A STUDY ON PROTEIN CONFORMATIONAL TRANSITIONS BY MONTE CARLO SIMULATIONS

by

Bahar Aydın

B.S., Ch.E, Istanbul Technical University, 2003

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

Graduate Program in Chemical Engineering

Boğaziçi University

2010

ACKNOWLEDGEMENTS

I owe my deepest gratitude to my thesis advisors, Prof. Pemra Doruker Turgut and Prof. Türkan Haliloğlu who believed in me throughout my thesis even though I work in a company during my thesis studies. In this long journey, they supported me with their encouraging, understanding and patient attitude and guidance from the initial to the final level. Their helpful comments, recommendations and knowledge enabled me to develop an understanding of the subject.

It is a pleasure to thank the thesis committee members: Prof. Viktorya Aviyente, Assist. Prof. Demet Akten Akdoğan, Assist. Prof. Elif Özkırımlı Ölmez, for the valuable time they have devoted reading and commenting on my thesis.

Special thanks to Arzu Uyar, who supported and encouraged me in any respect during the completion of this thesis. This thesis would not have been possible without her effort and cooperation especially with computer programs. She was both a colleague and a good friend with her positive approach in a number of ways.

I am grateful to my manager Reyhan Kasapoğlu for her support during my thesis studies. One can simply could not wish for a better and motivating manager.

Special thanks are due to my family for their endless love, patience and understanding I felt all of my life. I wish to express my deepest gratitude to all of them especially to my mother Alev Aydın for encouraging me to attend this master programme, grandmother Güzin Demirok for taking care of me since my childhood, my sister Yeşim Aydın for her smiling face and sense of humor and finally my father Cemil Aydın for his support. As always, they were right there whenever I needed. This thesis is indeed dedicated to them…

ABSTRACT

A STUDY ON PROTEIN CONFORMATIONAL TRANSITIONS BY MONTE CARLO SIMULATIONS

Exploration of conformational transitions between two different conformational states of proteins reveals significant information about their action mechanism and function. In this thesis, the recently developed anisotropic network model–Monte Carlo (ANM-MC) and targeted Monte Carlo (TMC) simulations are applied between two experimentally determined distinct conformations (initial/target) of 12 proteins with different sizes, number of chains, domains and motion types. The objective is to obtain a database of conformational transitions in forward/reverse directions of apo (free) to complex transitions. In ANM-MC, collective modes obtained from ANM are combined with the MC simulation approach. In TMC, the initial structure is forced towards the target structure without using collective modes. The simulation parameters were modified to perform better energy minimization and an automated version of the algorithm was developed to reduce the simulation time. The root mean-square deviation (RMSD) values between the initial and target states of the proteins studied fall between 4.1 and 15.6 Å. As a result of TMC, all proteins approach the target within an RMSD of 0.4 Å. On the other hand, the RMSD values between the predicted final structure in the forward ANM-MC and the target structure vary between 1.4 and 3.9 Å, whereas it varies between 1.8 and 4.7 Å in the reverse ANM-MC (excluding the unsuccessful case of diphtheria toxin). High initial overlap values between the selected modes and the target direction derive the conformations more to the target state. In general, the proteins with a hinge bending motion between two domains exhibit the most successful results and high overlap values. The most selected ANM modes in the initial stages of the simulation are the slowest five modes. Contact maps of the corresponding intermediates (snapshots) and the final structure are also analyzed. In all cases, the final structure is very similar to the target structure in terms of newly formed contacts and overall three-dimensional conformation.

ÖZET

MONTE CARLO SİMÜLASYONU İLE PROTEİNLERDE KONFORMASYONEL GEÇİŞLERİN İNCELENMESİ

Proteinlerde iki farklı yapı arasındaki konformasyonel geçişin incelenmesi, proteinlerin faaliyet mekanizmaları ve fonksiyonları hakkında önemli bilgiler ortaya çıkarmaktadır. Bu tezde, yeni geliştirilmiş Anisotropik ağ modeli-Monte Carlo (ANM-MC) ve Hedeflendirilmiş Monte Carlo (TMC) simülasyonları; farklı ebat, zincir, bölge sayısı ve hareket tipine sahip 12 proteinin deneysel olarak belirlenmiş iki farklı konformasyonuna (başlangıç/hedef) uygulanmıştır. Amaç; açık/serbest ve kompleks arasındaki düz ve ters yönlerdeki konformasyonel geçiş datasının elde edilmesidir. ANM-MC metodunda ANM'den elde edilen kolektif modlar MC simülasyon yaklaşımıyla birleştirilir. TMC'de kolektif modlar kullanılmadan ilk yapı hedef yapıya doğru itilir. Simülasyon parametreleri, ara yapılarda daha iyi enerji minimizasyonu sağlamak için değiştirildi ve simülasyon süresini kısaltmak için algoritmanın otomatik bir versiyonu geliştirildi. Çalışılan proteinlerin başlangıç ve hedef yapıları arasındaki kök ortalama kare sapması; 4.1 Å ve 15.6 Å arasında yer almaktadır. TMC simülasyonları sonucunda, tüm proteinler hedef yapıya hem düz hem de ters yönde 0.4 Å RMSD kadar yaklaşmıştır. Diğer yandan, düz ANM-MC sonucunda elde edilen son yapı, hedef yapıya 1.4 Å ve 3.9 Å arasında değişen RMSD ile yaklaşırken ters ANM-MC'de 2.3 Å ve 3.7 Å arasında yaklaşmıştır (problemli difteri toksin dışında). Seçilen modlar ve hedef yönler arasındaki yüksek overlap değeri, konformasyonları hedefe daha çok yaklaştırır. Genel olarak, iki domain arasında menteşe bükülme (hinge-bending) hareketine sahip proteinler yüksek overlap değerleri ile en başarılı sonuçları vermektedir. Simülasyonun ilk evrelerinde en çok seçilen ANM modları, en yavaş beş moddur. Son yapı ve ara yapılara ait kontak haritaları da analiz edilmektedir. Tüm simülasyonlarda elde edilen son yapı, üç boyutlu konformasyon ve oluşan yeni bağlantılar açısından son hedef yapıya benzerlik göstermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

Å	Angstrom
$\mathbf{R}$	Coordinate matrix of the initial structure
$\mathbf{R_{new}}$	Coordinate matrix of the target structure
$\mathbf{B_{rot}}$	Rotated (superimposed) coordinates of target structure
C^{α}	Alpha-carbon atom
$\mathbf{\Lambda}$	Diagonal matrix
$\mathbf{H}$	Hessian matrix
$\mathbf{H}_{ij}$	Super element matrix of the hessian matrix
$H(r_c - R_{ij})$	Heavyside step function
$\mathbf{I}$	Identity matrix
k	Damping factor
l_i	Bond length connecting atoms i-1 and I
l_i^S	Bond length connecting backbone and sidechain
m	Number of eigenvectors and corresponding conformations
n	Number of residues
N	Total number of residues
$\boldsymbol{Q}$	Target vector
r	random number
r_c	Cutoff radius
$\mathbf{R}_i$	Position vector of site I
$\mathbf{R_{new}}$	New coordinate matrix of the generated conformation
R_{ij}	Distance between site i and j
$R_{required}$	Desired approach value
$\Delta\mathbf{R}_i$	Fluctuation of vector of site i
Δx	Strength of perturbation
T	Absolute temperature
t	Transpose
$\mathbf{U}$	Orthogonal matrix

V	Potential energy
$\Delta\mathbf{x}_i$	Change in the x coordinate of the position vector of site i
$\Delta\mathbf{y}_i$	Change in the y coordinate of the position vector of site i
$\Delta\mathbf{z}_i$	Change in the z coordinate of the position vector of site i
α	Alpha helix
β	Beta strand
γ	Force constant
θ_i	bond angle between bond l_i and l_{i+1}
θ_i^S	bond angle between l_i and l_i^S, and ϕ_i^S
ϕ_i	torsional rotation of the bond l_i
ϕ_i^S	torsion angle defined by l_{i-1}, l_i and l_i^S
ϕ_i^-	torsional angle of backbone bond preceding i^{th} alpha carbon
ϕ_i^+	torsional angle of backbone bond succeeding i^{th} alpha carbon
Φ	New conformation
Φ_o	Original conformation
3D	Three-dimensional
ACM	Amplified Collective Motion
AK	Adenylate Kinase
ADP	Adenosine diphosphate
aMD	Accelerated Molecular Dynamics
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
ANM	Anisotropic Network Model
AP5	Adenosine-5'-)pentaphosphate
ApUp	Adenylyl 3'-5' uridine 3' monophosphate
BB	Backbone-backbone
BLZPACK	Block Lanczos Package
BME	Beta-mercaptoethanol
C	Catalytic

CaM	Calmodulin
CDK5	Cyclin-dependent Kinase 5
cMD	Conventional Molecular Dynamics
CO_3	Carbonate
com	Center of mass
Cryo-EM	Cryo-electron microscopy
DF	Deformation factor
DNA	Deoxyribonucleic acid
dSDNA	double stranded deoxyribonucleic acid
DT	Diphtheria Toxin
E	Energy
E.coli	Escherichia coli
ENM	Elastic Network Model
ENI	Elastic Network Interpolation
EOH	Ethanol
FRET	Fluorescence Resonance Energy Transfer
GBP	Glutamine Binding Protein
GDP	Guanosine-5'-diphosphate
GK	Glucokinase
Gly-Leu	Glycyl-L-leucine
GNM	Gaussian Network Model
GSP	5'-guanosine-diphosphate-monothiophosphate
HSTR	Human Serum Transferrin
LAO	Lysine/ Arginine/ Ornithine
L	Ligand
LR	Long-range
MENM	Mixed Elastic Network Model
MD	Molecular Dynamics
MC	Monte Carlo
MCS	Monte Carlo step
NHE	Ethane Sulfonic Acid
NH4	Ammonium

NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
NTP	Nucleotide Triphosphate
PDB	Protein Data Bank
PNM	Plastic Network Model
R	Receptor-binding
R2	Relaxed form
RG	Radius of gyration
RIP	Ribose
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
SB	Sidechain-backbone
SC	Side-chain
SR	Short-range
SS	Sidechain-sidechain
T	Transmembrane
TMC	Targeted Monte Carlo
TMD	Targeted Molecular Dynamics
V	Total potential energy

1. INTRODUCTION

Proteins perform biochemical functions such as catalysis, transport, contraction, protection, structure and metabolic regulation (Bhagavan, 2002). While proteins carry out their biological functions, they undergo conformational changes upon ligand binding (Zheng *et al.*, 2007). Thus, investigation of conformational transition pathways is important for understanding the structure-function-dynamics-relationship (Kim *et al.*, 2002a; Lei *et al.*, 2004).

Experimental techniques such as x-ray crystallography and nuclear magnetic resonance (NMR) are commonly used to determine detailed atomic structures of proteins. While experimental approaches reveal considerable information about specific protein conformations, they cannot detect less populated intermediate conformations or monitor protein conformational changes at atomic resolution as a function of time. Hence, computational techniques such as molecular dynamics (MD) simulations, Monte Carlo (MC) sampling methods, normal mode analysis (NMA) are used to investigate protein conformational transitions (McCammon and Harvey, 1987; Kitao and Go, 1999; Lei *et al.*, 2004; Bahar and Rader, 2005; Ma, 2005).

MD and NMA, which are commonly used to simulate protein dynamics by using all-atom empirical potentials (Brooks and Karplus, 1983), have time scale limitations to explore protein conformational changes in the range of microsecond-milliseconds to seconds. Hence, coarse-grained methods such as elastic network models (ENM) (Bahar and Rader, 2005; Ma, 2005) and MC simulations (Haliloglu and Bahar, 1998; Haliloglu, 1999; Kurt and Haliloglu, 1999) are efficient techniques for protein conformational analysis. Based on the recent studies, ENM methods have been proved to be successful to analyze protein conformational transitions (Mouawad and Perahia, 1996; Tama and Sanejouand, 2001; Delarue and Sanejouand, 2002; Tama and Brooks, 2002, 2006; Zheng and Brooks, 2005a, 2005b, 2006; Kim *et al.*, 2002a, 2005; Xu *et al.*, 2003; Mouawad *et al.*, 2002; Maragakis and Karplus, 2005; Krillova *et al.*, 2008; Kantarci-Carsibasi *et al.*, 2008). Most ENM based studies use x-ray structures of protein initial and target states to explore the protein conformational transitions.

Gaussian network model (GNM) (Bahar *et al.,* 1997a, 1998, Haliloglu *et al.,* 1997; Bahar and Jernigan, 1998; Bahar, 1999) and anisotropic network model (ANM) (Atilgan *et al.,* 2001) are the most widely used coarse-grained ENM methods. In GNM method, it is assumed that the residues undergo Gaussian-distributed isotropic fluctuations with no direction selection whereas anisotropic fluctuations are assumed to exist in ANM (Atilgan *et al.,* 2001). Latest computational studies proved that both methods can be used to describe low-frequency protein motion by generating collective modes and predicting atomic fluctuations for large biological systems based on harmonic potentials between close-neighboring residues in the protein structure (Kurkcuoglu *et al.,* 2004; Yildirim and Doruker, 2004). Based on recent studies (Doruker *et al.,* 2002), the computational time is reduced by hierarchical levels of coarse-graining and thus providing information about the structure-function relationship of very large biological systems such as GroEL-GroES complex (Keskin *et al.,* 2002), ribosome complex (Wang *et al.,* 2004; Kurkcuoglu *et al.,* 2008, 2009), RNA polymerase (Yildirim and Doruker, 2004).

ENM applications generate harmonic modes to illustrate protein collective motions, however they cannot provide information about the sequence of events and intermediate structures, hence only ENM is insufficient in exploring transitional pathways. Kantarci-Carsibasi *et al.* (2008) developed a new method called ANM-MC, in which the collective modes obtained from ANM are incorporated into an MC simulation method to explore conformational transition pathways and pathway intermediates between given initial and final states. In other words, NMA is continuously repeated through the simulation updating the normal modes of the present state followed by energy minimization by MC to produce a feasible pathway between two conformations. Firstly, the normal modes of the initial structure are calculated by ANM. Secondly, the slow mode overlapping the desired transition is selected among the overlap values of each mode with the target direction obtained by the superimposition of initial and target structures. The initial conformation is deformed along the selected mode to generate the next structure. Thirdly, the energy minimization of the new structure is performed by MC and the normal modes are recalculated. The method that was applied to adenylate kinase (AK) and hemoglobin and has proven to be successful in investigating the transition between open-closed conformations and sequence of events in transitional pathways (Kantarci-Carsibasi, 2009). RG-ANM-MC method, which is an extension of ANM-MC with an additional constraint

introduced on protein's radius of gyration (RG), was successfully applied to adenylate kinase (AK), human serum transferrin (HSTR) and LAO binding protein (Kantarci-Carsibasi, 2009). In RG-ANM-MC, only the initial state is provided as input and the target state is predicted by applying RG constraint. In both methods, normal modes are continuously generated and updated by iterative methodology during the computational simulation. The third methodology studied by Kantarci-Carsibasi *et al.* (2008) and Kantarci-Carsibasi (2009) was TMC (targeted MC) method, which is based on a forcing algorithm towards the target state without using collective modes. In TMC method, successive MC simulations are used without ANM calculations and the initial structure is forced towards the target structure along the desired direction

The objective of this study is to apply TMC and ANM-MC methods to relatively larger monomeric/dimeric protein structures with high root mean square distance (RMSD) values (> 4 Å) between their open/closed states. First, a database of protein structures is generated for this purpose to explore the applicability/limitations of the methods. The computational algorithm was revised to be applicable for larger proteins in a shorter period of time. Reverse transitions were also analyzed in the simulations beginning from the target structure towards the initial structure.

The plan of this thesis is as follows: an introduction to protein structure-dynamics-function relationship and conformational transitions together with information on the proteins to be studied will be presented in the following chapter. Details of ANM-MC and TMC techniques are given in the third chapter. Results and discussion of ANM-MC and TMC can be found in chapter four. Finally, the conclusion and recommendations will be given.

2. PROTEINS, DYNAMICS AND CONFORMATIONAL TRANSITIONS

2.1. Protein Structure and Function

Proteins are linear biopolymers built from sequences of amino acids. Each protein described by a specific sequence folds to a unique three-dimensional (3D) structure, called its native state. There are twenty common amino acids varying in size, hydrogen-bonding potential and charge. They are designed in several structures, ranging from the primary structure, over secondary structure, to tertiary and higher orders of structure due to the protein function. The primary structure represents the amino acid sequence. The secondary structure is maintained by the folding of polypeptide chains into structures like helices and beta sheets by repetitive hydrogen bonding. The three-dimensional tertiary structure is formed by ionic bonds, hydrogen bonds, -S-S- bridges, van der Waals forces and hydrophobic interactions. Proteins with two or more identical or different polypeptide chains called subunits have quaternary structures (Figure 2.1).

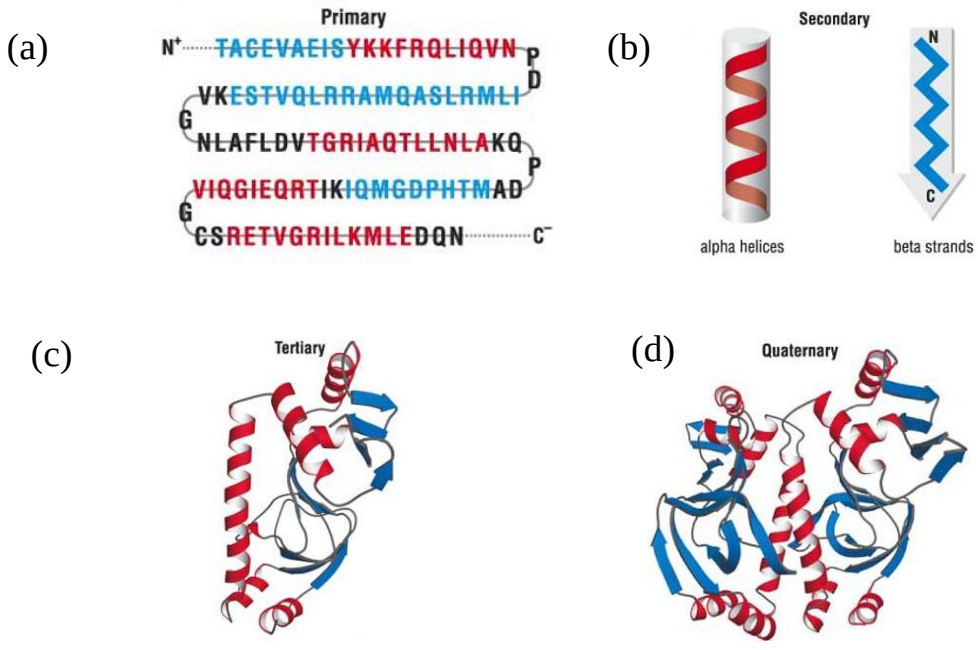


Figure 2.1. (a) Primary, (b) secondary, (c) tertiary, and (d) quaternary structures of proteins (Petsko and Ringe, 2004)

Each protein contain a backbone which is the main chain of the repeating units-the peptides with a side chain attached to each backbone. Protein backbone structure is as –C-C_α-N-. A side chain (SC) is bonded to the α-carbon and forms the specific residue for the amino acids. Amino acids are linked by covalent bonds to form linear polypeptide chains (Figure 2.2). The protein is surrounded by water molecules which are essential for the protein motion and the function.

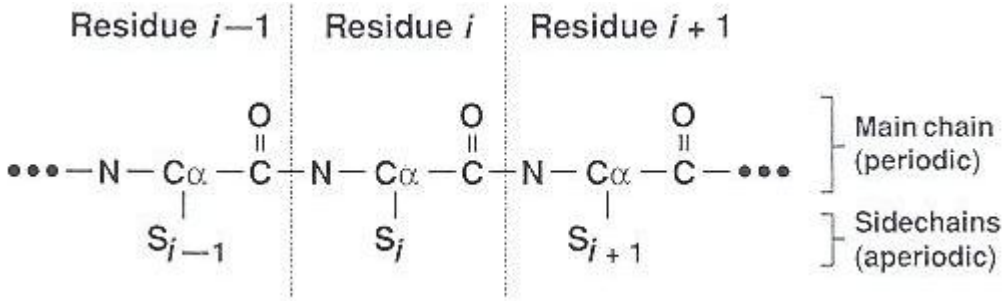


Figure 2.2. Formation of the polypeptide chain (Lesk, 2001)

Protein structure and dynamics determine the function of the protein. Proteins perform their function due to the side chains which have different sizes, shapes, hydrogen bonding capabilities and charge distributions (Leach, 2001). Proteins have several functions such as motion and locomotion of cells and organisms; the catalysis of biological reactions; and the transport of materials in body fluids. Moreover, the proteins contribute to the structure of cells, and the extracellular matrix they are embedded in. Besides, they also function as receptors for hormones and other signaling molecules and as the transcription factors that turn genes on and off to guide the differentiation of the cell (Stryer, 1988).

Protein perform the activities encoded by genes, which store the details of a protein sequence. During transcription and translation, a cell reads and uses the genetic information for protein construction. The resulting protein is then chemically changed (post-translational modification) before becoming functional. Proteins commonly contact with one another and form a complex to achieve a particular function (Stryer, 1988).

