COLLECTIVE DYNAMICS AND CONFORMATIONAL SAMPLING OF TRIOSEPHOSPHATE ISOMERASE

by

Sertan Cansu

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

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

2008

ACKNOWLEDGEMENTS

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

I would also like to express my appreciation to Prof. Türkan Haliloğlu for her contribution to enrichment of my knowledge and perspective and to Assist. Prof. Burak Alakent and Assist. Prof. Demet Akten Akdoğan for their comments and constructive criticisms on my thesis.

I gratefully thank to Bulent Balta for his guidance and advices on my work.

I am much indebted to my colleagues in Polymer Research Center, Ayşegül Özen, Ezgi Karaca, Özge Kürkçüoğlu, Burcu Aykaç Fas, Nevra Özer, Nigar Kantarcı Çarşıbaşı and Arzu Uyar for their helps and kind attitudes. I also want to thank Duygu Başaran for her friendship and support.

I acknowledge the financial support of TUBITAK (BIDEB SSA-2 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

COLLECTIVE DYNAMICS AND CONFORMATIONAL SAMPLING OF TRIOSEPHOSPHATE ISOMERASE

Molecular dynamics simulations (30-60 ns runs) are performed on free/apo triosephosphate isomerase (TIM) to determine any correlation between collective motions and loop 6 dynamics. Native TIM is active only as a homo-dimer even though cooperativity has not been observed between the two identical subunits. Both dimeric and monomeric (isolated from dimer) forms of TIM are simulated in explicit water at 300 K and 1 bar to inspect any differences between the structures in terms of fluctuation dynamics and functionally important loop 6 dynamics/closure. Significant cross-correlations between residue fluctuations are observed in the dimer, which result from the global counter-rotations of the two identical subunits in the essential modes of the dimer. Specifically, the first essential mode contributing to 34% of overall motion of the dimer is strongly coupled to the loop 6's closure over the active site. In contrast, such significant correlations cannot be observed in the monomeric structure, which maintains relatively localized motions of the loops in the essential modes. Thus, the onset of collective motions at ns timescale due to dimerization has functional implications as to the coordination of loop 6 closure. Additionally, a new technique for conformational sampling of proteins is proposed in this thesis, which combines elastic network models and energy minimization procedure. Different conformations generated along the slow harmonic normal modes are reverse-mapped by energy minimization using the generalized Born, implicit solvation model. The applicability and efficiency of this approach is demonstrated on dimeric TIM using the 60 ns MD trajectory.

ÖZET

TRİOZFOSFAT İZOMERAZIN KOLLEKTİF HAREKETLERİ VE KONFORMASYONEL KÜMELENMESİ

Triozfosfat izomeraz (TIM) molekülünün apo formunun kollektif hareketleri ile 'loop 6' bölgesinin hareketi arasındaki korelasyonu (bağlaşım, bağıntı) belirlemek için moleküler dinamik simülasyonları (30-60 ns) gerçekleştirilmiştir. TIM molekülünün özdeş monomerleri arasında eşgüdümlü bir hareket gözlenmese de, doğal yapısının homo-dimer olarak aktif olduğu bilinmektedir. Dimer yapının ve dimer yapısından izole edilen monomerlerin, belirtik çözücü içersinde 300 K sıcaklık, 1 bar basınç altında yapılan simülasyonları sayesinde, monomer ve dimer yapıların dalgalanma hareketleri ve fonksiyonel olarak önemli olan 'loop 6' bölgesi dinamiği/kapanması arasındaki farklılıklar incelenmiştir. Özdeş iki monomerin temel modlarındaki zıt dönme hareketleri sonucunda, dimer yapıdaki rezidü hareketleri arasında önemli karmaşık korelasyonlar gözlenmiştir. Özellikle dimer yapının toplam hareketinin % 34'ünü oluşturan ilk temel mod, 'loop 6' bölgesinin aktif bölge üzerine kapanma hareketi ile büyük ölçüde ilişkilidir. Buna karşılık, monomer yapının simülasyonunda bu tür belirgin korelasyonların yerine lokal hareketler yer almaktadır. Sonuç olarak dimer yapıdan kaynaklanan, ns zaman dilimindeki kollektif hareketler 'loop 6' bölgesinin koordine olarak kapanmasında fonksiyonel olarak önemli rol almaktadır. Buna ek olarak elastik ağ yapı modelleri ile enerji minimizasyonunun birleşmesi ile oluşan, proteinlerin konformasyon kümesinde örnekleme yapan yeni bir teknik, bu tez çalışması içerisinde önerilmiştir. Farklı yönlerde yavaş harmonik normal modlar doğrultusunda yaratılan değişik konformasyonlar, enerji minimizasyon yöntemi ile geri haritalanmıştır. Enerji minimizasyonu sırasında genel Born örtük çözücü modeli kullanılmıştır. Önerilen metodun uygulanabilirliği ve verimliliği 60 ns'lik dimer TIM simulasyonu üzerinde sınanmış ve açıklanmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

3D	Three-dimensional
a_i	Effective Born radii
a	Acceleration of an atom
C(i,j)	Orientational cross-correlations
C^{α}	Alpha-carbon
C^{β}	Beta-carbon
$\mathbf{F}_i$	Force on particle i
H	Hessian matrix
h(x)	Heavy side step function
k	mode
k_i	Force constant that describes the stiffness of the specific bond
k_B	Boltzman constant
l_i	Instantaneous bond length
$l_{i,0}$	Equilibrium length of a bond
M	Number of partial charges in implicit solvent
m_i	Mass of particle *i*
N	Number of Particles
Q	3Nx3N matrix of left singular vectors
q_i, q_j	Charges of atoms
$\mathbf{R}_i$	Position vector of residue i
r_c	Cutoff distance
ri	Interparticle distance
r_{ij}	The distance between partial charges
S	The diagonal matrix of the 3N singular values
T	Temperature
t	Time
u_k	Eigenvector
$\mathbf{v}_i$	Velocity of an atom i
V	Potential energy as a function form
V	The matrix of right singular vectors

w	Instantaneous torsion angle
x_i	The coordinate at which the particle is moving
ΔG^{GB}_{elec}	Generalized Born Solvation energy
$\Delta\mathbf{R}$	Matrix of instantaneous fluctuations of ΔR_i
$\Delta\mathbf{R}_i$	Displacement vector for R_i
$\Delta\mathbf{R}_{ij}$	Fluctuation in the distance vector
Å	Angstrom
α	Alpha helix
β	Beta strand
ε	The solvent dielectric constant
θ_i	Instantaneous angle
$\theta_{i,0}$	Equilibrium angle
γ	Phase factor
γ	Uniform Hookean force constant
σ_{ij}	Lennard-Jones collision parameter
ANM	Anisotropic Network Model
DHAP	Dihydroxyacetone phosphate
EM	Electron Microscopy
EN	Elastic Network
GAP	D-glyceraldehyde 3-phosphate -
GNM	Gaussian Network Model
MD	Molecular Dynamics
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
PGH	Phosphoglycolohydroxamate
RMSD	Root mean square deviation
TIM	Triosephosphate isomerase

1. INTRODUCTION

Biological macromolecules are flexible structures. Their flexibility is important for their specific function such as signal transduction and enzyme activity. In many cases binding of ligands to proteins lead to conformational changes, such as those between the open and close states of an enzyme. (Daniel *et al.*, 2003). Investigation of the relationship between enzyme activity and its conformational states is a challenging task that may be accomplished by combining molecular dynamics (MD) simulations and quantum mechanical calculations. MD simulations are useful in the conformational sampling of a molecule assuming that the MD trajectory is long enough to cover all the conformational space. The problem is the feasibility of running a finite-time trajectory with at least ten thousands of atoms (Daniel *et al.*, 2003). Moreover, an MD trajectory of nanoseconds duration at room temperature is unable to cross high-energy barriers. Hence, efficient techniques need to be devised as alternatives to MD simulations for sampling the conformational space of large molecules.

Although protein dynamics is commonly investigated by MD simulations that incorporate full atomistic detail including explicit solvent molecules, this technique is not computationally feasible for large systems such as supramolecular assemblages. In this respect, coarse-grained normal mode analysis (NMA) has become one of the standard techniques in studying the dynamics of biological macromolecules. Elastic network models, which provide analytical solutions to the equilibrium dynamics of biomolecules (Bahar et al., 1997; Atilgan et al. 2001), are simple, coarse-grained models for normal mode analysis. These models replace detailed inter-atomic potentials with uniform harmonic potentials between interacting atom or residue pairs. They provide high computational efficiency and accuracy in the prediction of functionally important motions of biological macromolecules. (Bahar & Rader, PMC 2006)

In this thesis a new technique, named as reverse-mapping is devised for the conformational space search of large biomolecules by using elastic network models and energy minimization. Alternative conformations are generated by deforming the native or

minimum energy conformation of a protein along the collective coordinates determined by elastic network model and subsequent energy minimization by using implicit solvation. (C Xu, D Tobi and I Bahar, 2003)

In the first part of the thesis, MD simulations are performed to elucidate the structure-function relationship of the enzyme triosephosphate isomerase (TIM), for which numerous experimental and computational studies exist (Daniel *et al.*, 2003; Wang *et al.*, 2001; Benkovic and Hammes-Schiffer, 2003). MD simulations focus on the difference in collective dynamics and catalytic activity between the monomer and dimeric forms of TIM. TIM is an important enzyme in glycolysis, catalyzing the interconvension between dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate (GAP). Its deficiency is a rare, autosomal recessive disorder that causes haemolytic anemia, cardiomyopathy, susceptibility to infections, severe neurological dysfunction, and, dearth in early childhood (Schneider, 2000).

In the second part of the thesis, new conformations of TIM are generated by forcing its native structure along the collective harmonic modes with different deformation values. The deformed structures are reverse-mapped by removing the steric overlaps via energy minimization and the energy and geometry of the new conformers are validated. This methodology, which is developed as an alternative to the MD simulations for conformational sampling of large biomolecules, may provide useful for producing flexible conformations for docking studies.

2. PROTEIN STRUCTURE AND FUNCTION

Proteins are biopolymers formed by the linkage of various combinations of 20 different amino acids. The linear arrangement of the amino acids that make up the protein determines its properties. Each amino acid has a central C^{α} atom to which are attached a hydrogen atom, an amino group ($-NH_2$), a carboxyl group (-COOH) and a side chain (-R). Depending on this side chain, each amino acid has different chemical properties (hydrophobicity, polarity, electrical charge 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

2.1. Protein Structure

The primary structure of the protein is the linear sequence of 20 kinds of amino acids. Chemical and physical properties of amino acids in the sequence determine the local

geometry of the protein -secondary structure- that results from local hydrogen bondings. 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 three dimensional (3D) folded form of the single protein chain, which is the native, or functional form. The interactions of the amino acid side chains determine the secondary and tertiary structure of the protein. Finally, the shapes or structures of large assemblies of proteins specify the quaternary structure.

2.2. Protein Function

The amino acid sequences specify the 3D structures, which in turn determine how the protein functions; i.e. the sequence **-** structure **-** function relationship. The formation of tertiary and quaternary structures can serve as large assemblies such as virus particles or muscle fibers or they can form protein binding sites for their ligands etc. (Introduction to Protein Strc. Carl Branden & John Toozee). Determining the three-dimensional structures, which are encoded in the amino acid sequences, is an important step towards understanding the functions of biomolecules. There are two methods with varying resolutions for determining the 3D structures of proteins, namely X-ray crystallography or NMR spectroscopy. These methods have been developed to obtain 3D models of relatively 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 rigid structures, but they fluctuate around their native states, i.e. the lowest energy conformation. With the recent advance in computational power, computational methods, especially molecular dynamics (MD) simulations, are being used to elucidate structure-function relationship in proteins (Zhiyong et al., 2003). Thus, calculations based on simple physical models could give new and testable insights into this dynamical system's internal motions. In addition to the molecular motion, MD simulations are also used for the refinement of macromolecular structures based on X-ray and NMR data and for sampling the conformational space of proteins. Since, MD simulations provide

detailed picture about dynamics of individual particles in a system as a function of time, one can obtain the microscopic details of protein motion as well as their functional significance far more easily than the experiments (Karplus and Petsko, 1990).

When protein motions in the micro to millisecond time scales are considered, MD technique is limited in applicability due to the huge computational effort involved. In this respect, it has been realized that the functionally important motions can also be determined by normal mode analysis (NMA). Studies performed by coarse grained NMA range from the prediction of biologically important relevant motions of proteins to the identification of conserved dynamic patterns within protein families, as well as the refinement of low-resolution structures. With the improvements in NMA, elastic network (EN) models have been introduced recently, which are computationally simpler methods. The advantage of these methods lies in the details of the force field used. EN models use harmonic springs between interacting atoms or residues instead of detailed atomic potentials used in MD focefields. Once the system is defined as nodes and springs, i.e. the elastic network is established, functionally important collective motions around the equilibrium state can be obtained by performing a coarse-grained normal mode analysis (Atilgan *et al.* 2007; Bahar *et al*. 1997).

2.4. Structure and Function of Triosephosphate Isomerase

Triosephosphate isomerase (TIM) is an important enzyme in glycolysis, which will be the focus of this thesis. TIM catalyzes the interconversion between dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate (GAP). Native TIM is active as a dimer, but no allostery or cooperativity between the identical monomers has been reported (Schnackerz and Gracy, 1991). Each monomer/subunit comprises of 248 residues and adopts a $(\alpha/\beta)_8$ fold, commonly named the TIM barrel. In Figure 2.1(a), the x-ray structure for apo form is shown together with the secondary structures in the $(\alpha/\beta)_8$ fold (each consecutive β strand, loop and α-helix are indicated by the same color, such as β1, loop1/L1 and α1 in red). There are three catalytically important residues (shown in magenta and stick representation in Figure 2.1(b)) located at the periphery of the β-core region of each TIM barrel: Glu165 and His95 participate in proton transfer and Lys13

hydrogen bonds weakly with the bridging oxygen of the ligand (Zhang *et al.*, 1994). The apo/open (Protein Data Bank, PDB code: 8TIM) and bound/closed (PDB code: 1TPH) structures are aligned in Figure 2.1(b). The ligand in the complex form (shown as black spheres) is phosphoglycolohydroxamate (PGH), an intermediate analogue having the same structure and orientation with the substrate DHAP in the crystal structures (Zhang *et al.*, 1994).

In Figure 2.1, X-ray structures of chicken TIM. a) Apo TIM secondary structures are indicated with each consecutive β strand, loop and α-helix shown in same color. b) Free (8TIM, light green) and bound (1TPH, light blue) TIM structures are aligned. Active site residues (Lys13, His95 and Glu165) and the inhibitor PGH are shown in magenta and black, respectively. Main difference between free and bound forms is the open and closed conformations of loop 6 indicated in dark green and dark blue, respectively.

Loop 6 is crucial for TIM's catalytic activity in that it closes over the bound ligand and protects it from solvent exposure during reaction (in Figure 2.1b, dark green: open/apo conformation, dark blue: closed/bound conformation). However, the loop also opens and closes as a natural motion of the enzyme in apo TIM, i.e. it is not ligand-gated (Williams and McDermott, 1995). The highly conserved loop 6 (residues Pro166-Ala176) comprises three hinge residues each at the N- and C-termini and a five residue tip region in between. Mutation of these hinge residues to Gly (excluding Pro166) adversely affects catalysis, resulting in the slowing down of loop motion due to sampling of many more conformations (Xiang *et al.*, 2004; Kempf *et al.*, 2007). Other than the two structures shown in Figure 1(b), there are many other structures of TIM present in the PDB (Berman *et al.*, 2000).

