi olarak isimlendirilen bölgeye bağlanan inhibitörler, enzimin global titreşimsel modlarında meydana getirdikleri etkilerin görece farklı olması ile dikkat çekmişlerdir. Akabinde, (üçü sadece enzim molekülünü, diğer üçü ise enzim ile bt10 molekülünden müteşekkil kompleks sistemi içeren) altı ayrı moleküler dinamik (MD) simülasyon sonucu elde edilen toplam 54,000 konformasyona MCG_ENM metodu uygulanarak sonuçlar analiz edilmiştir. Söz konusu teknik ile, tünel bölgesine inhibitor molekülünün bağlanmasının kısıtlayıcı etkisi sebebi ile kompleks yapının kollektif modlarındaki değişimler ve frekanslarında meydana gelen pozitif yöndeki sapmalar belirlenmiştir. Son olarak, ENM kollektif modlarında meydana getirdikleri kısıtlayıcı etkileri bakımından protein yan gruplarını incelemek amacıyla yeni bir hesapsal metot geliştirilmiştir. Söz konusu metotda, belirli bir rezidü yakınına ligand bağlanma etkisi, rezidü yan grubu atomlarına ekstra birer eleman eklenerek taklid edilmiştir. TIM üzerinde yapılan ENM-temelli araştırmalar, tünel bölgesinin enzimin ana hareketlerini etkileyebilecek önemli bir bağlanma bölgesi olduğunu ortaya çıkarmıştır. Diğer kimi enzimler kristal yapılarında gözlemlenen ligand pozisyonları üzerinden incelendiğinde, TIM için bulunan sınırlandırıcı etkiler bu enzimlerde de önemli ölçüde gözlemlenmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

C_α	Alpha Carbon
$\mathbf{H}$	Hessian Matrix
$\mathbf{H}_{i,j}$	Hessian Super Elements
R_c	Cutoff Value
R_{ij}	Distance Between Node i and j
R_{ij}^0	Distance Between Node i and j at equilibrium
$V_{harmonic}$	Harmonic Potential
$\mathbf{V}_i$	Arbitrary Vector i
$\|\mathbf{V}_1\|$	Length of Arbitrary Vector i
X_i	Position of Node i at Direction X
Y_i	Position of Node i at Direction Y
α	Alpha
$\beta_t(i)$	Beta
γ	Spring Constant
Δ	Difference Operator
$\mathbf{\Delta R}$	Potential Fluctuation Vector
$\Delta\lambda_i$	Eigenvalue Change for Mode i
$\langle\Delta\lambda\rangle$	Average Eigenvalue Change
λ_i	Eigenvalue of Mode i
λ_i^0	Eigenvalue of Reference Mode i

LIST OF ACRONYMS/ABBREVIATIONS

ADP	Adenosine Diphosphate
ATP	Adenosine Triphosphate
AdK	Adenylate Kinase
Arg	Arginine
Asn	Asparagine
BC	Biotin Carboxylase
ENM	Elastic Network Model
Gen	Generation
Glu	Glutamic Acid
Gly	Glycine
His	Histidine
IC_{50}	Half Maximal Inhibitory Concentration
Lig	Ligand
Lys	Lysine
MCG_ENM	Mixed Coarse-Grained Elastic Network Model
MD	Molecular Dynamics
MMTSB	Multiscale Modeling Tools for Structural Biology
OMPDC	Orotidine 5'-Monophosphate Decarboxylase
PDB	Protein Data Bank
Pro	Proline
RMSD	Root Mean Square Distance
RM	Reference Mode
Ser	Serine
TIM	Triosephosphate Isomerase

1. INTRODUCTION

Proteins are key players in vast majority of the biological activity in living organisms and they participate virtually in every cellular process. Enzymes, which are the focus of this thesis, form an important class of proteins that perform highly specific chemical catalysis in the cell. The molecules bound and acted upon by enzymes are called substrates. The part of the enzyme, where substrate molecules bind and chemically react, is called active site. There may also be allosteric site(s) of an enzyme distant from the active site, where effector molecules bind and regulate activity. The influence of such binding is called an allosteric effect.

The structure-function relationship of a protein is determined by its flexibility and conformational dynamics [1]. The sequence and three-dimensional structure of a protein can be accurately determined with experimental methods, however computer simulations are necessary in order to uncover its conformational dynamics. Both atomistic and coarse-grained simulation techniques can be utilized for this purpose. Coarse-grained techniques offer computational efficiency and are thus applicable to large protein complexes. Elastic network models are among the most effective approaches to calculate the collective vibrational modes of a protein, which are known to be related to biological function and conformational changes imposed by ligand binding.

Ligand binding has drastic effects on protein function, such as signal transduction or enzymatic reaction. Binding of a ligand may change the conformation of a protein by stabilizing a specific conformational state. However, even in cases where the conformation of the protein is not affected by ligand binding, the dynamic nature of the protein can still change upon binding. For example, the number of modes in the lowest frequencies and backbone flexibility have been observed to decrease due bound ligands for triosephosphate isomerase and dihydrofolate reductase [2]. The objective of this thesis is to develop and apply efficient ENM-based methodologies for allosteric or orthosteric ligand binding effects on vibrational dynamics of enzymes.

Chapters 2 and 3 of this thesis comprise the literature review and the computational methods, respectively. Results of computational studies performed in this thesis are mostly based on the enzyme TIM, which will be presented in Chapters 4 and 5. In the former chapter, the effect of ligand binding on TIM's collective modes are examined via mixed coarse-grained elastic network models, i.e. MCG_ENM [3,4]. The eigenvalue shifts and changes in mode shapes due to ligand binding are determined by performing MCG_ENM in the absence and presence of the ligand. Due to the computational efficiency of MCG_ENM, independent calculations could be performed on 1496 docking poses and 54,000 MD snapshots of TIM, both in the apo state and in the complex. The aim is to understand the inhibitory effect of certain ligands and their positions on TIM in terms of the changes they impose on TIM's vibrational dynamics.

In this thesis, a new ENM-based methodology is developed, where all residues of a protein are scanned one-by-one by placing extra nodes at side chain heavy atoms, in order to assess each residue's effect on protein's collective dynamics. The ultimate goal is to identify functionally important regions of the protein structure in terms of ligand binding. Applications of this scanning method to TIM and three other proteins are presented in Chapter 5.

2. LITERATURE REVIEW

2.1. Triosephosphate Isomerase

Triosephosphate isomerase (TIM) is a crucial enzyme of the glycolytic pathway. TIM catalyzes the interconversion between isomers dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate, which is an important reaction for the energy efficiency of glycolysis. TIM is found in almost all organisms. Deficiency of TIM could cause a neurological disorder in humans called triose phosphate isomerase deficiency [5].

TIM has a homo-dimeric structure, where each monomer has an independent catalytic site. Even though the enzyme cannot function in monomeric form, no co-operativity or allostery has been observed between the two catalytic sites [6]. TIM molecule used in this study belongs to the parasite Trypanosoma cruzi (an organism causing Chagas disease in humans) and molecule has 251 residues in each of its sub-units.

Each subunit has a (α/β) fold, commonly named the TIM barrel, and active sites are located at C-terminal end of the β-barrel. There are four catalytic residues were detected (Asn-12 and Lys-14 on loop 1, His-96 on loop 4, and Glu-168 on loop 6) on x-ray structure 1TCD [7]. In Figure 2.1 Catalytic loop 6 (Glu-168-Pro178), loop 7 Gly-212-Lys-218), and loop 8 (Gly-235-Lys-240) involved in the active-site geometry contributing via hydrogen (H)-bonding interactions with the substrate [8].

Each subunit has a barrel structure and active sites are located at C-terminal end of the β-barrel. There are four catalytic residues reported (Asn-12 and Lys-14 on loop 1, His-96 on loop 4, and Glu-168 on loop 6) on x-ray structure 1TCD [7]. In Figure 2.1, besides the so-called catalytic loop 6 (Glu-168-Pro178), loop 7 (Gly-212-Lys-218), and loop 8 (Gly-235-Lys-240) are also involved in the active-site geometry contributing via hydrogen (H)-bonding interactions with the substrate [8].

Catalytic loop 6 prevents solvent exposure during catalysis by closing over the substrate and stabilizes the reaction [9]. Dynamics of this loop has been found to be related with the global motion of the protein in previous computational studies, including both coarse-grained elastic network modeling (ENM) and classical MD simulations [3, 10, 11], suggesting that disturbance in global modes could affect catalytic site activity.

Allosteric binding sites may serve as species-specific drug targets and are especially important for curing parasitic diseases [12] since catalytic sites mostly conserved among species allosteric site may vary. In a previous study [13], several benzothiazole derivatives has been reported as candidate inhibitors for TcTIM. It was also indicated that the tunnel region at the interface is the allosteric region regulating the catalytic activity and the target for these inhibitors.

Five of these ligands (see figure 2.2) were chosen in order to be used in the docking study [14] from which the docking results used in chapter 4.1 was taken. Ligand 8, 9, and 10 are inhibitors and ligand 2 and 3 are non-inhibitor included in the study as a control group. In Figure 2.2 the ligands are shown and rotatable bonds are marked with a star. Also in Table 2.1. experimentally [13] determined inactivation and IC_{50} values for these ligands are given.

Table 2.1. Effectiveness of benzothiazoles used in dockings.

Ligand	**Inactivation (%)**	**IC50 for TcTIM (μM)**
2	0	0
3	0	0
8	95	33
9	91	56
10	95	8

The aim of the study [14] was to uncover the binding modes of benzothiazoles via performing blind dockings on both TcTIM and human TIM. The inhibitor found

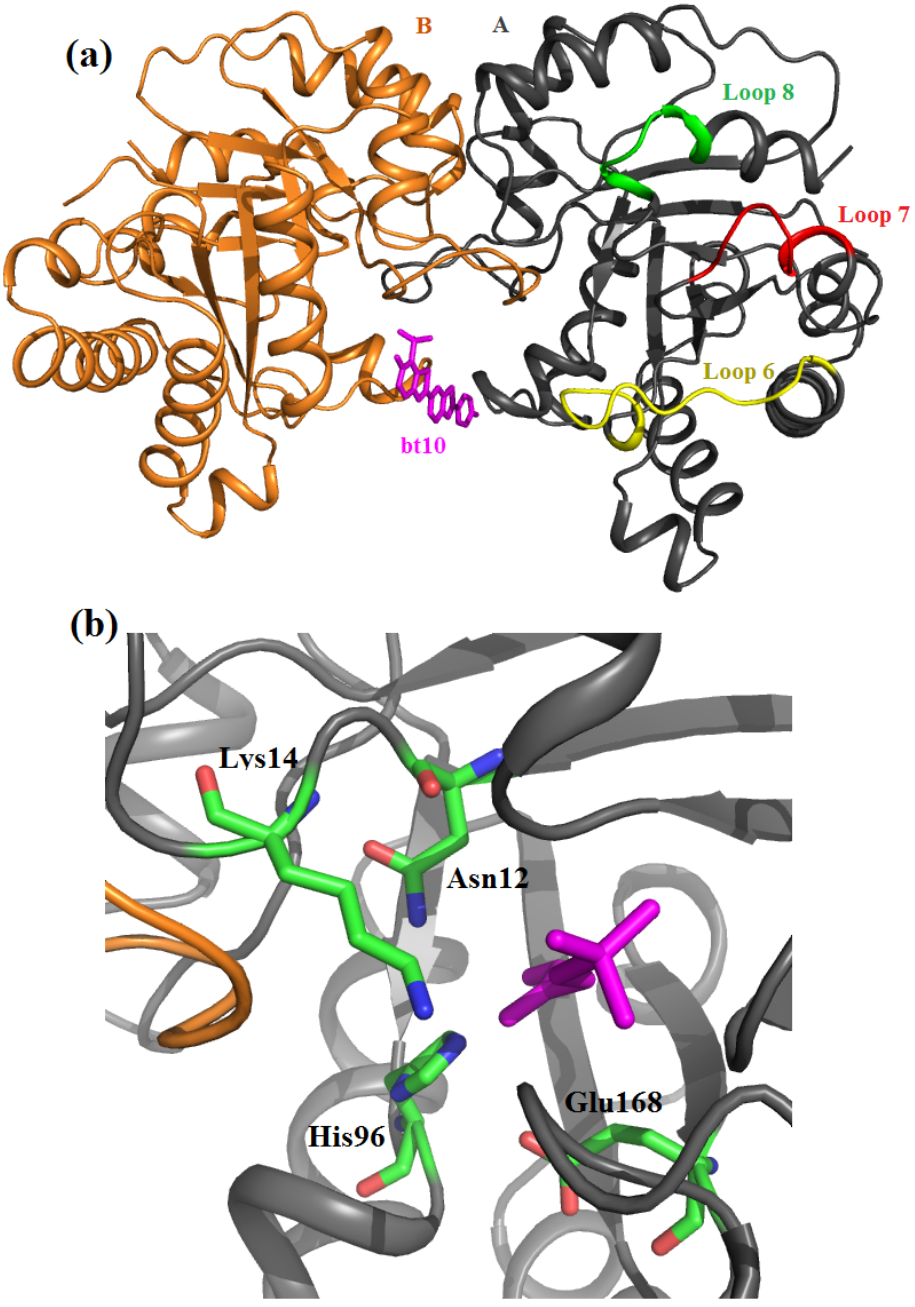

Figure 2.1. Structure of TIM; (a) dimeric structure showing important loops and inhibitor bt10, (b) catalytic residues and substrate analog.

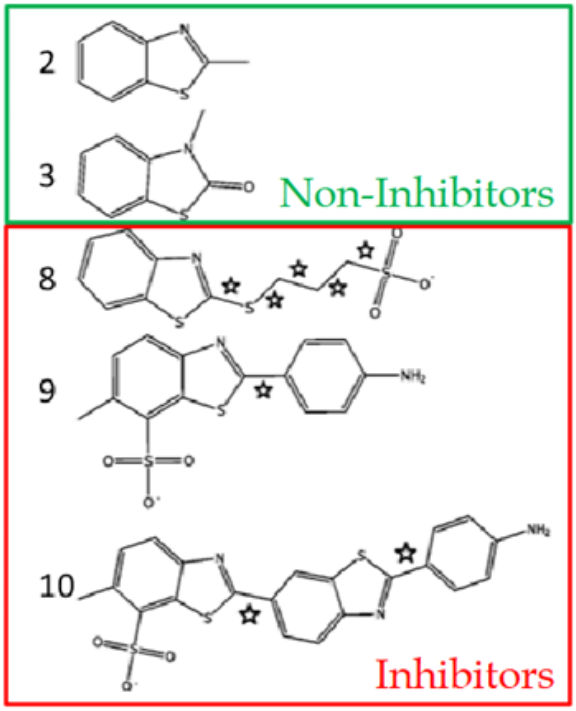

Figure 2.2. Ligands chosen for the docking study.

to be binding to the tunnel region for TcTIM but not for human TIM.

2.2. Effect of Ligand Binding On Globular Vibrational Modes

Motions of a protein occurring around a global energy minimum state is called the equilibrium dynamics, which can be efficiently obtained computationally by normal mode analysis of the equilibrium structure. Low frequency modes obtained through normal mode analysis corresponds to the global motions of a structure around it equilibrium state. The equilibrium dynamics got popularized in recent years by several studies [14–22], both experimental and computational, and these studies shows the significance of the equilibrium dynamics of proteins to achieve their catalytic function or allosteric regulation.

Allostery is the regulation of a protein's function by binding of an effector molecule to a site other than the active or catalytic site of a protein. Allosteric regulation could

take effect through various mechanisms, such as conformational change. However the scope of this work focuses on the regulation through the effect of the allosteric ligand on the global low-frequency motions.

In 2013 a study [23] , a foundational theory for how allosteric effectors can affect the function of structure via altering low-frequency dynamics without changing the conformation of the protein. The study shows that an allosteric control can occur without a change in the shape of a structure and with the effect of the ligand on globular fluctuation using experimental and computational methodologies.

In 2013 a study [23], a foundational theory for how allosteric effectors can affect the function of structure via altering low-frequency dynamics without changing the conformation of the protein. The study shows that an allosteric control can occur without a change in the shape of a structure and with the effect of the ligand on globular fluctuation using experimental and computational methodologies.

Another study [24] takes a more theoretical approach to the subject and examines the phenomenon of allosteric regulation without structural alteration. In the study multi-scale biomolecular models are used to show that modulation of the amplitude of thermal fluctuations around a mean structure, rather than a change in the conformation, is an effective mechanism in allostery.

In numerous studies it has been shown that, the experimentally observed conformational change correlates with at least one of the lowest frequency normal modes upon ligand binding [25–28]. Depending on these studies applying elastic network models (ENM), in describing protein conformational dynamics and transitions is encouraged.

In a 2010 study [2], it is aimed to see how global dynamics and flexibility of proteins change in different ligation states. MD simulations are performed for TIM and dihydrofolate reductase in bound and unbound state. Through time series analysis it is observed that the frequencies of motion shift to left when ligand is added to the simulation.

3. METHODS

3.1. Elastic Network Model (ENM)

Elastic network model [29,30] describes a three dimensional elastic network, where the protein is described as a set of nodes interconnected by Hookean springs. Generally, residue-based coarse-graining is employed by taking alpha-carbon atoms as nodes (one node for every residue) and close-neighboring node pairs in the structure are connected by harmonic springs. Springs are considered massless, whereas nodes have unit mass.

The formulation of potential energy for this model is given in the Equation 3.1, which is a summation over all harmonic springs.

$$V_{harmonic} = \frac{\gamma}{2} \sum_{i} \sum_{j} h(R_c - R_{ij})(R_{ij} - R_{ij}^0) \tag{3.1}$$

Here, $V_{harmonic}$ means total harmonic potential in the system. R_{ij} is the variable denoting to the distance between nodes i and j. R_{ij}^0 is the constant denoting to the distance between node i and j at equilibrium, which is taken as the distance in the initial folded structure. R_c is a cutoff distance and $h(R_c - R_{ij})$ is the heavy-side step function discarding interaction between residues that are faraway than R_c. γ is the force constant of the springs, which is generally taken as unity in standard ENM models.

$V_{harmonic}$ can also be represented as in the following form:

$$V_{harmonic} = \frac{1}{2} \Delta \mathbf{R}^{\mathbf{T}} \mathbf{H} \Delta \mathbf{R} \tag{3.2}$$

where $\Delta \mathbf{R}$ is the potential fluctuation vectors for 1 node which makes it a $3N$ by 1 matrix since every node has three spatial direction for its fluctuation vector. $\mathbf{H}$ is the Hessian matrix, which is the second derivate matrix of the overall potential function Equation 3.1. $\mathbf{H}$ is composed of super elements represented in the Equation 3.3, where

the off-diagonal super elements have the form in Equation 3.4.