The most widely used experimental methods to determine protein 3D structures are x-ray crystallography and nuclear magnetic resonance (NMR). Structures of large molecules can be solved by x-ray crystallography since there is a size limitation in NMR. X-ray crystallography was used to study nearly 56,000 protein structures available in RCSB Protein Data Bank (PDB) whereas only about 8000 structures were solved by NMR. Also cryo-EM was used to study approximately 290 protein structures (RCSB Protein Data Bank).

Electron microscopy (cryo-EM), fluorescence resonance energy transfer (FRET), and single particle tracking methods are other experimental techniques for investigating the protein conformational dynamics and protein interactions at low resolution. These tools give information about structural transitions and molecular mechanisms without atomic details. (Frank, 1996; Kasprzak *et al.*, 1988; Lakowicz, 1999; Lesk, 2001).

2.2. Protein Dynamics and Conformational Transitions

Protein dynamics and conformational changes play an important role in all biological processes, protein function and mechanism. Proteins undergo collective domain motions such as shear and hinge-bending motion depending on the biological activity and function (Gerstein *et al.*, 1994). Hinge-bending is the most commonly observed collective motion of proteins and takes place around one or multiple centers. Shear and hinge-bending motions are illustrated in Figure 2.3.

Hinge motion is usually observed in proteins with two domains linked by a hinge. Two parts of the structure moves as a rigid body towards each other with the axis of the rotation passing through the hinges. The main chain torsional angles at the hinge change during the hinge motion (Janin and Wodak, 1983; Perutz, 1989; Gerstein *et al.*, 1994; Gerstein and Krebs, 1998).

Shear-motion is a sliding motion, in which side chain torsional angles slightly change with no significant deformation of main chain torsional configuration. Proteins that undergo shear motion have a layered structure since a number of shear motions is

necessary to generate a large motion rather than an individual shear motion (Janin and Wodak, 1983; Perutz, 1989; Gerstein *et al.*, 1994; Gerstein and Krebs, 1998).

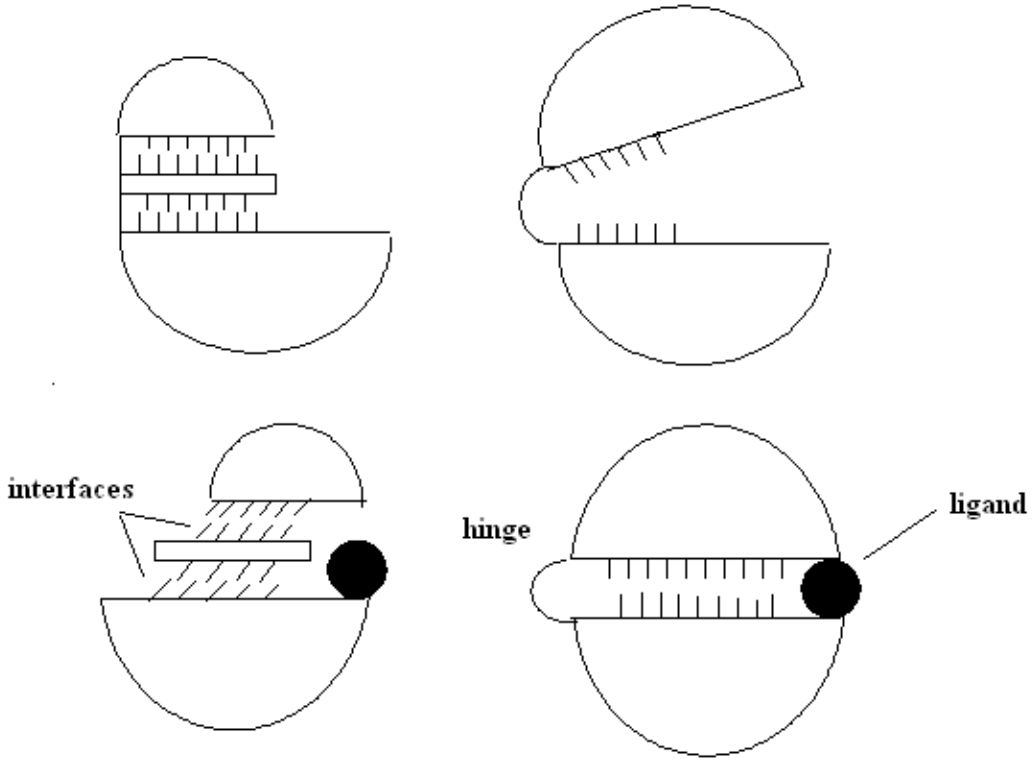


Figure 2.3. Mechanism of shear and hinge motions in proteins (Gerstein and Krebs, 1998)

While proteins perform their function, they undergo conformational transitions upon binding to ligands. One polypeptide may have many folded conformations with specific biological activity and these determine the protein function (Brooks et al., 1988). Proteins tend to have a unique apo/open conformation, i.e. the native state that corresponds to the global minimum of the free energy.

In as much as proteins are flexible molecules, they commonly change conformations in case of ligand binding. Specifically, flexible loop regions play an important role in enzymes. Many enzymes such as lysozyme change conformation to enclose the substrate in the right position for proper function. Immunoglobulins are highly flexible proteins, that can bind to a large range of ligands. G-proteins undergo a GTP-mediated conformational change with the dissociation of the α-domain from the rest of the

protein triggered by the binding of a hormone to its receptor. Allosteric proteins have special conformational transitions, in which a conformational change is triggered by ligand binding to one subunit altering the ligand affinity of the other subunits, thus manipulating the response of these proteins (Stryer, 1988). Antibody-antigen groups also exhibit important conformational changes with a different binding site topology, in which the receptor binds multiple ligands at the same region of the binding site (Keskin, 2007).

Protein-ligand interactions such as substrate binding to enzymes and antibody-antigen interactions are controlled by global motions of proteins (Amadei *et al.*, 1993; Hayward *et al.*, 1997; Tama and Sanejound, 2001; Keskin *et al.*, 2002; Tobi and Bahar, 2005; Keskin, 2007). Protein binding mechanisms are explained by several different models such as the lock and key model, the induced-fit model, and the pre-existing equilibrium model as shown in Figure 2.4. The first model proposed for the enzyme-substrate interaction was the lock and key model. In this analogy, only the key (substrate) with the correctly size fits into the key hole (active site) of the lock (enzyme). Experimental studies revealed another picture, the induced-fit model, where the enzyme is assumed to be flexible by adjusting its shape according to the substrate (Koshland, 1958). Pre-existing equilibrium is based on protein folding theories of the funnel energy landscape (Tsai *et al.*, 1999; Ma *et al.*, 2002; Li *et al.*, 2004; Frauenfelder, 1991; Karplus, 1997). In this mechanism, the protein possesses several conformations at its binding site. The ligand selects an active conformation to bind, moving the equilibrium towards the active conformation. The protein have multiple active-sites and functions due to possessing multiple structures (Goh *et al.*, 2004; Keskin, 2007).

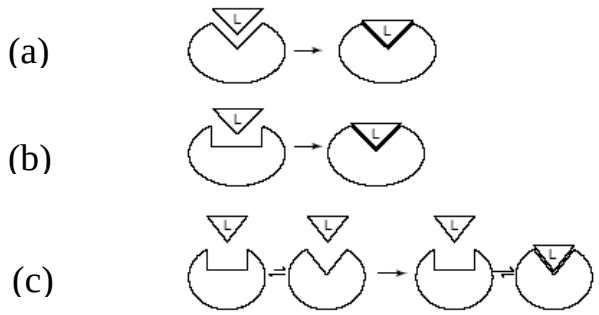


Figure 2.4. Models for protein binding mechanisms (a) Lock and key model, (b) Induced-fit model, (c) Pre-existing equilibrium model. L: Ligand (Goh *et al.*, 2004)

For allosteric proteins, ligand binding can affect the conformation of the protein at a distant region away from the binding site (Goh *et al.,* 2004). Kantarci-Carsibasi *et al.* (2008) studied allosteric conformational transitions of the enzyme adenylate kinase (AK) and hemoglobin. AK has two binding sites for binding both adenosine triphosphate (ATP) and adenosine monophosphate (AMP). Based on its allosteric behavior, the closure of the ATP-binding domain triggers the closure of the AMP-bind domain. However, the domain closure sequence in AK is still questionable and it is still studied whether the ATP is bound prior to AMP or not.

2.3. Molecular Modeling and Simulation Techniques

Various computational techniques are used to investigate molecular motions and protein dynamics. Molecular dynamics (MD) is one of the most commonly used simulation techniques. MD simulations generate the coordinates of each individual atom over time. The initial configuration is calculated from the experimentally determined structure, and put in its natural environment. Newton's equation of motion is then applied to this configuration (Leach, 2001). The sampling degree of conformational space of the protein is very important in MD simulations to generate the animation of output atom coordinates. In spite of the improvements in the simulation times of MD, still it is not sufficient to observe slow conformational changes and protein folding/unfolding in many important protein processes (Daggett, 2000). Therefore, more efficient techniques need to be developed.

NMA, especially coarse-grained approaches such as elastic network models (ENM) have been studied recently for large protein systems (Lei *et al.,* 2004; Bahar and Rader, 2005; Ma, 2005). Elastic network models are used to analyze transition pathways and intermediate structures between two molecular structures. The method uses low-frequency normal modes to generate the conformational changes. Collective coordinates are used to select functionally relevant modes of motion during the simulation (Kitao and Go, 1999). Normal Mode Analysis (NMA) approach is the most widely used atomistic technique which assumes that most positional protein fluctuation is observed along collective degrees of freedom (Levitt *et al.,* 1983; Go *et al.,* 1983; Brooks and Karplus, 1983). Proteins undergo slow to fast vibrations due to the vibration of atoms around covalent bonds. The vibrations at high frequencies represent fast motions (fast modes) whereas the vibrations at

lower frequencies, such as large domain motions stand for slow motions (slow modes). Active residues in the fastest modes have an important role in sustaining the structure since they offer a strong resistance to conformational changes. On the other hand, active residues in the slowest modes are easily affected by large scale (global) motions, thus slow motions determine the biological function (Bahar *et al.*, 1998). In NMA, the Hessian matrix is diagonalized by calculating the second derivative of the potential energy to obtain the collective variables in a local energy minimum. Tama and Sanejouand (2001) showed that NMA generates low frequency collective motions, which has a correlation with the experimental data during the conformational change.

Monte Carlo (MC) simulation technique is another commonly used tool to search for the conformational space, particularly in large proteins. It makes use of random numbers and probability statistics to sample conformational space and includes two steps; (1) generation of the new conformation and (2) decision of whether the new conformation will be accepted or not. In the first step, the new conformation is generated starting from an initial conformation by constructing moves such as multiple torsion moves and cartesian coordinate moves due to random numbers. In the second step, decision is made based on Metropolis criterion in which the low energy conformation is selected. If the energy of the new conformation is lower than the current conformation, then the new conformation is accepted as the current conformation with the relevant move. However; it is still possible for the new conformation, which has a higher energy than the current energy, to be accepted according to the Boltzmann factor. If the new conformation is rejected, the other new conformations will be generated until one of them is accepted. Coarse-grained MC models have shown that the native state protein dynamics can efficiently be simulated at a rate of one order of magnitude faster than atomistic models with multiple independent trajectories at long time scales (Haliloglu and Bahar, 1997; Bahar *et al.*, 1997c).

Experimental methods and atomistic simulations such as MD and NMA using all-atom empirical potentials are not efficient enough to analyze the conformational transitions, especially in large systems (Brooks and Karplus, 1983; Noe *et al.*, 2003).

Since MD simulations are not effective for exploring the conformational transitions in the time scale of microsecond/milliseconds to seconds, targeted MD simulations (TMD),

which samples the conformational space in a predefined direction (Schlitter *et al.,* 1993; Kong *et al.,* 2006, Van der Vaart and Karplus 2005, 2007) are used to simulate large conformational transitions. In TMD, MD simulation is applied successively to the initial state and RMSD from the final state is used to direct the protein (Schlitter *et al.,* 1993). TMD studies gave successful results in finding the transition pathways, whereas it fails to generate reversible pathways following the lowest energy pathway (Van der Vaart and Karplus, 2005). Zhang and coworkers applied (2006) conventional MD and TMD to glucokinase (GK) and its mutants in order to analyze the conformational transition from the closed state to the open state. Authors showed that the activator improves the enzyme activity through the conformational constraints even though the active site is far from the allosteric site by a distance of 20 Å. The conformational transition pathway was generated and the intermediate states were identified with the effect of kinetic analyses and mutagenesis as a result. This study is beneficial to understand the regulation mechanism of GK and to design drug for metabolic diseases.

Accelerated MD (aMD) simulations are used to observe transitions of systems having large energetic barrier between end states since it cannot be achieved by conventional MD (cMD) simulation. In this method, the accessible time scale of cMD simulations is extended due to the change in the underlying potential energy surface of the system. The simulation is accelerated by the addition of a non-negative boost potential, which increases the energy within basins (Hamelberg *et al.*, 2004). As a result, a trajectory generated on this modified surface accelerates the transitions from state to state. The simulation was applied to RAS protein to understand the GTP- and GDP-bound conformational states. The study revealed results about the conformational change and the intermediate structures, thus describing the RAS function and the signaling pathway (Grant *et al*., 2009).

Low resolution coarse-grained approaches such as ENM (Mouawad and Perahia, 1996; Tama and Sanejouand, 2001; Delarue and Sanejouand 2002; Tama and Brooks, 2002, 2006; Zheng and Brooks, 2005a, 2005b, 2006; Kim *et al.* 2002a, 2005; Xu *et al.*, 2003; Mouawad *et al.,* 2002; Maragakis and Karplus, 2005; Krillova *et al.,* 2008) and MC simulation techniques are efficient tools to analyze the conformational transition of large proteins and their complexes as alternative to atomistic models. Krebs *et al.* (2002)

reported an NMA based computational study on a set of different proteins with different binding mechanisms. According to the results, conformational transition between the open and closed states of proteins follow the lowest frequency normal modes.

Mouawad and Perahia (1996) applied single NMA on the T structure of hemoglobin along both positive and negative directions. The authors proposed four intermediate structures between T-R by the continuous deformation of T structure along first 3 slowest modes sequentially followed by energy minimization. The final RMSD value to R state was 1.82 Å. Xu and coworkers (2003) studied the transition from T to relaxed form (R2) in hemoglobin and demonstrated that this transition is aligned with the slowest mode by reporting an approach of 2.4 Å to R2 state.

Zheng and Brooks (2005a; 2005b; 2006) developed an ENM based approach that iteratively minimizes the error of fitting the given distance constraints and the energy cost by using the crystal structure of the initial state and distance constraints for the final state. The transitions in a set of 16 protein structures were generated successfully in this study. Recently, a mixed elastic network model (MENM) was also developed by Zheng and coworkers (2007) to analyze the large scale conformational changes of KIF1A kinasin and myosin II. The model combines the elastic network potentials of the initial and final states by adding their respective partition functions. The initial and final structures are retained as local minima on the MENM surface and connected to generate the MENM energy function and the transition paths.

Maragakis and Karplus (2005) developed the plastic network model (PNM) to analyze the transitions via a minimum energy path between two end structures. The transitional pathway from open to closed conformations of adenylate kinase (AK) was studied by this methodology and intermediate crystal structures in the pathway were generated.

The elastic network interpolation (ENI) algorithm was developed by Kim and coworkers (2002b) to generate conformational transition intermediates between two states of lactoferrin. The ENI methodology is a coarse-grained ENM in which it is assumed that the optimal intermediates can be generated by uniformly interpolating the distances in two

different conformations (Kim *et al.,* 2002b; 2005). By using the ENI methodology, Feng *et al.* (2009) recently studied the conformational transitions of AK. In this study, the authors performed CHARMM energy minimization, analyzed the energy profiles and finally generated the intermediate structures. The study showed that the computed intermediates along the transition pathway are similar to the known pdb structures.

In another recent study, large-amplitude conformational changes of AK were studied by combining ENM and geometric path planning algorithms developed in robotics research (Krillova *et al.,* 2008). The conformations were explored by using the combined method and the structure is directed by means of collective motions generated by low-frequency modes.

There have been other studies combining different approaches mentioned so far to improve the conventional methods and to eliminate the constraints. Amplified Collective Motion (ACM) is one of these methods, which was proposed by Zhang *et al*. (2003). ENM-derived normal modes are used in the method to increase MD simulation efficiency (Zhang *et al.,* 2003; He *et al.,* 2003). The method was applied to bacteriophage T4 lysozyme and villin headpiece sub domain (HP-36). Both applications achieved successful results in sampling low energy conformations and the coverage of the conformational space, thus accelerating the conformational sampling.

Berteotti et al. (2008) studied the large-scale conformational motions and the closure mechanism of a kinase by atomistic simulations. They presented a computational method to simulate the atomistic dynamics of the open-to-closed movement of the cyclin-dependent kinase 5 (CDK5). A new sampling method was used to find the minimum free-energy between the initial and the final state. The free-energy profile associated with the global motion and a CDK5 intermediate was also estimated.

Finally, Kantarci-Carsibasi *et al.* (2008) developed ANM-MC methodology to analyze the conformational transitional pathways and intermediate structures. Collective modes obtained from ANM are combined with MC simulation in this methodology. In ANM-MC methods, normal modes are continuously generated and updated by iterations in the computational simulation. The transitions between initial and final structures of AK

and hemoglobin together with the sequence of events in transitional pathways were investigated successfully. The obtained intermediate states were compared with the known intermediate structures and proved to have very close RMSD values. The second methodology studied was TMC method, which is based on a forcing algorithm towards the target state without using collective modes.

3. MATERIALS AND METHODS

3.1. Anisotropic Network Model (ANM)

ANM (Atilgan *et al.*, 2001), which is a coarse-grained NMA technique, is used to determine the vibrational motions in proteins. Directions of collective motions are predicted in ANM based on the 3D anisotropy of the residue fluctuations. It selects 3*N*-6 internal modes for a structure with N-residues and calculates the magnitude and direction vectors of residue displacements for these modes. C^{α} atoms of amino acids are assumed to be interaction centers in the coarse-grained representation of protein.

The total potential energy (V) of the folded protein structure is given as the summation of all harmonic interactions of (*i*, *j*) pairs, which are connected by harmonic springs with a universal constant γ in a cutoff distance, r_c as

$$V = (\gamma / 2)\left[\sum_{i,j}^{N} (R_{ij} - R_{ij}^{0})^2 H(r_c - R_{ij})\right] \qquad (3.1)$$

where $H(r_c - R_{ij})$ is the heavy side step function and equals 1 if $x \geq 0$, and zero otherwise]; and selects all residue pairs within the cutoff distance of r_c. γ is the universal force constant and R_{ij} is the distance between sites *i* and *j* in the native structure of protein. Thus, the overall potential formula includes the fluctuations for all components. The overall potential energy of a structure with *N* interaction sites is simplified as

$$V = (1/2)\, \Delta\mathbf{R}^{\mathrm{T}}\, \mathbf{H}\, \Delta\mathbf{R} \qquad (3.2)$$

where, $\Delta\mathbf{R}$ is a 3*N*-dimensional vector of the fluctuations $\Delta\mathbf{R}_i$ in the position vectors $\mathbf{R}_i$ of all sites ($1 \leq i \leq N$), $\Delta\mathbf{R}^{\mathrm{T}}$ is its transpose, and **H** is the Hessian matrix. For a structure with N residues, the second derivatives of the overall potential are calculated in (*N*x*N*) super-elements of size (3x3) which forms the (3*N*x3*N*) Hessian matrix, **H**, which is expressed as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H_{11}} & \mathbf{H_{12}} & \bullet & \bullet & \mathbf{H_{1N}} \\ \mathbf{H_{21}} & & & & \mathbf{H_{2N}} \\ \bullet & & & & \bullet \\ \bullet & & & & \bullet \\ \mathbf{H_{N1}} & & & & \mathbf{H_{NN}} \end{bmatrix} \tag{3.3}$$

where the ij^{th} super-element $\mathbf{H}_{ij}$ of $\mathbf{H}$ is,

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

where X_i, Y_i and Z_i being the components of $\mathbf{R}_i$. the symmetric Hessian matrix $\mathbf{H}$ is converted into the form in Equation 3.5 to calculate the normal modes of the elastic network,

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

$\mathbf{\Lambda}$ is the ($3N$ x $3N$) diagonal matrix including the eigenvalues or squared normal mode frequencies as diagonal elements. $\mathbf{U}$ is an ($3N$ x $3N$) orthogonal matrix and $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$, the identity matrix. The columns of this matrix are the normalized eigenvectors giving the normal mode directions of motion. In normal mode calculations, $3N$-6 individual internal modes are considered to describe the overall motion since the overall rotational and translational motion of the molecule due to corresponding six zero eigenvalues are excluded.

Previous studies have shown that basic fluctuation characteristics and important collective mode shapes can be generated by coarse-grained ANM method efficiently, i.e. the required computational time is much shorter when compared to the atomistic simulation techniques like molecular dynamics (MD) simulations (Bahar *et al.,* 1997a, 1998; Doruker *et al.,* 2000, 2002). The collective motions at low frequencies produced by ANM and essential motions from MD match with high overlaps (Doruker *et al.,* 2006; Cansu and Doruker, 2008).

3.2. Monte Carlo (MC) Simulation Technique

A coarse-grained Monte Carlo (MC) simulation method (Haliloglu and Bahar, 1998) has been developed to simulate protein dynamics at different time scales. Simulation results have been shown to be aligned with NMR measurement results, such as order parameters and hydrogen exchange data (Kurt and Haliloglu, 1999). This method provides efficient simulation of protein dynamics with multiple independent trajectories at long time scales.