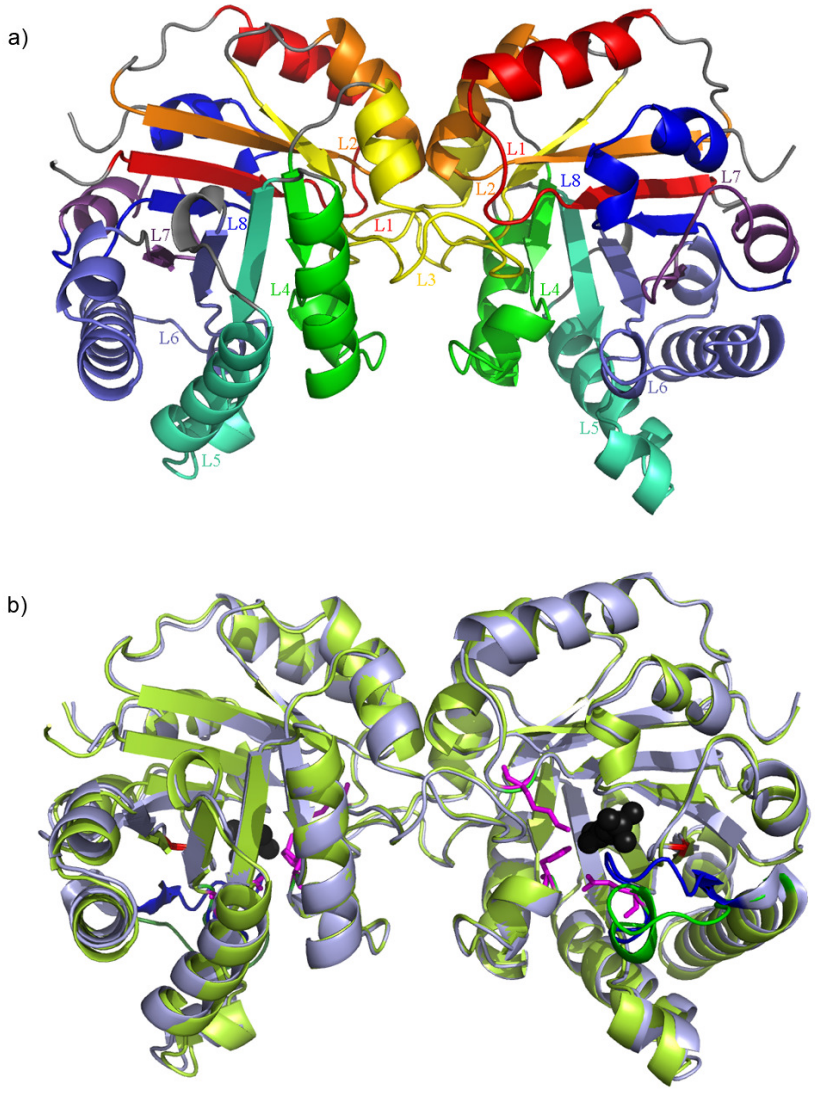

Figure 2.1. X-ray structures of chicken TIM. a) Apo TIM secondary structures b) Free (8TIM) and bound (1TPH) TIM structures superimposed

3. THE MODEL AND THE METHODS

The basic concepts of the methods used in this work, which are molecular dynamics simulations , energy minimization, and the Anisotropic Network Model (ANM) (Atilgan et. al. 2001) are introduced here.

3.1. Molecular Dynamics Simulations

In order to predict the time-dependent events occurring in a molecular system on the atomistic scale, molecular dynamics simulations are widely used. In MD simulations, atoms are allowed to interact with each other using empirical potential energy functions or forcefields, from which the forces on atoms are extracted for a given configuration. Successive configuration of the system is obtained by the integration of Newton's equation of motion, which is:

$$\frac{d^2\mathbf{R_i}}{dt^2} = \frac{\mathbf{F_i}}{m_i} \tag{3.1}$$

In the equation, the motion of a particle with a mass of m_i along the direction of $\mathbf{R}_i$ under the force of $\mathbf{F}_i$ in that direction is described.

3.1.1. Forcefield

Forcefields describe the potential energy of a system as a function of the atomic positions/coordinates. MD simulations are based on an empirical model of interactions within a system involving stretching, bonds and rotation of bonds, as well as non-bonded interactions within a system.

$$V(\mathbf{R}_1,...,\mathbf{R}_N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 +$$

"Bond stretching"

$$\sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 +$$

"Angle bending"

$$\sum_{torsions} \frac{V_n}{2}(1 + \cos(nw - \gamma)) +$$

"Bond rotation (torsion)" (3.2)

$$\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right)$$

"Nonbonded interactions"

In the above equation, $V(\mathbf{R}_1,...,\mathbf{R}_N)$ denotes the potential energy, which is a function of the positions ($\mathbf{R}_i$) of N atoms or particles. The first term in the equation describes the interaction of pairs of bonded atoms, where l_i is the bond length. The second term is similarly the summation over all the angles in the molecule modeled using a harmonic potential, where θ_i is the angle between the three successive atoms. Torsional potential describes the change in energy when a bond rotates, and is depicted with the third term in the equation. The fourth contribution in the equation is for the non-bonded atoms, which are separated by at least three atoms. The non-bonded interactions are defined by two different potentials. The former one is the Lennard-Jones 12-6 potential function that accounts for van der Waals interactions, whereas the latter one is the Coulomb potential for electrostatic interactions. There may be more complicated terms in the force fields other than these basic four components (Leach, 2001).

3.1.2. Energy Minimization

Potential energy of a macromolecular system is a multi-dimensional function of the atomic coordinates; hence, protein fluctuates in a multi-dimensional energy surface. To predict the geometries of the system at the minimum points, minimization algorithms are used. Energy minimization prior to the MD simulation provides a better starting conformation, removes the steric overlaps in the structure and relaxes the bond lengths

and angles. The task is to minimize the energy of the system according to 3N atomistic coordinates, which is not a trivial task. Steepest descent and conjugate gradient are widely used minimization algorithms to solve this nonlinear optimization problem.

Steepest descent method is performed prior to the conjugate gradient method due to its quick convergence ability in finding the minima. And then to determine the exact location of the minimum point, the minimization is switched to conjugate gradient method.

3.1.3. Initialization of the System and Integration

In classical MD simulations, the initial configuration of the system should be introduced by specifying 3N atomistic coordinates ($\mathbf{R}_i$) of the structure. This structure is generally obtained from the experimental data, such as X-ray or NMR structure of a protein. It is meaningful to select a starting conformation that is close to the desired state of the protein, generally minimum energy/native state. Furthermore any high-energy interactions in the system may cause instabilities during the simulation, therefore an energy minimization is performed prior to the simulation.

In order to emphasize boundary effects in the simulation, periodic boundary conditions are used. By the utilization of periodic boundary conditions it is possible to include the solvent effect with a relatively small number of particles. In periodic boundary conditions, particles are enclosed in a solvent box, this box is replicated to infinity by rigid translation in all the three cartesian directions, completely filling the space. The shape of the solvent box may be a truncated octahedron, a cube or a hexagonal prism depending on the shape of the initial structure.

When the initial configuration of the system is minimized in a solvent box, it is required to assign initial velocities at $t=0$. The initial velocities of the system are assigned according to the Maxwell-Boltzmann distribution at the initial temperature. (Leach, 2001)

After the system is initialized, i.e. put in a solvent box and assigned initial velocities, the potential energy of the system can be calculated, and hence the force on each atom from the derivative of potential energy is determined by,

$$\mathbf{F}_i = -\nabla V_i(\mathbf{R}_1,...,\mathbf{R}_N) = -\frac{\partial V(\mathbf{R}_1,...,\mathbf{R}_N)}{\partial \mathbf{R}_i} \tag{3.3}$$

Once the forces on each atom at the current time *t* are calculated, the next step is to produce the new conformation at time $t+\Delta t$ according to the equation:

$$\frac{d^2\mathbf{R}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} \tag{3.4}$$

In order to perform the integration of equation (3.4) there are several numerical algorithms. Verlet algorithm is one of the most commonly used algorithms in MD simulations (Verlet, 1967) Verlet algorithm is based on the addition and subtraction of the Taylor series expansions for the time dependence of the coordinates $\mathbf{R}_i$ at times $t-\Delta t$ and $t+\Delta t$.

$$\begin{aligned}\mathbf{R}_i(t+\Delta t) &= \mathbf{R}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 + \cdots \\ \mathbf{R}_i(t-\Delta t) &= \mathbf{R}_i(t) - \mathbf{v}_i(t)\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 - \cdots\end{aligned} \tag{3.5}$$

Adding these equation togetger and combining with (3.4) produces:

$$\mathbf{R}_i(t-\Delta t) \approx -\mathbf{R}_i(t-\Delta t) + 2\mathbf{R}_i(t) + \frac{\Delta t^2}{m_i}\mathbf{F}_i \tag{3.6}$$

And the difference of the positions at $t-\Delta t$ and $t+\Delta t$ gives the velocity as:

$$v_i(t) \approx \frac{1}{2\Delta t}\left[\mathbf{R}_i(t+\Delta t) - \mathbf{R}_i(t-\Delta t)\right] \tag{3.7}$$

Slightly modified versions of the Verlet algorithm have been proposed to increase the accuracy of the calculation of positions and velocities. Leap-frog algorithm (Hockney, 1970) is a modified version which increases the accuracy in calculations.

3.1.4. Essential Dynamics from MD Trajectories

MD trajectory during the simulation is extracted as the fluctuation trajectory matrix of order *3N* x *n*.

$$\Delta\mathbf{R} = \begin{bmatrix} \Delta\mathbf{R_1}(t_1) & \Delta\mathbf{R_1}(t_2) & \dots & \Delta\mathbf{R_1}(t_n) \\ \Delta\mathbf{R_2}(t_1) & \Delta\mathbf{R_2}(t_2) & \dots & \Delta\mathbf{R_2}(t_n) \\ \Delta\mathbf{R_3}(t_1) & \Delta\mathbf{R_3}(t_2) & \dots & \Delta\mathbf{R_3}(t_n) \\ \vdots & \vdots & \vdots & \vdots \\ \Delta\mathbf{R_N}(t_1) & \Delta\mathbf{R_N}(t_2) & \dots & \Delta\mathbf{R_N}(t_n) \end{bmatrix} \tag{3.8}$$

The *N* number of C^{α} of the protein is described by the fluctuation matrix at successive times t_1, t_2,…,t_n. $\Delta\mathbf{R}_i(t_j)$ is the 3x1 column that represents the change from the average structure in the position vector for the i^{th} C^{α} at the j^{th} timestep.

Using the singular value decomposition (SVD) (Kitao *et. al.*, 1991) , $\Delta\mathbf{R}$ is decomposed into the product of three matrices.

$$\Delta\mathbf{R} = \mathbf{Q}\,\mathbf{S}\,\mathbf{V}^{\mathbf{T}} \tag{3.9}$$

Q is the *3N* x *3N* matrix of left singular vectors $\mathbf{q}_i$, which corresponds to the principal axes (PA) in principal component analysis (PCA). **V** is the *n* x *3N* matrix of right singular vectors, and **S** is the diagonal matrix of the *3N* singular values of s_i, which are also called eigenvalues. The eigenvalues represent the relative contributions of each mode to the overall motion of the protein during the MD simulation.

After decomposition of $\Delta\mathbf{R}$, it is possible to describe the motion by focusing on the essential modes with the highest eigenvalues. (see Doruker *et. al.*, 2000)

3.2. Anisotropic Network Model

NMA has become one of the standard techniques in studying the dynamics of biological macromolecules. The ANM is a simple yet powerful model for normal mode analysis of proteins.

This model, which incorporates the anisotropy of residue fluctuations, is a 3-D extension of the Gausssian network model with isotropic fluctuations in three directions. The potential energy of a structure with N interaction sites is expressed as a Gaussian form (Atilgan *et al.*, 2001)

$$V = (\gamma/2)\, \Delta\mathbf{R}^{T}\, \mathbf{H}\, \Delta\mathbf{R} \tag{3.9}$$

Here $\Delta\mathbf{R}$ is a $3N$-dimensional vector of the fluctuations $\Delta\mathbf{R}_i$ in the position vectors $\mathbf{R}_i$ of all sites ($1 \leq i \leq N$), $\Delta\mathbf{R}^{T}$ being its transpose, and $\mathbf{H}$ is the Hessian matrix. Following Tirion's formulation (Tirion, 1996), a universal force constant $\boldsymbol{\gamma}$ is adopted for all the interactions in the system among close-neighboring bonded and non-bonded pairs. Note that in this definition, the components of the Hessian matrix do not contain $\boldsymbol{\gamma}$. More specifically, all pairs of sites that are closer than a cutoff distance, r_c, are connected by springs, which sum up to the potential energy:

$$V = (\gamma/2) \sum_i \sum_j h(r_c - R_{ij})\, (\Delta\mathbf{R}_j - \Delta\mathbf{R}_i)^2 \tag{3.10}$$

$h(x)$ is the Heaviside step function [$h(x) = 1$ if $x \geq 0$, and zero otherwise], R_{ij} is the distance between sites i and j.

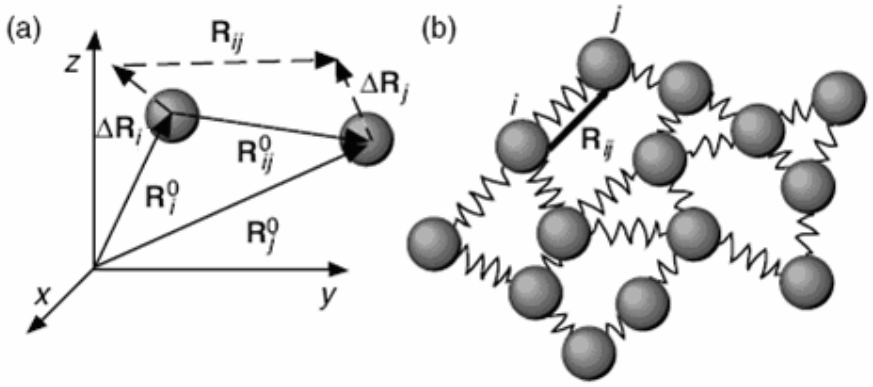

Figure 3.1. Fluctations ΔR_i and ΔR_j in the position vectors of residue sites *i* and *j*.