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \cdots & \mathbf{H}_{1,N} \\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} & \cdots & \mathbf{H}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{N,1} & \mathbf{H}_{N,2} & \cdots & \mathbf{H}_{N,N} \end{bmatrix} \tag{3.3}$$

$$\mathbf{H}_{i,j} = \begin{bmatrix} \frac{\partial^2 V}{\partial X_i \partial X_j} & \frac{\partial^2 V}{\partial X_i \partial Y_j} & \frac{\partial^2 V}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Y_i \partial X_j} & \frac{\partial^2 V}{\partial Y_i \partial Y_j} & \frac{\partial^2 V}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Z_i \partial X_j} & \frac{\partial^2 V}{\partial Z_i \partial Y_j} & \frac{\partial^2 V}{\partial Z_i \partial Z_j} \end{bmatrix} \tag{3.4}$$

Off-diagonal super elements are explicitly given in Equations 3.5 and 3.6.

$$\frac{\partial^2 V}{\partial X_i \partial X_j} = -\gamma \frac{(X_i - X_j)^2}{R_{ij}} \tag{3.5}$$

$$\frac{\partial^2 V}{\partial X_i \partial Y_j} = -\gamma \frac{(X_i - X_j)(Y_i - Y_j)}{R_{ij}} \tag{3.6}$$

Diagonal super elements are given in the following two equations.

$$\frac{\partial^2 V}{\partial X_i^2} = \gamma \sum_j \frac{(X_i - X_j)^2}{R_{ij}} \tag{3.7}$$

$$\frac{\partial^2 V}{\partial X_i \partial Y_i} = \gamma \sum_j \frac{(X_i - X_j)(Y_i - Y_j)}{R_{ij}} \tag{3.8}$$

Eigenvalues and corresponding eigenvectors of the Hessian matrix give normal modes of the system. However, first six modes correspond to three translation and three rotation motion and does not contain any useful knowledge therefore these modes are discarded

at the beginning. Eigenvalues are proportional to the square of the frequency of the corresponding mode of motion and eigenvectors give the directions of node/residue fluctuations. In this thesis, eigenvalues are used in analyses instead of frequency values for simplicity. It was well established in the literature that lowest modes of fluctuations correspond to the global motions of the macromolecular system, which are related to the biological functional of that system.

Overall, this method is an appropriate simplification of a complex molecular system. By analyzing slowest modes obtained from ENM one can understand the global dynamics of a macromolecular system. Real distinguishing advantage of using this coarse-grained method compared to other methods available for studying dynamics of a macromolecular system is its simplicity and low computational cost. This computational efficiency makes ENM methodology most suitable for its application to more than thousands of structures, either docking poses or MD snapshots, as will be presented in this thesis.

3.2. Mixed Coarse-Grained ENM

Mixed coarse-grained ENM [3, 4] follows the similar methodology of classical ENM, but with the addition of some atomistic or high resolution nodes for certain residues or ligands in the structure. In classical ENM, all residues represented in the model according to the one-node-per-residue basis, whereas in the mixed coarse-grained model both low resolution and high resolution regions exist. While low resolution region residues still represented as one-node-per-residue in the model, the high resolution region residues represented as many-nodes-per-residue and corresponding mass and spring constant values differ accordingly.

In this work, this approach is found especially suitable for investigating collective dynamics of protein-ligand complexes. For this purpose, the coordinates of ligand's all heavy atoms are taken into account and the protein is modeled at low-resolution, i.e. coarse-grained at the residue level, by placing nodes at residue centroids. In the high-resolution region, each heavy atom has unit mass and pairs of atoms that fall

within a specified cutoff radius (7 Å) are connected by identical harmonic springs with force constant of unity. In the low-resolution region, the force constant between any two coarse-grained nodes equals to the total number of heavy atom pairs (from two different residues) that fall within the same atomistic cutoff. And the mass of each coarse-grained node equals to the number of heavy atoms it has.

This method is computationally efficient and useful when consideration of connections between heavy atoms of the atomistic structure is required. Therefore, mixed course-grained model are used rather than standard ENM in Chapter 4 where analyses of docking results are made. In chapter 4, ligands bound the protein structures are taken as high resolution region where protein structure itself taken as a low resolution region with a cutoff value of 7 Å. For each MD snapshot in Chapter 4.2 (54,000 in total), MCG_ENM was performed to extract the slowest eigenvectors and eigenvalues, which represent global mode shapes and their squared frequencies, respectively. As a result, the effect of ligand binding on mode shapes and frequencies can be determined by comparing apo and complex runs.

3.3. Frequency Shift Scanning

Detecting ligand binding sites, whether allosteric or orthosteric, is an important goal in protein studies. As will be presented in Chapter 4, binding of a ligand to a functionally important site could constrain the global motions of an enzyme and thereby affect its functionality. This constraining effect could be observed through a normal mode analysis, which indicates positive shifts in the frequencies of collective modes.

Here we introduce a new method based on standard (residue/C_α-based) ENM, where the side-chain heavy atoms of a specific residue are included as extra nodes in order to mimic and pronounce the presence of a ligand that may interact with that specific site/residue. Residues of the protein are scanned one-by-one by taking each residue's heavy atoms as extra nodes. A residue-based cutoff distance of 15 Å is used between all pairs of nodes, whether backbone C_α of each residue or the specific residue's

side chain atoms. Thus the effect of side chains is overemphasized here.

This analysis mostly carried out via examining the change in the frequencies of global modes, which are slowest modes. Since the measure of frequency directly related to the eigenvalues of the hessian matrix, it is possible observe the changes through eigenvalues directly and this change is called eigenvalue shift.

3.4. Analysis Of Changes In Collective Modes

General equation for calculating the eigenvalue shift of mode i due to ligand binding or the addition of side chain atoms as extra nodes is given below.

$$\Delta\lambda_i = \frac{\lambda_i - \lambda_i^0}{\lambda_i^0} \times 100 \tag{3.9}$$

Here $\Delta\lambda_i$ denotes eigenvalue shift of mode i as percentage. λ_i^0 is the original or unperturbed eigenvalue of mode i, when only C_α atoms (coarse-grained residues) are included in the model without ligand or side chains. And λ_i is the eigenvalue of mode i obtained either in the presence of the bound ligand (MCG_ENM) or due to the addition of side chain heavy atoms as extra nodes in residue scanning.

In some analyses, average of eigenvalue shift of a few slowest modes for every residue was used in order to drive general conclusions about the importance of that residue. The way of calculating the averages are given below.

$$\langle\Delta\lambda\rangle = \frac{1}{N}\sum_{i}^{N}\Delta\lambda_i \tag{3.10}$$

$\langle\Delta\lambda\rangle$ is the average value obtained for a residue or docking pose. N is the number of modes chosen to be averaged over. Number N can be changed for different cases according to the how the structure reacted to the method (see chapter 4 or 5 for detail).

Throughout the study, overlap values are calculated as the absolute value of the

dot product of the normalized fluctuation vectors of the backbone atoms (see Equation 3.11) which give a value between 0 and 1 also corresponds to the absolute value of the cosine of the angle between the vectors. 0 means the vectors are orthogonal 1 means the vectors are parallel thus this parameter is used in order to compare characters of two modes of motion.

$$O = \left| \frac{\mathbf{V}_1.\mathbf{V}_2}{\|\mathbf{V}_1\|\|\mathbf{V}_2\|} \right| \tag{3.11}$$

O denotes overlap value. $\mathbf{V}_1$ is one of the vectors and $\|\mathbf{V}_1\|$ is the length of that vector. This parameter was used for crating overlap matrices in Chapter 4 and for evaluating the effect of a residue on the character of a global motion in Chapter 5.

4. LIGAND EFFECT ON TRIOSEPHOSPHATE DYNAMICS

TIM plays an important role in glycolysis catalyzing the reversible interconversion of the triose phosphate isomers dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate. That makes TIM vital for forms of life which perform glycolysis. TIM is a dimer with two identical subunits, composed of eight central β-strands surrounded by eight α-helices. TIM molecule used in this study belongs to the parasite Trypanosoma cruzi and molecule has 251 residues in each of its subunits and sometimes referred as TcTIM. So called tunnel region at the interface is a known allosteric region for TIM molecule. Three (ligand 8, 9, and 10) of the five different benzathiazoles used in the first section of this chapter are known inhibitors and other two are used as control group [13, 14] (see Figure 2.2). In the later section, only ligand 10, which also called bt10 and most effective inhibitor among five benzathiazoles, is used.

4.1. An ENM Analysis of Benzathiazole Dockings on TIM

In a previous study [14] (see Chapter 2 for details), five different benzathiazoles were docked on three different conformers of TcTIM, i.e. triosephosphate isomerase from the parasite Trypanosoma cruzi. This docking study was intended to be a blind docking study, without any specified region targeted on the structure instead an entire domain was targeted. One hundred docking poses were obtained for each ligand on a specific protein conformation. As a result, a total of 1496 docking poses were analyzed here that include all combinations of five ligands and three distinct TIM conformers.

In the blind docking study [13, 14], four distinct binding regions were favored by the ligands, which are shown in Figure 4.1 from three different points of view. The red region corresponds to the tunnel at the interface of the identical monomers. The blue region comprises catalytic residues Lys14, His96, and Glu168. A and B regions represent other preferred sites, for which no functional significance has been attributed

so far. In Figure 4.2, the relationship between ligand type and binding region preference is presented by the number of poses, in which each ligand binds to specific sites. It can be seen that, ligands 2, 3 and 8 do not have a strong preference for any region. On the other hand, ligands 9 and 10 strongly favor the tunnel region. Furthermore, the percentage of binding to the tunnel region is correlated with the ligand size.

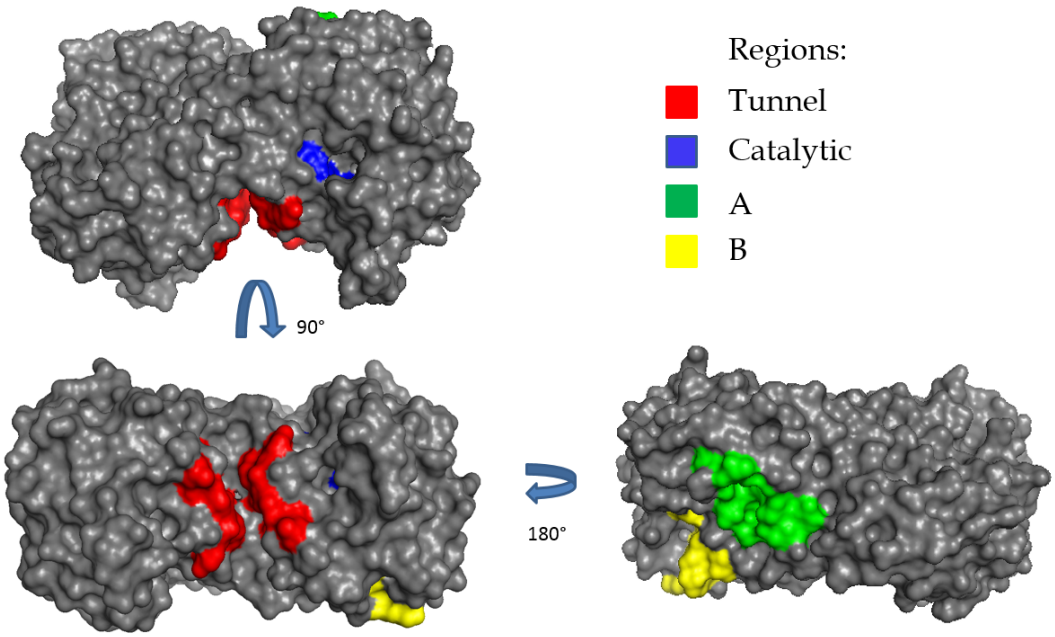

Figure 4.1. Loop structures of TIM.

In this thesis, these blind docking results on TIM were analyzed by mixed coarse-grained ENM in order to understand the effects of binding of inhibitors and non-inhibitors on TIM's collective dynamics. Specifically, we investigated the extent to which the binding of ligands to specific regions affects the global modes of motion.

In order to study the changes in protein functional dynamics due to ligand binding, global modes (i.e. lowest-frequency modes) from ENM were calculated with and without the ligand. Percentage change in eigenvalues were calculated in comparison to the eigenvalues of the initial conformer without the docked ligand. Thus, the shift in eigenvalue (or squared frequency) of the kth collective mode was calculated due ligand binding according to Equation 3.9.

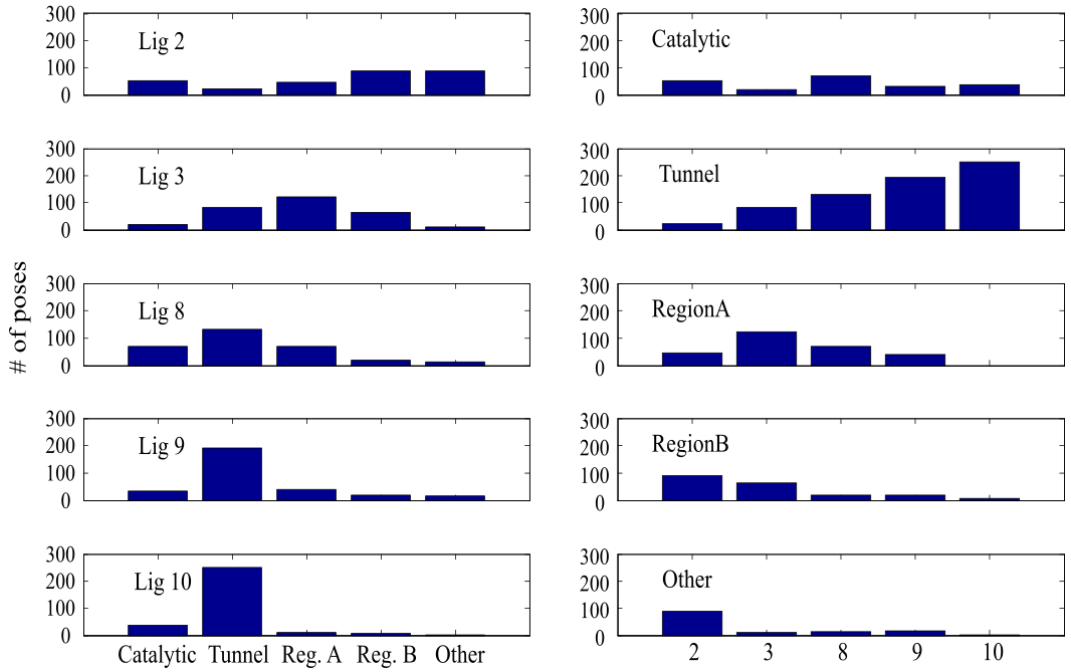

Figure 4.2. The number of docking poses with the ligand bound to a specified region.

Figure 4.3 gives the eigenvalue shift in first 30 modes of motion for specific binding regions. The percentage shift is reported as an average over the poses where any ligand is bound to that specific region. The poses with the ligand bound to the tunnel region indicate pronounced shifts especially in the first two modes, which implies significant restriction on global motions of the structure. Higher modes of the tunnel region and the catalytic region also exhibit shifts to a certain extent. In contrast no effect of binding can be detected for other regions including A and B. The distributions (Box plots) are shown in Figure 4.4 for the tunnel and catalytic regions. In box plots red line in the box represents median, bottom and top lines of the box are the first and third quartiles of the data and the whiskers represent the lowest and highest data points within 1.5 interquartile range (length of the box) lower and higher than lower and upper quartile (edges of the box). Rest of data are showed as red crosses as outliers.

When we group the small ligands (2 and 3) and the large ligands (8, 9 and, 10) in two different distributions (Figure 4.4), the small ligands at the tunnel result in some eigenvalue shift in the first two modes but not as much as that caused by the

larger ligands. On the other hand, separating the distributions according to different conformers of TIM in Figure 4.5 does not indicate much difference between D1, D2 and D3.

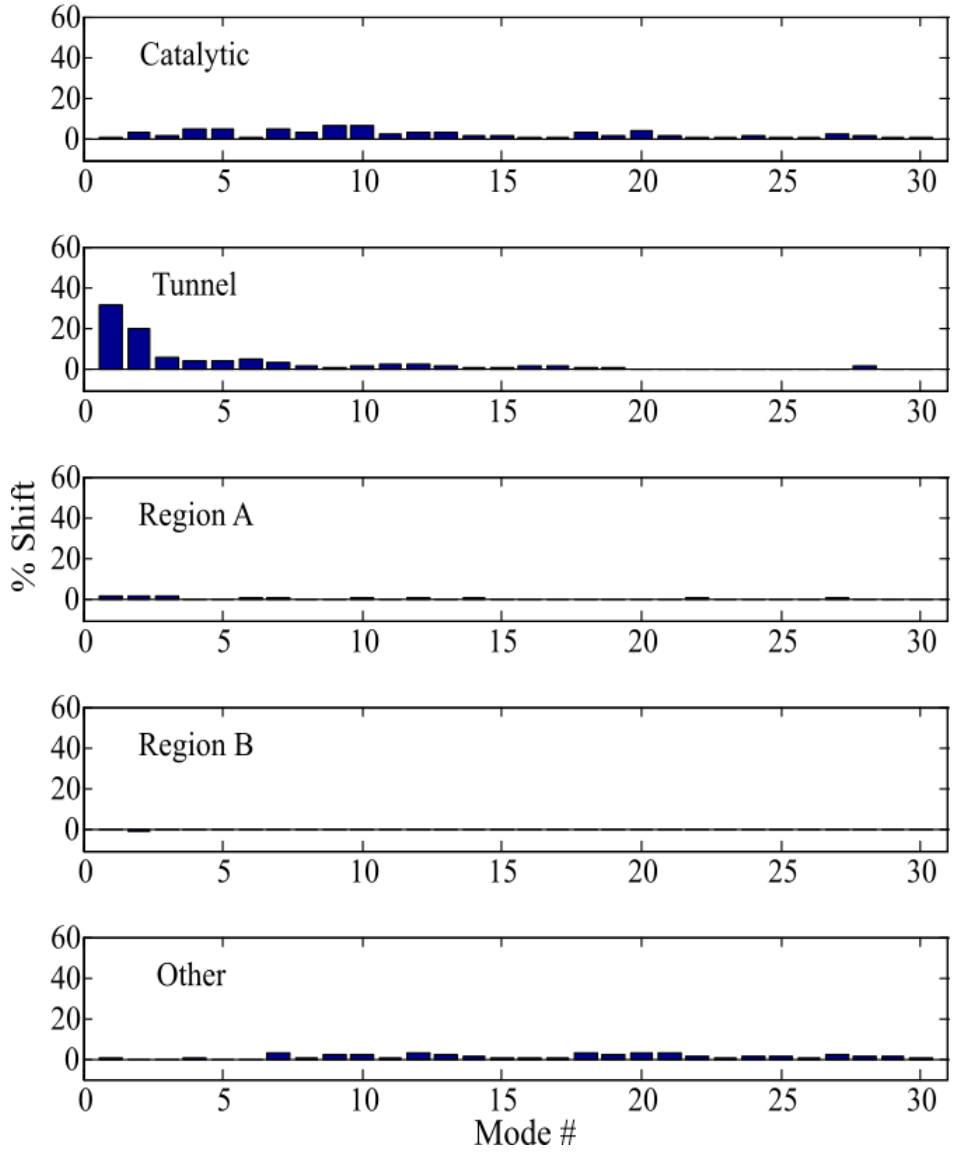

Figure 4.3. Average percent eigenvalue shift in first 30 modes.