In MC model, the backbone of the protein structure is represented by the virtual bond model, which was originally proposed by Flory (Flory, 1969). In this representation, each residue has two interaction sites, its alpha-carbon atom and side chain center of mass, as shown in Figure 3.1. In this representation of the protein, 3*N*-6 variables define the conformation of the backbone: *N*-1 backbone virtual bonds l_i (between alpha-carbon atoms i-1 and *i*), *N*-2 bond angles θ_i (the angle between l_i and l_{i+1}), and *N*-3 dihedral angles ϕ_i (the torsional rotation of the bond l_i). The side chain conformation is expressed by generalized variables $\{l_i^S, \theta_i^S, \phi_i^S\}$, where l_i^S is the bond length connecting backbone and side chain interaction site, θ_i^S is the bond angle between l_i and l_i^S, and ϕ_i^S is the torsion angle defined by l_{i-1}, l_i and l_i^S.

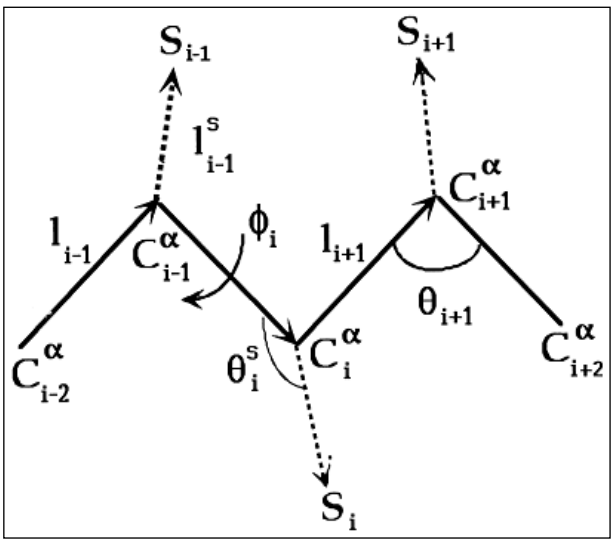


Figure 3.1. Schematic representation of the virtual bond model (Flory, 1969)

The energy $E(\Phi)$ of a given protein conformation Φ is calculated as the sum of two types of interactions: long-range (LR) interactions between non-bonded close residues, and short-range (SR) interactions between covalently bonded residues along the chain sequence as,

$$E(\Phi) = E_{LR}(\Phi) + E_{SR}(\Phi) \tag{3.6}$$

Bahar and Jernigan (1997) studied the long-range potentials to evaluate $E_{LR}(\Phi)$ according to Equation 3.7:

$$E_{LR}(\Phi) = \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} E_{SS}(r_{ij}) + \sum_{i=1}^{N-4}\sum_{j=i+4}^{N} E_{SB}(r_{ij}) + \sum_{i=1}^{N-5}\sum_{j=i+5}^{N} E_{BB}(r_{ij}) \tag{3.7}$$

where r_{ij} represents the distance between sites *i* and *j* in conformation Φ. The first term is the potential between side chain sites (SS); the second term is the potential between the side chain and backbone sites (SB); and the last term is the potential between two backbone sites (BB) of residues *i* and *j*, respectively.

The short-range conformational energy of the backbone is calculated from the addition of statistical potentials extracted from protein structures based on virtual bond model in Equation 3.8 (Bahar *et al.*, 1997b; Haliloglu and Bahar, 1998).

$$E_{SR}(\Phi) = \sum_{i=2}^{N} E(l_i) + \sum_{i=2}^{N-1} E(\theta_i) + \sum_{i=3}^{N-1}\left[E(\phi_i^-)/2 + E(\phi_i^+)/2 + \Delta E(\phi_i^-, \phi_i^+)\right] + \sum_{i=3}^{N-1}\left[\Delta E(\theta_i, \phi_i^-) + \Delta E(\theta_i, \phi_i^+)\right] \tag{3.8}$$

In this formula, the first summation term denotes the potential during the stretching of the virtual backbone bonds, the second summation term represents the bending of backbone bond angles; and the third term refers to the bond torsions ϕ_i^- and ϕ_i^+, which are the rotational angles of the virtual backbone bonds preceding and succeeding the i^{th} α-carbon, respectively. Pair wise interdependence of the torsion and/or bond angle bending is

also included in the equation. The statistical potentials, extracted from the probability distributions of side chain packing in low resolution models, are used to calculate the short-range conformational energy of the side chains (Kurt *et al.*, 2003). The energy of a residue type *A* associated with a side chain bond angle at state θ_i is expressed as,

$$E_A(\theta_i) = -RT\, ln\left[P_A(\theta) / P_A^0(\theta)\right] \tag{3.9}$$

where $P_A(\theta)$ is the statistical probability of the bond being at angle θ and $P_A^0(\theta)$ is the background probability assuming a uniform distribution. Similar equations of side chain conformation energy are derived in terms of side chain bond length and torsions as

$$E_{SR}^S(\Phi) = \sum_{i=1}^{N} E\left(l_i^S\right) + \sum_{i=2}^{N} E\left(\theta_i^S\right) + \sum_{i=2}^{N} E\left(\phi_i^S\right) \tag{3.10}$$

where l_i^S, θ_i^S and ϕ_i^S are the bond length, bond angle and torsion angle of side chain *I* respectively.

In the low resolution MC/Metropolis simulation technique, either an alpha-carbon or a side chain site is chosen randomly and is subjected to a differential perturbation based on a uniformly distributed random number generator. The strength of the perturbation, Δx, determined as,

$$\Delta x = k(2r - 1) \tag{3.11}$$

where r is the random number variable in the range $0 \le r \le 1$, and *k* is a damping factor that depends on the response at a given temperature. The move is accepted according to the Metropolis criterion, in which a new random number is generated and the new conformation Φ is accepted if the following condition is satisfied,

$$\exp\left\{-\left[E(\Phi) - E(\Phi)_0\right] / RT\right\} \le r \tag{3.12}$$

where Φ_o represents the original conformation.

One MC step (MCS) is expressed as *N* perturbations and can be viewed as the average time for *N* residues tending to move. Even though an MC algorithm with only local moves can give insight about the real time of a process, the implementation of both global deformations on collective modes and local moves shows no correspondence between MCS and real time. The analyzes are considered to be more reliable if multiple independent runs are applied for efficient sampling of the conformational space.

3.3. ANM-MC Simulation

Kantarci-Carsibasi *et al.,* (2008) developed an algorithm, in which collective moves are generated by ANM, and energy minimization of the deformed conformations is performed by short MC runs. Schematic flowchart of ANM-MC method is given in Figure 3.2. The initial and final/target structures of the protein are used as inputs to the algorithm after alignment. The initial and final states are also named as apo/open and bound/closed conformations of the protein, respectively.

In the first step (STEP 1) of the algorithm, atomistic structures of the initial and target states are coarse-grained by considering two nodes for each residue which are the alpha-carbon and the center of mass (com) of the side chain aligned with the knowledge based potentials used. Glycine has only one node (one alpha-carbon).

In the second step (STEP 2), the target structure is superimposed onto the initial structure based on the backbone carbon (alpha-carbon) coordinates to determine the target direction (***Q***). ***Q*** is a 3*N* dimensional vector including the alpha-carbon coordinate deformations and is calculated by subtracting the initial coordinates from the aligned target coordinates.

In the third step (STEP 3), m lowest frequency eigenmodes (m=10 in ANM-MC) are generated after ANM is applied to the initial structure by using BLZPACK (Block Lanczos Package), which is a Fortran 77 implementation of the block Lanczos algorithm (Grimes *et al.,* 1994; Marques, 1995, 2001).

In the fourth step (STEP 4), the dot products of the m eigenvectors with the target direction are calculated including both the positive and negative directions. The cutoff distance r_c in ANM is considered to be 18 Å, based on alpha-carbon coordinates. The eigenvector $\mathbf{u_i}$ (Eq. 3.5), which gives the highest dot product / overlap with the target direction $\boldsymbol{Q}$ is determined.

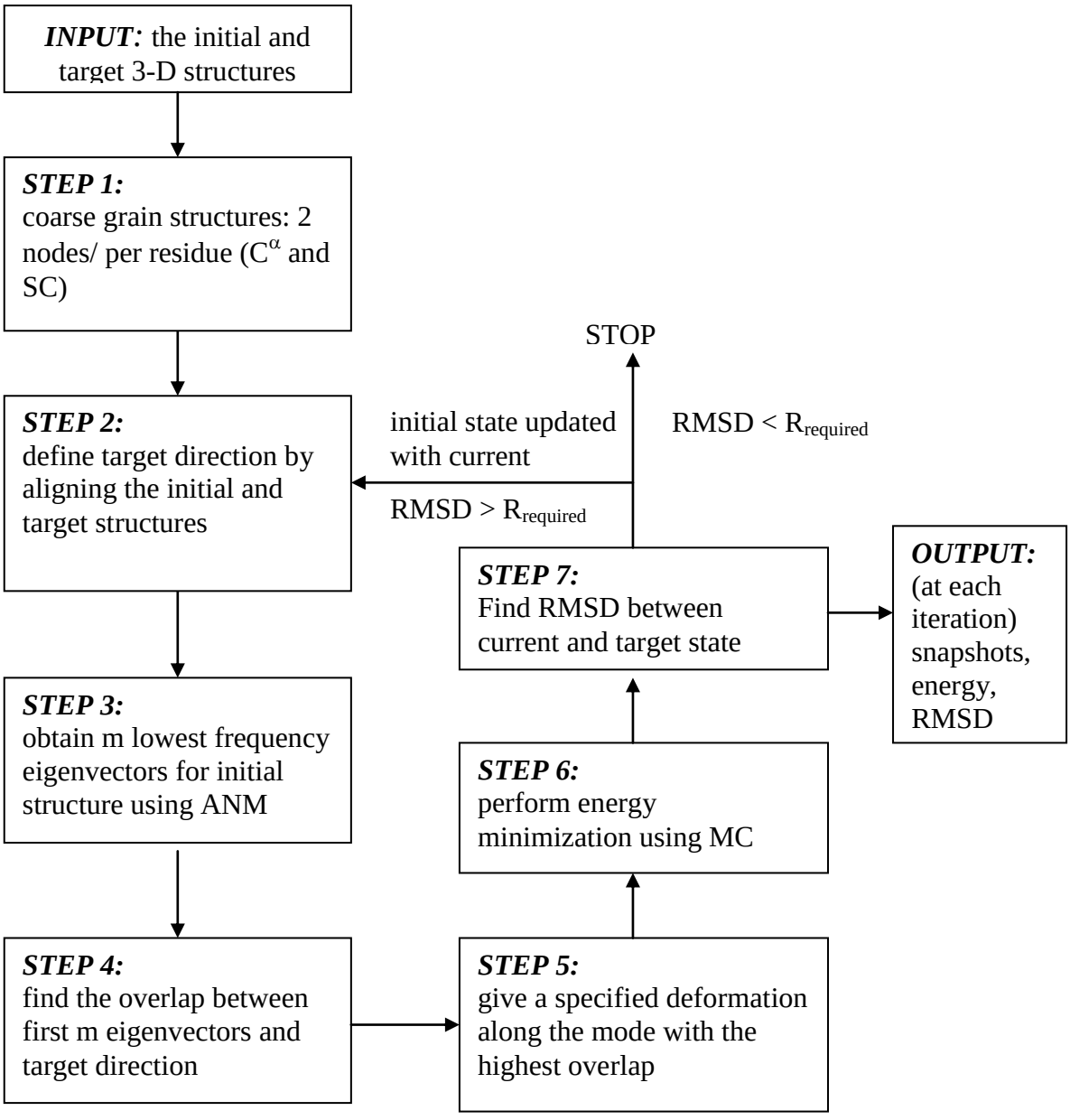


Figure 3.2. Flowchart of the ANM-MC algorithm (Kantarci-Carsibasi *et al., 2008*)

In the fifth step (STEP 5), the initial conformation is subjected to a specified deformation (deformation factor DF=0.2 Å) along the selected mode direction i to generate a new conformation which is given as: $\mathbf{R}_{new} = \mathbf{R} \pm \mathbf{u_i} * DF$. In this notation, $\mathbf{R}_{new}$ and $\mathbf{R}$ indicate the coordinate matrix of the new conformation and the initial conformation respectively. $\mathbf{u_i}$ is the i^{th} eigenvector which is multiplied by a rescale factor DF (DF=0.2 Å) to provide the corresponding RMSD (RMSD=0.2 Å) between the new and the old conformations. Deformation is applied to the alpha-carbon and the side chain nodes of each residue along the selected mode direction.

In the sixth step (STEP 6), MC algorithm is applied to the deformed conformation for energy minimization allowing to relax by 100 or 1000 MC steps.

In the final step (STEP 7), the RMSD between the target structure and the energy-minimized new structure is calculated by superimposing both structures according to Equation 3.13. Here, B_{rot} denotes the aligned (rotated) coordinates of the target structure, whereas $\mathbf{A}$ denotes the coordinates of the new structure. Only the alpha-carbon coordinates are taken into account in this calculation.

$$\text{RMSD} = \left[\frac{\sum_{i=1}^{N} [(x_{A,i} - x_{B_{rot},i})^2 + (y_{A,i} - y_{B_{rot},i})^2 + (z_{A,i} - z_{B_{rot},i})^2]}{N} \right]^{1/2} \quad (3.13)$$

If the RMSD is smaller than the desired value ($R_{required}$=0.1), the simulation stops; if it is greater than 0.1, then the updated initial structure and target structure enters a new iteration in STEP 2 to define the new target direction.

The ANM-MC simulation uses the initial and final structures of a protein and requires the input parameters; the deformation factor (DF), the number of MC steps (MCS) and perturbation strength for each iteration, and the cutoff distance is taken as 18 Å for ANM calculations (r_c = 18 Å). According to the previous study by Kantarci-Carsibasi *et al* (2008), the parameter adjustment was done as DF=0.2 Å and MCS=1000. In this thesis;

DF and MCS are considered to be 0.2 Å and 100, respectively. The perturbation strength was taken as 0.1.

3.4. Targeted Monte Carlo (TMC) Simulation

TMC simulations (Kantarci-Carsibasi *et al.*, 2008) differs from ANM-MC methodology with the fact that normal mode directions are not incorporated in TMC. Firstly, the protein is moved towards the target state starting from the initial state by deforming the coordinates along the target direction. In this application, collective modes are not used, so the new conformation can be calculated by $\mathbf{R}_{new} = \mathbf{R} \pm \boldsymbol{Q} * DF$. Next, MC is applied to the new conformation and the following algorithm is the same as ANM-MC (Figure 3.2). TMC is also similar to morphing algorithms (Krebs and Gerstein, 2000), in which linear interpolation is applied to the path between two conformations together with energy minimization. DF and MCS are the main parameters in TMC methodology. By using TMC simulations, the effect of incorporating normal mode directions in ANM-MC can be analyzed.

3.5. Proteins Studied

TMC and ANM-MC methods were applied to proteins consisting of one or two subunits; these proteins are adenylate kinase, antiviral protein (cyanovirin-N), calmodulin (CaM), dipeptide binding protein, diphtheria toxin, d-ribose binding protein, glutamine binding protein (GBP), helicase, human serum transferrin (HSTR), lactoferrin, lysine/arginine/ornithine (LAO) binding protein and small G-protein Arf6. The structural (Murzin *et al.*, 1995; Orengo *et al.*, 1997; Greene *et al.*, 2007) and functional data (RCSB Protein Data Bank) including the motion type, number of chains and residues; the pdb codes of the open/closed structures and the RMSD values between them are given in Table 3.1 and Table 3.2. The pdb files of the open and closed structures are taken from RCSB Protein Data Bank (RCSB Protein Data Bank) and the RMSD values between them are calculated by using the PyMOL Molecular Graphics System (DeLano, 2002). The structural and functional information of these proteins are given with a comparison of the open and closed structures in the following subsections. In this section, the structural domains based on CATH structural classification are indicated by different colors in

Figures 3.3-3.14 (Orengo *et al.*, 1997; Greene *et al.*, 2007). All proteins are illustrated in PyMOL (DeLano, 2002).

Moreover, dynamic domains in the proteins are determined by using the Dyndom Program, which determines fixed and moving domains, hinge axes and hinge bending residues based on two distinct protein conformations (Hayward *et al.*, 1997; Hayward and Berendsen, 1998). In all figures, the blue regions denote the fixed domain, whereas others denote the moving domain(s) (DeLano, 2002).

Table 3.1. The pdb codes of the initial/target structures, RMSD values, number of chains and residues of the proteins studied (RCSB Protein Data Bank; DeLano, 2002).

Protein Name	**Initial pdb**	**Target pdb**	**RMSD (Å)**	**Number of chains**	**Total number of residues**
Adenylate Kinase	4ake	1ake	7.13	1	214
Antiviral protein (cyanovirin-N)	1l5e	1l5b	7.01	2	202
Calmodulin	1cll	1a29	15.01	1	143
Dipeptide-binding protein	1dpe	1dpp	6.53	1	507
Diphtheria toxin	1ddt	1mdt	15.63	1	522
D-Ribose binding protein	1urpA	2dri	4.06	1	271
Glutamine binding protein	1gggA	1wdnA	5.34	1	220
Helicase	8ohm	1cu1A	4.39	1	435
Human Serum Transferrin	1bp5A	1a8e	6.70	1	328
Lactoferrin	1lfh	1lfg	6.41	1	690
LAO binding protein	2lao	1lst	4.70	1	238
Small G-protein Arf6	1e0s	2j5x	4.17	1	164

Table 3.2. The structural and functional data of the proteins studied (Orengo *et al.*, 1997; Greene *et al.*, 2007; Murzin *et al.*, 1995; Hayward *et al.*, 1997; Hayward and Berendsen, 1998)

Protein Name	**Motion Type**	**Function**	**Number of CATH Domains**	**CATH Domains**	**SCOP Classification**	**Dynamic Domains**	**Number of dynamic domains**
Adenylate kinase	Hinge bending	Nucleotide, ATP, AMP binding	3	Core domain LID domain AMP-bind domain	Alpha & beta	Fixed domain Moving Domain Moving Domain	3
Antiviral protein (cyanovirin-N)	Hinge bending	Sugar binding	2	C-terminal domain N-terminal domain	All beta	Fixed domain Moving Domain	2
Calmodulin	Hinge bending	Calcium binding	2	N-terminal domain C-terminal domain	All alpha	Fixed domain Moving Domain	2
Dipeptide-binding protein	Hinge bending	Dipeptide binding	2	Domain 1 Domain 2	Alpha & beta	Fixed domain Moving Domain	2
Diphtheria toxin	Hinge bending	Transferase activity	3	R domain T domain C domain	Alpha & beta Mainly alpha Mainly beta	Fixed domain Fixed domain Moving Domain	2
D-Ribose binding protein	Hinge bending	Sugar binding	2	Domain 1 Domain 2	Alpha & beta	Fixed domain Moving Domain	2
Glutamine binding protein	Hinge bending	Glutamine binding	2	Large domain Small domain	Alpha & beta	Fixed domain Moving Domain	2
Helicase	Hinge bending	Enzyme for binding nucleic acid	3	NTPase domain RNA binding domain Helical domain	Alpha & beta Alpha & beta Mainly alpha	Fixed domain Moving Domain Fixed domain	3
Human Serum Transferrin	Hinge bending	Iron binding	2	Domain 1 Domain 2	Alpha & beta	Fixed domain Moving Domain	2
Lactoferrin	Hinge bending	Iron binding	4	Domain 1 Domain 2 Domain 3 Domain 4	Alpha & beta	Fixed domain Moving Domain Moving Domain Moving Domain	3
LAO binding protein	Hinge bending	Transport activity	2	Domain 1 Domain 2	Alpha & beta	Fixed domain Moving Domain	2
Small G-protein Arf6	Shear motion	Nucleotide binding	2	Domain 1 Domain 2	Alpha & beta	Fixed domain Moving Domain	2

3.5.1. Adenylate Kinase (AK)

AK is a monomer with 214 residues. It undergoes a large conformational change of hinge bending during their catalytic cycle. Its function is to catalyze the transfer of the terminal phosphoryl group from adenosine trio phosphate (ATP) to adenosine monophosphate (AMP). The free structure from *Escherichia coli* was determined at 2.2 A resolution. It has three domains; core domain (blue), ATP-binding (LID) domain (red) and AMP-binding (yellow) domain. Upon substrate binding, the enzyme increases its chain mobility in a region remote from the active center (Müller *et al.*, 1992). The closed structure, which is ligated with the two-substrate-mimicking inhibitor P1,P5-bis (adenosine-5'-) pentaphosphate (AP5), has been determined at 1.9 A and contains two copies of the complex. The RMSD value between the initial and target structures is 7.13 Å.