In Figure 3.1, the equilibrium position vectors with respect to the frame XYZ are denoted as $\mathbf{R}_i^0$ and $\mathbf{R}_j^0$ and their instantaneous values are $\mathbf{R}_i$ and $\mathbf{R}_j$. $\mathbf{R}_{ij}^0$ and $\mathbf{R}_{ij}$ are the equilibrium and instantaneous separation vectors between sites *i* and *j*. The change in the separation with respect to the equilibrium coordinates is $\mathbf{R}_{ij} - \mathbf{R}_{ij}^0 = \Delta\mathbf{R}_j - \Delta\mathbf{R}_i$

If we define $V_S = V/\boldsymbol{\gamma}$, then $\mathbf{H}$ is composed of N^2 submatrices $\mathbf{H}_{ij}$ of the form

$$\mathbf{H}_{ij} = \begin{bmatrix} \partial^2 V_s/\partial X_i \partial X_j & \partial^2 V_s/\partial X_i \partial Y_j & \partial^2 V_s/\partial X_i \partial Z_j \\ \partial^2 V_s/\partial Y_i \partial X_j & \partial^2 V_s/\partial Y_i \partial Y_j & \partial^2 V_s/\partial Y_i \partial Z_j \\ \partial^2 V_s/\partial Z_i \partial X_j & \partial^2 V_s/\partial Z_i \partial Y_j & \partial^2 V_s/\partial Z_i \partial Z_j \end{bmatrix} \tag{3.11}$$

with X_i, Y_i, and Z_i being the components of $\mathbf{R}_i$. Note that :

$$\frac{\partial^2 V_s}{\partial X_i \partial Y_j} = -\frac{\partial^2 V_s}{\partial X_j \partial Y_i} = -\frac{(X_j - X_i)(Y_j - Y_i)}{R_{ij}^2} \text{ for } i \neq j, \text{ and} \tag{3.12}$$

$$\frac{\partial^2 V_s}{\partial X_i \partial Y_i} = \sum_j \frac{(X_j - X_i)(Y_j - Y_i)}{R_{ij}^2} \text{ for } i=j \tag{3.13}$$

where the summation is performed over all neighbors of node *i* within the cutoff distance.

The inverse of the Hessian matrix gives correlations between the fluctuations at sites i and j as

$$< \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j >= \left(\frac{k_B T}{\gamma} \right) tr\left[\mathbf{H}^{-1} \right]_{ij} \tag{3.14}$$

where k_B is the Boltzman constant, T is the absolute temperature and $tr[\mathbf{H}^{-1}]_{ij}$ is the trace of the ij th submatrix $[\mathbf{H}^{-1}]$ of $\mathbf{H}^{-1}$.

By performing modal decomposition of the Hessian matrix, the overall motion can be expressed as a sum over the (3N-6) individual internal fluctuations modes, i.e.,

$$< \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j >= \sum_{i=1}^{3N-6}{}_k \left[\Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \right]_k \tag{3.15}$$

The contribution of the k^{th} mode correlation is :

$$\left[\Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \right]_k = \left(\frac{k_B T}{\gamma} \right) tr\left[\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_\mathbf{k}^\mathbf{T} \right]_{ij} \tag{3.16}$$

where k is the k^{th} nonzero eigenvalue of $\mathbf{H}$ and $\mathbf{u}_k$ is the corresponding eigenvector. Each eigenvalue is related to the frequency of an individual mode, $w_k = (\gamma\lambda_k)^{1/2}$, and the corresponding eigenvector describes how the positions of the N sites of the structure are affected as a result. Thus, the displacement of each site driven by a specific mode can be added to its equilibrium position vector and, as a result, pairs of alternative conformations of the molecule can be viewed under the action of specific modes. (Doruker and Jernigan, 2003)

Despite their simplicity, coarse-grained NMA with ANM is proved in many applications to be a promising tool for describing the collective dynamics of a wide range of biomolecular systems. (Doruker and Jernigan, 2003)

3.3. Reverse Mapping of ENM Conformations

Figure 3.2 represents our new methodology of performing conformational sampling, which can be used as an alternative to MD simulations. New realistic protein conformations are generated using the collective normal modes from the anisotropic network model. Here we employ atomistic ANM, where the positions of the nodes are defined by the coordinates of the heavy atoms in the folded structure of the protein. The node pairs that fall within a specific cutoff distance (10 Å) are connected by identical harmonic springs representative of the bonded and nonbonded interactions between the atom pairs. The overall potential energy of the harmonic system is given by equation 3.10.

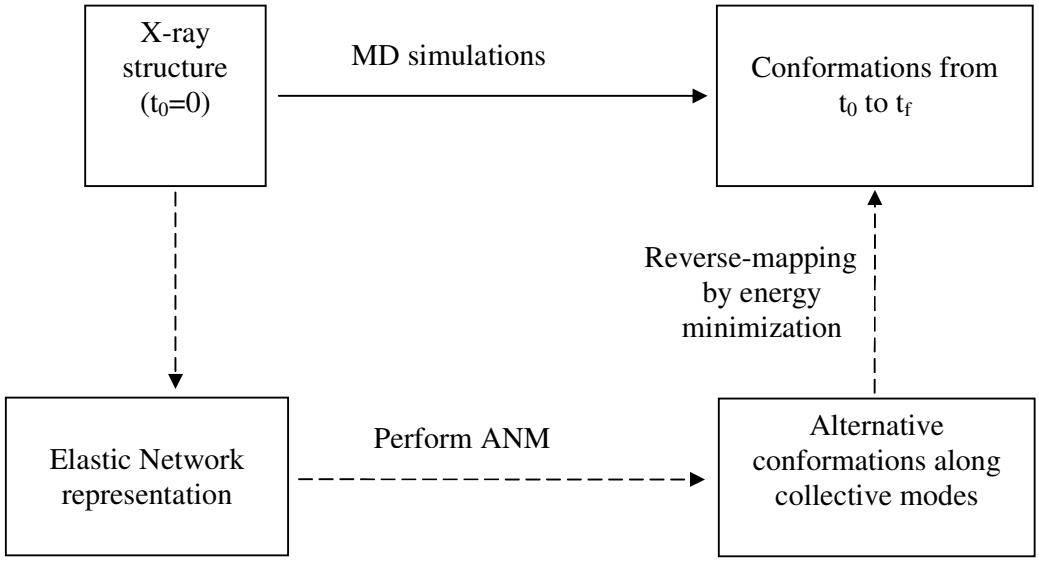

Figure 3.2. Reverse-mapping procedure

Harmonic vibrational analysis of the protein network is performed around an equilibrium/ native structure of the protein. New conformations are generated by deforming the native structure along both positive and negative directions of the low-frequency eigenvectors. At least two different deformation factors (DF) are used along each direction to overcome the energy barriers and produce two corresponding conformations. However, these new conformations have high energies due to close contacts between atoms and distorted local geometries. In order to relax the deformed

structures into realistic conformations, brief energy minimization is performed using the program AMBER (Case *et al.*, 2004; Case *et al.*, 2005) with the ff03 forcefield parameters with implicit solvation model (Duan *et al.*, 2007).

3.4. Implicit Solvent Minimization

The generalized Born (GB) solvation model can be used instead of explicit solvation models. In the GB model, the electrostatic contribution to the free energy of solvation of a solute consists of two terms (equation 3.17). In the equation, the change in coulombic interaction energy on moving a set of M partial charges from vacuum to solvent and the Born solvation self-energy change of the M partial charges are introduced.

$$\Delta G_{elect}^{GB} = -\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{M}\sum_{j=i+1}^{M}\frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{M}\frac{q_i^2}{a_i} \qquad (3.17)$$

where q_i and q_j are atomic charges, r_{ij} is the distance between partial charges i and j, ε is the solvent dielectric constant, and a_i is the effective atomic born radius, which describes the distance from a charge to the dielectric boundary. (Wang and Wade, 2003)

The effective Born radii are the key parameters in the GB model, and it can be computed by a numerical integration procedure (Still et al.,1990) For an efficient computer based model, till now four approximations have been made for the effective Born radius calculation. These are the pairwise descreening approximation introduced by Hawkins et al.(Hawkins *et al.*,1995; Hawkins *et al.*, 1996)., the pairwise approximation with empirical parameters (P1–P5) introduced by Qiu et al. (Qiu *et al.*,1997), the pairwise Gaussian atomic volume approximation introduced by Schaefer and Karplus (Schaefer and Karplus, 1996), and the surface integral formulation introduced by Ghosh et al (Ghosh *et al.*,1998). In this thesis, the AMBER 8.0, with The Hawkins, Cramer, Truhlar (Hawkins *et al.*,1995; Hawkins *et al.*, 1996) pairwise GB model is used.

4. RESULTS AND DISCUSSION

Protein flexibility and dynamics are crucial for enzyme activity. In many cases binding of ligands to enzymes lead to conformational changes, such as those between the open and closed states of proteins. It is still a challenging and fundamental task to uncover the relationship between an enzyme's activity and dynamics at the molecular level because protein motions cover a broad range of length and time scales. (Daniel *et al.*, 2003). In this thesis, the structure-function relationship of the enzyme triosephosphate isomerase (TIM) is addressed, for which numerous experimental and computational studies exist. (Daniel *et al.*, 2003; Wang *et al.*, 2001; Benkovic and Hammes-Schiffer, 2003) via Molecular Dynamics Simulations and a new technique we have developed, called reverse-mapping.

4.1. Molecular Dynamics Simulations of TIM

Other than the two structures shown in Figure 2.1(b), there are many other structures of TIM present in the PDB (Berman *et al.*, 2000). When available x-ray structures are aligned, mainly the conformational flexibility of loop 6 between open and closed states is observed. As a result, molecular dynamics (MD) (Brown and Kollman, 1987; Joseph *et al.*, 1990; Massi *et al.*, 2006) and Langevin dynamics (Derreumaux and Schlick, 1998) simulation studies have so far concentrated on this loop region without taking the conformational flexibility of the whole protein into account. Recently, dominant collective motions of TIM- possibly being coupled to loop opening/closure- have been revealed by elastic network model analysis of the intact dimeric structure (Kurkcuoglu *et al.*, 2006).

In current thesis, for the first time –at least to our knowledge- classical MD simulations of appreciable length (30-60 ns) are performed on apo TIM in explicit solvent without constraining any parts of the protein. The aim is to extract collective motions and to investigate their correlation with active site loop dynamics. In addition, MD simulations on monoTIM (isolated monomers from dimeric TIM) are carried out in order to observe any changes in equilibrium and dynamic properties of the monomer, which could explain why native TIM is active as a dimer.

4.1.1. Simulation Details

Molecular dynamics simulations were carried out using the program AMBER (Case *et al.*, 2004; Case *et al.*, 2005) with the ff03 forcefield parameters (Duan *et al.*, 2003). Three independent runs were performed for apo TIM at 300 K: (i) 60 ns run of the dimer structure (subunits A and B), (ii) 30 ns run of the A-subunit, and (iii) 60 ns run of the B-subunit. The apo structure of chicken TIM (PDB code: 8TIM) with 2.58 Å resolution was used as the starting conformation for all MD simulations. The simulation details are summarized in Table 4.1 and the procedure is as follows.

Table 4.1: Simulation system details

Run	Simulation lengths (ns)	Box dimensions (Å)	Number of atoms	No of H_2O molecules	Number of Cl- ions
Dimer	60	107	57,532	16,674	8
Monomer A	30	77	23,066	6,437	4
Monomer B	60	77	23,066	6,437	4

A periodic truncated octahedron box was used for solvation of the protein in explicit TIP3P water molecules (Jorgensen et al., 1983). Energy minimization was performed using 50 cycles of steepest descent algorithm, followed by conjugate gradient until the RMS gradient per atom reaches 0.01 kcal/mol/Å. Each trajectory started with velocity assignments according to the Boltzmann distribution at either 10 or 20 K and the temperature was gradually raised to 300 K. NPT simulations were performed at 300 K and 1 bar using the weak coupling algorithm for both pressure and temperature (Berendsen *et al.*, 1984). Constant pressure periodic boundary conditions are used with isotropic position scaling. A time step of 2 fs was used by the implementation of SHAKE algorithm (Ryckaert *et al.*, 1977) for the bonds involving hydrogens. The Ewald summation technique with the particle-mesh method (Essman *et al.*, 1995) was used to calculate long-range electrostatic interactions with a cutoff distance of 9 Å.

The essential dynamics of TIM (Amadei et al., 1993) is extracted from the MD simulation trajectories by using singular value decomposition of the fluctuation matrix. The

fluctuation matrix is formed based on the C^{α} atom coordinates of the residues after removal of translational and rotational motions of the protein by superimposing each snapshot onto the initial conformation.

4.1.2. Fluctation Dynamics

The root-mean square deviation (RMSD) of each snapshot from the initial, energy-minimized structure is calculated after alignment based on C^{α} atoms (Figure 4.1). RMSD values fall within reasonable limits throughout the runs, i.e. less than 2-2.5 Å in monomer and dimer runs. These values denote reliable conformations, which have not deviated extensively from the starting conformation.

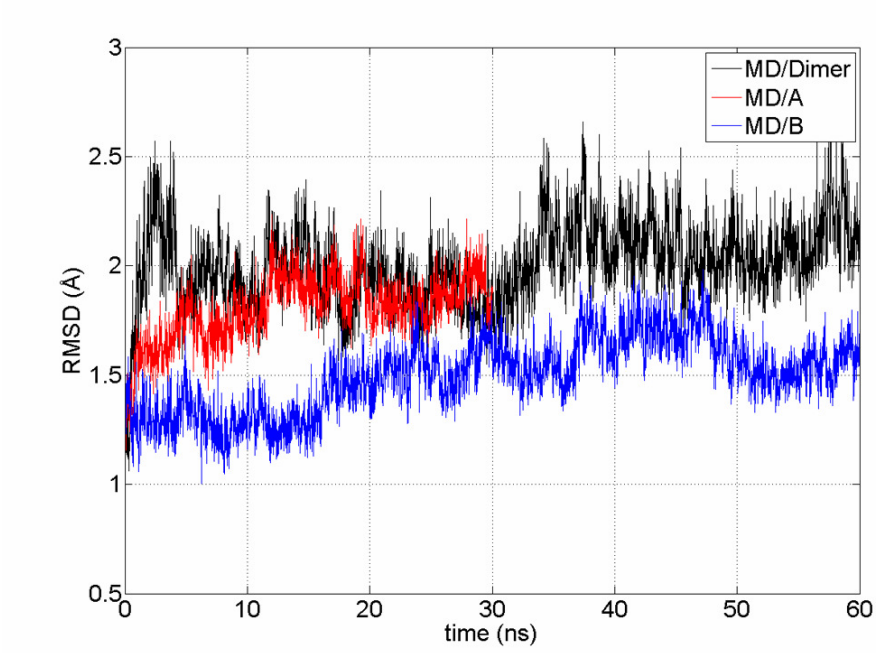

Figure 4.1. RMSD of the snapshots from the initial energy-minimized structure for dimer and monomer runs.

In Figure 4.2(a), the mean square fluctuations (MSF) of C^{α} atoms about their average positions are plotted as function of residue index for the dimer run together with the x-ray data for apo structure (8TIM). The residue fluctuations are averaged over the identical A and B monomers in all cases. In general there is good agreement between simulation and

experiment. The loops (particularly loops L5 and L6) and helix α1 exhibit higher fluctuations in solvent environment, i.e. in the absence of crystal contacts.

Figure 4.2 (b) is a comparison of dimer and monomer simulations, again results are averages over A and B subunits.