Figure 4.6 shows the scatter plot of eigenvalue shift vs. binding energy (or score) obtained from AUTODOCK [31, 32] calculations for all docking poses. As expected, only the first three modes show significant shifts. Average eigenvalue shift of the first three modes successfully differentiates the tunnel from other regions in Figure 4.7. The

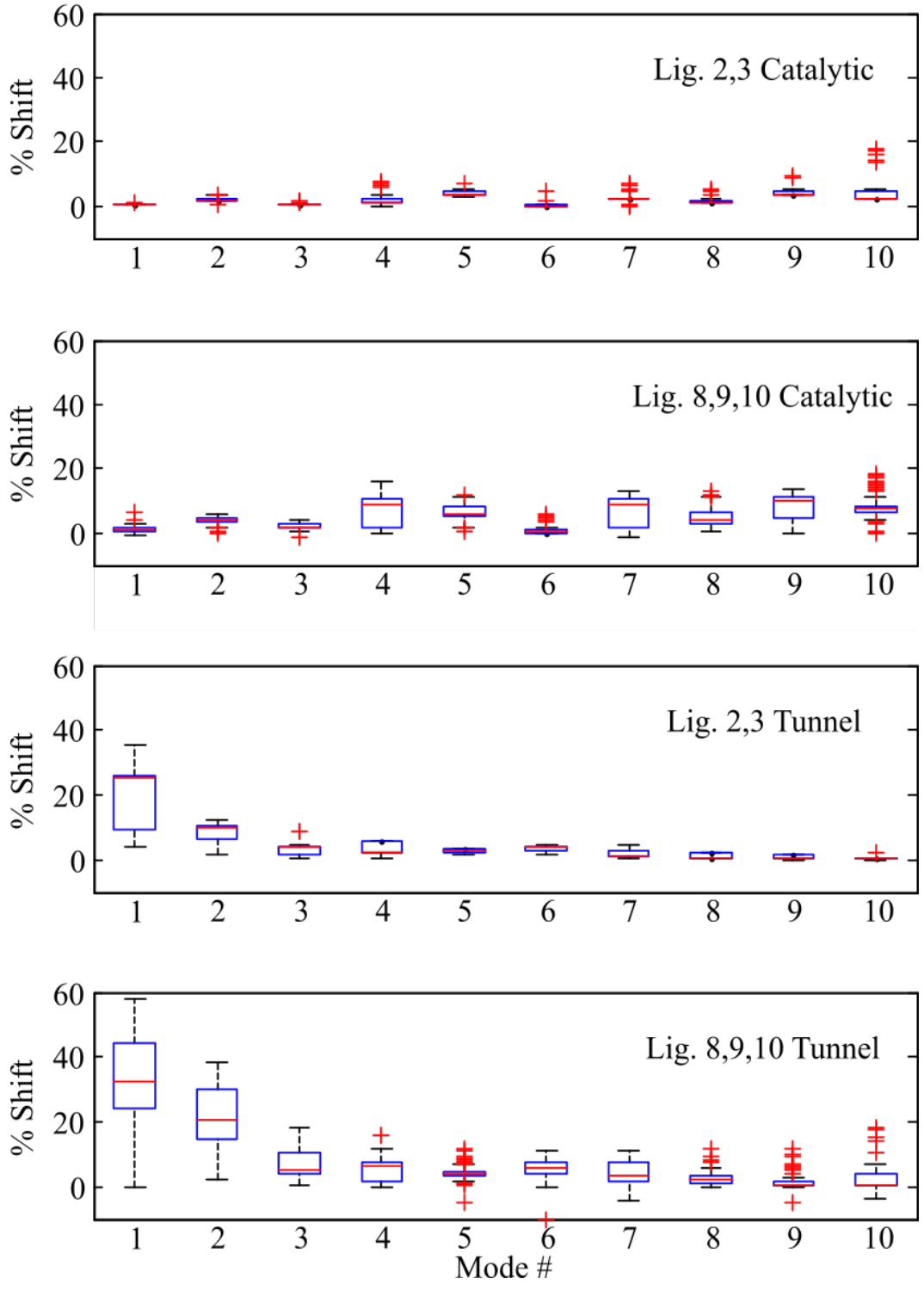

Figure 4.4. Distributions of eigenvalue shift in the first 10 modes for small and large ligands.

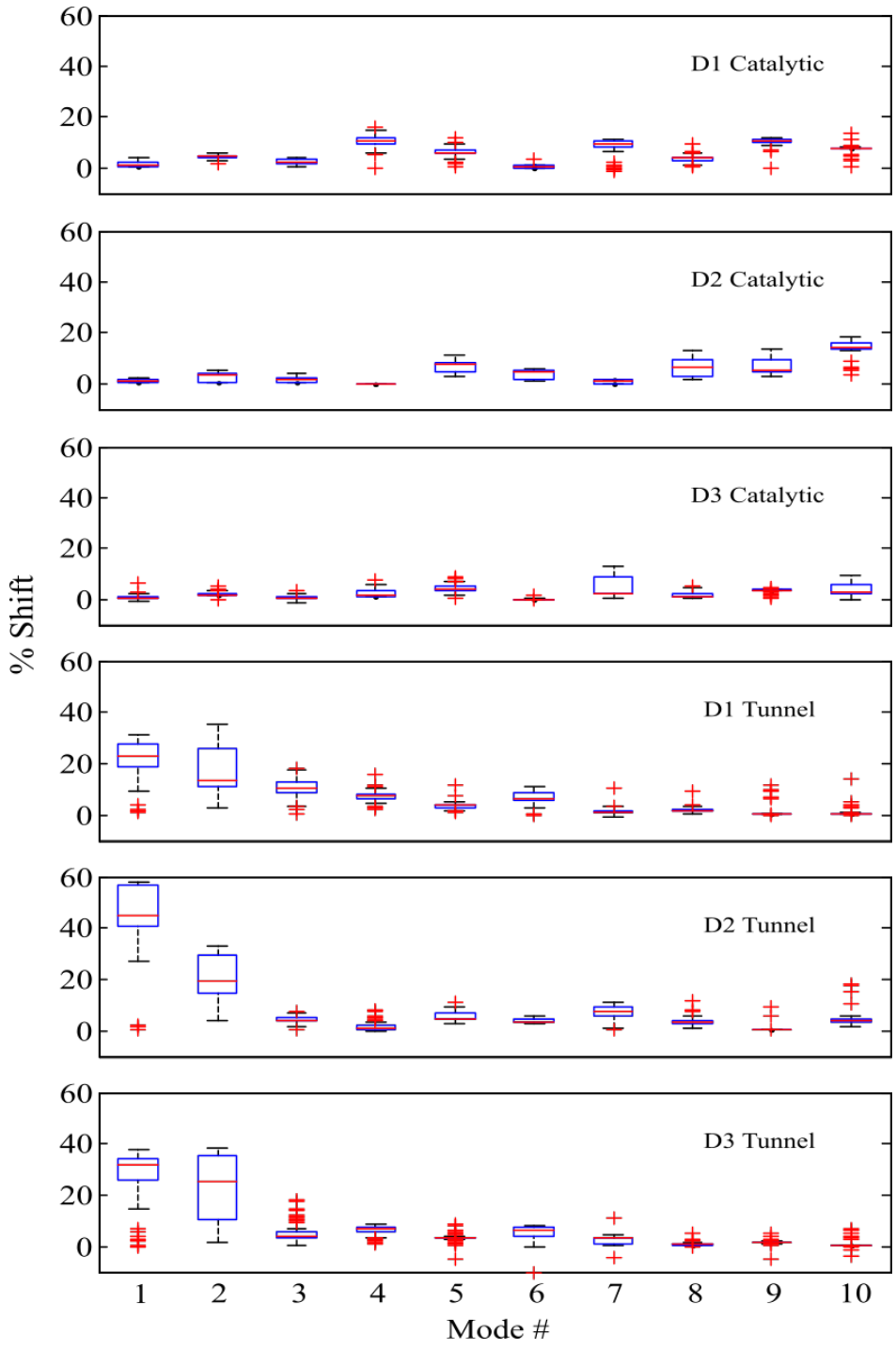

Figure 4.5. Distributions of eigenvalue shift in the first 10 modes for different conformers (D1, D2 and D3).

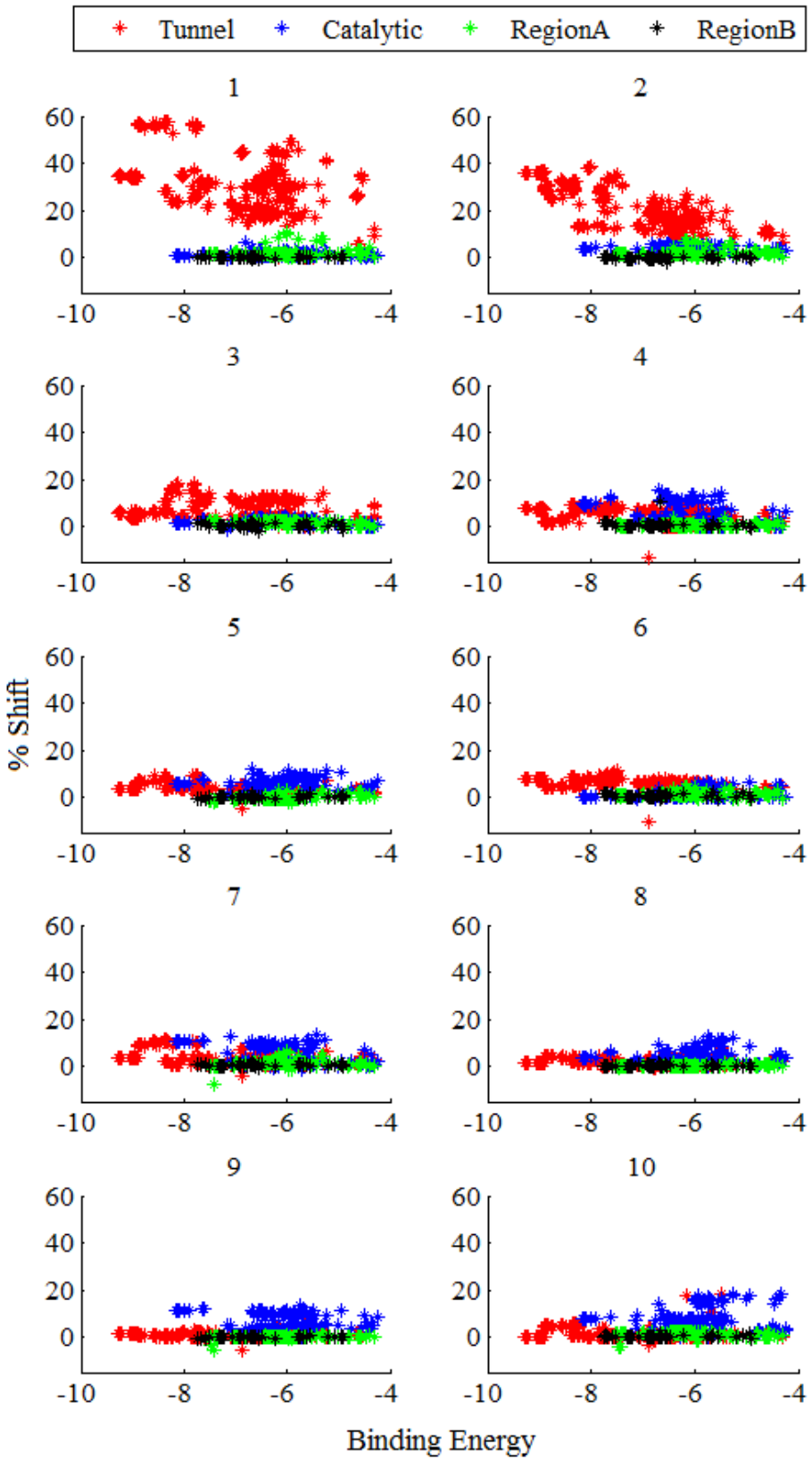

Figure 4.6. Scatter of eigenvalue shift vs. binding energy for different modes.

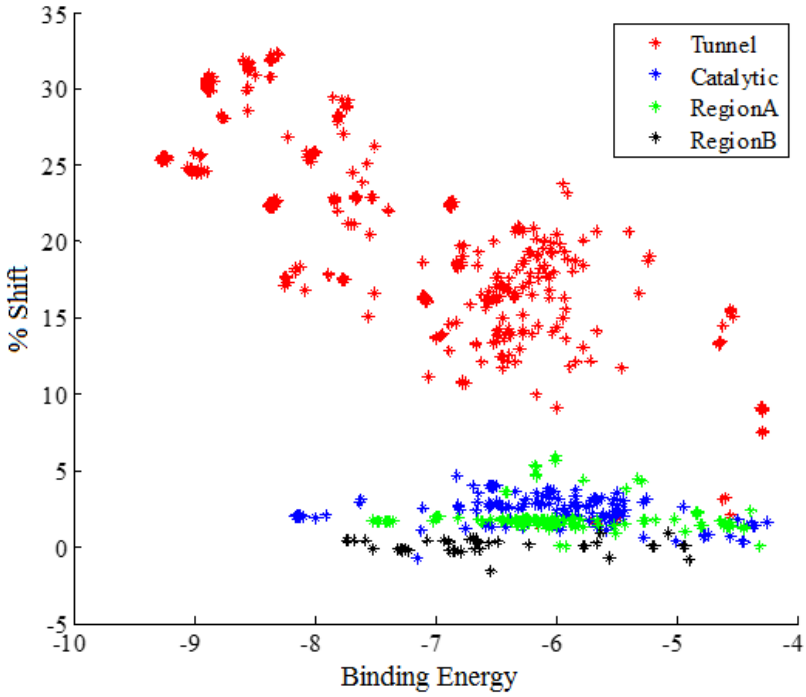

Figure 4.7. Scatter of eigenvalue shift vs. binding energy for average of first three modes.

eigenvalue shift is not only higher for tunnel region but also correlates with binding energy. This implies that stronger binding to tunnel will considerably affect the overall dynamics of the molecule, which makes this region a prime target for allosteric drug designs.

When each ligand is considered separately, the inhibitors can also be distinguished (Figure 4.8). The profiles of non-inhibitors (2 and 3) are different from those of inhibitors (8, 9 and 10). For effective ligands, the eigenvalue shifts can distinguish the tunnel even though the scores (or ΔG values) of some poses at the tunnel are quite low. For only the tunnel region, the eigenvalue shift progressively increases with ligand effectiveness and the profiles evolves into a binding energy and eigenvalue shift correlation. Moreover, profiles for other regions seems inseparable.

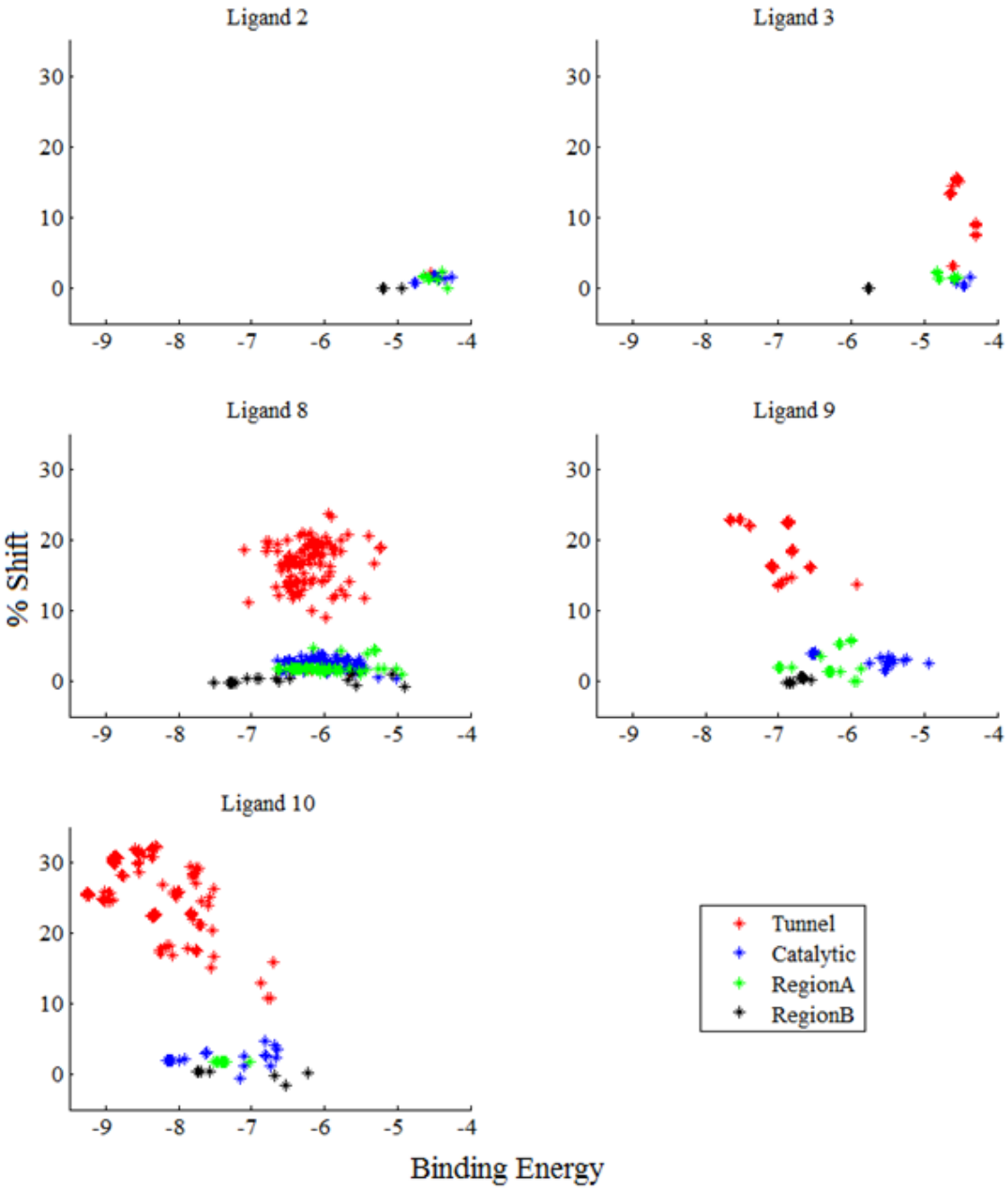

Figure 4.8. Scatter of average eigenvalue shift of first three modes vs. binding energy for each ligand separately.

4.2. An Analysis Of TIM Complex Dynamics

Ligand 10, which is also named as bt10, is the most effective inhibitor among all five ligands considered. ENM analysis of the docking results in the previous section also agrees with these experimental findings on the effectiveness of these ligands (see Table 2.1). In this chapter, six independent 100 ns MD simulation runs [33] for TIM molecule in apo form (3 runs) and TIM-bt10 complex (3 runs) were examined using the same methodology, i.e. MCG_ENM. Apo runs are called 'Apo 1', 'Apo 2', and 'Apo 3'; complex runs are called 'Complex 1', 'Complex 2', and 'Complex 3' The bt10 stays bound to the tunnel region in all runs of the complex.

These MD runs [33] of apo TIM and TIM-bt10 complex have been previously carried out using AMBER9 [34] with ff03 force field parameters [35]. Antechamber module of AMBER was used in order to determine the force field parameters of ligand 10. The runs were 100 ns long and first apo run were started from the structure with Protein Data Bank [36] id 1TCD [7] and other runs started from later snapshots.

MCG_ENM was performed to extract the slowest eigenvectors and eigenvalues for each MD snapshot. A total of 9,000 snapshots were considered for each run excluding the equilibration period. This sums up to 54,000 snapshots, i.e. independent ENM calculations, including the six independent runs. Slowest mode eigenvectors and eigenvalues represent direction of global motion and their squared frequencies, respectively. As a result, the effect of ligand binding on mode shapes and frequencies can be determined by comparing apo and complex runs.

For all snapshots of apo or complex runs, the eigenvalue distribution of first nine modes are shown in Figure 4.9. When the ligand is bound to the interface, the first three slowest modes' eigenvalues (squared frequencies) indicate significant shifts to the right. Thus, the inhibitor seems to act as a constraint on global modes of the protein, which agrees with the blind docking results in the previous chapter.