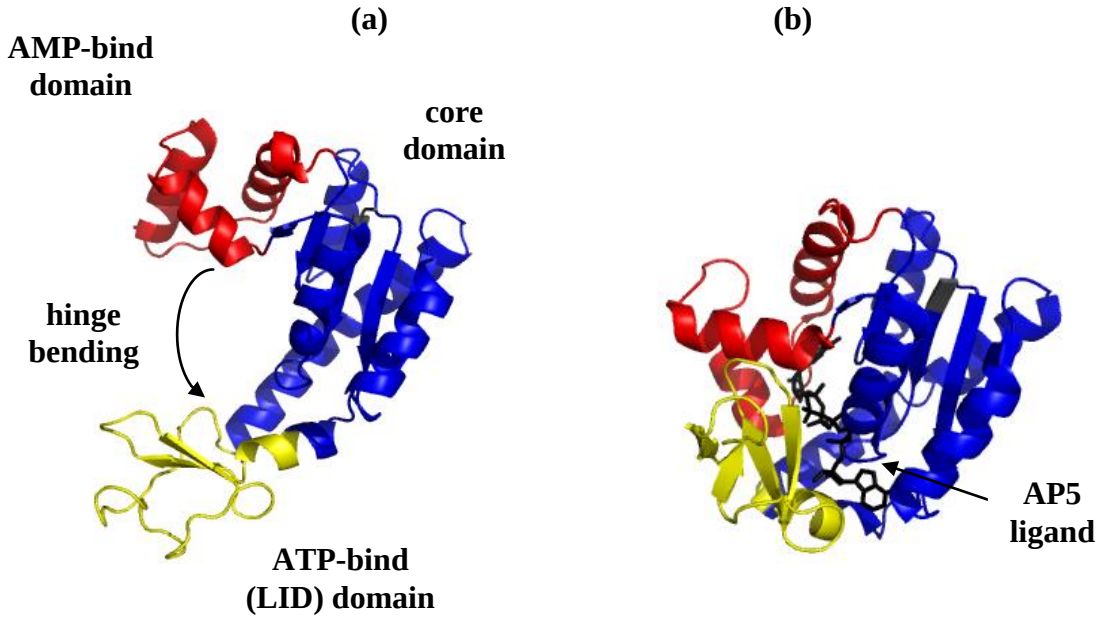


Figure 3.3. Ribbon diagrams for (a) apo/open conformation (PDB code: 4AKE) and (b) bound/closed conformation (PDB code: 1AKE) of adenylate kinase

3.5.2. Antiviral Protein

Antiviral protein (cyanovirin-N) from *Cyanobacterium* is a sugar binding protein with two chains of 101 residues for each. The protein is found to be a domain-swapped

dimer by X-ray crystallography. The dimer structure is metastable and kinetically trapped at neutral pH and room temperature. Its SCOP structural classification is “all beta sheets”. It interacts selectively and non-covalently with any mono-, di- or trisaccharide carbohydrates. The structure has hinge bending movement. The closed structure is a complex with Na^{+} ion and 2-[N-Cyclohexylamino] Ethane Sulfonic Acid (NHE) ligand. RMSD value between the initial and the target state is 7.01 Å (Barrientos *et al.,* 2002). Each domain, indicated by a different color according to the SCOP structural classification, contains two chains. Both domains are moving domains.

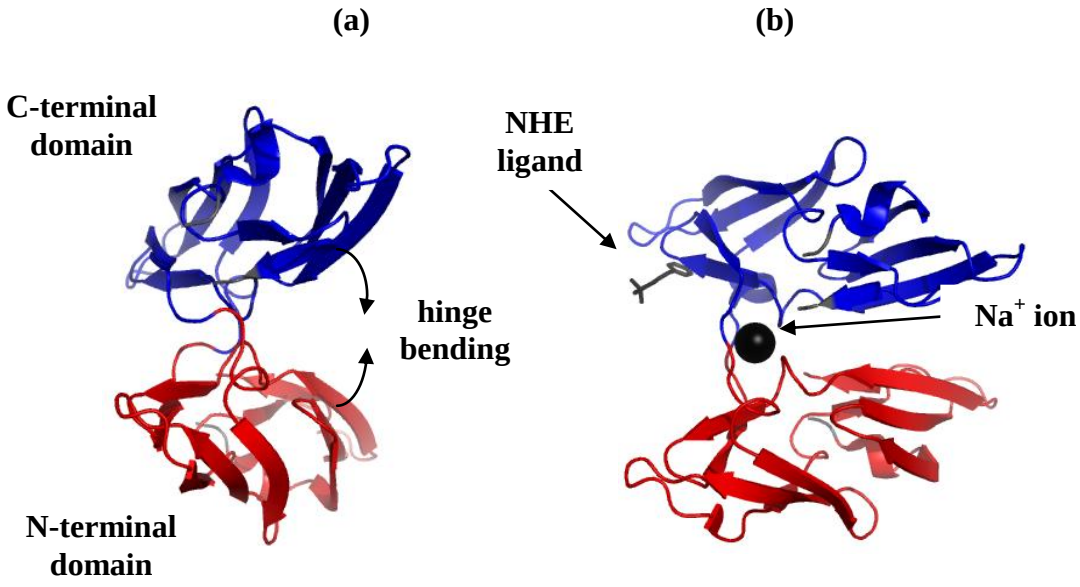


Figure 3.4. Ribbon diagrams for (a) apo/open conformation (PDB code: 1L5E) and (b) bound/closed conformation (PDB code: 1L5B) of antiviral protein

3.5.3. Calmodulin (CaM)

CaM is a calcium (Ca^{+2}) binding protein, which selects and non-covalently binds to calcium ions. The crystal structure of CaM was determined and refined at 1.7 Å resolution. It is a dumb-bell-shaped and monomeric molecule, with similar lobes connected by a central alpha-helix. It consists of two EF-hand units, each including two helices connected with calcium-binding loop. It has two domains; N-terminal and C-terminal domains. It is classified as "all alpha" protein in terms of SCOP structure. There is a large hinge bending movement, while the drug trifluoperazine and Ca^{+2} ions are bound to the CaM forming the closed structure. The initial closed structure is Ca^{+2} ions and ethanol

(EOH) bound structure. The RMSD value between bound/bound structures is 15.01 Å. The protein has 146 residues, including the first three missing residues (Chattopadhyaya R. *et al.*, 1992, Vertessy *et al.*, 1998, Kuboniwa *et al.*, 1995).

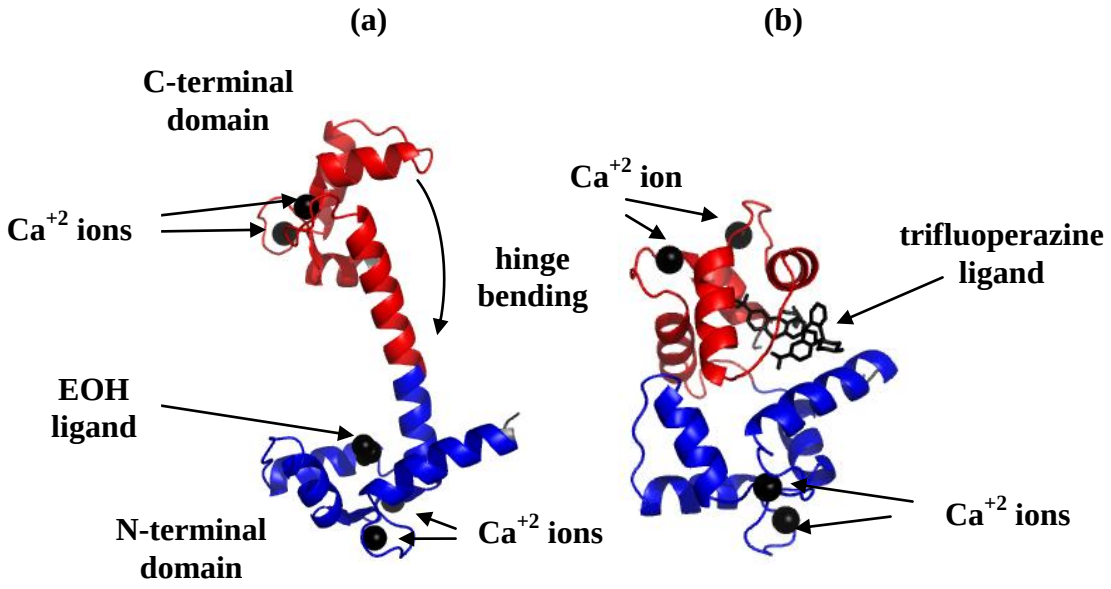


Figure 3.5. Ribbon diagrams for (a) initial bound conformation (PDB code: 1CLL) and (b) target bound conformation (PDB code: 1A29) of CaM

3.5.4. Dipeptide Binding Protein

Dipeptide binding protein is a monomeric transport protein with 507 residues. It functions as the initial receptor for permeases and active transport systems. The free structure contains two distinct domains, which are connected by two hinges, and has a bound Cd^{+2} ion. The closed structure of the dipeptide binding protein in complex with Gly-Leu (glycyl-L-leucine) has been determined at 3.2 Å resolution. Complex protein is used in peptide transport and taxis toward peptides. This protein prefers to bind shorter peptides. The protein structure is classified as "alpha and beta" including mainly parallel beta sheets (beta-alpha-beta units). The protein exhibits a conformational change of hinge bending type. The initial state has an RMSD of 6.53 Å from the target state (Dunten *et al.*, 1995; Nickitenko *et al.*, 1995).

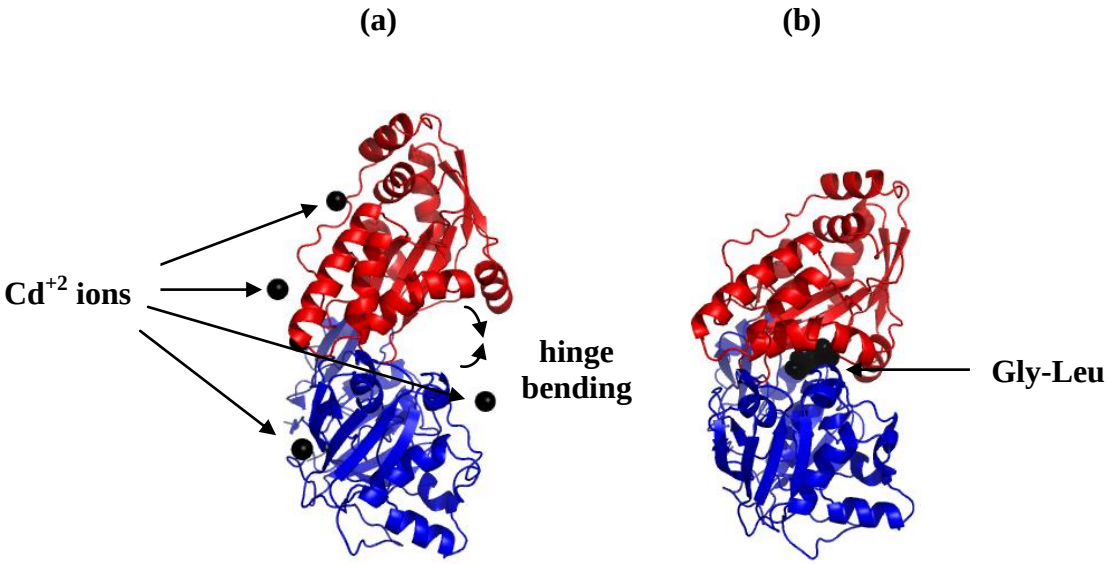


Figure 3.6. Ribbon diagrams for (a) apo/open conformation (PDB code: 1DPE) and (b) bound/closed conformation (PDB code: 1DPP) of dipeptide binding protein

3.5.5. Diphtheria Toxin (DT)

The refined structure of dimeric diphtheria toxin was determined at 2.0 Å resolution. The protein has 535 residues with 12 missing residue between 188-199th residues. The refined model contains 522 amino acid residues, 1 molecule of the bound dinucleotide inhibitor adenylyl 3'-5' uridine 3' monophosphate (ApUp), and 405 well-ordered water molecules. ApUp is bound by a long loop (residues 34-52) crossing the active site. Glu 148 was identified as playing a catalytic role ADP-ribosylation (Adenosine diphosphate-ribosylation) of elongation factor 2 by DT. The closed structure was determined at 2.3 Å resolution.

The protein has three domains; the receptor-binding (R) domain (blue), the transmembrane (T) domain (light blue) and the catalytic (C) domain (red). DT changes conformations of hinge motion while opening or closing the receptor-binding (R) domain. Structure-based mechanism of intoxication includes five steps. In the first step, proteolysis of a disulfide-linked surface loop (residues 186-201) between the catalytic (C) and transmembrane (T) domains takes place. The second step is the binding of a beta-hairpin loop protruding from the R domain to the DT receptor, leading to receptor-mediated

endocytosis. In the third step, open monomer is formed at low pH and cause the exposure of apolar surfaces in the T domain, which insert into the endosomal membrane. In the fourth step, the C domain is translocated into the cytosol. In the last step, catalysis was performed by the C domain of ADP-ribosylation of elongation factor 2 (Bennett *et al.*, 1994a; 1994b).

The monomeric protein has a RMSD value of 15.63 Å between open and closed conformations and hinge bending movement takes place. MC simulations were applied to this structure in two different approaches. In the first approach, the structure was assumed to be monomeric with a long bond of 8 Å connecting the missing residues. In the latter case, it was designed as dimeric protein with the missing residue region comprising the chain ends. Missing residue region is shown on the open structure.

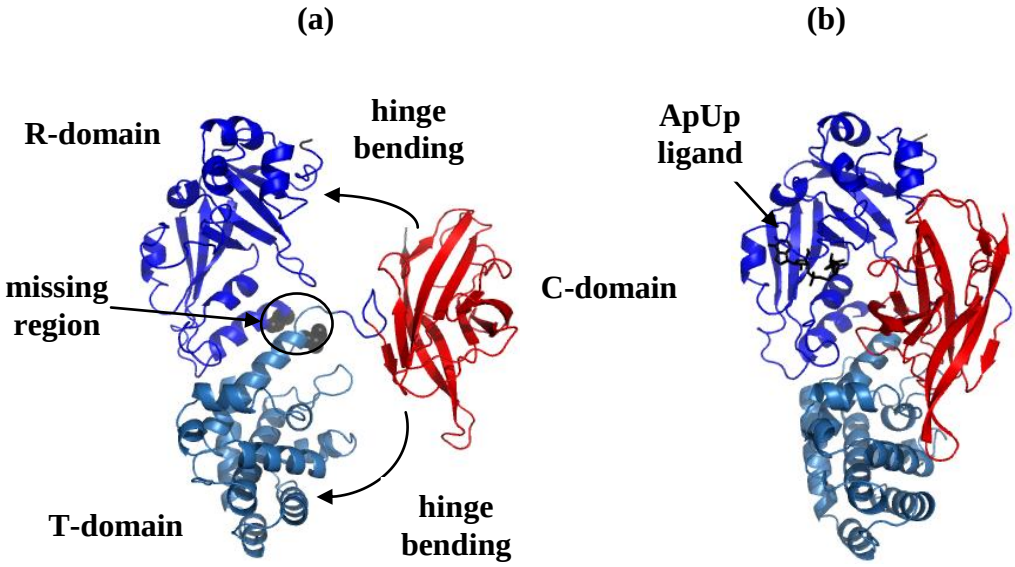


Figure 3.7. Ribbon diagrams for (a) apo/open conformation (PDB code: 1DDT) and (b) bound/closed conformation (PDB code: 1MDT) of diphtheria toxin

3.5.6. D-Ribose Binding Protein

D-ribose binding protein is a monomeric transport and sugar binding protein with 271 residues. Open (ligand-free) form of the *Escherichia coli* ribose-binding protein was determined by X-ray crystallographic studies at 2.3 Å resolution. The protein is classified

as "alpha and beta" due to the SCOP structure. Ribose (RIP) ligand is bound to the closed state. There is a hinge bending movement between open and closed states with an RMSD of 4.06 Å (Bjorkman *et al.*, 1994; 1998).

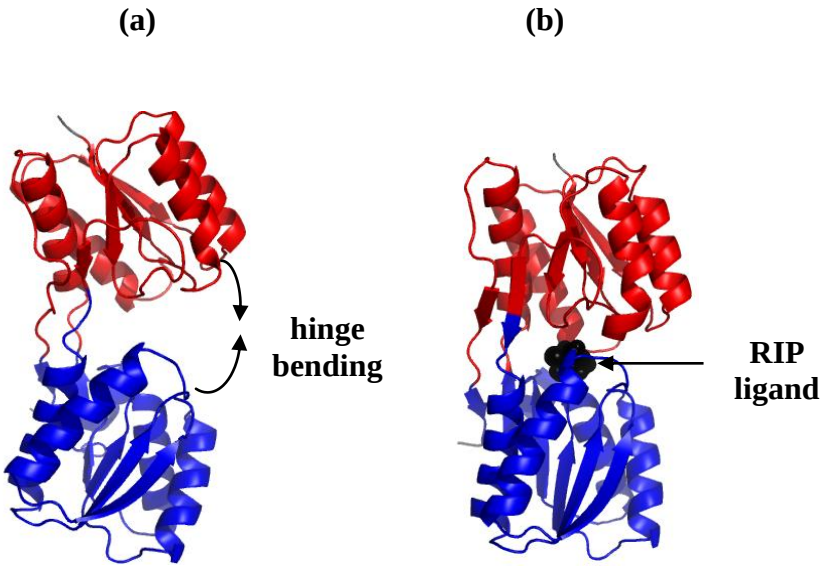


Figure 3.8. Ribbon diagrams for (a) apo/open conformation (PDB code: 1URP-A chain) and (b) bound/closed conformation (PDB code: 2DRI) of d-ribose binding protein

3.5.7. Glutamine Binding Protein (GBP)

Glutamine binding protein from *Escherichia coli* is a monomer with 220 residues and two domains. Domains have a similar super secondary structure and linked by two antiparallel beta-strands. The small domain contains three alpha-helices, four parallel and one antiparallel beta-strands, whereas the large domain contains two additional alpha-helices and three more short antiparallel beta-strands. The protein is classified as "alpha and beta" due to the SCOP structure. GBP complexed with its ligand Gln is the closed structure. The open structure was defined at 2.3 Å resolution, whereas the complex was refined to 1.94 Å resolution. The RMSD value between open and closed structures is 5.34 Å due to the hinge bending movement (Sun *et al.*,1998; Hsiao *et al.*, 1996).

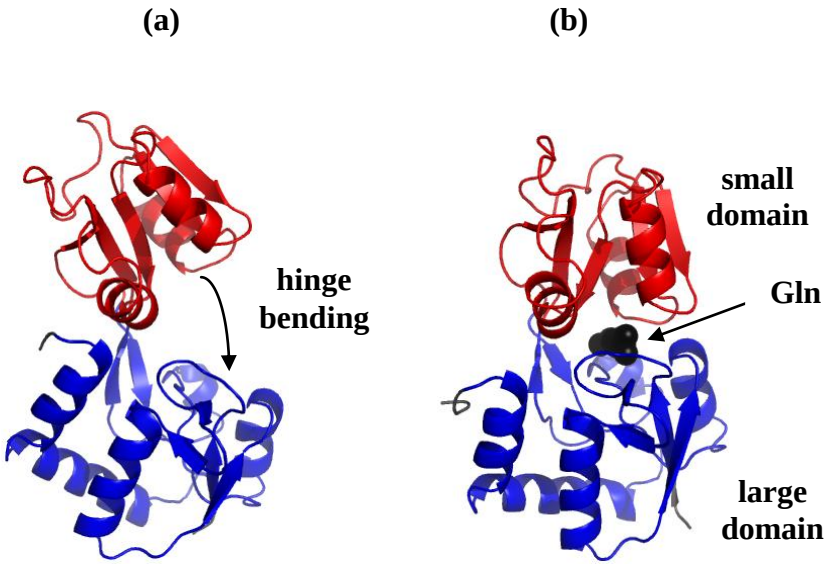


Figure 3.9. Ribbon diagrams for (a) apo/open conformation (PDB code: 1GGG-A chain) and (b) bound/closed conformation (PDB code: 1WDN-A chain) of GBP

3.5.8. Helicase

Helicases are enzymes which catalyze strand separation of double stranded deoxyribonucleic acid (dsDNA)1 or double stranded ribonucleic acid (dsRNA) coupled with hydrolysis of NTP. They perform their function for many cellular events including transcription, RNA processing, translation, and DNA or RNA replication (,1-3). The crystal structure of RNA helicase domain from genotype 1b hepatitis C virus was determined at 2.3 Å. The protein has 435 residues. PO_4^{-3} ion is bound to the closed structure. The RMSD value between the open and closed states is 4.39 Å. Three domains in the structure form a Y-shaped molecule. The domain that binds the RNA binding is distinctively separated from the other two domains. The SCOP structure is classified as "alpha and beta". Helicase undergoes hinge bending motion (Cho *et al.*, 1998; Yao *et al.*, 1999).