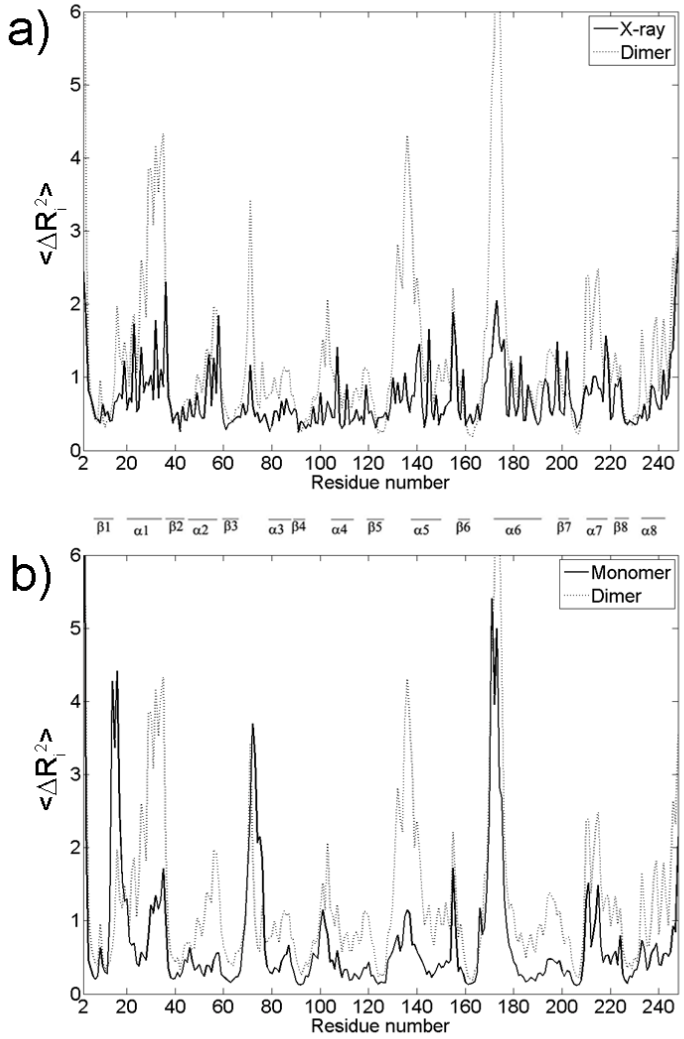

Figure 4.2. Mean square fluctuation ($Å^2$) about the average position plotted for C^{α} atoms of residues

Loop 1, which lies at the interface of the dimer and contains the catalytic residue Lys13, becomes highly mobile in isolated monomer simulations. This should have an adverse effect on catalysis, which has been expected in previous protein engineering

studies on monomeric TIM and has lead to the redesign of loop 1 (Thanki *et al.*, 1997). There is a slight broadening of the mobility of loop 3, also forming part of the interface. Deletion of loop 3 has produced a stable monomeric TIM variant with reduced catalytic activity (Borchert *et al.*, 1994). Other than these changes at the no-longer existing interface of the monomer, higher fluctuations are observed predominantly in dimeric structure- α1 and α2, and loop 5 being the most pronounced regions. Almost all helices and loops (excluding L1-L3) show higher fluctuations in dimeric form, the reasons of which will be discussed in upcoming sections. The β-strands forming the core of each TIM barrel fold are at the minima, consistently in experiments and all simulations.

4.1.3. Loop Closure and Dynamics

Three snapshots from dimer (showing the loop of subunit A) and monomer B simulations are displayed in Figure 4.3 (a) and (b), respectively. All the snapshots (green, orange and blue) are aligned with the bound/closed x-ray structure (1TPH), which is also shown in gray, in panels (a) and (b). Loop 6 is indicated by darker colors (closed x-ray conformation for loop 6 in black). The snapshots are chosen such that loop 6 can be observed in different conformations- open, intermediate and nearly closed. Green/ orange/ blue structures represent the snapshots at 7.3/ 43.3/ 34.6 ns for the dimer and 2.5/ 7.6/ 20.9 ns for the monomer B runs. For example, the tip of loop 6 moves more than 7Å during loop closure in the dimer (panel a). In Figure 4.3 panels (c) and (d), four other snapshots are displayed focusing on the loop region from the alternative subunits for dimer and monomer runs, respectively. Specifically, blue/ gray/ pink/ green structures represent the snapshots at 18.6/ 16.4/12.3/26.4 ns for the dimer run (B-chain) and 20.0/ 4.6/ 25.6/10.8 ns for the monomer A run. The differences between the dimer and monomer runs can be summarized as follows: Global deformations of the whole protein structure are observed predominantly in the dimer structure. The loop motion seems less targeted between open and closed conformations in the monomer, based on all snapshots sampled (only seven shown in the figure).

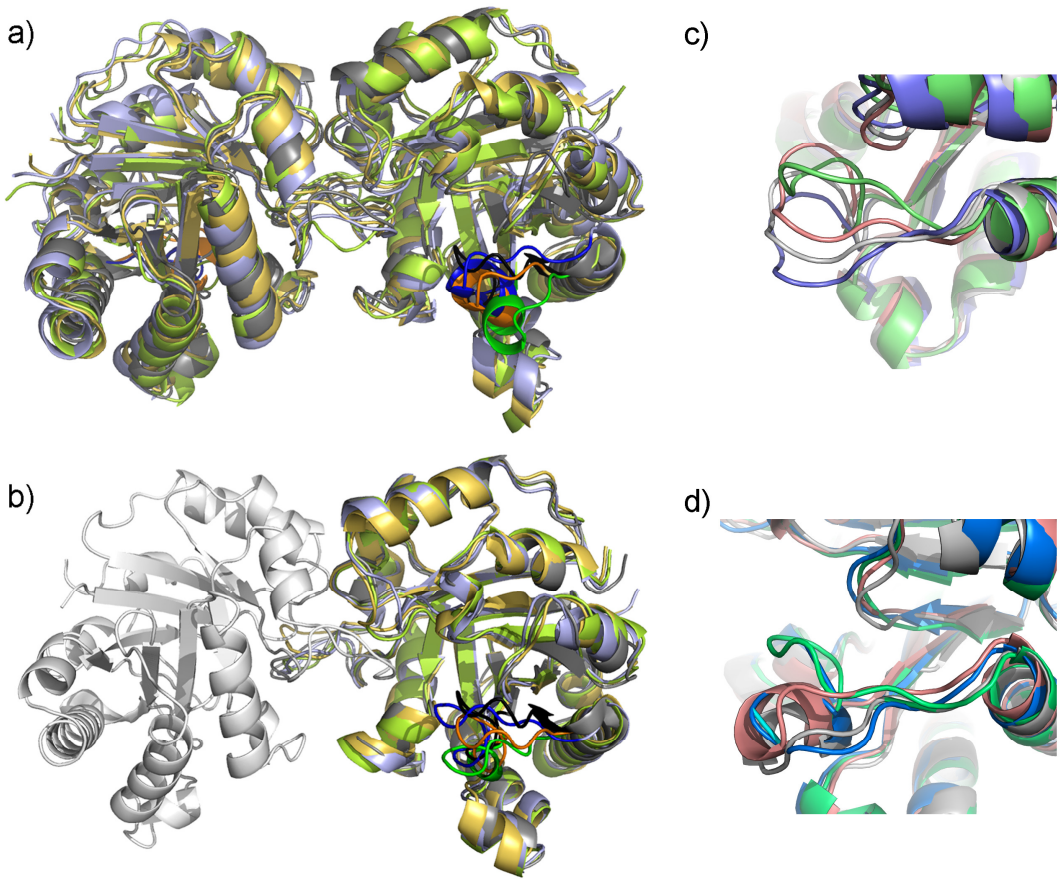

Figure 4.3. Several snapshots obtained from simulations aligned on the x-ray structure (1TPH-bound TIM).

To assess the extent of loop closure on a quantitative basis, same as in the work of Massi *et al.* (Massi *et al.*, 2006), the distances between three residues located at the tip of loop 6 (Ile170, Gly171 and Thr172) and Tyr208 on loop 7(red residue in Figure 2.1 b) are calculated. The reason for the choice of Thy208 is that it stays quite stationary during the runs (see Figure 4.2).

In Figure 4.4, these inter-residue distances based on C^{α} atoms of loop 6 tip residues (Ile170, Gly171, and Thr172) and Tyr208, which is on loop 7 and displays low mobility in the simulations, are plotted as a function of time. The dashed and solid horizontal lines represent the distances for the same residue pairs in open (8TIM) and closed (1TPH) x-ray structures, respectively. In the dynamics of dimer structure (8TIM), snapshots with

distances comparable to the closed form are encountered much more frequently than in monomer simulations (specifically looking at black and red lines for residues 170 and 171). In both single chain simulations, the loop 6 region tends to remain primarily in open form (dashed lines).

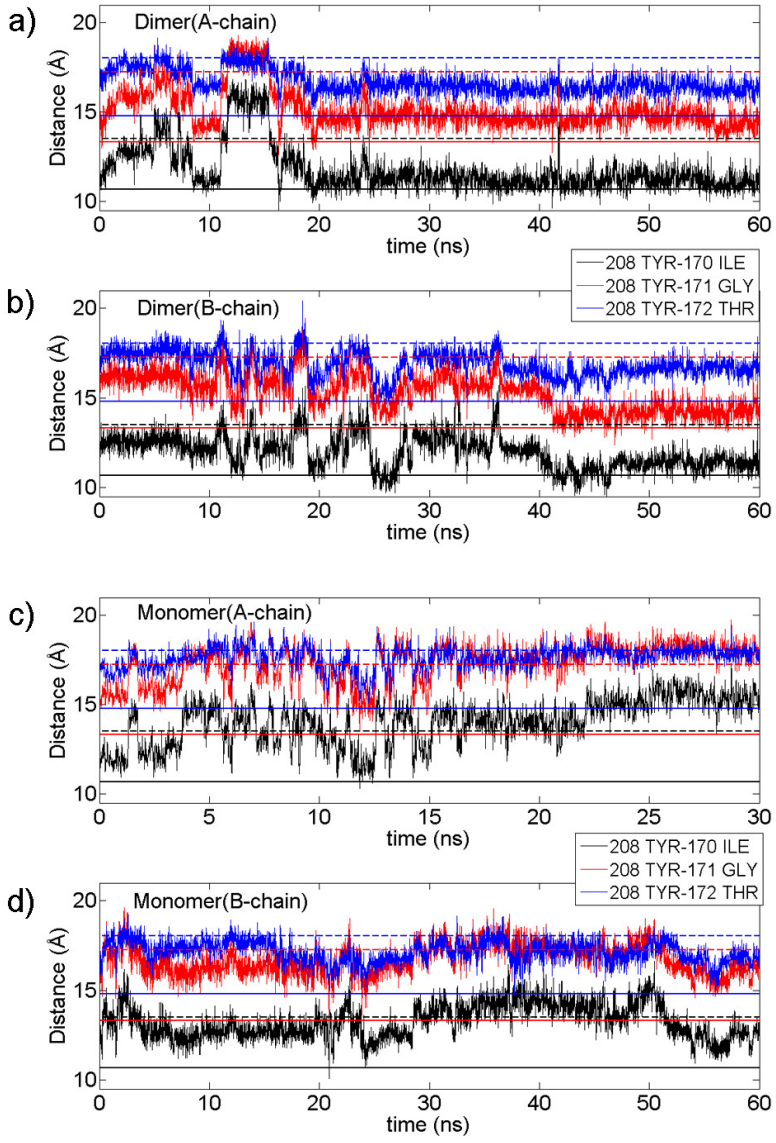

Figure 4.4. Loop opening and closing is observed by plotting the distance between the C^{α} atoms of loop 6 tip residues (Ile170, Gly171, and Thr172) and Tyr208

In free TIM, the catalytic residue Glu165 makes hydrogen bonding with Ser96. Upon PGH binding (Zhang *et al.*, 1994), Glu165 moves 2-3 Å towards the substrate analog, assuming the catalytically important conformation. Analysis of our trajectories indicate that the carboxylate oxygens of Glu165 are making hydrogen bonds with either the backbone amide nitrogen and the side-chain hydroxyl group of Ser96 during a significant portion of each run. Specifically, the percentage of snapshots, in which the carboxylate oxygens are hydrogen bonded to Ser96, are 92.7% and 87.9% for dimer and monomer runs, respectively (averaged over A and B chains and the two oxygens in each case). Of course, complete closure of the loop and movement of Glu165 towards the ligand observed in the reference structure (1TPH) necessitates the presence of a ligand, which is not present in our apo structures. Moreover, the timescale of loop 6 motion has been found experimentally to be on the order of microseconds (Williams and McDermott, 1995; Kempf et al., 2007; Massi et all., 2006; Rozovsky and McDermott, 2001; Desamero et al., 2003). As a result, we cannot expect to see complete closure and reopening during our runs at the ns scale.

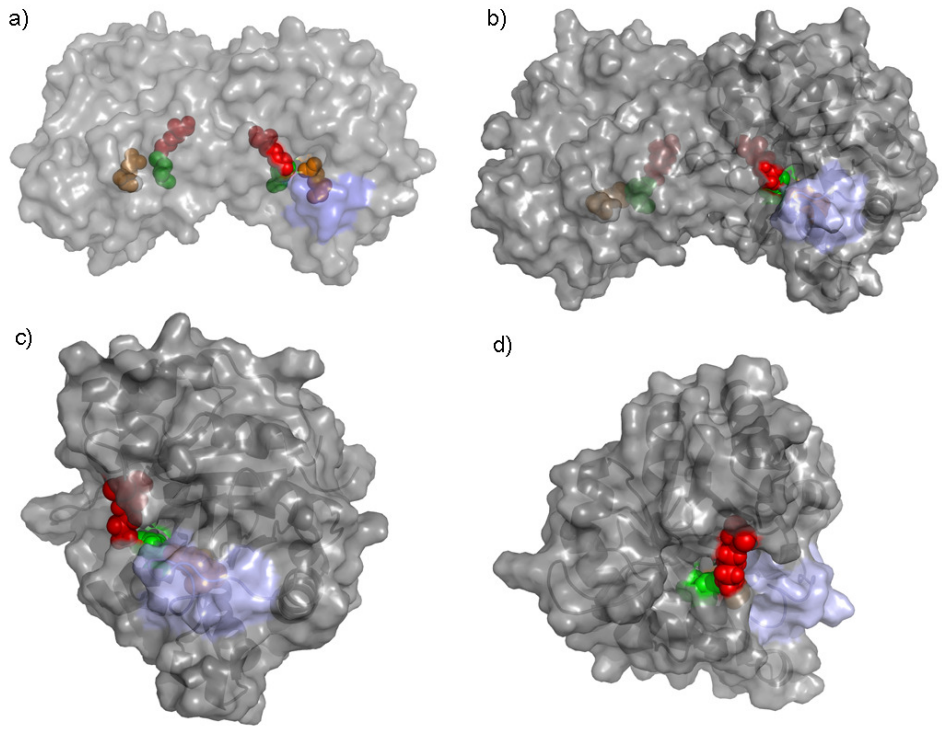

Figure 4.5. Surface representations of x-ray structure (1TPH), and relatively closed snapshots from dimer and monomer simulations

In Figure 4.5, the surface representations of the closed x-ray structure (1TPH, panel a) is compared with one of the relatively closed snapshots chosen from the dimer (at 34.6 ns, panel b) and monomer B (at 20.9 ns, panels c and d showing two different views of the same snapshot) runs. Loop 6 residues are colored in blue, and active site residues Lys13, His95 and Glu165 are shown in red, green and orange, respectively. The solvent exposure of the active site region in dimer snapshot (panel b) resembles that of the x-ray structure, where the ligand PGH has been removed for comparison of the surfaces. In contrast, in the monomer snapshot Lys13 and His95 are more solvent exposed at the no longer existing interface region (panel d).