In order to analyze if there are also changes in the shape/character of modes or

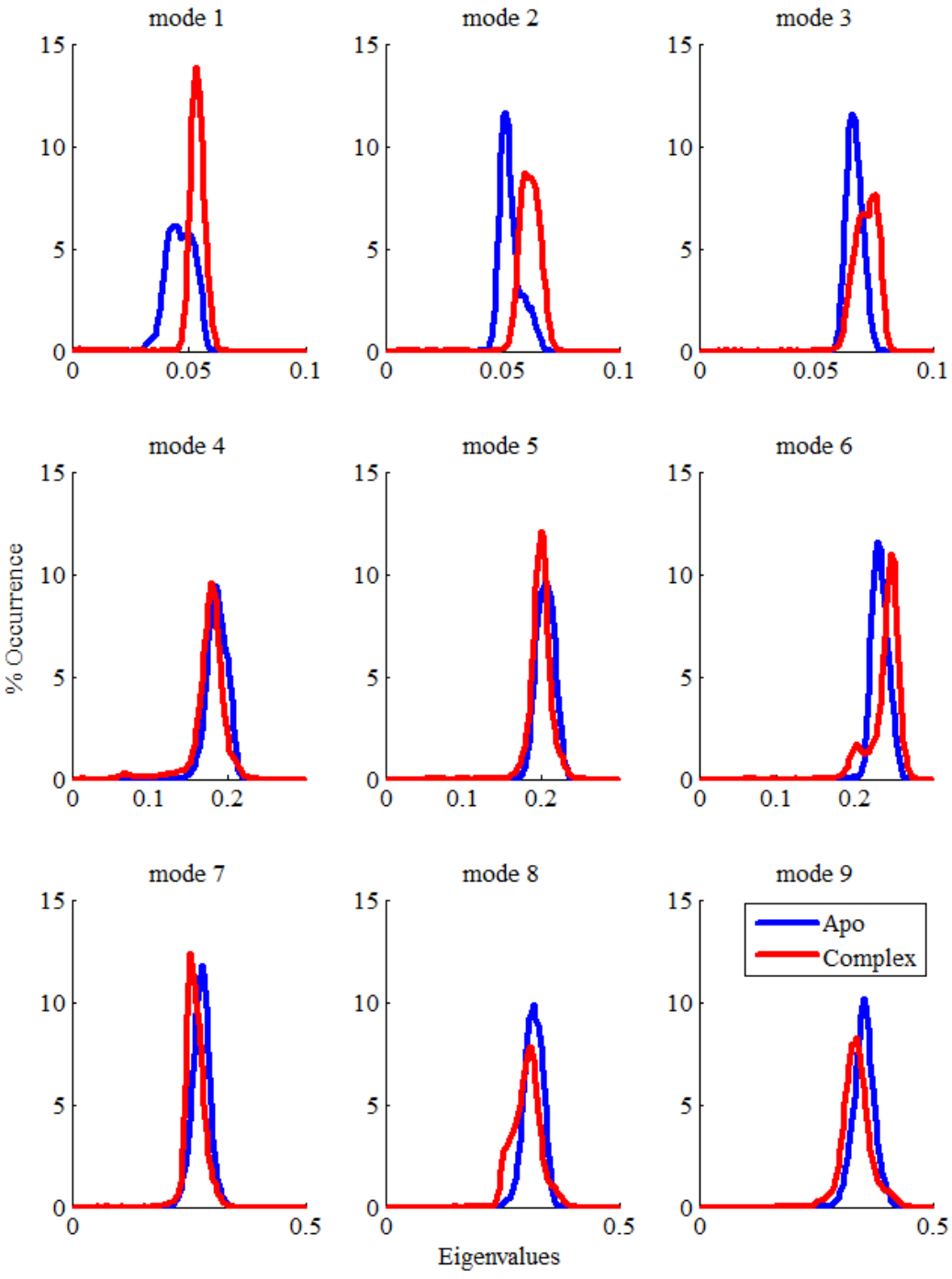

Figure 4.9. Eigenvalue distributions of apo and complex snapshots for the first nine modes.

swapping among the modes due to ligand binding, the lowest-frequency eigenvectors for each snapshot were analyzed with respect to some fixed/reference modes based on a specific structure/snapshot. For this aim, the ENM modes of Apo1 run (10th ns snapshot after the equilibration period) were taken as reference modes (RM), which are shown in Figure 4.10. Here, RM1 represents the bending of the subunits, which is denoted as the front view. RM2 is the counter-rotation (twisting) of the subunits, which is coupled to catalytic loop opening/closure (side view shown). RM3, another bending type motion, is presented from the top after rotating the dimer 90 with respect to the front view. In fact, the ENM modes of the TcTIM crystal structure (PDB id: 1TCD) are very similar to the RMs, as shown in Figure 4.11, which indicate that the 10th ns snapshot of Apo1 run is an appropriate reference.

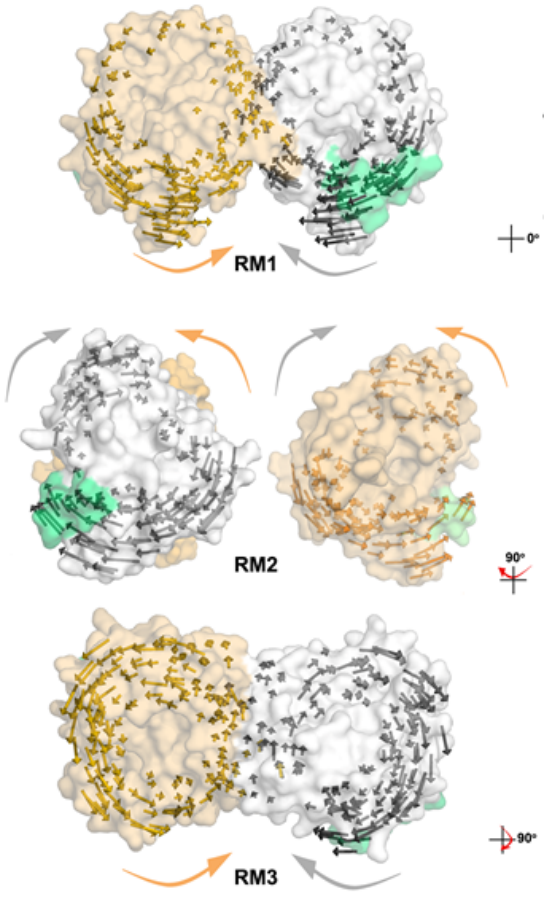

Figure 4.10. Vector representations of reference modes (RM) that are calculated from Apo1 (snapshot at 10th ns).

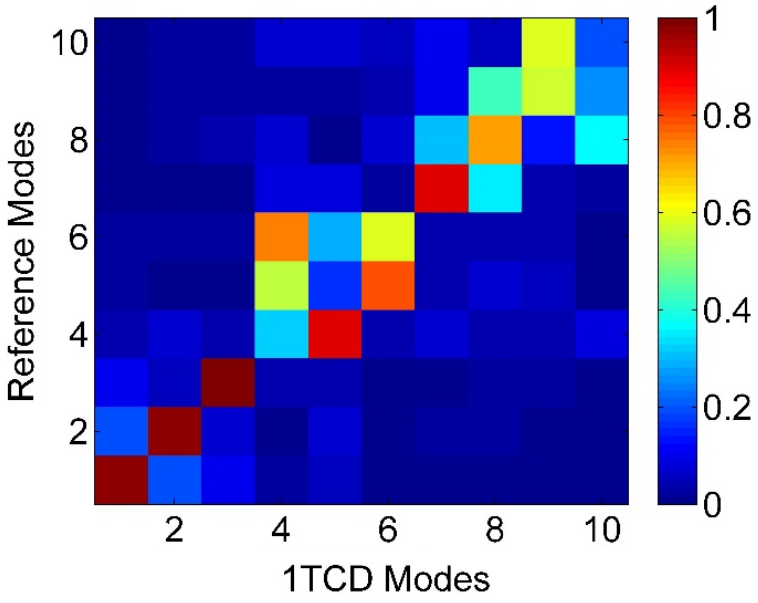

Figure 4.11. Mean overlap matrices between MCG_ENM modes of each snapshot and the reference modes.

In Figure 4.12, the correspondence between the eigenvectors of the first 10 modes of the MD snapshots with the RMs taken from initial structure of Apo1 run are shown. Each overlap matrix is an average over 9000 frames of a specific run. The first four RMs are almost always preserved in apo simulations (average overlap values>0.95). In contrast, overlap values with RMs decrease significantly in the complex and there is a scattered behavior rather than a one-to-one correspondence.

In the "% of mode shape conservation" column of Table 4.1, the percentage of MD snapshots that present at least one ENM mode (among its first 10 modes) that highly correlates (overlap>0.7) with a specific RM is reported for apo and complex runs. Through this calculation, changes in mode shapes can be detected because mode swapping is not considered. In the "% of mode index conservation" column of the same table, the percentage of MD snapshots that preserve both the mode shape (overlap>0.7) and the mode index are reported with reference to each RM. The extent of mode swapping can be inferred from these numbers.

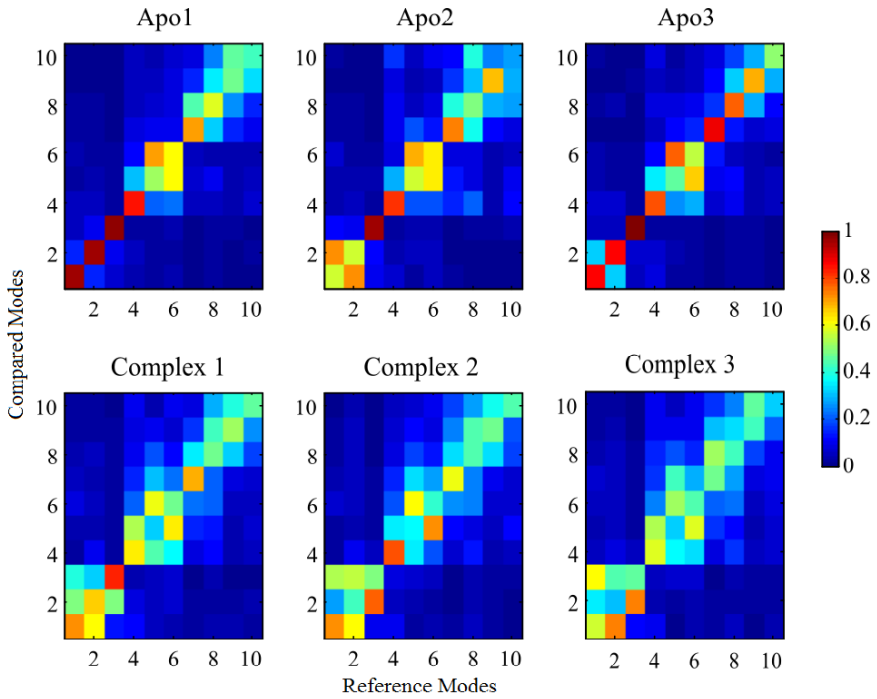

Figure 4.12. Clustering of MD snapshots.

When Table 4.1 is examined, the complex runs indicate more alterations in mode shapes and/or swapping among the slowest modes. In contrast to the preservation of first four modes in apo simulations, RM1, RM2 and RM4 cannot be observed in 25-30% of the snapshots in the complex. RM3 on the other hand is highly conserved even in the presence of inhibitor.

The aligned MD snapshots were clustered according to their pairwise RMSDs, which results in nine clusters shown in Figure 4.13. For this purpose, 5400 snapshots at 100ps intervals, including both apo and complex runs, were considered in k-means clustering algorithm implemented in kclust module of MMTSB Ensemble Tools [37]. Here, the apo runs span a broader conformational space than the complex runs, which is consistent with the constraining effect due to the ligand bound at the tunnel.

Detailed analysis of these MD trajectories (not within the scope of this thesis) indicates changes in dynamics of the catalytic site and loop. Allosteric effect of bt10 bound on tunnel region of TcTIM and its effect on motions of the loop region are

Table 4.1. Conservation of RMs in MD snapshots using MCG_ENM.

	% of mode shape conservation		% of mode index conservation	
RM Index	**Apo**	**Complex**	**Apo**	**Complex**
1	98	74	70	47
2	99	71	70	21
3	100	99	100	50
4	94	73	85	54
5	87	69	11	0
6	64	53	16	3
7	91	67	78	46
8	58	33	45	17
9	67	41	52	24
10	13	11	11	10

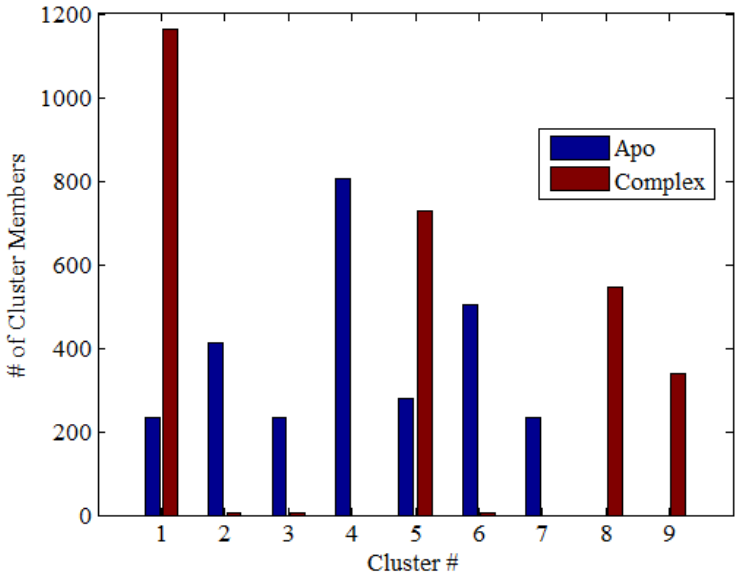

Figure 4.13. Mean overlap matrices between MCG_ENM modes of each snapshot and the reference modes.

examined in the recently published study [33].

5. MIMICKING LIGAND EFFECTS ON COLLECTIVE DYNAMICS

This chapter focuses on mimicking the effect of ligand binding on global modes in the absence of a bound ligand. In order to achieve this goal, a new methodology, named as frequency shift scanning, is developed based on standard (residue/C_α-based) ENM. As explained in Chapter 3.3, the side-chain heavy atoms of a specific residue are included as extra nodes in order to imitate the presence of a ligand that may interact with that specific site. Residues of the protein are scanned one-by-one by taking each residue's heavy atoms as extra nodes to calculate the resulting shift in eigenvalues of slowest modes. After scanning all residues of the protein, a color-coded map of the protein structure is obtained according to their effect on eigenvalues. Binding of a ligand to regions causing high eigenvalue shifts may be expected to restrict the collective dynamics and as a result affect protein's function.

5.1. TIM

The frequency shift scanning analysis is first performed on the T. cruzi TIM, which is the same apo state as the one in the previous chapter (PDB id: 1TCD). Here the aim is to observe if the tunnel region stands out in terms of its effect on collective modes. Figure 5.1 presents color-coded representations of TIM (from three different points of view), according to the average eigenvalue shifts based on first 10 modes. Red residues are the ones that act as significant constraints on the system, thereby shifting the collective mode frequencies to the right (increasing the vibrational frequencies). Blue residues correspond to solvent-exposed protruding loops or chain ends, shifting the eigenvalues to the left when extra masses are included.

As a result of the analysis, the tunnel region shows distinguishably positive shifts in eigenvalues. Especially the residues Tyr102 and Tyr103 give the highest eigenvalue shift, which correspond to the aromatic residues [14, 38, 39] that interact with the

benzothiazoles, such as bt10, discussed in the previous section. Other regions at the interface, such as the opposite side of the tunnel, do not indicate such significant effects as that of the tunnel region. Such a picture is in line with the findings of the previous chapter based on docking and MD simulations. Moreover, this method could possibly be utilized to detect allosteric sites that alter functional dynamics.

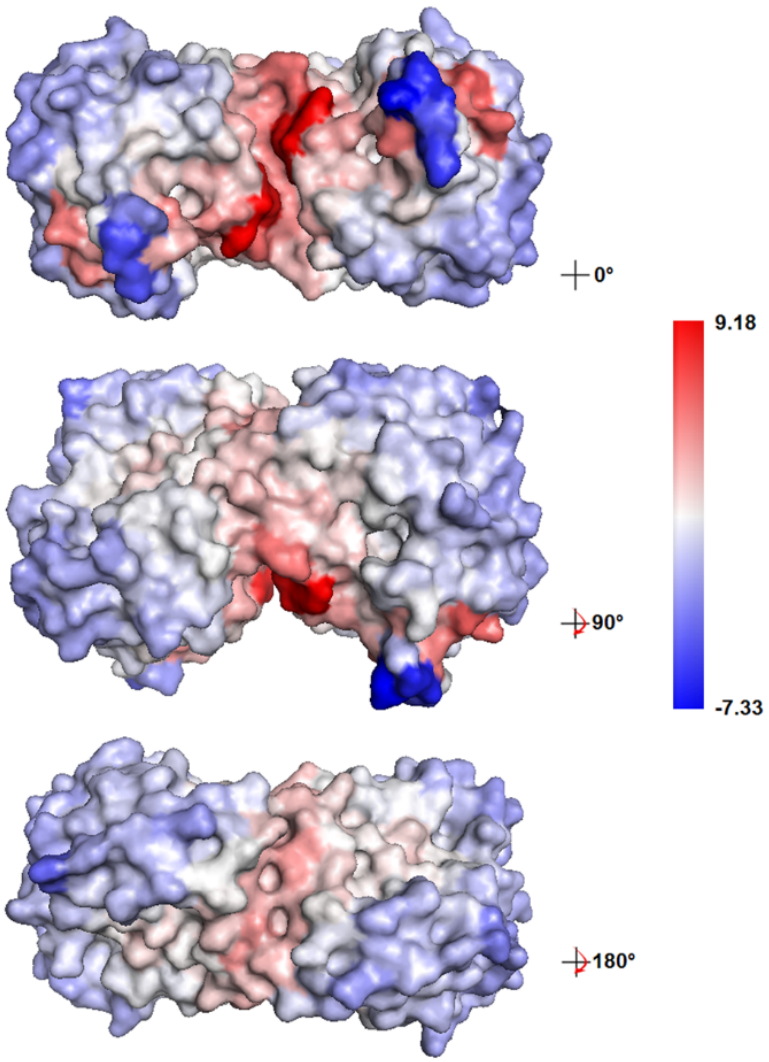

Figure 5.1. Frequency shift scanning of TIM molecule based on the first 10 modes.

The above analysis was performed with a standard cutoff (R_c) value of 15 Å in ENM. However, different cutoff values in the range of 7 to 15 Å were also used to analyze any dependence of the methodology on cutoffs. The cutoff value of 7 Å gives

more than six zero eigenvalues, which indicates that this value is not suitable for an alpha-carbon based ENM. Figure 5.2 shows the average change of first 10 eigenvalues based on inclusion of each residue's side chain atoms. Results for four different cutoff values are consistent in terms of the resulting profiles. Here, the vertical line divides the identical monomers of TIM that seem to present almost symmetric profiles. As a result, 15 Å cutoff was chosen for further calculations in order to avoid zero eigenvalue problem. This value has been also commonly used in literature [30, 40].

