ELASTIC NETWORK MODEL BASED APPROACHES FOR CONFORMER GENERATION AND DOCKING APPLICATIONS

by

Zeynep Kürkçuoğlu Soner

B.S., Chemical Engineering, Boğaziçi University, 2008

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Chemical Engineering
Boğaziçi University

2015

ELASTIC NETWORK MODEL BASED APPROACHES FOR CONFORMER GENERATION AND DOCKING APPLICATIONS

APPROVED BY:

Prof. Pemra Doruker Turgut (Thesis Supervisor)	.....
Prof. Canan Atılgan	.....
Prof. Mehmet C. Çamurdan	.....
Prof. Türkan Haliloğlu	.....
Assoc. Prof. E. Demet Akten Akdoğan	.....

DATE OF APPROVAL: 24.07.2015

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Prof. Pemra Doruker for her guidance and support throughout my PhD thesis. The words cannot describe the thankfulness that I would like to express, it was an honour to work with her. She always encouraged me to explore knowledge whenever it is possible; throughout the courses, scientific meetings or by gathering with people who are experts in the field.

I would also like to thank Prof. Canan Atilgan, Prof. Mehmet C. Camurdan, Prof. Turkan Haliloglu and Assoc. Prof. E. Demet Akten Akdogan for sparing their time to read and comment about my thesis. Their valuable suggestions and discussions contributed a lot to this study. Many thanks to Assist. Prof. Bulent Balta for his contributions to molecular dynamics studies.

It was a great pleasure for me to collaborate with Prof. Ivet Bahar who has such an inspiring personality. Her remarkable comments and suggestions shaped in a great extent this thesis.

I would like to acknowledge the financial support from TÜBİTAK BİDEB 2210-2211, TÜBİTAK Project 109M213 and 113M237, DPT Project 2009k120520, Bogazici University BAP 5714 and Betil fund.

I would like to thank to all my friends from Polymer Research Center for all the fun we had beside fruitful academic discussions; I will miss our breakfast gatherings along Bosphorus, enjoying summer in Kilyos and Bogazici pool in Bebek. I would also like to thank to my dear friends from Galatasaray and Bogazici for making my life colorful, meaningful and enjoyable. Special thanks to Yudum for being my "guarantor" during the fellowship.

And last but not least, I am thankful to my family for always being there for me with endless love, encouraging and supporting me in any circumstances. Special

thanks to my sister Ozge for her guidance and motivating me during my academic life. And I am grateful to my husband Seren for all his help in writing this thesis and for being my life partner with his unconditional love and support.

I am so lucky to be surrounded with such wonderful people...

ABSTRACT

ELASTIC NETWORK MODEL BASED APPROACHES FOR CONFORMER GENERATION AND DOCKING APPLICATIONS

The dynamic nature of proteins poses challenging problems in computational biology, especially in terms of conformational sampling and transitions. In this thesis, an Elastic Network Model (ENM)-based computational method, namely ClustENM, was developed for sampling large conformational changes of biomolecules with various sizes and oligomerization states. ClustENM is an iterative method that combines ENM with energy minimization and clustering steps. It is an unbiased technique, which necessitates only an initial structure of the biomolecule as input but no information on target. To test the performance of ClustENM in conformational sampling, it was applied to six systems, namely adenylate kinase (AK), calmodulin, p38 MAP kinase, HIV-1 reverse transcriptase, triosephosphate isomerase (TIM), and supramolecule 70S ribosome. The generated atomistic conformers were found to be in agreement with experimental data (971 structures) and molecular dynamics (MD) simulations. ClustENM was used to model the trigger factor-50S subunit of ribosome complex, leading to structures consistent with the data from cryo-EM. Additionally, ligand effects on TIM conformational dynamics were investigated based on MD simulations of its apo form and complexes with an inhibitor or its substrate. Generated conformers from ClustENM were further used in docking applications for AK, LAO-binding protein, dipeptide binding protein and biotin carboxylase. Close-to-native ligand binding poses were obtained especially in the first three cases. Thus, ClustENM emerges as a computationally efficient method applicable to extremely large systems or transitions. Its utility relies on the generation of a manageable number of atomistic conformers that are entropically accessible to a folded starting structure, which can also assist ligand docking applications.

ÖZET

KONFORMASYON TARAMASI VE YERLEŞTİRME UYGULAMALARI İÇİN ELASTİK AĞ YAPI MODELİ TABANLI YAKLAŞIMLAR

Proteinlerin dinamik yapısı, hesaplamalı biyolojide, özellikle konformasyonel örnekleme ve geçişler açısından zorlu problemler oluşturmaktadır. Bu tezde ClustENM adında Elastik Ağ Yapı Modeli (ENM) bazlı bir metot, özellikle büyük yapısal değişiklik geçiren, çeşitli boyutlarda ve oligomerik haldeki biyomoleküllere uygulanmak üzere geliştirilmiştir. Enerji minimizasyonu, kümelenme ve ENM'yi birleştiren ClustENM, iteratif bir yöntemdir. Girdi olarak sadece biyomolekülün deneysel bir yapısına gereksinim duyan, tarafsız bir tekniktir. ClustENM'nin performansını konformasyonel örneklemede sınamak için altı sistem kullanılmıştır: adenilat kinaz (AK), kalmodulin, p38 MAP kinaz, HIV-1 ters transkriptaz, triosefosfat izomeraz (TIM) ve bir süpramolekül olan 70S ribozom. Bu yöntemle üretilen atomistik çözünürlükteki yapıların, mevcut deneysel veriler (971 yapı) ve moleküler dinamik (MD) simülasyonları ile uyumlu olduğu gözlemlenmiştir. ClustENM, ribozomun 50S altbirimi ve tetikleyici faktör kompleks yapısına da uygulanmış olup, üretilen yapıların kriyo-elektron mikroskobundan elde edilen verilerle tutarlı olduğu görülmüştür. Ayrıca, bir inhibitor ve substratın TIM dinamiğine olan etkisini gözlemlemek amacıyla bağımsız MD verilerine detaylı bir inceleme yapılmıştır. ClustENM ile yaratılan yapılar, dört farklı protein için yerleştirme uygulamalarında kullanılmıştır: AK, LAO bağlayıcı protein, dipeptit bağlayıcı protein ve biyotin karboksilaz. Özellikle ilk üç sistem için kristal yapılardaki ligand konumlanmasına yakın sonuçlar elde edilmiştir. Böylece, ClustENM'in çok büyük sistemler veya geçişler için kullanılabilecek verimli bir yöntem olduğu görülmüştür. Yöntem, başlangıç yapısını kullanarak entropik açıdan erişilebilir yapıları üretmektedir ve bu yapılar yerleştirme çalışmalarında kullanılabilmektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii
ABSTRACT v
ÖZET vi
LIST OF FIGURES x
LIST OF TABLES xvi
LIST OF SYMBOLS xviii
LIST OF ACRONYMS/ABBREVIATIONS xx
1. INTRODUCTION 1
2. CONFORMATIONAL SEARCH TECHNIQUES FOR PROTEIN STRUCTURE AND DYNAMICS 4
2.1. Energy Landscape 4
2.2. Experimental Techniques for Determination of Protein Structure and Conformational Changes 6
2.3. Computational Techniques for Conformational Sampling and Transitions 8
2.3.1. Molecular Dynamics (MD) and Related Methods 8
2.3.2. Normal Mode Analysis (NMA) and Related Techniques 9
2.3.2.1. Classical NMA 10
2.3.2.2. Coarse-Grained NMA 11
2.3.2.3. Applications of ENM in Small Ligand-Protein Docking 15
2.3.3. Other Techniques for Conformational Search 17
2.3.4. Computational Techniques Used for Investigating Ribosome Conformational Search and Dynamics 19
3. MATERIALS AND METHODS 22
3.1. Anisotropic Network Model (ANM) 22
3.2. ClustENM-I for Conformational Sampling 25
3.3. ClustENM-II for Docking Applications 30
3.4. Molecular Dynamics Simulations 32
3.5. Ensemble Analysis 34
3.6. Softwares and Parameters Used for Protein-Ligand Docking 36

3.7. Mean Square Fluctuations 36
4. CONFORMATIONAL SAMPLING USING CLUSTENM-I FOR STRUCTURES UNDERGOING LARGE CONFORMATIONAL CHANGES 38
4.1. Adenylate Kinase 38
4.2. Calmodulin 46
4.3. Mitogen-activated protein kinase p38 49
4.4. HIV1-Reverse Transcriptase 50
5. APPLICATION OF CLUSTENM TO BACTERIAL RIBOSOME 54
5.1. 70S Ribosome 54
5.2. 50S ribosome-Trigger Factor (TF) Complex 60
6. TRIOSEPHOSPHATE ISOMERASE CONFORMATIONAL DYNAMICS . 64
6.1. Application of ClustENM-I on TIM 64
6.2. Inhibitor effect on TIM global dynamics and its specific interactions with the enzyme 68
6.2.1. Root Mean Square Deviation and Mean Square Fluctuations for Apo and bt10-Bound MD Runs 69
6.2.2. Specific interactions 70
6.2.3. Collective and Loop 6 Dynamics 75
6.3. Substrate DHAP effect on TIM global and catalytic loop dynamics . . 82
7. CLUSTENM-II: APPLICATION IN SMALL LIGAND DOCKING 85
7.1. Adenylate kinase 85
7.2. Lysine-arginine-ornithine-binding protein 90
7.3. Dipeptide Binding Protein 92
7.4. Biotin Carboxylase 95
8. CONCLUSIONS 103
8.1. Conformational sampling using ClustENM 103
8.2. Docking applications 104
8.3. Computational efficiency 105
8.4. Choice of parameters in ClustENM 105
8.5. Advantages/Limitations of ClustENM 106
9. FUTURE WORK 107

APPENDIX A: LIST OF PDB STRUCTURES USED IN ENSEMBLE ANALYSIS AND DETAILED DOCKING RESULTS FOR LAO BINDING PROTEIN . . . 108
REFERENCES 113

LIST OF FIGURES

Figure 2.1. A schematic representation of the change on the energy landscape upon ligand binding. 5

Figure 2.2. A state among many states with equivalent energy can be stabilized upon interaction with the partner. 5

Figure 2.3. Reduced energy barrier between two states due to environmental effects such pH, temperature etc. 6

Figure 3.1. Flowchart of ClustENM Procedure I. 27

Figure 3.2. Structures used in the application of ClustENM Procedure I. . . . 29

Figure 3.3. Flowchart of ClustENM Procedure II. 31

Figure 3.4. Open (left) and ligand bound (right) conformers of proteins used in application of ClustENM Procedure II. 33

Figure 4.1. Starting structures for AK. 39

Figure 4.2. Conformational sampling of AK. (a) LID-CORE and NMP-CORE angles. (b) Independent samplings based on four distinct starting structures (squares). (c) Representative surface that merges four distinct simulations (in panel b) into a single one. 40

Figure 4.3. NMP-CORE and LID-CORE angles cannot distinguish fully-open crystal structure and a generated conformer that have an RMSD of 8 Å. 43

Figure 4.4. RMSD histograms for AK conformers to different states, using starting conformer as (a) fully-open, (b) fully-closed, (c) closed-LID, (d) closed-NMP. 44

Figure 4.5. Accessible states when the starting structure is selected as fully-open, fully-closed, closed-LID or closed-NMP (LID domain is shown as blue, NMP as red). 45

Figure 4.6. Extended and closed states of CAM, wheren N-lobe (blue), linker (green) and C-lobe (red) are shown. 47

Figure 4.7. Conformational sampling for CAM. (a) Projection of experimental data (blue) onto the subspace spanned by the first three PCs of generated conformers (red) of CAM. (b) MSF for generated conformers and experimental data. 48

Figure 4.8. Initial/final RMSD range of the starting structure/generated conformers to the NMR models of 2k0e (blue). 49

Figure 4.9. p38 MAP kinase. (a) Projection of generated conformers (red), experimental (blue) and MD (black and green) data onto subspace spanned by the first two PCs of generated conformers (red). (b) MSF for generated (red) and MD (black) data. 50

Figure 4.10. p38 MAP kinase conformers from x-ray (blue), generated clusters (red) and MD (green) data. In all cases, the opaque structure corresponds to the starting structure. 51

Figure 4.11. HIV1-RT. (a) Projection of generated conformers (red), experimental data (blue) and MD data (black and green) onto subspace spanned by the first two PCs of generated conformers (red). (b) MSF for generated (red) and MD (black) data. 52

Figure 4.12. HIV1-RT conformers from x-ray (blue), generated clusters (red) and MD (green) data. In all cases, the opaque structure corresponds to the starting structure. 53

Figure 5.1. Mutual RMSD for 70S ribosome x-ray structures, based on P atoms of 16S and 23S only. 55

Figure 5.2. Closeness/farness of the generated conformers (columns) to crystal structures for 70S ribosome (rows). 56

Figure 5.3. 70S ribosome (a) Initial/final RMSD histogram. (b) PCA of generated conformers. 56

Figure 5.4. The dynamics of 70S from 32 generated structures with higher weights. 57

Figure 5.5. Mobility of tRNA together with L1 stalk. 58

Figure 5.6. Close-up view of protein dynamics located at the exit tunnel of 70S ribosome. 59

Figure 5.7. Energetically minimized 50S ribosome-TF complex structure (side and top views). 61

Figure 5.8. 50S ribosome-TF complex dynamics from generated conformers using ClustENM. 62

Figure 5.9. 50S ribosome-TF generated conformers (blue) aligned onto cryo-EM map (surface) and experimental TF structure (red). 63

Figure 6.1. TIM structure from *Trypanosoma cruzi*. 65

Figure 6.2. TIM PCA and MSF. (a) Subspace spanned by the first two PCs of generated conformers (red) with projected experimental data (blue) and MD data (black, green and yellow) (b) MSF for generated (red) and MD (black) conformers. 66

Figure 6.3. X-ray (blue), generated (red), MD (green) conformers of TIM. . . 67

Figure 6.4. Inhibitor bt10 structure. 68

Figure 6.5. Dimeric and monomeric RMSD profiles for MD runs of TIM. . . . 69

Figure 6.6. MSF of apo and complex TIM for each run. 71

Figure 6.7. Blue (red) residues' MSFs decrease (increase) upon bt10 binding by more than 20% compared to apo runs. 72

Figure 6.8. Aromatic interactions on bt10-bound tunnel region in TIM interface. 72

Figure 6.9. Change in H-bonding interactions upon bt10 binding to the interface. 75

Figure 6.10. PCA on TIM runs: (a) PC overlap matrices with reference to Apo1 PCs, (b)PC1 and PC2 for Apo1. 76

Figure 6.11. Residue-residue orientational cross-correlations based on the cumulative action of first five PCs for (a) apo, (b) complex and more than 100 PCs contributing to 90% of total motion for (c) apo and (d) complex, reported as averages over three independent runs. . . 78

Figure 6.12. Loop 6 orientational cross-correlations based on tip residues I173-G174-T175 for subunit A (a, c, e) and subunit B (b, d, f). 79

Figure 6.13. Loop 6 opening and closure observed in the monomers based on the distance between loop's tip residue G174 and a relatively immobile residue Y211. 80

Figure 6.14. (a) Loop 6 opening/closure event from G174-Y211 distance histogram in apo and complex. (b) RMSF for pseudodihedral angle in apo and complex runs. 81

Figure 6.15. Residue MSF for apo and DHAP complex simulations. 83

Figure 6.16. Distance histograms for I170-Y208, G171-Y208 and T172-Y208 apo and complex simulations. 84

Figure 7.1. RG vs RMSD correlation for AK (energy-based search). 87

Figure 7.2. AK docking results. Docking poses are shown for (c) apo, (d) gen1, (e) gen4, (f) gen7. 89

Figure 7.3. RG vs RMSD correlation for LAO binding (energy-based search). 92

Figure 7.4. LAO binding protein conformers used in HADDOCK: (a) gen1, gen4, apo, (b) gen6, ligand-bound form (gray). LAO docking poses are shown for (c) apo, (d) gen1, (e) gen4 and (f) gen6. 93

Figure 7.5. RG vs RMSD correlation for DBP protein. 96

Figure 7.6. DBP conformers used in HADDOCK (a) gen 1 (green) and gen3 (orange) aligned on apo (blue), (b) gen5 (magenta) aligned on closed form (gray). DBP docking poses (purple) for (c) apo, (d) gen1, (e) gen3 and (f) gen5 aligned on closed crystal structure (gray). Ligand in crystal structure and docking pose is shown in magenta and cyan stick representation, respectively. 97

Figure 7.7. RG vs RMSD correlation for BC protein. 99

Figure 7.8. BC conformers used in ATP docking (a) apo (blue), gen1 (green), gen2 (orange), (b) gen3 (yellow), gen4 (magenta), ligand-bound form (grey). 100

Figure 7.9. Ligand poses for (a) apo cluster 1, (b) gen1 cluster1, (c) gen2 cluster 1, (d) gen3 cluster 4 and (e) gen4 cluster 2 aligned on ligand-bound complex (grey). Ligand positioning in crystal and docked structures is shown in magenta and cyan stick representation, respectively. 101

LIST OF TABLES

Table 3.1. Structures used in ClustENM-I. 28

Table 3.2. Proteins and their ligands used in ClustENM-II. 32

Table 3.3. MD simulations system details for TIM, HIV1-RT and p38 kinase. 35

Table 4.1. RMSD table of representative conformers to the four states. 42

Table 6.1. Percentage of intact $\pi - \pi$ interactions in apo and complex runs. . 73

Table 6.2. Occurrence % of H-bonding residue pairs differing more than 15%. 74

Table 6.3. Individual and cumulative contributions of PCs to total motion. . 77

Table 7.1. Number of AK conformers from blind search/energy-based selection. 86

Table 7.2. Docking results for AK using AutoDock v4. 88

Table 7.3. Number of LAO conformers from blind search/energy-based selection. 91

Table 7.4. HADDOCK docking results for LAO conformers (summary). . . . 94

Table 7.5. Number of DBP conformers obtained using blind search. 95

Table 7.6. HADDOCK docking results for DBP. 96

Table 7.7. Number of BC conformers from blind search (monomer A). 98

Table 7.8. Number of BC conformers from blind search (monomer B). 98

Table 7.9. BC binding protein docking results details. 102

Table A.1. PDB ids of proteins used in ensemble analysis, selected based on 90% sequence similarity. 108

Table A.2. LAO binding protein docking results details from HADDOCK software. 111

LIST OF SYMBOLS

a_j	Coefficient $-1, 0, 1$
$\mathbf{d}_i$	Displacement vector
DF	Deformation RMSD
g	Generation
$h(R_c - R_{ij})$	Heavyside step function
$\mathbf{H}$	$3N$ by $3N$ Hessian matrix
k_B	Boltzmann's constant
N	Total number of residues (or atoms)
$\bar{R}$	Average position of the Cα atom of the i^{th} residue
$\mathbf{\Delta R}$	Positional fluctuation vector for N nodes
R_c	Cut-off radius
$R_i(t)$	Position of the Cα atom of the i^{th} residue at the t^{th} snapshot
R_{ij}	Instantaneous distance between nodes i and j
R^0_{ij}	Equilibrium distance between nodes i and j
$\mathbf{S}$	Orthogonal matrix with its columns being the normalized eigenvectors that give the directions (or shape) of normal modes
T	Absolute temperature
$\mathbf{u}_{ik}$	k^{th} eigenvector for the i^{th} node
$\mathbf{u}_j$	j^{th} eigenvector
$\mathbf{V}_1$	Combination of slowest modes when all modes are present
V_{ANM}	ANM potential energy
$\mathbf{V}_i$	Combination of slowest modes
α	Alpha
β	Beta
γ	Force constant for the harmonic spring
λ_k	k^{th} eigenvalue

$\mathbf{\Lambda}$	$3N$ by $3N$ diagonal matrix with diagonal elements as the squared normal mode frequencies
π	Pi

LIST OF ACRONYMS/ABBREVIATIONS

3D	Three Dimensional
A	Adenine
AK	Adenylate Kinase
AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
ANM	Anisotropic Network Model
ANM-MC	Anisotropic Network Model-Monte Carlo
BC	Biotin Carboxylase
CAM	Calmodulin
CG MD	Coarse-grained Molecular Dynamics
coMD	Collective Molecular Dynamics
DNA	Deoxyribonucleic Acid
E	Exit
EF-G	Elongation Factor G
EM	Electron Microscopy
ENM	Elastic Network Model
FIRST	Floppy Inclusions and Rigid Substructure Topography
FRET	Förster resonance energy transfer
GNM	Gaussian Network Model
GPU	Graphical Processing Unit
HENM	Hybrid Elastic Network Model
HIV-1 RT	Human Immunodeficiency Virus-1 Reverse Transcriptase
LAO	Lysinine-Arginine-Ornithine-Binding Protein
MAP	Mitogen Activated Protein
MC	Monte Carlo
MD	Molecular Dynamics
mRNA	Messenger Ribonucleic Acid
MSF	Mean-Square Fluctuation

NMA	Normal Mode Analysis
NMR	Nucleic Magnetic Resonance
p38	Mitogen Activated protein kinase p38
P	Peptidyl
PC	Principal Component
PCA	Principal Component Analysis
PDB	Protein Data Bank
PRS	Perturbation-Response Scanning
RG	Radius of Gyration
RMSD	Root-Mean-Square Deviation
RNA	Ribonucleic Acid
TF	Trigger Factor
TIM	Triosephosphate Isomerase
tRNA	Transfer Ribonucleic Acid

1. INTRODUCTION

Proteins may undergo small or large conformational changes in order to perform their cellular functions, such as binding, catalysis, switching and serving as structural elements of living organisms. The conformational flexibility of proteins poses one of the challenging problems in molecular biology. Efficient computational algorithms are necessary to sample protein conformations for more accurate prediction of binding sites and affinities in docking studies and also for studying the structure-dynamics-function relationship of proteins.

There are experimental techniques, such as X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy, which provide information about the structure and functional dynamics of proteins, but these techniques generally provide static pictures of populated states but not intermediates or have applicability problems for large systems [1]. In this respect computational techniques complement experiments by providing more detailed information on conformational transitions. Among atomistic computational techniques, molecular dynamics (MD) simulations are commonly used for studying protein motion. Normal mode analysis (NMA) based on classical force fields is another computational tool for studying conformational dynamics of proteins. However as the system sizes increases, both of these atomistic techniques suffer from computational cost [2]. Elastic network models (ENM), which are coarse-grained approaches employing pair-wise harmonic interactions, are computationally efficient tools that can provide insight about the vibrational dynamics of supramolecular systems [3–5]. The fact that collective ENM modes of the system are highly correlated with the conformational transition directions has encouraged their usage for facilitating standard conformational search algorithms. Similarly, ENMs have also become popular in generating multiple conformers for drug design studies.

This thesis describes a new ENM-based methodology called ClustENM, which can be used in sampling of large conformational changes of proteins and/or extremely large complexes, such as the ribosome. These conformers can further be utilized in

protein-ligand docking studies. This iterative method integrates ENM with energy minimization and clustering in order to generate atomistic conformers. Basically, the native structure is deformed along the linear combinations of slowest normal modes, which are related with the global motions of the protein. Generated structures are clustered and conformers taken from these clusters are energetically minimized using implicit solvent model in order to obtain relaxed structures. The procedure is repeated for several cycles so that large conformational changes along the collective directions can be sampled using this computational efficient algorithm.

In the first part of the thesis, applicability of ClustENM in conformational search for structures with various size and oligomeric states will be illustrated using six different systems: adenylate kinase (AK), calmodulin (CAM), human immunodeficiency virus-1 reverse transcriptase (HIV1-RT), triosephosphate isomerase (TIM), mitogen-activated protein (MAP) kinase p38 and supramolecule 70S ribosome. The generated conformers will be compared with the available experimental data and MD simulations (Chapter 4). Additionally, the method will also be applied on the complex of 50S subunit of ribosome and trigger factor (TF, a chaperone binding to the nascent peptide exit tunnel of the ribosome) to shed light onto the dynamics of this complex, which has not been studied yet (Chapter 5).

The thesis also contains a separate chapter for TIM (Chapter 6), describing the dynamics extracted from the generated conformers and also from MD simulations. Moreover, the effect of an inhibitor bound to the interface region of this dimeric enzyme will be revealed by detailed analysis of six independent 100 ns-long MD simulations of apo enzyme and inhibitor bound complex. The impact of the inhibitor presence on collective and catalytic loop dynamics, as well as specific interactions (e.g. aromatic, hydrogen bonding) will be explained in detail. Simalarly, the effect of TIM substrate bound to catalytic site on collective and local dynamics will be summarized.

Finally, the application of ClustENM in small ligand docking will be shown for four proteins expressing hinge-bending type motion: adenylate kinase, lysine-arginine-ornithine binding protein, dipeptide binding protein and biotin carboxylase. The gener-

ated conformers starting from apo structure will be used in docking to reveal the ligand positioning, which will be compared with the available crystal structure of ligand-bound complex (Chapter 7).

A general review of experimental and computational techniques used in conformational sampling and studying protein dynamics will be given in the following chapter (Chapter 2). Methods Chapter consists of the detailed description of ClustENM procedures and the list of systems, on which the methodology is applied. After Methods, results will be presented and the thesis will be concluded by Conclusions and possible applications of the methodology as Future Work.

2. CONFORMATIONAL SEARCH TECHNIQUES FOR PROTEIN STRUCTURE AND DYNAMICS

2.1. Energy Landscape

Energy landscape of a protein is a multidimensional surface, which may be described as a function of collective variables. It contains the information about the free energy required to change the thermodynamic state of the system [6]. The ability of a protein to sample alternate conformations, the probability of the sampling are examples of information that can be extracted from the energy landscape [7].

The energy landscapes for proteins are highly rugged or rough, and there are multiple minima on the surface, which correspond to stable states of the system [7–10]. A transition on the surface is possible from one minimum to another, by passing through a saddle point. This process is equivalent to conformational change of the protein. If a minimum is deep, it means that it is thermodynamically stable. The height of the energy barrier that surrounds the minima defines the kinetic stability of the protein. A deep minimum surrounded by high barriers may be highly populated during folding [11, 12]. It may also belong to ensemble of states under a particular condition, such as pH, temperature, concentration of partners (e.g. ligand). Sampling alternate stable conformations is vital for a protein to execute its function such as binding to partners and allosteric regulation. However, there can be also chance for creating dysfunctional interactions with partners upon visiting the alternative conformations [7].

The changes on the energy landscape are illustrated in Figures 2.1 to 2.3 [7]. Figure 2.1 shows the stabilization of a near-native state upon ligand binding. In another example shown in Figure 2.2, energy surface indicates that the protein may sample alternative conformations with equal energy. However one of these states may have higher affinity for a certain partner, and this may lead to further stabilization upon interaction with partner.

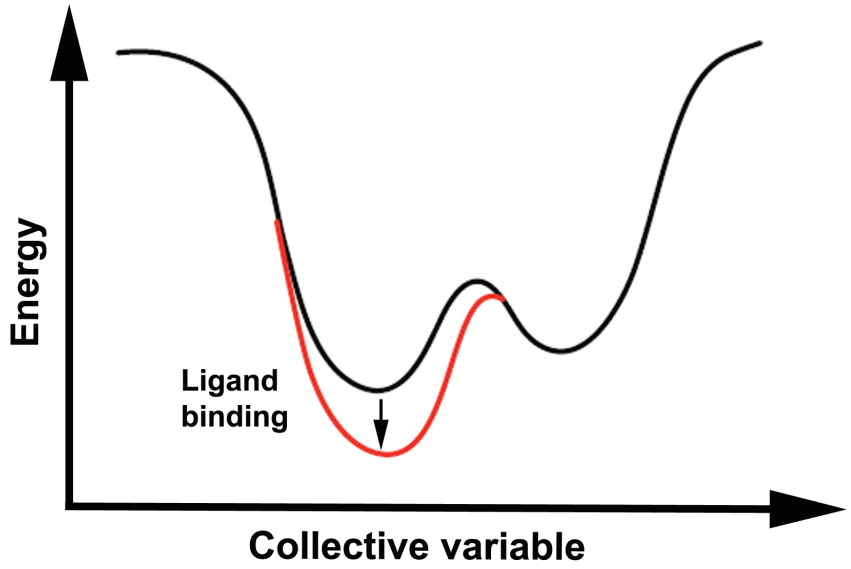

Figure 2.1. A schematic representation of the change on the energy landscape upon ligand binding.

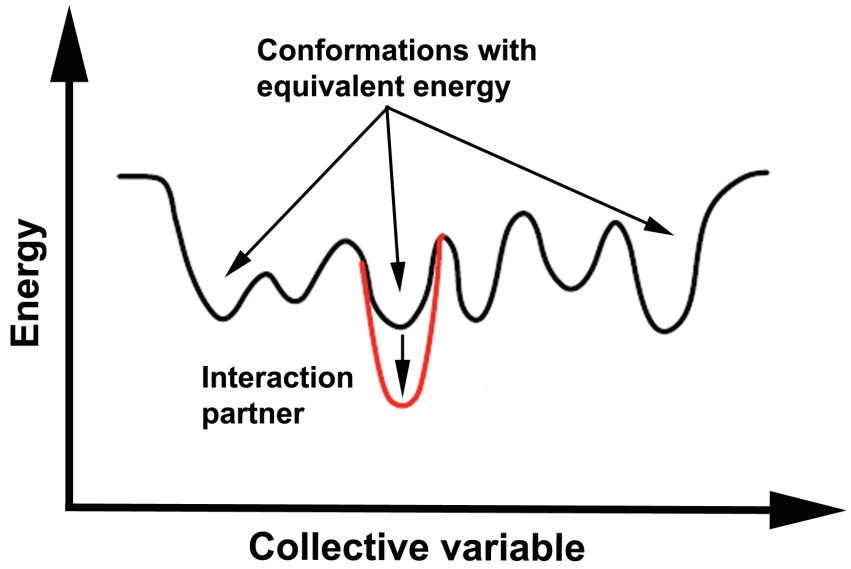

Figure 2.2. A state among many states with equivalent energy can be stabilized upon interaction with the partner.

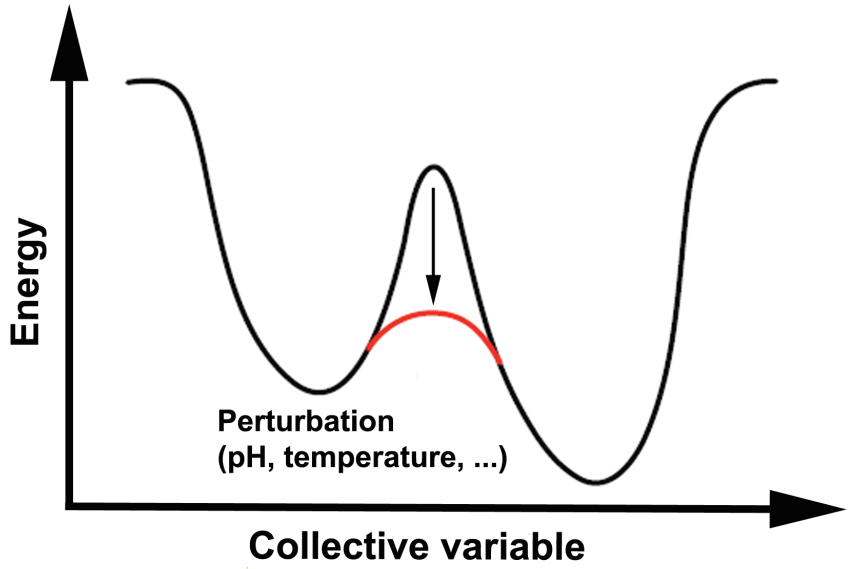

Figure 2.3. Reduced energy barrier between two states due to environmental effects such pH, temperature etc.

The environmental changes such as pH, temperature can also induce alternations on the energy surface; such as the reduction of high energy energy barrier between two states, as shown in Figure 2.3.

It is not easy to visualize the energy landscape of a protein. For a system of N atoms, the energy is a function of $3N$-6 internal and $3N$ Cartesian coordinates [10]. However, it is possible to reduce the complexity of this multidimensional space into a lower dimensional one by using "collective variables", which can be selected for example heuristically, by trial and error or by choosing essential variables extracted from the methods like normal mode analysis or principal component analysis, etc. [6].

2.2. Experimental Techniques for Determination of Protein Structure and Conformational Changes

As a conventional technique in protein structure determination, X-ray crystallography provides atomic-resolution picture of macromolecules and gives information

about the mobility of the protein by means of temperature factors [13]. Advances in crystallographic methods lead to crystal structure isolation of structures as large as 80S eukaryotic ribosome complex (with molecular mass about 3.2 MDa) with 3.0 Å resolution [14]. Still, obtaining high resolution structure becomes difficult with increasing molecular size, using this technique [13]. Moreover, crystallography provides only the average conformation for a specific crystallographic condition, thus gives a limited knowledge about the conformational space of the protein [1, 13]. Nucleic magnetic resonance (NMR) spectroscopy is another experimental technique that provides information about the functional dynamics of the proteins providing tens of conformations [15]. However, conventional solution NMR is limited with proteins of relatively small in size, up to 50 kDa [1, 15, 16]. The problem lies in the deterioration of both sensitivity and resolution of solution NMR spectra due to the line broadening of peaks as the molecular weight of the studied structure increases. To overcome this issue, NMR techniques such as transverse relaxation-optimized spectroscopy, magic-angle spinning solid-state NMR can be employed to gather information about the structure and the dynamics of larger systems [16]. The structures determined by these experimental techniques are deposited in Protein Data Bank [17], which currently holds over 100,000 entries including membrane proteins.

Structures of large assemblies such as molecular chaperones and ribosome can be visualized using cryo-electron microcopy (cryo-EM). The molecular structure is determined from the three dimensional (3D) density maps obtained from cryo-EM experiments [18]. Depending on the map resolution, the data obtained from cryo-EM can be further investigated using rigid-body fitting of atomic coordinates from X-ray crystallography or NMR, segmentation algorithms, ab initio model building. The resolution of 3D cryo-EM reconstruction has been constantly improved over the years and currently, 2908 maps of resolution ranging from 80 to 2.8 Å are deposited in EMDataBank [18].

Small angle solution X-ray and neutron scattering can be used as a complementary technique in structural biology, in terms of revealing oligomeric states and domain organizations in solution. It has also applications in low-resolution shape model-

ing, intrinsic disorder, protein-protein or protein-nucleic acid assembling processes [19]. There are also other techniques for studying protein conformational changes; namely hydrogen-deuterium exchange, time-resolved X-ray crystallography, fluorescence spectroscopy, inelastic neutron scattering [1].

2.3. Computational Techniques for Conformational Sampling and Transitions

2.3.1. Molecular Dynamics (MD) and Related Methods

MD is a valuable tool for investigating protein's behavior in a solvent (explicit or implicit) and sampling its conformations in atomistic resolution. It uses molecular mechanics force field, i.e. an empirically derived potential energy function that describes all molecular interactions [20]. Protein's motion with respect to time is determined by numerically solving Newton's equation of motions for a system of interacting particles. One drawback of this atomistic approach is that classical MD simulations still have high computational cost; typically months are needed for conducting a standard, "long enough" MD simulation, together with the need of significant time for analysis of the massive data generated [2].

The use of graphical processing units (GPU) resulted in accelerations between 10- and 1000-fold over central processing units-only implementations [20]. Still, the general purpose GPU-accelerated MD simulations do not provide a match to Anton, a special-purpose machine, which is designed and constructed to run MD simulations, resulting in great acceleration in simulation times [21]. For example, by conducting MD simulations using Anton, it was possible to reach millisecond level of simulated biological time for bovine pancreatic trypsin inhibitor with 58 residues [22]. Moreover, the ability of MD force fields to reproduce the true potential energy surfaces of proteins is still questioned, and additional refinements of parameters are needed in order to increase the accuracy of current force fields in describing long term structural dynamics of proteins [2,20,23].

Another issue is that, classical MD may provide limited conformational sampling

due to slow crossing of energy barriers [24, 25]. To overcome this problem, enhanced sampling methods are developed to increase the efficiency of MD. Targeted MD simulation is one of these methods, driving the protein from one structure (conformation) to another one by an external force [26]. Accelerated MD [27], metadynamics [28], replica-exchange MD [29] are among the methods that smooth the energy landscape with a biasing potential.

Coarse-grained MD (CG MD) is another tool to study the dynamics of especially large biomolecules. Using this method, it is possible to simulate mesoscopic systems for longer time scales [30]. CG MD models are based on the representation of the system as pseudo-atoms. For example, in MARTINI force field [31], the structure is modeled as one pseudo-atom for four heavy atoms. Then the dynamics is simulated using a potential function (simulation or experimental data based). The microsecond dynamics of viral capsids [32], protein-induced membrane bending by six N-BAR molecules [33] and the ribosome [34] are among the systems that are studied by CG MD. However, parametrization of CG MD potential and also the ability of this model to capture pairwise interactions are still questioned [30].

MD trajectories can also be used to create an ensemble through pharmacophore modeling, which have been proved to be successful in discriminating known HIV-1 inhibitors from drug-like non-inhibitors in docking studies [35]. In Zacharias' study, principal component analysis (PCA) is applied on MD trajectory and during ligand–receptor docking, the protein is allowed to relax in the direction of the soft modes obtained from PCA [36].

2.3.2. Normal Mode Analysis (NMA) and Related Techniques

Although the timescale reached by MD is biologically significant to describe protein folding phenomena (e.g. millisecond scale) for proteins having residues less than 100, there is still room for simplified and computationally efficient methods. This is especially important for studying large system sizes and systems undergoing large conformational changes, which take place in longer time scales [2]. For this purpose, the

popularity of normal modes and related techniques is still increasing in the study of biological macromolecular dynamics [2, 37].

2.3.2.1. Classical NMA. Classical NMA is a full-atom approach that analyzes vibrational motions of the native structure around a local energy minimum. In classical NMA, any atomistic empirical potential energy function can be represented as a sum of the quadratic terms in the displacements to describe the system. The assumption is that, the system displays solid-like behavior, which is in fact the case for proteins at 100-200 K [38–40]. After performing a thorough energy minimization of the folded protein structure, vibrational modes of the system are determined by solving eigenvalue problem for the force constant matrix together with atomic masses, resulting in $3N$ atoms-6 non-zero normal frequencies and eigenvectors for a system of N atoms in this harmonic regime [2, 41]. In the high frequency modes, the displacements occur mostly in covalently bonded atoms (local motions), conversely low frequency modes yield the collective movements involving large parts of the structure [42]. However, the computational efficiency of classical NMA decreases as N becomes larger [2].

Miloshevsky and Jordan employed normal mode following technique [43] combined with a Monte Carlo (MC) scheme on gramicidin A [44], KcsA K+ channel [45] and ApcT amino acid transporter [46] for investigating conformational transitions without providing the closed structure. The technique uses all-atom normal modes and CHARMM22 potential energy.

Perahia and coworkers explored the functional motions of HIV-1 protease (homodimer with 198 residues) by defining consensus normal modes, from multiple-minima NMA [47]. NMA is applied on a set of energy-minimized structures obtained from short MD simulation. A new consensus covariance matrix is obtained from averaging all individual covariance matrices and the modes and their frequencies are extracted from this matrix. They reported that low frequency consensus modes yield robust description of biologically relevant motions upon the binding of inhibitors.

In another study [48], transition pathway and free energy profile of adenylate kinase, calmoduline and p38α kinase are investigated using the combination of NMA in internal coordinates [49] and umbrella sampling MD [50,51] in the presence and absence of the ligands. Using this method, it was possible to obtain transient conformations, consistent with previous experimental and computational studies. Moreover, calculated free energy profiles reveal the intrinsic flexibility of adenylate kinase and calmodulin without obvious energy barriers.

2.3.2.2. Coarse-Grained NMA. Coarse grained models of proteins suggested that the lowest frequency normal modes remain robust despite the elimination of interatomic interaction details within a protein, meaning that they are a global property of the structure and shape [2, 52–57]. Following the pioneering work of Tirion [58], in which Lennard-Jones and electrostatic interactions are replaced by Hookean springs for atom pairs within a cutoff distance, the elastic network model (ENM) [3–5] has been introduced as a coarse-grained approach for studying protein dynamics. Due to coarse-graining that can be applied at hierarchical levels [57], the computational cost can be significantly reduced compared to classical NMA so that application to very large complexes, i.e. supramolecules, becomes possible. In standard or residue-based ENM, the α-carbons of the residues are taken as nodes and the node pairs that fall within a specified cutoff distance are connected by harmonic springs with a uniform force constant. If the residue fluctuations are assumed to be isotropic, the model is Gaussian Network Model (GNM) [3, 59]. If the anisotropic effect is taken into account, one can have information about the three dimensional character of the system, thus this model is known as the anisotropic network model (ANM) [5].