Another issue is whether loop 6 dynamics resembles that of a rigid lid. This could be determined by looking at the pseudo-dihedral angles of the loop region, i.e. the internal degrees of freedom of the chain backbone. The i^{th} pseudo-dihedral angle (ϕ_i) connecting residues i and i+1 is calculated by using the C^{α} coordinates of residues i-1, i, i+1 and i+2. Comparison of apo and bound crystal structures has indicated that the loop acts as a rigid body with two hinges, namely N-terminal hinge located at ϕ_i = 166, 167, and C-terminal hinge located at ϕ_i = 174, 175, 176 (Joseph *et al.*, 1990). Recent stochastic boundary MD simulations have shown that RMS fluctuations of ϕ_i increase towards the C-terminus of the loop (Massi *et al.*, 2006). Simulation results given in Figure 4.6 also confirm that the loop flexibility increases towards the C-terminus on the ns time-scale, both for mono- and dimeric TIM with the maximum flexibility observed at the C-terminus hinge. Both curves give RMS fluctuations of ϕ_i averaged over chains A and B, with the monomer and dimer runs showing a similar trend. The increasing trend of fluctuations from the N-terminal to the C-terminal hinge is also consistent with the ns-dynamics obtained by solution NMR study (Kempf *et al.*, 2007).

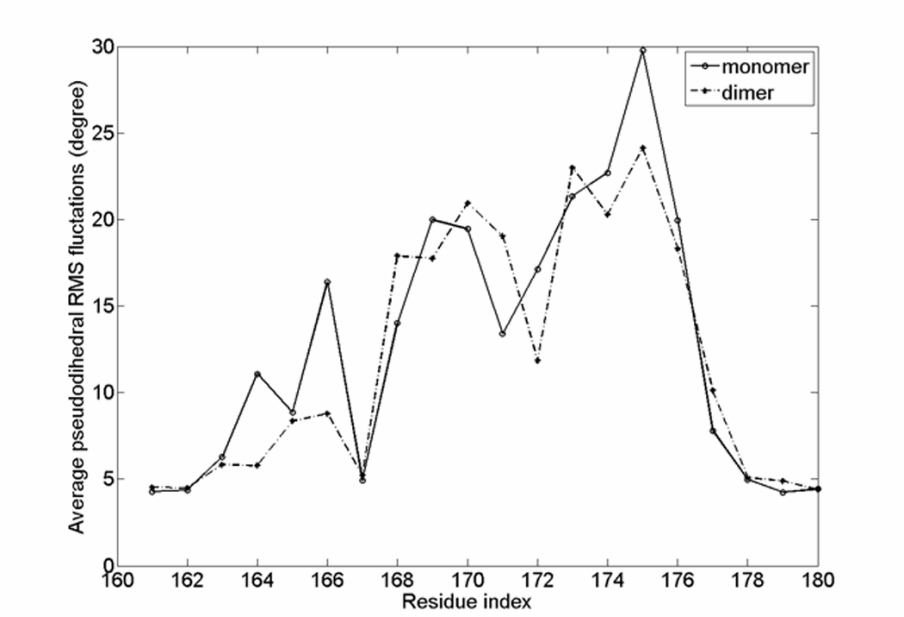

Figure 4.6. RMS fluctuations of pseudo-dihedral angles as a function of residue number for the region encompassing loop 6

4.1.4. Collective Motions Emerging in the Dimeric Structure

Figure 4.7 gives the normalized orientational cross-correlations *C(i,j)* between residue fluctuations defined as

$$C(i,j) = \frac{\left\langle \Delta \mathbf{R}_i . \Delta \mathbf{R}_j \right\rangle}{\left[\left\langle \Delta \mathbf{R}_i . \Delta \mathbf{R}_i \right\rangle \left\langle \Delta \mathbf{R}_j . \Delta \mathbf{R}_j \right\rangle \right]^{1/2}} \tag{4.1}$$

where $\Delta \mathbf{R}_i$ is the fluctuation in the position vector $\mathbf{R}_i$ of site *i*. Strong positive and negative correlations are observed in the (Figure 4.7a) dimer compared to relatively localized correlations in (Figure 4.7b) monomer A and (Figure 4.7c) monomer B. The brackets represent time averages over recorded snapshots. The cross-correlations vary in the range [-1, 1] with the lower and upper limits indicating fully anti-correlated and correlated fluctuations in terms of orientation, respectively. *C(i,j)* = 0 gives uncorrelated fluctuations in terms of orientation. In the dimer simulation (Figure 4.7a) there are strong positive and negative correlations, both within each subunit (intra-subunit, squares on the diagonal) and

between different subunits (inter-subunit, off-diagonal). These distinct regions with high +/- correlations are indicative of collective domain motions that dominate the overall dynamics of the dimer. In contrast, such strong correlations can not be observed in single monomer A and B (Figure 4.7 b and c) simulations. We should state that the level and pattern of correlations that emerge from TIM dimer simulation is very significant compared to previous MD results of ns duration, such as on EcoRI-DNA complex (Doruker *et al.*, 2006).

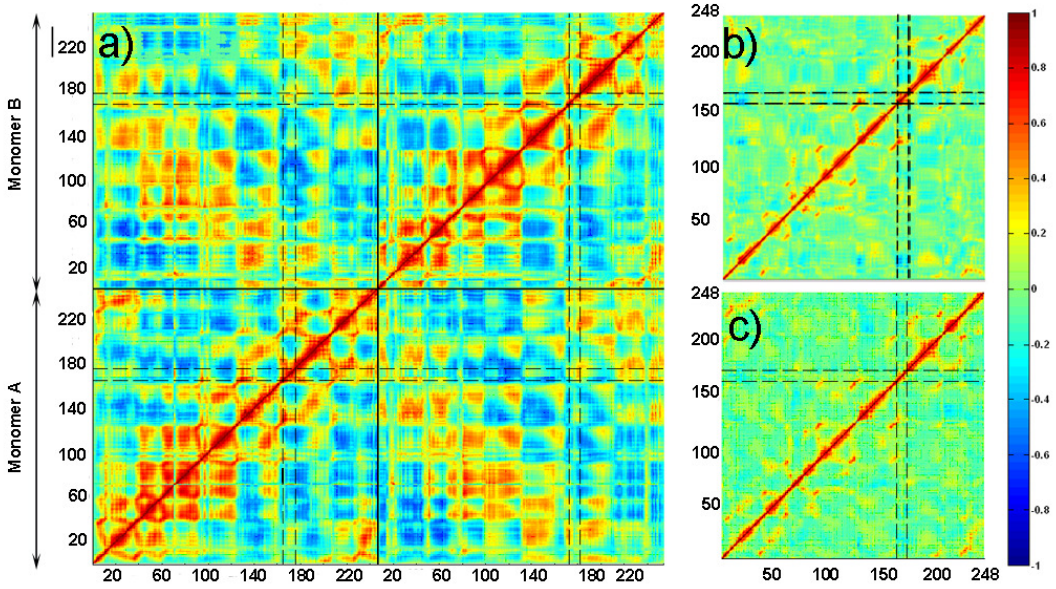

Figure 4.7. Normalized cross-correlations between residue fluctuations (eq. 4.1) including all modes

Based on symmetry considerations, it is expected that dimerization should give rise to rigid body motions of the two monomers, which may, in turn, be crucial for enzymatic activity. To get a better insight about the cooperative dynamics in the structure, we will look at the cross-correlations based on the first five modes extracted by singular value decomposition. The five modes that form the essential subspace (Amadei *et al.*, 1993) constitute 58%, 39% and 42% of the overall motion in dimer, monomer A and monomer B runs, respectively.

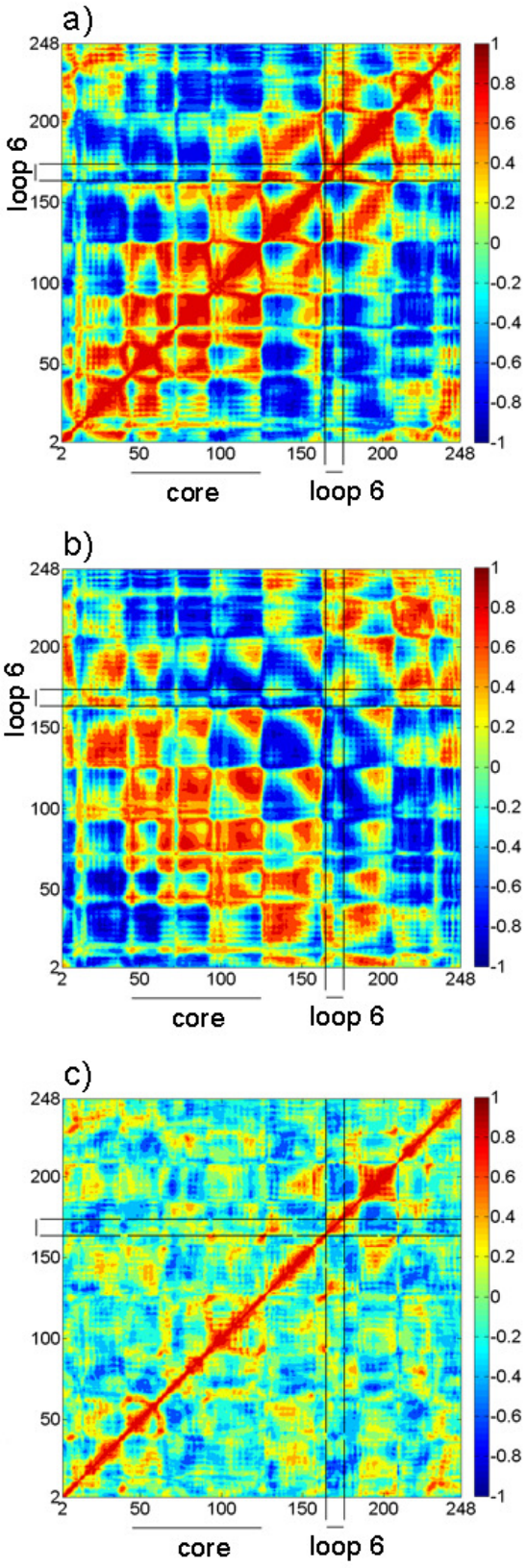

Figure 4.8. Normalized cross-correlations between residue fluctuations in the first five essential modes (averages over A and B chains)

Figure 4.8 gives the cross-correlations in the essential subspace of these runs, specifically (a) intra-subunit (dimer), (b) inter-subunit (dimer) and (c) monomer (averaged over A and B). Since the correlation patterns are quite similar for A and B chains in the dimer, the cross-correlations have been averaged over the two chains in the dimer. The strongly correlated (red) and anti-correlated (blue) regions resulting from the collective motions are clearly observed. Even though some negatively correlated and few positively correlated regions appear in monomer simulations, they are not as noteworthy as in the dimeric structure. Clearly there is a core region/domain that runs across the interface of the dimer. This core region is made up of positively correlated residues ~ 45-125 from both chains A and B, i.e. including intra-subunit and inter-subunit interactions. This region comprises the residues starting from $\alpha 2$ running to the end of $\beta 5$ together with loops 3 and 4.

Figure 4.9 shows the first mode deformations in dimer and monomer B runs. A global twisting motion of the two monomers in opposite directions can be observed in panel a. There is a separate rotation axis for each monomer that passes through the core of each TIM barrel. The two axes are not aligned due to accompanying bending motion of the monomers. This collective motion in dimer assists loop closure (red). In contrast, the first mode of the monomer describes independent - not globally coordinated- deformations of the loops (panel b). Two alternative conformations (green and yellow) from the dimer simulation given in panels c and d depict the loop motion resulting from first mode. Partial loop closure is observed in the dimer (the tip residue Thr172 moves about 5.6 Å, Figure 4.9c), whereas the loop motion (red and blue), is along the lateral direction in the monomer (tip movement of 2.0 Å, Figure 4.9d), i.e. not along loop closure. In other essential modes, such as second, third and fourth modes (not shown), similar counter-rotation-type motion with different axes is observed in the dimer. These collective motions are also consistent with previous elastic network results (Kurkcuoglu *et al.*, 2006).

At this point a comment will be made about the recent work by Van Wynsberghe and Cui (Van Wynsberghe and Cui, 2006), where they have stated that a large number of low-frequency normal modes are necessary for identification of correlated motions using residue cross-correlations. In contrast to the case of harmonic normal modes, a small number of anharmonic modes extracted from MD simulations can describe a very high percentage of the protein motion. As stated above, the first five modes describe ~ 60% of

the overall motion in the dimer and ~ 40% in the monomer runs. However, if a coarse-grained normal mode analysis of TIM is performed as in previous work (Kurkcuoglu *et al.*, 2006), ~50 (~140) modes are needed to describe at least 30% (40%) of the overall motion. In general, it is expected that correlations should get weaker as more modes are included; still the correlation plots based on all modes (Figure 4.7) and on five modes (Figure 4.8) exhibit very similar features. Thus, it seems justified to focus on the first five essential modes in this thesis.

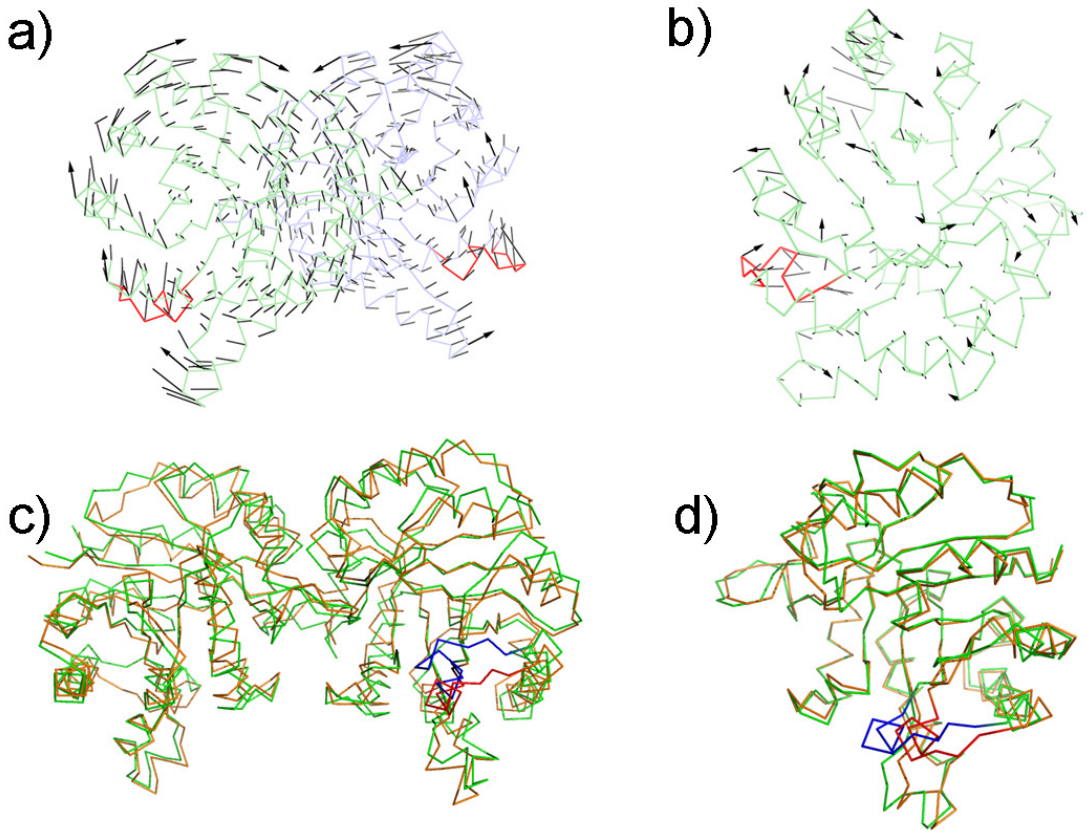

Figure 4.9. Deformations in the first essential mode from different point of views

4.1.5. Correlation between Collective Motions and Loop Dynamics

For each essential (or principal) mode, the overlap values are calculated by taking the dot product of eigenvectors for the tip region (169A-173G) of loop 6 with the deformation vector that is in the direction of loop closure. The deformation vector is calculated based

on aligned open (8TIM) and closed (1TPH) x-ray structures, again for the specified tip residues of loop 6. In Table 4.2, the overlap values are tabulated for the first five principal modes, together with the contribution of each mode to the overall motion. In single chain dynamics (monomer A or B) overlap values are lower than the ones in dimeric structure. Overlap values close to one are indicative of similar conformational changes to those observed in crystal structures and values above 0.6 are highlighted in the table. Specifically, the first mode of dimer contributing to 34% of overall motion exhibits high overlap values (A chain: 0.86, B chain: 0.67). Thus, first mode describes strong coupling between the collective motions of TIM and loop opening/closing. In contrast, high values are observed for modes 2,3 and 5 only in monomer B, but the contribution of these modes to the overall motion is comparatively lower- each being less than 10%.