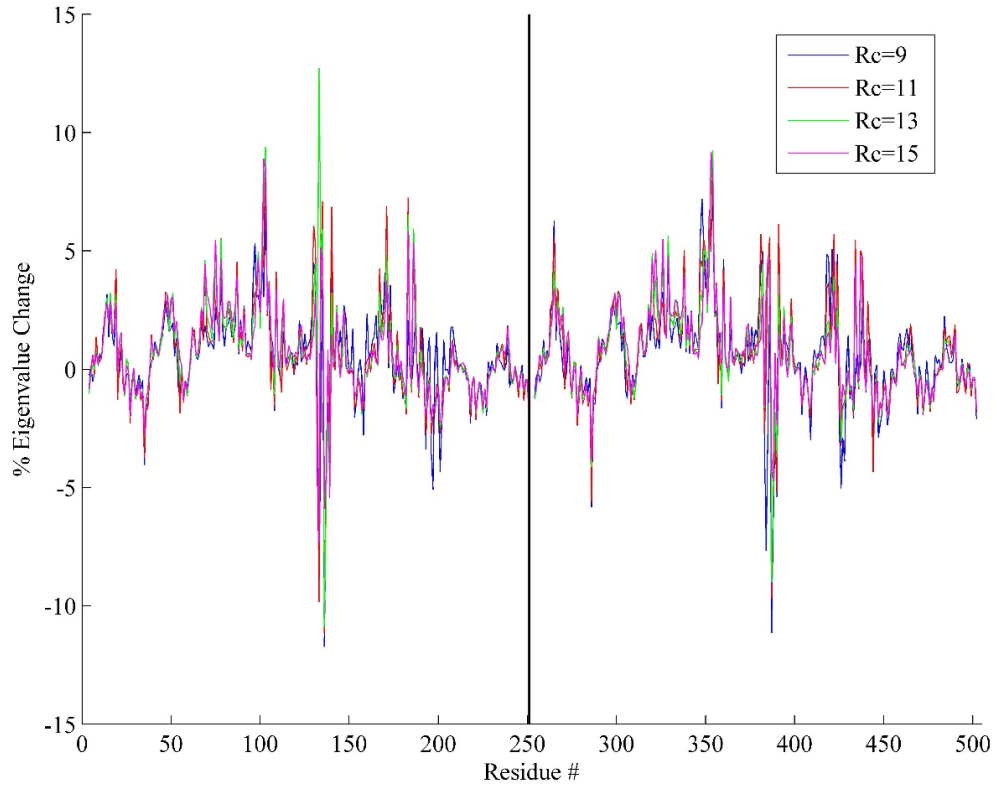

Figure 5.2. Average eigenvalue change of 10 slowest modes for different cutoff values.

In Figure 5.3, percent change in eigenvalues are given as box plots, where a distribution including all residues is provided separately for each mode. In this presentation each data point corresponds to a shift value obtained by carrying out the procedure for a single residue. Here, the median corresponds to almost zero for each mode. Both positive and negative eigenvalue shifts are observed for each mode, which seems like a contradiction with the predominant positive shifts observed in docking results.

However, the docking algorithms provide favorable poses, i.e. positions, in terms of interactions on the protein surface, whereas this methodology scans every residue in the protein without any previous bias in terms of binding or solvent exposure. As discussed in Chapter 4, the positive shifts may have functional importance due to their constraining effect on collective modes. They may indicate potential target regions for ligand binding, which will be further analyzed in the following sections on other proteins.

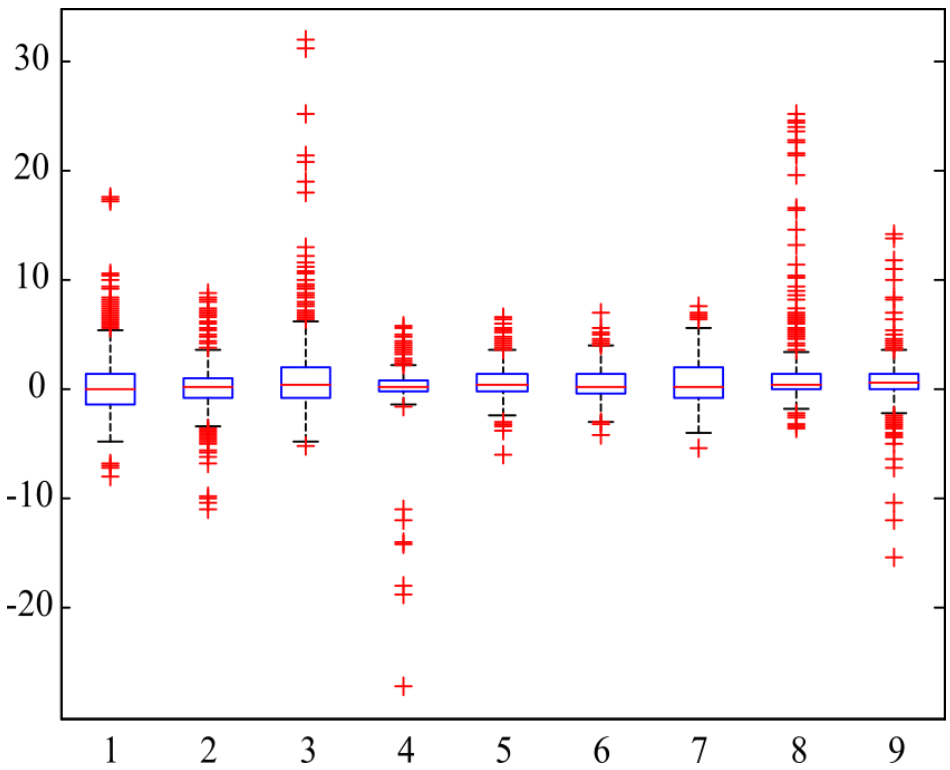

Figure 5.3. Distribution of eigenvalue shifts for the first nine modes.

As a counterpart of Figure 5.3, Figure 5.4 shows color-coded shifts for each mode on TIM molecule from front (tunnel region), top, and back points of view, respectively. Coloring was done in the range of minimum to maximum eigenvalue shifts observed in Figure 5.3.

In the first three global modes, which correspond to RM1, RM2 and RM3 presented in Figure 4.10, highest positive changes in eigenvalues are observed for the residues located in the tunnel region, which is consistent with the previous docking

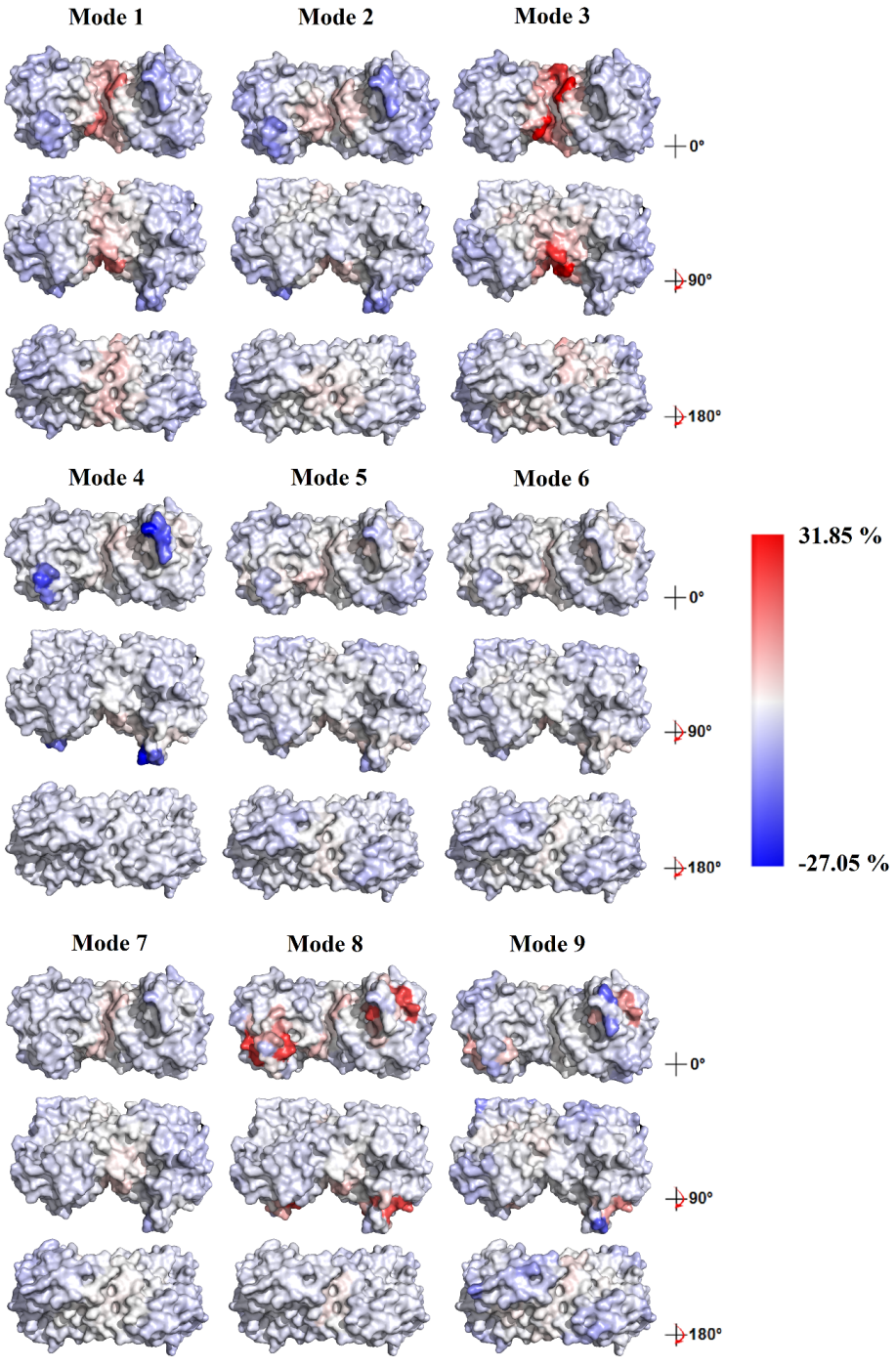

Figure 5.4. Shift of eigenvalues in the first nine modes colored-coded on TIM's surface.

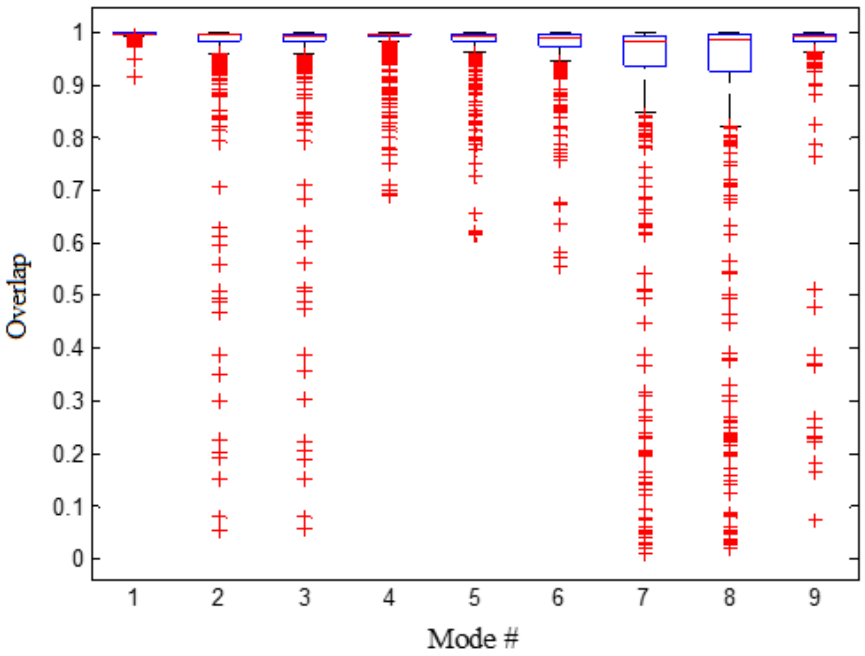

Figure 5.5. Distribution of overlap values for the first nine modes.

and MD results in Chapter 4. Negative shifts mostly occur at protrusions or bulge regions on the protein surface, whereas positive shifts mostly correspond to cavity regions. Although many cavities exist on the surface, none of these regions show as high a shift as the tunnel region, which points to the tunnel as a special target for ligands. Modes 4, 8 and 9 exhibit more localized negative and/or positive shifts around solvent-exposed α-helix regions in TIM. Especially in mode 8, some residues, which do not lie on a major cavity, cause high positive shifts because they pose a constraining effect on the protrusion's mobility. Modes five to seven do not exhibit any high shifts.

It is further analyzed whether the extra nodes added to each side chain alter the mode shape, i.e. the eigenvector, of a specific mode. Figure 5.5 shows the distributions of overlap values calculated by taking the dot product between ith mode eigenvectors obtained from standard ENM (without any extra nodes) and ENM with additional side chain atoms. Each data point corresponds to the effect of a single residue on a specific mode. Overlap value close to 1 indicates that the character of the mode is preserved, whereas a value close to 0 means that the mode shape has totally changed, which may

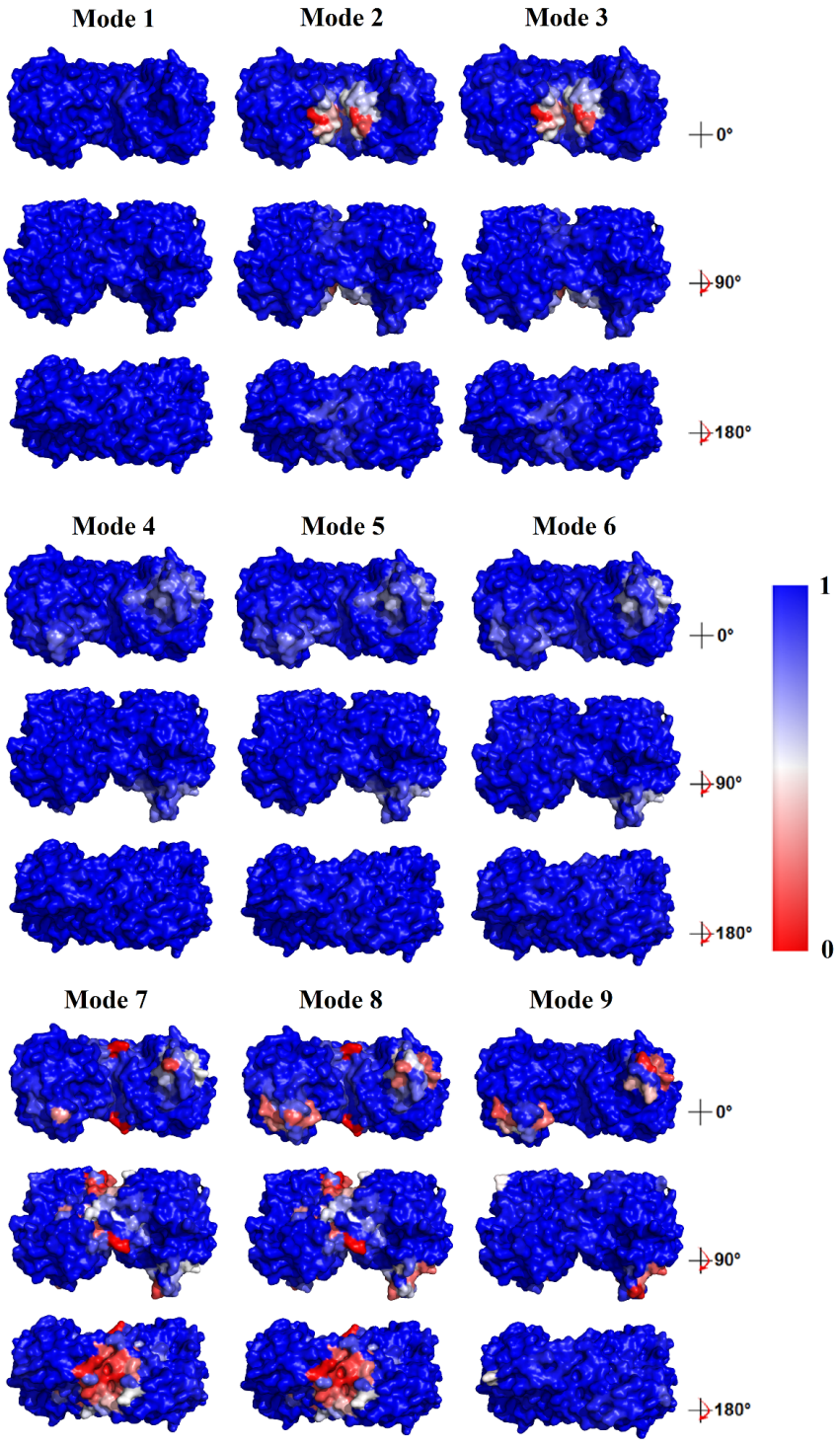

Figure 5.6. Overlap values for the first nine modes colored-coded on TIM's surface.

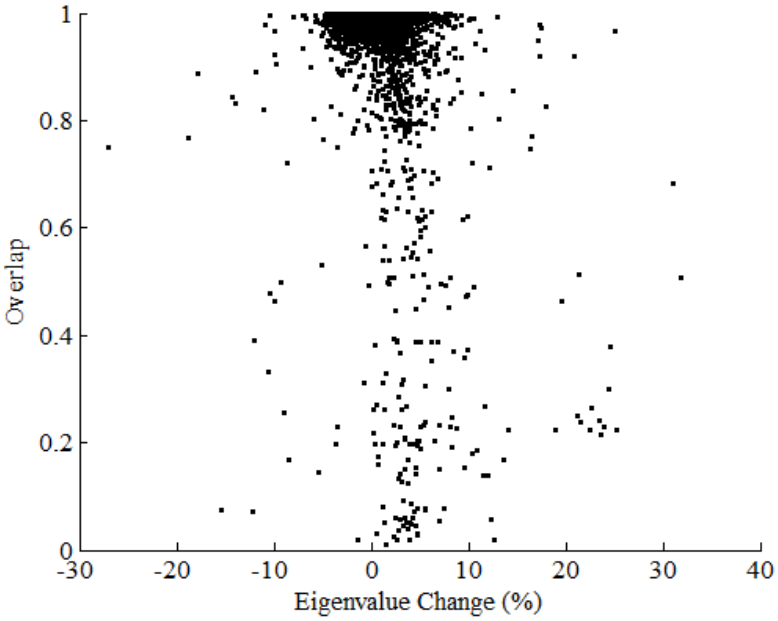

Figure 5.7. Scatter plot for overlap vs. eigenvalue change.

result from mode swapping. Even though the distributions generally lie closer to one especially for the most global modes, there are still some outliers.

Figure 5.6 provides the color-coded overlap values for each mode on surface representation of TIM. For most residues and in most of the modes there is almost no character change. When Figures 5.4 and 5.6 are examined together, the regions that show high changes mostly do not match in both figures except some regions in modes 8 and 9. In mode 2 and 3 some residues in the tunnel region induce some character change, however those residues do not directly match with the eigenvalue shifting residues in the tunnel region.

In Figure 5.7, a scatter plot of overlap values and eigenvalue change for first 10 modes of every residue was made in order to investigate any relation between eigenvalue changes and overlap values. The correlation coefficient between these two parameters is -0.34, which does not indicate any strong correlation. Thus, the eigenvalue shift and character change are mostly independent from each other.

5.2. Orotidine 5'-Monophosphate Decarboxylase (OMPDC)

The scanning method was applied to another enzyme, orotidine 5'-monophosphate decarboxylase (OMPDC), which takes role in pyrimidine biosynthesis. OMPDC is a homo-dimeric enyzme like TIM. In the apo structure PDB id: 3GDK), the catalytic loop, which in fact closes on the active site during reaction, is open conformation. Surface representation of 3GDK is shown in Figure 5.8, onto which the ligand orotidine 5'-phosphate before decarboxylation (from PDB structure 3GDT) is superposed to clarify the location of its catalytic site on the apo structure.