The simplicity of the ENM has given rise to its wide application in analysis of conformational dynamics and transitions. Examination of proteins with known open and ligand bound-forms [60–63] have revealed that one or a few low-frequency normal modes largely overlap with the motion of protein occurring during the transition from the unbound to the ligand-bound state. Inspired from these results, selecting the native structure as a starting point and searching for possible conformations along the

displacement of normal modes has been extensively studied for many cases.

In a study by Ahmed *et al.* [64], directional information obtained from ENM was incorporated into a geometric simulation algorithm to explore biologically relevant conformational transitions in proteins. In this method, graph theoretical approach Floppy Inclusions and Rigid Substructure Topography (FIRST) [65], was used to determine the rigid and flexible regions of the protein. From the resulting coarse grained representation of the structures, the low frequency ENM modes were obtained. In the generation of new conformations, rigid parts of the protein were allowed only in rigid body motion, whereas links between them were fully flexible. Backbone motions were biased toward the direction of random linear combination of low frequency modes and side chains were moved according to experimental rotamer information. By this method, in 7 out of 8 cases, conformations similar to the ligand bound structures (with RMSD less than 3.1 Å) were able to be sampled starting from the initial structure by applying multiple unbiased runs, which results in numerous produced conformers. If the information of the target structure is known, that can be also included into the simulation. In this work, the largest protein was cytrate synthase having 860 residues and the smallest one was calmodulin with 148 residues. Calmodulin was the unsuccessful case in unbiased sampling.

For large conformational changes, Jimenez-Roldan *et al.* [66] combined ENM, rigidity analysis of FIRST and geometric simulation of protein motion. They explored conformational change along normal mode vectors by modeling the protein structure as a coarse grained simple harmonic network to produce an eigenvector for low frequency motion. Using FIRST rigidity analysis, non-covalent interactions were identified and the protein's rigid and flexible parts were labeled. Finally, geometric simulation was used to explore conformational space using a low frequency normal mode and rigidity information. They applied the method on a set of six proteins with different sizes (from 58 to 1605 residues) and structural characteristics and they have showed that the method identified intrinsic motion specific to the protein and its amplitude limits. Large amplitude motions can be explored until these are limited by bonding or steric constraints.

Low frequency normal modes obtained from residue-based ENM was used to enhance conformational sampling of maltose binding protein and nucleotide binding domains of a maltose-transporter in temperature-accelerated molecular dynamics [25]. Single or a combination of two to three collective modes were found enough to describe the conformational transition in temperature-accelerated MD generated pathways.

Normal modes are also used in targeted sampling, once the end conformers are available. For example, Kim *et al.* introduced elastic network interpolation (ENI) [67] and its extension rigid-cluster ENI [68] to generate transition pathways between two end states. The method was based on a coarse-grained ENM and generated intermediate structures by interpolating the distance of connected residues between two distance values specified by end conformations, by using a simple quadratic cost function. In rigid-cluster ENI, point masses were replaced by rigid cluster in order to reduce computational cost. ENI was also used in Feng *et al.*'s work [69] for generating transition pathway between adenylate kinase open and closed form, and the steric clashes in the intermediate structures were removed using CHARMM energy minimization. Using this method, it was possible to obtain intermediates close to the available experimental structures (such as the intermediate structure where LID domain is closed and NMP domain is open).

Plastic network model [70] is another method developed for studying conformational transition between two end states, following lowest elastic energy path. The method was an extension of ENM since it adapted a free energy function based on elastic theory, introduced by Tirion [58]. The method was applied to adenylate kinase by taking open and closed states of the protein as end points. Along the transition pathway, LID domain closure preceded NMP domain closure and the intermediates close to experimental structures were generated (again, closed LID-open NMP domain structures).

Zheng and Brooks introduced an ENM based method for predicting conformational changes of proteins by using a starting structure and distance constraints of the end state [71, 72]. In this method, a perturbation was given to the harmonic Hamilto-

nian that additionally included pairwise distance constraints. As a result, the predicted conformational change, which was a linear combination of multiple low-frequency normal modes, was computed as a response displacement induced by the perturbation. For most of the 22 test cases, the transition from initial to end states was successful. The method was extended in a following study [73], to accurately predict both the direction and amplitude of the conformational change. The conformational search was driven by iterative minimization of the error of fitting the distance constraints and enforcing the restraint of low elastic energy. The method was tested on 16 pairs of protein structures and structures near to native end state were obtained (RMSD 1-2 Å).

Another hybrid methodology called ANM-MC [74, 75] combines collective moves obtained from ENM and local moves via knowledge-based MC simulation for generating transition pathways. In the targeted version of the algorithm, the structure was iteratively deformed along the direction of the collective mode that overlaps with the conformational transition direction and minimization of conformational energy was performed using MC simulation. The method successfully approached the target and produced intermediates for a set of 10 proteins undergoing large hinge-bending type motions [75]. Moreover, the unbiased version of ANM-MC, which uses only the radius of gyration of the target as input, was successful in predicting the closed state for eight (out of 10) of these hinge-bending proteins.

Kirillova *et al.* [76] and Al-Bluwi *et al.* [77] used NMA as a bias in the conformation exploration by motion planning based 'Rapidly-exploring Random Tree' algorithm to compute a conformational transition path for a large amplitude conformational transition. Coarse-grained representation of the protein (one bead per tripeptide) is used for further reducing computational cost. The method is tested on ten proteins including adenylate kinase and GroEL.

Collective molecular dynamics (coMD) is another method combining ENM and MD, introduced by Gur *et al.* [78] for generation of transition pathways between two end points. The method is based on deforming the structure along the collective modes

predicted by ENM that are selected by Monte Carlo/Metropolis algorithm from the pool of all accessible modes. Then, the energetics are provided by means of full-atomic MD simulation protocol. The method is applied for exploring adenylate kinase conformational transition, using open and closed states of the enzyme as end points.

Das *et al.* also developed a method called ANMPathway [79] that uses a simple two-state anisotropic network model for constructing an energetically favorable pathway between two endpoints of a conformational transition. The endpoints are known experimental structures, which are labeled as stable states of the system. The ENM representation of these end states is used in the construction of two-state potential, which has a cusp hypersurface in the configuration space. Linear interpolation between two end states is performed in order to identify a structure which has equal energy values according to both surfaces. This structure is treated as the transition state. Starting from this transition state, steepest descent algorithm is applied on both surfaces and obtained conformers constitute the transition pathway. The method is applied on adenylate kinase, ATP-driven calcium pump SERCA, leucine transporter and glutamate transporter, yielding good agreement with other similar methods and MD.

2.3.2.3. Applications of ENM in Small Ligand-Protein Docking. One of the challenging problems in drug design studies is the incorporation of flexible nature of proteins. ENM has gained recognition for solving this problem, in both protein-small ligand and protein-protein docking for including especially backbone flexibility into docking applications. In this section, we will only focus on usage of ENM in small ligand docking.

One of the earlier examples was Cavasotto *et al.*'s work of protein-ligand docking, where receptor flexibility was incorporated through deforming the receptor structure along a combination of "relevant" modes which are related with the selected regions of interest [80]. These relevant modes were not necessarily the low-frequency modes. In their method, the original structure was deformed along the combination of relevant normal modes, and the receptor ensemble was complexed with known binders.

This was followed by global energy minimization for the side chain optimization using flexible ligand-flexible side chain approach, employing biased probability Monte Carlo method. Finally, the validation procedure consisted of receptor ensemble docking against the generated multiple receptor conformers. Using this methodology, Cavasotto *et al.* reported the increase in the docking accuracy in cAMP-dependent protein kinase dockings and the enhancement of the discrimination between binders and nonbinders. However, Dietzen *et al.* [81] have shown that application of ENM does not significantly improve the docking performance in the case of local ligand-specific induced fit movements. In their work, they used "binding pocket restricted" normal modes for reconstructing the holo structures of 433 proteins from apo structures.

In May and Zacharias' protein-ligand docking work in 2008, ENM was used for accounting receptor global backbone flexibility, by means of relaxation in a few precalculated soft modes [82]. In 2011, Leis and Zacharias included receptor backbone flexibility obtained from ENM in grid-based protein-ligand docking [83], and implemented the method in protein-ligand docking software AutoDock [84, 85]. The basic idea was that instead of dealing with one input structure for the receptor, the code was modified to include the structural information obtained from several deformations along a slow (soft/low frequency) normal mode. The method was tested on apo protein kinase A (PKA) dockings, which indicated significant improvement compared to rigid PKA docking [83].

Rueda *et al.* used ENM on heavy atoms of residues within a distance of 10, 15, 20 Å from the cognate ligand for generating multiple receptor conformations to be used in docking studies [86]. The ENM normal mode space (slowest 100 modes) was used for the generation of Cartesian displacement using Metropolis MC algorithm with a Hamiltonian. At each iteration, one of the 100 modes was displaced randomly following Metropolis criteria. Their algorithm focused on equilibrium conformations of 2 Å RMSD near the original crystallographic structure. An RMSD-based clustering was performed on the side-chains of heavy atoms of the generated structures, using thresholds ranging between 0.5-1 Å. As a result, in a benchmark containing 28 proteins, their method predicted near native ligand poses 20% more efficiently compared to single

receptor docking.

In another study by Akten *et al.* [87], a reverse-mapping technique was used on cyclophilin A (CypA) by applying ENM at different resolutions (atomistic and coarse-grained). The original structure was deformed along each of the seven lowest frequency modes, both in negative and positive directions. Generated conformers were energetically minimized using implicit solvent method in order to eliminate steric clashes and then used for ensemble docking. Realistic conformations with energies comparable to the crystal structure were generated and further used in docking, which lead to the prediction of correct binding modes.

2.3.3. Other Techniques for Conformational Search

Apart from MD and NMA, there are other computational methods that explore the conformational space. One example is the CONCORD method, which is based on the geometrical description of a given protein structure [88,89]. Geometrical description of the protein consists of topological constraints such as bonds, angles and noncovalent constraints e.g. hydrogen bonds, salt bridges and hydrophobic clusters. Using these constraints, the structures are iteratively built starting from random coordinates, until all constraints are satisfied [89]. This method is also used in Seeliger *et al.*'s study in 2010 [90] for the prediction of holo structure using the knowledge of apo structures, the ligand and the radius of gyration (RG) of the holo structure. the algorithm uses CONCORD methodology for generation of structures, AutoDock VINA [91] for ligand docking and MD for refinement. The method is tested on ten proteins with conformational arrangement up to 7 Å. In 8 out of 10 cases, close-to-native ligand binding poses were obtained. One limitation of the protocol was that it was unable to predict a change in the secondary structure between apo and holo states, where a helix refolds into a loop in one case.

Yesylevskyy *et al.* employs hierarchical clustering of correlation patterns technique for domain identification and treat these domains as rigid-body like clusters. Then the clusters are allowed to rotate relative to the other for simulating slow pro-

tein dynamics in globular proteins [92]. It is applied on ten hinge-bending proteins to blindly search for their closed states, resulting in realistic closed conformation for 8 out of 10 cases.

Another study concerning the prediction of ligand binding motions in hinge-bending proteins describes the conformation explorer technique [93]. Starting from open monomeric structure, holo structures are predicted for five different proteins by first identifying the hinges in the structure, then applying Euler rotation to one of the domains about the hinge and finally applying short MD for equilibrating protein-ligand complex and scoring the generated structures using a fitness function, which favors closed or holo structures. Before the application of short MD, the ligand is docked using AutoDock software.

Perturbation-response scanning technique (PRS, [94]) introduced by Atilgan and coworkers is used together with MD to display conformational motions experienced by various proteins such as hinge-bending, shear, allosteric [95, 96]. Based on linear response theory, PRS consists of the systematical application of forces at singly selected residues, and then measuring the linear response of the whole protein by means of the magnitude and directionality of the displacements that the residues exhibit [94]. Application of the method on a set of 25 proteins revealed that in some proteins binding-induced conformational change may be the consequence of the perturbation of residues confined in a specific region and in others the ensemble of residues can be located on different regions on the protein. Moreover, PRS-calculated atomic displacement vectors are overlapping with the ones obtained from experimental data.

Recently, Gipson *et al.* have introduced a hybrid method for rapid large scale conformational analysis, called as Structured Intuitive Move Selector [97]. The method is applied for determination of active residues for cyanovirin-N, exploring conformational changes of ribose binding protein and transient conformational state of maltose-binding protein.

2.3.4. Computational Techniques Used for Investigating Ribosome Conformational Search and Dynamics

The supramolecular assembly ribosome (with mass above 2 MDa), which synthesizes proteins using the genetic information of messenger ribonucleic acid (mRNA) in the translation process, is a macromolecular machine of interest and widely studied by experimental and computational means. Although available experimental data such as cryo-EM and x-ray crystallography reveal functionally important conformational changes during translation (e.g. ratchet-like relative rotation of two subunits namely 50S and 30S or large displacement of L1 stalk region in bacterial ribosome [98–101]), these data mostly provide static pictures of this large assembly. Computational methods combined with experimental data provide a complementary approach for understanding the intermediates and dynamics of ribosome during the translation process.

One of the earliest studies that employ computational techniques for studying ribosome dynamics is performed by Tama *et al.* [102] using ENM analysis, rotation-translation block method [52], cryo-EM and x-ray crystallographic data. In this work, coarse grained ENM analysis (based on alpha carbons and phosphate atoms) revealed that lowest-frequency normal modes correspond to functionally important motions. For example, lowest mode describes the motion of L1 stalk and third mode corresponds to ratchet-like rearrangement of subunits, which is experimentally observed in response to the binding of elongation factor G [99]. The results hint to the robustness of functionally important motions, which is intrinsically accessible to ribosomal structure.

Wang *et al.* also reported the global motions of 70S ribosome, revealed by ENM [103]. Their results agree with Tama *et al.* [102] especially for large domain motions. However, several modes are also identified that facilitate the exit (E)-transfer RNA (tRNA) exiting from the assembly. Also, the adenine (A)-site tRNA and peptidyl (P)-site tRNA are described to be positively correlated, hinting that their translocation occurs simultaneously, whereas E-site tRNA translocation possibly does not occur simultaneously.

ENM is also used for investigating the dynamics of 70S with and without tRNA, elongation factor Tu and the ribosomal proteins [104]. The results revealed the ratchet-like rotational motions of the subunits, head rotation of small subunit and L1-L7/L12 stalk movements are among the global motions of the ribosome, which are present even in the absence of the ribosomal proteins, suggesting that these are topology-inherited motions.

Collective dynamics of ribosomal nascent polypeptide exit tunnel is also investigated using ENM by examining the individual lowest-frequency modes and the linear combination of these modes [105]. The entrance, the neck and the exit domains of the tunnel exhibits different domain motions; entrance and exit domains moving in the exit direction whereas neck domain displaying rotational motion. The analysis also revealed the anti or non-correlated motions of L4 and L22 proteins located at the narrowest region of the tunnel, which may have an important role in polypeptide gating mechanism.

The global motions of 70S ribosome is also explored by performing 500 ns of coarse-grained molecular dynamics, using a low resolution anharmonic network model [34]. Simulation revealed the anticorrelated motion of L7/L12 and L1 lateral stalks, widening of the tRNA cleft and the rotation of the small subunit together with the movement of L1 stalk.

Targeted MD simulations that were used together with available experimental data shed light onto the movement of tRNA through the ribosome [106, 107]. Moreover, excited states of ribosome translocation such as P/E-A/P* and P/E-A/P positioning of two tRNAs in the presence of elongation factor G were investigated via MDfit methodology [108], which employs modified MD simulations to generate configurations of excited states integrated with information obtained from x-ray crystallography, cryo-EM and biochemical data.

Ribosome dynamics is also studied by Seo *et al.* using hybrid elastic network model (HENM, [109]). In this coarse grained method, all representative atoms are

connected via harmonic springs like in the case of classical ENM, then the rigid clusters are identified (such as secondary structures), and finally flexible regions (e.g. hinges and loops) are left as point masses. Besides extracting the slow modes such as ratchet-like motion, L1 stalk movement using HENM, hybrid elastic network interpolation (HENI) based on the modes from HENM were used for simulating the transition pathway of ribosome,by taking reference the two-step mechanism for translocation proposed by Frank and Agrawal [99]. In this proposition, first step is the rotation of 30S subunit with respect to 50S, accompanied by the motion of tRNA as well as mRNA towards P and E sites from A and P sites, respectively. Second step is the sole rotationof 30S back to initial position.

3. MATERIALS AND METHODS

3.1. Anisotropic Network Model (ANM)

The anisotropic network model [5, 56, 57, 60] is an ENM based on inter-residue contact topology of native conformations of proteins, which are obtained from experimental techniques such as x-ray crystallography, nuclear magnetic resonance (NMR) or cryo-electron microscopy. This coarse-grained method incorporates the anisotropy of residue fluctuations in three dimensions, giving information on both the magnitude and the direction of collective motions using normal mode analysis.

In ANM, the structure is represented as a network of nodes, generally coarse-grained at one-node-per-residue level. Node pairs that fall within a specified cutoff distance are connected via harmonic springs of uniform force constant. The nodes are commonly located at the Cα atoms of amino acids and the P atoms of nucleotides [5, 103], resulting in a coarse-grained representation of the structure.

ANM potential energy for the system of N nodes is the summation of all harmonic interactions in the structure:

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

In Eqn 3.1, γ is force constant, R_{ij} and R_{ij}^0 are the instantaneous and equilibrium distances between nodes i and j, respectively (where $i, j = 1 \ldots N$). Based on the cutoff distance R_c, the heavyside step function $h(R_c - R_{ij})$, is equal to one if $R_c \geq R_{ij}$ and zero, otherwise.

The potential energy in Eqn 3.1 can be also expressed in the following form:

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

Where $\Delta\mathbf{R}$ is the positional fluctuation vector for N nodes ($3N$-dimensional), $\mathbf{H}$ is the ($3N$ by $3N$) Hessian matrix, whose elements are determined from the second derivative of the potential energy with respect to mass weighted coordinates. The mass of each node is taken as unity in classical ANM. Each "super element" ' of $\mathbf{H}$ is defined as:

$$H_{ij} = \begin{bmatrix} \frac{\partial^2 V_{ij}}{\partial x_i \partial x_j} & \frac{\partial^2 V_{ij}}{\partial x_i \partial y_j} & \frac{\partial^2 V_{ij}}{\partial x_i \partial z_j} \\ \frac{\partial^2 V_{ij}}{\partial y_i \partial x_j} & \frac{\partial^2 V_{ij}}{\partial y_i \partial y_j} & \frac{\partial^2 V_{ij}}{\partial y_i \partial z_j} \\ \frac{\partial^2 V_{ij}}{\partial z_i \partial x_j} & \frac{\partial^2 V_{ij}}{\partial z_i \partial y_j} & \frac{\partial^2 V_{ij}}{\partial z_i \partial z_j} \end{bmatrix} \tag{3.3}$$

It can be expanded as:

$$H_{ij} = \frac{-\gamma}{R_{ij}^2} \begin{bmatrix} (x_j - x_i)(x_j - x_i) & (x_j - x_i)(y_j - y_i) & (x_j - x_i)(z_j - z_i) \\ (y_j - y_i)(x_j - x_i) & (y_j - y_i)(y_j - y_i) & (y_j - y_i)(z_j - z_i) \\ (z_j - z_i)(x_j - x_i) & (z_j - z_i)(y_j - y_i) & (z_j - z_i)(z_j - z_i) \end{bmatrix} \tag{3.4}$$

To extract normal modes of elastic network, symmetric $\mathbf{H}$ matrix is diagonalized into the form:

$$\mathbf{S}^T\mathbf{H}\mathbf{S} = \mathbf{\Lambda} \tag{3.5}$$

$\mathbf{\Lambda}$ is ($3N$ by $3N$) diagonal matrix with diagonal elements are the squared normal mode frequencies; $\mathbf{S}$ is an orthogonal matrix with its columns being the normalized

eigenvectors that give the directions (or shape) of normal modes. The orthogonal transformation of **H** matrix results in $3N-6$ internal normal modes together with three translational and three rotational modes. The eigenvalues for translational and rotational modes are equal to zero. Thus, overall motion can be expressed as the summation of $3N-6$ internal normal modes. Diagonalizing **H** for large systems such as ribosome can be computationally exhaustive. To overcome this issue, the eigenvalue problem can be solved using software package BLZPACK [110] with block Lanczos [111] algorithm, which is efficient especially for sparse matrices like **H**.

The mean-square fluctuation of i^{th} node based on the k^{th} normal mode is obtained from the following expression:

$$\left\langle \Delta R_i^2 \right\rangle_k = \left(\frac{k_B T}{\gamma} \right) \frac{\left(\mathbf{u}_{ik}\right)^2}{\lambda_k} \tag{3.6}$$

Where k_B is the Boltzmann's constant, T is the absolute temperature, $\mathbf{u}_{ik}$ is the k^{th} eigenvector for i^{th} node and λ_k is the k^{th} eigenvalue.

In this thesis, we employ a modified version of standard ANM, which is named as the mixed-resolution ENM [112]. In this version, the structure can be modelled as a mixture of low and high resolution regions. In the low-resolution regions, the nodes are placed at the residue centroid coordinates, whereas heavy atoms represent the nodes in the high resolution region. The force constants of the springs connecting any two coarse-grained nodes are scaled based on the total number of atomistic pairwise interactions between residue pairs (bonded or non-bonded) that are within an atomistic cut-off radius of 10 Å for proteins (13 Å for ribosome). The eigenvectors and eigenvalues of H are extracted using software package BLZPACK using the block Lanczos algorithm. In this thesis, we only use low-resolution nodes that correspond to amino acid or nucleotide residues. However, we included this version in the ClustENM methodology, for future possible applications to cases like ligand bound complex structures, where ligand or region of interest can be modeled as high resolution region and the rest of the structure in low resolution.

3.2. ClustENM-I for Conformational Sampling

We developed a new iterative methodology called ClustENM for generating atomistic conformers with possible application to large protein complexes and/or conformational changes. Each cycle or generation of the iterative methodology consists of the following steps:

(i) Energy minimization with implicit solvent model is applied on the starting structure (and each representative conformer at successive iterations). For the minimization, AMBER12 [113] is used with ff03 force field parameters for proteins [114] and ff10 [113] for ribosome. Pairwise generalized Born model [115, 116] is used for the minimization with implicit solvent using 16 Å (30 Å for ribosome) cutoff for non-bonded interactions. Using a modified generalized Born theory based on the Debye-Hückel limiting law for ion screening of interactions [117], the concentration of 1-1 mobile counterions in solution is set to 0.1 M. For proteins, 500 cycles of steepest descent are followed by conjugate gradient with a convergence criterion for the energy gradient (drms) set to 0.01 kcal/mol/Å. For ribosome, instead of setting drms value, maximum step for minimization is set to 1000 to save computational time due to large system size.

(ii) ENM is applied to energetically minimized conformer(s) to extract the slowest modes. The nodes are placed at the residue centroid coordinates (whole structure is modelled as low resolution).

(iii) Eigenvectors $\mathbf{u}_j$ of the m slowest modes extracted from ENM, scaled based on their frequency $\lambda_j^{\frac{1}{2}}$ and are linearly combined using three coefficients $\mathbf{a}_j$ [105]:

$$\mathbf{V}_i = \sum_j \frac{a_j \mathbf{u}_j}{\lambda_j^{\frac{1}{2}}} \tag{3.7}$$

First m slowest modes are taken into account up to the point, where a significant increase (jump) is observed in the eigenvalues of the initial structure.

(iv) Native structure is deformed along the direction vectors $\mathbf{d}_i$ obtained from the combination of slow modes ($\mathbf{V}_i$), by taking $\mathbf{V}_1$ (where all m modes are present)

as reference with a specified deformation RMSD (DF):

$$\mathbf{d}_i = \frac{DF}{N^{\frac{1}{2}}} \left(\frac{\mathbf{V}_i}{|V_1|} \right) \tag{3.8}$$

with N being the total number of nodes. For each residue, the deformation of its centroid is applied on its backbone and side chain atoms, thereby generating new atomistic conformers (resulting in 3^m new conformers based on starting conformer).

(v) Generated conformers are clustered using kclust module of MMTSB toolset [118], based on mutual RMSD values. RMSD cut-off for clusters is set to $0.75*DF$.

(vi) Conformer with the lowest RMSD to the average structure of each cluster is selected as the representative conformer.

The above procedure, starting with Step 1, is repeated for each representative structure in subsequent cycles/generations. In Step 5, the clustering is applied on all generated conformers belonging to generation "g" (where g is equal to zero for the cycle of starting structure, one for first iteration, etc.-see Figure 3.1). Thus, at each iteration, number of clusters grows due to increased number of generated conformers. In the final generation, the weight of each representative conformer is calculated based on the number of elements in its cluster over the total number of generated conformers before clustering. It is assumed that each representative conformer represents a conformational state with the assigned weight.

During the generation of conformers, the case where $a_j = 0$ (Eqn 3.7) for all eigenvectors results in the presence of "parent" (initial structure on which the procedure is applied) among the generated conformers. The parent conformer(s) is (are) clustered together with deformed conformers along $\mathbf{d}_i$, and representatives are selected from all of the clusters, including the parent(s)'s. Thus, the information of parents (i.e. representative structures in earlier generations) is kept in the final generation.

ClustENM Procedure I is applied on a dataset of six diverse structures. Hinge bending proteins adenylate kinase (AK) and calmodulin (CAM) are monomeric; and

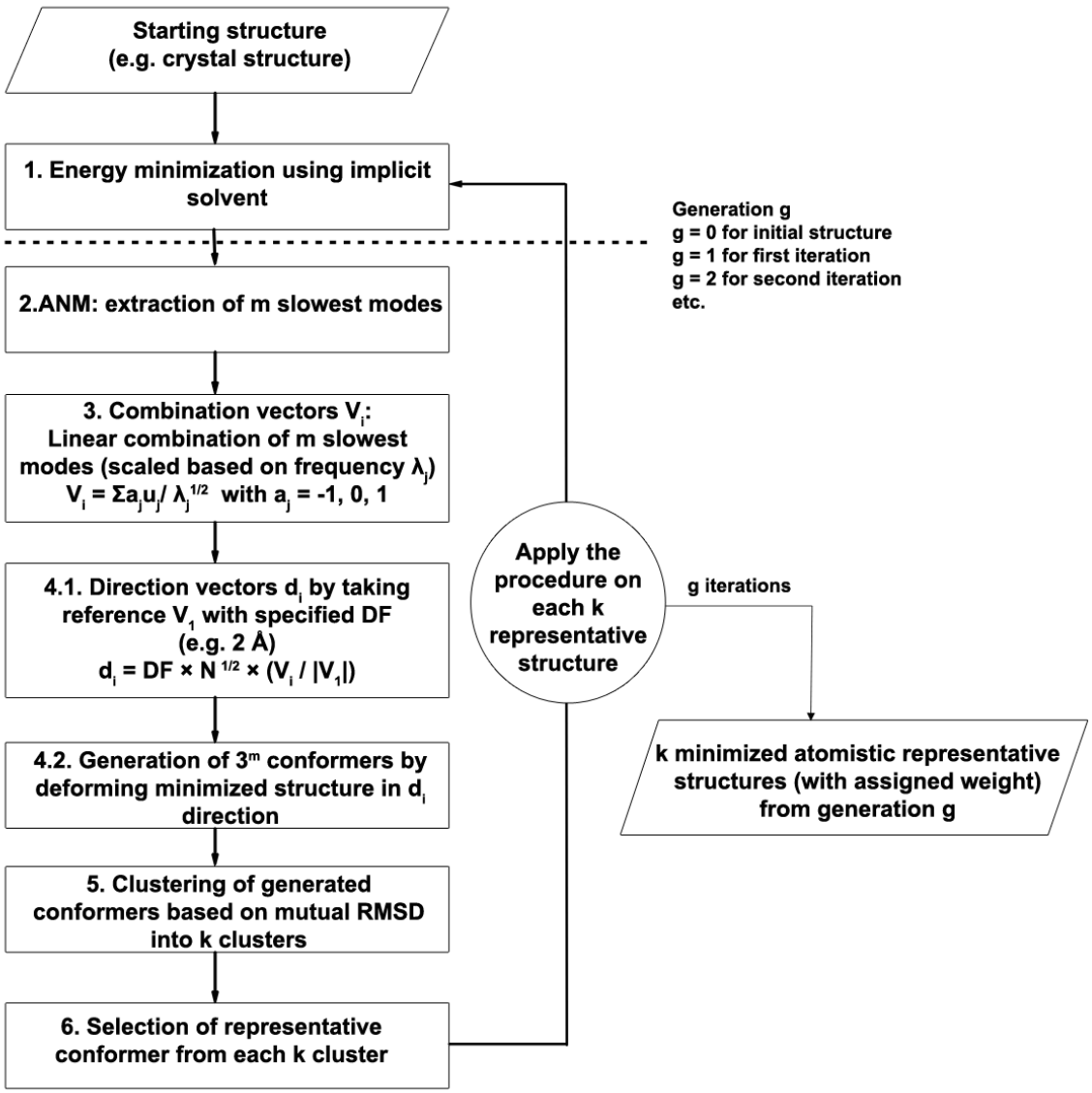

Figure 3.1. Flowchart of ClustENM Procedure I.

HIV1 reverse transcriptase (HIV1-RT) is a heterodimeric protein. Triosephosphate isomerase is a homodimeric enzyme with a catalytic loop and p38 kinase present comparatively localized motions. Finally, 70S ribosome is a supramolecule with diverse global motions. These structures are selected due to available experimental and computational data and to show the applicability of the method to diverse in terms of size and topology. Experimental structures for comparison are chosen based on 90% sequence similarity to starting structures.

The list of these structures (seen in Figure 3.2) and system properties are given in Table 3.1. For AK, the number of experimental structures that have 90% sequence similarity to 4ake/1ake equals to nine. But except 4ake, all of these structures are fully closed and similar to 1ake. Even though we use only two of the AK's crystal structures for our purpose here, AK's conformational transitions have been extensively studied by experimental and computational means, therefore we will compare our results with available data in the literature.

Table 3.1. Structures used in ClustENM-I.

Structure	N	DF (Å)	m	No of conformers	g	Initial structure	No of exp. structures
AK	214	2	3	71	5	4ake	2
				226	5	1ake	
				222	5	closed-NMP	
				113	5	closed-LID	
CAM	144	3	3	195	5	1cll	550
HIV1-RT	958	2	5	50	2	1rtj	155
P38	345	1	5	41	2	1p38	221
TIM	497	1	3	24	2	1tcd	16
70S Ribosome	11336	3	5	101	2	4kdk-4kdj	27

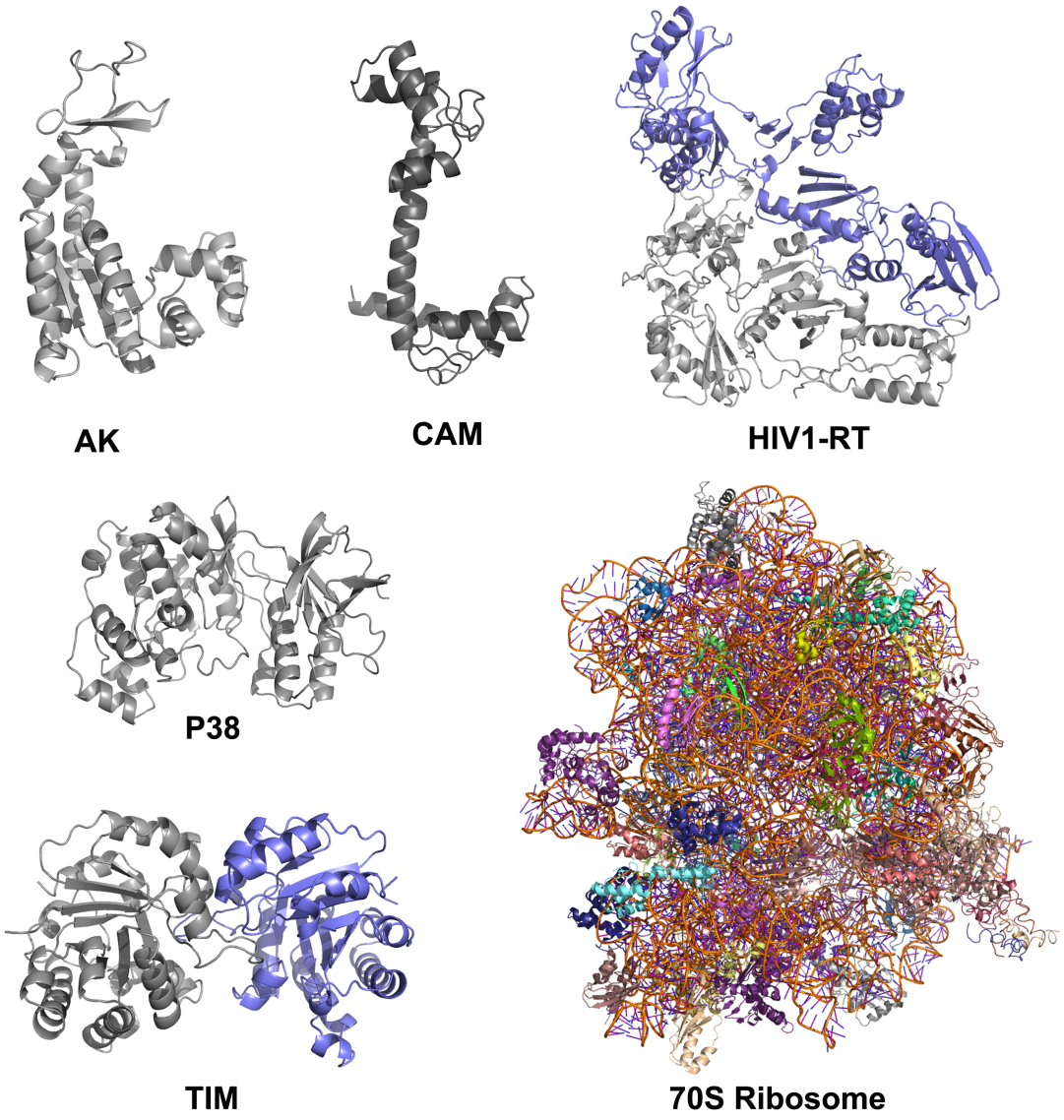

Figure 3.2. Structures used in the application of ClustENM Procedure I.

3.3. ClustENM-II for Docking Applications

A variation of ClustENM is applied on four proteins undergoing large conformational changes for docking applications. Minimization-deformation-clustering scheme is similar to Procedure I, with following modifications:

(i) Same as Procedure I-Step 1.
(ii) Same as Procedure I-Step 2.
(iii) Same as Procedure I-Step 3.
(iv) Similar to Procedure I-Step 4 with a slight modification: Fixed DF of 2 Å is applied without scaling based on $\mathbf{V}_1$.
(v) Similar to Procedure I-Step 5 with a slight modification: RMSD cutoff is set to 2 Å.
(vi) Similar to Procedure I-Step 6 with a modification: clusters containing parent structure(s) are excluded. Representative structures are chosen from clusters different than parent structure(s) in order to sample as many distinct conformers as possible.

The above procedure, starting with Step 1, is repeated for each minimized representative structure in subsequent stages/iterations, as seen in the flowchart given in Figure 3.3.

The procedure starts by performing ENM on the minimized native unbound protein conformation. In consequent iterations (or generations) ENM is performed on each representative structure of the generated clusters if the blind search mode (without any filtering or constraints) is used. If an energy criteria is considered (energy-based selection), in Step 6, representative conformers that have lower energy compared to minimized open/unbound protein (starting structure) are selected for the generation of next cycle conformers. At the end of the procedure, additional post-filtering criteria can be applied. For example, we use RG of the apo structure for filtering out the structures having RG larger than the apo, since our dataset consists of hinge-bending proteins which undergo conformational transition from open to closed structure. ClustENM

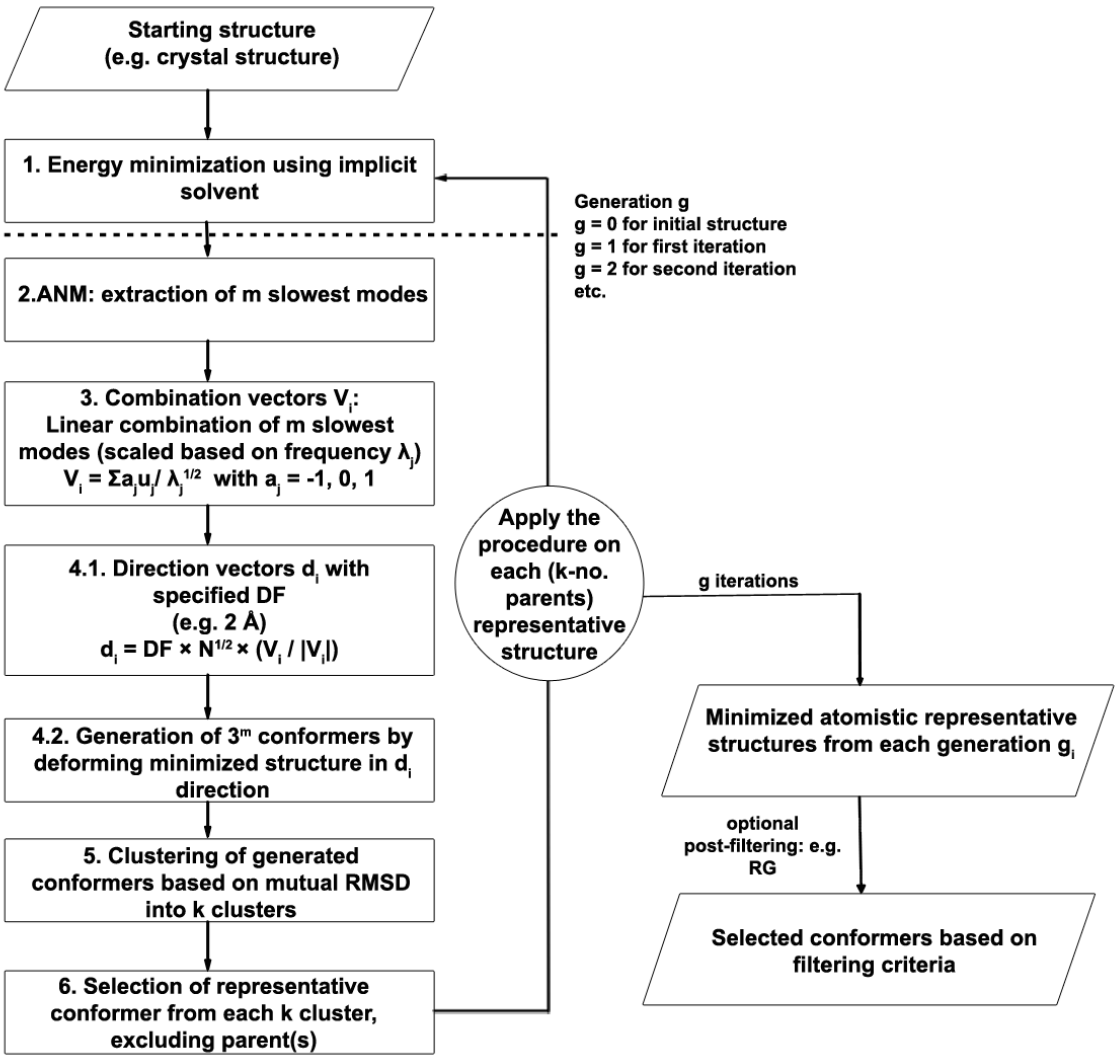

Figure 3.3. Flowchart of ClustENM Procedure II.

Procedure II is applied on proteins listed in Table 3.2. Additionally, the open and closed (ligand bound) states of the structures are shown in Figure 3.4

Table 3.2. Proteins and their ligands used in ClustENM-II.

Protein	PDB id	Number of residues	RMSD between open and closed structure (Å)	Ligand
AK	4ake (open)	214	7.1	Bis(adenosine)-5'-pentaphosphate (AP5) (in 1ake)
Lysine-arginine-ornithine-binding protein (LAO)	2lao (open)	238	4.77	Lysine (1lst)
Dipeptide-binding protein (DBP)	1dpe (open)	507	6.5	Glycyl-leucine (1dpp)
Biotin Carboxylase (BC)	1dv1 (open)	894	4.1/4.6 (monomeric / dimeric)	Adenosine triphosphate (ATP) (1dv2)

3.4. Molecular Dynamics Simulations

MD simulations in explicit solvent are carried out for TIM, p38 and HIV1-RT in order to compare the conformational space explored by MD and ClustENM Procedure I. Inhibitor bound TIM complex simulations are also performed in order to investigate the effect of the ligand on conformational dynamics of the enzyme.