Table 4.2. Overlap of the loop's motion with the loop closure direction in the essential modes.

Modes	Monomer A		Monomer B		Dimer		
	Overlap (A)	Mode contribution	Overlap (B)	Mode contribution	Overlap (A-chain)	Overlap (B-chain)	Mode contribution
1	0.22	0.13	0.07	0.17	**0.86**	**0.67**	0.34
2	0.08	0.09	**0.60**	0.09	0.35	0.20	0.09
3	0.24	0.07	**0.63**	0.08	**0.62**	0.14	0.08
4	0.31	0.06	0.30	0.05	**0.80**	**0.82**	0.04
5	0.25	0.04	**0.71**	0.04	**0.86**	0.34	0.03

Figure 4.10 gives the normalized probability of snapshots along the open/closed transition coordinate based on the distance between Tyr208 and Gly171 (the tip of loop 6) for loop 6. Calculations are based on the distance between the tip of loop 6 (Gly171) and Tyr208 (same as the red lines in Figure 4.4). The corresponding distances in open (8TIM) and closed (1TPH) x-ray structures are 17.3 and 13.3 Å, respectively. First (a, b, c) and second (d, e, f) row panels refer to the loop conformations in subunit A and B for the dimer run, respectively. For comparison, third (g, h, i) and fourth (j, k, l) rows belong to monomer runs A and B. Panels in the first (a,d,g,j), second (b,e,h,k) and third (c,f,i,l) columns refer to the first mode, the cumulative effect of first five modes and the overall motion, respectively. In the first mode, the loops can make significant progress towards

closure only in the dimeric structure. Similar trend is observed based on the cumulative action of first five modes (second column). For comparison, the progress towards closure is presented for the overall motion- all modes- as extracted from Figure 4.4. In summary, the loop closure is to a great extent facilitated by the collective motions in dimeric TIM. This could be one of the reasons- overlooked so far- why native TIM is catalytically active as a dimer.

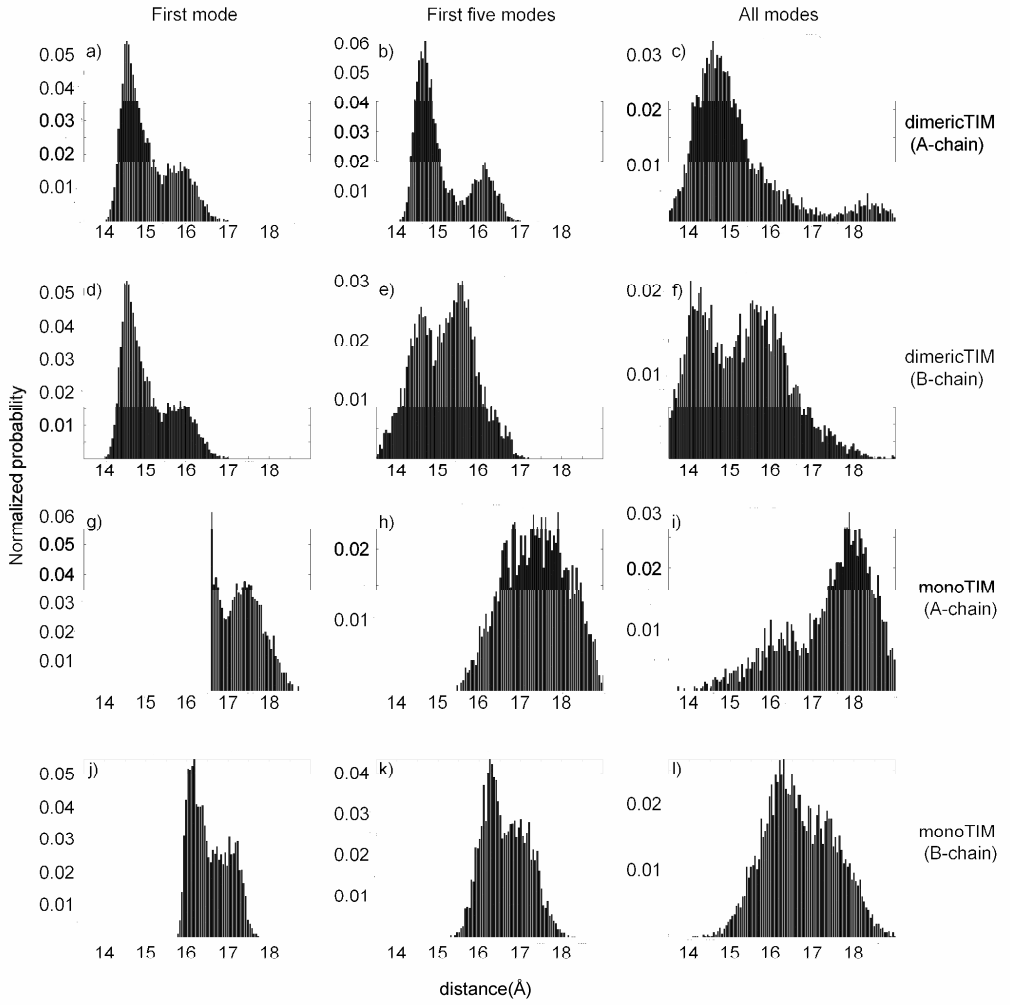

Figure 4.10. Normalized probability distribution of the open/closed transition coordinate based on the distance between Tyr208 and Gly171

Finally, loop motion in the different subunits is positively correlated in the essential modes, which is observed in the inter-subunit orientational correlation plot for the dimer (Figure 4.8b). Normalized cross-correlations, *C(i,j),*for the loop residues from different subunits are listed in Table 4.3 for the first mode and the cumulative first five modes.

Specifically, loop residues, excluding the N-terminal hinge residues and 171, exhibit high positive correlations in the cumulative first five modes.

Table 4.3. Inter-subunit cross-correlation values for the loop residues in the first mode and for the first five modes

Residues from chains A and B	Cross Correlations	
	First Mode	First five modes
166	-0.2357	0.1768
167	-0.3677	-0.2058
168	-0.2524	-0.0322
169	0.1807	0.6488
170	0.4531	0.8866
171	-0.1791	-0.2938
172	0.2893	0.7683
173	0.3902	0.6496
174	0.5528	0.8451
175	0.5025	0.9482
176	0.4122	0.802

4.2. Conformational Sampling of TIM by Reverse-mapping

In this part of the thesis, reverse-mapping of TIM will be performed. The conformers generated along the slow modes will be analyzed in detail to observe the accuracy of the method. Geometric consistency of the conformers are checked via a web server called MolProbity (Simon *et al.*, 2003; Davis *et al.*, 2004) and for validation further calculations are also performed which are reported in the following sections.

4.2.1. Properties of Reverse-mapped Conformers

Table 4.4 gives the potential energies of the new structures before and after implicit minimization procedure, which allows a direct comparison of energies for different conformers. Low and high deformation factors (DF) are used for each mode to find distinct conformers. X-ray is the initial crystal structure minimized with AMBER forcefield ff03. L and H refer to low and high deformation factors, respectively. Table 4.4 also gives the RMSD values calculated with respect to native structure (X-ray) before and after minimization. The initial energies, of the structures deformed along the harmonic modes are quite high due to steric overlaps. These overlaps are removed by employing energy minimization and reasonable values are obtained after minimization compared to the X-ray structure's energy. For computational efficiency, the minimization criterion- RMS gradient per atom- is selected as 0.1 kcal/mol/Å. The average number of steps required to attain the desired criterion is 1127 requiring about 10 mins of CPU time on eight processors each with 1.5 GHz/6MB L3 Cache Intel ® Itanium2 processor.

Ramachandran plots are used to visualize the dihedral angles of all residues. The reason for plotting these maps is to validate the new conformers geometrically, because C^{α} location is very important for determining any distortion in geometry as it connects the side chain with the backbone (Simon *et al.*, 2003). The dihedral angles of amino acid residues in reverse-mapped structures seem to be reasonable compared to the native structure's in their Ramachandran plots (Figure 4.11).

A more strict validation of the new conformers is performed using MolProbity web server. (Simon *et al.*, 2003; Davis *et al.*, 2004) It is a general-purpose web server that provides detailed analysis of all atom contacts and any steric problems within the molecule. The models originally used in the construction of this server come from high-resolution X-ray crystallography and NMR, but even these data sometimes is not perfect and may have some local errors due to misinterpretation of complicated experimental data (Davis *et al.*, 2007). In the analysis with MolProbity reviewed in Table 4.5, though there seems to be some rotamer and Ramachandran outliers, the overall structures are scored as 100^{th} percentile, which means they are the best among structures of comparable resolution. Besides, even the X-ray structure of TIM with 2.5 Å resolution gives similar outliers after

energy minimization. Hence, these outliers probably stem from the original dataset used in the determination of allowable region's borders.

Table 4.4. Energy and RMSD of new structures before and after minimization.

	Energies (kcal/mol)		**RMSD (Å)**	
Conformers	**Initial**	**Final**	**Initial**	**Final**
Xray	5,603	-12,771	0.0	0.0
L-1a	10,333	-12,834	0.8	0.9
L-1b	12,506	-12,832	0.8	0.9
L-2a	14,816	-12,807	0.8	0.8
L-2b	9,445	-12,793	0.8	0.8
L-3a	9,613	-12,799	0.7	0.8
L-3b	14,496	-12,821	0.7	0.8
H-1a	17,718	-12,772	1.2	1.3
H-1b	20,569	-12,798	1.2	1.3
H-2a	25,389	-12,806	1.2	1.2
H-2b	15,900	-12,755	1.2	1.2
H-3a	16,448	-12,801	1.1	1.1
H-3b	23,974	-12,764	1.1	1.1

Another reason for these outliers could be the brief minimization procedure employed for computational efficiency. In the current work, energy gradient per atom for minimization criteria is taken as 0.1 kcal/mol/Å. When this value is decreased to 0.01 kcal/mol/Å, i.e. a more strict minimization is performed for H-1a conformer, rotamer outliers, Ramachandran outliers, Ramachandran favored and C^{β} deviations further decrease to 3.26, 0.21, 95.80 and 0, respectively.

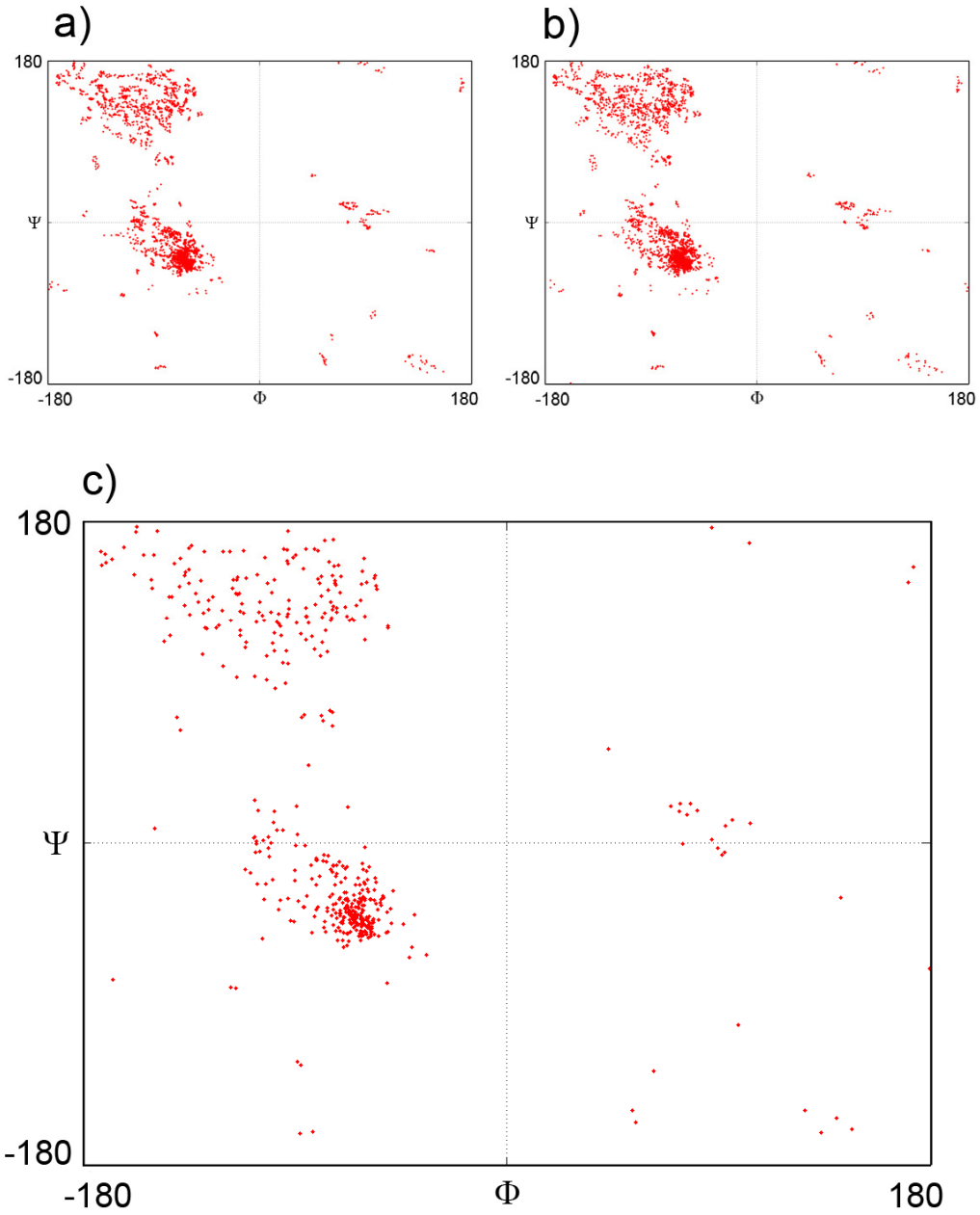

Figure 4.11. Ramachandran plot for the TIM conformers a) Low DF, b) High DF, c) Native conformation after energy minimization.