Figure 5.9 and 5.10 show the distribution of eigenvalue shifts and overlaps given for the first 10 modes. These distributions are quite similar to those observed for TIM (Figure 5.3), except for the presence of a residue with a large negative change. This residue (Lys209A), which can be clearly observed in Figure 5.11, lies on an extended loop and also leads to changes in mode shapes (see red colored residue on Figure 5.12b). This result is in parallel with results from TIM section, in which the lumpier regions give negative changes. However, the residues with positive change, which constrain the dynamics of the molecule, need to be focused on based on the observations so far. Therefore, in Figure 5.12a, the color-coding is performed only using positive shifts and all negative changes are considered zero.

In Figure 5.12a, residues causing positive eigenvalue shifts are mostly located at concave regions and the highest shifts corresponds to active site, i.e. the substrate binding region. However, not all concave regions are observed to create positive shifts in global modes, as in the case of TIM. These results indicate a potential for the method for detecting active site regions on the protein structures.

Residues (Lys59A, His61A, Lys93A, and Ser154A) that make hydrogen bonds with the ligand in the crystal structure 3GDT all give positive shifts. Especially the residue Ser154A gives very high positive shift. High positive shift giving residues can be seen in Figure 5.11, where average of first ten eigenvalue change vs. residue number is given for open loop domain (subunit A) of 3GDT.

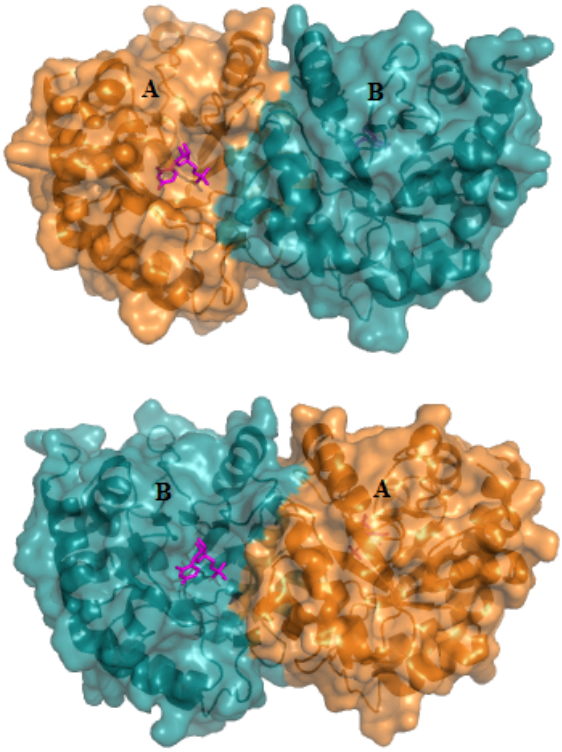

Figure 5.8. Structure of OMPDC with ligand.

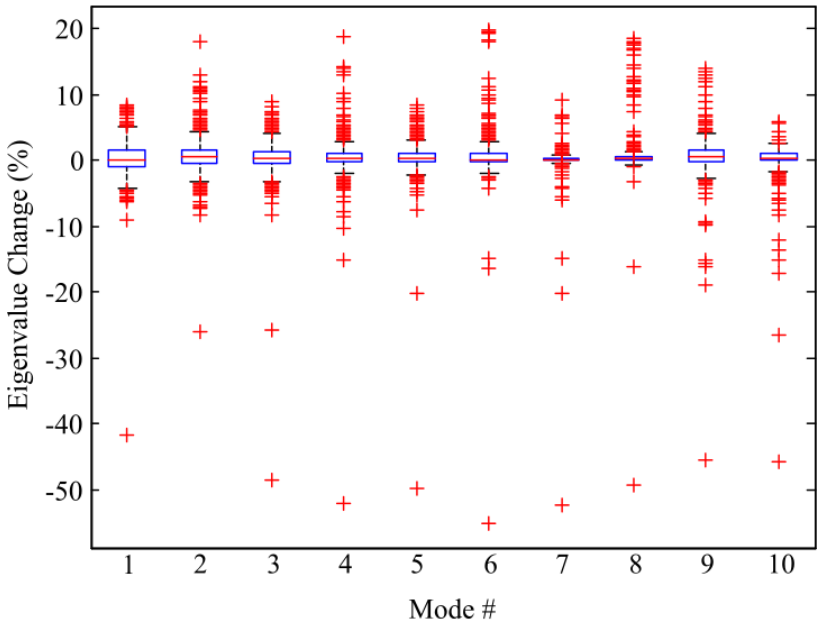

Figure 5.9. Distributions of eigenvalue shift (%) for 3GDK.

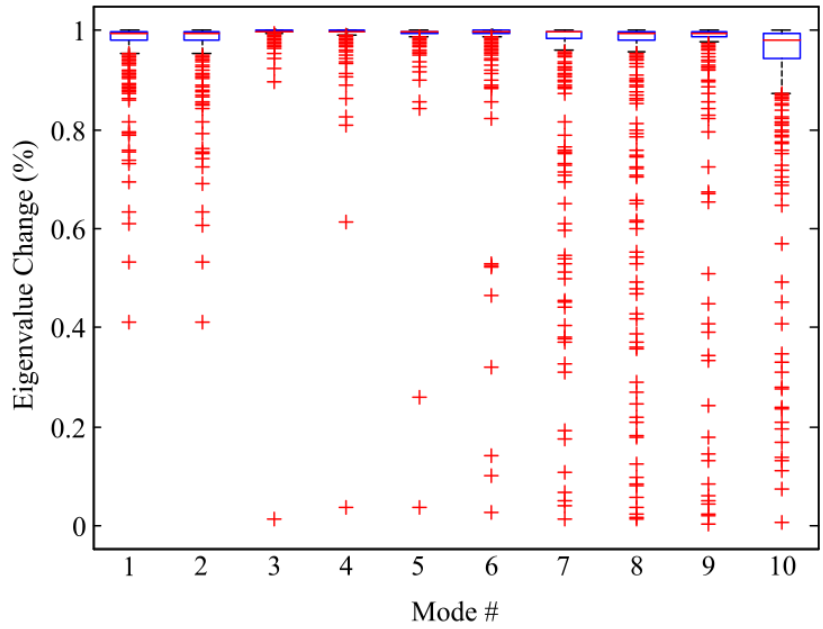

Figure 5.10. Distributions of overlap for 3GDK.

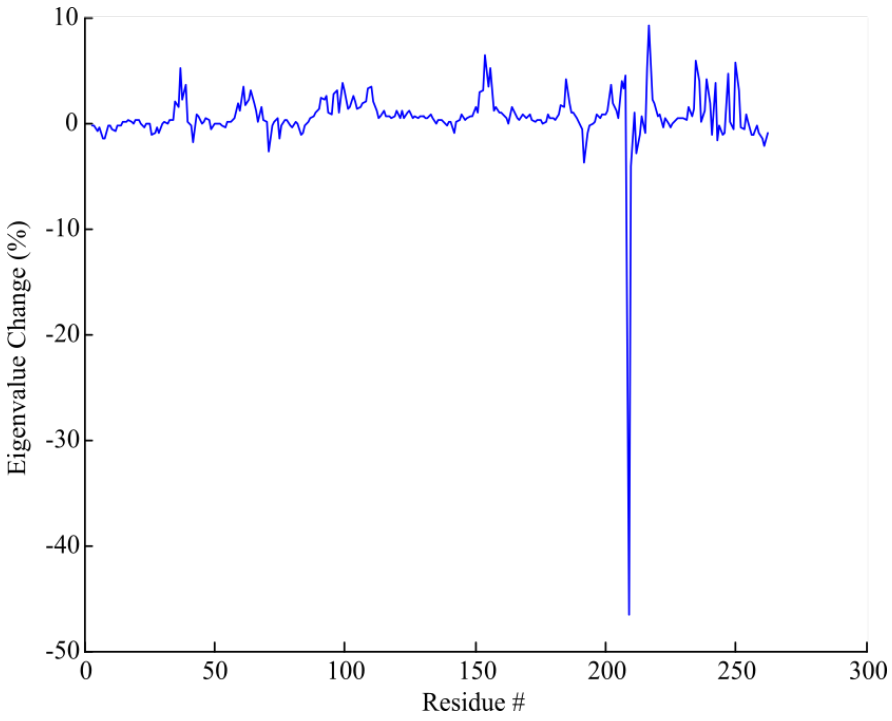

Figure 5.11. Eigenvalue change vs. Residue # for 3GDK.

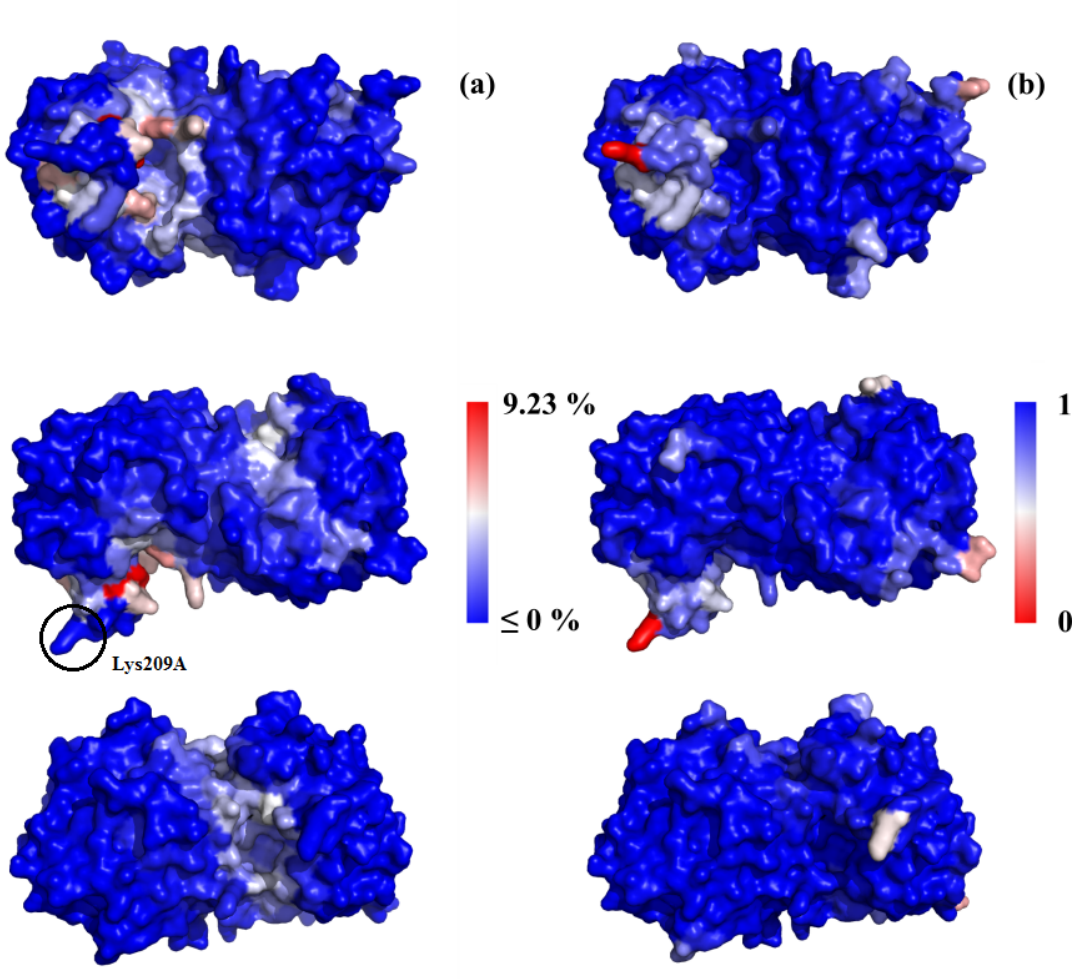

Figure 5.12. (a) Eigenvalue shift (%) and (b) overlap average for first 10 modes for 3GDK.

In the scanning analysis of OMPDC, only the average of overlap values were considered in the character change analysis (see Figure 5.12b). Character changing residues are mostly on the loop region and most character changing one is the Lys209A which is also the residue showing most negative change in eigenvalues. However, similar to the results for TIM, character change does not correspond to any distinctive property of the structure we have considered.

5.3. Adenylate Kinase (AdK)

The method was also applied on different conformers of AdK, which is an enzyme that undergoes large conformational changes. AdK is involved in cellular energy homeostasis by catalyzing the interconversion of adenine nucleotides. AdK shows a maximum RMSD change of 7.14 Å between its apo/open (4AKE) and bound/closed structures (1AKE). In Figure 5.13 surface and cartoon representations for these structures are given. In Figure 5.13a, the open structure (4AKE) does not originally contain any ligand however the ligand (ATP) was taken from the closed structure (1AKE) and superposed on open structure in order to give reader an idea about where the active site is. Various AdK conformers have been previously generated [41] using an ENM-based algorithm, which sample states between its open and closed structures and used in this work. These conformers have been named starting from 'gen1' to 'gen7', which have respective RMSDs of 7.1 Å and 2.4 Å to the closed structure, respectively. They represent a sequence of conformations starting from the 4AKE (closest to gen1) and sequentially getting closer to 1AKE (closest to gen7).

In Figure 5.14 results for various conformers of AdK with ranging RMSDs from crystal structure 4AKE superposed each other. Eigenvalue changes for most of the residues do not vary in different conformers, whereas for some residues eigenvalue changes show great differences from one another. The median of the distributions for eigenvalue shifts is around zero for each mode (Figure 5.16).

Character change of the modes examined through one-to-one overlap of the eigenvectors of the system and generally follow same profile with occasional drops (see Figure

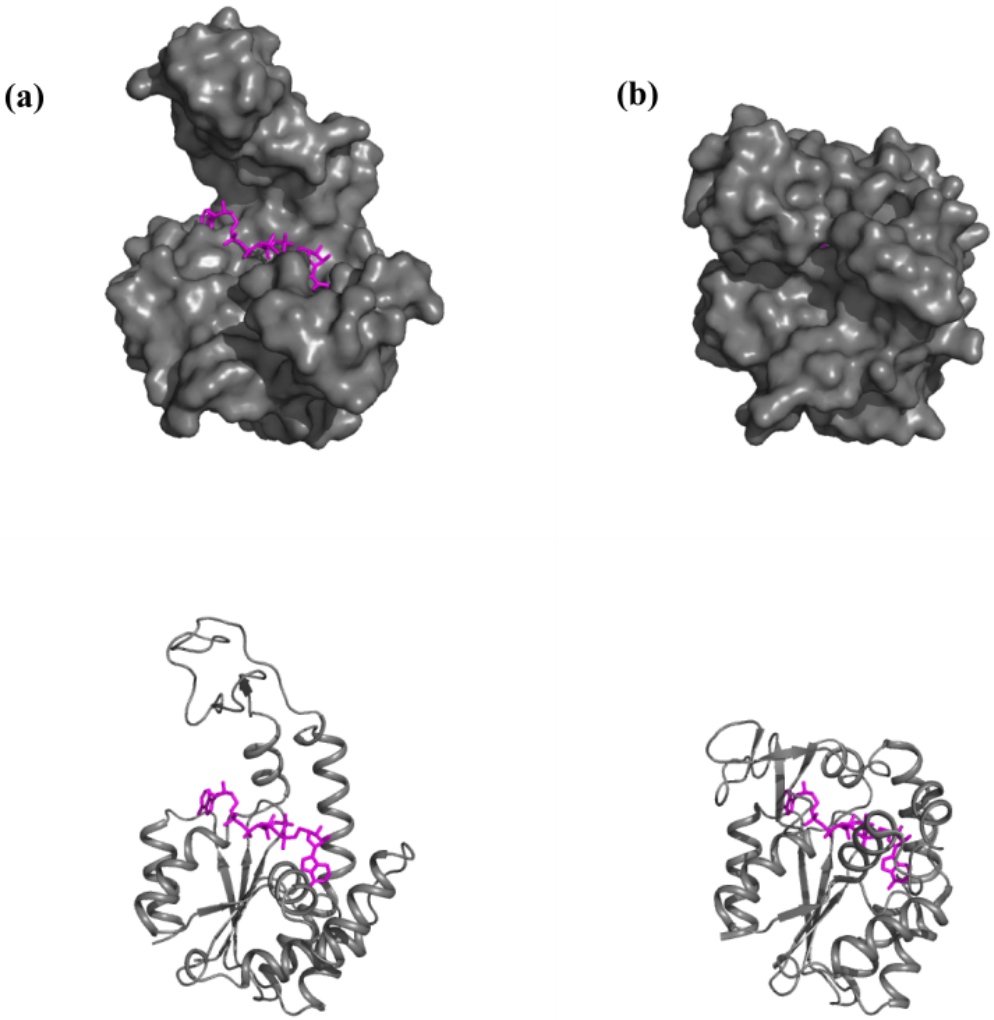

Figure 5.13. Open (a) and closed (b) structures of AdK with ligand.

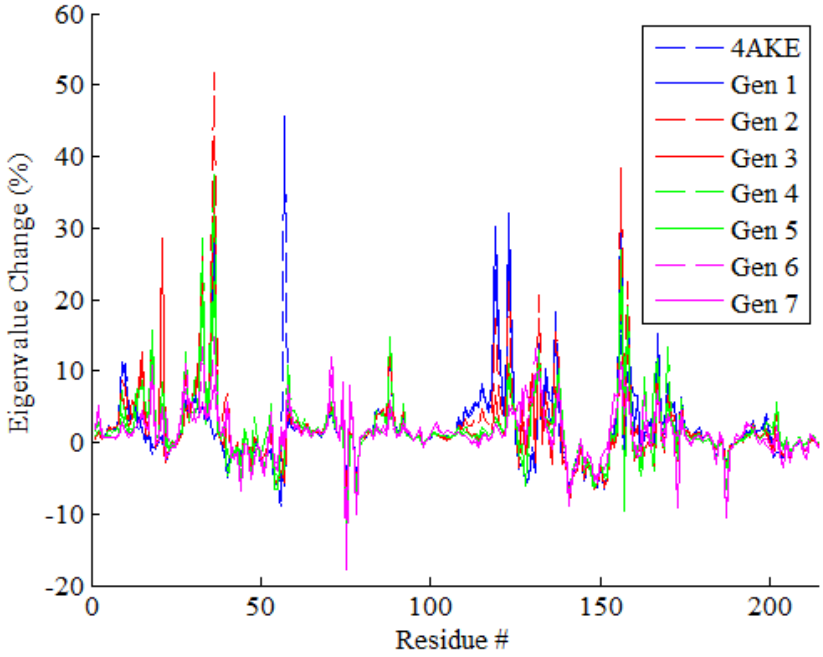

Figure 5.14. Eigenvalue shift (%) vs. residue # for various conformers of AdK.