Three independent 100 ns apo TIM and three independent 100 ns inhibitor-bound TIM simulations are performed with the ff03 force field [114]; one 100 ns and one 50 ns p38 kinase; and two 100 ns hiv1-rt runs with ff10 force field parameters in

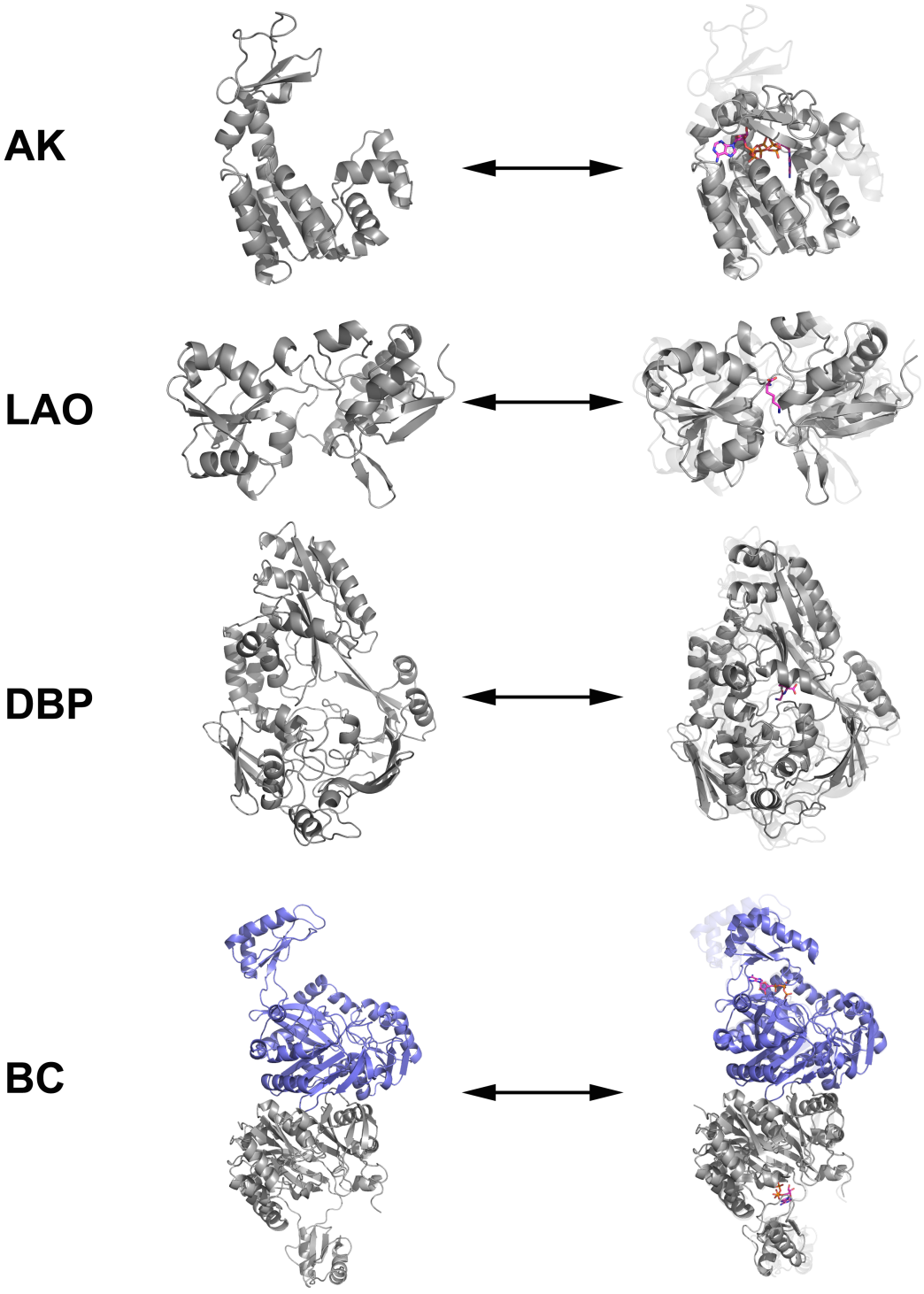

Figure 3.4. Open (left) and ligand bound (right) conformers of proteins used in application of ClustENM Procedure II.

AMBER12 [113]. The force field parameters for the inhibitor, which is a benzothiazole derivative 2-2(2-(4-aminophenyl) benzothiazole)-6-methylbenzothiazole-7-sulfonic acid, sodium salt, are obtained using antechamber module of AMBER, after determining RESP partial charges using Gaussian03 [119] at the B3LYP/6-31+G** level.

In all runs, energy minimization is performed starting with 500 cycles of steepest descent algorithm, followed by conjugate gradient with a convergence criterion for RMS gradient per atom set to 0.01 kcal/mol/Å. Initial velocities are assigned according to Maxwellian distribution at 10 K, then the temperature is gradually increased to 300 K. All runs are carried out as NPT simulations with isotropic scaling at 300 K and 1 atm, using Langevin dynamics for first 1 ns with collision frequency 1 ps-1 in order to homogenously dissipate the heat, then switched to the weak-coupling algorithm [120] for both temperature and pressurefor the rest of the simulation, with default relaxation time constants of 1 ps each. The truncated octahedron periodic box is filled with TIP3P water molecules [121] and neutralized with counter ions (Cl- for all proteins except p38 MAP kinase). A time step of 2 fs is used due to the application of SHAKE algorithm [122]. Ewald summation technique with the particle-mesh method [123] is applied with a cutoff distance of 9 Å for long-range electrostatic interactions. MD details for all systems are summarized in Table 3.3.

3.5. Ensemble Analysis

For ClustENM Procedure I, ProDy [125] is used to superimpose experimental structures which have 90% sequence similarity to starting structure (Table A.1), prior to principal component analysis (PCA). The alignment of the generated conformers onto these structures is also performed using this software. This procedure is carried on for CAM, p38 kinase, HIV1-RT and TIM.

PCA is performed on the generated conformers by using Cα coordinates for proteins and Cα and P coordinates of ribosome. Aligned experimental and MD data are projected onto the subspace generated by conformers PCs, in order to observe whether the subspace spanned by the PCs includes these data.

Table 3.3. MD simulations system details for TIM, HIV1-RT and p38 kinase.

MD runs	Total number of atoms	Number of water molecules	Box dimensions (Å)	Initial structure
Apo TIM MD1	54980	15753	89	1tcd
Apo TIM MD2	54830	15711	89	2 ns of TIM MD1
Apo TIM MD3	57554	16611	90	11 ns of TIM MD3
Complex TIM MD1	59274	17170	91	7 ns of Apo1, inhibitor docked [124]
Complex TIM MD2	61282	17842	92	13.5 ns of Apo1, inhibitor docked [124]
Complex TIM MD3	59848	17364	91	0.5 ns of Complex2
HIV1-RT MD1	134615	39541	128	1rtj
HIV1-RT MD2	152405	45471	128	5 ns of HIV1-RT MD1
P38 MD1	45462	13267	90	1p38
P38 MD2	44202	12847	90	50 ns of P38 MD1

3.6. Softwares and Parameters Used for Protein-Ligand Docking

In ClustENM Procedure II, docking of the ligands onto generated receptor conformers is performed using AutoDock [84, 85] and HADDOCK [126] softwares. For peptide dockings, we preferred HADDOCK since its scoring function and minimization protocol is suitable for amino acids. For the remaining ligands we selected to perform dockings using AutoDock.

In ligand dockings onto AK an BC, both the receptor and the ligands are prepared using AutoDockTools [85]. The ligands are kept almost rigid for AK (2 out of 22 bonds are active in AP5 ligand) and completely rigid for ATP in BC case. 100 runs are performed using the Lamarckian genetic algorithm of AutoDock v4.0 to explore the conformational space3. Each run consists of 25x106 energy evaluations. Grid box is located at the binding site of the ligand, covering receptor residues interacting with the ligand. The box dimensions are 64 x 72 x 80 with 0.375 Å spacing for AK and 80 x 80 x 80 with 0.375 Å spacing for BC. Results of all runs are clustered using an RMSD cut-off of 2 Å.

For peptide dockings onto LAO and DBP proteins, HADDOCK web server Easy Interface is used with default parameters. Residues interacting with the ligand in the complex crystal structure are selected as active residues.

3.7. Mean Square Fluctuations

Mean square fluctuations for MD runs and generated conformers are calculated based on the following formula:

$$MSF = \frac{1}{T}\sum_{t=1}^{T}(R_i(t) - \bar{R})^2 \tag{3.9}$$

where T is number of snapshots, $R_i(t)$ is the position of the Cα atom of i^{th} residue

at t^{th} snapshot and $\bar{R}$ is the average (mean) position of Cα atom of i^{th} residue. For generated conformers, weigth of the representative conformers are taken into account.

4. CONFORMATIONAL SAMPLING USING CLUSTENM-I FOR STRUCTURES UNDERGOING LARGE CONFORMATIONAL CHANGES

Proteins and nucleic acid complexes undergo local or domain-wise transitions between stable states in order to perform their function such as binding, catalysis, switching and serving as structural elements of living organisms [127]. It is essential to understand such transitions in order to elucidate protein function. We applied ClustENM Procedure I described in Methods, on the dataset of structures given in Table 3.1 to reveal the accessible states for a given initial state (e.g. determined from x-ray crystallography). This chapter will summarize results on the first four proteins (AK, CAM, HIV1-RT, and P38), whereas conformational sampling of TIM and ribosome will be discussed in the following chapters.

4.1. Adenylate Kinase

AK is a protein that is widely studied by many experimental and computational studies concerning large scale conformational transitions (a recent review about AK is in [6]). It is a phosphotransferase enzyme that plays an important role in cellular energy homeostasis; catalyzing the interconversion of ATP and AMP into two ADP molecules within the cells. AK consists of three domains: the LID domain where ATP molecule binds, the NMP domain also called the AMP-binding domain and the CORE domain. The LID and NMP domains undergo large conformational changes (about 7 Å from fully open to fully closed state) around the CORE domain, to provide solvent free environment for the substrates. In fact, the open and closed states of AK are accessible even in the absence of ligand, as suggested by experimental single-molecule Förster resonance energy transfer (FRET) studies [128, 129].

ClustENM-I is applied on AK to explore its conformational space, starting from four different states of the enzyme. Two of these are crystal structures that represent

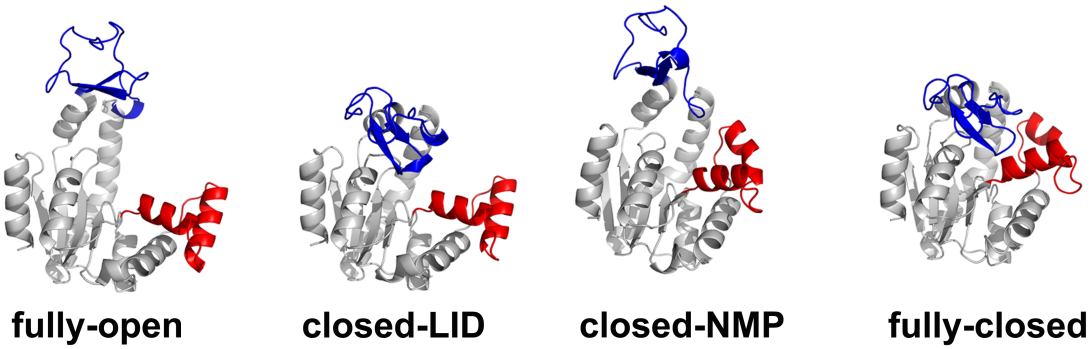

Figure 4.1. Starting structures for AK.

the fully-open state with open LID and NMP domains (pdb id: 4ake) and the fully-closed state with both domains closed (pdb id: 1ake). The other two structures, which correspond to closed-LID (open NMP), and closed-NMP (open LID) are selected among the generated structures in order to have same sequence as 4ake and 1ake. Although there are crystal structures that represent closed-LID and closed-NMP states such as 1dvr and 2ak3, respectively, we could not use these since their sequence identity to 1ake and 4ake are below 50%. The starting structures are presented in Figure 4.1.

We used LID-CORE and NMP-CORE angles defined in Beckstein *et al.*'s work [130] to describe the inter-domain conformational degrees of freedom of AK. The angles are calculated based on the definition given in the same study (shown in Figure 4.2a), where angle LID-CORE is formed from the centers of geometry (blue dots) of the backbone and Cβ atoms in residues 179-185 (CORE), 115-125 (CORE-LID) and 125-153 (LID). NMP-CORE is calculated similarly but for residues 115-125 (CORE-LID), 90-100 (CORE) and 35-55 (NMP).

In Figure 4.2b, the clusters from the fifth generation are shown for each distinct starting structure on the angle space. Representative conformer of each cluster corresponds to a dot with its diameter proportional to the cluster size (weight). Even though each independent simulation preferentially samples the quadrant closer to its starting structure, there is a large overlapping region sampled by multiple runs. These independent samplings are merged into a single graph (Figure 4.2c), which is color-coded

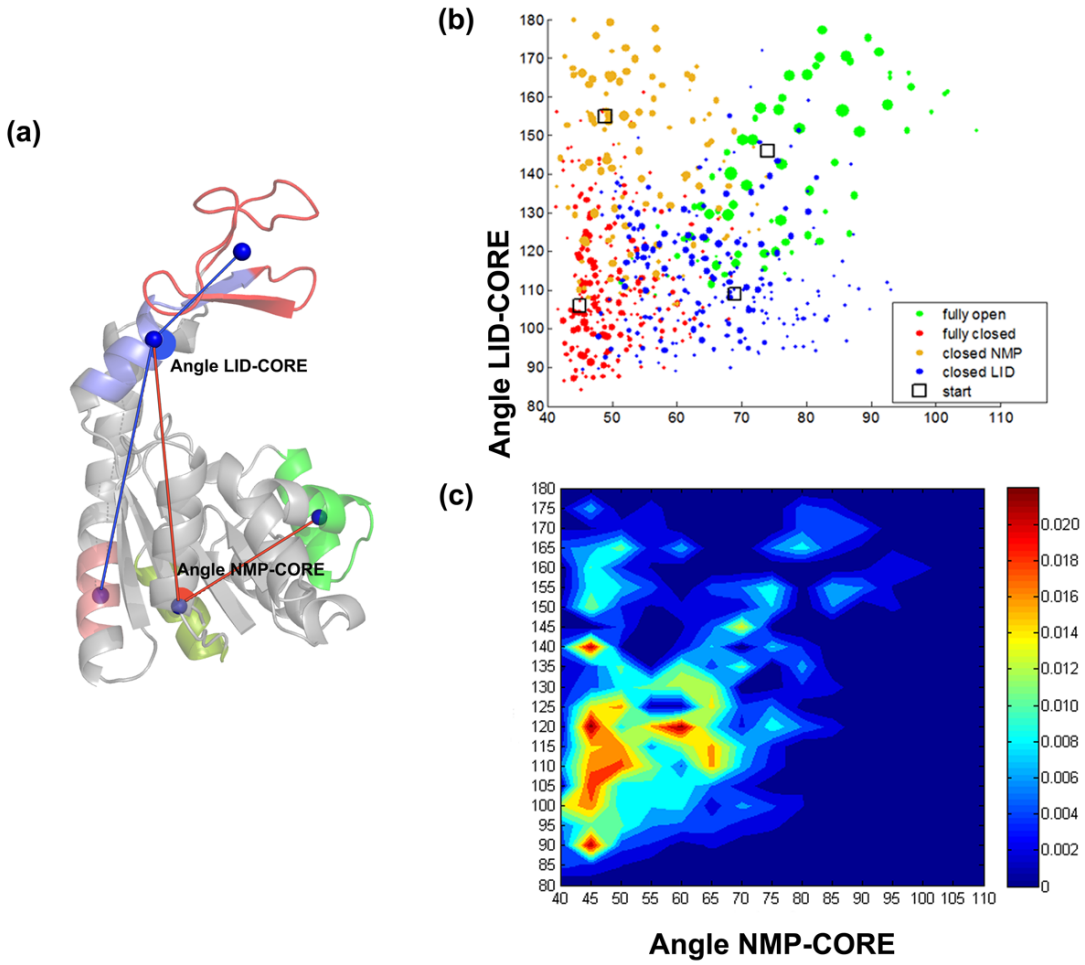

Figure 4.2. Conformational sampling of AK. (a) LID-CORE and NMP-CORE angles. (b) Independent samplings based on four distinct starting structures (squares). (c) Representative surface that merges four distinct simulations (in panel b) into a single one.

according to overall weights obtained for the four runs. The generated conformers cover a broader range of NMP-CORE and LID-CORE angles than the free energy surface reported for AK by Beckstein *et al.* [130], where the respective ranges were 40-75° and 100-155°. More interestingly, the highly populated regions in our surface correspond to the lowest free energy regions in the mentioned study.

Figure 4.2c also reveals that the region around the fully-open state (NMP-CORE: 74°, LID-CORE: 146°) is not highly populated. This is in agreement with studies stating that fully-open state of the apo enzyme is not energetically favorable [69, 74, 130–134]. Beside the closed-LID, closed–NMP intermediate states, there is an additional half open/half-closed state, which is highly populated (NMP-CORE: 55-60°, LID-CORE: 115-120°) in Figure 4.2c. This state is observed in both experimental and computational studies [6], which claim it as an intermediate state in the transition pathway between open and closed conformations of the enzyme.

The sampled regions in Figure 4.2c are also in agreement with the transition pathways studies. For example, ANMPathway [79] and coMD [78] report the closing of LID, followed by NMP closure in the transition from fully-open to fully-closed state. Same behavior is reported in the studies of Daily *et al.* [132], Matsunaga *et al.* [135], Beckstein *et al.* [130], Uyar *et al.* [75]. In the reverse direction, again the dominance of LID opening followed by NMP opening is observed in coMD [78], Beckstein *et al.*'s work [130]. In numerous computational studies [74, 130, 133, 135–137], it is reported that LID domain movement of apo AK occurs on a relatively flat energy surface. LID motion is also dominantly observed in small-angle X-ray scattering [138]. The sampled regions in our results shows that LID domain angle can have values between 105-135° when the NMP is open (65°), and between 90-125° when the NMP domain is closed (45°). However, we also observe NMP movement in our generated conformers, which is reflected from the highly populated region on Figure 4.2c where NMP angle changes between 45-65° when LID angle is around 120°.

While analyzing the different conformers on the surface, we observed that commonly used LID-CORE and NMP-CORE angles do not uniquely describe the inter-

domain flexibility of AK. Specifically two conformers falling on same spot of the conformational space may represent quite distinct structures, such as those shown in the Figure 4.3. In the figure, the fully-open crystal (4ake) structure and one generated conformer are shown, which are close to each other on the angle space but have an overall RMSD of 8 Å. The difference stems from the LID positioning, inducing a chirality (both have LID-CORE angle of 147°). Thus, angle metric is inadequate to distinguish such conformational differences. Still we employed this representation in order to compare our results with the available free energy surface [130] and angle-based results presented in other computational studies [75, 78, 79]. For the other proteins analyzed in this thesis, however, we will base our results on PCA.

Besides the angle analysis, we calculated the RMSD of the cluster representatives with fully-open, fully-closed, closed-LID and closed-NMP states for each simulation, in order to investigate the closeness of the representative structures of the final generation to these states. The lowest RMSDs to the specific states are reported in Table 4.1. In parenthesis, the number of conformers that have RMSD less than 3 Å in parenthesis are given (see Figure 4.4 for detailed RMSD analysis as histograms).

Table 4.1. RMSD table of representative conformers to the four states.

Starting structure	Sampled states			
	Fully-open	Fully-closed	Closed-LID	Closed-NMP
Fully-open	-	2.2 (2)	1.4 (15)	4.6
Fully-closed	4.6	-	2.7 (6)	1.2 (21)
Closed-LID	1.4 (8)	2.0 (17)	-	3.9
Closed-NMP	4.3	1.0 (14)	3.1	-

According to our results, starting from the fully-open state, it is possible to reach the closed-LID and fully-closed states, but not the closed-NMP state at the end of fifth generation. From the fully-closed state in reaching closed-LID and closed-NMP, but not fully-open. From closed-LID, it is possible to obtain both fully-closed and open structures, and structures which are closed-NMP-like (RMSD 3.9 Å). Interestingly,

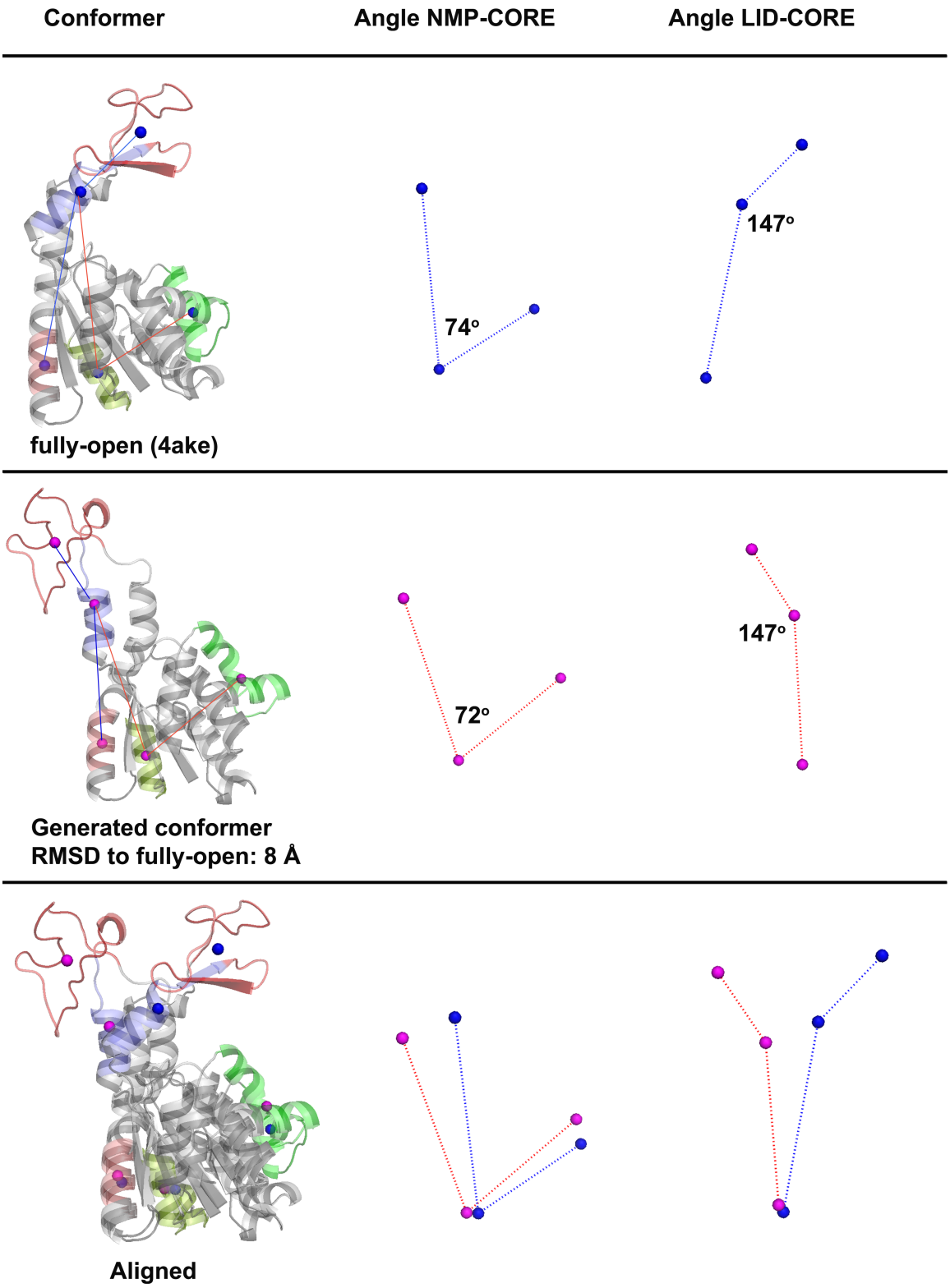

Figure 4.3. NMP-CORE and LID-CORE angles cannot distinguish fully-open crystal structure and a generated conformer that have an RMSD of 8 Å.

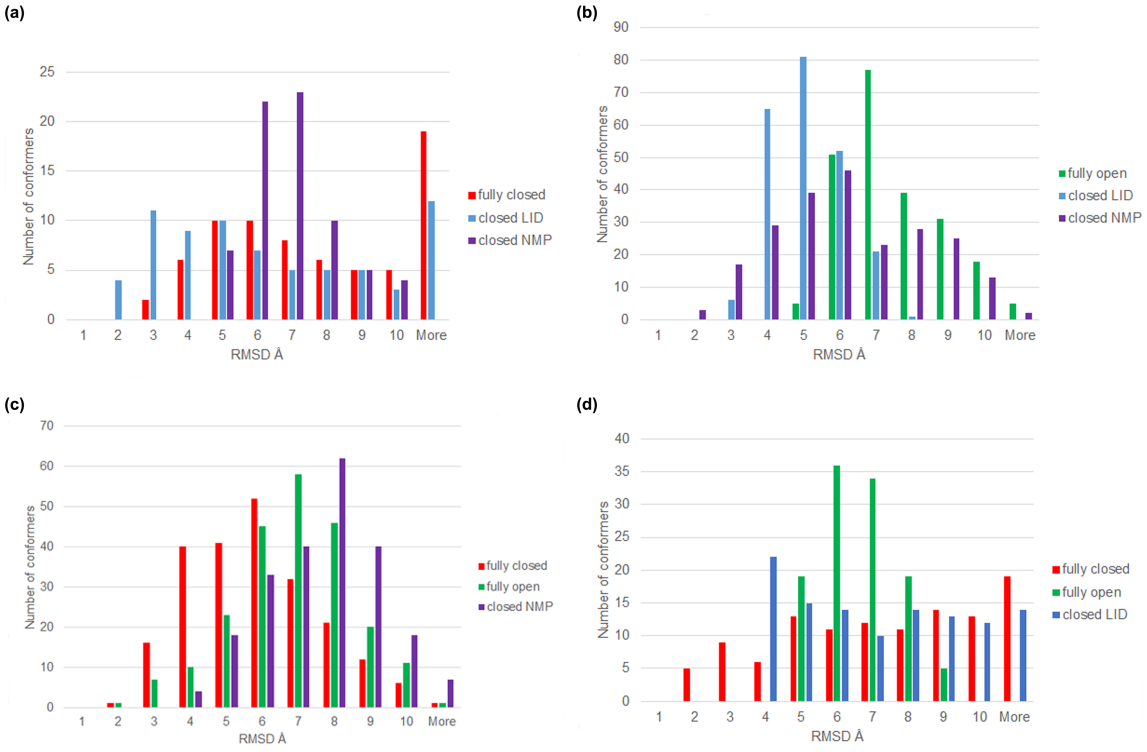

Figure 4.4. RMSD histograms for AK conformers to different states, using starting conformer as (a) fully-open, (b) fully-closed, (c) closed-LID, (d) closed-NMP.

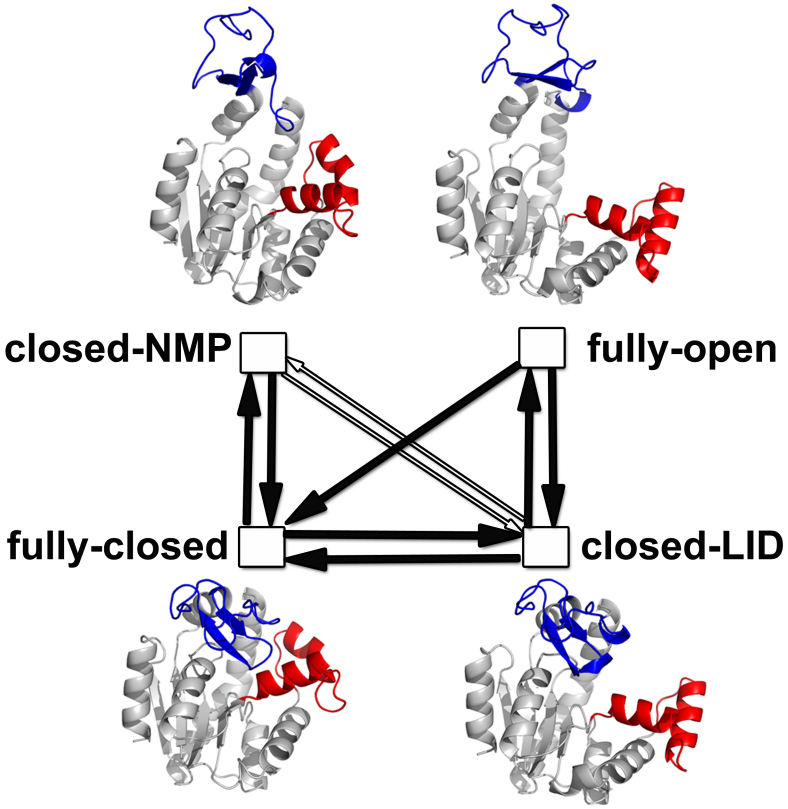

Figure 4.5. Accessible states when the starting structure is selected as fully-open, fully-closed, closed-LID or closed-NMP (LID domain is shown as blue, NMP as red).

starting from closed-NMP results in fully-closed and closed-LID-like (RMSD 3.1 Å), but not fully-open. Figure 4.5 summarizes these findings based on accessible states from each starting structure. Filled arrows are drawn between the starting states if any of the generated conformers have an RMSD less than 3 Å to a specific state (tip of the arrow). Empty arrows imply a similar situation, but with an RMSD between 3-4 Å. It should be stated that each arrow implies that such a conformational change is realizable within five generations, but does not describe the transition pathway. For example, our analysis cannot answer the following question: Does the transition from fully-open to fully-closed state pass through half closed states or the closed-LID state or both?

4.2. Calmodulin

Calcium binding protein CAM functions as a multipurpose intracellular Ca2+ receptor that is expressed in all eukaryotic cells. CAM consists of a single polypeptide chain with four Ca2+ binding sites (two on each lobe). It has 148 residues which constitute three distinct regions in the structure: N-lobe consists of residues 1-68, C-lobe residues are 92-148 and the remaining residues form the linker which is helical as seen in Figure 4.6. CAM may undergo large conformational transition, such as the change from the extended state to the fully collapsed state is about 15 Å (shown in Figure 4.6). Having an important role in calcium signaling pathways, it can bind to various target proteins in the cell to alter their activity.

We used the extended conformer of CAM (pdb id: 1cll) as the starting structure for exploration of its conformational space. At the end of fifth generation, 195 clusters (i.e. representative conformers) were generated and later used for PCA. Subsequently, 550 experimental structures, including NMR models and crystal structures, were projected onto the subspace spanned by the first three PCA modes in Figure 4.7a. Interestingly, the horse-shoe shaped distribution of the generated structures closely overlaps with that spanned by experimental data. Thus, the topology-driven modes have effectively guided the conformational space search for CAM. The highly populated region that is not sampled by generated conformers corresponds to the ligand-bound fully-collapsed state of CAM (closest generated structure has an RMSD of 9.1 Å, starting with initial RMSD of 15 Å). We should note that the helical structure of the linker is generally preserved in the generated conformers although a slight bending is also observed. The end-to-end linker distance defined in Aykut *et al.*'s work [139], (which corresponds to the distance between carbon alpha atoms of residues 69 and 91) changes between 30-34 Å range. This range is far from the bent linker end-to-end distance which is equal to 22 Å in fully closed state. We suppose that more generations are needed to be performed in order to observe pronounced bending of the linker.

Due to the availability of experimental data describing the distinct states of CAM (including intermediates), we also calculated and compared mean square fluctuations

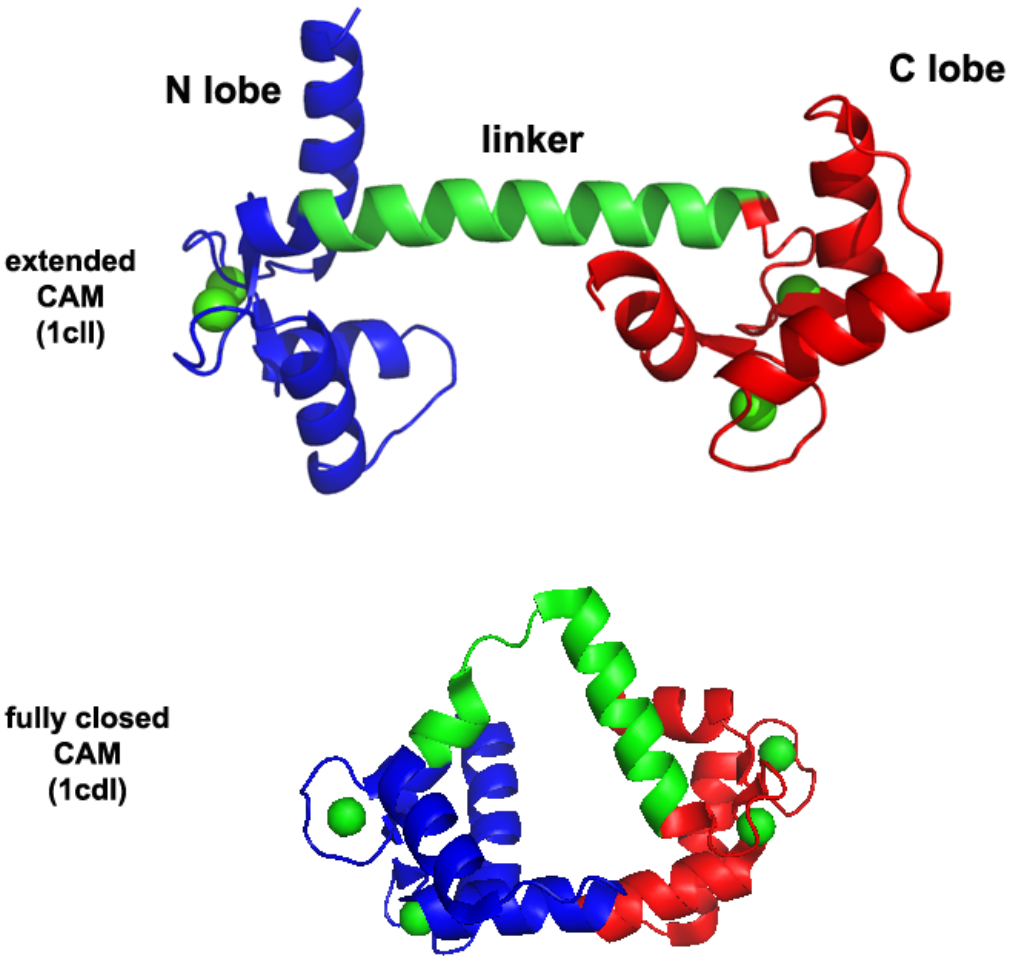

Figure 4.6. Extended and closed states of CAM, wheren N-lobe (blue), linker (green) and C-lobe (red) are shown.

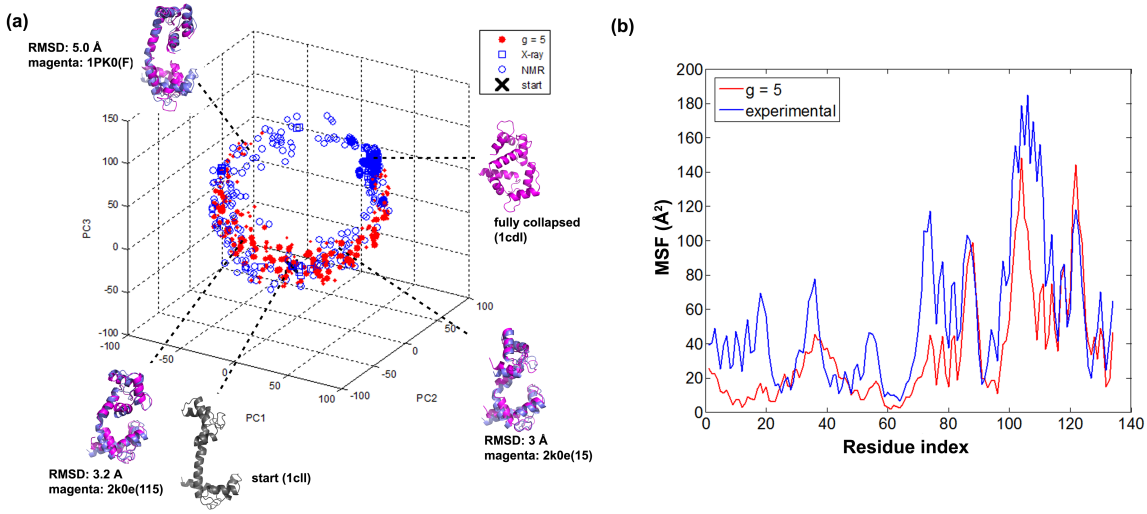

Figure 4.7. Conformational sampling for CAM. (a) Projection of experimental data (blue) onto the subspace spanned by the first three PCs of generated conformers (red) of CAM. (b) MSF for generated conformers and experimental data.

based on the ensembles of generated and experimental conformers as seen in Figure 4.7b. The trend for both data is similar, except residues 1-20 which seem more mobile in experimental data. As seen from the figure, N-lobe and C-lobe fluctuation scale is different, although CAM is a symmetric structure. This may stem from the overall alignment of the structures.

We wanted to observe how close our conformers get to the NMR ensemble (pdb id: 2k0e) reported for apo CAM [140], which contain 160 models. In Figure 4.8, two distributions are compared: blue histogram plots each NMR model's RMSD from the initial extended structure, whereas the red one reports the lowest RMSD value between each NMR model and the generated conformers. The initial RMSD range of 7-15 Å decreases to 3-10 Å, as the generated conformers that lie closest to NMR models are considered. Three of these conformers are shown in Figure 4.7a.

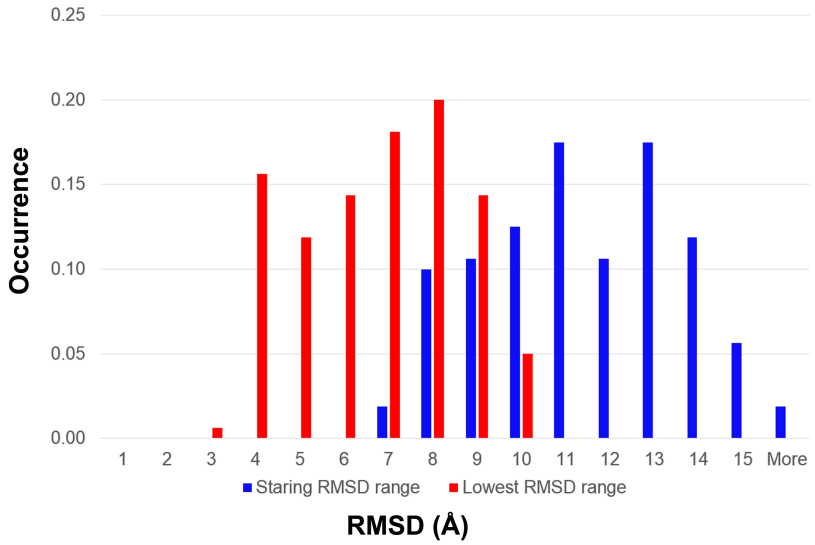

Figure 4.8. Initial/final RMSD range of the starting structure/generated conformers to the NMR models of 2k0e (blue).

4.3. Mitogen-activated protein kinase p38

Mitogen-activated protein kinase p38 has an important role as a signal transduction mediator and it is activated in response to extracellular stimuli such as osmotic shock, ultraviolet light, growth factors. Its link to inflammation, cell cycle, cell death, tumorigenesis in specific cell types has been reviewed elsewhere (Zarubin and Han, Cell Research, 2005).