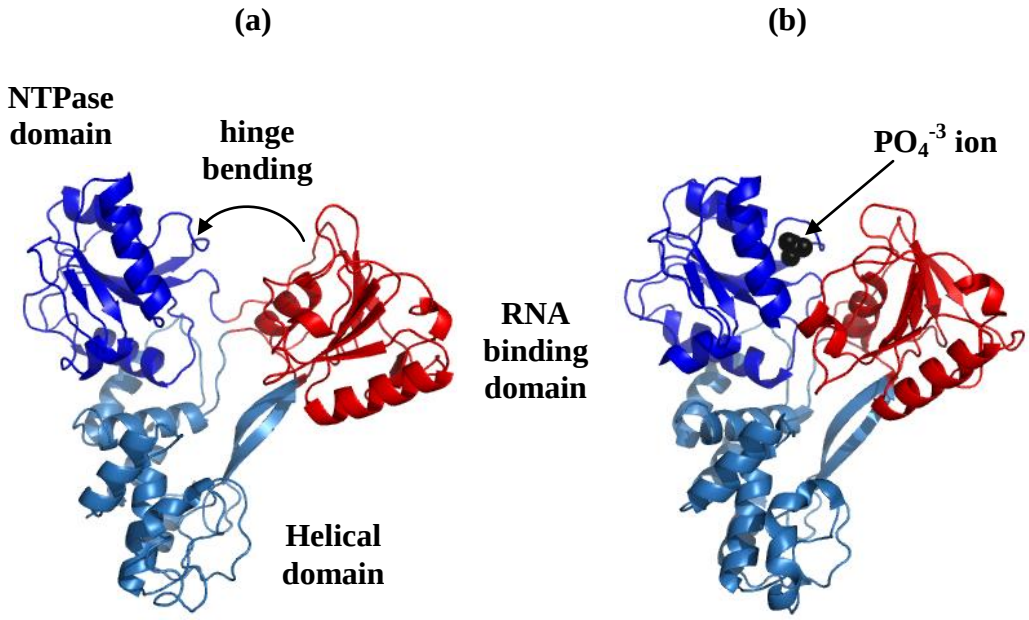


Figure 3.10. Ribbon diagrams for (a) apo/open conformation (PDB code: 8OHM) and (b) bound/closed conformation (PDB code: 1CU1-A chain) of helicase

3.5.9. Human Serum Transferrin (HSTR)

Human Serum Transferrin binds to ferric ions in the bloodstream and transports them to cells, where they are released depending on pH. To understand iron release process, the large conformational change of the open structure was studied by X-ray crystallography. The protein structure is classified as "alpha and beta. The protein has a conformational change by hinge motion. The conformational change is a rigid-body motion combined with several local adjustments. It is a monomer containing 331 residues (first 3 residues are missing). Fe^{+2} and CO_3^{-2} ions are bound to the closed state. There is an RMSD value of 6.7 Å between open and closed structures (Jeffrey *et al.*, 1998; Macgillivray *et al.*, 1998).

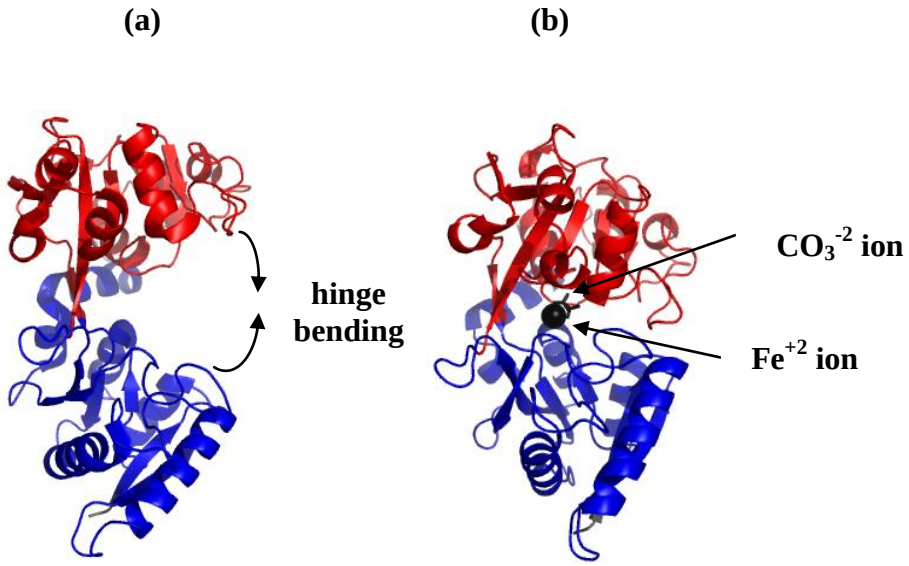


Figure 3.11. Ribbon diagrams for (a) apo/open conformation (PDB code: 1BP5-A chain) and (b) bound/closed conformation (PDB code: 1A8E) of HSTR

3.5.10. Lactoferrin

Lactoferrin, which is found in secretions such as milk and tears, has an iron-binding function. It is a single chain with 691 residues and the structure was determined at 2.8 Å resolution. The protein contains four domains. The protein structure is classified as "alpha and beta. Cl^- ion is bound to the open state, whereas Fe^{+2} ion, carbonate ligand (CO_3^{-2}), are bound to the closed state. Lactoferrin has a conformational change of hinge motion. It has two hinge regions at residues 89-92 and 249-252. The RMSD value between open and closed forms is 6.41 Å (Norris G.E. *et al.*, 1991).

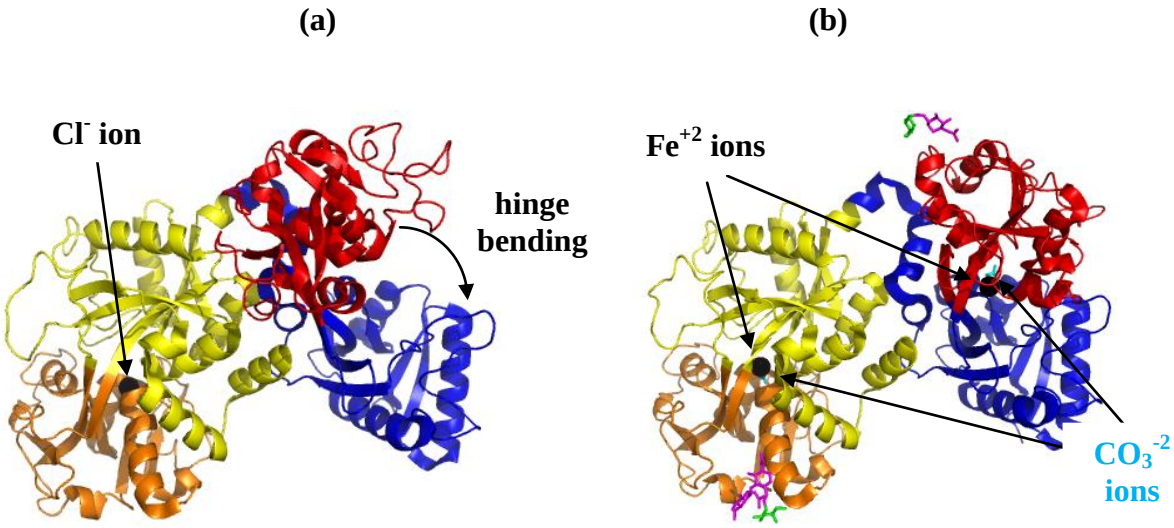


Figure 3.12. Ribbon diagrams for (a) apo/open conformation (PDB code: 1LFH) and (b) bound/closed conformation (PDB code: 1LFG) of lactoferrin (green ligands are FUC ligands, magenta ligands are NAG ligands, cyan-colored CO_3^{-2} ions are behind Fe^{+2} ions)

3.5.11. LAO binding protein

LAO binding protein is responsible for binding Lysine, Arginine or Ornithine amino acids in the cell. It has one chain of 238 residues. The three-dimensional structures of LAO binding protein from *Salmonella typhimurium* with and without lysine were determined by x-ray crystallographic methods at 1.8 and 1.9 Å resolution, respectively. The RMSD value between both structures is 4.7 Å. The open structure has two lobes, which are far apart but in contact with each other in the closed structure. The protein is classified as "alpha and beta" based on SCOP structure. The protein has a hinge bending motion (Oh *et al.*, 1993).

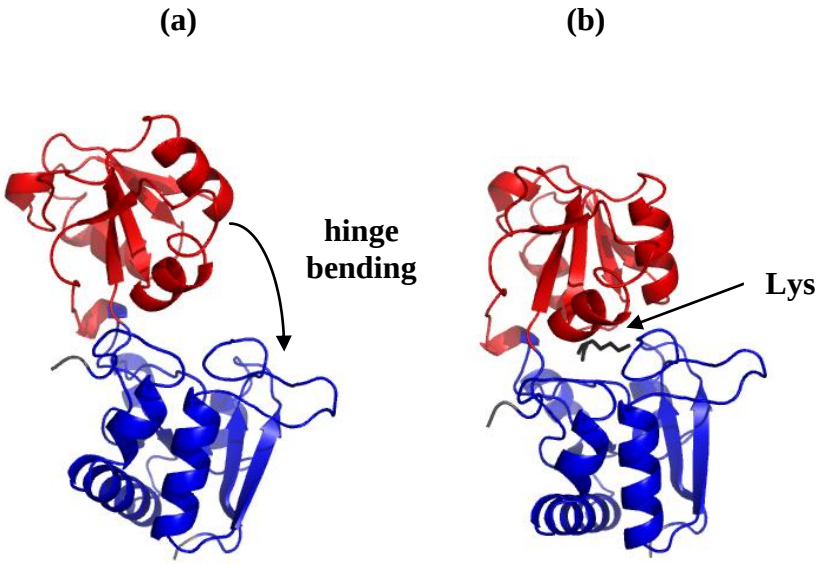


Figure 3.13. Ribbon diagrams for (a) apo/open conformation (PDB code: 2LAO) and (b) bound/closed conformation (PDB code: 1LST) of LAO binding protein

3.5.12. Small G-protein Arf6

Small G-protein Arf6 has an essential role in endocytotic pathways at the periphery of the cell. The crystal structure of Arf6-GDP was determined at 2.3 Å resolution. The protein is a monomer with 174 residues (first 10 residues are missing). Mg^{+2} metal ion and 5'-guanosine-diphosphate-monothiophosphate (GSP) ligand are bound to the final state whereas ligands guanosine-5'-diphosphate (GDP), beta-mercaptoethanol (BME) and ammonium (NH_4^+) ion are bound to the initial state. In fact, both states are bound and closed states. However, there is a transition between both structures. The protein is classified as "alpha and beta" according to the SCOP structure. The motion type is shear motion. The RMSD value between open and closed structures is 4.17 Å (Vitagliano *et al.*, 2000; Pasqualato *et al.*, 2001).

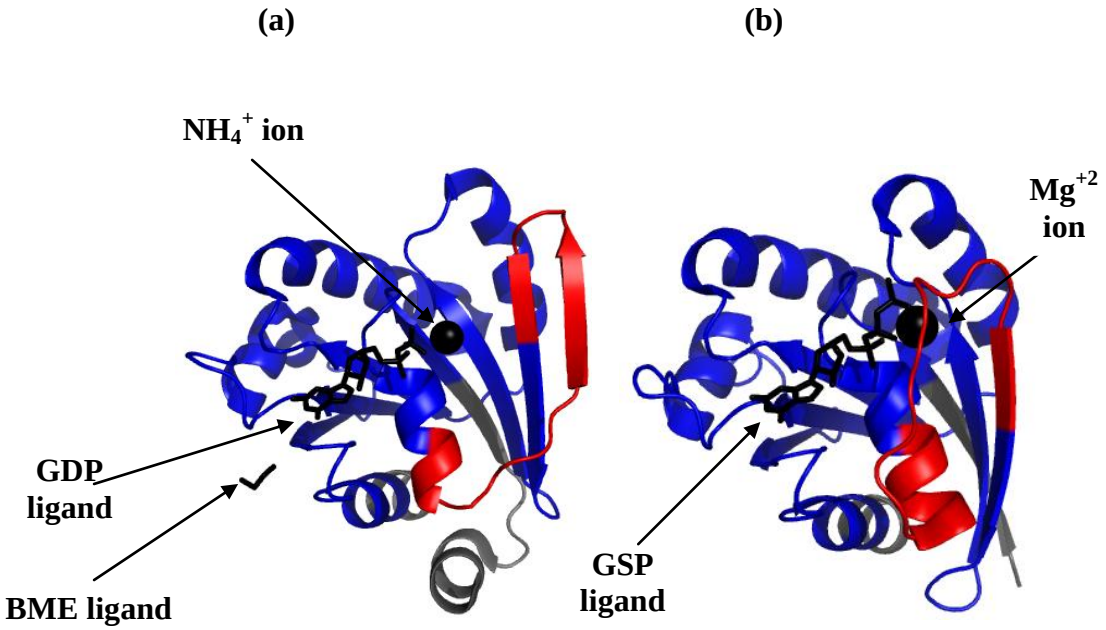


Figure 3.14. Ribbon diagrams for (a) initial bound/closed conformation (PDB code: 1E0S) and (b) target bound/closed conformation (PDB code: 2J5X) of small G-protein Arf6

4. TMC AND ANM-MC SIMULATION RESULTS

In this chapter, TMC and ANM-MC simulations are used to analyze the conformational transitions for the proteins; AK, antiviral protein (cyanovirin-N), CaM, dipeptide binding protein, diphtheria toxin, d-ribose binding protein, glutamine binding protein, helicase, HSTR, lactoferrin, LAO binding protein and small G-protein Arf6 based on the previous study by Kantarci-Carsibasi *et al*. (2008) and Kantarci-Carsibasi (2009). The simulations used in the relevant study was revised to reduce the iteration time and to be applicable to large proteins and those with high conformations which indicates high RMSD values (>4Å) from the target structure. The water molecules and the ligand atoms are not taken into consideration in the simulation.

Forward ANM-MC and reverse ANM-MC simulations were applied at MCS100 (Monte Carlo step=100) iterations. MCS=100 is preferred in most cases due to shorter computational times to increase the computational efficiency. DF is taken as 0.2 for ANM-MC simulations.

RMSD values from the target structure, in other words the degree of approach to the target is calculated for each MC and ANM-MC simulation as given in Table 4.1. The RMSD and potential energy profiles are plotted in Figures 4.10-21, where "r" denotes the reverse runs. The overlap value and selected modes are investigated for ANM-MC runs. These results are also presented in Table 4.2 and also in Figures 4.10-4.21. The x-axis of the graphs represents the number of iterations. "r" denotes the reverse The conformational transition pathways are analyzed by illustrating the simulation snapshots and the contact maps at specific iterations. The x-axis of the graphs represents the number of iterations. In ANM-MC simulations, one iteration includes the ANM calculation followed by deformation along the selected mode with the specified DF, and then followed by an MC simulation of 100 steps (MCS=100).

4.1 Parameter Adjustment

In ANM-MC and TMC simulations, there are three adjustable parameters, namely the deformation factor, the perturbation strength (k in Eq. 3.11), and the number of Monte Carlo steps (MCS). In the initial ANM-MC and TMC studies (Kantarci-Carsibasi *et al.*, 2008; Kantarci-Carsibasi, 2009), these parameters have been set at DF = 0.2, perturbation strength (k = 0.01), and MCS = 100. However, virtual bond-lengths for few residues in the highly mobile loop regions have exceeded their equilibrium values due to insufficient energy minimization. In order to see the effect of more stringent energy minimization on the generation of intermediate structures, the strength of perturbation applied to the nodes, i.e. local moves during MC simulation, is increased from 0.01 to 0.1 in this thesis recently. For DF = 0.2, various MCS values are tested both to attain sufficient energy minimization and computational efficiency.

For this parameter adjustment stage, the conformational transition of adenylate kinase (AK) has been focused on for comparison with previous simulations (DF = 0.2 and MCS = 100 for both cases). Figure 4.1 exhibits almost coinciding RMSD profiles that decrease to a plateau value of about 2.4 Å in both recent and previous simulation parameters. However, the RMSD value between the two final structures obtained at the plateau is 1.76 Å. Similarly, Figure 4.2 indicates quite similar overlap and selected mode profiles in both recent and previous simulations. In Figure 4.3, the comparison of the total potential energy profiles is given for previous (red dashed-line) and recent (red solid line) ANM-MC100 forward simulations, As expected, the energies after MCS=100 (at each iteration) are much lower in the recent simulations employing higher perturbation strength. The bond-length for the virtual bond connecting the backbone alpha-carbons of residues 126 and 127 is plotted in Figure 4.4 as a function of time. Fluctuation of virtual backbone bond-lengths around 3.8 Å can be clearly observed in the recent study, whereas an unrealistic increasing trend has been obtained in the previous study. With the new parameter set, all virtual bonds for backbone and side chains fluctuate around equilibrium values during the course of the simulations.

When the contact maps for the target x-ray structure and the final structures obtained from simulations are compared, more native contacts are formed with the new

parameter set (Figure 4.5). In terms of transition pathway, the closure of LID domain precedes that of the AMP domain, consistently in both simulations. The initial contacts formed between the LID and the core domains are indicated by the red circle in Figure 4.6. According to these results, perturbation strength value of 0.1 seems suitable for performing recent simulations.

Figure 4.7 displays the effect of several MCS values on energy minimization, namely for 20 (blue), 50 (black), and 100 (red) MCS. For DF = 0.2 and perturbation strength = 0.1, the results for MCS = 50 and 100 are more satisfactory in terms of energy minimization. In terms of contact maps, same profiles similar to that of target structure are obtained using all MCS values (Figure 4.8). As a result, MCS = 100 value is chosen and used in all simulations in this thesis.

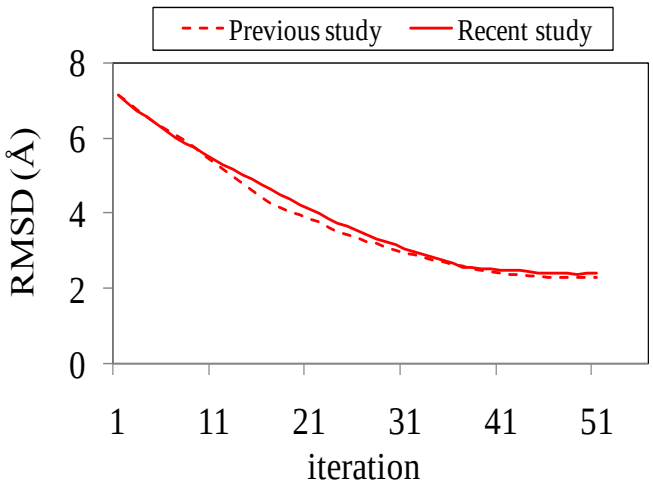
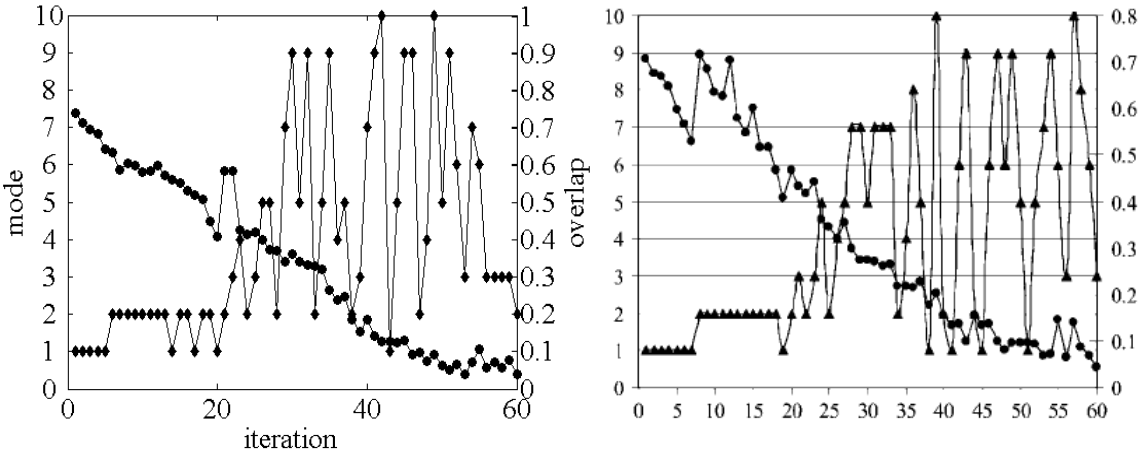


Figure 4.1. Comparison of RMSD profiles, previous study (dashed-line) and recent study (line)



(a) recent study (b) Previous study

Figure 4.2. Overlap and selected modes

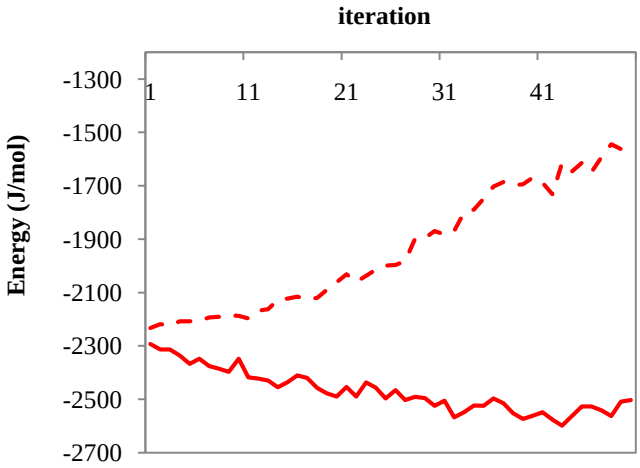


Figure 4.3. Total potential energy profiles, previous study (red dashed-line) and recent study (red line)