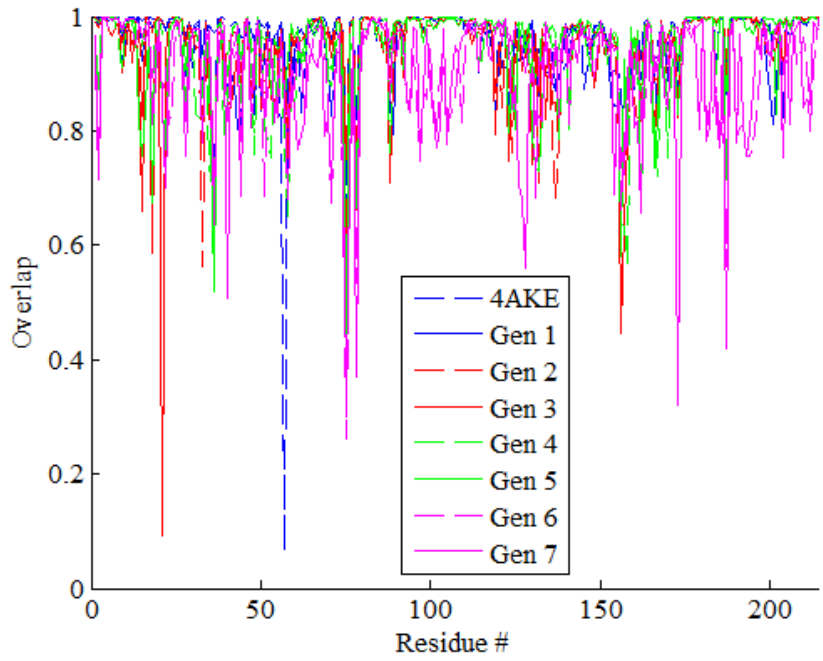

Figure 5.15. Overlap vs. residue # for various conformers of AdK.

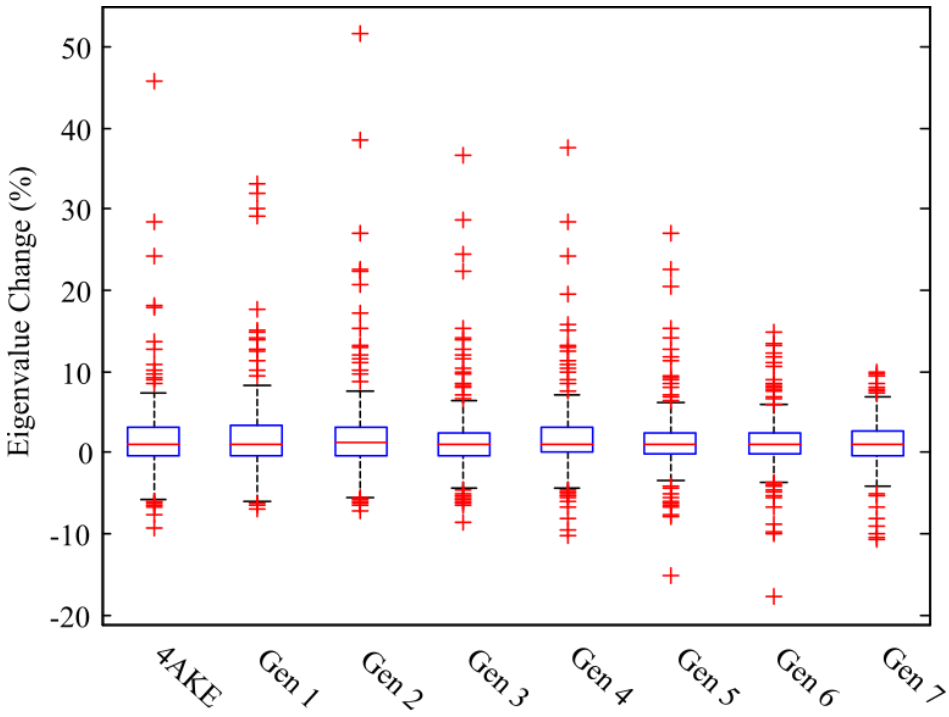

Figure 5.16. Distribution of average eigenvalue shift of first ten modes (%) for various conformers of AdK.

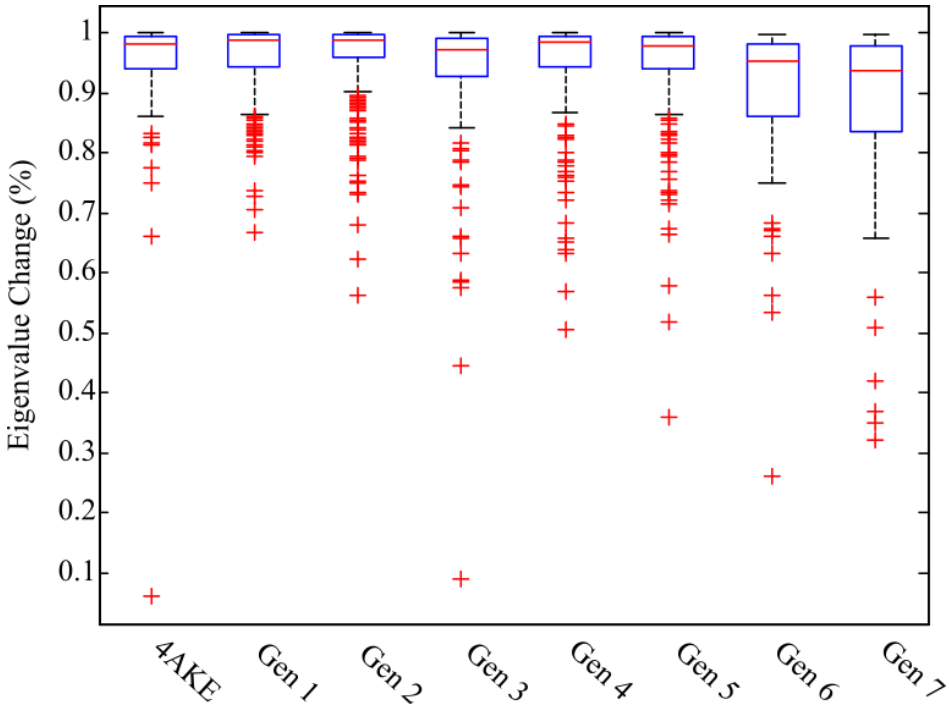

Figure 5.17. Distribution of overlap values for various conformers of AdK.

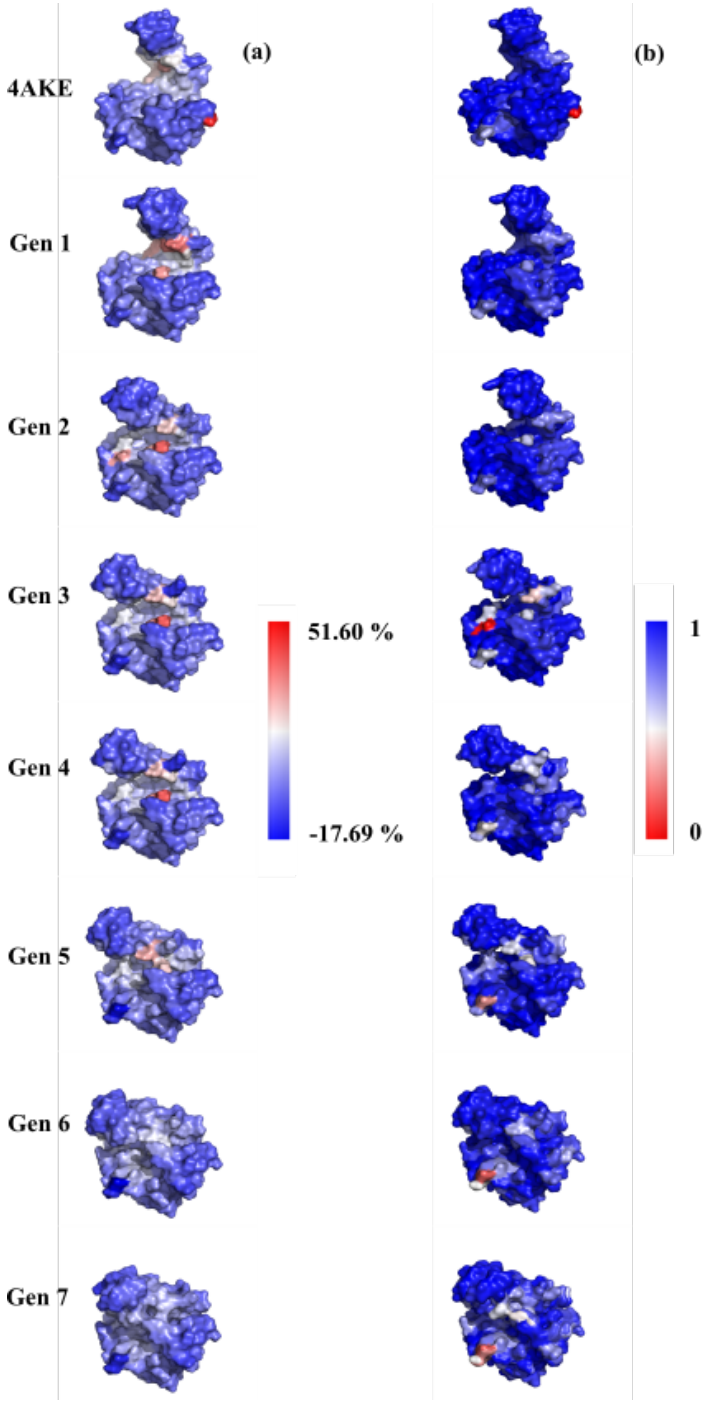

Figure 5.18. (a) Eigenvalue shift (%) (b) Overlap for various conformers of AdK.

5.15). However, overlap values do not directly corresponds to the eigenvalue shift likewise in the case of TIM and OMPDC. Moreover, character change does not show the same profile as the eigenvalue change which is middle generations showing more shift in the ligand binding region. On the other hand, Overlap values tilted towards zero as the molecule became closed (see Figure 5.17 and Figure 5.18). Also, expanded regions does not create character change as it was the case in OMPDC.

Figures 5.16 and figure 5.17 show distributions of eigenvalue shift and overlap values of ten mode average for different conformer. In Figure 5.18, the eigenvalue shifts and overlap values are color-coded on the corresponding structure, which are from different generations or the open crystal structure of AdK (see figure for counterparting distributions). The ligand binding regions of the crystal (open) structure and the later generations (gen 6, 7; closed structures) do not show meaningful positive change. However in the middle generations (gen 1-5), some residues in the binding region show significant positive change. In this example, results suggest that the middle half-open-half-closed structures are more suitable to be a receptor for a ligand.

Arg36, Lys57, Arg119, Arg123, and Arg156 are the residues showing five highest positive changes (above 30%) in all of the AdK conformers. Although the ligand is an exceptionally big one and interacting with many of the residues, it forms hydrogen bond with residues Arg36, Lys57, Arg123, and Arg156 and contact with Arg119.

5.4. Biotin Carboxylase

Biotin carboxylase (BC) takes role in ATP-dependent carboxylation of biotin in fatty acid synthesis. BC is relatively large protein with 894 residues, and undergoes a conformational change of 4.1 Å to grasp ATP.

BC has two catalytic sites located on distinct subunits of its homo-dimeric structure. Figure 5.19 focuses the active site of BC with the bound ligand (ADP). Four different conformers (named as generations) of BC between its open structure (PDB id: 1dv1) and closed structure (PDB id: 1dv2) are analyzed here using the same

methodology applied throughout the chapter.

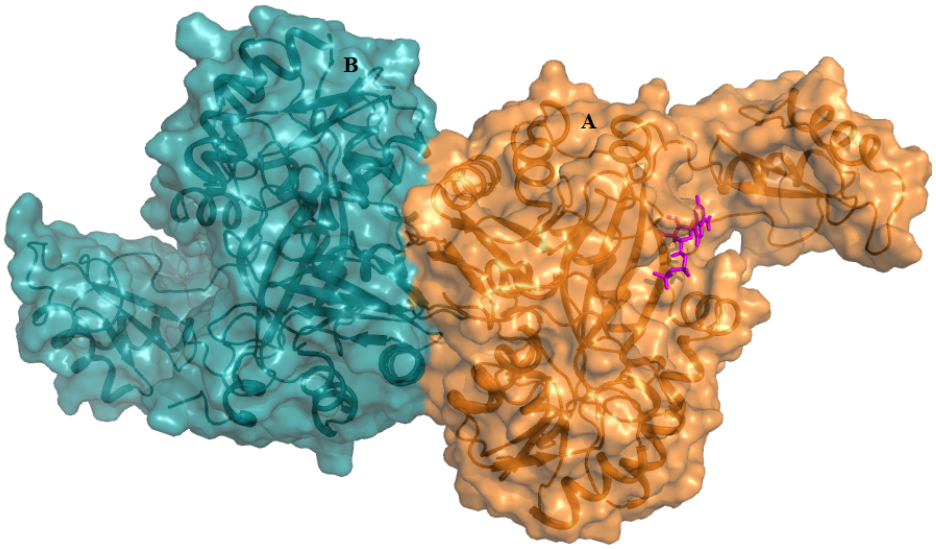

Figure 5.19. Ligand (ADP) superposed on structure Gen 1.

Average of the first ten modes' eigenvalue changes calculated are plotted in Figures 5.20 and 5.21. Among the four structures, the first conformer (gen1) carries more peak residues, some of which are in catalytic region. In all of the conformers, a peak is observed for Lys159A, which is on a loop contacting with the ligand ADP.

Figures 5.22 and 5.23 indicate color-coded structures based on the eigenvalue shift and overlap values as averages over the first ten modes. Both eigenvalue change and overlap values peak at the same residue Lys159A, which has not been unobserved in previous sections. In general, eigenvalues fluctuate around zero and adopt higher shifts near the active region.

5.5. Overview of scanning methodology

All in all, this methodology is tried on TIM, OMPDC, AdK, and BC. A detailed analysis made for TIM and obtained successful and conclusive results. Tunnel region, the known allosteric region, is highlighted in first three modes of global motion showing that a constraining affect will be on vibrational dynamics of the structure. This is the

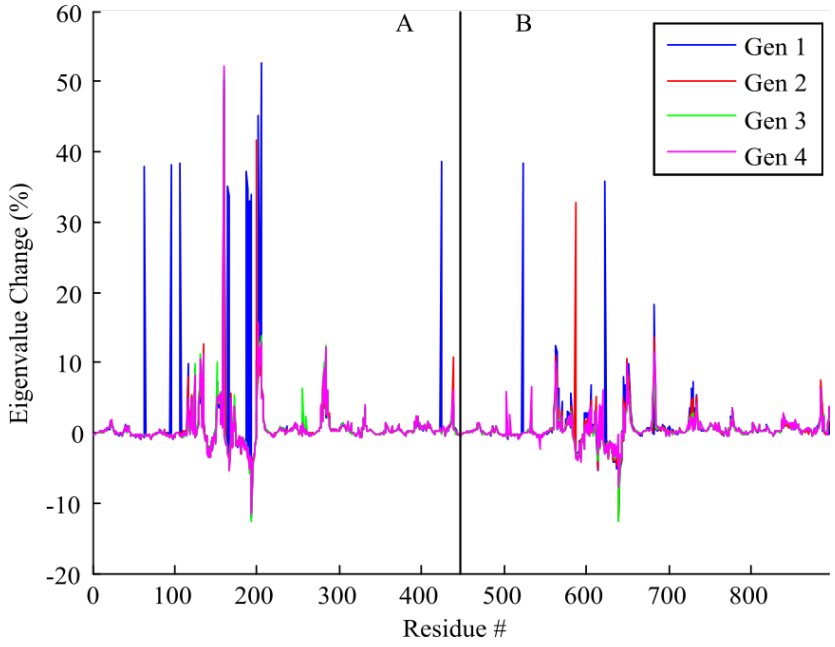

Figure 5.20. Average of first ten eigenvalue change (%) vs. residue # for conformers of BC.

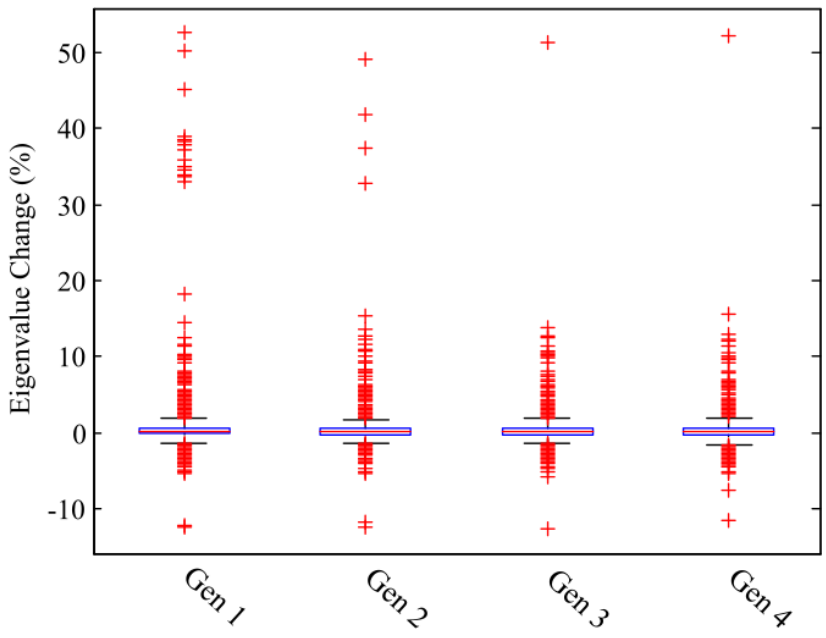

Figure 5.21. Distribution of average of first ten eigenvalue change (%) vs. residue # for conformers of BC.

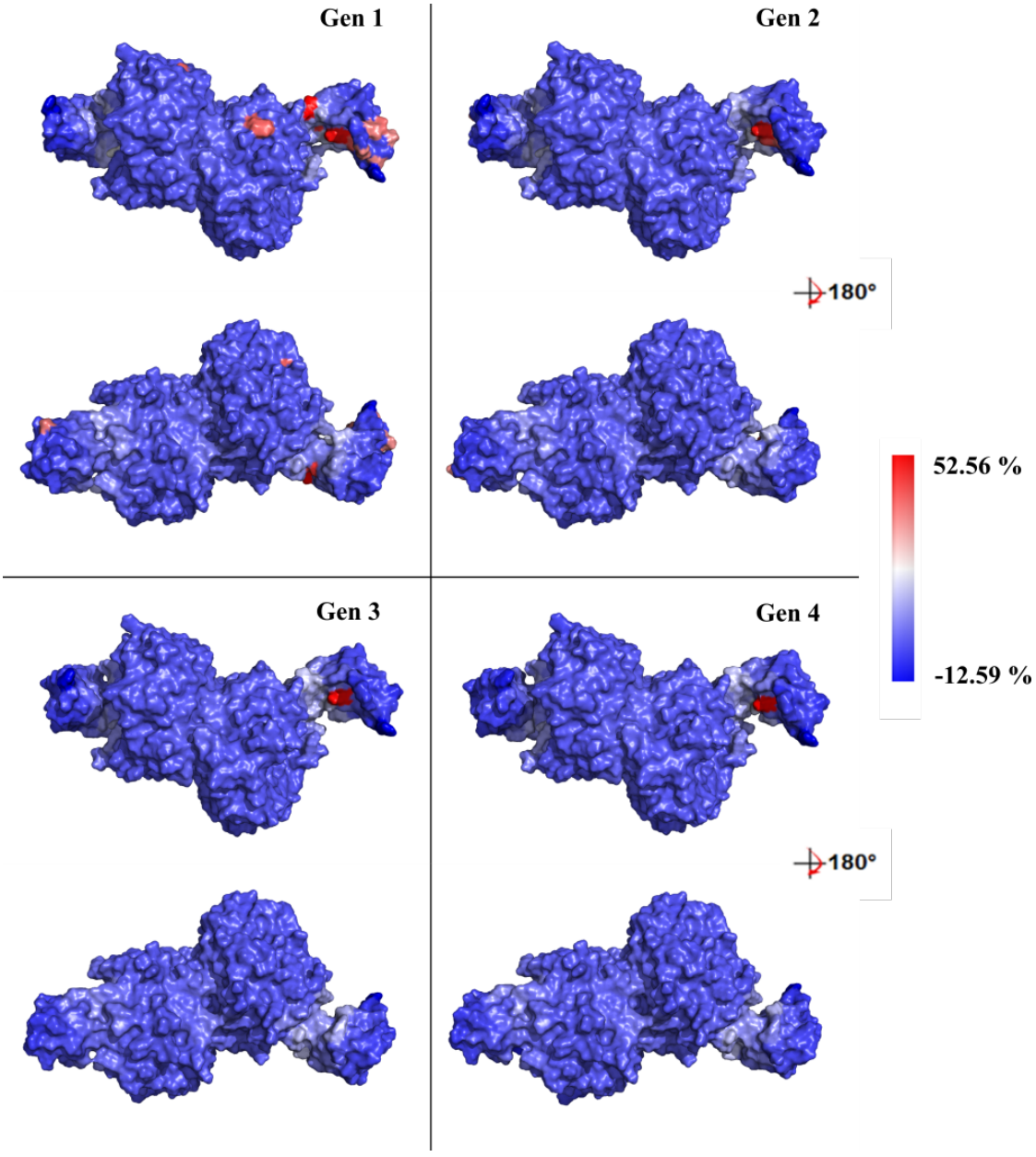

Figure 5.22. Average of eigenvalue shift (%) for various conformers of BC.