Table 4.5. MolProbity results for geometric validation

Conformers	Rotamer outliers	Ramachandran outliers	Ramachandran favored	C^{β} deviations > 0.25 Å
	Goal: <1%	Goal: <0.2%	Goal: >98%	Goal: 0
X-ray	6.25	0.42	95.17	3
L-1a	5.71	0.42	95.59	1
L-1b	4.89	0.21	95.8	1
L-2a	4.89	0.42	95.38	1
L-2b	5.98	0.42	95.38	3
L-3a	5.98	0.21	95.59	1
L-3b	5.16	0.42	95.59	2
H-1a	5.98	0.42	95.59	2
H-1b	5.43	0.21	96.01	1
H-2a	4.35	0.42	95.38	2
H-2b	6.52	0.21	95.38	2
H-3a	5.43	0.21	95.59	0
H-3b	5.16	0.42	95.59	2

4.2.2. Comparison of Reverse-mapped and MD Conformations

In previous section, it is shown that the reverse-mapped conformers have reasonable energies and geometries. In order to report the success of the reverse mapping method in conformational sampling, the structures generated are compared with the 60 ns MD trajectory for dimeric TIM. Each conformer (L-1a to H-3b) has been superimposed -based on C^{α} atoms- onto the snapshots obtained from the MD trajectory (6000 snapshots from 57 ns trajectory that excludes the 3 ns equilibration time). In Table 4.6, the average and minimum values of RMSDs over the trajectory are displayed for each conformer. Minimum RMSD values between 1.3 and 1.7 mean that the new conformers are close enough to the conformational space sampled by MD results.

Table 4.6. The average and the minimum of RMSD over the trajectory

Conformers	RMSD (Å)	
	Average RMSD	Minimum RMSD
X-ray	1.846	1.377
L-1a	1.895	1.434
L-1b	2.142	1.419
L-2a	2.258	1.515
L-2b	1.731	1.321
L-3a	2.084	1.521
L-3b	1.902	1.377
H-1a	2.051	1.516
H-1b	2.391	1.511
H-2a	2.535	1.698
H-2b	1.808	1.262
H-3a	2.290	1.685
H-3b	2.040	1.415

Even though there is appreciable similarity between the reverse-mapped conformers and MD snapshots, the new conformers should also differ among each other to some extent. Thus, the similarity measure -RMSD- for the new conformers is tabulated in matrix form in Appendices A and B for the overall structure and loop 6 regions.

When the dynamics of TIM is considered, it is known that in the MD study there are both open and closed forms of the structure. Therefore, it should be noted whether the new conformers are capable of covering the loop opening/closing motion besides being close to the conformational space sampled with MD. For this aim, in Figure 4.12 new conformers with low df (a) and high df (b) are superimposed and loop6 mobility is emphasized. Among these new conformers, H-1a and H-1b are selected due to their loop6 regions motion. In panel c, red and green colored loops are H-1a and H-1b respectively, and the conformer with black colored is the native structure. Glu171 and Thr172-the tip region for the loop6- moves 4.39 Å and 5.13 Å, respectively. In d panel, a side view of TIM is shown, where the direction of motion between H-1a and H-1b is indicated. The direction of motion

indicates the clockwise rotation of the A-subunit and counter-clockwise of B-subunit from H-1b which is depicted with indistinct lines.

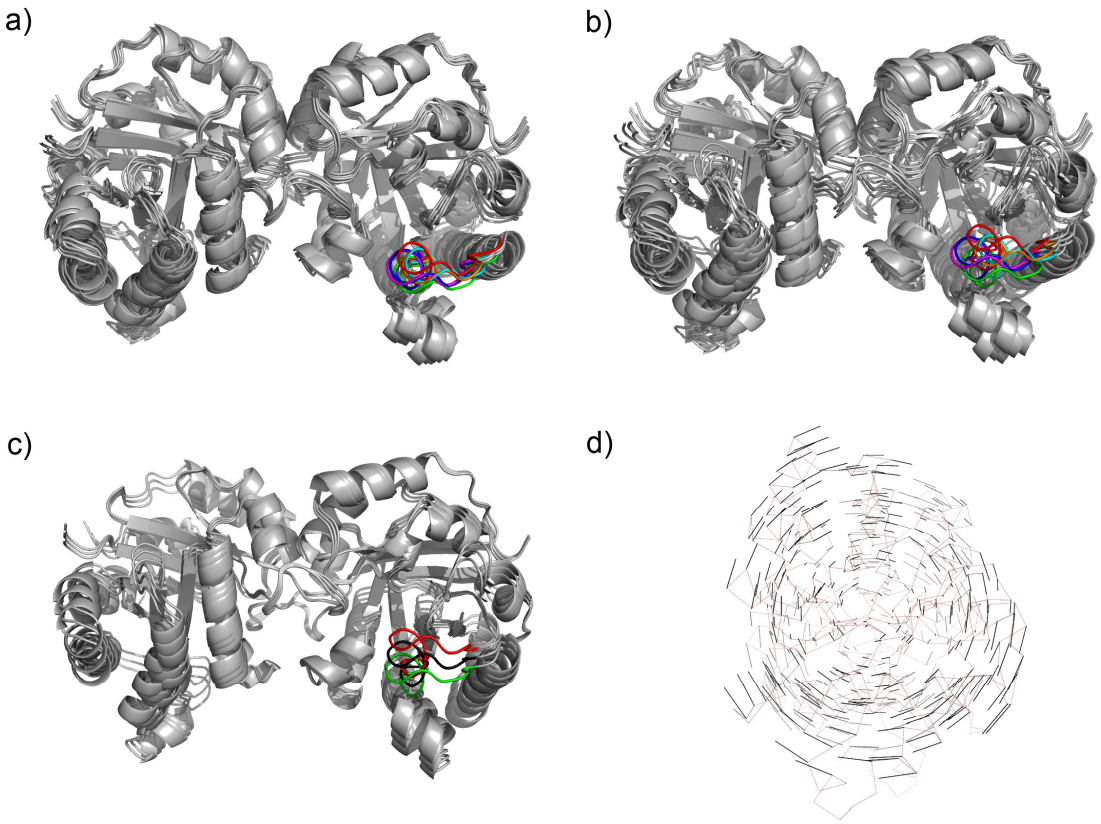

Figure 4.12. New conformers with low df (a) and high df (b) are superimposed and loop6 mobility is emphasized. And the loop closing/opening in from front (c) and side view(d) with a different representation.

To see the extend of loop6 closure, the H-1a conformer is taken as the initial structure and reverse-mapping method is applied again. And conveniently, the new structure, H-1a', (Figure 4.12), closes more over the active site. The results can be given on a quantative basis: in H-1a', the distance between Glu171 and Thr 208 is 15.60 Å, whereas the same distance is 15.06 Å for h-1a (before the second reverse-mapping). Figure 4.10 (a) and (d) gives the normalized probability of the same distances in the first mode of essential dynamics from MD trajectory. The corresponding distances in open (8TIM) and closed (1TPH) X-ray structures are 17.3 Å and 13.3 Å. Hence, reverse-mapping procedure can

sample conformations of loop 6 along the closing direction, which is observed in the slowest ANM mode.

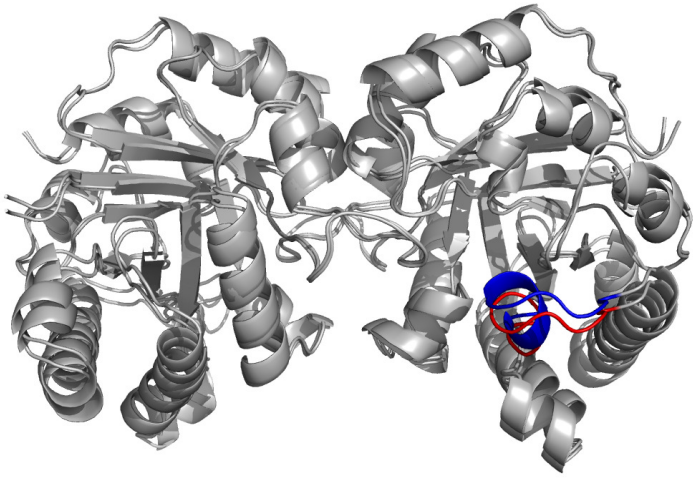

Figure 4.12. The H-1a structure (with red loop6) is further deformed along the first mode. The new structure is called H-1a'.

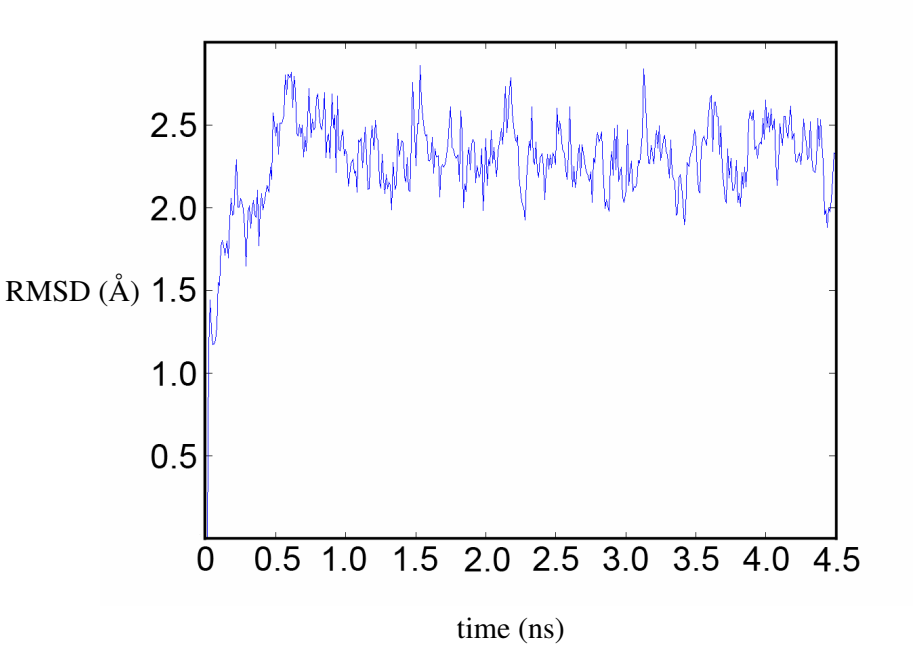

Figure 4.13. RMSDs from MD simulation of H-1a.

After generating new structures along harmonic modes, implicit minimization provides removal of steric overlaps and brings the structure to a local minima, but does not guarantee a stable structure. Hence, the stability of reverse-mapped structures are analyzed by MD simulation. For this reason, H-1a structure is provided as the initial frame for the MD simulation of 4.5 ns duration using the same parameters section 4.1.1. The resulting RMSD plot (Figure 4.13) indicates no conformation instability in the presence of explicit solvent molecules.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

Molecular simulations on TIM have so far concentrated on the dynamics of active site and loop 6 (Brown and Kollman, 1987). In contrast, recent study with elastic network models (Kurkcuoglu *et al.*, 2006) has shown the existence of dominant domain motions in TIM with functional importance. Accordingly, conformational freedom of the whole apo TIM structure has been taken into account for the first time in the current MD simulations - both for monomeric and dimeric TIM. In conformity with elastic network model results, significant collective motions are observed in the essential modes describing the counter-rotation of the identical subunits in the dimer (Figure 4.9). In contrast, monoTIM structure displays rather independent dynamics of the outer loops without major cooperative deformations in the whole monomer. More importantly, the global twisting in the first mode of dimer simulation that accounts for 34% of total motion is strongly coupled to the loop 6 dynamics along the loop closure direction, whereas no such action is observed in monoTIM.

Figure 4.10 clearly indicates that loop 6 makes significant progress towards closure based on the action of essential modes- most importantly the first mode- in the dimeric structure. Thus, the effective coordination of loop 6 by the global motions in apo TIM emerges as an inherent feature of the native enzyme. A recent NMR relaxation study (Eisenmesser, 2005) has shown similar characteristic motions of cyclophilin A during catalysis and in apo form, indicating that the intrinsic nature of collective dynamics may affect the catalytic turnover rate.

As has been expected in previous work on monomeric TIM (Thanki, 1997), the simulations also show that the absence of the second subunit in monoTIM increases solvent exposure of active site residues (Lys13 and His95 in Figure 5), which should negatively affect catalytic activity. At the same time, the onset of rigid-body motions-

predominantly twisting- of the two monomers in dimeric TIM has functional importance in terms of targeting loop 6's closure over the active site, which is not observed in monoTIM in the time-scale of our simulations. Thus, dimeric TIM captures the features of a functional scaffold, which controls the motion of outer helices and loops about the β-strand core region of each barrel. Since closure of loop 6 over the active site is necessary for catalysis, its proper coordination should –at least to some extent- contribute to the high catalytic efficiency of TIM.

As summarized above, classical MD simulations have provided a detailed picture of the ns dynamics of TIM and its correlation with loop 6 closure. In the second part of the thesis, new TIM conformers are generated by reverse-mapping ANM conformations. With this sampling procedure, different conformations between the closed and open states could be observed without time dependent information on the process. Reverse-mapped structures are found to be similar to the conformers encountered during MD simulations. These conformers are also validated in detail and scored to be in the 100 percentile via MolProbity server (Simon *et al.*, 2003; Davis *et al.*, 2004) (Table 4.5), i.e. rated the best among structures of comparable resolution. Finally, the high computational efficiency of this reverse-mapping technique should be stressed, which takes less than 10 minutes for each conformer generation compared to a 30 ns MD run that takes a month for dimeric TIM using the same computational platform.

5.2. Recommendations

Although the dimer and monomer forms of apo TIM are studied extensively in the first part of this thesis, further simulations on ligand-bound forms should be performed to see the changes in dynamics. The solvent accessible area in monoTIM simulations should be calculated with the ligand-bound form, because the existence of ligand would probably effect the dynamics. But of course, the absence of the second subunit is still expected to affect the collective motions, which is important for conformational dynamics.

In the second part, where the new conformations are generated along the harmonic modes, the method needs further development for larger systems. However, with the fully

atomistic representation used in this thesis, the methodology would not be feasible due to system size limitations. In order to work with supramolecular assemblages, such as ribosome, the network should be constructed in the mixed coarse-grained resolution with atomistic and low-resolution regions (Kurkcuoglu *et al.*, 2006) and the reverse-mapping procedure should be performed using constraints on the low-resolution regions.