Like in the case of CAM, PCA is applied on the generated representative conformers (41 structures) of p38 kinase. Before projecting experimental data onto the subspace spanned by PCs of generated conformers, experimental structures are iteratively aligned, based on the alpha carbons of residues 5 to 31, 36 to 114, 122 to 169, 185 to 351, which are resolved in at least 90% of the dataset [141], using ProDy software. Although 155 experimental structures are projected onto the subspace spanned by the first two PCs of generated conformers, the data still tend to accumulate within a narrower region compared to generated conformers and two independent 50-100 ns MD runs (Figure 4.9a, MD1 is 100 ns and MD2 is 50 ns long). The “grid-like” trend of the generated conformers stems from the persistent selection of a specific direction,

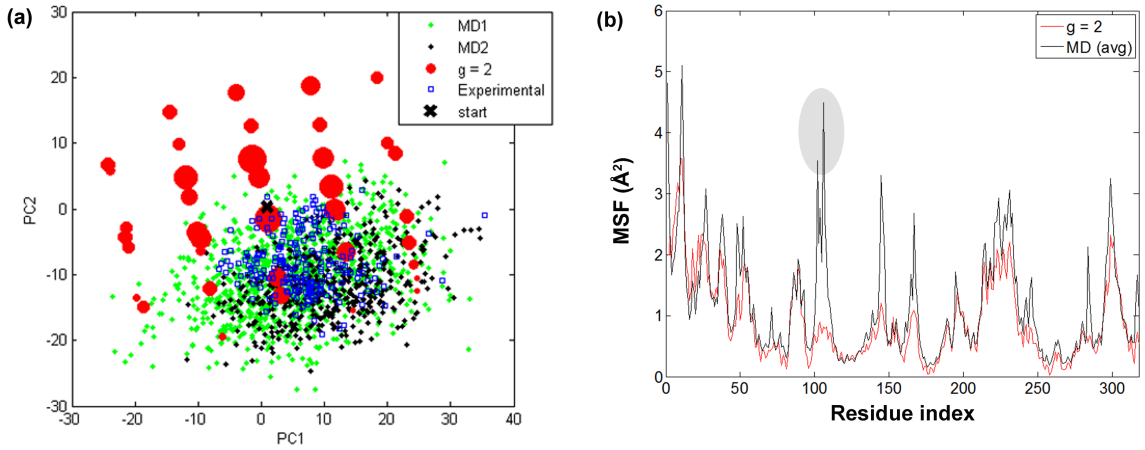

Figure 4.9. p38 MAP kinase. (a) Projection of generated conformers (red), experimental (blue) and MD (black and green) data onto subspace spanned by the first two PCs of generated conformers (red). (b) MSF for generated (red) and MD (black) data.

which can be also observed from Figure 4.10c (red conformers). Nevertheless, Figure 4.9a shows the drift of MD data to the regions occupied by the generated conformers.

Moreover, the trends in MSF (Figure 4.9b) for generated conformers and MD runs show consistency in overall. We observe higher fluctuations in MD compared to generated conformers only for residue index 100-115 region which corresponds to the loop indicated by an arrow on Figure 4.10.

Figure 4.10 shows the dynamics from experimental data with mostly localized motions (blue), generated conformers (red) with global motions and MD (green), which is a combination of both. In this figure, the loop which do not reflect fluctuation in experimental data belong to a region left out from the alignment (residues 170-184), but it is present in both MD and generated conformers.

4.4. HIV1-Reverse Transcriptase

Reverse transcriptase enzyme is essential for retrovirus HIV to convert its viral RNA to a double stranded DNA during the infection process of the host cell. Being

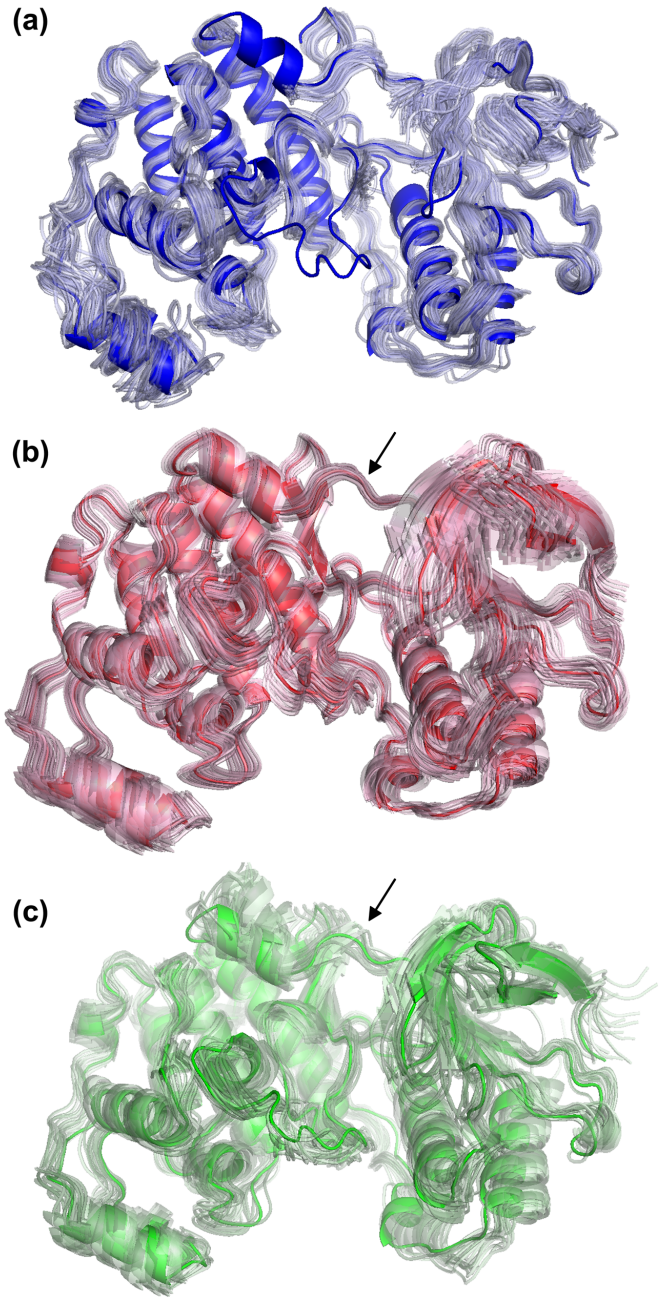

Figure 4.10. p38 MAP kinase conformers from x-ray (blue), generated clusters (red) and MD (green) data. In all cases, the opaque structure corresponds to the starting structure.

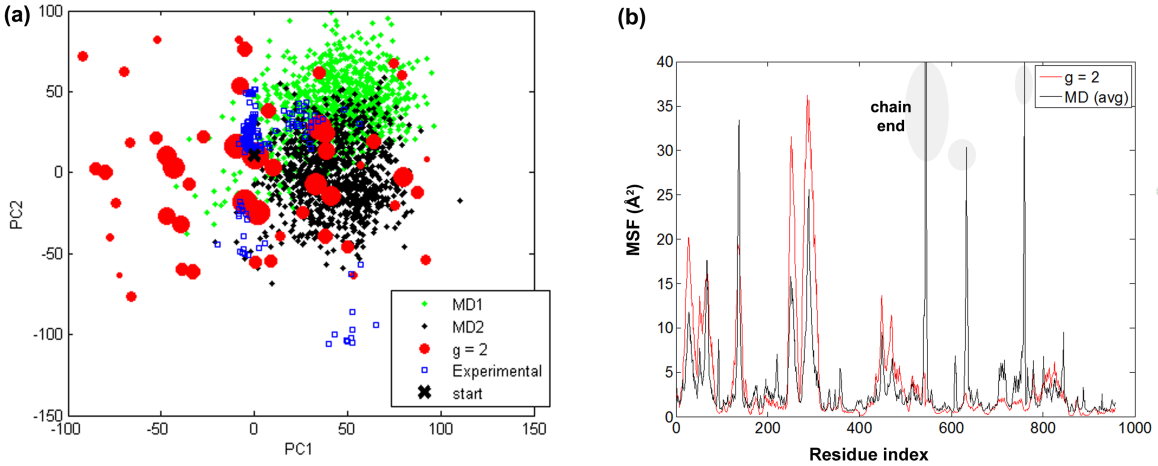

Figure 4.11. HIV1-RT. (a) Projection of generated conformers (red), experimental data (blue) and MD data (black and green) onto subspace spanned by the first two PCs of generated conformers (red). (b) MSF for generated (red) and MD (black) data.

a target in anti-HIV drug development, HIV1-RT is an asymmetric heterodimer composed of p66 and p51 subunits. The structure has four domains named as "fingers", "palm", "thumb" and "connection" [142] and p66 has a COOH-terminal RNase H. Both polymerase and RNase active sites are located on p66 subunit

Figure 4.11a shows the PC subspace spanned by generated conformers, which covers most of the experimental and MD data. Even though two independent MD runs (100 ns each) mostly sample the right-hand side of the conformational space (PC1 $>$ 0), some of the snapshots are observed on the left-hand side that is sampled by the generated conformers, similar to p38 case. One cluster of experimental data (located about PC1 = 50, PC2 = -100), which is not sampled by either the generated clusters or the MD snapshots, corresponds to the closed state of the clamp (observed in Figure 4.12a). MSF for generated conformers, MD and experimental data indicate a satisfactory level of correspondence, even though the experimental data cluster in a more confined region in the Figure 4.11b. High peaks on MSF figure corresponds either chain ends or the loop motions in subunit B, showed with arrows in Figure 4.12. Although the clamp motion is emphasized in the dynamics from experimental data in Figure 4.12 (blue); generated (red) and MD (green) structures express more global

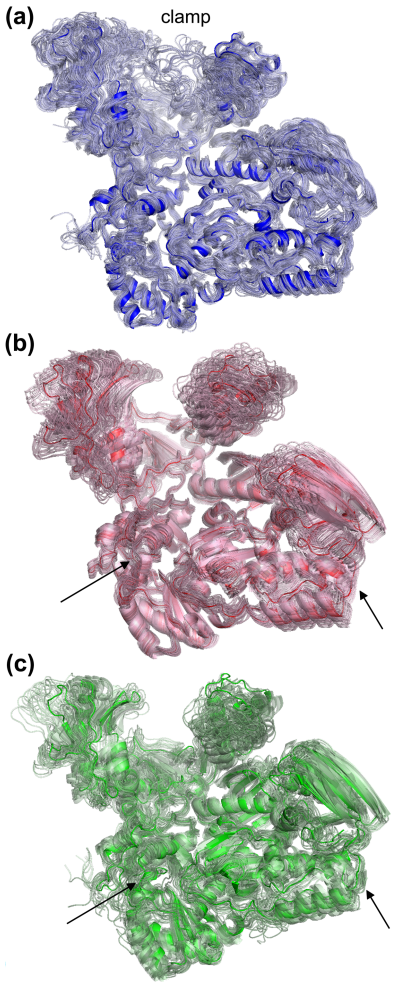

Figure 4.12. HIV1-RT conformers from x-ray (blue), generated clusters (red) and MD (green) data. In all cases, the opaque structure corresponds to the starting structure.

motions and behave similarly.

5. APPLICATION OF CLUSTENM TO BACTERIAL RIBOSOME

5.1. 70S Ribosome

Ribosome is a supramolecular machine that performs protein synthesis in all living cells. 70S bacterial ribosome is formed by the association of small (30S) and large (50S) subunits, which in total contain more than 50 different ribosomal proteins and several ribosomal RNAs. Protein synthesis is accompanied by several conformational changes of the ribosome, which are challenging to observe by classical MD. The translocation process still needs clarification in terms of molecular mechanism [107, 143]. Several methods have been used to investigate ribosome dynamics including ENM [102–105, 109], coarse-grained MD [34] and MD using a multi-basin structure-based model [107].

In order to demonstrate the applicability of our sampling method (ClustENM-I) to supramolecules, we used a recently reported crystal structure of 70S bacterial ribosome complexed with elongation factor G, which is trapped in an intermediate state of translocation [144], called "chimeric pe*/E" state (pdb id: 4KDK (50S)-4KDJ (30S)). The complex contains 50 ribosomal proteins, mRNA, tRNA and EF-G besides the 23S, 16S and 5S RNA. Using the minimized 70S structure, we generated 101 representative conformers in the second cycle.

Due to the missing chains and/or residues in experimental structures, we calculated mutual RMSDs (among experimental data or between experimental and generated structures) based on P atoms of the common chains 23S and 16S only, using PyMOL software [145]. Accordingly, available 27 experimental structures of 70S including the initial structure fall into five major clusters (Figure 5.1).

Comparison of generated conformers with experimental structures indicates that almost half of the conformers approach towards one of the clusters other than the initial

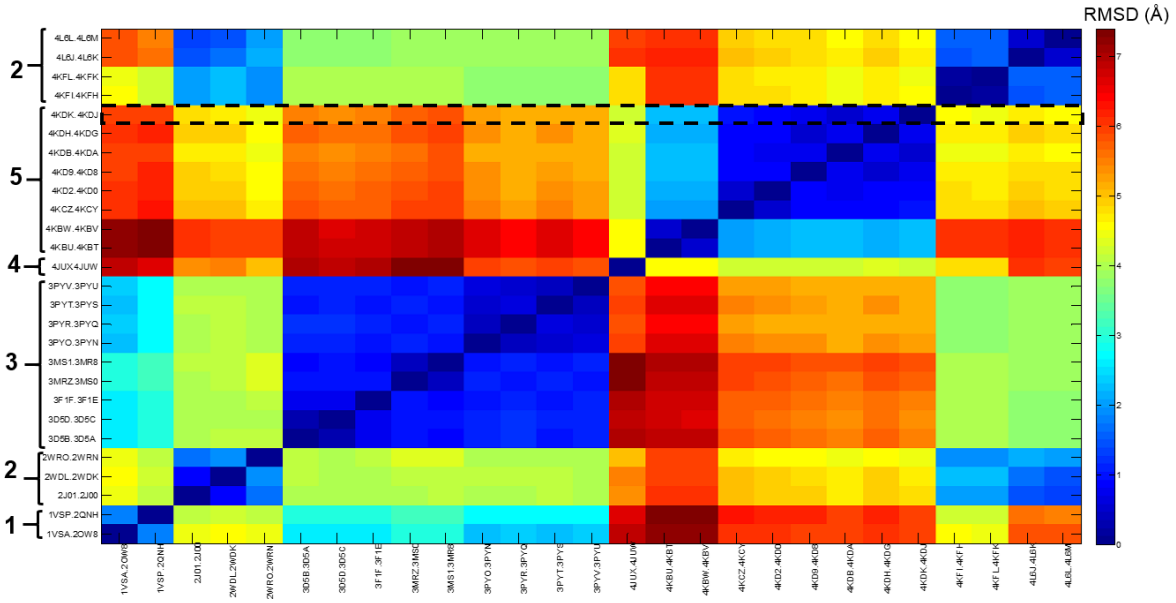

Figure 5.1. Mutual RMSD for 70S ribosome x-ray structures, based on P atoms of 16S and 23S only.

structure's cluster. In in Figure 5.2, the starting structure's RMSD to experimental structures is subtracted from the generated conformer's RMSD to the same structures, to observe the closeness/farness of the generated conformers. Red-orange colored regions correspond to structures that are closer to crystal structures than the starting structure (colorbar on the right shows RMSD). Clusters of experimental structures are separated by dashed lines. Cluster no.5 contains the initial structure 4kdj-4kdk.

Moreover, Figure 5.3a indicates that the starting RMSD range (of initial structure with experimental structures) shifts to lower values for the generated conformers (lowest RMSD for each generated structure is used in the histogram).

PCA is applied on generated 70S ribosome structures and the distribution of conformers on the subspace spanned by PC1 and PC2 can be observed in Figure 5.3b. PC1 and PC2 describe the ratchet motion of subunits and the 30S head motion, respectively, including high mobility of L1, consistent with previously reported results in numerous studies [34, 102–105, 107, 109, 143]. These motions and the mobility of the complex can be depicted in Figure 5.4.

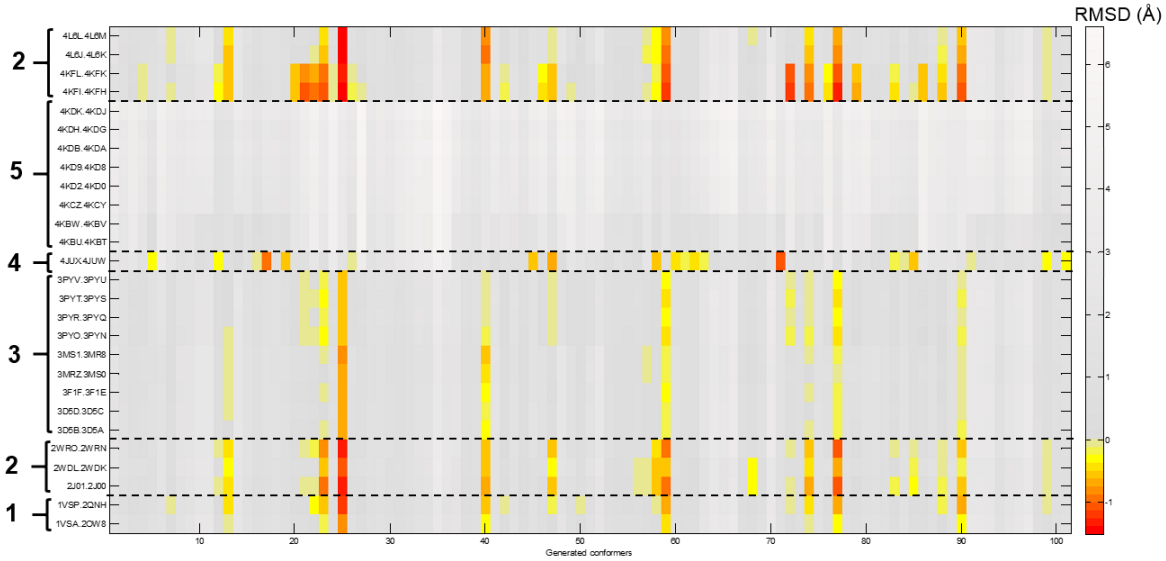

Figure 5.2. Closeness/farness of the generated conformers (columns) to crystal structures for 70S ribosome (rows).

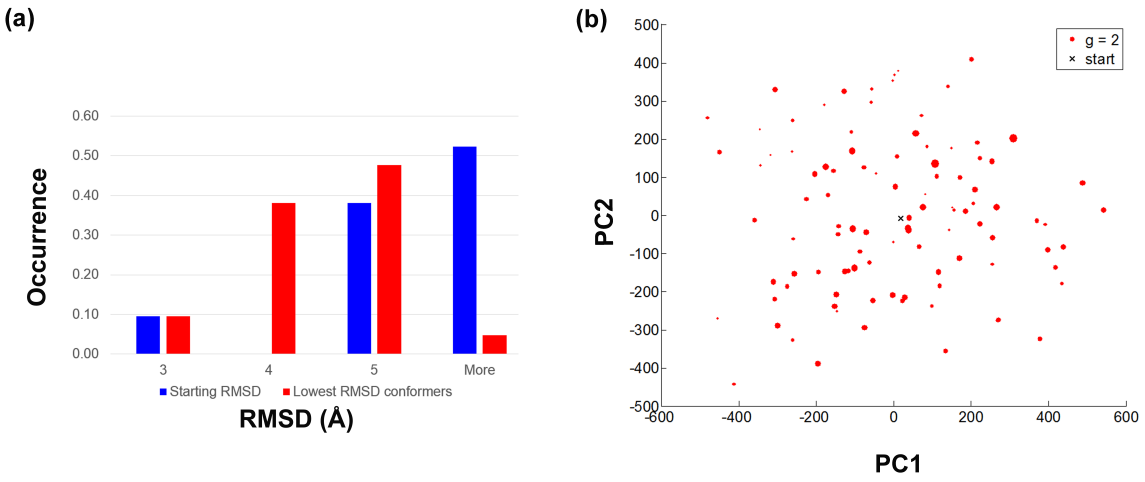

Figure 5.3. 70S ribosome (a) Initial/final RMSD histogram. (b) PCA of generated conformers.

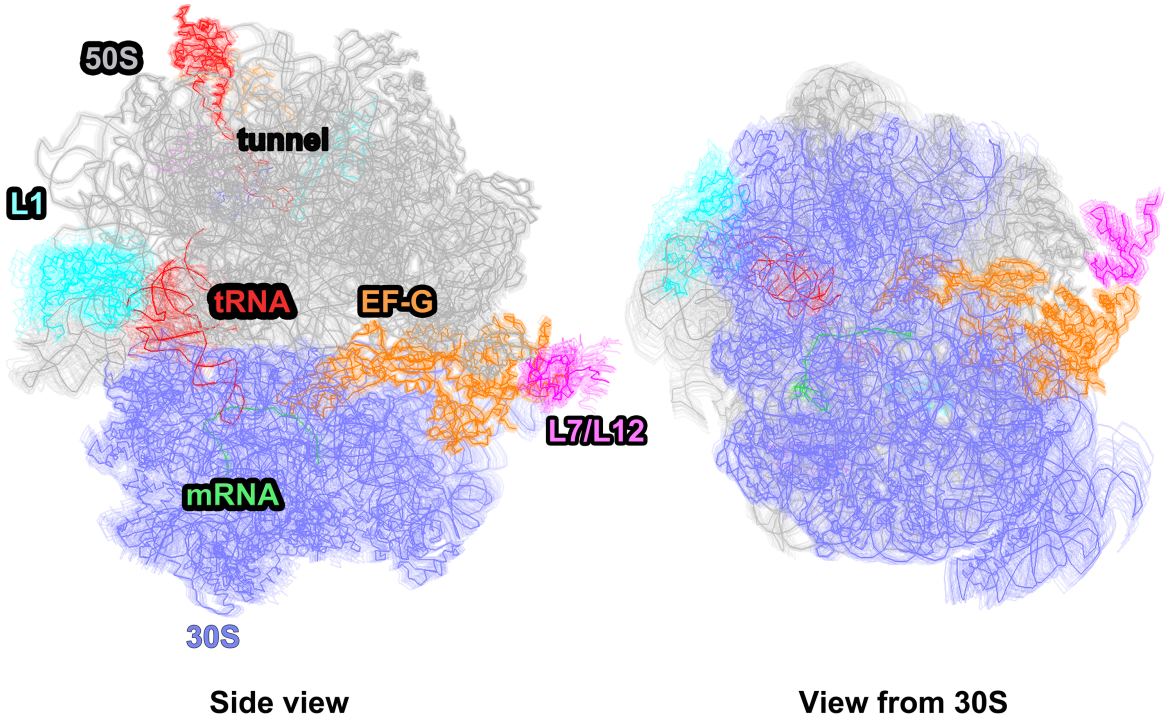

Figure 5.4. The dynamics of 70S from 32 generated structures with higher weights.

In generated conformers, high mobility of tRNA is apparent together with L1 stalk, as seen in Figure 5.5. tRNA also has the tendency to move towards to the exit site, i.e. away from the P site. We can speculate that this may due to the presence of EF-G, which accelerates the translocation in forward direction [108]. To further analyze this situation, new generations can be created without EF-G, and the differences in the dynamics of tRNA can be investigated in more detail.

Ribosomal tunnel on 50S subunit, through which the nascent polypeptide chain exits, is shown with nearby proteins L4, L22, L23, L24 and L29 in Figure 5.6. In our previous study regarding ribosomal nascent polypeptide exit tunnel, we observed L4 and L22 anti- or non-correlated motions at the narrowest and important region of the tunnel [105]. In this work we also observe the flexibility of L4 and L24 in generated conformers, however L22-23 and L29 exhibit mostly rigid-body type translations. Increasing DF in the generation of conformers for such a complex may lead to more detailed observation of the these proteins'dynamics, which is local.

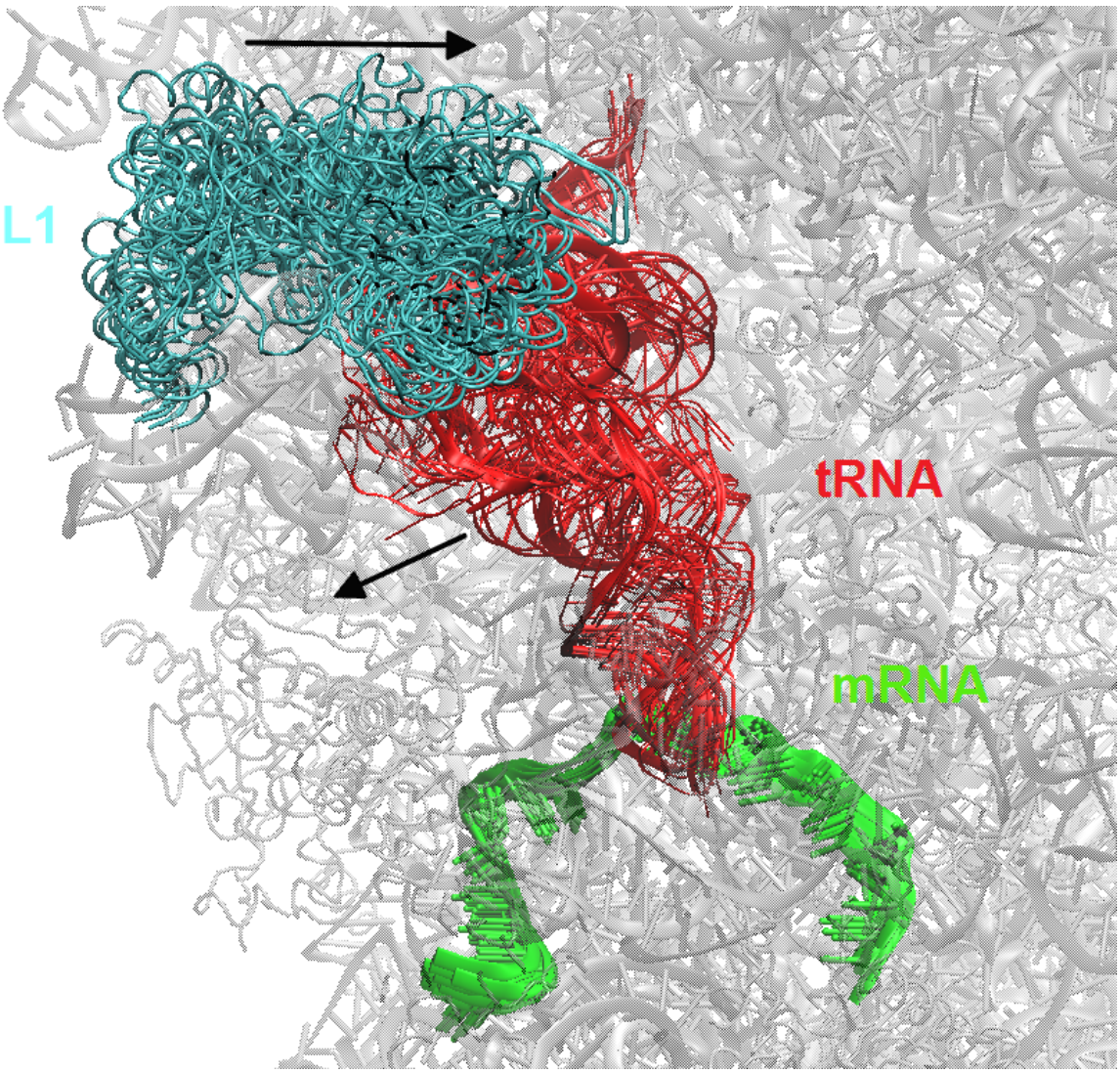

Figure 5.5. Mobility of tRNA together with L1 stalk.

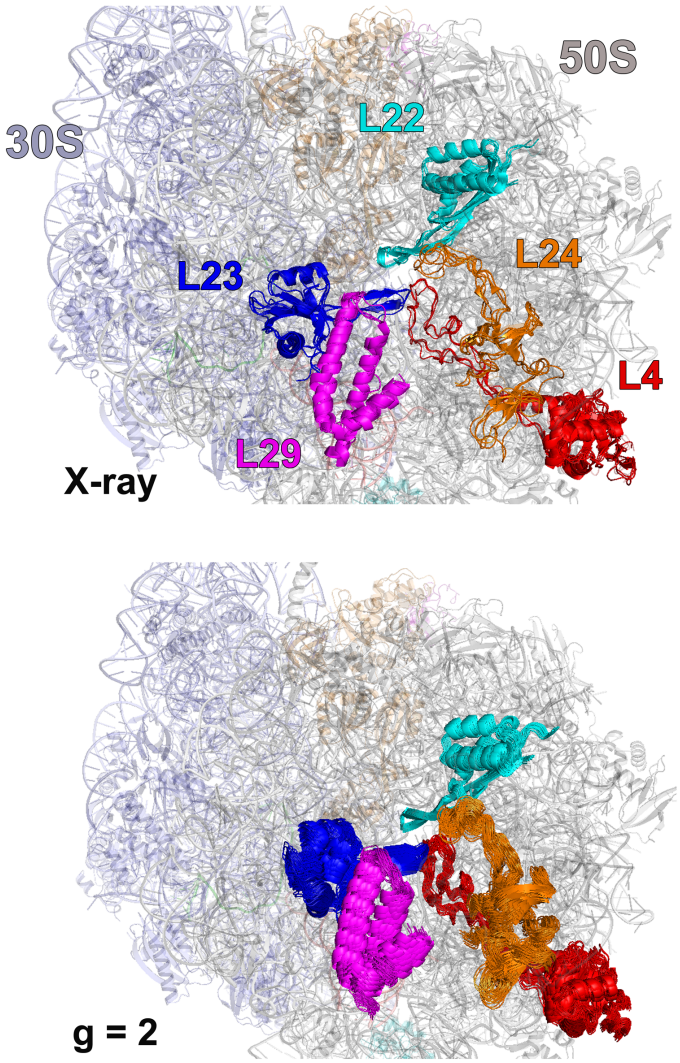

Figure 5.6. Close-up view of protein dynamics located at the exit tunnel of 70S ribosome.

5.2. 50S ribosome-Trigger Factor (TF) Complex

Bacterial TF binds to the exit tunnel of ribosome [146–150] and acts as a chaperone facilitating the folding of newly synthesized proteins, emerging from this tunnel [151, 152]. E. coli TF with 432 residues consists of three domains; namely the ribosome-binding domain (residues 1-112, "tail" domain), the peptidyl prolyl isomerase domain (residues 150-246, "head" domain) and the C-terminal domain, which is the core domain (residues 113-149 and 247-432 where "arms" are located). TF binds in monomeric form to the ribosome, whereas unliganded TF can also exits in monomeric and dimeric form, in the solution (100 kDa).

In this study, we applied ClustENM methodology (Procedure II) on 50S ribosome-TF complex, in order to shed light on the dynamics of such a complex. Currently, the intact structure of ribosome-TF complex is not available in Protein Data Bank. However, three crystal structures of 50S ribosome, containing a portion of TF binding domain are present. One of them is from *Haloarcula Marismortui* (pdb id: 1w2b), which has whole 50S unit together with ribosomal proteins; however this structure has only 35 residues of the TF binding domain. The other is from *Deinococcus Radiodurans* (pdb id: 2aar), which carries a coarse-grained but longer part of the binding domain (113 Cα atoms) on 50S ribosome, however this structure lacks most of the ribosomal proteins (only L23 and L29 are present). Another is again from *Deinococcus Radiodurans* (pdb id: 2d3o) at 3.35 Å resolution, which contains TF N-terminal binding domain (112 residues) in complex with 50S subunit, where L24, L29 proteins are also available. To obtain a 50S ribosome structure with all ribosomal proteins and maximum number of TF domain residues, we aligned 2aar and 1w2b structures on top of each other using PyMOL. Then, we extracted a TF conformer from MD simulations of a previous study [153], performed with ionic strength of 150 mM. We selected the binding domain of this conformer and aligned this region on top of 113 Cα atoms-binding domain of TF from 2aar. At the end, we obtained an atomistic 1w2b-TF complex. Resulting 50S ribosome-TF complex, which has 156,991 atoms in total (corresponding to 7023 residues), was subjected to energy minimization using implicit solvent with parameters described in Methods. The minimized complex shown in Figure 5.7, is

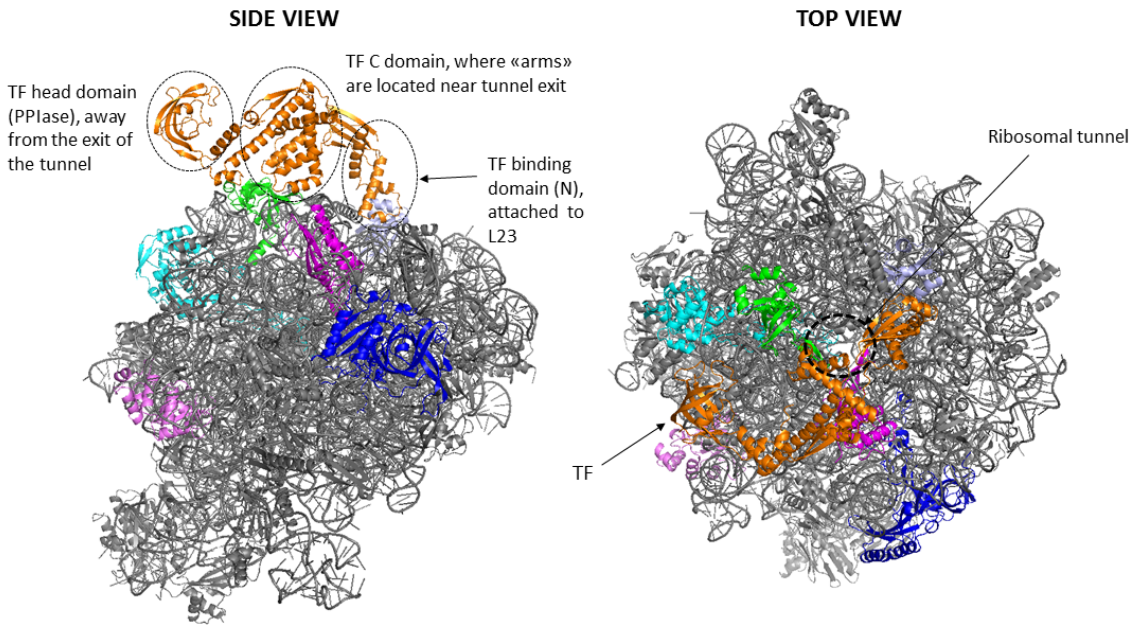

Figure 5.7. Energetically minimized 50S ribosome-TF complex structure (side and top views).

used as starting structure in ClustENM. As seen from the figure, TF (orange) binding domain is next to the L23 (light blue) ribosomal protein. Other proteins surrounding the ribosomal exit tunnel, namely L3 (blue), L4 (cyan), L22 (magenta), L24 (green) and L29 (violet) are also shown on 50S ribosome (gray) in the same figure.

ClustENM-II is applied on the energetically minimized complex with DF of 2 Å, using first five slowest modes in the blind conformational search. 17 distinct representative structures are obtained in total from two generations. The structures are superimposed in Figure 5.8, revealing high mobility of head domain. The binding domain and the arms located at C domain remain relatively stable on ribosomal exit tunnel, which is contrary to their significant mobility observed in previous MD simulations of monomeric apo TF in solution, at different ionic strengths [153–156]. This supports the view that N-terminal tail and C-terminal arms provide a protected folding space for the nascent polypeptide chain [147]. Moreover, high mobility of head domain is in conformity with crystal and cryo-EM data, which indicate the necessity of the rotation of head domain by 24 degrees towards the C-terminal arms in order to fit the

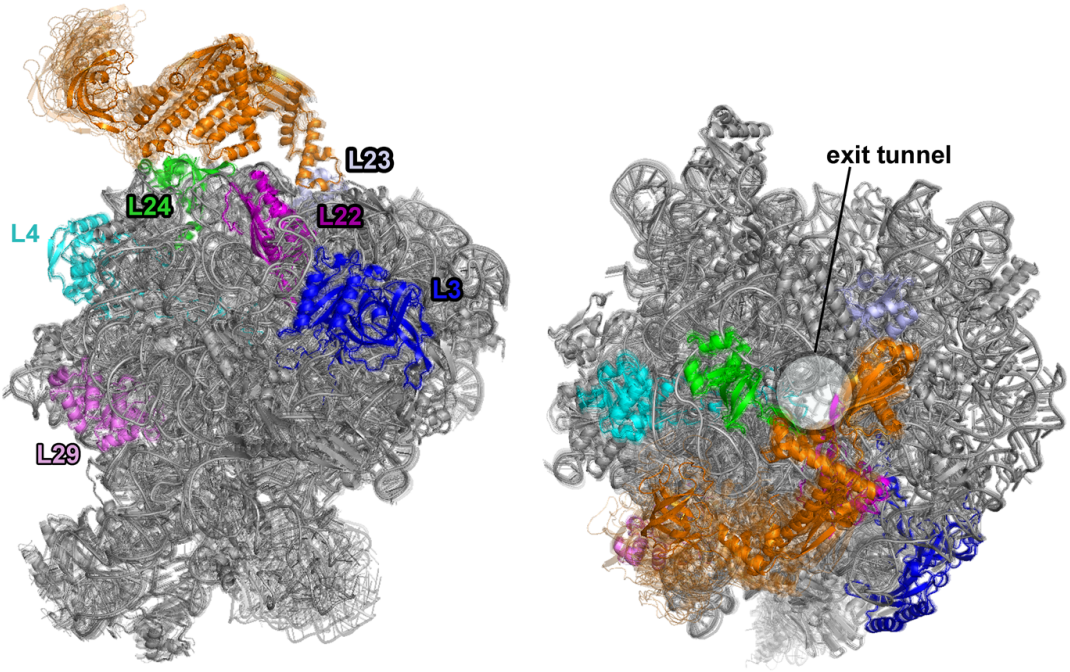

Figure 5.8. 50S ribosome-TF complex dynamics from generated conformers using ClustENM.

density [150].

We also aligned four of the generated structures on this cryo-EM map (EMDB map 1499) using Chimera, which has also one aligned experimental structure of TF (pdb: 2vrh), shown as red in Figure 5.9. The generated structures (shown in blue colors) are in conformity with the experimental structures; binding and core domains are relatively stable and head domain is flexible.

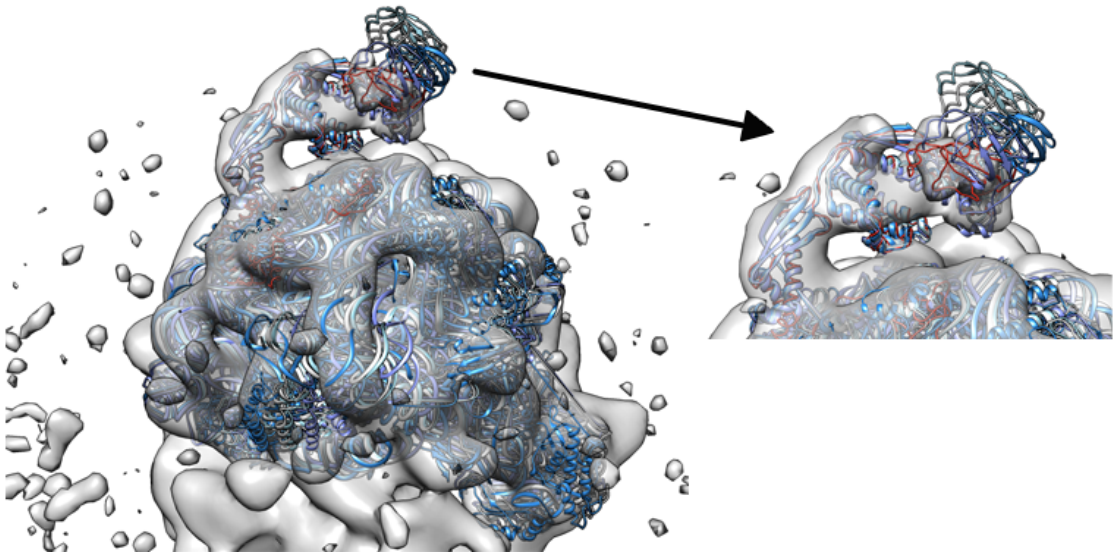

Figure 5.9. 50S ribosome-TF generated conformers (blue) aligned onto cryo-EM map (surface) and experimental TF structure (red).

6. TRIOSEPHOSPHATE ISOMERASE CONFORMATIONAL DYNAMICS

6.1. Application of ClustENM-I on TIM

TIM is a crucial enzyme in the glycolytic pathway, catalyzing the interconversion of dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate (GAP) by isomerization reaction. It is active in dimeric form, although each identical subunit possesses its own catalytic site. Each subunit is composed of 251 residues (in TIM from Trypanosoma cruzi, TcTIM) adopting the TIM-barrel topology. The active site is located at the C-terminal end of the β-barrel. There are four catalytic residues, namely Asn12 and Lys14 on loop 1, His96 on loop 4 and Glu168 on loop 6, which are shown on one subunit of the X-ray structure 1tcd [157] in Figure 6.1. Catalytic loop 6 (Glu168-P178), loop 7 (Gly212-Lys218) and loop 8 (Gly235-Lys240) also contribute to the active site geometry via H-bonding interactions with the substrate [158]. Interdigitating loop 3 (Gln66-Val79) plays a crucial role in the stability of the dimer by interacting with other subunit [159–161]. Allostery or cooperativity has not been observed between the identical active sites [162], located on both subunits.