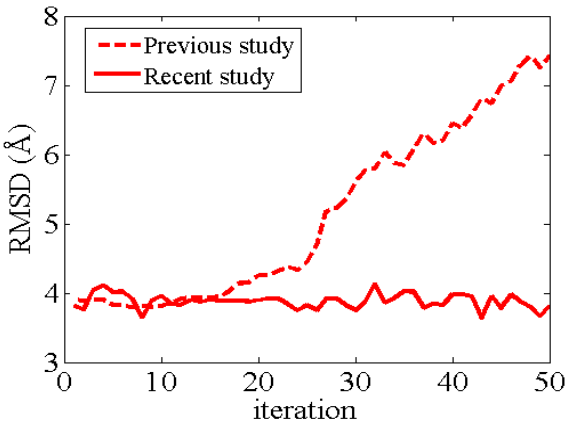
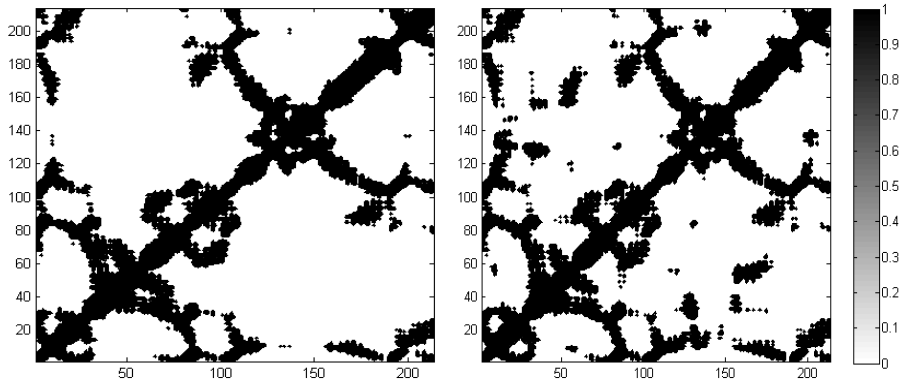
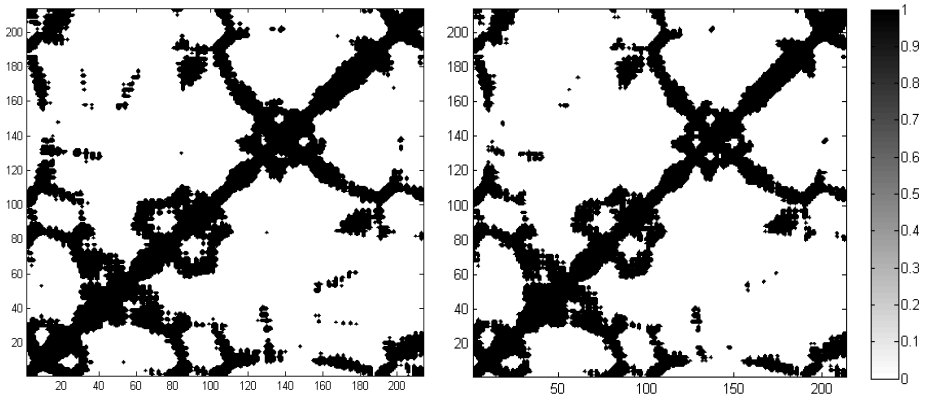


Figure 4.4. Virtual bond-length between 126th -127th residues, previous study (red dashed-line), recent study (red line)



(a) Initial structure (b) Target structure



(c) Recent study ANM-MC100 (d) Previous study ANM-MC100

Figure 4.5. Contact maps of the final structure in the recent and previous studies

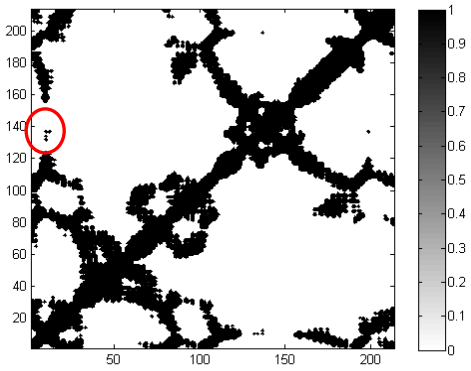


Figure 4.6. Contact map showing contact formation in LID domain (snap 3 from ANM-MC100)

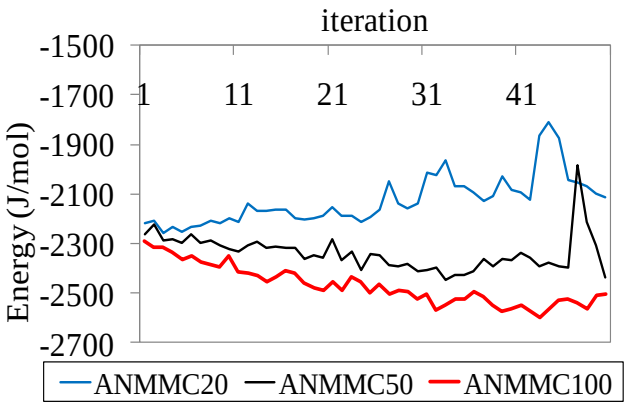
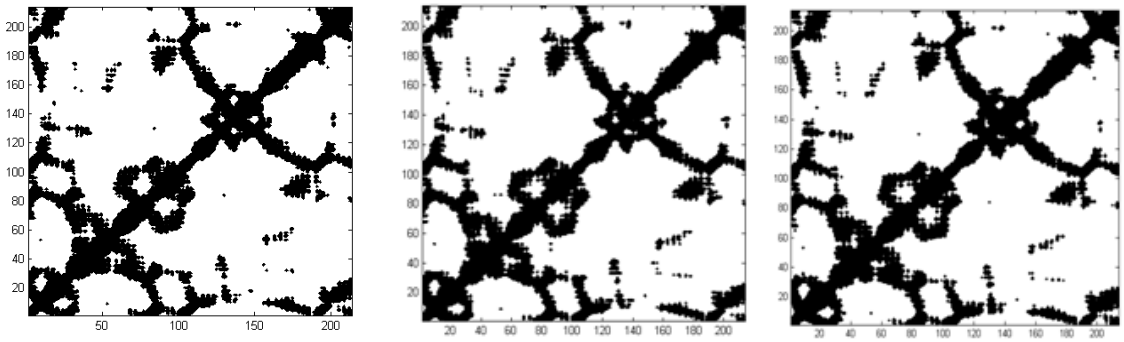


Figure 4.7. Total potential energy profiles for ANM-MC, MCS20 (blue), MCS50 (black), and MCS100 (red)



(a) ANM-MC20 (b) ANM-MC50 (c) ANM-MC100

Figure 4.8. Contact maps at different MCS

4.2 RMSD Profile

In TMC simulations, all proteins approached the target structure with an RMSD value less than 0.4 Å. In TMC, the final RMSD value reach a plateau at the end of the simulation in less iterations and thus in a shorter period when compared to ANM-MC iterations. The computational time also depends on the initial RMSD value of the protein, which is shown in Figure 4.9. RMSD values obtained in forward and reverse TMC simulations follow a similar profile towards the target state.

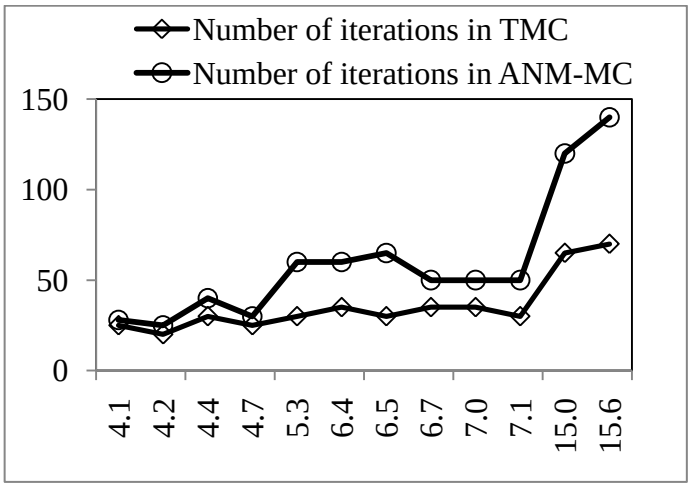


Figure 4.9. Iteration number versus initial RMSD for the proteins studied

In ANM-MC simulations, proteins approached the target structure with RMSD values varying between 1.4 Å-3.9 Å and 1.8 Å-10.5 Å in the forward and reverse runs, respectively. The RMSD values between open and closed structures fall between 4.1 Å and 15.6 Å. In general, the final structures go closer to the target in forward ANM-MC than reverse ANM-MC simulations.

The dipeptide binding protein, d-ribose binding protein, glutamine binding protein, HSTR and LAO binding protein, which have similar hinge motion between two "alpha & beta" domains, exhibit the most successful approaches with RMSD values varying between 1.4 Å - 1.9 Å and 1.8 Å -3.5 Å in the forward and reverse ANM-MC runs, respectively. The initial RMSD values from the target state fall between 4.1 Å and 6.7 Å for the relevant proteins.

The simulation was modified to be applicable for CaM, which has a high RMSD value of 15.0 Å. The final structure can approach the target structure by an RMSD of 2.9 Å. On the other hand reverse ANM-MC run can only progress the target state with an RMSD of 3.7 Å. This may be caused by the fact that both structures considered in the simulation are bound structures and the overlap values are low at 0.41 and 0.32 in the forward and reverse ANM-MC run, respectively.

AK, helicase and lactoferrin both have 3 domains and "alpha & beta" structures. The RMSD values from the target states are 7.1 Å, 4.4 Å and 6.4 Å, respectively. The target states are approached with RMSD values of 2.4, 1.7 Å and 2.2 Å in the forward ANM-MC runs and 4.7 Å, 2.3 Å and 3.1 Å in the reverse ANM-MC runs for AK, helicase and lactoferrin, respectively. In the reverse ANM-MC case of AK, the final state can only approach the target by 4.7 Å, which is one of the unsuccessful cases.

Dimeric antiviral protein cannot go further than 2.8 Å and 2.9 Å towards the target state in the forward and reverse ANM-MC simulations, respectively. The initial RMSD between the open and the closed structures is 7.0 Å. The reason for the high final RMSD values obtained in ANM-MC simulations result from the all-beta domain-swapped structure.

In TMC simulations of small G-protein Arf6, the target state is approached with an RMSD value of 0.3 Å, whereas the initial structure cannot approach the target state more than 3.1 Å in ANM-MC runs due to the shear motion mechanism.

The simulation is also applied to Diphtheria toxin (DT) which has a missing region between 188^{th}-199^{th} residues and an initial RMSD of 15.6 Å. The protein is assumed to be a monomer in the first simulation by completing the missing residue coordinates with the next coordinates. In the second simulation, the protein is separated at the missing residue region as two chains and given to the simulation as a dimer. Both the monomer and the dimer structures reach the target state with an RMSD of 0.4 Å in TMC runs. In ANM-MC forward runs, monomer and dimer protein approach the target state with RMSD values of 3.7 Å and 3.9 Å, respectively. On the other hand, RMSD values in reverse ANM-MC runs do not progress further than 10.5 Å, respectively.

Table 4.1. Final RMSD values from the target structure for the proteins studied (r denotes the reverse runs for both TMC and ANMMC)

Name	Initial RMSD (Å)	Motion Type	Number of chains	Number of residues	Final RMSD (Å)			
					TMC100	TMC100r	ANMMC100	ANMMC100r
Adenylate kinase	7.1	Hinge bending	1	214	0.1	0.3	2.4	4.7
Antiviral protein (cyanovirin-N)	7.0	Hinge bending	2	202	0.4	0.4	2.9	2.8
Calmodulin	15.0	Hinge bending	1	143	0.3	0.3	2.9	3.7
Dipeptide binding protein	6.5	Hinge bending	1	507	0.2	0.1	1.9	2.9
Diphtheria toxin (1)	15.6	Hinge bending	1	535	0.4	0.4	3.7	10.5
Diphtheria toxin (2)	15.6	Hinge bending	2	535	0.4	0.4	3.9	10.5
D-Ribose binding protein	4.1	Hinge bending	1	271	0.3	0.3	1.4	1.8
Glutamine binding protein	5.3	Hinge bending	1	220	0.4	0.4	1.9	2.6
Helicase	4.4	Hinge bending	1	435	0.3	0.4	1.7	2.3
Human Serum Transferrin	6.7	Hinge bending	1	328	0.1	0.1	1.9	3.5
Lactoferrin	6.4	Hinge bending	1	690	0.1	0.4	2.2	3.1
LAO binding protein	4.7	Hinge bending	1	238	0.1	0.3	1.6	3.0
Small G-protein Arf6	4.2	Shear motion	1	164	0.3	0.3	3.1	3.4

4.3. Energy Profile

The total potential energy values have a decreasing trend varying in a similar range in TMC and ANM-MC runs as shown in Figures 4.10-4.21. This shows that the energy minimization is applied successfully in MC algorithm during the iterations. On the other

hand, the energy values begin to increase and have fluctuations after the plateau region. This shows that the energy is not minimized any more once the final RMSD value reach a plateau and stabilize. In TMC cases of diphtheria toxin, total potential energy reach up to a maximum value around snapshot 30 due to the missing residues.

4.4. Overlap and Selected Modes

The initial overlap value between the mode selected in the first MCS and the target direction together with the most selected modes in the first 10 iterations (given in parenthesis) are given in Table 4.2. and Figures 4.10-21. The mode existing only once in the first 10 iterations are not taken into consideration. The overall results show that the slowest five modes are the most preferred modes during the simulations, especially at the beginning of the iterations. The modes have high fluctuations when the plateau is reached. The overlap values have a decreasing trend during the simulation except the calmodulin case, which exhibits a slight increase followed by a decrease. There is a direct proportion between the overlap values and the degree of approach to the target state. In the cases that do not progress towards the target state, low overlap values are generated. When the final RMSD value reaches a plateau, the overlap falls below 0.3.

Dipeptide-binding, d-ribose binding, glutamine binding proteins and HSTR show high initial overlap values between 0.68-0.90 in the forward ANM-MC runs, thereby generate the most successful results. Only the first two slowest modes (Mode 1 and Mode 2) are selected in these cases. When compared to Zheng and Brook's study (2006), the overlap value (0.80) of d-ribose binding protein is lower than that (0.93) in the relevant study. The overlap (0.68) of glutamine binding protein are also not consistent with Zheng and Brooks' study (2006), which generated an overlap value is 0.87. Zheng and Brook built an elastic network model by using a harmonic potential with a single force constant to account for pair wise interactions between all Ca atoms within a cutoff distance of 10 Å. However, the cutoff distance is taken as 18 Å in this thesis. The difference in overlap values is caused by different cutoff distances in both studies. On the other hand, forward ANM-MC runs of HSTR and LAO binding protein generate similar overlap values (0.9 and 0.81, respectively) with those (0.86 and 0.89) in the study of Zheng and Brooks (2006).

For dimeric antiviral protein, Mode 1 is selected during the first 10 snapshots in both forward and reverse ANM-MC runs. However, the selected mode varies afterwards and the overlap values decrease at a relatively higher rate. In this period, the selected modes in the reverse run show more fluctuations than the forward ANM-MC run.

In CaM simulations, Mode 3 is the most selected mode during forward ANM-MC simulation in which the initial overlap is 0.41. During the iterations, the overlap slightly increases till 0.56 and then decreases at an average of 0.50. The increase derives the approach to the target state starting from an RMSD of 15.0 Å. The overlap value is about 0.32 and there is a fluctuation in the selected modes varying between Mode 2, 5 and 6 in the reverse ANM-MC run. The low overlap values and the variation in selected modes show the final RMSD value to the target do not progress any more. The initial overlap values generated in the forward and reverse runs are similar to the given results in Tama and Sanejouand's study (2001) which are 0.50 and 0.37, respectively. Tama and Sanejouand used a cutoff of 0.8 Å in normal mode calculations.

The most selected mode in the forward ANM-MC of AK is the first and second slowest modes with an initial overlap value of 0.74. On the other hand, low overlap value (0.56) shows the low degree of approach in the reverse case and the first mode is the most selected one. In helicase runs, Mode 2 is the most selected mode during the first 10 iterations of forward ANM-MC. The initial overlap (0.78) is lower than the overlap value (0.96) obtained in Zheng and Brook's study (2006). On the other hand, Mode 1, 4 and 5 are selected most in the reverse ANM-MC of lactoferrin. The overlap value generated in the forward and reverse ANM-MC runs are 0.52 Å and 0.63 Å, respectively. The overlap result are consistent with the result (0.60) in Tama and Sanejouand's study (2001).

In ANM-MC runs of small G-protein Arf6, the overlap values are within 0.27, which shows that the initial structure does not undergo a considerable conformational transition. Moreover, there is a fluctuation in the modes selected during the iterations.

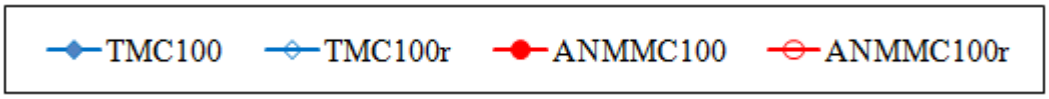
Mode 2 is the most selected mode in the forward runs of diphtheria toxin, which has a missing residue region. The overlap values obtained in the forward and reverse ANM-MC simulations (0.58-0.32) and are similar to those (0.58 and 0.37 respectively) in

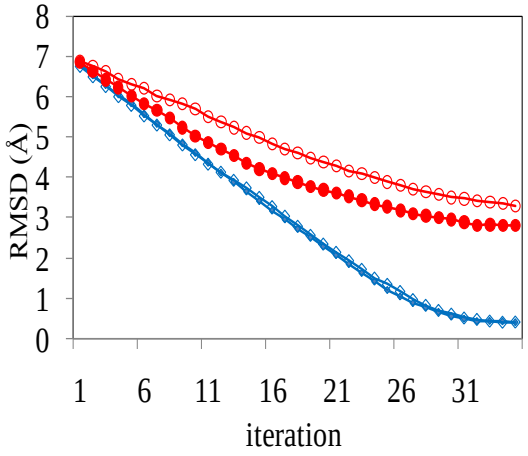
the study of Tama and Sanejouand (2001). The RMSD from the target starts from 15.6 Å and reaches down to a plateau of 10.5 Å, which is supported by the low overlap values in the reverse ANM-MC runs.

Table 4.2. Overlap values and the most selected modes (in parenthesis) for the proteins studied (r denotes the reverse runs for both TMC and ANMMC)

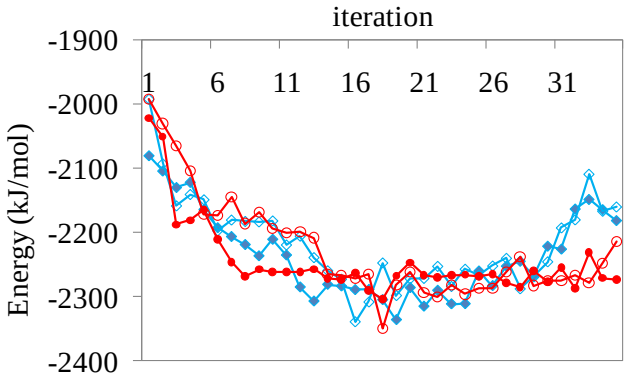
Name	**Initial RMSD (**Å)	**Motion Type**	**Number of chains**	**Number of residues**	**Overlap (Modes selected)***	
					ANMMC100	**ANMMC100r**
Adenylate kinase	7.1	Hinge bending	1	214	0.74 (1,2)	0.56 (1)
Antiviral protein (cyanovirin-N)	7.0	Hinge bending	2	202	0.84 (1)	0.59 (1,6)
Calmodulin	15.0	Hinge bending	1	143	0.41 (3)	0.32 (2)
Dipeptide binding protein	6.5	Hinge bending	1	507	0.84 (1)	0.70 (1)
Diphtheria toxin (monomeric)	15.6	Hinge bending	1	535	0.58 (2)	0.32 (2, 4, 5)
Diphtheria toxin (dimeric)	15.6	Hinge bending	2	535	0.58 (2)	0.32 (2, 4, 5)
D-Ribose binding protein	4.1	Hinge bending	1	271	0.80 (1, 2)	0.70 (1, 2)
Glutamine binding protein	5.3	Hinge bending	1	220	0.68 (2)	0.55 (2, 3)
Helicase / Hepatit C (HCV Helicase)	4.4	Hinge bending	1	435	0.78 (2)	0.40 (1, 5, 7)
Human Serum Transferrin	6.7	Hinge bending	1	328	0.9 (2)	0.54 (1, 3)
Lactoferrin	6.4	Hinge bending	1	690	0.52 (1, 4, 5)	0.63 (1, 4)
LAO binding protein	4.7	Hinge bending	1	238	0.81 (2)	0.53 (2, 3, 4)
Small G-protein Arf6	4.2	Shear motion	1	164	0.27 (1, 10)	0.22 (2, 5, 7, 9)

* The modes selected in first 10 iterations are considered (modes selected once are ignored)