APPENDIX A: Root Mean Square Deviation of the Conformers Before Energy Minimization

Table A.1. RMSD before energy minimization

	Xray	L-1a	L-1b	L-2a	L-2b	L-3a	L-3b	H-1a	H-1b	H-2a	H-2b	H-3a	H-3b
Xray	0.0	0.8	0.8	0.8	0.8	0.7	0.7	1.2	1.2	1.2	1.2	1.1	1.1
L-1a	0.8	0.0	1.7	1.1	1.1	1.1	1.1	0.4	2.1	1.4	1.4	1.4	1.4
L-1b	0.8	1.7	0.0	1.1	1.1	1.1	1.1	2.1	0.4	1.4	1.4	1.4	1.4
L-2a	0.8	1.1	1.1	0.0	1.6	1.1	1.1	1.5	1.5	0.4	1.9	1.4	1.4
L-2b	0.8	1.1	1.1	1.6	0.0	1.1	1.1	1.5	1.5	1.9	0.4	1.4	1.4
L-3a	0.7	1.1	1.1	1.1	1.1	0.0	1.5	1.4	1.4	1.4	1.4	0.4	1.9
L-3b	0.7	1.1	1.1	1.1	1.1	1.5	0.0	1.4	1.4	1.4	1.4	1.9	0.4
H-1a	1.2	0.4	2.1	1.5	1.5	1.4	1.4	0.0	2.5	1.7	1.7	1.7	1.7
H-1b	1.2	2.1	0.4	1.5	1.5	1.4	1.4	2.5	0.0	1.7	1.7	1.7	1.7
H-2a	1.2	1.4	1.4	0.4	1.9	1.4	1.4	1.7	1.7	0.0	2.3	1.6	1.6
H-2b	1.2	1.4	1.4	1.9	0.4	1.4	1.4	1.7	1.7	2.3	0.0	1.6	1.6
H-3a	1.1	1.4	1.4	1.4	1.4	0.4	1.9	1.7	1.7	1.6	1.6	0.0	2.2
H-3b	1.1	1.4	1.4	1.4	1.4	1.9	0.4	1.7	1.7	1.6	1.6	2.2	0.0

Table A.2. RMSD before energy minimization for loop6 (averaged over two subunits)

	Xray	**L-1a**	**L-1b**	**L-2a**	**L-2b**	**L-3a**	**L-3b**	**H-1a**	**H-1b**	**H-2a**	**H-2b**	**H-3a**	**H-3b**
Xray	0.0	1.4	1.4	1.1	1.1	1.1	1.1	2.0	2.0	1.6	1.6	1.7	1.7
L-1a	1.4	0.0	2.7	1.9	1.5	1.8	1.8	0.7	3.4	2.4	1.8	2.2	2.2
L-1b	1.4	2.7	0.0	1.5	1.9	1.8	1.8	3.4	0.7	1.9	2.4	2.2	2.2
L-2a	1.1	1.9	1.5	0.0	2.2	1.3	1.4	2.5	2.1	0.5	2.7	1.7	1.9
L-2b	1.1	1.5	1.9	2.2	0.0	1.4	1.3	2.1	2.5	2.7	0.5	1.9	1.7
L-3a	1.1	1.8	1.8	1.3	1.4	0.0	2.3	2.3	2.3	1.7	1.7	0.6	2.8
L-3b	1.1	1.8	1.8	1.4	1.3	2.3	0.0	2.3	2.3	1.7	1.7	2.8	0.6
H-1a	2.0	0.7	3.4	2.5	2.1	2.3	2.3	0.0	4.1	2.9	2.3	2.7	2.6
H-1b	2.0	3.4	0.7	2.1	2.5	2.3	2.3	4.1	0.0	2.3	2.9	2.6	2.6
H-2a	1.6	2.4	1.9	0.5	2.7	1.7	1.7	2.9	2.3	0.0	3.3	1.9	2.0
H-2b	1.6	1.8	2.4	2.7	0.5	1.7	1.7	2.3	2.9	3.3	0.0	2.0	1.9
H-3a	1.7	2.2	2.2	1.7	1.9	0.6	2.8	2.7	2.6	1.9	2.0	0.0	3.4
H-3b	1.7	2.2	2.2	1.9	1.7	2.8	0.6	2.6	2.6	2.0	1.9	3.4	0.0

APPENDIX B: Root Mean Square Deviation of the Conformers After Energy Minimization

Table B.1. RMSD after energy minimization

	Xray	**L-1a**	**L-1b**	**L-2a**	**L-2b**	**L-3a**	**L-3b**	**H-1a**	**H-1b**	**H-2a**	**H-2b**	**H-3a**	**H-3b**
Xray	0.0	0.9	0.9	0.8	0.8	0.8	0.8	1.3	1.3	1.2	1.2	1.1	1.1
L-1a	0.9	0.0	1.7	1.2	1.2	1.1	1.1	0.4	2.1	1.5	1.5	1.4	1.4
L-1b	0.9	1.7	0.0	1.2	1.2	1.1	1.1	2.1	0.4	1.5	1.5	1.4	1.4
L-2a	0.8	1.2	1.2	0.0	1.6	1.1	1.1	1.5	1.5	0.4	2.0	1.4	1.4
L-2b	0.8	1.2	1.2	1.6	0.0	1.1	1.1	1.5	1.5	2.0	0.4	1.4	1.4
L-3a	0.8	1.1	1.1	1.1	1.1	0.0	1.5	1.5	1.5	1.4	1.4	0.4	1.9
L-3b	0.8	1.1	1.1	1.1	1.1	1.5	0.0	1.5	1.5	1.4	1.4	1.9	0.4
H-1a	1.3	0.4	2.1	1.5	1.5	1.5	1.5	0.0	2.5	1.8	1.8	1.7	1.7
H-1b	1.3	2.1	0.4	1.5	1.5	1.5	1.5	2.5	0.0	1.8	1.8	1.7	1.7
H-2a	1.2	1.5	1.5	0.4	2.0	1.4	1.4	1.8	1.8	0.0	2.4	1.7	1.7
H-2b	1.2	1.5	1.5	2.0	0.4	1.4	1.4	1.8	1.8	2.4	0.0	1.7	1.7
H-3a	1.1	1.4	1.4	1.4	1.4	0.4	1.9	1.7	1.7	1.7	1.7	0.0	2.3
H-3b	1.1	1.4	1.4	1.4	1.4	1.9	0.4	1.7	1.7	1.7	1.7	2.3	0.0

Table B.2. RMSD after energy minimization for loop6 (averaged over two subunits)

	Xray	**L-1a**	**L-1b**	**L-2a**	**L-2b**	**L-3a**	**L-3b**	**H-1a**	**H-1b**	**H-2a**	**H-2b**	**H-3a**	**H-3b**
Xray	0.0	1.4	1.4	1.1	1.1	1.1	1.2	2.1	2.1	1.7	1.7	1.7	1.8
L-1a	1.4	0.0	2.8	2.0	1.6	1.8	1.8	0.7	3.5	2.4	1.9	2.2	2.2
L-1b	1.4	2.8	0.0	1.6	2.0	1.8	1.8	3.5	0.7	1.9	2.4	2.2	2.2
L-2a	1.1	2.0	1.6	0.0	2.2	1.3	1.4	2.6	2.2	0.6	2.8	1.7	1.9
L-2b	1.1	1.6	2.0	2.2	0.0	1.4	1.3	2.1	2.6	2.8	0.6	1.9	1.7
L-3a	1.1	1.8	1.8	1.3	1.4	0.0	2.3	2.4	2.4	1.8	1.8	0.6	2.9
L-3b	1.2	1.8	1.8	1.4	1.3	2.3	0.0	2.4	2.4	1.7	1.8	2.9	0.6
H-1a	2.1	0.7	3.5	2.6	2.1	2.4	2.4	0.0	4.2	3.0	2.3	2.7	2.7
H-1b	2.1	3.5	0.7	2.2	2.6	2.4	2.4	4.2	0.0	2.4	3.0	2.7	2.7
H-2a	1.7	2.4	1.9	0.6	2.8	1.8	1.7	3.0	2.4	0.0	3.3	1.9	2.1
H-2b	1.7	1.9	2.4	2.8	0.6	1.8	1.8	2.3	3.0	3.3	0.0	2.1	1.9
H-3a	1.7	2.2	2.2	1.7	1.9	0.6	2.9	2.7	2.7	1.9	2.1	0.0	3.5
H-3b	1.8	2.2	2.2	1.9	1.7	2.9	0.6	2.7	2.7	2.1	1.9	3.5	0.0

REFERENCES

Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen, "Essential dynamics of proteins", *Proteins,* Vol. 17, pp. 412-425, 1993.

Atilgan A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, I. Bahar, "Anisotropy of Fluctation Dynamics of Proteins with an Elastic Network Model", *Biophyscial Journal*, Vol. 80, pp. 505-515, 2001.

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

Benkovic, S. J. and S. Hammes-Schiffer, "A perspective on enzyme catalysis", *Science*, Vol. 301, pp. 1196-1202, 2003.

Berendsen, H.J.C., J. P. M. Postma, W. F. Van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath", *Journal of Chemical Physics*, Vol. 81, pp. 3684-3690, 1984

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

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

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

Brown, F.K. and P. A. Kollman, "Molecular dynamics simulations of "loop closing" in the enzyme triose phosphate isomerase", *Journal of Moecular. Biology*, Vol. 198, pp. 533-546, 1987.

Case, D. A., T. A. Darden, T. E. I. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, B. Wang, D. A. Pearlman, *et al.,* 2004, *AMBER 8*, University of California, San Francisco

Cui, Q., I. Bahar, Editors, 2005, *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, CRC Press, Boca Raton, FL

Case, D.A., T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. Woods, "The Amber biomolecular simulation programs", *Journal of Computational Chemistry*, Vol. 26, pp. 1668-1688, 2005

Daniel, R. M., R. V. Dunn, J. L. Finney, and J. C. Smith, "The role of dynamics in enzyme activity", *Annual Review of Biophysics and Biomolecular Structure*, Vol. 32, pp. 69–92, 2003.

Derreumaux, P. and T. Schlick, "The loop opening/closing motion of the enzyme triosephosphate isomerase", *Biophysical Journal*, Vol. 74, pp. 72-81, 1998.

Desamero, R., S. Rozovsky, N. Zhadin, A. McDermott, and R. Callender, "Active site loop motion in Triosephosphate Isomerase: T-jump relaxation spectroscopy of thermal activation", *Biochemistry*, Vol. 42 pp. 2941-2951, 2003.

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* , Vol. 40, pp. 512-524, 2000.

Doruker P. and R. L. Jernigan, "Functional Motions Can Be Extracted from On-Lattice Construction of Protein Structures", *Proteins,* Vol. 53, pp.174-181, 2003.

Doruker, P., L. Nilsson, and O. Kurkcuoglu, "Collective dynamics of EcoRI-DNA complex by elastic network model and molecular dynamics simulations", *Journal of Biomolecular Structure and Dynamics,* Vol. 24, pp. 1-15, 2006.

Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo and T. Lee, "A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins", *Journal of Computational Chemistry,* Vol. 24, pp. 1999-2012, 2003.

Eisenmesser, E. J., O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J.J. Skalicky, L. E. Kay and D. Kern, "Intrinsic dynamics of an enzyme underlies catalysis", *Nature*, Vol 438, pp. 117-121, 2005.

Erijman, L., G. H. Lorimer, and G. Weber, "Reversible dissociation and conformational stability of dimeric ribulose biphosphate carboxylase", *Biochemistry*, Vol.32, pp. 5187-5195, 1993.

Ghosh, A., C. S. Rapp, R. A. Friesner, "Generalized Born model based on a surface integral formulation", *Journal of Physical Chemistry*, Vol. 102, pp. 10983–10990, 1998.

Hawkins, G. D., C. Cramer, D. G. Truhlar, "Pairwise solute descreening of solute charges from a dielectric medium", *Chemical Physical Letters*, Vol. 246, pp. 122–129, 1995.

Hawkins, G. D., C. J. Cramer, D. G. Truhlar, "Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium", *Journal of Physical Chemistry*, Vol. 100, pp.19824–19839, 1996.

Hockney, R. W., "The Potential Calculations on Some Applications", *Methods is Computational Physics*, Vol. 9, pp. 136-143, 1970.

Davis, I. W., L. W. Murray, J. S. Richardson, D. C. Richardson, "MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes", *Nucleic Acids Research,* Vol. 32, pp. W615-W619, 2004.

Davis, I. W., A. L. Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall III, J. Snoeyink, J. S. Richardson and D. C. Richardson, "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids", *Nucleic Acids Research*, Vol 35, pp. W375-W383, 2007.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water", *Journal of Chemical Physics*, Vol. 79, pp. 926-935, 1983.

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

Karplus, M.,and G. A. Petsko, "Molecular Dynamics simulations in biology", *Nature*, Vol. 347, page 631-639, 1990.

Kempf, J. G., J. Y. Jung, C. Ragain, N. S. Sampson and J. P. Loria, "Dynamic requirements for a functional protein hinge", *Journal of Molecular Biology,* Vol. 368, pp. 131-149, 2007.

Kitao, A., F. Hirata, N. Go, "The effects of solvent on the conformation and the collective motions of a protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum", *Chemical Physics*, Vol. 158, pp. 447–472, 1991.

Kurkcuoglu, O., R. L. Jernigan, and P. Doruker, "Loop motions of triosephosphate isomerase observed with elastic networks", *Biochemistry*, Vol. 45, pp. 1173-1182, 2006.

Leach, A.R., 2001, *Molecular Modeling: Principles and Application (2nd Edition)*, Prentice Hall,

Lovell, C. C., I. W. Davis, W. B. Arendall III, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, D. C. Richardson, "Structure validation by C-alpha geometry:

phi, psi, and C-beta deviation", *Proteins: Structure, Function, and Genetics*, Vol. 50, pp. 437-450, 2003.

Massi, F., C. Wang, and A. G.Palmer III, "Solution NMR and computer simulation studies of active site loop motion in triosephosphate isomerase", *Biochemistry,* Vol. 45, pp. 10787-10794, 2006.

Qiu, D, P. S. Shenkin, F. P. Hollinger, W. C. Still, "The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii.", *Journal of Physical Chemistry A*, Vol. 101, pp. 3005–3014, 1997.

Rozovsky, S., A. E. McDermott, "The Time scale of the Catalytic Loop Motion in Triosephosphate Isomerase", *Journal of Molecular Biology,*Vol. 310, pp. 259-270, 2001.

Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes", *Journal of Computational Physics,* Vol. 23, pp. 327-341, 1977.

Schaefer, M, and M. A. Karplus, "Comprehensive analytical treatment of continuum electrostatics", Journal of Physical Chemistry, Vol. 100, pp. 1578–1599, 1996.

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

Schneider, A. S., "Triosephosphate isomerase deficiency: historical perspectives and molecular aspects", *Best Practice & Research Clinical Haematology*, Vol. 13, pp. 119-140, 2000.

Still, W.C., A. Tempczyk, R. C. Hawley, T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics", *Journal of American Chemical Society*, Vol. 112, pp. 6127-6129, 1990.

Thanki, N., J. P. Zeelen, M. Mathieu, R. Jaenicke, R. A. Abagyan, R. K. Wierenga and W. Schliebs, "Protein engineering with monomeric triosephosphate isomerase (monoTIM): the modeling and structure verification of a seven-loop residue", *Protein Engineering,* Vol. 10, pp. 159-167, 1997.

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

Van Wynsberghe, A. W. and Q. Cui, "Interpreting correlated motions using normal mode analysis", *Structure,* Vol. 14, pp. 1647-1653, 2006.

Verlet, L., "Computer Experiments of Classical Fluids I. Thermodynamical Properties of Lennard-Jones molecules", *Physical Review*, Vol. 159, pp 98-106, 1967.

Wang, T. and R. C. Wade, "Implicit Solvent Models for Flexible Protein-Protein Docking by Molecular Dynamics Simulation", *Proteins*, Vol. 50, pp. 158-169, 2003.

Wang, W., O. Donini, C. M. Reyes, and P. A. Kollman, "Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions", *Annual Review of Biophysics & Biomolecular Structure,*. Vol. *30*, pp. 211–243, 2001.

Williams, J. C. and A. E. McDermott, "Dynamics of the flexible loop of triosephoshate isomerase: the loop motion is not ligand gated", *Biochemistry*, Vol. 34, pp. 8309-8319, 1995.

Xiang, J., J. Jung and N. S. Sampson, "Entropy effects on protein hinges: the reaction catalyzed by triosephosphate isomerase", *Biochemistry*, Vol. 43, pp. 11436-11445, 2004.

Zhang, Z., S. Sugio, E. A. Komives, K. D. Liu, J. R. Knowles, G. A. Petsko, and D. Ringe, "Crystal structure of recombinant chicken Triosephosphate isomerase-

Phosphoglycolohydroxamate complex at 1.8Å resolution", *Biochemistry*, Vol. 33, pp. 2830-2837, 1994.

Zhang, Z., Y. Shi, and H. Liu, "Molecular Dynamics Simulations of Peptides and Proteins with Amplified Collective Motions", *Biophysical Journal*, Vol. 84, pp. 3583-3593, 2003.