The available experimental structures of TIM reveal the relatively localized opening/closure motion of loop 6. This functional loop 6 (residues E168-P181 in 1tcd) closes over the active site and protects the substrate from solvent during the reaction. However, it is also known that loop 6 opening/closure is not ligand-gated and takes place in the apo state, as well as when substrate is bound [163–165]. There are also crystal structures (e.g 1tsi, 1lzo) where the loop 6 is in open conformation despite the presence of the ligand on the catalytic site.

Several MD simulations and ENM studies have stressed the presence of collective motions in TIM, such as counter-rotation or bending of the subunits, which are coupled to loop 6 dynamics but not evident in experimental structures [164–169]. In

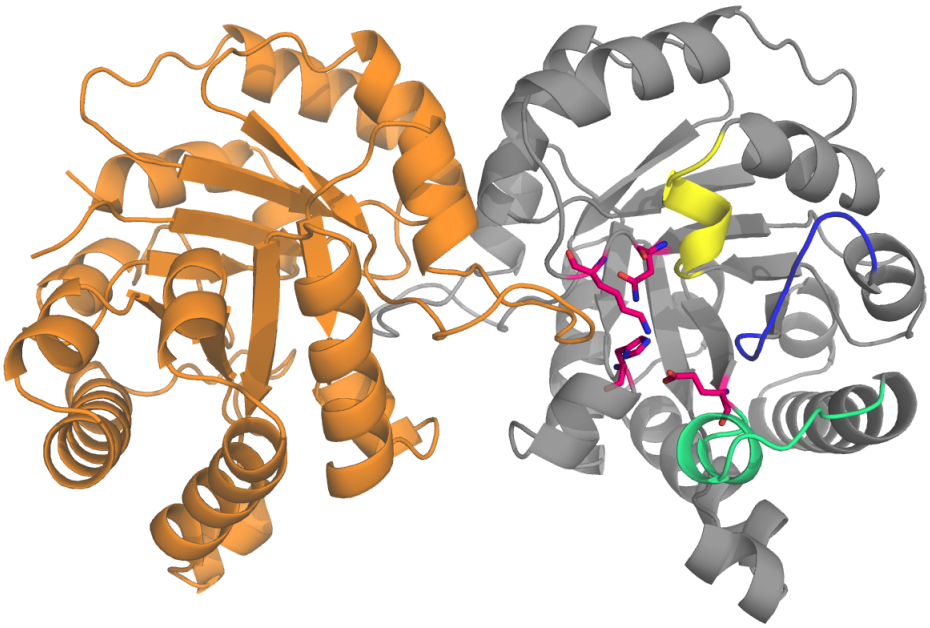

Figure 6.1. TIM structure from *Trypanosoma cruzi*.

conformity with these findings, the conformers generated here using first three collective modes cover a wider space in comparison to the projected experimental structures (blue squares), which tend to cluster in a confined region of the PC subspace, as seen in Figure 6.2a. We also projected the MD snapshots obtained from three independent 100 ns runs onto the same PC subspace, which indicate similar subspaces being sampled with the generated conformers.

High correlation is observed in residue MSF between generated conformers and MD simulations as seen in Figure 6.2-b. However the fluctuations of loop 6 indicated by the circles on the figure are more pronounced in MD runs. The dynamics from experimental, generated and MD data can be visualized in Figure 6.3, where only the catalytic loop motion (opening/closure) is observed in x-ray (blue), but global dynamics is present in both generated (red) and MD (green) structures.

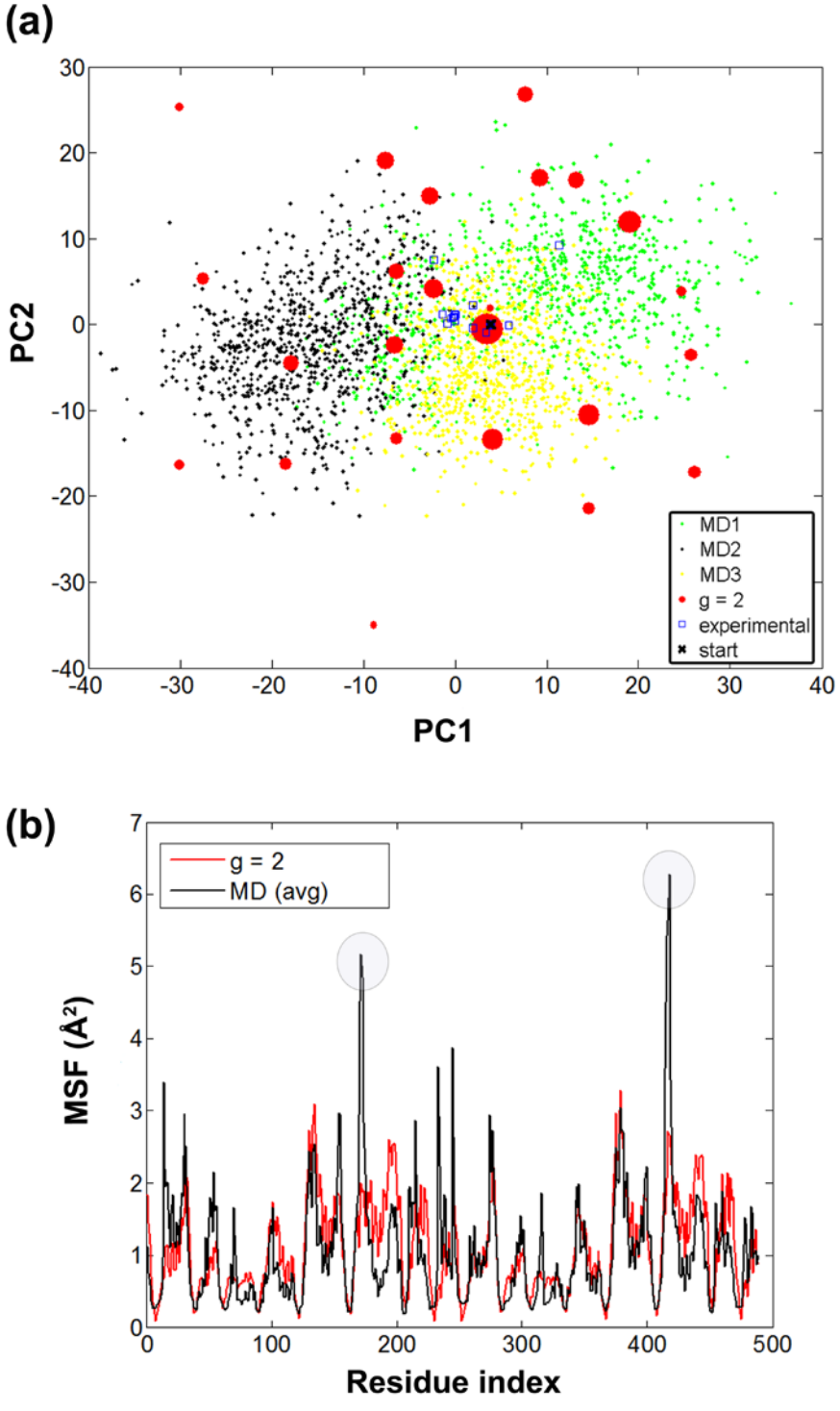

Figure 6.2. TIM PCA and MSF. (a) Subspace spanned by the first two PCs of generated conformers (red) with projected experimental data (blue) and MD data (black, green and yellow) (b) MSF for generated (red) and MD (black) conformers.

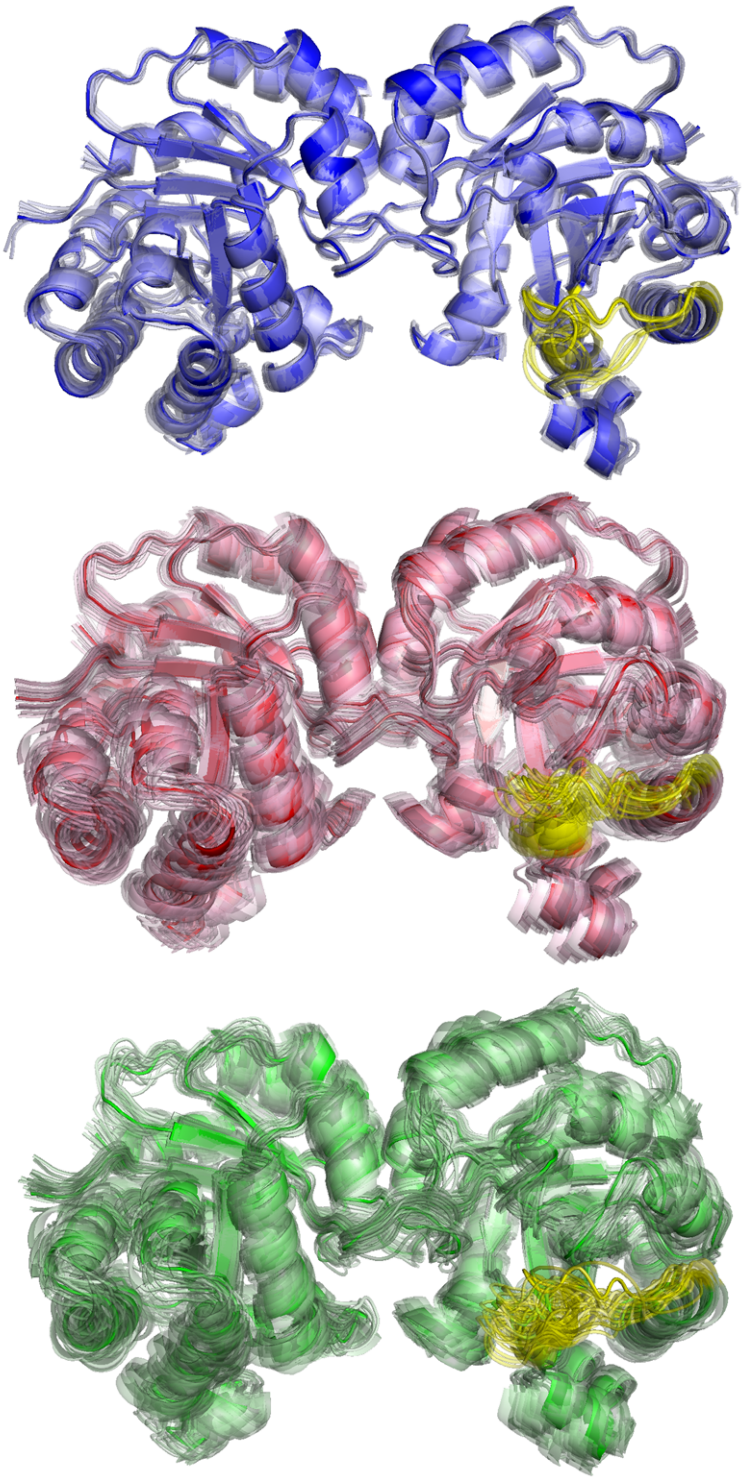

Figure 6.3. X-ray (blue), generated (red), MD (green) conformers of TIM.

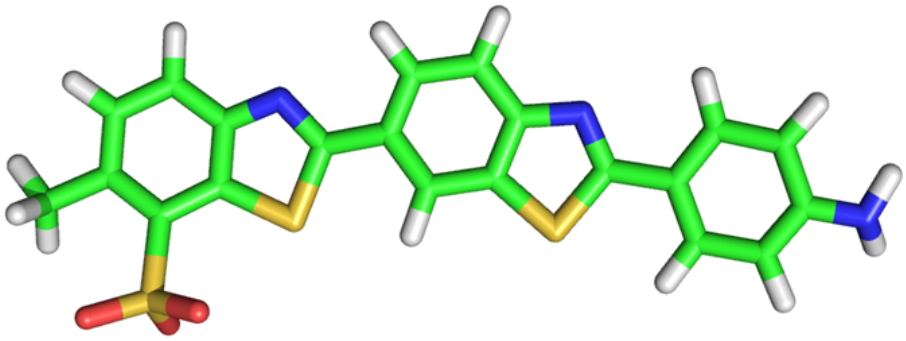

Figure 6.4. Inhibitor bt10 structure.

6.2. Inhibitor effect on TIM global dynamics and its specific interactions with the enzyme

In drug design studies against the parasitic diseases, subunit interfaces of oligomeric enzymes including TIM may serve as species-specific drug target sites, since they are generally less conserved than the active site [159,170,171]. Several benzothiazole derivatives have been reported as drug candidates against the Chagas disease by inhibiting TcTIM's activity [160, 171, 172]. One such example is an inhibitor named as bt10 (Figure 6.4) that binds to the specific tunnel region at TcTIM's interface [160, 171]. Previous works [124,173,174] have revealed that bt10 interacts with the aromatic clusters formed by Phe75 of one subunit and Tyr102-103 of the other subunit on the interface in the so-called tunnel region formed by dimerization. Our previous blind docking study [124], has indicated that bt10 binds selectively to the tunnel region in dimer, whereas its specificity for the interface decreases in the monomer. However, the specific inhibitory mechanism of this derivative is still not clear. One proposition on this issue has been the destabilization of the interface upon binding and subsequent dissociation into monomers [173, 174]. However, our MD simulations as will be presented here show that the interface region occupied by bt10 enhances the network of interactions and at the same time modifies the overall dynamics of the protein. As such, there seems to be an allosteric effect on the catalytic site and loop 6.

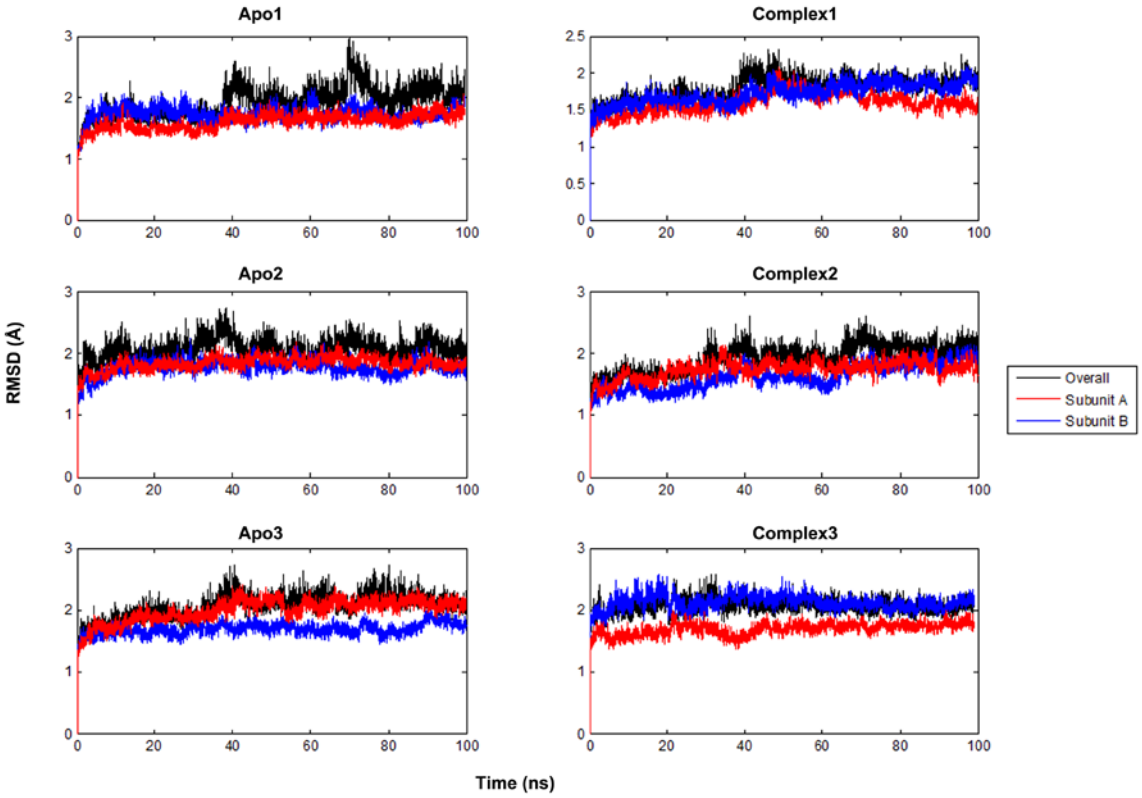

Figure 6.5. Dimeric and monomeric RMSD profiles for MD runs of TIM.

6.2.1. Root Mean Square Deviation and Mean Square Fluctuations for Apo and bt10-Bound MD Runs

To understand the inhibitor effect on the enzyme, we performed six independent 100 ns-long MD simulations. Three of these simulations involved the apo enzyme and the rest correspond to bt10-bound complex. The RMSDs of MD snapshots with respect to the minimized initial structure of each run (heavy-atom based) are confined within 2.5 Å, indicating stability over 100 ns (Figure 6.5). Based on RMSD profiles of dimer and monomers, first 10 ns of each run were discarded as equilibration period from further analysis.

To analyze the inhibitor effect on TIM dynamics, first we calculated residue MSF for apo and bt10-bound TIM, and shown in Figure 6.6 for each run. Two-way ANOVA is applied on residue-based MSFs, with null hypothesis that ligand has no effect on residue MSF. The null hypothesis is rejected with p-value of 0.0185 (less than 0.05).

We show the change in MSF values upon ligand binding in Figure 6.7. If the MSF of any residue in the complex increases (decreases) by more than 20% with reference to the apo runs, it is shown in red (blue) based on the average over three runs. In the presence of bt10, the residue mobility increases in subunit A, whereas it decreases in subunit B. This effect is due to the asymmetric shape of bt10.

6.2.2. Specific interactions

In order to gain insight on intra- and inter-molecular interactions, we analyzed π stacking and H-bonding in the presence and the absence of the ligand bt10. We first concentrated on two identical aromatic clusters (Figure 6.8) at the interface formed by Phe75 of one subunit and Tyr102, 103 of the other subunit, which have been stated to be important for the stability of the dimer [171,173,174]. These clusters are located at the ends of the tunnel-shaped region in the interface (Figure 6.8). It has been proposed that benzothiazole inhibitors bound to this region destabilize the aromatic clusters, eventually leading to disruption of the dimer and inactivation of the enzyme [173,174].

The inhibitor bt10 interacts with the aromatic clusters throughout our MD runs. Contrary to the disruption hypothesis, our π-π distance analysis indicates that binding of the ligand enhances stability of the aromatic clusters. Table 6.1 summarizes the percentages of intact π-π interactions within these clusters during simulations. Here the criterion for interaction is taken as the distance between two aromatic rings' centroids being less than 7 Å [175]. π stacking is enhanced more in cluster 1. A slight decrease is observed between Tyr102 (A) - Tyr103 (A) (cluster 2), but still both aromatic clusters are stable when the ligand is bound. Tyr103 of both subunits predominantly interacts with the ligand as seen in Figure 6.8.

The total number of H-bonds between subunits A and B, which is calculated as an average over all snapshots, increases from 6 $\pm$ 2 in apo to 7 $\pm$ 2 in complex runs. There are 1 $\pm$ 1 additional H-bonds between the inhibitor and interface residues in each snapshot. Specifically, SO3- group of bt10 mostly interacts with Arg71 (A), Lys113 (A) and Tyr103 (B) shown as green in Figure 6.9. These findings also confirm

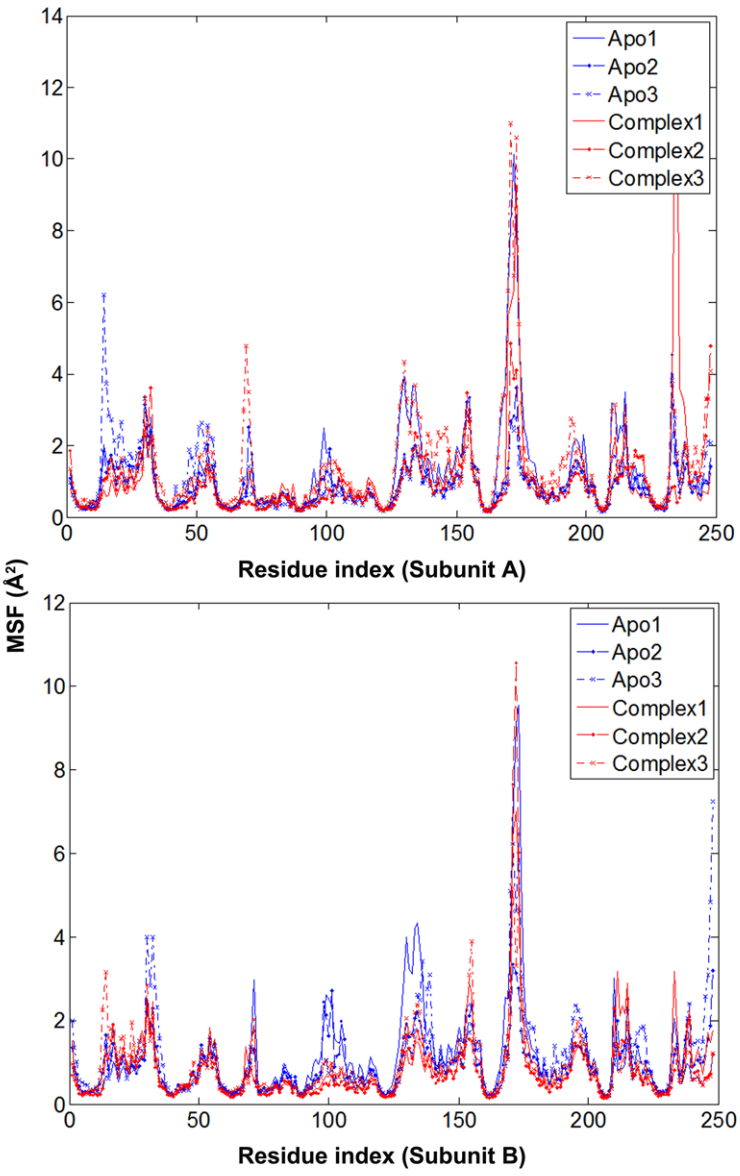

Figure 6.6. MSF of apo and complex TIM for each run.

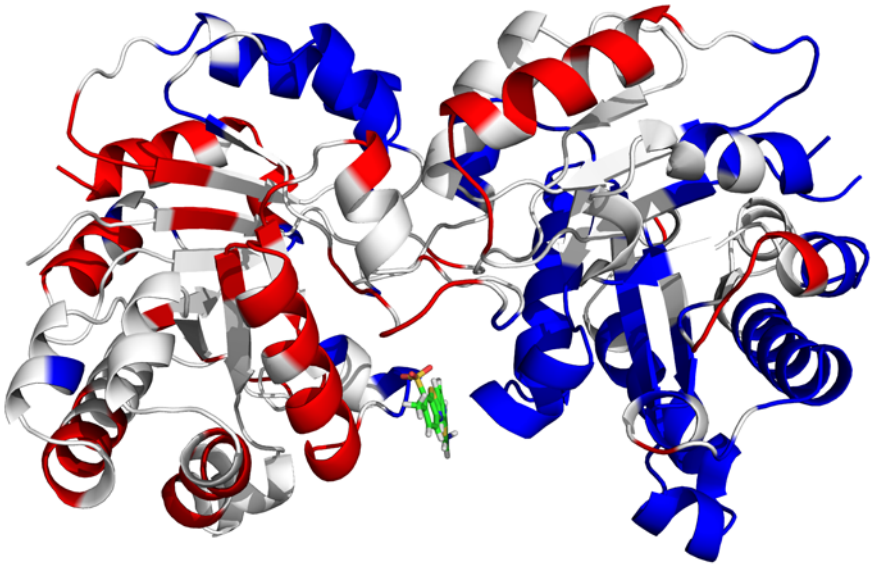

Figure 6.7. Blue (red) residues' MSFs decrease (increase) upon bt10 binding by more than 20% compared to apo runs.

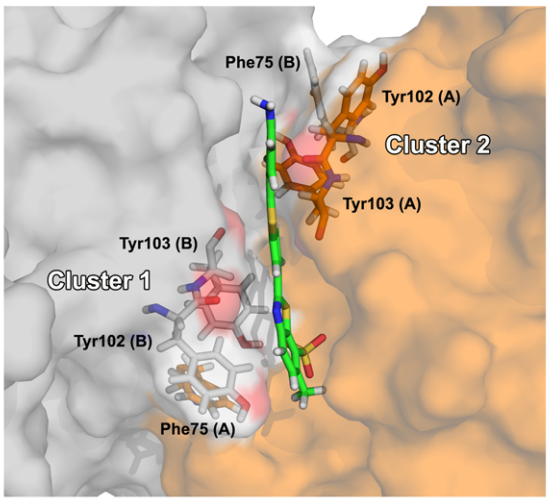

Figure 6.8. Aromatic interactions on bt10-bound tunnel region in TIM interface.

Table 6.1. Percentage of intact $\pi - \pi$ interactions in apo and complex runs.

Aromatic cluster	$\pi - \pi$ pair	Apo	Complex
1	Phe75 (A)-Tyr102 (B)	43	95
1	Phe75 (A)-Tyr103 (B)	95	90
1	Tyr102 (B)-Tyr103 (B)	22	66
2	Phe75 (B)-Tyr102 (A)	79	96
2	Phe75 (B)-Tyr103 (A)	85	91
2	Tyr102 (A)-Tyr103 (A)	70	53

the stability of dimeric structure in the presence of ligand.

We further investigated changes in the H-bonding network of the enzyme upon ligand binding. If the occurrence of H-bonds between any two residues changes by more than 15%, those residues are colored as red (increase by 20%), orange (increase by 15%), cyan (decrease by 15%) and blue (decrease by 20%) in the Figure 6.9 and summarized in Table 6.2. Especially, H-bonding between the subunits are influenced, whose effect propagates towards loop 6 of both subunits. H-bonds are enhanced mostly towards subunit B, which may explain the reduced MSF values of the same subunit.

Three loop 3 residues (Thr76, Gly77, Glu78) of subunit A exhibit increased H-bonds with their partners from subunit B, when bt10 is bound (shown in bold in Table 2). It has been stated each Thr76 contributes to the H-bonding network of the other subunit's active site [158]. Inter-subunit H-bonding between Thr76 and catalytic residue Asn12 is present in about 10% of the snapshots and we do not observe a significant change in the complex. Instead there is an enhancement in H-bonding between Thr76 (A) and Glu98 (B). As for the catalytic loop 6, the H-bonding occurrence increases for Ser97-Glu168 pair on subunit B and decreases for Thr171-Trp175 on subunit A. H-bonding pairs were extracted using VMD Hydrogen Bonds analysis, where the default criterion of 3 Å distance between donor and acceptor and 20 degrees supplementary angle is used.

Table 6.2. Occurrence % of H-bonding residue pairs differing more than 15%.

Donor Residue	Acceptor Residue	Apo	Complex	Color
Glu19 (B) Main	Asp86 (A) Side	37	13	Blue
Cys40 (A) Main	Lys60 (A) Main	7	24	Orange
Thr45 (B) Side	Glu78 (A) Main	12	44	Red
Thr45 (A) Side	Glu78 (B) Main	35	12	Blue
Thr52 (B) Side	His48 (B) Main	35	51	Orange
Thr52 (A) Side	His48 (A) Main	31	53	Red
Thr76 (A) Side	Glu98 (B) Side	55	81	Red
Gly77 (A) Main	Gln66 (B) Side	28	46	Orange
Ser97 (B) Side	Glu168 (B) Side	38	57	Orange
Arg99 (B) Side	Glu105 (B) Side	43	6	Blue
Lys113 (B) Side	Glu105 (B) Side	26	42	Orange
Thr131 (A) Side	Glu134 (A) Side	49	14	Blue
Thr131 (B) Side	Glu134 (B) Side	36	16	Cyan
Arg139 (A) Side	Glu134 (A) Side	17	44	Red
Arg139 (B) Side	Glu134 (B) Side	24	39	Orange
Thr175 (A) Side	Trp171 (A) Main	35	17	Cyan

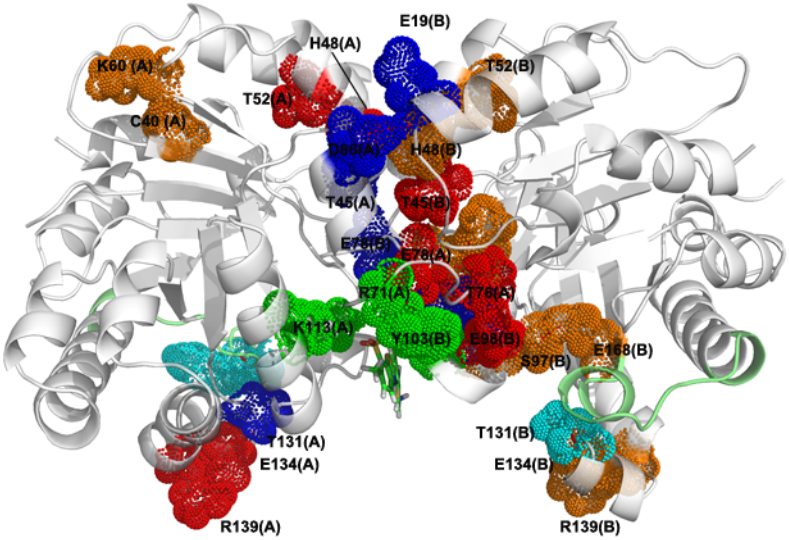

Figure 6.9. Change in H-bonding interactions upon bt10 binding to the interface.

Overall, the increase in both the number of inter-subunit $\pi-\pi$ interactions and H-bonds indicate a stronger interaction network, i.e. a constrained region at the interface, which also affects the collective dynamics significantly.

6.2.3. Collective and Loop 6 Dynamics

We applied PCA on each trajectory to extract the essential anharmonic motions (PCs) of both apo and complex enzyme. In previous MD and ENM studies [164–166, 169], counter-rotation of the subunits (Figure 6.10) has appeared as a common, global mode driving loop 6 opening/closure over the active site in apo TIM of two different species, namely chicken and Trypanosoma cruzi. In current work, this mode can also be detected among the first three or four PCs of the apo and complex runs (Figure 6.10). Another common PC is the bending of the subunits. Details of PCA are summarized in Figure 6.10 and Table 6.3.

As it can be observed from pairwise overlap of PCs, although there are overlap values as high as 0.75, our 100 ns runs are not long enough for the convergence of collective motions extracted from PCA, since different regions of the conformational

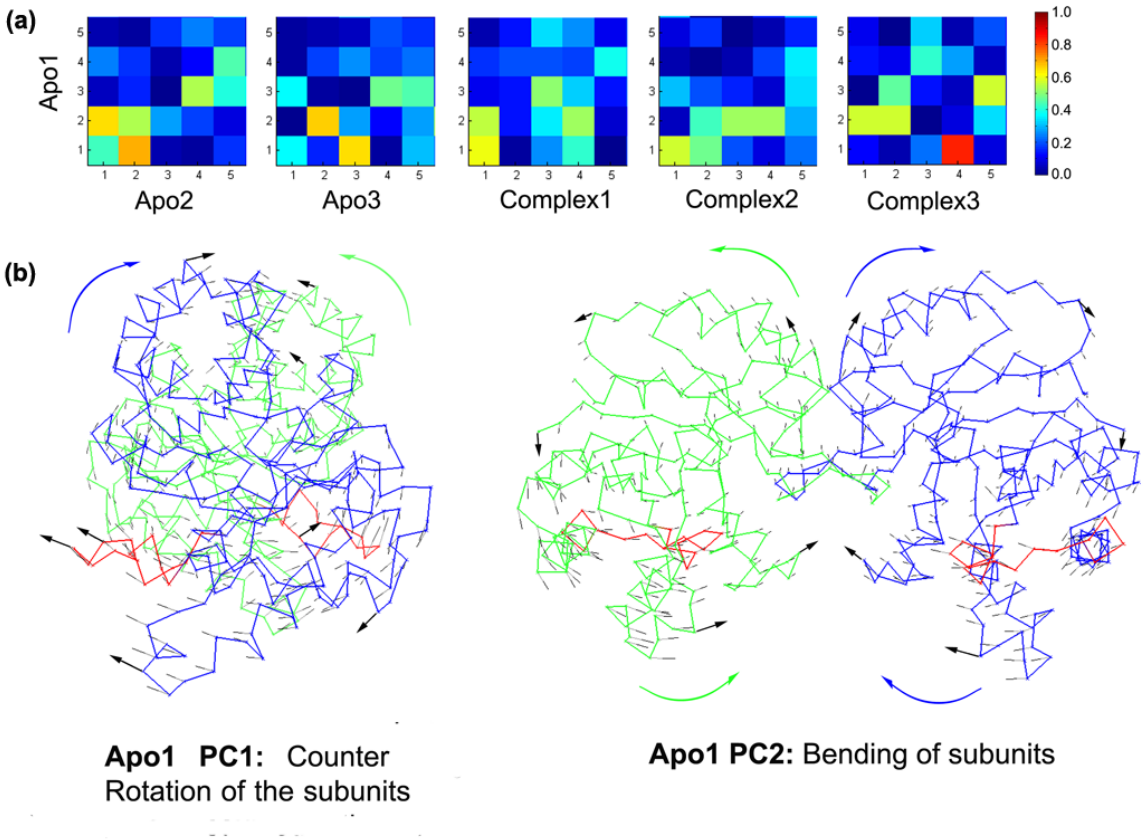

Figure 6.10. PCA on TIM runs: (a) PC overlap matrices with reference to Apo1 PCs, (b)PC1 and PC2 for Apo1.

space is sampled (see Figure 6.2). To solve this issue, we also applied ENM on the average structure of each run, and calculated the overlap between PCs and ENM modes, which describe long-term global dynamics of the system. We observed reasonable similarity between lowest frequency ENM and PCs, indicating that global vibrational modes are in fact present in our simulations.

Table 6.3. Individual and cumulative contributions of PCs to total motion.

MD runs	PC1	PC2	PC3	PC4	PC5	Cumulative contribution of first 5 PCs	# of PCs that explain 90% of motion
Apo1	36.5	9.5	7.5	3.6	3.0	60	100
Apo2	17.0	15.6	6.6	5.0	4.0	48	139
Apo3	19.9	13.7	9.1	5.3	4.2	52	122
Complex1	21.2	14.5	6.0	3.9	3.0	49	134
Complex2	17.6	9.0	7.1	5.4	3.9	43	153
Complex3	27.2	10.5	6.6	5.1	4.0	53	109

Moreover, residue orientational cross-correlation maps are extracted based on the cumulative action of the first five PCs, and also more than 100 PCs that contribute to 90% of total motion in Figure 6.11a, Figure 6.11b and Figure 6.11c, Figure 6.11d, respectively. The figures indicate that strength of correlations for both intra- and inter-subunits decrease in the presence of bt10.

We also investigated the effect of the ligand on catalytic loop dynamics. We extracted the cross-correlations of the loop 6 tip residues, namely I173-G174-T175 with other regions on the enzyme. As seen in loop 6 cross-correlation plots in Figure 6.12, the presence of inhibitor occludes the cooperativity of the catalytic loop. Below the graphs, the structure is colored based on the correlation values of the catalytic loop, in the presence and absence of the inhibitor. Especially, the positive correlations of loop 6 with loops 7 and 8, emphasized in terms of functional relevance in a recent study by Katebi and Jernigan [176], decrease in the complex.

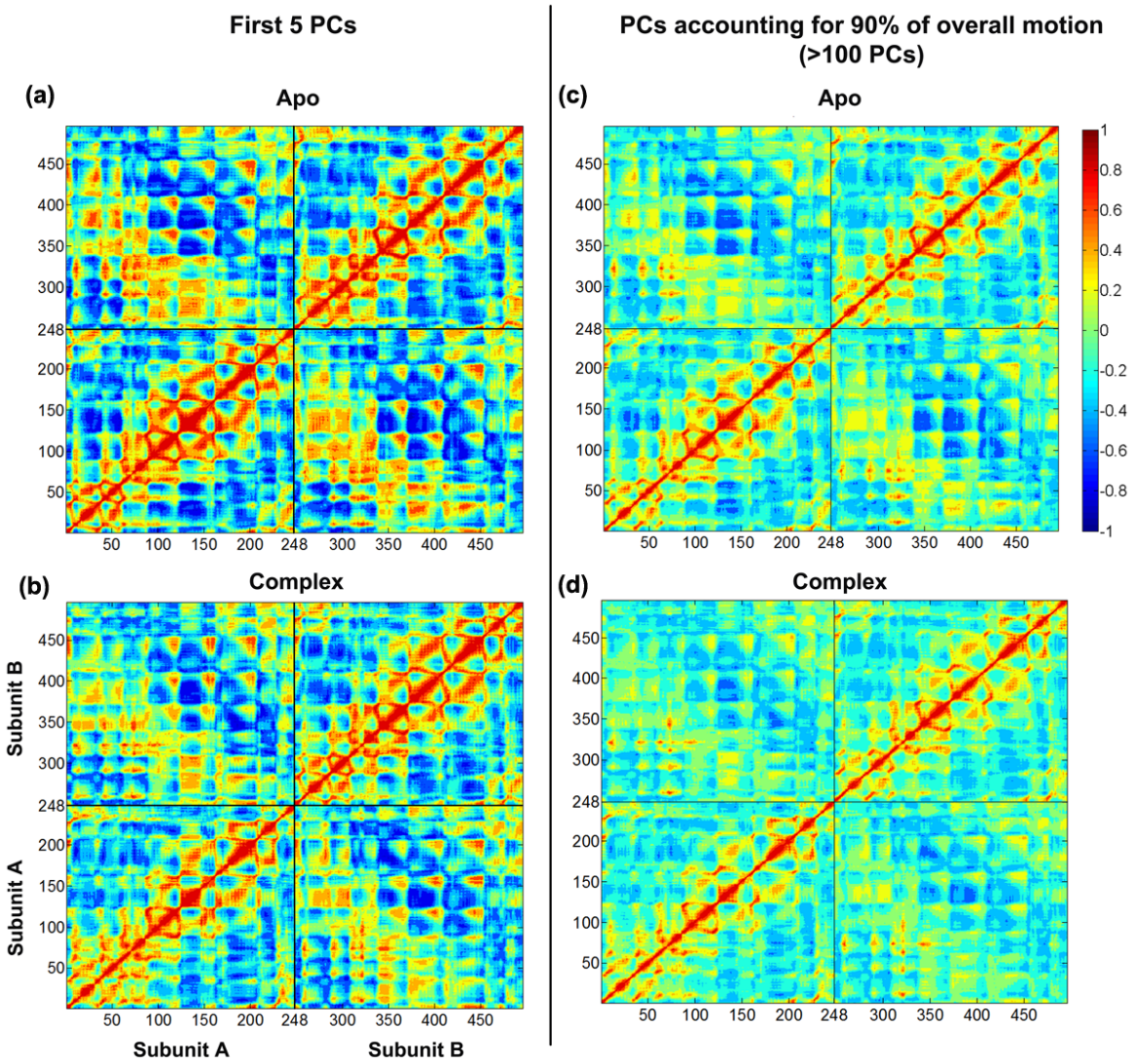

Figure 6.11. Residue-residue orientational cross-correlations based on the cumulative action of first five PCs for (a) apo, (b) complex and more than 100 PCs contributing to 90% of total motion for (c) apo and (d) complex, reported as averages over three independent runs.