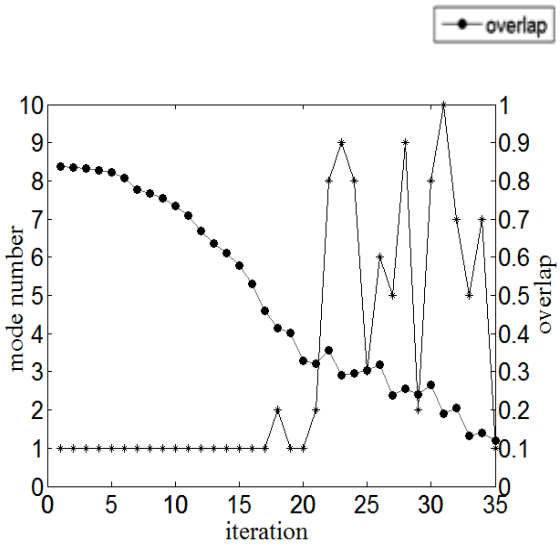




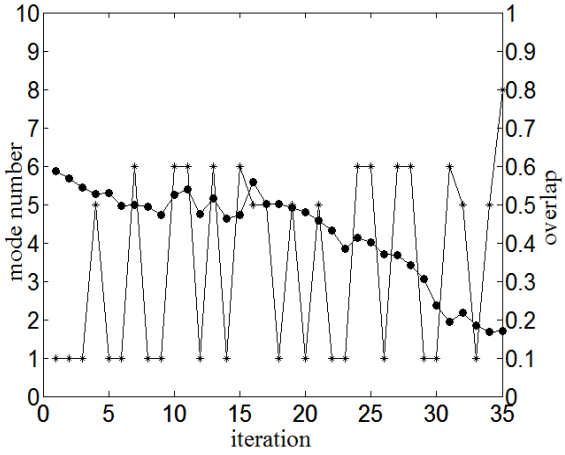
(a) RMSD from target



(b) Total potential energy

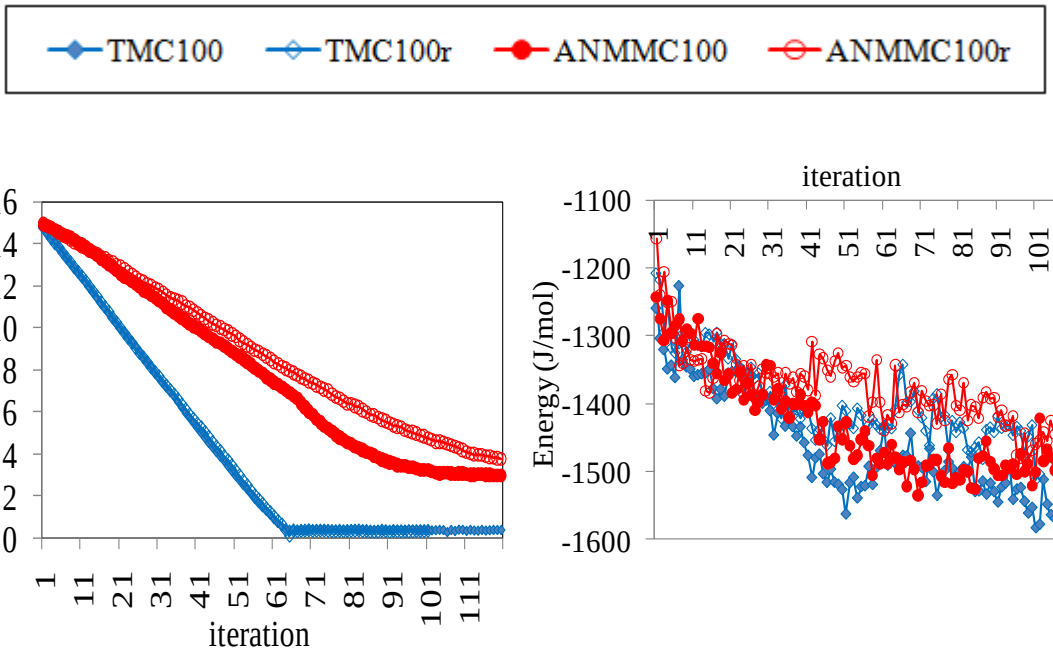


(c) overlap and selected modes in ANM-MC



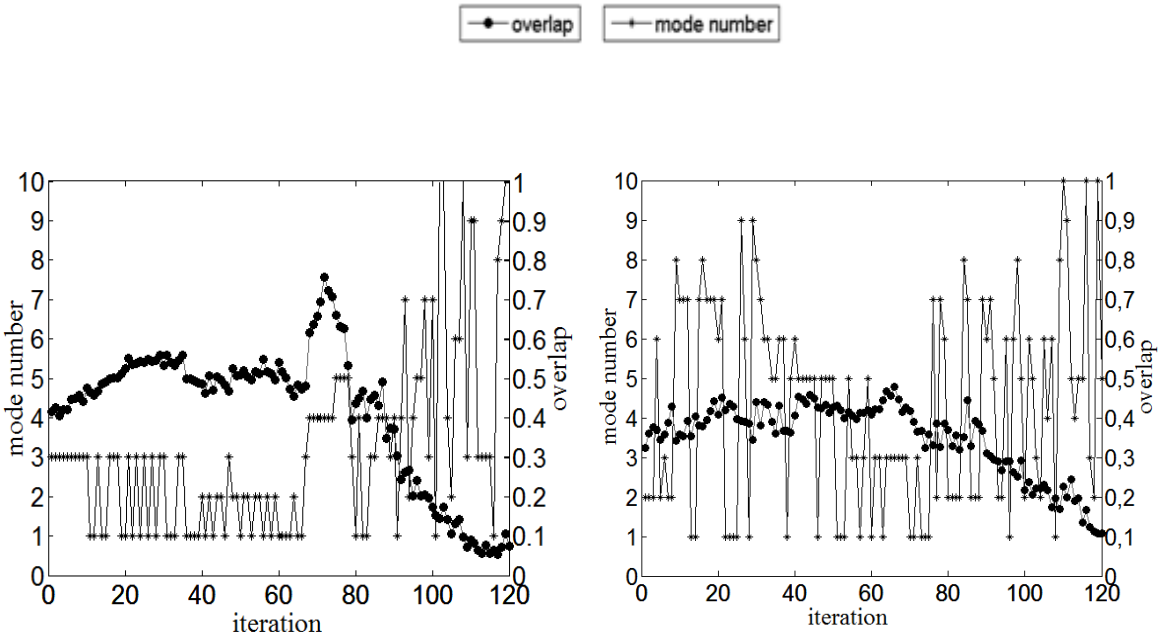
(d) overlap and selected modes in ANM-MCr

Figure 4.10. ANM-MC and TMC results (profiles) as a function of iteration number for Antiviral protein (reverse denoted by "r").



(a) RMSD from target

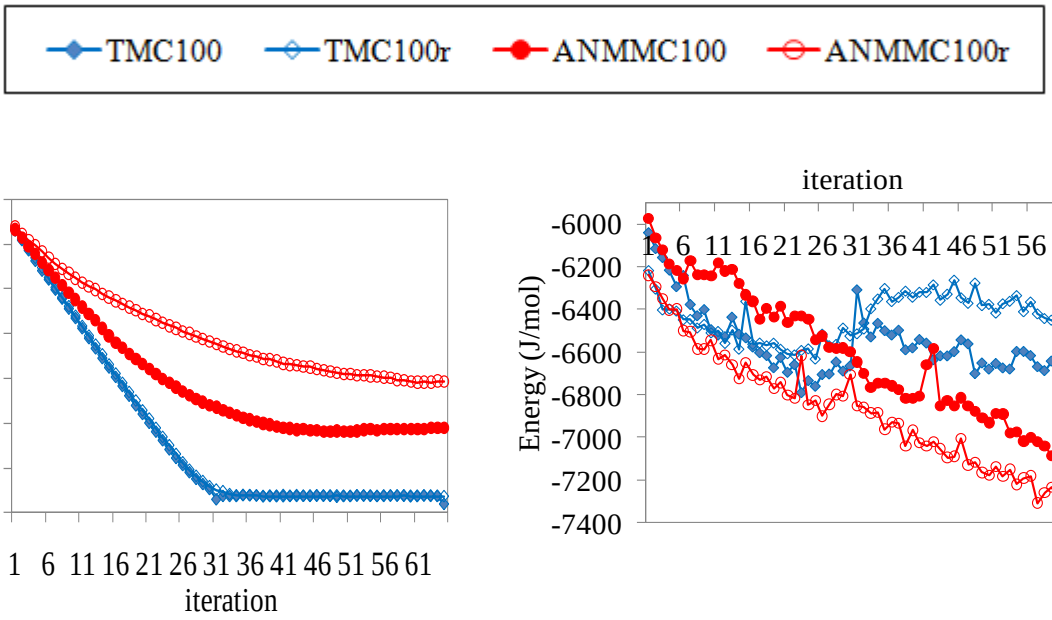
(b) Total potential energy



(c) overlap and selected modes in ANM-MC

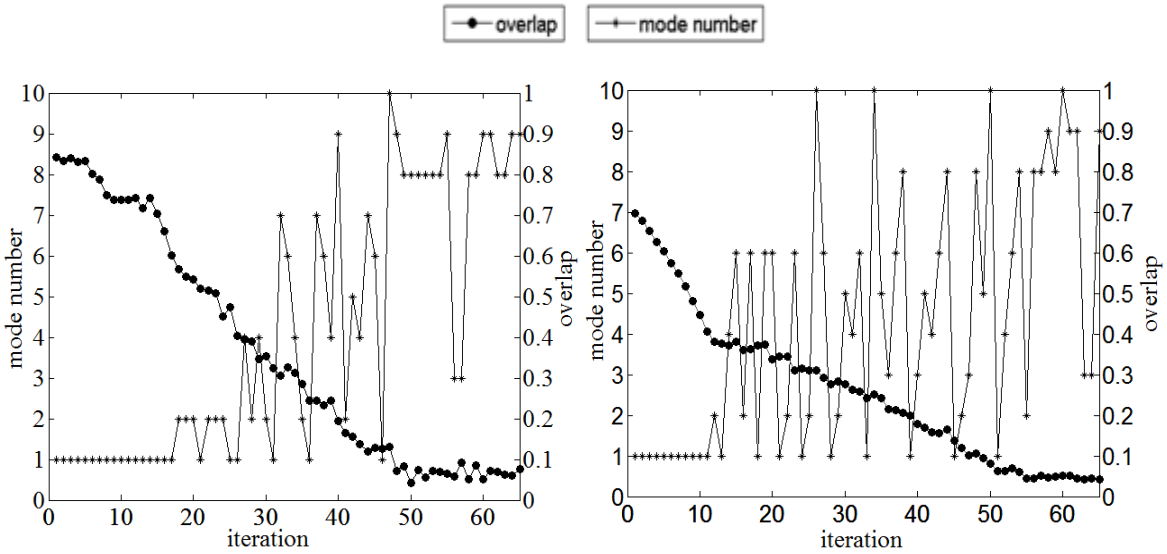
(d) overlap and selected modes in ANM-MCr

Figure 4.11. ANM-MC and TMC results (profiles) as a function of iteration number for Calmodulin protein (reverse denoted by "r").



(a) RMSD from target

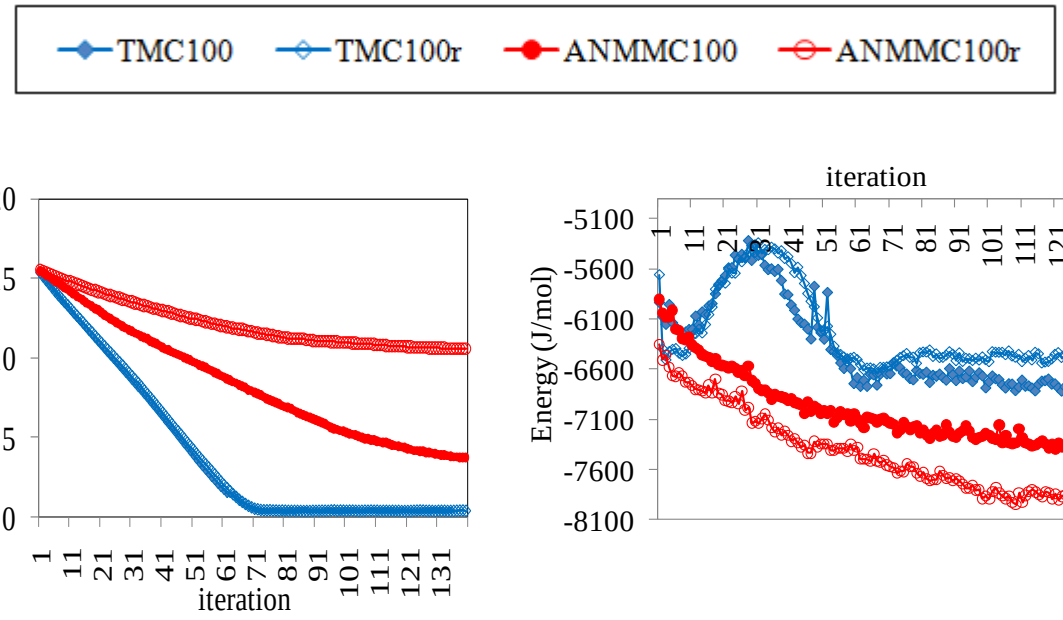
(b) Total potential energy



(c) overlap and selected modes in ANM-MC

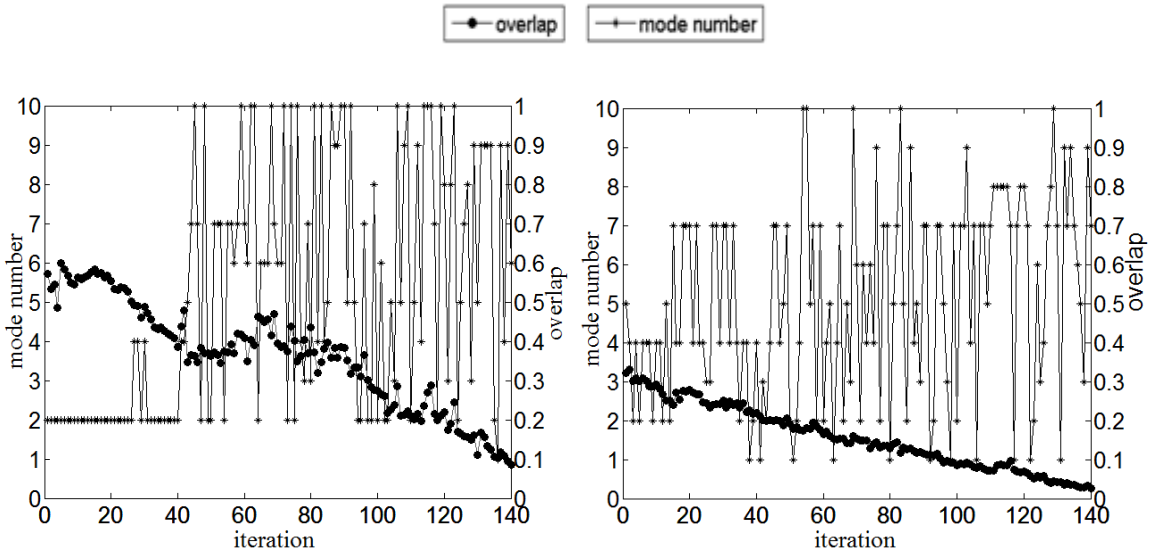
(d) overlap and selected modes in ANM-MCr

Figure 4.12. ANM-MC and TMC results (profiles) as a function of iteration number for dipeptide binding protein (reverse denoted by "r").



(a) RMSD from target

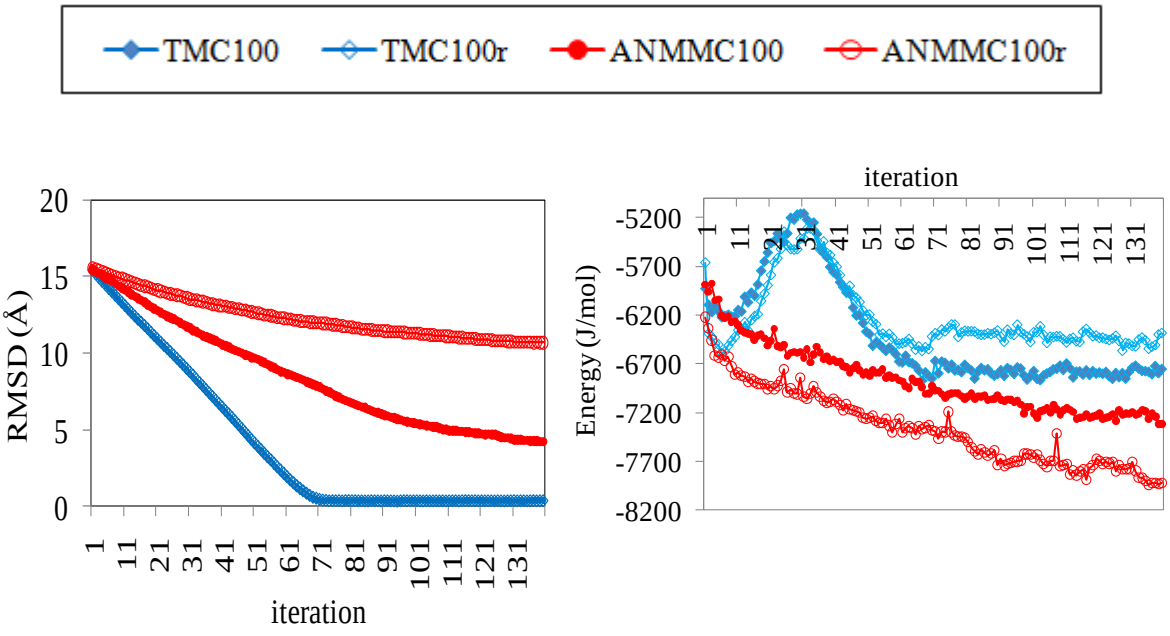
(b) Total potential energy



(c) overlap and selected modes in ANM-MC

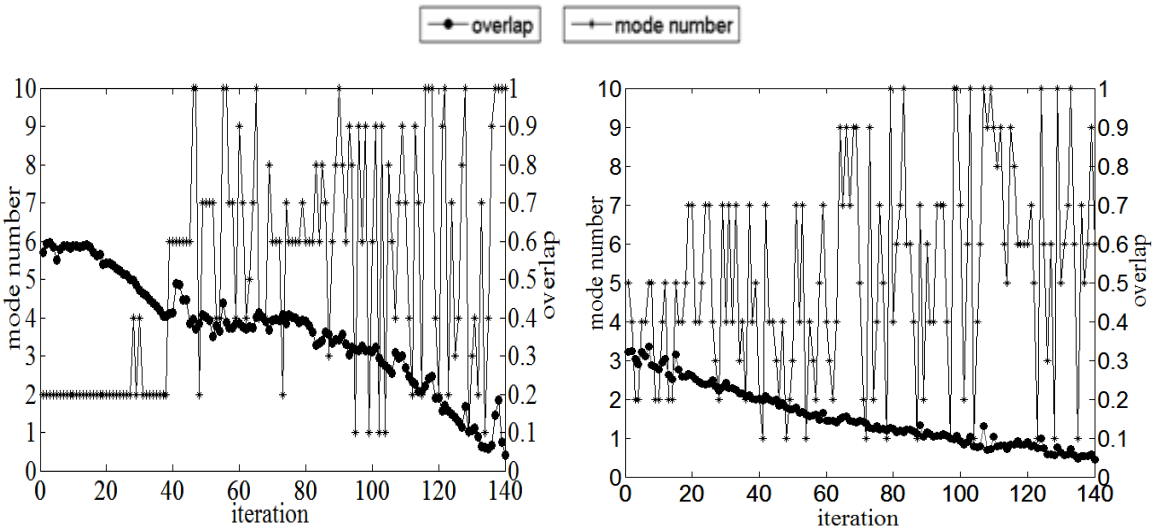
(d) overlap and selected modes in ANM-MCr

Figure 4.13. ANM-MC and TMC results (profiles) as a function of iteration number for monomeric Diphtheria Toxin protein (reverse denoted by "r").



(a) RMSD from target

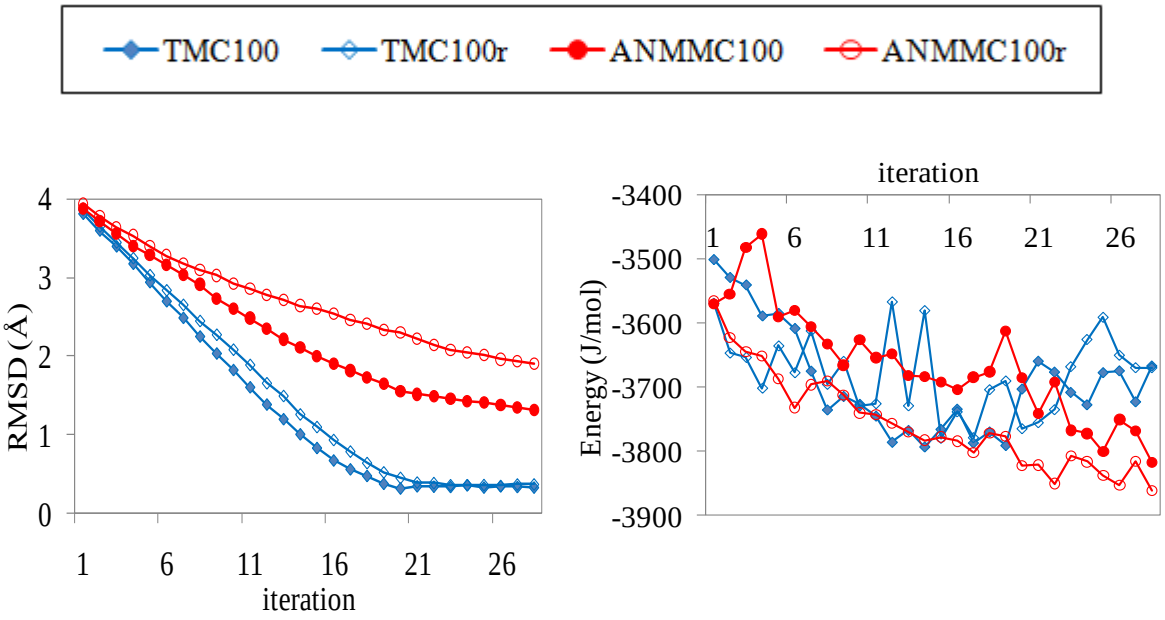
(b) Total potential energy



(c) overlap and selected modes in ANM-MC

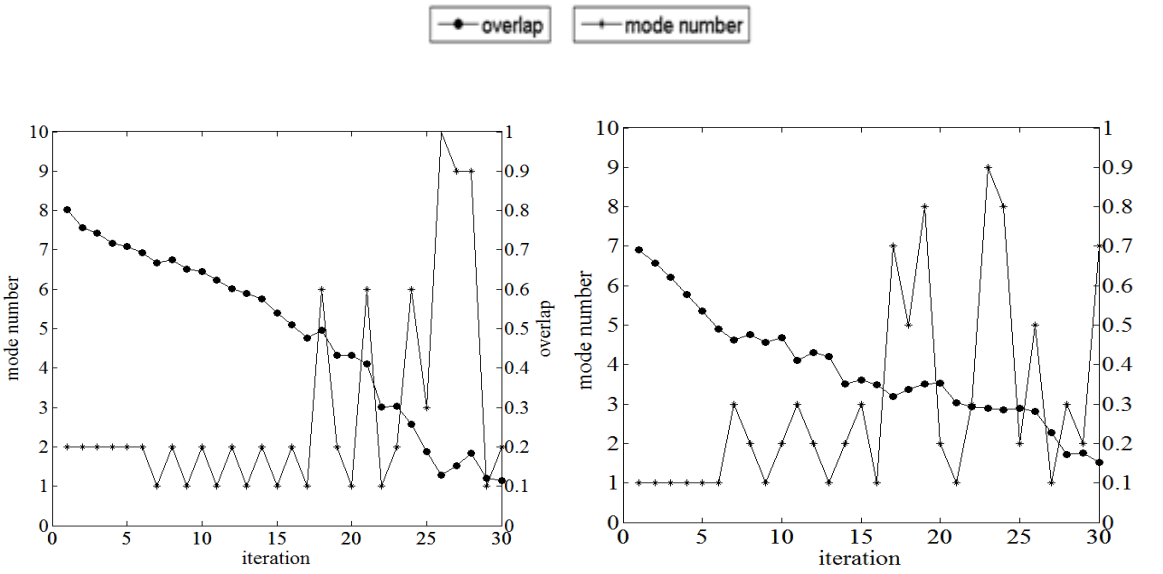
(d) overlap and selected modes in ANM-MCr

Figure 4.14. ANM-MC and TMC results (profiles) as a function of iteration number for dimeric Diphtheria Toxin protein (reverse denoted by "r").