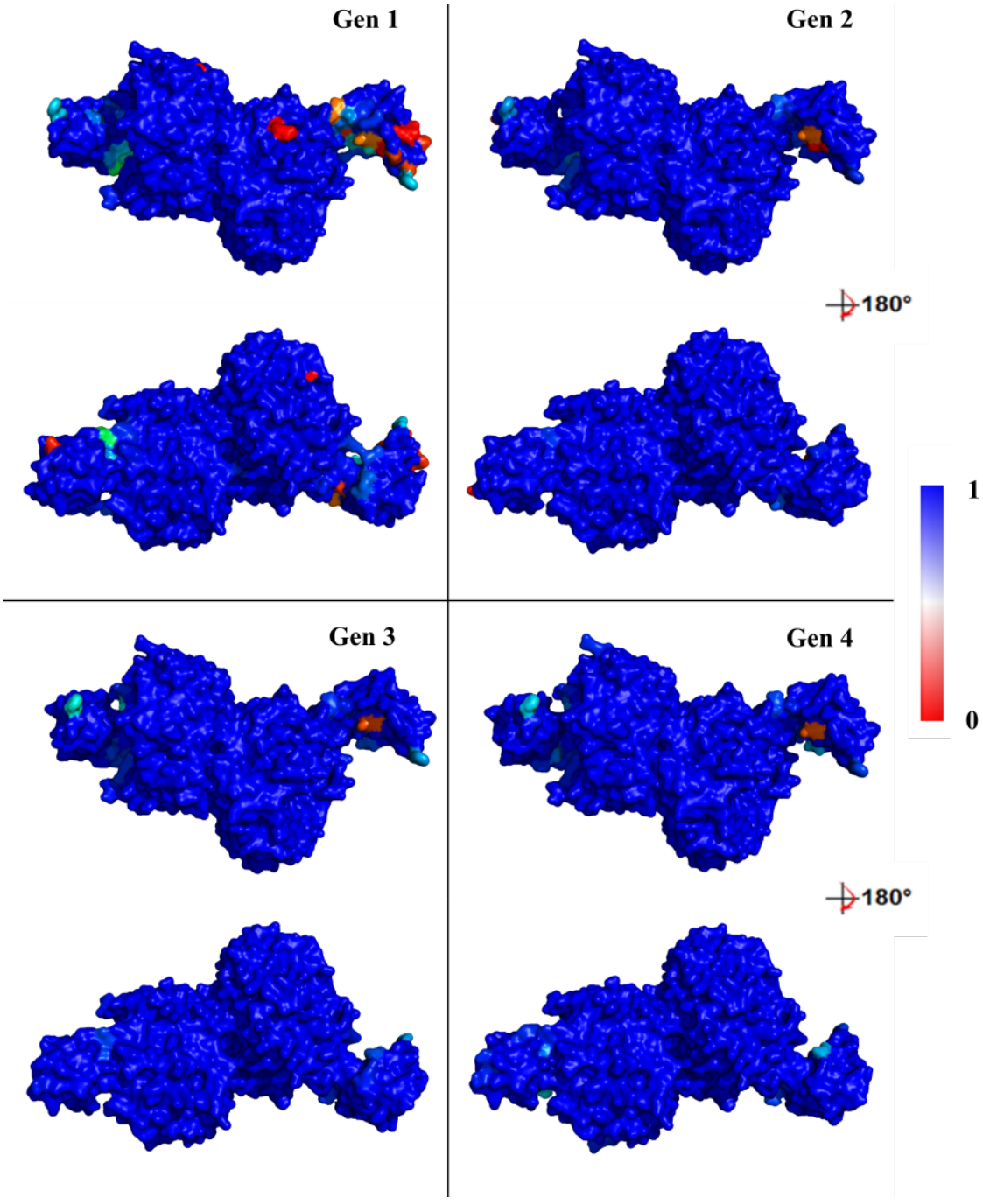

Figure 5.23. Average of overlap values for various conformers of BC.

exact result obtained in the Chapter 4 while comparing docking poses where ligand bound to the tunnel region with docking poses where ligand bound elsewhere. Also it is observed that this eigenvalue shift does not caused by a character change of modes and character changing residue does not seem to correspond any meaningful region. Then the method is tried on OMPDC, a similar structure to TIM. For this structure active site of the protein is highlighted which is not the case with TIM. Many conformers of AdK molecule are examined. Again ligand binding regions show more positive shift (constraining effect) than rest of the structure and half-open-half-closed middle conformers illustrate this effect more clearly. However the results for BC seem inconclusive. The method shows some high peak at some residues in contrast with their neighboring residues where in previous cases the values shows much more smooth changes. It should also be mentioned that the BC is considerably larger than other structures and could be root of the problem. Almost in all cases, eigenvalue shift does not correlate with overlap values meaning eigenvalue changes does not caused by character change but rather by a change of frequency of the mode.

It also needs to be emphasized that this scanning technique high depends on the computationally efficient ENM algorithms which gives global vibrational motions of the any system with a comparatively low cost. With a more time consuming method it would not plausible to scan every residue of a given structure.

6. CONCLUSION

6.1. Conclusions

This thesis focuses on the analysis of ligand effect on vibrational dynamics of TIM based on elastic network models. For this aim, the computational efficient MCG_ENM [3, 4] is utilized here for the first time to consider the ligand explicitly in protein complexes. This method systematically considers all heavy atoms of the ligand and as a result represents the inter-molecular interactions better than the alternative approach, where an arbitrary number of coarse-grained node(s) is assigned for the ligand.

In the first section, a total of 1496 blind docking poses including five different types of ligands (two non-inhibitors, three inhibitors) on TIM were analyzed with MCG_ENM. Larger ligands (inhibitors 8, 9, and 10) prefer tunnel region on TIM interface more than smaller ligands (non-inhibitors 2 and 3). Inhibitors located at the tunnel region lead to relatively higher frequency shifts, which can be related to their experimentally determined inhibitory effect. Especially, the first two slowest modes are significantly affected, which have been shown [4] to be functionally important. This study may provide insight on the allosteric coupling between the tunnel region at the interface and the catalytic site of TIM.

In the second part, a total of 54,000 MD snapshots coming from six independent MD runs (three on apo TIM and three on TIM's complex with bt10) were examined with MCG_ENM. Analysis of MD snapshots shows that two global modes, namely counter-rotation and bending of the subunits, are altered significantly in the presence of the inhibitor. The vibrational frequencies of these modes shift to the right emphasizing a constraining effect, which may potentially alter substrate entrance, substrate positioning in the active site and/or product release steps. Moreover, clustering of MD snapshots indicates that the apo runs span a broader conformational space than the complex runs, which is consistent with the constraining effect due to the ligand bound

at the tunnel. In summary, the tunnel region at the interface of TIM presents a key site for modifying its conformational dynamics at global and local scales.

Scanning of residues by modifying the side chains in standard ENM [29,30] analysis is introduced in the last chapter as a new method to determine sites/residues that might potentially alter global dynamics as a result of ligand binding. The scanning of TIM residues indicates that constraining the tunnel region significantly changes global vibrational modes, which is in line with MCG_ENM analyses on TIM and may be related with an allosteric regulation. The method also identifies the functional sites for proteins undergoing conformational changes. For such cases intermediate conformers with relatively mobile domains give much more constraining impact on the global modes than conformers with relatively closed domains.

Overall this study underlines that binding of ligands to key sites at subunit interfaces may act as constraints for the folded network of interactions, and thereby affect enzyme dynamics and function. Moreover a potential new method for identifying effective key regions on the protein structures for altering functional vibrational motions is introduced.

6.2. Recommendations

The analysis made for docking poses could be widened by applying to other proteins and docking benchmark in order to observe its potential use as a post-processing procedure for docking poses obtained with conventional algorithms. MCG_ENM may also be helpful in quantifying the extent to which alternative ligand positions alter functional modes.

The success of scanning method introduced here depends on the computational efficiency of ENM analysis. Other types of ENM based scanning methods could also be developed such as extending cutoff radius of every residue or altering mass and spring constants of every residue one by one. Moreover application of further coarse graining methodology via considering a fraction of residues in scanning would increase

the method's computational efficiency and enable its application to large multimeric systems or supramolecules.

REFERENCES

1. Ben-Avraham, D., "Vibrational Normal-Mode Spectrum of Globular Proteins", *Physical Review B*, Vol. 47, No. 21, p. 14559, 1993.

2. Alakent, B., S. Baskan and P. Doruker, "Effect of Ligand Binding on the Intraminimum Dynamics of Proteins", *Journal of Computational Chemistry*, Vol. 32, No. 3, pp. 483–496, 2011.

3. Kurkcuoglu, O., O. T. Turgut, S. Cansu, R. L. Jernigan and P. Doruker, "Focused Functional Dynamics of Supramolecules by Use of a Mixed-Resolution Elastic Network Model", *Biophysical Journal*, Vol. 97, No. 4, pp. 1178–1187, 2009.

4. Kurkcuoglu, O., R. L. Jernigan and P. Doruker, "Loop Motions of Triosephosphate Isomerase Observed with Elastic Networks", *Biochemistry*, Vol. 45, No. 4, pp. 1173–1182, 2006.

5. Orosz, F., J. Olah and J. Ovadi, "Triosephosphate Isomerase Deficiency: Facts and Doubts", *IUBMB Life*, Vol. 58, No. 12, pp. 703–715, 2006.

6. Schnackerz, K. D. and R. W. Gracy, "Probing the Catalytic Sites of Triosephosphate Isomerase by 31P-NMR With Reversibly and Irreversibly Binding Substrate Analogues", *European Journal of Biochemistry*, Vol. 199, No. 1, pp. 231–238, 1991.

7. Maldonado, E., M. Soriano-Garca, A. Moreno, N. Cabrera, G. Garza-Ramos, M. T. de Gomez-Puyou, A. Gomez-Puyou and R. Perez-Montfort, "Differences in the Intersubunit Contacts in Triosephosphate Isomerase from Two Closely Related Pathogenic Trypanosomes", *Journal of Molecular Biology*, Vol. 283, No. 1, pp. 193–203, 1998.

8. Wierenga, R., E. Kapetaniou and R. Venkatesan, "Triosephosphate Isomerase: A Highly Evolved Biocatalyst", *Cellular and Molecular Life Sciences*, Vol. 67, No. 23,

pp. 3961–3982, 2010.

9. Zhai, X., M. K. Go, A. C. O'Donoghue, T. L. Amyes, S. D. Pegan, Y. Wang, J. P. Loria, A. D. Mesecar and J. P. Richard, "Enzyme Architecture: The Effect of Replacement and Deletion Mutations of Loop 6 on Catalysis by Triosephosphate Isomerase", *Biochemistry*, Vol. 53, No. 21, pp. 3486–3501, 2014.

10. Cansu, S. and P. Doruker, "Dimerization Affects Collective Dynamics of Triosephosphate Isomerase", *Biochemistry*, Vol. 47, No. 5, pp. 1358–1368, 2008.

11. Kurkcuoglu, Z. and P. Doruker, "Substrate Effect on Catalytic Loop and Global Dynamics of Triosephosphate Isomerase", *Entropy*, Vol. 15, No. 3, pp. 1085–1099, 2013.

12. Perez-Montfort, R., M. de Gomez-Puyou and A. Gomez-Puyou, "The Interfaces of Oligomeric Proteins as Targets for Drug Design Against Enzymes from Parasites", *Current Topics in Medicinal Chemistry*, Vol. 2, No. 5, pp. 457–470, 2002.

13. Tellez-Valencia, A., S. Avila-Rios, R. Perez-Montfort, A. Rodriguez-Romero, M. T. de Gomez-Puyou, F. Lopez-Calahorra and A. Gomez-Puyou, "Highly Specific Inactivation of Triosephosphate Isomerase from Trypanosoma Cruzi", *Biochemical and Biophysical Research Communications*, Vol. 295, No. 4, pp. 958–963, 2002.

14. Kurkcuoglu, Z., G. Ural, E. Demet Akten and P. Doruker, "Blind Dockings of Benzothiazoles to Multiple Receptor Conformations of Triosephosphate Isomerase from Trypanosoma Cruzi and Human", *Molecular Informatics*, Vol. 30, No. 11-12, pp. 986–995, 2011.

15. Changeux, J.-P. and S. J. Edelstein, "Allosteric Mechanisms of Signal Transduction", *Science*, Vol. 308, No. 5727, pp. 1424–1428, 2005.

16. Henzler-Wildman, K. and D. Kern, "Dynamic Personalities of Proteins", *Nature*, Vol. 450, No. 7172, pp. 964–972, 2007.

17. Bahar, I., T. R. Lezon, L.-W. Yang and E. Eyal, "Global Dynamics of Proteins: Bridging Between Structure and Function", *Annual Review of Biophysics*, Vol. 39, p. 23, 2010.

18. Bahar, I., C. Chennubhotla and D. Tobi, "Intrinsic Dynamics of Enzymes in the Unbound State and Relation to Allosteric Regulation", *Current Opinion in Structural Biology*, Vol. 17, No. 6, pp. 633–640, 2007.

19. Bahar, I., T. R. Lezon, A. Bakan and I. H. Shrivastava, "Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins", *Chemical Reviews*, Vol. 110, No. 3, pp. 1463–1497, 2009.

20. Tokuriki, N. and D. S. Tawfik, "Protein Dynamism and Evolvability", *Science*, Vol. 324, No. 5924, pp. 203–207, 2009.

21. Hammes-Schiffer, S. and S. J. Benkovic, "Relating Protein Motion to Catalysis", *Annu. Rev. Biochem.*, Vol. 75, pp. 519–541, 2006.

22. Smock, R. G. and L. M. Gierasch, "Sending Signals Dynamically", *Science*, Vol. 324, No. 5924, pp. 198–203, 2009.

23. Rodgers, T. L., P. D. Townsend, D. Burnell, M. L. Jones, S. A. Richards, T. C. McLeish, E. Pohl, M. R. Wilson and M. J. Cann, "Modulation of Global Low-Frequency Motions Underlies Allosteric Regulation: Demonstration in CRP/FNR Family Transcription Factors", *PLoS Biol*, Vol. 11, No. 9, p. e1001651, 2013.

24. McLeish, T. C., T. Rodgers and M. R. Wilson, "Allostery Without Conformation Change: Modelling Protein Dynamics at Multiple Scales", *Physical Biology*, Vol. 10, No. 5, p. 056004, 2013.

25. Tama, F. and Y.-H. Sanejouand, "Conformational Change of Proteins Arising From Normal Mode Calculations", *Protein Engineering*, Vol. 14, No. 1, pp. 1–6, 2001.

26. Xu, C., D. Tobi and I. Bahar, "Allosteric Changes in Protein Structure Computed by a Simple Mechanical Model: Hemoglobin T↔ R2 transition", *Journal of Molecular Biology*, Vol. 333, No. 1, pp. 153–168, 2003.

27. Bakan, A. and I. Bahar, "The Intrinsic Dynamics of Enzymes Plays a Dominant Role in Determining the Structural Changes Induced Upon Inhibitor Binding", *Proceedings of the National Academy of Sciences*, Vol. 106, No. 34, pp. 14349–14354, 2009.

28. Yang, L., G. Song and R. L. Jernigan, "How Well Can We Understand Large-Scale Protein Motions Using Normal Modes of Elastic Network Models?", *Biophysical Journal*, Vol. 93, No. 3, pp. 920–929, 2007.

29. Doruker, P., A. R. Atilgan and I. Bahar, "Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to Alpha-Amylase Inhibitor", *Proteins: Structure, Function, and Bioinformatics*, Vol. 40, No. 3, pp. 512–524, 2000.

30. Atilgan, A., S. Durell, R. Jernigan, M. Demirel, O. Keskin and I. Bahar, "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model", *Biophysical Journal*, Vol. 80, No. 1, pp. 505–515, 2001.

31. Huey, R., G. M. Morris, A. J. Olson and D. S. Goodsell, "A Semiempirical Free Energy Force Field with Charge-Based Desolvation", *Journal of Computational Chemistry*, Vol. 28, No. 6, pp. 1145–1152, 2007.

32. Morris, G. M., R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, "Autodock4 and Autodocktools4: Automated Docking with Selective Receptor Flexibility", *Journal of Computational Chemistry*, Vol. 30, No. 16, pp. 2785–2791, 2009.

33. Kurkcuoglu, Z., D. Findik, E. D. Akten and P. Doruker, "How an Inhibitor Bound to Subunit Interface Alters Triosephosphate Isomerase Dynamics", *Biophysical*

Journal, 2015.

34. Case, D. A., T. Darden, T. E. Cheatham III, C. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley *et al.*, "AMBER 9", *University of California, San Francisco*, Vol. 45, 2006.

35. Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee *et al.*, "A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations", *Journal of Computational Chemistry*, Vol. 24, No. 16, pp. 1999–2012, 2003.

36. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, "The Protein Data Bank", *Nucleic Acids Research*, Vol. 28, No. 1, pp. 235–242, 2000.

37. Feig, M., J. Karanicolas and C. L. Brooks, "MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology", *Journal of Molecular Graphics and Modelling*, Vol. 22, No. 5, pp. 377–395, 2004.

38. Espinoza-Fonseca, L. M. and J. G. Trujillo-Ferrara, "Exploring the Possible Binding Sites at the Interface of Triosephosphate Isomerase Dimer as a Potential Target for Anti-Tripanosomal Drug Design", *Bioorganic & Medicinal Chemistry Letters*, Vol. 14, No. 12, pp. 3151–3154, 2004.

39. Espinoza-Fonseca, L. M. and J. G. Trujillo-Ferrara, "Structural Considerations for the Rational Design of Selective Anti-Trypanosomal Agents: The Role of the Aromatic Clusters at the Interface of Triosephosphate Isomerase Dimer", *Biochemical and Biophysical Research Communications*, Vol. 328, No. 4, pp. 922–928, 2005.

40. Eyal, E., L.-W. Yang and I. Bahar, "Anisotropic Network Model: Systematic Evaluation and a New Web Interface", *Bioinformatics*, Vol. 22, No. 21, pp. 2619–2627, 2006.

41. Soner, Z. K., *Elastic Network Model Based Approaches for Conformer Generation and Docking Applications*, Ph.D. Thesis, Boğaziçi University, 2015.