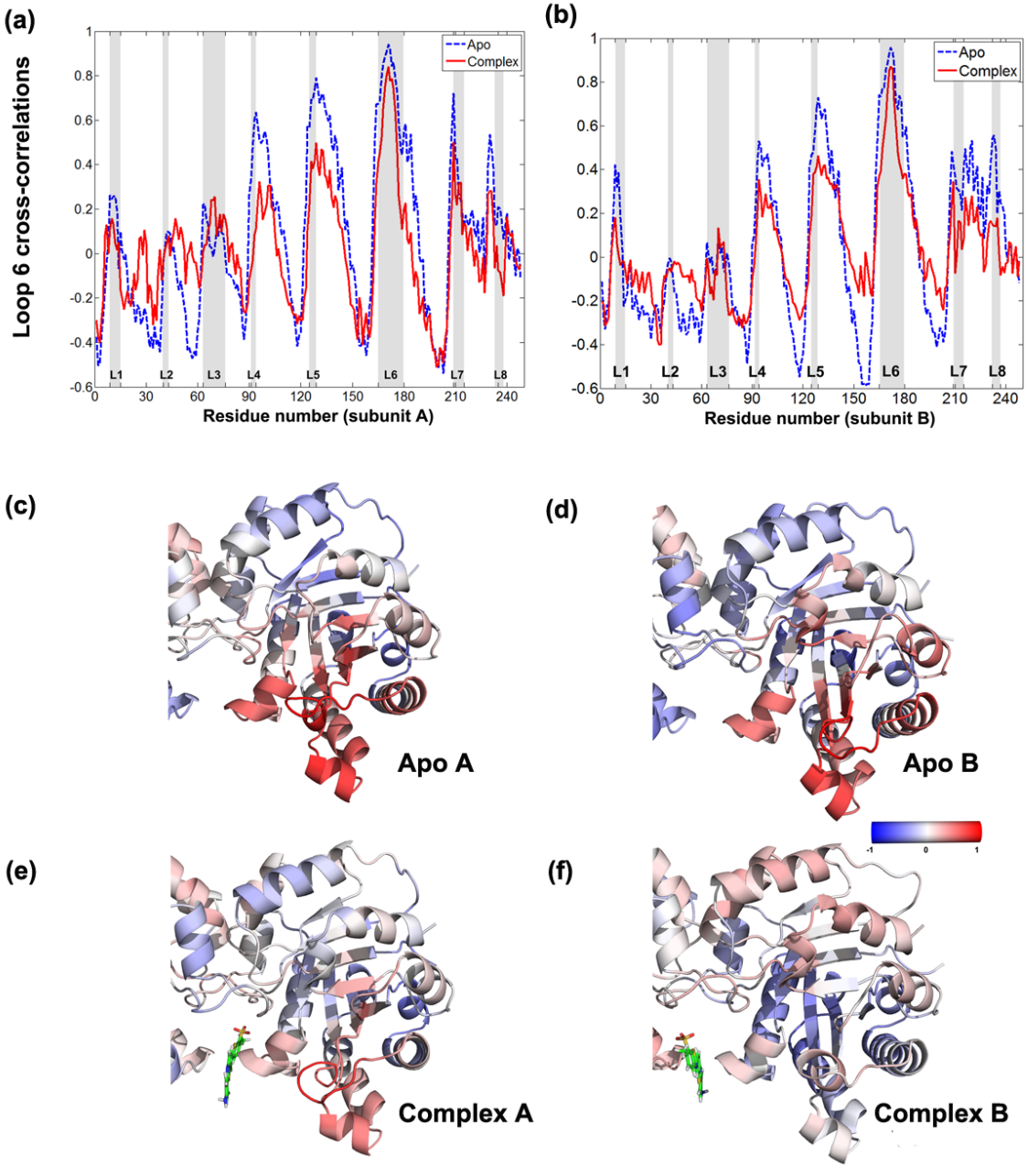

Figure 6.12. Loop 6 orientational cross-correlations based on tip residues I173-G174-T175 for subunit A (a, c, e) and subunit B (b, d, f).

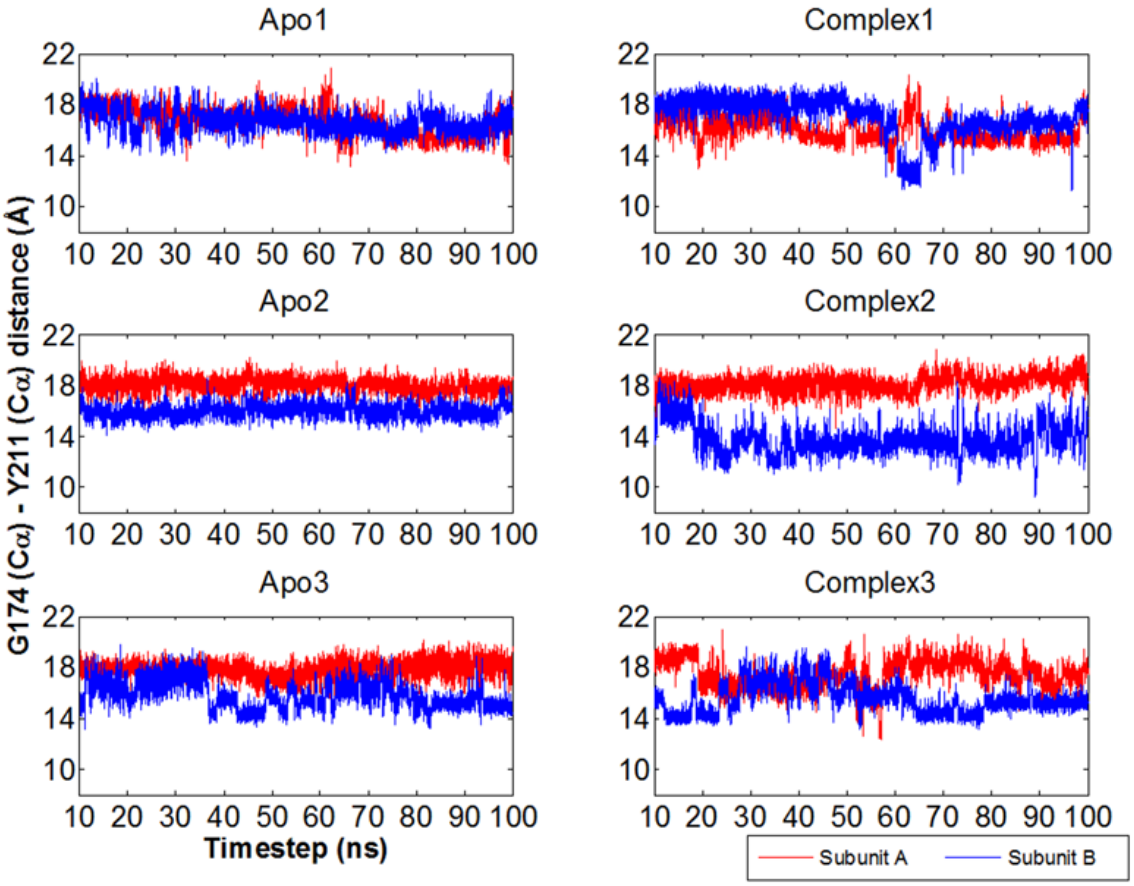

Figure 6.13. Loop 6 opening and closure observed in the monomers based on the distance between loop's tip residue G174 and a relatively immobile residue Y211.

To assess the effect of the inhibitor presence on loop 6 opening/closure event, the distance between alpha-carbons of loop 6 tip residue (Gly174) and a relatively immobile residue Tyr211 at the beginning of loop 7 [164, 165, 177] is calculated throughout the simulations. In our current apo and complex runs, multiple opening and closure events of the loop are observed during 100 ns (Figure 6.13). Figure 6.14a gives normalized distributions for Gly174-Tyr211 distance, which indicate a wider range (9.5-21 Å) being sampled in complex runs, compared to 13-20 Å in apo.

We assessed the flexibility of loop 6 by calculating its pseudo-dihedral angles in apo and complex runs. The ith pseudodihedral angle connecting residues i and $i+1$ is calculated based on the Cα\$$coordinates of residues$i-1,i,i $i+1$ and $i+2$. Both curves in Figure 6.14b give RMS fluctuations of pseudo-dihedral angles averaged over chains A

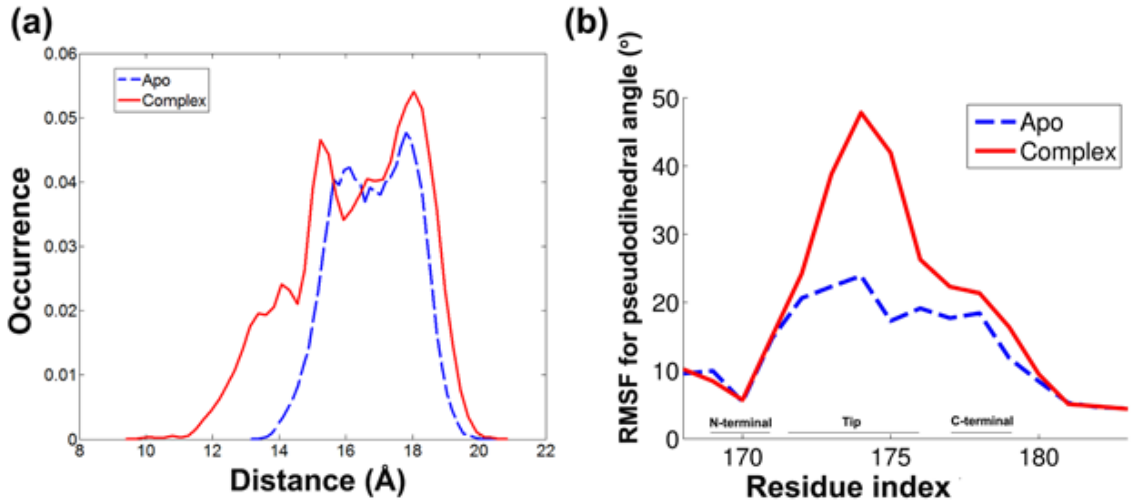

Figure 6.14. (a) Loop 6 opening/closure event from G174-Y211 distance histogram in apo and complex. (b) RMSF for pseudodihedral angle in apo and complex runs.

and B. Previously, N- and C-terminal residues (P169-W171, K177-A179 respectively, shown in the figure) have been detected as hinges that drive the rigid lid closure upon binding of the substrate, based on comparison of one apo and one bound crystal structure [178]. However here, the loop flexibility in the apo state is apparent in the tip and C-terminal residues, in conformity with previous computational and NMR studies [164, 165, 176, 177, 179]. In the presence of the inhibitor, the loop's tip residues (especially Ile173, Gly174, Thr175) become even more flexible.

Moreover, solvent accessible surface area values of the catalytic residues Asn12, His96 and Glu165 reduce in the presence of inhibitor and also sample less number of conformational states in the presence of ligand (results not shown). This is in line with the constraining effect of the inhibitor observed from the MSF.

Stabilization of the aromatic clusters at the interface and enhancement of inter-subunit H-bonds indicate an allosteric effect of the inhibitor, rather than destabilization of the dimeric structure. Inhibitor also modifies collective dynamics of the enzyme, as well as the catalytic loop's dynamics, flexibility and correlations, and solvent exposure of catalytic residues.

6.3. Substrate DHAP effect on TIM global and catalytic loop dynamics

The opening/closure of the catalytic loop 6 over the active site in apo TIM has been previously shown to be coupled with the global motions of the enzyme, specifically the counter rotation of the subunits [164, 166–169, 180]. We investigated at which extent DHAP presence in the catalytic site affects this coupling and TIM dynamics, by analyzing independent apo and DHAP-bound chicken TIM MD simulations. We carried out two apo and two complex simulations where DHAP is present at one or both catalytic sites (60 ns each).

Residual MSF profiles for the apo and complex shown in Figure 6.15 reflect the constraining of the enzyme overall mobility in the presence of the DHAP. Interestingly, this effect is not apparent in residues that are in close contact with DHAP (see inset in Figure 6.15). In contrast, mobility of relatively distant residues that include the interface between subunits and some outlying helices (shown in blue in the inset) are effectively reduced. Specifically, only three residues (shown in red) that are located on loop 6 exhibit higher MSF in the complex: P166 at the N-terminus (23% increase in MSF), T175 and A176 at the C-terminus (45%).

In all simulations, multiple opening/closure events of loop 6 take place even in the presence of DHAP. The distances between the catalytic loop tip residues I170, G171, T172 and residue Y208 reach higher values when DHAP is bound, meaning that the loop adopts more open positions compared to the apo enzyme, as seen in Figure 6.16. Moreover, the average pseudodihedral RMSF shows that the conformational flexibility of the loop 6 is increased in the complex, especially toward the C-terminus.

We also carried our PCA to reveal dominant global motions in the simulations. We reported the findings in the work of Kurkcuoglu and Doruker [165]. Basically, PCA indicate that in apo enzyme, the dominant mode is the counter rotation of subunits that drives the opening/closure of the catalytic loop, in conformity with our previous findings [164, 166, 169]. In contrast, this mode is suppressed in the complex and the bending of the subunits appears in the dominant PCs coupled to loop opening/closure.

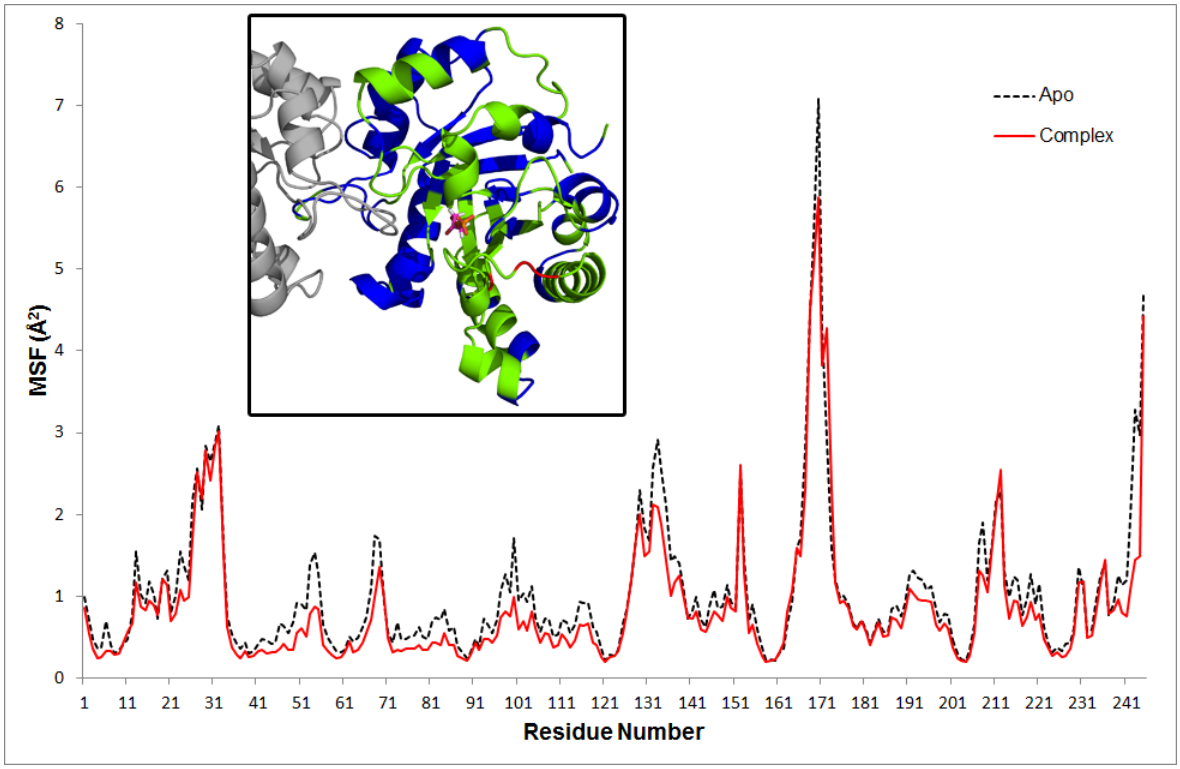

Figure 6.15. Residue MSF for apo and DHAP complex simulations.

Furthermore, the normalized orientational cross-correlations are altered into two anti-correlated blocks, as a result of ligand binding. Moreover, the correlations of the loop with the rest of the enzyme are also affected to a great extent.

In summary, the specific interactions of DHAP modify the global and loop dynamics to a certain extent, which have not been observed in previous studies based on purely entropic ENM [169]. In another MD study, the frequencies of intra-minimum motions based on short MD runs have been shown to be altered in the presence of DHAP [180]. Still our simulations indicate that multiple loop opening/closure events are present and coupled to the global motions of the enzyme, both in apo and complex form. Therefore, entropic effects seem to have a significant role in the mechanisms of DHAP binding/release in TIM.

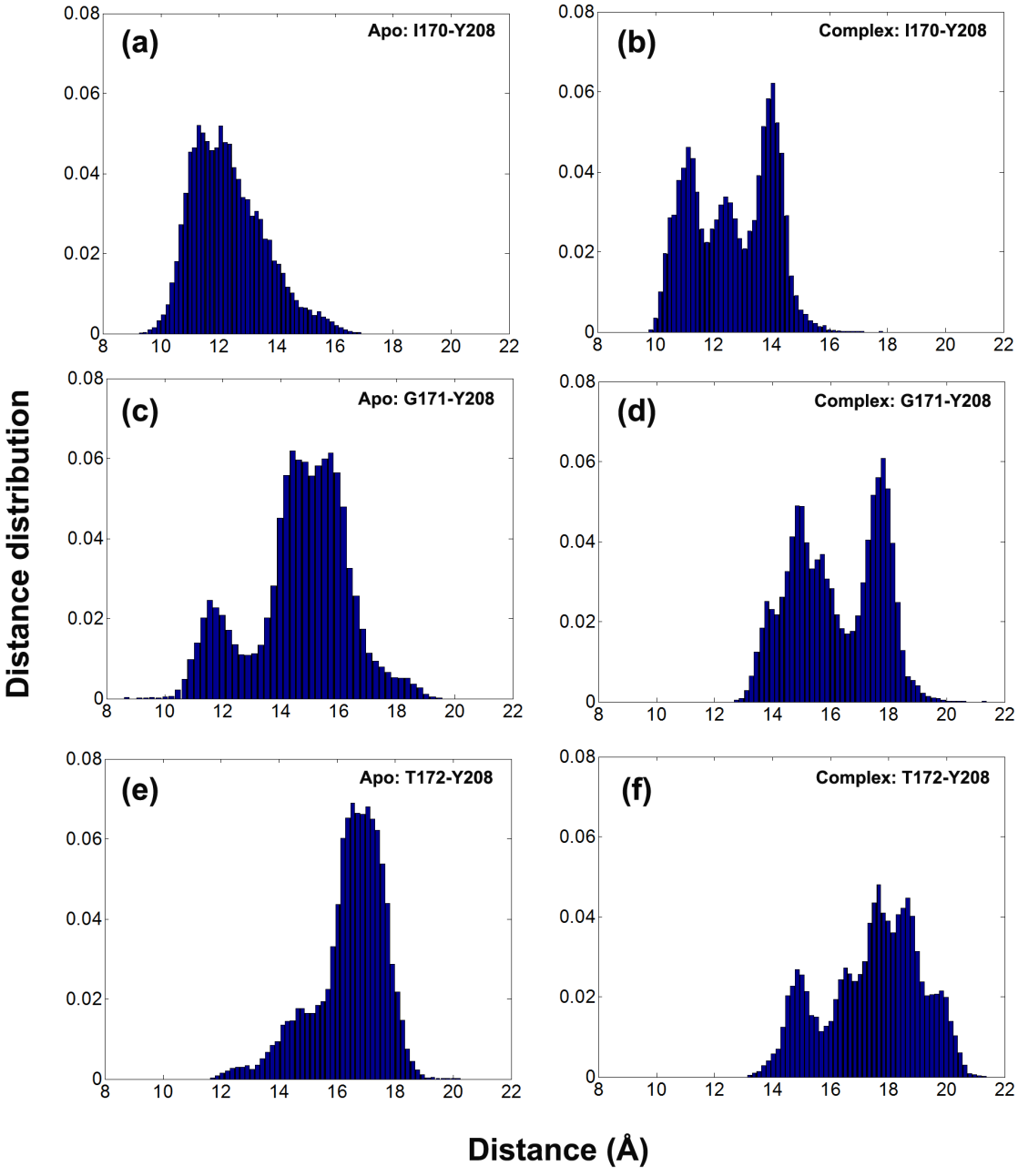

Figure 6.16. Distance histograms for I170-Y208, G171-Y208 and T172-Y208 apo and complex simulations.

7. CLUSTENM-II: APPLICATION IN SMALL LIGAND DOCKING

The flexible nature of proteins is one of the challenging problems in drug design studies since even small conformational changes can affect the nature of ligand-protein interactions [181]. Thus, efficient computational algorithms are necessary to sample protein conformations for more accurate prediction of binding sites and affinities in docking studies. Accounting for the backbone flexibility is still challenging for small ligand-protein docking, as well as for the protein-protein and protein-DNA/RNA cases. We applied ClustENM Procedure II to four proteins undergoing large conformational changes, namely AK, LAO binding protein, DBP and BC, to assess the usage of generated conformers in small ligand docking applications, as well as to consider the protein flexibility in docking. As explained in Methods, the main difference in Procedure II is the exclusion of parent structures in order to sample as many distinct states as possible.

7.1. Adenylate kinase

Energetically-minimized open structure of AK (4ake) was used for generating the first cycle of protein conformers. Conformational search of AK was initially performed using blind search approach, i.e. without any filtering criteria. In the first generation, three distinct atomistic conformers were obtained. Then, the procedure was repeated until the completion of seven generations. The closeness of distinct conformers generated at each cycle to ligand bound crystal structure is reported in Table 7.1, based on their backbone RMSD. As the number of iterations increases, closer structures to complex are obtained, but the total number of generated conformers also increases. For example at the 7th iteration, there are 8 conformers relatively closer to the bound structure (RMSDs < 4 Å) among 149 conformers. At the same time, we get 100 conformers at this final iteration that lie farther away from the ligand-bound structure than the starting apo form (initial RMSD of apo structure is 7.2 Å). We will refer to these conformers as "outliers" in this text.

In an alternative conformational search, we applied the energy-based selection by comparing the energy values (obtained from minimization in implicit solvent) of the minimized structures to the energy of the minimized open crystal structure. The conformers having lower energy than apo structure were selected as parent conformers for creating the next generation and those with higher energy were discarded. Again, RMSD values with respect to the closed structure are reported in Table 7.1 (on the right-hand side of each cell). At the 7^{th} iteration, we obtain 7 structures out of 38 with RMSDs < 4 Å.

Table 7.1. Number of AK conformers from blind search/energy-based selection.

Generation/ cycle	Total number of conformers in each cycle	Number of conformers within specific RMSD range to closed structure (Å)					
		$2-3$	$3-4$	$4-5$	$5-6$	$6-7.2$	>7.2
1	3/3	0/0	0/0	0/0	0/0	1/1	2/2
2	10/9	0/0	0/0	0/0	2/2	1/1	7/6
3	18/9	0/0	0/0	4/3	1/0	3/2	10/4
4	35/13	0/0	4/1	2/3	4/1	3/4	22/4
5	55/22	1/1	4/2	3/5	1/3	10/3	36/8
6	102/27	3/2	3/4	8/6	7/3	12/5	69/7
7	149/38	2/1	6/6	10/6	12/6	19/7	100/12
All cycles	372/121	6/4	17/13	27/23	27/15	49/23	246/43

At the end of 7^{th} cycle of the blind search, the structure with closest RMSD value to bound structure has an RMSD of 2.8 Å and this value is 2.4 Å in the energy-based search. However in 5^{th} and 6^{th} cycles of both search algorithms, structures in the range of 2.7-2.8 Å are also obtained. At the end, total number of distinct conformers obtained from the blind search and energy-based selection are 372 and 121, respectively. Most of the outliers are eliminated when the energy criterion is applied and 14% of distinct conformers have an RMSD < 4 Å to complex (compared to 6% in blind search).

Furthermore, RG of the apo structure can be used as a post-filtering criterion in

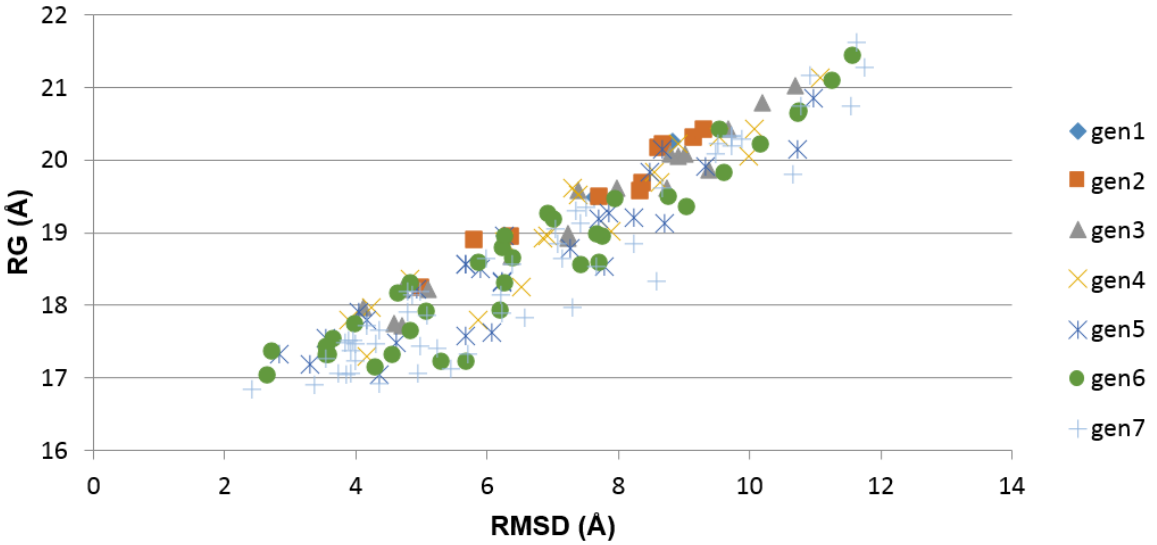

Figure 7.1. RG vs RMSD correlation for AK (energy-based search).

order to eliminate conformers with larger RGs. Application of RG filtering on energy-based search discards 29 conformers, all being outliers, and increases the percentage of close-to-bound conformers to 18%.

In fact, when RG vs. RMSD-to-closed crystal structure values are plotted for the generated conformers of energy-based search, a direct correlation is observed meaning that as the conformer gets closer to the bound structure, there is a reduction in its RG. Figure 7.1 shows this direct correlation, which is consistently observed for all four proteins undergoing large hinge-bending type motion, used in this study. Thus, discarding conformers with larger RG than the apo form seems to be reasonable if one is interested in conformers closer to the bound structure or the intermediate states between open and closed forms. Moreover, the conformer with lowest RG value can be used as a starting point for docking, which is explained below.

For the docking of inhibitor AP5 on the generated AK conformers, we selected the conformer from 7^{th} cycle of energy-based search with the lowest RG (gen7) among all conformers and traced back its parents up to the first generation (gen6, gen5,..., gen1). We performed dockings on this sequence of conformers (shown in Figure 7.2a, b) together with apo structure, and the results are summarized in Table 7.2. The RMSD

of the docked ligand with respect to its positioning on closed crystal structure (1ake) is reported for the first cluster of docking, which has the highest score, meaning the lowest binding energy value. Total number of poses is equal to 100 in one docking run. in Table 7.2, number of elements in lowest energy cluster and total number of clusters are also given in the table. Figure 7.2 shows docking poses for gen1, gen4 and gen7, aligned on ligand bound crystal structure (the ligand positioning in crystal structures is shown in magenta stick).

Table 7.2. Docking results for AK using AutoDock v4.

Conformer	Backbone RMSD (Å) to complex	Ligand RMSD (Å) RMSD (Å)	Binding energy (kcal/mol)	Number of elements/ poses	Number of clusters
apo	7.1	7.4	-2.96	100	1
gen1	6.3	7.8	-2.96	100	1
gen2	5.0	4.8	-2.49	98	2
gen3	4.1	3.4	-3.97	100	1
gen4	4.2	2.9	-7.06	100	1
gen5	3.5	6.5	-1.23	100	1
gen6	2.7	15.6	-3.24	45	6
gen7	2.4	5.1	-2.71	95	2
alt1	4.2	2.6	-6.05	100	1
alt2	2.8	5.3	-2.55	100	1
alt3	2.8	5.3	-5.11	8	2

Even though the conformer gen7 obtained in the last cycle lies close to the bound receptor structure, the docked ligand RMSD is 5 Å. We traced back the parents of gen7 and performed dockings on all of them, including apo structure, to see whether lower ligand RMSDs could be obtained. Interestingly, the highest docking score obtained in gen4 at the same time corresponds to a pose with satisfactory ligand RMSD of 2.9 Å. As opposed to the bound crystal structure with fully closed LID and NMP domains, gen4 resembles the closed-LID state where the LID domain is partially closed and

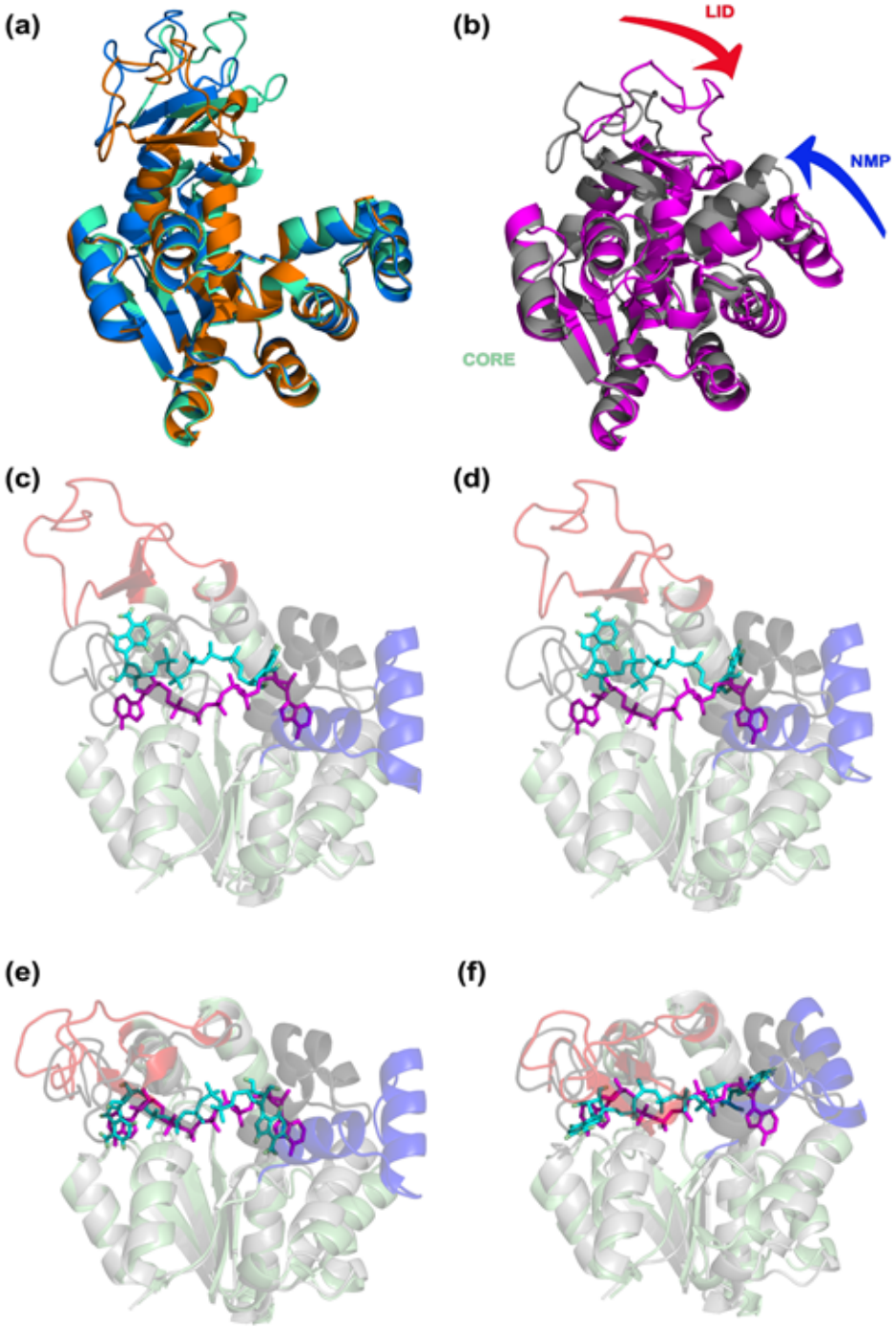

Figure 7.2. AK docking results. Docking poses are shown for (c) apo, (d) gen1, (e) gen4, (f) gen7.

NMP domain is open. In order to check the consistency of these results, we performed additional dockings on three other conformers (denoted as alt1, alt2 and alt3 in Table 7.2). Alt1 that is similar to gen4 leads to successful results with low ligand RMSD and high docking score. The ligand RMSDs for alt2 and alt3 that are closer to the bound structure are found to be around 5 Å, consistent with gen7 results.

Our conformer generation procedure leads to successful docking poses especially for 'in-between' AK state with a partially closed LID and open NMP (closed-LID state) and average success for fully closed domains. As explained in the ClustENM Procedure I AK results, closed-LID state is one of the highly populated intermediate states observed in AMP/ATP binding to AK. Receptor residues interacting with the ligand in gen4 are mainly located on the CORE domain (H-bonds with Gly10, Gly12, Gly14 and Lys200; ionic interactions with Arg167) and one residue on LID domain (Arg123, making hydrogen and ionic bonds).

7.2. Lysine-arginine-ornithine-binding protein

LAO and DBP, which will be discussed in the next section, both belong to the periplasmic binding protein (PBP) superfamily. PBPs serve to transport a wide variety of sugars, amino acids, peptides and inorganic ions into bacteria. Both PBPs consist of two domains connected by a hinge region at the interface, where the ligand binding site is located. Large bending motion around the hinge facilitates the alternation between open/apo and closed/ligand-bound conformations [182]. Specifically, LAO exhibit open-to-closed structural rearrangements amounting to 4.7 Å.

Same procedure was applied starting with the minimized apo crystal structure of LAO binding protein using blind and energy-based search. The results including 7 cycles are summarized in Table 7.3.

At the end of the blind and energy-based searches, the conformer closest to the bound structure has an RMSD of 2.2 Å and 1.7 Å, respectively. In the energy-based selection, the search stops at the 6th cycle because all generated conformers at the 7th

Table 7.3. Number of LAO conformers from blind search/energy-based selection.

Generation/ cycle	Total number of conformers in each cycle	Number of conformers within specific RMSD range to closed structure (Å)			
		$1-2$	$2-3$	$3-4.7$	>4.7
1	4/2	0/0	0/0	0/0	4/2
2	8/3	0/0	0/0	0/0	8/3
3	17/6	0/0	0/0	1/1	16/5
4	26/3	0/0	0/0	3/1	23/2
5	45/3	0/0	1/1	7/1	37/1
6	75/4	0/1	3/2	13/0	59/1
7	114/0	0/0	3/0	15/0	96/0
All cycles	289/21	0/1	7/3	39/3	243/14

cycle have higher energy values compared to open crystal structure. Energy screening reduces the total number of distinct conformers from 289 (in blind search) to 21 and eliminates most of the outliers. As a result, 19% of conformers in energy-based search have an RMSD < 3 Å to the closed structure (compared to 2% in blind search). Application of further RG filtering on energy-based conformers reduces the number of conformers to eight with four having RMSD < 3 Å. RG-RMSD correlation is given in Figure 7.3, displaying direct correlation.

In the case of LAO, energy-based search is more effective than the blind search in various respects: (i) the number of outliers is excessively reduced, (ii) the structures with closest RMSD values to the closed form are obtained sooner than the blind search, and (iii) energy-based selection imposes a "stop criterion" to the conformational search.

For the docking study of LAO binding protein, we again selected a conformer with low RG and traced back its parents throughout the generations. The results for the dockings of lysine on apo, gen1 to 6 structures are summarized in Table 7.4 (details are given in Table A.2). Some of these conformers (gen1 (green), gen4 (orange), apo

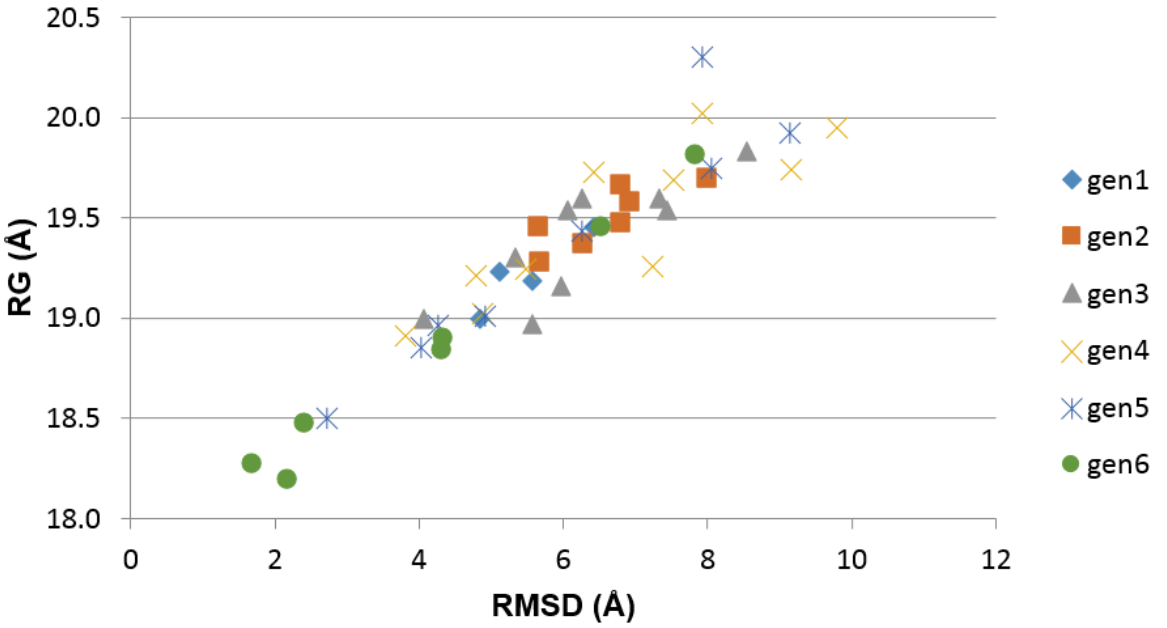

Figure 7.3. RG vs RMSD correlation for LAO binding (energy-based search).

(blue), (b) gen6 (magenta), ligand-bound form (gray)) and docking poses (ligand in cyan stick) for apo, gen1, gen4 and gen6 are shown in Figure 7.4.

In Table 7.4, HADDOCK score and number of elements are only reported for the cluster with the highest score. Total number of docked poses (elements) is equal to 200. Detailed results for all clusters are given in Table A.2.

The conformers closer to the closed structure (gen5 and 6 with backbone RMSD ≤ 2.7 Å) yield highest scores (at the same time most populated clusters) and close-to-native ligand binding poses, i.e. low ligand RMSDs. Moreover, the number of clusters in last three dockings is reduced compared to apo, gen1-3. So peptide binding to LAO seems to be facilitated by almost-full domain closure in the apo state in contrast to the previous case of AK.

7.3. Dipeptide Binding Protein

In the first generation of DBP (which undergoes a conformational change 6.5 Å between open and closed states), three distinct conformers were obtained and en-

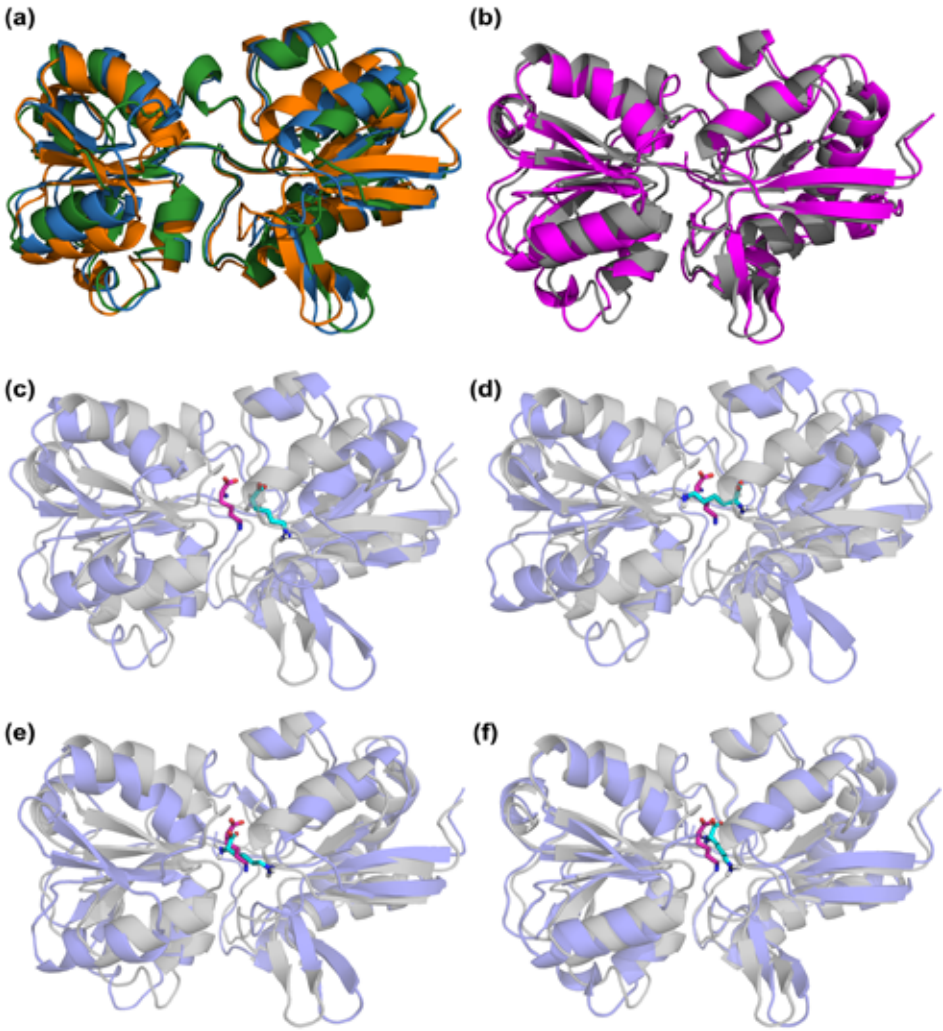

Figure 7.4. LAO binding protein conformers used in HADDOCK: (a) gen1, gen4, apo, (b) gen6, ligand-bound form (gray). LAO docking poses are shown for (c) apo, (d) gen1, (e) gen4 and (f) gen6.

Table 7.4. HADDOCK docking results for LAO conformers (summary).

Conformer	Backbone RMSD (Å) to complex	Ligand RMSD (Å)	HADDOCK score	Number of elements/ poses	Number of clusters
apo	4.7	4.2	-28.6±1.8	41	8
gen1	6.4	5.6	-33.5±4.8	30	8
gen2	7.0	8.3	-33.5±6.1	46	8
gen3	5.3	4.4	-38.1±6.7	43	8
gen4	3.8	2.1	-37.2±0.6	116	3
gen5	2.7	2.2	-46.2±0.7	200	1
gen6	1.7	2.0	-48.5±1.8	122	3

ergetically minimized. However, the energy value of each new conformer was higher compared to that of minimized open crystal structure, therefore energy-based selection was not possible. So we applied only blind search for five generations producing a total number of 100 conformers. If we select conformers having an RG less than apo RG, 37 conformers remain and most of the outliers (61 structures having an RMSD > 6.5 Å) are discarded. RG filtering increases the percentage of the conformers close to ligand-bound conformation (RMSD < 4 Å) from 11% (pure blind search) to 30%. Two closest conformers at the end of the 5th cycle have RMSD values of 1.5 and 1.9 Å. Total number of conformers in each cycle and number of conformations in RMSD intervals are reported in Table 7.5 for blind search.

For the docking of ligand glycyl-leucine, we again selected the conformer with the lowest RG from the last cycle and traced its parents to the beginning (gen1-5, RG vs RMSD plot is shown in Figure 7.5). Docking results for are summarized in Figure 7.6 and Table 7.6. Docking to gen5 conformer, which is closest to the closed crystal structure, yields the highest score and the lowest ligand RMSD (1.5 Å). Thus, in the case of both periplasmic binding proteins, as the conformer gets closer to the ligand-bound form, higher docking scores are obtained and ligand positioning becomes closer

Table 7.5. Number of DBP conformers obtained using blind search.

Generation/ cycle	Total number of conformers in each cycle	Number of conformers within specific RMSD range to closed structure (Å)					
		$1-2$	$2-3$	$3-4$	$4-5$	$5-6.5$	>6.5
1	3	0	0	0	1	0	2
2	8	0	0	0	1	2	5
3	17	0	1	1	1	4	10
4	24	0	1	2	0	4	17
5	48	2	0	4	1	5	36
All cycles	100	2	2	7	4	15	70

to the native form.

7.4. Biotin Carboxylase

In fatty acid synthesis, BC catalyzes ATP-dependent carboxylation of biotin. It has a homo-dimeric structure with each subunit having its own catalytic site. Each subunit consists of three domains; namely A, B, C, where B domain is the ATP grasp domain closing over the globular section formed by A and C domains (undergoing a conformational change of 4 Å). Open and closed forms of BC are shown in Methods section.

The first generation of BC consists of three distinct conformers. We applied blind and energy-based search for four generations, which resulted in the generation of same conformers (i.e. there are no discarded structures in the case of energy-based selection). At the end of both procedures, we obtained a total of 35 distinct conformers. 26 of these conformers have more open structures than the starting conformer in both subunits. At the end of the fourth generation, we obtain conformers closed to the bound structure (lowest RMSDs are 2.3 Å for subunit A, 1.7 Å for subunit B, respectively).

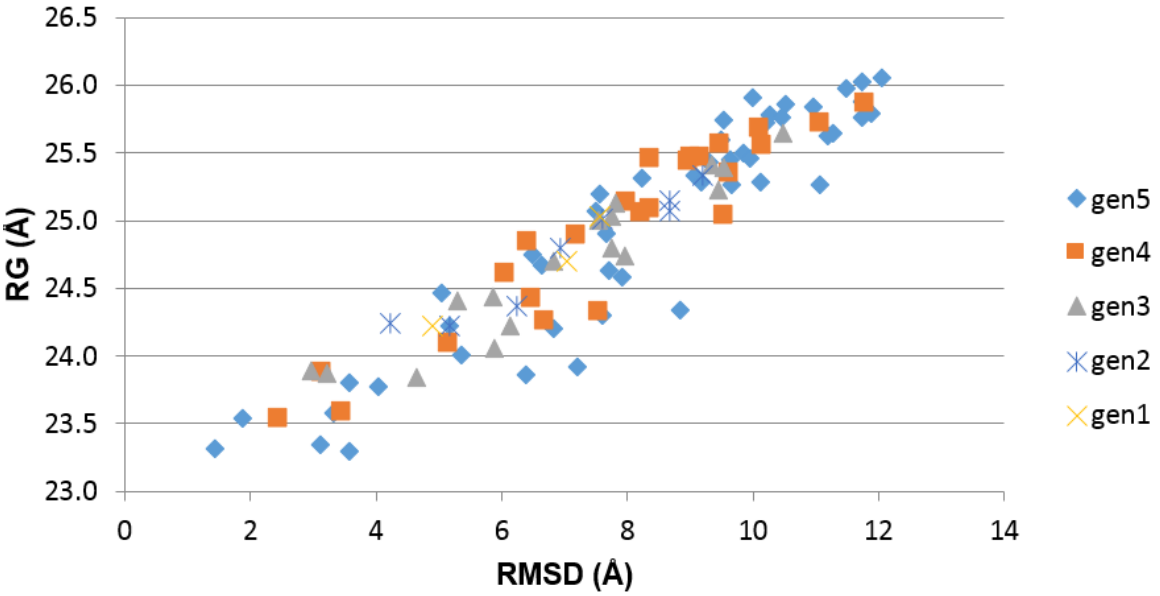

Figure 7.5. RG vs RMSD correlation for DBP protein.

Table 7.6. HADDOCK docking results for DBP.

Conformer	Backbone RMSD (Å) to complex	Ligand RMSD (Å) RMSD (Å)	HADDOCK score	Number of elements/ poses	Number of clusters
apo	6.5	6.1	-19.1±3.3	195	1
gen1	4.9	4.9	-19.1±1.5	199	1
gen2	4.2	3.8	-21.0±1.3	198	1
gen3	3.0	2.2	-20.2±1.9	198	1
gen4	2.5	3.2	-24.9±1.6	200	1
gen5	1.4	1.5	-31.4±2.6	193	2

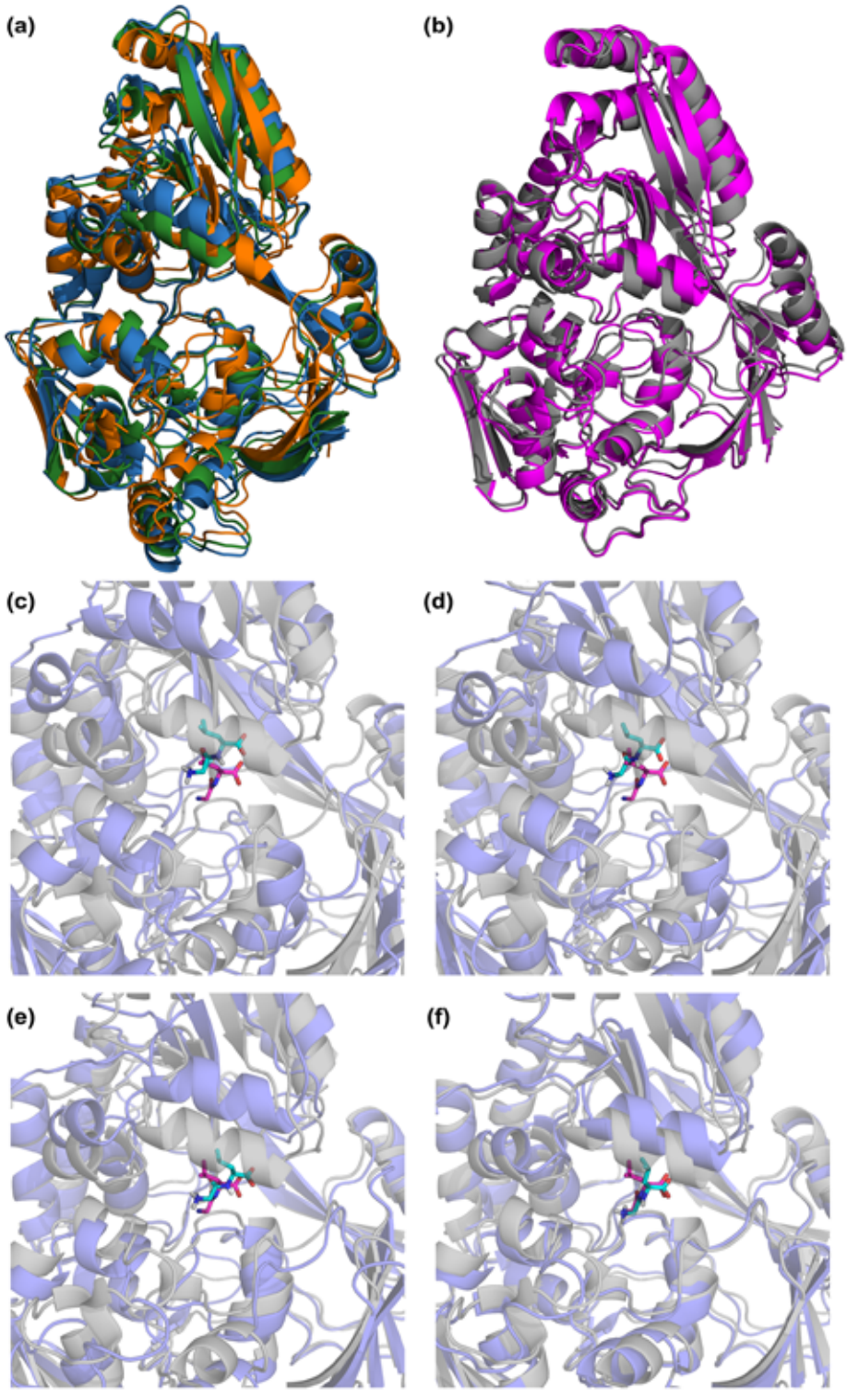

Figure 7.6. DBP conformers used in HADDOCK (a) gen 1 (green) and gen3 (orange) aligned on apo (blue), (b) gen5 (magenta) aligned on closed form (gray). DBP docking poses (purple) for (c) apo, (d) gen1, (e) gen3 and (f) gen5 aligned on closed crystal structure (gray). Ligand in crystal structure and docking pose is shown in magenta and cyan stick representation, respectively.

Moreover, if we filter the generated conformers based on their monomeric RG, 12 conformers remain for subunit A with 3 of them being outliers, and 9 conformers for subunit B with 1 outlier. Thus, post-filtering using RG results in elimination of outliers like in previous cases. Total number of conformers and their RMSD range based on the bound structure are given separately for subunits A and B in Table 7.7 and 7.8, respectively. RG-RMSD plot in Figure 7.7 reflects direct correlation.

Table 7.7. Number of BC conformers from blind search (monomer A).

Generation/ cycle	Total number of conformers in each cycle	Number of conformers within specific RMSD range to closed structure (Å)			
		$1-2$	$2-3$	$3-4.1$	>4.1
1	3	0	0	0	3
2	7	0	0	1	6
3	8	0	2	1	5
4	17	0	3	2	12
All cycles	35	0	5	4	26

Table 7.8. Number of BC conformers from blind search (monomer B).

Generation/ cycle	Total number of conformers in each cycle	Number of conformers within specific RMSD range to closed structure (Å)			
		$1-2$	$2-3$	$3-4.1$	>4.1
1	3	0	0	1	2
2	7	0	1	1	5
3	8	0	1	0	7
4	17	3	0	2	12
All cycles	35	3	2	4	26

We selected the conformer with lowest RG (for subunit B) among the generated conformers and traced back its parents for gathering the conformers to be used in ATP docking to BC subunit B. The conformers are shown in Figure 7.8, together with apo

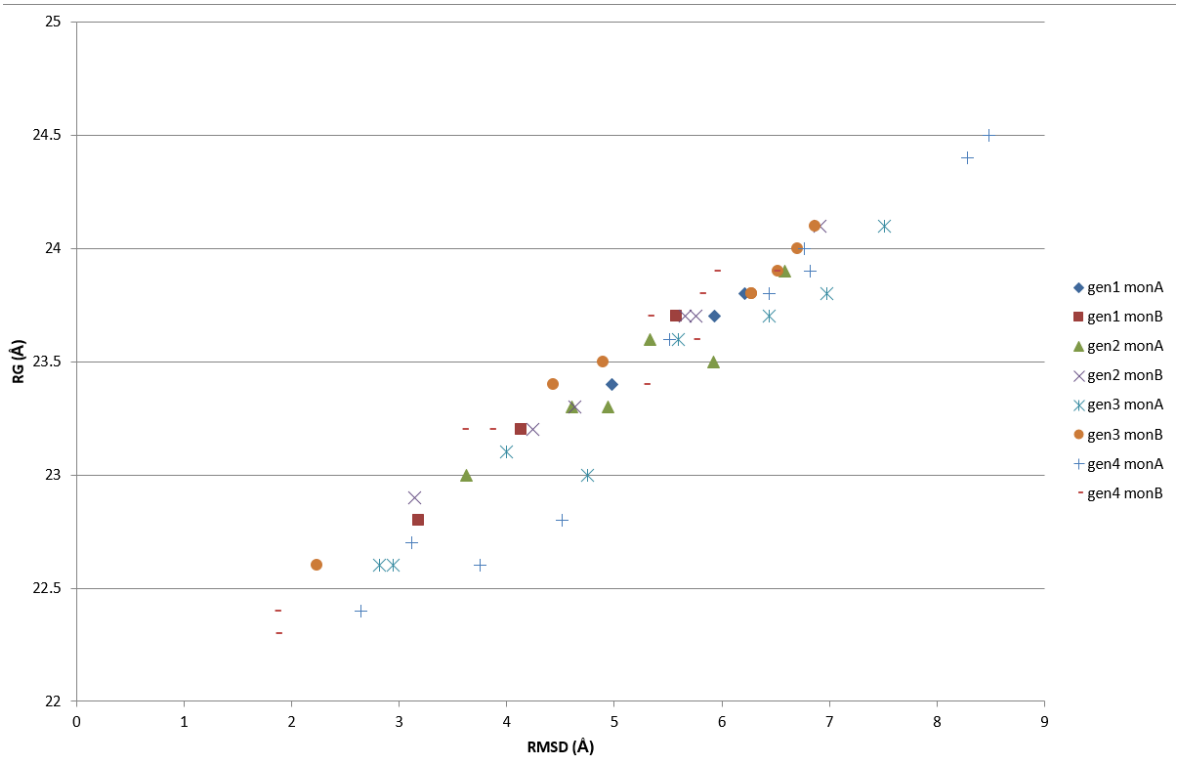

Figure 7.7. RG vs RMSD correlation for BC protein.

and closed structures. Conformers from gen3 and gen4 yield the lowest ligand RMSD, which correspond to most populated clusters, as seen in Table 7.9 (an additional cluster with ligand RMSD of 4.9 Å is also present in gen4 dockings, but it is not populated). Docking poses for all conformers are given in Figure 7.9, indicating that as the B domain closes over the A-C domain, ligand positioning closer to bound structure is observed. Like in the case of peptide binding proteins LAO and DBP, closed conformers yield ligand poses closer to bound structure.

In the current study, we aimed to assess the performance of our ENM-based conformer generation method in small ligand docking, specifically for proteins undergoing large conformational changes (such as AdK, LAO and DBP). Using this method, we were able to obtain conformers close to ligand-bound forms. The method has the advantage of requirement of only the apo structure and the ligand, not the target structure. Beside the low computational cost of the method explained in Conclusions, the number of conformers generated to be used in docking is manageable; one can perform the docking to all generated conformers.

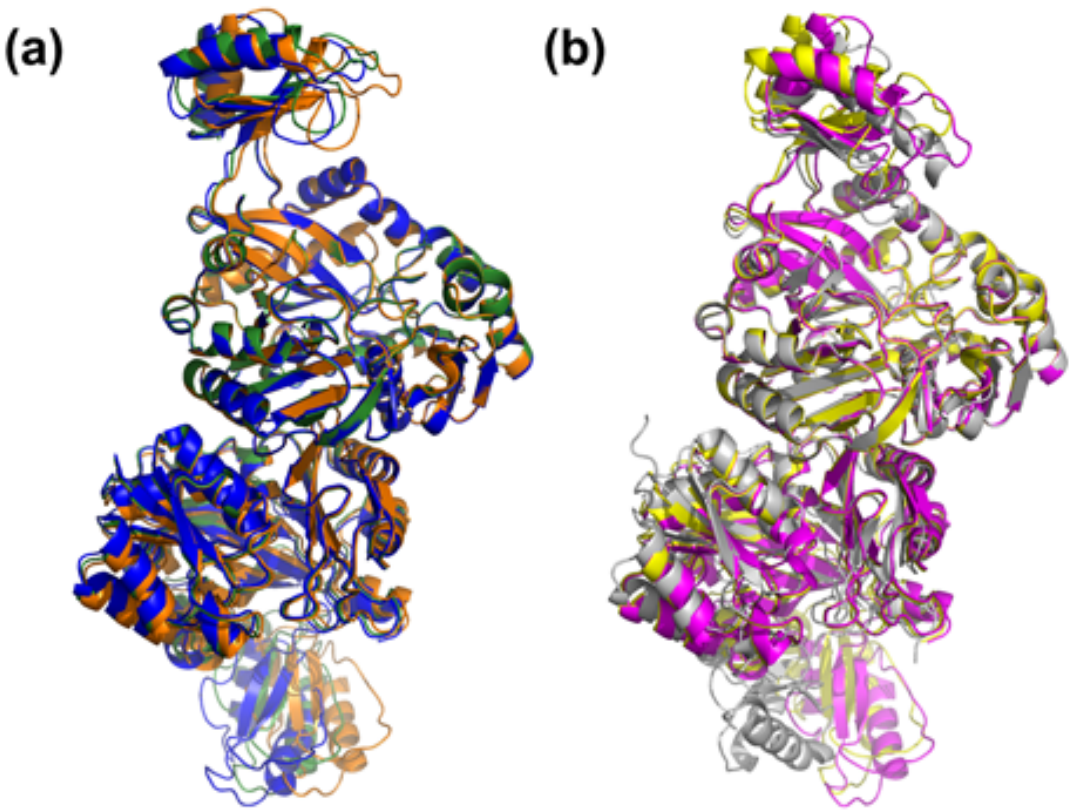

Figure 7.8. BC conformers used in ATP docking (a) apo (blue), gen1 (green), gen2 (orange), (b) gen3 (yellow), gen4 (magenta), ligand-bound form (grey).

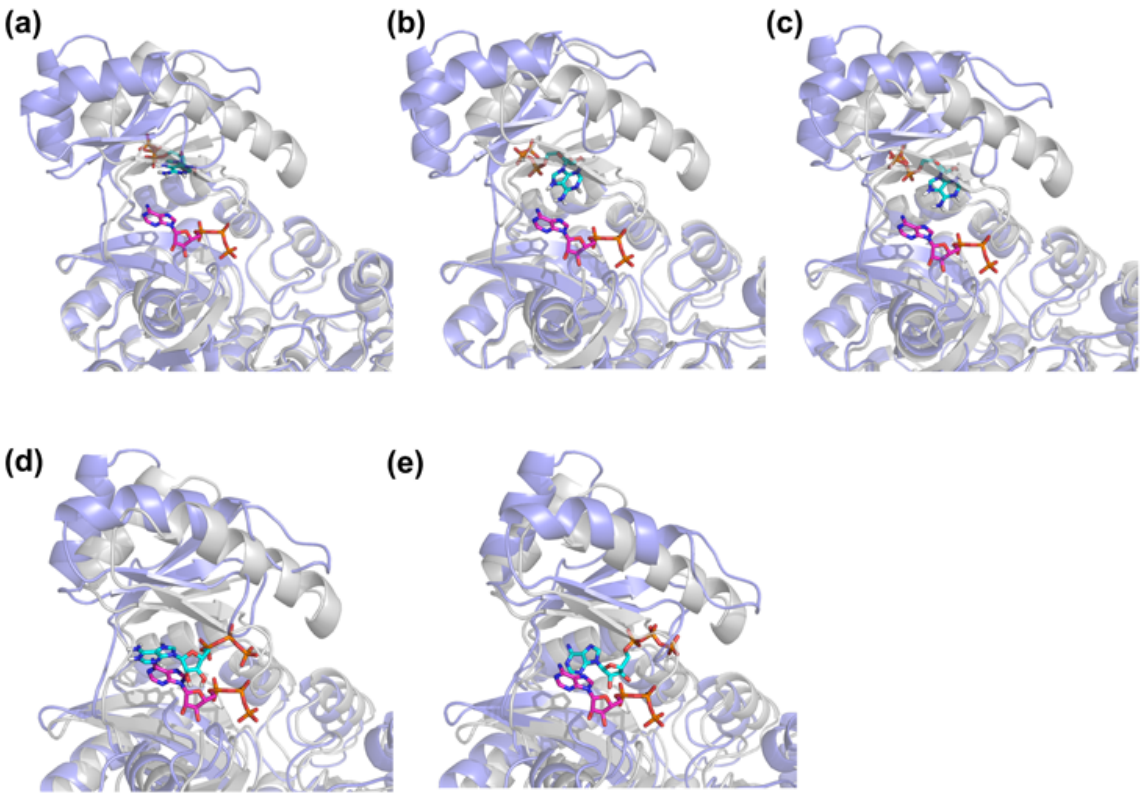

Figure 7.9. Ligand poses for (a) apo cluster 1, (b) gen1 cluster1, (c) gen2 cluster 1, (d) gen3 cluster 4 and (e) gen4 cluster 2 aligned on ligand-bound complex (grey). Ligand positioning in crystal and docked structures is shown in magenta and cyan stick representation, respectively.

Table 7.9. BC binding protein docking results details.

	Backbone RMSD (Å) of to complex (monomeric/dimeric)	Conformer clusters	Binding energy (kcal/mol)	Number of elements/ poses	Ligand RMSD (Å)
Apo	4.1/4.6	Cluster 1	-3.00	72	15
		Cluster 2	-2.59	2	16
		Cluster 3	-2.27	20	10
		Cluster 4	-2.12	6	9.8
Gen 1	4.2/5.7	Cluster 1	-3.50	98	15
		Cluster 2	-3.31	1	15
		Cluster 3	-2.42	1	15
Gen 2	3.2/5.3	Cluster 1	-3.10	84	15
		Cluster 2	-2.71	2	15
		Cluster 3	-2.43	1	16
		Cluster 4	-2.32	13	6.9
Gen 3	2.3/6.1	Cluster 1	-2.61	12	15
		Cluster 2	-2.58	4	17
		Cluster 3	-2.43	6	16
		Cluster 4	-2.42	78	5.0
Gen 4	1.9/5.3	Cluster 1	-3.76	2	18
		Cluster 2	-3.04	69	7.4
		Cluster 3	-3.03	5	16
		Cluster 4	-2.97	2	16
		Cluster 5	-2.70	5	15
		Cluster 6	-2.25	10	8.8
		Cluster 7	-2.13	2	12
		Cluster 8	-2.09	5	4.9

8. CONCLUSIONS

In this thesis, an ENM-based iterative methodology called ClustENM is developed for the sampling of states that are entropically accessible from a starting structure and to generate atomistic conformers for supramolecules or proteins undergoing large conformational changes. The methodology consists of the application of ENM on an energetically minimized native structure and extraction of the global modes; then the deformation of the native structure along the combination of these modes and obtaining new structures; clustering these structures and selecting representative conformers from these clusters. These representatives constitute a "generation". Minimization-ENM-deformation-clustering steps are applied iteratively on the generations, producing new offsprings.

8.1. Conformational sampling using ClustENM

ClustENM is applied for conformational sampling of seven systems, expressing local to large conformational changes and consisting of 140-residues (CAM) to 11,000-residues (supramolecule 70S ribosome). The assessment of generated conformers is performed by comparison with experimental and MD data, using PCA, MSF analysis and angle-based collective variables in the case of AK. Although in p38 MAP kinase and HIV1-RT cases, there are regions where MD and experimental data do not exist, but sampled from ClustENM, still these representative conformers have comparable energy values (obtained from energy minimization in implicit solvent) to the starting structure's. So, we can speculate that these states are entropically accessible by the initial structure, but high energy barriers may exist between these states and the starting state.

Moreover, generated atomistic conformers give insight about the dynamics of 70S ribosome, for which classical MD simulations cannot be performed. The results are consistent with the previously observed ribosomal subunits motions. Therefore, the method can be successfully applied on supramolecules, for which classical MD sim-

ulations cannot be efficiently performed. In all seven systems, we obtained conformers that are consistent with the experimental and MD data.

We also applied ClustENM on 50S ribosomal subunit-TF complex. The dynamics of this complex still needs the clarification, and the generated conformers revealed that TF head domain is highly mobile when it is bound onto the ribosome, and its core and binding domains are relatively stable. This is in contrast to monomeric apo TF high flexibility in explicit solvent MD simulations.

Beside the comparison between generated conformers, experimental and MD data, we also investigated the effect of an inhibitor, bound to TIM interface, on global and catalytic loop dynamics of the enzyme. We also analyzed the changes in the specific interactions like aromatic and H-bonding due to inhibitor presence in order to clarify the inhibition mechanism. The inhibitor constrains the overall dynamics of TIM, modifies catalytic loop dynamics, enhances aromatic interactions in the tunnel region of the interface and induces changes in H-bonding network.

8.2. Docking applications

The methodology is also used in generation of conformers for docking applications (Procedure II). Four proteins displaying large hinge-bending type motions are utilized as test cases. During generation of conformers, clusters containing parent conformers are discarded in order to sample as many distinct conformers as possible. Two type of exploration techniques, namely blind and energy based search are employed and their performances are compared. Open apo conformers are used as initial structure for the generations, and the post-filtering based on apo RG is also applied in order to eliminate conformers more open than the starting structure, which turned to be successful in all four cases. Lowest RG conformer along its parent conformers are used in small ligand docking, resulting in close-to-native bound ligand poses. AK-AP5 dockings have not been performed before, as to our knowledge.

8.3. Computational efficiency

ClustENM is a computationally efficient methodology, which can be applied on various systems with different oligomerization states and conformational changes. It does not require a target conformation, and produces relatively manageable number of structures, as a result of the clustering steps.

The time required for generation of conformers depends basically on the system size and number of generations. For example, for AK fully-open state's fifth generation in Procedure I, production of 71 representative conformers (resulted from clustering of 1134 conformers, $42*3^3$ conformers) completes in 34 minutes using 3 nodes, where each node holds 2 Intel Xeon X5670 CPU's, each having 6 cores. Most of this time is spent for the minimization of the structures (32 minutes).

For the supramolecule 70S ribosome however, this time drastically increases. For the second generation, ENM and creation of 2430 conformers (from 10 conformers of first generation, $10*3^5$ conformers are produced) completes in 7 hours, clustering them completes in 3 hours using one node. Finally, the minimization of 101 representative structures is finished in 17 days using 8 nodes (approximately 4 hours for each structure). Still, the computational cost is significantly cheaper than the classical MD simulations.

8.4. Choice of parameters in ClustENM

The parameters such as number of modes to be used (m) and number of generations (g) which are determined by the user also affect the time required for the simulations. For example, one can specify m as 10 modes, leading to the number of conformers per initial structure produced in one generation being 310 (equal to 59049), which is a computationally demanding number of conformers for clustering and minimization, especially for large molecules. Therefore, in this study, the decision on m is based on the jump between the eigenvalues of slowest modes, in order to incorporate global modes as many as possible with low computational cost. "g" is another user-

defined parameter which extends the simulation time. Moreover, criteria for the energy minimization in implicit solvent can be relaxed in order to reduce the time spent for minimization.

Another issue is the choice of DF, which is also user-dependent and can differ according to system type. For example, taking into account the large conformational change of adenylate kinase, DF can be set accordingly to 2 Å and g can be set to 5. Setting DF to smaller values may result in exploring a smaller area on conformational space. On the other hand, TIM expresses relatively localized motions, therefore DF is set to 1 Å and g can be as low as 2. Using very large DF in this case, may introduce large scale distortions, which may not be optimized using minimization. Another option for determining DF for the systems that the user is unfamiliar with, can be investigating the dynamics of the system (initial structure) using ANM, in order to have an idea about the type of motion (local or large scale) that the structure may express potentially, depending on its topology.

8.5. Advantages/Limitations of ClustENM

One of the major advantages of ClustENM is that the method requires only a starting structure and is able to perform unbiased sampling. This sampling basically results in the entropically accessible states of a starting structure, since ENM is purely topological model. Another advantage is the applicability of the method for generating atomistic structures, from systems of one hundred residues to supramolecules like ribosome. Low computational cost of the method is another benefit. As for the limitations, DF and g are user-dependent and it is not always easy to guess these parameters as explained above. For example in CAM case, five generations with 3 Å DF was not enough to reach fully collapsed state.

9. FUTURE WORK

One proposition about the future application of ClustENM may be the usage of generated conformers as a starting point for various MD simulations. Explicit solvent simulations can be performed using different states of the protein as an initial structure, and energy landscape of the system can be extracted.

Application of mixed-coarse grained ENM on ligand-bound systems can be also performed by modeling the ligand in high resolution (atomistically) and the protein in low resolution (coarse-graining) to sample the conformations of such a system. TcTIM-bt10 structure can be used as a test for this purpose.

The performance of generated conformers using ClustENM can be evaluated in protein-protein or protein/nucleic acid docking using available softwares like HADDOCK.

Moreover, instead of using only few number of conformers in docking, the ligand can be docked on all of the generated conformers. The conformers used in the docking can be specified based on post-filtering criteria like RG, FRET distances or constraints obtained from NMR studies for example.

APPENDIX A: LIST OF PDB STRUCTURES USED IN ENSEMBLE ANALYSIS AND DETAILED DOCKING RESULTS FOR LAO BINDING PROTEIN

Table A.1. PDB ids of proteins used in ensemble analysis, selected based on 90% sequence similarity.

Structure	PDB Codes
AK	*4ake*, *1ake*
CAM	1cdl, 1s26, 1y6w, 2r28, 3bya, 3g43, 4dck, 2k61, 2moj(NMR, 20), *1cll*, 1sk6, 1yr5, 2v01, 3dve, 3hr4, 4gow, 2kne(NMR, 20), 2mok(NMR, 20), 1ctr, 1wrz, 1zuz, 2v02, 3dvj, 3j41, 4l79, 2l7l, 2m55(NMR, 20), 1iwq, 1xfu, 2be6, 2vay, 3dvk, 3oxq, 2jzi(NMR, 20), 2l53(NMR, 20), 2mg5(NMR, 20), 1l7z, 1xfv, 2f3y, 2w73, 3dvm, 3sui, 2k0e(NMR, 160), 2lgf, 1lvc, 1xfy, 2f3z, 2wel, 3ewt, 4bw7, 2k0f(NMR, 160), 2ll6(NMR, 20), 1pk0, 1xfz, 2lv6, 2y4v, 3ewv, 4bw8, 2k0j, 2ll7(NMR, 20)
TIM	1ag1, 1iig, 1kv5, *1tcd*, 1tsi, 2j27, 4hhp, 5tim, 6tim, 1ci1, 1iih, 1sux, 1tpf, 2j24, 3tim, 4tim

Table A.1. PDB ids of proteins used in ensemble analysis, selected based on 90% sequence similarity (continued)

Structure	PDB Codes
p38	1a9u, 1wbt, 2okr, 3fi4, 3gfe, 3kf7, 3nnx, 3s3i, 4eh3, 1bl6, 1wbv, 2onl, 3fkl, 3gi3, 3kq7, 3nww, 3s4q, 4eh4, 1bl7, 1wbw, 2oza, 3fkn, 3ha8, 3l8s, 3o8p, 3tg1, 4eh5, 1bmk, 1wfc, 2puu, 3fko, 3hec, 3l8x, 3o8t, 3u8w, 4eh6, 1di9, 1yw2, 2qd9, 3fl4, 3heg, 3lfa, 3o8u, 3uvp, 4eh7, 1ian, 1ywr, 2rg5, 3fln, 3hl7, 3lfb, 3obg, 3uvq, 4eh8, 1kv1, 1zyj, 2rg6, 3flq, 3hll, 3lfc, 3obj, 3uvr, 4eh9, 1kv2, 1zz2, 2y8o, 3fls, 3hp2, 3lfd, 3oc1, 3zs5, 4ehv, 1lew, 1zzl, 2yis, 3flw, 3hp5, 3lfe, 3ocg, 3zsg, 4ewq, 1lez, 2baj, 2yiw, 3fly, 3hrb, 3lff, 3od6, 3zsh, 4f9w, 1m7q, 2bak, 2yix, 3flz, 3hub, 3lhj, 3ody, 3zsi, 4f9y, 1ouk, 2bal, 2zaz, 3fmh, 3huc, 3mgy, 3odz, 3zya, 4fa2, 1ouy, 2baq, 2zb0, 3fmj, 3hv3, 3mh0, 3oef, 4a9y, 4geo, 1ove, 2ewa, 2zb1, 3fmk, 3hv4, 3mh1, 3p4k, 4aa0, 4ka3, 1oz1, 2fsl, 3bv2, 3fml, 3hv5, 3mh2, 3p5k, 4aa4, 4kin, *1p38*, 2fsm, 3bv3, 3fmm, 3hv6, 3mh3, 3p78, 4aa5, 4kip, 1r39, 2fso, 3bx5, 3fmn, 3hv7, 3mpa, 3p79, 4aac, 4kiq, 1r3c, 2fst, 3c5u, 3fsf, 3hvc, 3mpt, 3p7a, 4dli, 4l8m, 1w7h, 2gfs, 3ctq, 3fsk, 3iph, 3mvl, 3p7b, 4dlj, 4loo, 1w82, 2ghl, 3d7z, 3gc7, 3iw5, 3mvm, 3p7c, 4e5a, 4lop, 1w83, 2ghm, 3d83, 3gcp, 3iw6, 3mw1, 3pg3, 4e5b, 4loq, 1w84, 2gtm, 3ds6, 3gcq, 3iw7, 3new, 3qud, 4e6a, 1wbn, 2gtn, 3e92, 3gcs, 3iw8, 3nnu, 3que, 4e6c, 1wbo, 2i0h, 3e93, 3gcu, 3k3i, 3nnv, 3rin, 4e8a, 1wbs, 2npq, 3fc1, 3gcv, 3k3j, 3nnw, 3roc, 4eh2

Table A.1. PDB ids of proteins used in ensemble analysis, selected based on 90% sequence similarity (continued)

Structure	PDB Codes
HIV-1 RT	1bqm, 1hqe, 1lw0, 1rth, 1tkz, 2opr, 3dlg, 3isn, 3lp1, 1bqn, 1hqu, 1lw2, 1rti, 1tl1, 2ops, 3dlk, 3ith, 3lp2, 1c0t, 1hvu, 1lwc, *1rtj*, 1tl3, 2rf2, 3dm2, 3jsm, 3m8p, 1c0u, 1hys, 1lwe, 1s1t, 1tv6, 2rki, 3dmj, 3jyt, 3m8q, 1c1b, 1ikv, 1lwf, 1s1u, 1tvr, 2vg5, 3dok, 3kjv, 3mec, 1c1c, 1ikw, 1n5y, 1s1v, 1uwb, 2vg6, 3dol, 3kk1, 3med, 1dlo, 1ikx, 1n6q, 1s1w, 1vrt, 2vg7, 3drp, 3kk2, 3mee, 1dtq, 1iky, 1qe1, 1s1x, 1vru, 2wom, 3drr, 3kk3, 3meg, 1dtt, 1j5o, 1r0a, 1s6p, 2b5j, 2won, 3drs, 3kle, 3nbp, 1eet, 1jkh, 1rev, 1s6q, 2b6a, 2ykm, 3dya, 3klf, 3qip, 1ep4, 1jla, 1rt1, 1s9e, 2ban, 2ykn, 3e01, 3klg, 3qo9, 1fk9, 1jlb, 1rt2, 1s9g, 2be2, 2zd1, 3ffi, 3klh, 1fko, 1jlc, 1rt3, 1suq, 2hmi, 2ze2, 3hvt, 3kli, 1fkp, 1jle, 1rt4, 1sv5, 2i5j, 3bgr, 3i0r, 3lak, 1hmv, 1jlf, 1rt5, 1t03, 2iaj, 3c6t, 3i0s, 3lal, 1hni, 1jlg, 1rt6, 1t05, 2ic3, 3c6u, 3ig1, 3lam, 1hnv, 1jlq, 1rt7, 1tkt, 2jle, 3di6, 3irx, 3lan, 1hpz, 1klm, 1rtd, 1tkx, 2opp, 3dle, 3is9, 3lp0
70S Ribosome-50S	1vsa, 2wdl, 3d5d, 3ms1, 3pyt, 4kbu, 4kd2, 4kdh, 4kfl, 1vsp, 2wro, 3f1f, 3pyo, 3pyv, 4kbw, 4kd9, *4kdk*, 4l6j, 2j01, 3d5b, 3mrz, 3pyr, 4jux, 4kcz, 4kdb, 4kfi, 4l6l
70S Ribosome-30S	2ow8, 2wdk, 3d5c, 3mr8, 3pys, 4kbt, 4kd0, 4kdg, 4kfk, 2qnh, 2wrn, 3f1e, 3pyn, 3pyu, 4kbv, 4kd8, *4kdj*, 4l6k, 2j00, 3d5a, 3ms0, 3pyq, 4juw, 4kcy, 4kda, 4kfh, 4l6m

Table A.2. LAO binding protein docking results details from HADDOCK software.

	Conformer clusters	HADDOCK score	Number of elements	Ligand RMSD Å
Apo	Cluster 2	-28.6±1.8	41	4.2
	Cluster 3	-28.3±2.0	37	4.0
	Cluster 4	-28.1±2.5	31	5.0
	Cluster 6	-27.3±3.9	14	2.4
	Cluster 1	-25.8±1.5	48	4.4
	Cluster 5	-24.3±6.5	15	3.7
	Cluster 7	-18.7±4.9	9	4.9
	Cluster 8	-16.1±3.8	5	5.0
Gen 1	Cluster 4	-33.5±4.8	30	5.6
	Cluster 3	-26.0±1.1	35	4.8
	Cluster 1	-25.2±2.2	46	5.7
	Cluster 5	-24.4±2.7	22	6.1
	Cluster 2	-23.0±2.4	41	6.3
	Cluster 6	-17.6±6.0	11	5.2
	Cluster 7	-8.4±2.9	7	5.5
	Cluster 8	-8.3±7.0	5	5.1
Gen 2	Cluster 1	-33.5±6.1	46	8.3
	Cluster 5	-31.6±1.8	18	5.7
	Cluster 2	-30.3±1.5	39	5.4
	Cluster 3	-29.8±4.6	38	5.3
	Cluster 4	-29.1±1.7	28	5.2
	Cluster 8	-29.1±9.6	6	5.6
	Cluster 7	-25.9±3.4	9	4.3
	Cluster 6	-19.8±4.1	13	4.9

Table A.2. LAO binding protein docking results details from HADDOCK software (continued)

	Conformer clusters	HADDOCK score	Number of elements	Ligand RMSD Å
Gen 3	Cluster 2	-38.1±6.7	43	4.4
	Cluster 1	-33.6±4.6	50	5.4
	Cluster 3	-33.3±2.7	32	4.5
	Cluster 4	-31.8±3.2	29	4.4
	Cluster 7	-27.8±2.0	13	4.3
	Cluster 6	-27.6±4.3	14	4.0
	Cluster 5	-26.1±4.1	14	4.3
	Cluster 8	-9.5±5.5	4	4.2
Gen 4	Cluster 2	-37.2±0.9	78	3.3
	Cluster 1	-37.2±0.6	116	2.1
	Cluster 3	-21.7±6.8	5	3.4
Gen 5	Cluster 1	-46.2±0.7	200	2.2
Gen 6	Cluster 1	-48.5±1.8	122	2.0
	Cluster 2	-47.4±2.3	67	1.9
	Cluster 3	-29.0±6.4	7	5.5

REFERENCES

1. Grant, B. J., A. A. Gorfe and J. A. McCammon, "Large Conformational Changes in Proteins: Signaling and Other Functions", *Current Opinion in Structural Biology*, Vol. 20, pp. 142–147, 2010.

2. Mahajan, S. and Y.-H. Sanejouand, "On the Relationship Between Low-Frequency Normal Modes and the Large-Scale Conformational Changes of Proteins", *Archives of Biochemistry and Biophysics*, Vol. 567, pp. 59–65, 2015.

3. Bahar, I., A. R. Atilgan and B. Erman, "Direct Evaluation of Thermal Fluctuations in Proteins Using a Single Parameter Harmonic Potential", *Folding and Design*, Vol. 2, pp. 173–181, 1997.

4. Hinsen, K., A. Thomas and M. J. Field, "Analysis of Domain Motions in Large Proteins", *Proteins: Structure, Function and Bioinformatics*, Vol. 34, pp. 369–382, 1999.

5. Atilgan, A., S. R. Durell, R. L. Jernigan, M. Demirel, O. Keskin and I. Bahar, "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model", *Biophysical Journal*, Vol. 80, pp. 505–515, 2001.

6. Seyler, S. L. and O. Beckstein, "Sampling Large Conformational Transitions: Adenylate Kinase as a Testing Ground", *Molecular Simulation*, Vol. 40, pp. 855–877, 2015.

7. Gershenson, A., L. M. Gierasch, A. Pastore and S. E. Radford, "Energy Landscapes of Functional Proteins are Inherently Risky", *Nature Chemical Biology*, Vol. 10, pp. 884–891, 2014.

8. Fraunfelder, H., S. G. Sligar and P. G. Wolynes, "The Energy Landscapes and Motions of Proteins", *Science*, Vol. 254, pp. 1598–1603, 1991.

9. Onuchic, J. N., Z. Luthey-Schulten and P. G. Wolynes, "Theory of Protein folding: The Energy Landscape Perspective", *Annual Review of Physical Chemistry*, Vol. 48, pp. 545–600, 1997.

10. Leach, A. R., *Molecular Modelling Principles and Applications*, Pearson Eduction Limited, Harlow, England, 2001.

11. Sutto, L., J. Latzer, J. A. Hegler, D. U. Ferreiro and P. G. Wolynes, "Consequences of Localized Frustration for the Folding Mechanism of the IM7 protein", *Proceedings of the National Academic Science of the United States of America*, Vol. 104, pp. 19825–19830, 2007.

12. Zheng, W., N. P. Schafer and P. G. Wolynes, "Frustration in the Energy Landscapes of Multidomain Protein Misfolding", *Proceedings of the National Academic Science of the United States of America*, Vol. 110, pp. 1680–1685, 2013.

13. Wall, M. E., S. C. Gallagher and J. Trewhella, "Large Scale Shape Change in Proteins and Macromolecular Complexes", *Annual Review of Physical Chemistry*, Vol. 51, pp. 355–380, 2000.

14. Jenner, L., S. Melnikov, N. G. de Loubresse, A. Ben-Shem, M. Iskakova, A. Urzhumtsev, A. Meskauskas, J. Dinman, G. Yusupova and M. Yusupov, "Crystal Structure of the 80S Yeast Ribosome", *Current Opinion in Structural Biology*, Vol. 22, pp. 759–767, 2012.

15. Kay, L. E., "NMR Studies of Protein Structure and Dynamics", *Journal of Magnetic Resonance*, Vol. 173, pp. 193–207, 2005.

16. van Ingen, H. and A. M. J. J. Bonvin, "Information-Driven Modeling of Large Macromolecular Assemblies Using NMR Data", *Journal of Magnetic Resonance*, Vol. 241, pp. 103–114, 2014.

17. 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.

18. Lawson, C. L., M. L. Baker, C. Best, C. Bi, M. Dougherty, P. Feng, G. van Ginkel, B. Devkota, I. Lagerstedt, S. J. Ludtke, R. H. Newman, T. J. Oldfield, I. Rees, G. Sahni, R. Sala, S. Velankar, J. Warren, J. D. Westbrook, K. Henrick, G. J. Kleywegt, H. M. Berman and W. Chiu, "EMDataBank.org: Unified Data Resource for CryoEM", *Nucleic Acids Research*, Vol. 39, pp. 456–464, 2011.

19. Chadhuri, B. N., "Emerging Application of Small Angle Solution Scattering in Structural Biology", *Protein Science*, Vol. 24, pp. 267–276, 2015.

20. Freddolino, P. L., C. B. Harrison, Y. Liu and K. Schulten, "Challenges in Protein Folding Simulations: Timescale, Representation, and Analysis", *Nature Physics*, Vol. 6, pp. 751–758, 2010.

21. Shaw, D. E., R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan and B. Towles, "Millisecond-scale Molecular Dynamics Simulations on Anton", *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, SC '09, pp. 39:1–39:11, ACM, New York, NY, USA, 2009.

22. Shaw, D. E., P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan and W. Wriggers, "Atomic-Level Characterization of the Structural Dynamics of Proteins", *Science*, Vol. 330, pp. 341–346, 2010.

23. Raval, A., S. Piana, M. P. Eastwood, R. O. Dror and D. E. Shaw, "Refinement of Protein Structure Homology Models via Long, All-Atom Molecular Dynamics Simulations", *Proteins*, Vol. 80, pp. 2071–2079, 2012.

24. May, A. and M. Zacharias, "Accounting for Global Protein Deformability During Protein–Protein and Protein–Ligand Docking", *Biochimica et Biophysica Acta*, Vol. 1754, pp. 225–231, 2005.

25. Vashisth, H. and C. L. Brooks, "Conformational Sampling of Maltose-Transporter Components in Cartesian Collective Variables Is Governed by the Low-Frequency Normal Modes", *The Journal of Physical Chemistry Letters*, Vol. 3, pp. 3379–3384, 2012.

26. Schlitter, J., M. Engels and P. Kruger, "Targeted Molecular Dynamics: A New Approach for Searching Pathways of Conformational Transitions", *Journal of Molecular Graphics*, Vol. 12, pp. 84–89, 1994.

27. Hamelberg, D., J. Mongan and J. A. McCammon, "Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules", *Journal of Chemical Physics*, Vol. 120, pp. 11919–11929, 2004.

28. Bussi, G., A. Laio and M. Parrinello, "Equilibrium Free Energies from Nonequilibrium Metadynamics", *Physical Review Letters*, Vol. 96, p. 090601, 2006.

29. Hansmann, U. H. E., "Parallel Tempering Algorithm for Conformational Studies of Biological Molecules", *Chemical Physics Letters*, Vol. 281, pp. 140–150, 1997.

30. Rader, A. J., "Coarse-Grained Models: Getting More with Less", *Current Opinion in Pharmacology*, Vol. 10, pp. 753–759, 2010.

31. Marrink, S. J., H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations", *The Journal of Physical Chemistry B*, Vol. 11, pp. 7812–7824, 2007.

32. Arkhipov, A., P. L. Freddolino and K. Schulten, "Stability and Dynamics of Virus Capsids Described by Coarse-Grained Modeling", *Structure*, Vol. 14, pp. 1767–1777, 2006.

33. Arkhipov, A., Y. Yin and K. Schulten, "Membrane-Bending Mechanism of Amphiphysin N-BAR Domains", *Biophysical Journal*, Vol. 97, pp. 2727–2735, 2009.

34. Trylska, J., V. Tozzini and J. A. McCammon, "Exploring Global Motions and Corelations in the Ribosome", *Biophysical Journal*, Vol. 89, pp. 1455–1463, 2005.

35. Meagher, K. L. and H. A. Carlson, "Incorporating Protein Flexibility in Structure-Based Drug Discovery: Using HIV-1 Protease as a Test Case", *Journal of the American Chemical Society*, Vol. 126, pp. 13276–13281, 2004.

36. Zacharias, M., "Rapid Protein-Ligand Docking Using Soft Modes from Molecular Dynamics Simulations to Account for Protein Deformability: Binding of FK506 to FKBP", *Proteins: Structure, Function and Bioinformatics*, Vol. 54, pp. 759–767, 2004.

37. 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, pp. 23–42, 2010.

38. Knapp, E., S. Fischer and F. Parak, "Protein Dynamics from Mössbauer Spectra. The Temperature Dependence", *The Journal of Physical Chemistry*, Vol. 86, pp. 5042–5047, 1982.

39. Zaccai, G., "How Soft is a Protein? A Protein Dynamics Force Constant Measured by Neutron Scattering", *Science*, Vol. 288, pp. 1604–1607, 2000.

40. Roh, J. H., V. N. Novikov, R. B. Gregory, J. E. Curtis, Z. Chowdhuri and A. P. Sokolov, "Onsets of Anharmonicity in Protein Dynamics", *Physical Review Letters*, Vol. 95, p. 038101, 2005.

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

42. Cui, Q. and I. Bahar, "Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems", *C&H/CRC Mathematical & Computational Biology Series*.

43. Kolossvary, I. and W. C. Guida, "Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides", *Journal of the American Chemical Society*, Vol. 118, pp. 5011–5019, 1996.

44. Miloshevsky, G. V. and P. C. Jordan, "The Open State Gating Mechanism of Gramicidin A Requires Relative Opposed Monomer Rotation and Simultaneous Lateral Displacement", *Structure*, Vol. 14, pp. 1241–1249, 2006.

45. Miloshevsky, G. V. and P. C. Jordan, "Open-State Conformation of the KcsA K+ Channel: Monte Carlo Normal Mode Following Simulations", *Structure*, Vol. 15, pp. 1654–1662, 2007.

46. Miloshevsky, G. V., A. Hassanein and P. C. Jordan, "Antiport Mechanism for Cl−/H+ in ClC-ec1 from Normal-Mode Analysis", *Biophysiscal Journal*, Vol. 98, pp. 999–1008, 2010.

47. Batista, P. R., C. H. Robert, J.-D. Marechal, M. B. Hamida-Rebai, P. G. Pascutti, P. M. Bischa and D. Perahia, "Consensus Modes, a Robust Description of Protein Collective Motions from Multiple-Minima Normal Mode Analysis-Application to the HIV-1", *Physical Chemistry Chemical Physics*, Vol. 12, pp. 2850–2859, 2010.

48. Wang, J., Q. Shao, Z. Xu, Y. Liu, Z. Yang, B. P. Cossins, H. Jiang, K. Chen, J. Shi and W. Zhu, "Exploring Transition Pathway and Free-Energy Profile of Large-Scale Protein Conformational Change by Combining Normal Mode Analysis and Umbrella Sampling Molecular Dynamics", *The Journal of Physical Chemistry B*, Vol. 118, pp. 134–143, 2014.

49. Lopez-Blanco, J. R., J. I. Garzon and P. Chacon, "iMod: Multipurpose Normal

Mode Analysis in Internal Coordinates", *Bioinformatics*, Vol. 27, pp. 2843–2850, 2011.

50. Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, "The Weighted Histogram Analysis Method for FreeEnergy Calculations on Biomolecules. I. The method", *Journal of Computational Chemistry*, Vol. 13, pp. 1011–1021, 1992.

51. Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, "Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method", *Journal of Computational Chemistry*, Vol. 16, p. 1339−1350, 1995.

52. Tama, F., F.-X. Gadea, O. Marques and Y.-H. Sanejouand, "Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules", *Proteins*, Vol. 41, pp. 1–7, 2000.

53. Levy, R. M., D. Perahia and M. Karplus, "Molecular Dynamics of an α-Helical Polypeptide: Temperature Dependence and Deviation from Harmonic Behavior", *Proceedings of the National Academic Science of the United States of America*, Vol. 79, pp. 1346–1350, 1982.

54. Li, G. and Q. Cui, "A Coarse-Grained Normal Mode Approach for Macromolecules: An Efficient Implementation and Application to Ca(2+)-ATPase", *Biophysical Journal*, Vol. 83, p. 2457–2474, 2002.

55. Durand, P., G. Trinquier and Y.-H. Sanejouand, "A New Approach for Determining low-Frequency Normal Modes in Macromolecules", *Biopolymers*, Vol. 34, pp. 759–771, 1994.

56. Doruker, P., A. R. Atilgan and I. Bahar, "Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to a-Amylase Inhibitor", *Proteins: Structure, Function and Bioinformatics*, Vol. 40, pp. 512–524, 2000.

57. Doruker, P., R. L. Jernigan and I. Bahar, "Dynamics of Large Proteins through Hierarchical Levels of Coarse-Grained Structures", *Journal of Computational Chemistry*, Vol. 23, pp. 119–127, 2002.

58. Tirion, M. M., "Large Amplitude Elastic Motions in Proteins from a Single Parameter, Atomic Analysis", *Physical Review Letters*, Vol. 77, pp. 1905–1908, 1996.

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

60. Tama, F. and Y.-H. Sanejouand, "Conformational Change of Proteins Arising from Normal Mode Calculations", *Protein Engineering*, Vol. 14, pp. 1–6, 2001.

61. Krebs, W. G., V. Alexandrov, C. A. Wilson, N. Echols, H. Yu and M. Gerstein, "Normal Mode Analysis of Macromolecular Motions in a Database Framework: Developing Mode Concentration as a Useful Classifying Statistic", *Proteins*, Vol. 48, pp. 682–695, 2002.

62. 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, pp. 920–929, 2007.

63. Wako, H. and S. Endo, "Ligand-Induced Conformational Change of a Protein Reproduced by a Linear Combination of Displacement Vectors Obtained from Normal Mode Analysis", *Biopysical Chemistry*, Vol. 159, pp. 257–266, 2011.

64. Ahmed, A., F. Rippmann, G. Barnickel and H. Gohlke, "A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins", *Journal of Chemical Information and Modeling*, Vol. 51, pp. 1604–1622, 2011.

65. Jacobs, D., A. J. Rader, L. Kuhn and M. F. Thorpe, "Protein flexibility predictions using graph theory", *Proteins: Structure, Function and Bioinformatics*,

Vol. 44, pp. 150–165, 2001.

66. Jimenez-Roldan, L. E., R. B. Freedman, R. A. Romer and S. Awells, "Rapid Simulation of Protein Motion: Merging Flexibility, Rigidity and Normal Mode Analyses", *Physical Biology*, Vol. 9, pp. 016008–1–12, 2012.

67. Kim, M. K., R. L. Jernigan and G. S. Chirikjian, "Efficient Generation of Feasible Pathways for Protein Conformational Transitions", *Biophysical Journal*, Vol. 83, pp. 1620–1630, 2002.

68. Kim, M. K., R. L. Jernigan and G. S. Chirikjian, "Rigid-Cluster Models of Conformational Transitions in Macromolecular Machines and Assemblies", *Biophysical Journal*, Vol. 89, pp. 43–55, 2005.

69. Feng, Y., L. Yang, A. Kloczkowski and R. L. Jernigan, "The Energy Profiles of Atomic Conformational Transition Intermediates of Adenylate Kinase", *Proteins*, Vol. 77, pp. 551–558, 2009.

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

71. Zheng, W. and B. R. Brooks, "Identification of Dynamical Correlations within the Myosin Motor Domain by the Normal Mode Analysis of Elastic Network Model", *Biophysical Journal*, Vol. 88, pp. 3109–3117, 2005.

72. Zheng, W. and B. R. Brooks, "Normal Modes Based Prediction of Protein Conformational Changes Guided by Distance Constraints", *Biophysical Journal*, Vol. 88, pp. 3109–3117, 2005.

73. Zheng, W. and B. R. Brooks, "Modeling Protein Conformational Changes by Iterative Fitting of Distance Constraints Using Reoriented Normal Modes", *Biophysical Journal*, Vol. 90, pp. 4327–4336, 2006.

74. Kantarci-Carsibasi, N., T. Haliloglu and P. Doruker, "Conformational Transition Pathways Explored by Monte Carlo Simulation Integrated with Collective Modes", *Biophysical Journal*, Vol. 95, pp. 5862–5873, 2008.

75. Uyar, A., N. Kantarci-Carsibasi, T. Haliloglu and P. Doruker, "Features of Large Hinge-Bending Conformational Transitions. Prediction of Closed Structure from Open State", *Biophysical Journal*, Vol. 106, pp. 2656–2666, 2014.

76. Kirillova, S., J. Cortes, A. Stefaniu and T. Simeon, "An NMA-Guided Path Planning Approach for Computing Large-Amplitude Conformational Changes in Proteins", *Proteins*, Vol. 70, pp. 131–143, 2008.

77. Al-Bluwi, I., M. Vaisset, T. Simeon and J. Cortes, "Modeling Protein Conformational Transitions by a Combination of Coarse-Grained Normal Mode Analysis and Robotics-Inspired Methods", *BMC Structural Biology*, Vol. 13, p. S2, 2013.

78. Gur, M., J. D. Madura and I. Bahar, "Global Transitions of Proteins Explored by a Multiscale Hybrid Methodology: Application to Adenylate Kinase", *Biophysical Journal*, Vol. 105, p. 1643–1652, 2013.

79. Das, A., M. Gur, M. H. Cheng, S. Jo, I. Bahar and B. Roux, "Exploring the Conformational Transitions of Biomolecular Systems Using a Simple Two-State Anisotropic Network Mode", *Protein Science*, Vol. 24, pp. 267–276, 2014.

80. Cavasotto, C. N., J. A. Kovacs and R. A. Abagyan, "Representing Receptor Flexibility in Ligand Docking through Relevant Normal Modes Binding Site", *Journal of American Chemical Society*, Vol. 127, pp. 9632–9640, 2005.

81. Dietzen, M., E. Zotenko, A. Hildebrandt and T. Lengauer, "On the Applicability of Elastic Network Modes in Small-Molecule Docking", *Journal of Chemical Information and Modeling*, Vol. 52, pp. 844–856, 2012.

82. May, A. and M. Zacharias, "Protein-Ligand Docking Accounting for Receptor Side

Chain and Global Flexibility in Normal Modes: Evaluation on Kinase Inhibitor Cross Docking", *Journal of Medicinal Chemistry*, Vol. 51, pp. 3499–3506, 2008.

83. Leis, S. and M. Zacharias, "Efficient Inclusion of Receptor Flexibility in Grid-Based Protein–Ligand Docking", *Journal of Computationall Chemistry*, Vol. 32, pp. 3422–3439, 2011.

84. 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, pp. 1145–1152, 2007.

85. 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 Flexiblity", *Journal of Computational Chemistry*, Vol. 16, pp. 2785–2791, 2009.

86. Rueda, M., G. Bottegoni and R. Abagyan, "Consistent Improvement of Cross-Docking Results Using Binding Site Ensembles Generated with Elastic Network Normal Modes", *Journal of Chemical Information and Modeling*, Vol. 49, pp. 716–725, 2009.

87. Akten, E. D., S. Cansu and P. Doruker, "Docking Study Using Atomistic Conformers Generated via Elastic Network Model for Cyclosporin A/Cyclophilin A Complex", *Journal of Biomolecular Structure & Dynamics,*, Vol. 27, pp. 13–25, 2009.

88. de Groot, B. L., D. M. F. van Aalten, R. M. Scheek, A. Amadei, G. Vriend and H. J. C. Berendsen, "Prediction of Protein Conformational Freedom from Distance Constraints", *Proteins: Structure, Function and Bioinformatics*, Vol. 29, p. 240–251, 1997.

89. Seeliger, D. J., J. Haas and B. L. de Groot, "Geometry-Based Sampling of Conformational Transitions in Proteins", *Structure*, Vol. 15, p. 1482–1492, 2007.

90. Seeliger, D. J. and B. L. de Groot, "Conformational Transitions upon Ligand Binding: Holo-Structure Prediction from Apo Conformations", *PLoS Computational Biology*, Vol. 6, p. e1000634, 2010.

91. Trott, O. and A. J. Olson, "AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading", *Journal of Computational Chemistry*, Vol. 31, pp. 455–461, 2009.

92. Yesylevskyy, S. O., V. N. Kharkyanen and A. P. Demchenko, "The Blind Search for the Closed States of Hinge-Bending Proteins", *Proteins*, Vol. 71, pp. 831–843, 2008.

93. Flores, S. C. and M. B. Gerstein, "Predicting Protein Ligand Binding Motions with the Conformation Explorer", *BMC Bioinformatics*, Vol. 12, p. 417, 2011.

94. Atilgan, C. and A. R. Atilgan, "Perturbation-Response Scanning Reveals Ligand Entry-Exit Mechanisms of Ferric Binding Protein", *PLoS Computational Biology*, Vol. 5, p. e1000544, 2009.

95. Atilgan, C., Z. N. Gerek, S. B. Ozkan and A. R. Atilgan, "Manipulation of Conformational Change in Proteins by Single-Residue Perturbations", *Biophysical Journal*, Vol. 99, pp. 933–943, 2010.

96. Atilgan, A. R., A. O. Aykut and C. Atilgan, "Subtle pH Differences Trigger Single Residue Motions for Moderating Conformations of Calmodulin", *The Journal of Chemical Physics*, Vol. 135, p. 155102, 2011.

97. Gipson, B., M. Moll and L. E. Kavraki, "SIMS: A Hybrid Method for Rapid Conformational Analysis", *PLoS One*, Vol. 8, p. e68826, 2013.

98. Agrawal, R. K., P. Penczek, R. A. Grassucci, N. Burkhardt, K. H. Nierhaus and J. Frank, "Effect of Buffer Conditions on the Position of tRNA on the 70 S Ribosome as Visualized by Cryo-electron Microscopy", *Journal of Biological*

Chemistry, Vol. 274, p. 8723–8729, 1999.

99. Frank, J. and R. K. Agrawal, "A Ratchet-Like Inter-Subunit Reorganization of the Ribosome During Translocation", *Nature*, Vol. 406, pp. 318–322, 2000.

100. Gomez-Lorenzo, M. G., C. M. Spahn, R. K. Agrawal, R. A. Grassucci, P. Penczek, K. Chakraburtty, J. P. Ballesta, J. L. Lavandera, J. F. Garcia-Bustos and J. Frank, "Three Dimensional Cryo-electron Microscopy Localization of EF2 in the Saccharomyces Cerevisiae 80S Ribosome at 17.5 Å Resolution", *EMBO Journal*, Vol. 19, p. 2710–2718, 2000.

101. Harms, J., F. Schluenzen, R. Zarivach, A. Bashan, S. Gat, I. Agmon, H. Bartels, F. Franceschi and A. Yonath, "High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium", *Cell*, Vol. 107, pp. 679–688, 2001.

102. Tama, F., M. Valle, J. Frank and C. L. Brooks, "Dynamic Reorganization of the Functionally Active Ribosome Explored by Normal Mode Analysis and Cryo-Electron Microscopy", *Proceedings of the National Academic Science of the United States of America*, Vol. 100, p. 9319–9323, 2003.

103. Wang, Y., A. J. Rader, I. Bahar and R. L. Jernigan, "Global Ribosome Motions Revealed with Elastic Network Model", *Journal of Structural Biology*, Vol. 147, pp. 302–314, 2004.

104. Kurkcuoglu, O., P. Doruker, T. Z. Sen, A. Kloczkowski and R. L. Jernigan, "The Ribosome Structure Controls and Directs mRNA Entry, Translocation and Exit Dynamics", *Physical Biology*, Vol. 5, p. 046005, 2008.

105. Kurkcuoglu, O., Z. Kurkcuoglu, P. Doruker and R. L. Jernigan, "Collective Dynamics of the Ribosomal Tunnel Revealed by Elastic Network Modeling", *Proteins*, Vol. 75, pp. 837–845, 2009.

106. Sanbonmatsu, K. Y. and C.-S. Tung, "Large-Scale Simulations of the Ribosome:

A New Landmark in Computational Biology", *Journal of Physics*, Vol. 46, pp. 334–342, 2006.

107. Whitford, P. C. and K. Y. Sanbonmatsu, "Simulating Movement of tRNA through the Ribosome During Hybrid-State Formation", *The Journal of Chemical Physics*, Vol. 139, p. 121919, 2012.

108. Whitford, P. C., A. Ahmed, Y. Yu, S. P. Hennelly, F. Tama, C. M. T. Spahn, J. N. Onuchic and K. Y. Sanbonmatsu, "Excited States of Ribosome Translocation Revealed through Integrative Molecular Modeling", *Proceedings of the National Academic Science of the United States of America*, Vol. 108, pp. 18943–18948, 2011.

109. Seo, S., Y. Jang, P. Qian, W. K. Liu, J.-B. Choi, B. S. Lim and M. K. Kim, "Efficient Prediction of Protein Conformational Pathways Based on the Hybrid Elastic Network Model", *Journal of Molecular Graphics and Modelling*, Vol. 47, pp. 25–36, 2014.

110. Marques, O. A., "BLZPACK: Description and User's Guide", 1997.

111. Grimes, R. J., J. Lewis and H. Simon, "A Shifted Block Lanczos Algorithm for Solving Sparse Symmetric Generalized Eigenproblems", *SIAM Journal on Matrix Analysis and Applications*, Vol. 15, pp. 228–272, 1994.

112. 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, p. 1178–1187, 2009.

113. Case, D. A., T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews,

M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, "AMBER 12", , 2012, `http://ambermd.org/`.

114. Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, "A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations", *Journal of Computational Chemistry*, Vol. 24, pp. 1999–2012, 2003.

115. Hawkins, G. D., C. J. Cramer and D. G. Truhlar, "Pairwise Solute Descreening of Solute Charges from a Dielectric Medium", *Chemical Physics Letters*, Vol. 246, pp. 122–129, 1995.

116. Hawkins, G. D., C. J. Cramer and D. G. Truhlar, "Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium", *The Journal of Physical Chemistry*, Vol. 100, p. 19824–19839, 1996.

117. Srinivasan, J., M. W. Trevathan, P. Beroza and D. A. Case, "Application of a Pairwise Generalized Born Model to Proteins and Nucleic Acids: Inclusion of Salt Effects", *Theoretical Chemistry Accounts*, Vol. 101, pp. 426–434, 1999.

118. Feig, M., J. Karanicolas and C. L. Brooks, "MMTSB Tool Set", MMTSB NIH Research Resource, The Scripps Research Institute, 2001.

119. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador,

J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, "Gaussian 03, Revision D.01", Gaussian, Inc., Wallingford, CT, 2004.

120. 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.

121. 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.

122. 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.

123. Essman, U., L. Perera, M. L. Berkowitz, T. A. Darden, H. Lee and L. G. Pedersen, "A Smooth Particle Mesh Ewald Method", *Journal of Chemical Physics*, Vol. 103, p. 8577–8593, 1995.

124. Kurkcuoglu, Z., G. Ural, E. D. Akten and P. Doruker, "Blind Dockings of Benzothiazoles to Multiple Receptor Conformations of Triosephosphate Isomerase from Trypanosoma cruzi and Human", *Molecular Informatics*, Vol. 30, pp. 986–995, 2011.

125. Bakan, A., L. M. Meireles and I. Bahar, "ProDy: Protein Dynamics Inferred from Theory and Experiments", *Bioinformatics*, Vol. 27, pp. 1575–1577, 2011.

126. Vries, S. J. D., M. van Dijk and A. M. J. J. Bonvin, "The HADDOCK Web Server

for Data-Driven Biomolecular Docking", *Nature Protocols*, Vol. 5, pp. 883–897, 2010.

127. Petsko, G. A. and D. Ringe, *Protein Structure and Function*, New Science Press, London, England, 2004.

128. Hanson, J., K. Duderstadt, L. P. Watkins, S. Bhattacharyya, J. Brokaw, J.-W. Chu and H. Yang, "Illuminating the Mechanistic Roles of Enzyme Conformational Dynamics", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 104, pp. 18055–18060, 2007.

129. Henzler-Wildman, K. A., V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner and D. Kern, "Intrinsic Motions Along an Enzymatic Reaction Trajectory", *Nature*, Vol. 450, pp. 838–844, 2007.

130. Beckstein, O., E. J. Denning, J. R. Perilla and T. B. Woolf, "Zipping and Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open $\leftrightarrow$ Closed Transitions", *Journal of Molecular Biology*, Vol. 394, pp. 160–176, 2009.

131. Bhatt, D. and D. M. Zuckerman, "Heterogeneous Path Ensembles for Conformational Transitions in Semiatomistic Models of Adenylate Kinase", *Journal of Chemical Theory and Computation*, Vol. 6, p. 3527–3539, 2010.

132. Daily, M. D., G. N. Phillips and Q. Cui, "Many Local Motions Cooperate to Produce the Adenylate Kinase Conformational Transition", *Journal of Molecular Biology*, Vol. 400, p. 618–631, 2010.

133. Lou, H. and I. Cukier, "Molecular Dynamics of Apo-Adenylate Kinase: A Distance Replica Exchange Method for the Free Energy of Conformational Fluctuations", *Journal of Physical Chemistry B*, Vol. 110, p. 24121–24137, 2006.

134. Lu, Q. and J. Wang, "Single Molecule Conformational Dynamics of Adenylate

Kinase: Energy Landscape, Structural Correlations, and Transition State Ensembles", *Journal of American Chemical Society*, Vol. 130, p. 4772–4783, 2008.

135. Matsunaga, Y., H. Fujisaki, T. Terada, T. Furuta, K. Moritsugu and A. Kidera, "Minimum Free Energy Path of Ligand-Induced Transition in Adenylate Kinase", *PLoS Computational Biology*, Vol. 8, p. e1002555, 2012.

136. Jana, B., B. V. Adkar, R. Biswas and B. Bagchi, "Dynamic Coupling Between the LID and NMP Domain Motions in the Catalytic Conversion of ATP and AMP to ADP by Adenylate Kinase", *Journal of Chemical Physics*, Vol. 134, p. 035101, 2011.

137. Arora, K. and C. L. Brooks, "Large-Scale Allosteric Conformational Transitions of Adenylate Kinase Appear to Involve a Population-Shift Mechanism", *Proceedings of the National Academic Science of the United States of America*, Vol. 104, p. 18496–18501, 2007.

138. Daily, M. D., L. Makowski, G. N. Phillips and Q. Cui, "Large-Scale Motions in the Adenylate Kinase Solution Ensemble: Coarse-Grained Simulations and Comparison with Solution X-ray Scattering", *Chemical Physics*, Vol. 396, pp. 84–91, 2012.

139. Aykut, A., A. R. Atilgan and C. Atilgan, "Designing Molecular Dynamics Simulations to Shift Populations of the Conformational States of Calmodulin", *PLoS Computational Biology*, Vol. 9, p. e1003366, 2013.

140. Gsponer, J., J. Christodoulou, A. Cavalli, J. M. Bui, B. Richter, C. M. Dobson and M. Vendruscolo, "A Coupled Equilibrium Shift Mechanism in Calmodulin-Mediated Signal Transduction", *Structure*, Vol. 16, p. 736–746, 2008.

141. 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 of the United States of America*,

Vol. 106, pp. 14349–14354, 2011.

142. Kohlstaedt, L. A., J. Wang, J. M. Friedman, P. A. Rice and T. A. Steitz, "Crystal Structure at 3.5 A Resolution of HIV-1 Reverse Transcriptase Complexed with an Inhibitor", *Science*, Vol. 256, pp. 1783–1790, 1992.

143. Munro, J., K. Y. Sanbonmatsu, C. M. T. Spahn and S. C. Blanchard, "Navigating the Ribosome's Metastable Energy Landscape", *Trends in Biochemical Sciences*, Vol. 34, pp. 390–400, 2009.

144. Zhou, J., L. Lancaster, J. P. Donohue and H. F. Noller, "Crystal Structures of EF-G–Ribosome Complexes Trapped in Intermediate States of Translocation", *Science*, Vol. 340, p. 1236086, 2013.

145. Schrödinger, LLC, "The PyMOL Molecular Graphics System, Version 1.7.0.5", 2010.

146. Kramer, G., T. Rauch, W. Rist, S. Vorderwülbecke, H. Patzelt, A. Schulze-Specking, N. Ban, E. Deuerling and B. Bukau, "L23 Protein Functions as a Chaperone Docking Site on the Ribosome", *Nature*, Vol. 419, p. 171–174, 2002.

147. Ferbitz, L., T. Maier, H. Patzelt, B. Bukau, E. Deuerling and N. Ban, "Trigger Factor in Complex with the Ribosome Forms a Molecular Cradle for Nascent Proteins", *Nature*, Vol. 431, pp. 590–596, 2004.

148. Ferbitz, L., T. Maier, H. Patzelt, B. Bukau, E. Deuerling and N. Ban, "Structure of Trigger Factor Binding Domain in Biologically Homologous Complex with Eubacterial Ribosome Reveals its Chaperone Action", *Proceedings of the National Academy of Science of the United States of America*, Vol. 102, p. 12017–12022, 2005.

149. Schlünzen, F., D. N. Wilson, P. Tian, J. M. Harms, S. J. McInnes, H. A. Hansen, R. Albrecht, J. Buerger, S. M. Wilbanks and P. Fucini, "The Binding Mode of

the Trigger Factor on the Ribosome: Implications for Protein Folding and SRP Interaction", *Structure*, Vol. 13, p. 1685–1694, 2005.

150. Merz, F., D. Boehringer, C. Schaffitzel, S. Preissler, A. Hoffmann, T. Maier, A. Rutkowska, J. Lozza, N. Ban, B. Bukau and E. Deuerling, "Molecular Mechanism and Structure of Trigger Factor Bound to the Translating Ribosome", *EMBO Journal*, Vol. 27, p. 1622–1632, 2008.

151. Bukau, B., E. Deuerling, C. Pfund and E. Craig, "Getting Newly Synthesized Proteins into Shape", *Celll*, Vol. 101, pp. 119–122, 2000.

152. Hartl, F. and M. Hayer-Hartl, "Molecular Chaperones in the Cytosol: From Nascent Chain to Folded Protein", *Science*, Vol. 295, p. 1852−1858, 2002.

153. Ezeroglu, G., *A Computational Study on Bacterial Trigger Factor and Ribosome Dynamics*, M.Sc. Thesis, Bogazici University, 2014.

154. Can, M. T., *An in Silico Study on Ribosome-Binding Trigger Factor*, M.Sc. Thesis, Bogazici University, 2012.

155. Thomas, A., S. Mao and A. H. Elcock, "Flexibility of the Bacterial Chaperone Trigger Factor in Microsecond-Timescale Molecular Dynamics Simulations", *Biophysical Journal*, Vol. 105, pp. 732–744, 2013.

156. Singhal, K., J. Vreede, A. Mashaghi, S. J. Tans and P. G. Bolhuis, "Hydrophobic Collapse of Trigger Factor Monomer in Solution", *PLoS One*, Vol. 8, p. e59683, 2013.

157. Maldonado, E., M. Soriano-Garcia, A. Moreno, N. Cabrera, G. Garza-Ramos, M. T. Gomez-Puyou and R. Gomez-Puyou, "Differences in the Intersubunit Contacts in Triosephosphate Isomerase from Two Closely Related Pathogenic Trypanosomes", *Journal of Molecular Biology*, Vol. 283, p. 193–203, 1998.

158. Wierenga, R. K., E. G. Kapetaniou and R. Venkatesan, "Triosephosphate Isomerase: A Highly Evolved Biocatalyst", *Cellular and Molecular Life Sciences*, Vol. 67, pp. 3961–3982, 2010.

159. Perez-Montfort, R., M. T. 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, p. 457–470, 2002.

160. Tellez-Valencia, A., S. Avila-Rios, R. Perez-Montfort, A. Rodriguez-Romero, M. T. Gomez-Puyou, F. Lopez-Calahora and A. Gomez-Puyou, "Highly Specific Inactivation of Triosephosphate Isomerase from Trypanosoma cruzi", *Biochemical and Biophysical Research Communications*, Vol. 295, p. 958–963, 2002.

161. Olivares-Illana, V., A. Rodriguez-Romero, M. B. I. Becker, J. Garcia, R. Perez-Montfort, N. Cabrera, F. Lopez-Calahorra, M. T. Gomez-Puyou and A. Gomez-Puyou, "Perturbation of the Dimer Interface of Triosephosphate Isomerase and its Effect on Trypanosoma cruzi", *PLoS Neglected Tropical Diseases*, Vol. 1, pp. 1–8, 2007.

162. Schnackerz, K. D. and R. W. Gracy, "Probing the Catalytic Sites of Triosephosphate Isomerase by 31P-NMR with Reversibly and Irreversibly Binding Substrate Analogs", *European Journal of Biochemistry*, Vol. 199, pp. 231–238, 1991.

163. 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.

164. Cansu, S. and P. Doruker, "Dimerization Affects Collective Dynamics of Triosephosphate Isomerase", *Biochemistry*, Vol. 47, p. 1358–1368, 2008.

165. Kurkcuoglu, Z. and P. Doruker, "Substrate Effect on Catalytic Loop and Global Dynamics of Triosephosphate Isomerase", *Entropy*, Vol. 15, pp. 1085–1099, 2013.

166. Kurkcuoglu, Z., A. Bakan, D. Kocaman, I. Bahar and P. Doruker, "Coupling Between Catalytic Loop Motions and Enzyme Global Dynamics", *PLoS Computational Biology*, Vol. 8, p. e1002705, 2012.

167. Zimmermann, M. T. and R. L. Jernigan, "Protein Loop Dynamics are Complex and Depend on the Motions of the Whole Protein", *Entropy*, Vol. 14, p. 687–700, 2012.

168. Skliros, A. M., M. T. Zimmermann, D. Chakraborty, S. Saraswathi, A. R. Katebi, S. P. Leelananda, A. Kloczkowski and R. L. Jernigan, "The Importance of Slow Motions for Protein Functional Loops", *Physical Biology*, Vol. 9, p. 014001, 2012.

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

170. Tellez-Valencia, A., V. Olivares-Illiana, A. Hernandez-Santoyo, R. Perez-Montfort, M. Costas, A. Rodriguez-Romero, F. Lopez-Calahora, M. T. Gomez-Puyou and A. Gomez-Puyou, "Inactivation of Triosephosphate Isomerase from Trypanosoma cruzi by an Agent that Perturbs its Dimer Interface", *Journal of Molecular Biology*, Vol. 341, p. 1355–1365, 2004.

171. Olivares-Illana, V., R. Perez-Montfort, F. Lopez-Calahorra, M. Costas, A. Rodriguez-Romero, M. T. Gomez-Puyou and A. Gomez-Puyou, "Structural Differences in Triosephosphate Isomerase from Different Species and Discovery of a Multitrypanosomatid Inhibitor", *Biochemistry*, Vol. 45, p. 2556–2560, 2006.

172. Papadopoulou, M. V., W. D. Bloomer, H. S. Rosenzweig, M. Kaiser, E. Chatelain and J.-R. Ioset, "Novel 3-nitro-1H-1,2,4-triazole-based piperazines and 2-amino-1,3-Benzothiazoles as Antichagasic Agents", *Bioorganic & Medicinal Chemistry*, Vol. 21, p. 6600–6607, 2013.

173. Espinoza-Fonseca, L. M. and J. G. Trujillo-Ferrera, "Exploring the Possible Bind-

ing Sites at the Interface of Triosephosphate Isomerase Dimer as a Potential Target for Anti-Tripanosomal Drug Design", *Bioorganic & Medicinal Chemistry Letters*, Vol. 14, p. 3151–3154, 2004.

174. Espinoza-Fonseca, L. M. and J. G. Trujillo-Ferrera, "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, p. 922–928, 2005.

175. Burley, S. K. and G. A. Petsko, "Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization", *Science*, Vol. 229, pp. 23–28, 1985.

176. Katebi, A. R. and R. L. Jernigan, "The Critical Role of the Loops of Triosephosphate Isomerase for its Oligomerization, Dynamics, and Functionality", *Protein Science*, Vol. 23, pp. 213–228, 2014.

177. Massi, F., C. Wang and A. G. Palmer, "Solution NMR and Computer Simulation Studies of Active Site Loop Motion in Triosephosphate Isomerase", *Biochemistry*, Vol. 45, pp. 10787–10794, 2006.

178. Joseph, D., G. A. Petsko and M. Karplus, "Anatomy of a Conformational Change: Hinged Lid Motion of the Triosephosphate Isomerase Loop", *Science*, Vol. 249, pp. 1425–1428, 1990.

179. 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, p. 131–149, 2007.

180. Alakent, B., S. Baskan and P. Doruker, "Effect of Ligand Binding on the Intraminimum Dynamics of Proteins", *Journal of Computational Chemistry*, Vol. 32, pp. 483–496, 2011.

181. Halperin, I., B. Ma, H. Wolfson and R. Nussinov, "Principals of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions", *Proteins: Structure, Function and Bioinformatics*, Vol. 47, pp. 409–443, 2002.

182. Dwyer, M. A. and H. W. Hellinga, "Periplasmic Binding Proteins: A Versatile Superfamily for Protein Engineering", *Current Opinion in Structural Biology*, Vol. 14, p. 495–504, 2004.