(a) RMSD from target

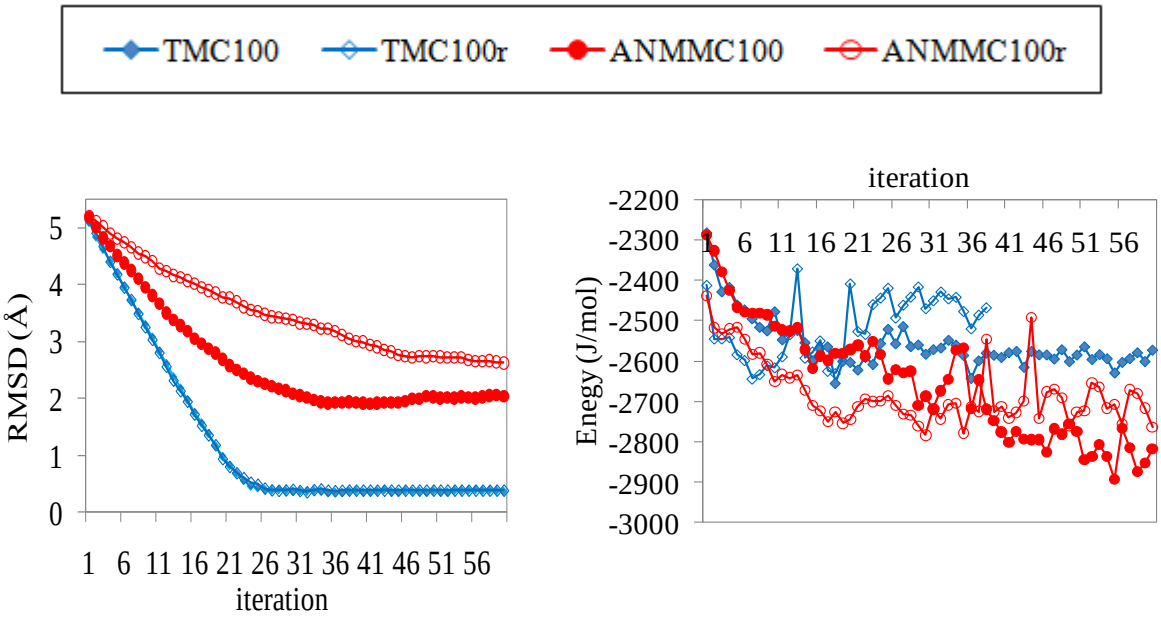
(b) Total potential energy



(c) overlap and selected modes in ANM-MC

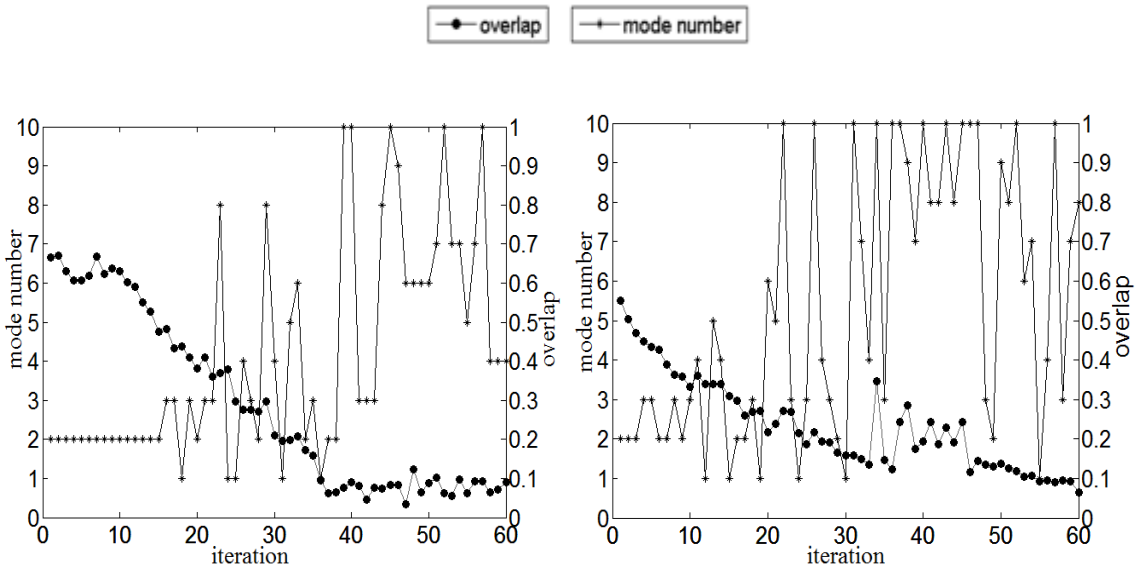
(d) overlap and selected modes in ANM-MCr

Figure 4.15. ANM-MC and TMC results (profiles) as a function of iteration number for d-ribose binding protein (reverse denoted by "r").



(a) RMSD from target

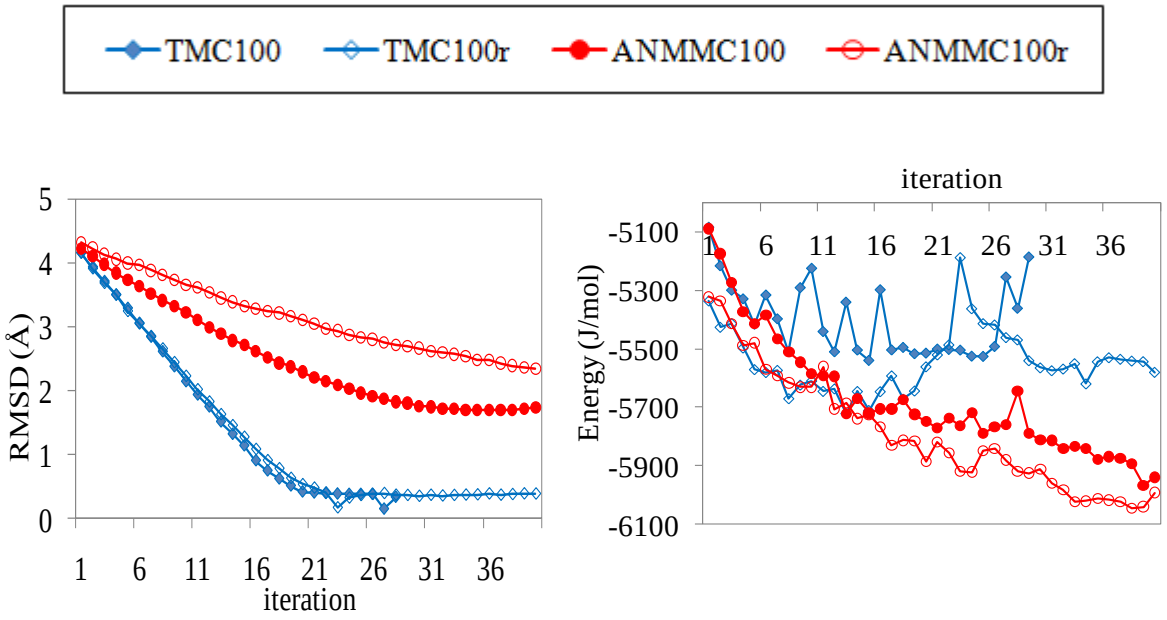
(b) Total potential energy



(c) overlap and selected modes in ANM-MC

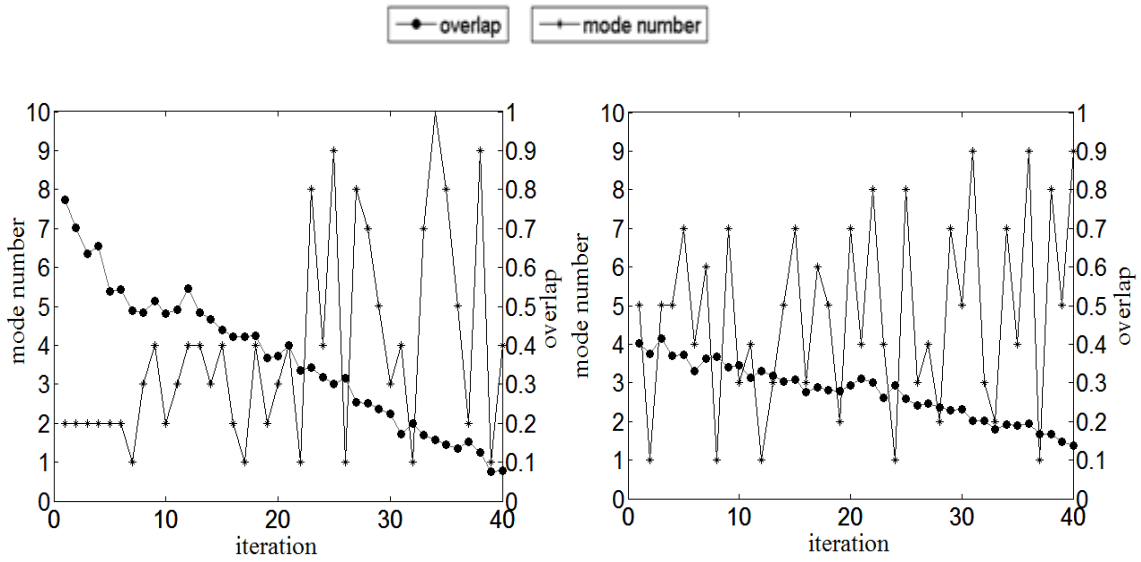
(d) overlap and selected modes in ANM-MCr

Figure 4.16. ANM-MC and TMC results (profiles) as a function of iteration number for glutamine binding protein (reverse denoted by "r").



(a) RMSD from target

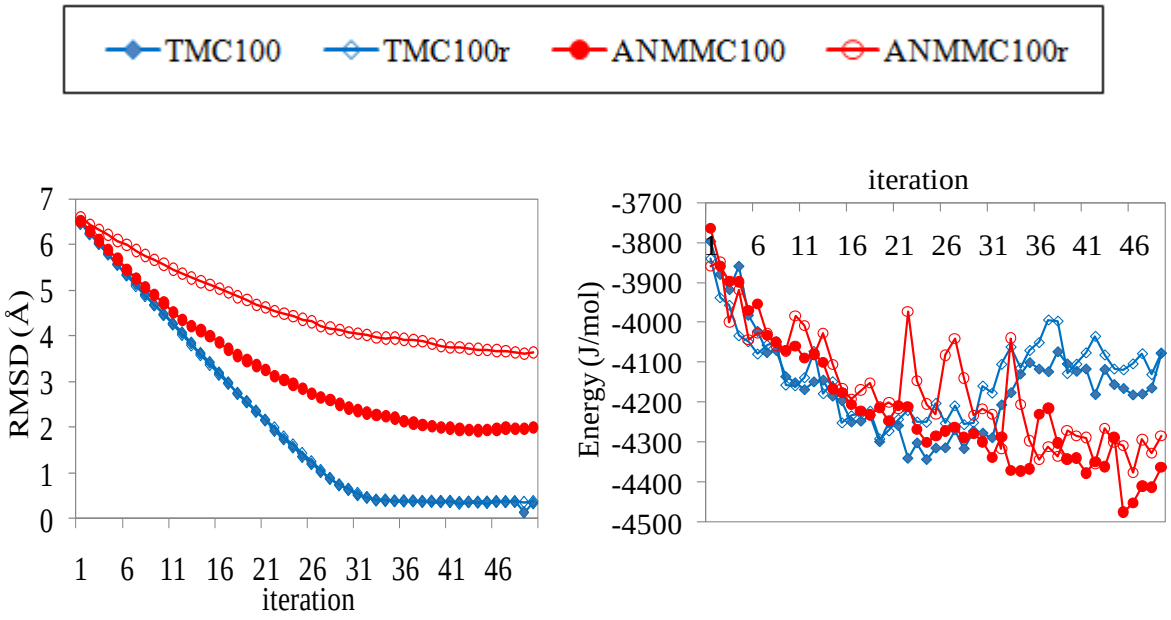
(b) Total potential energy



(c) overlap and selected modes in ANM-MC

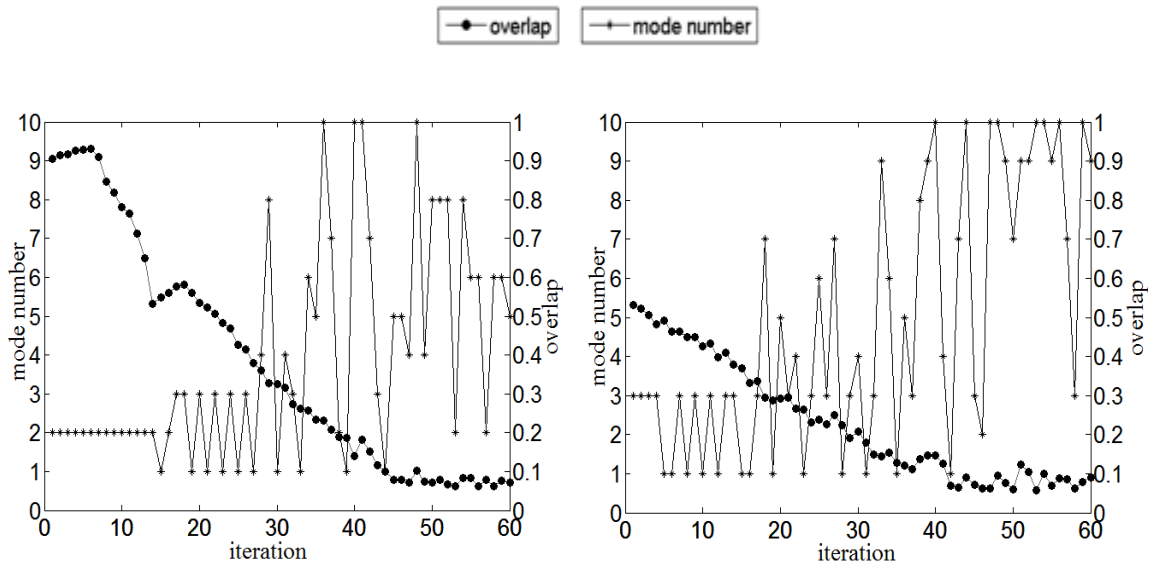
(d) overlap and selected modes in ANM-MCr

Figure 4.17. ANM-MC and TMC results (profiles) as a function of iteration number for helicase (reverse denoted by "r").



(a) RMSD from target

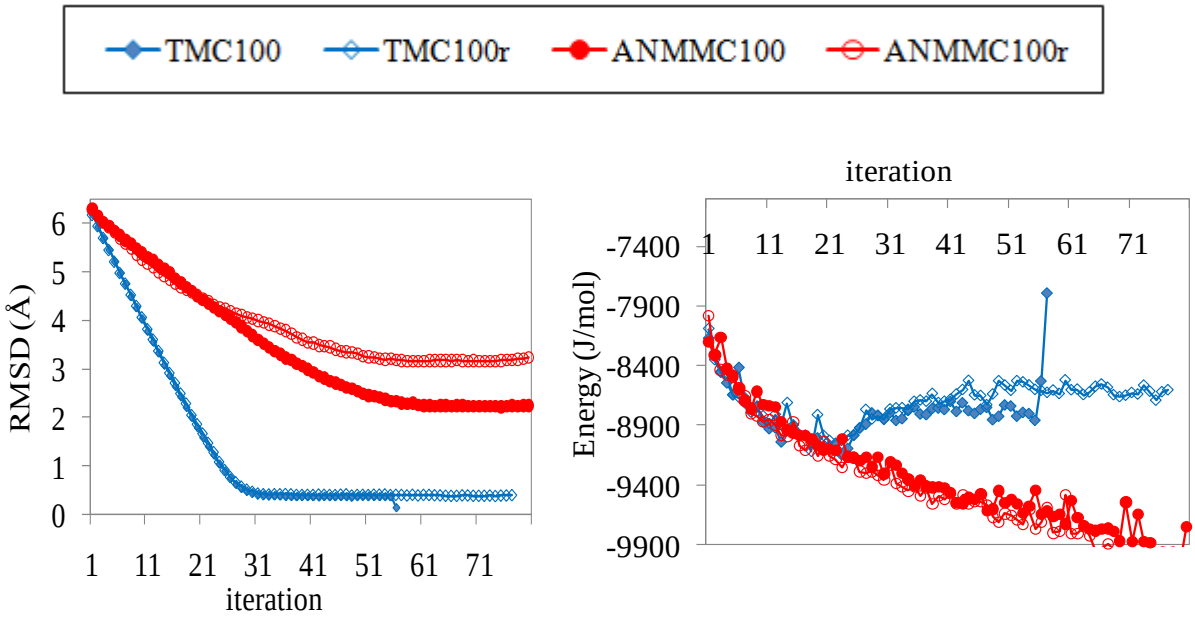
(b) Total potential energy



(c) overlap and selected modes in ANM-MC

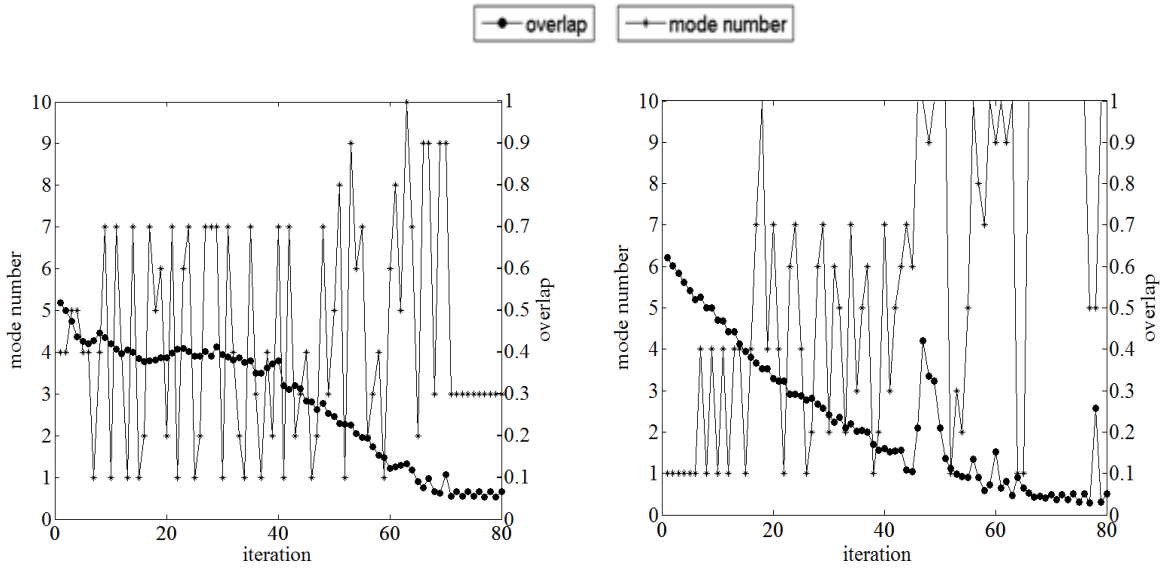
(d) overlap and selected modes in ANM-MCr

Figure 4.18. ANM-MC and TMC results (profiles) as a function of iteration number for HSTR (reverse denoted by "r").



(a) RMSD from target

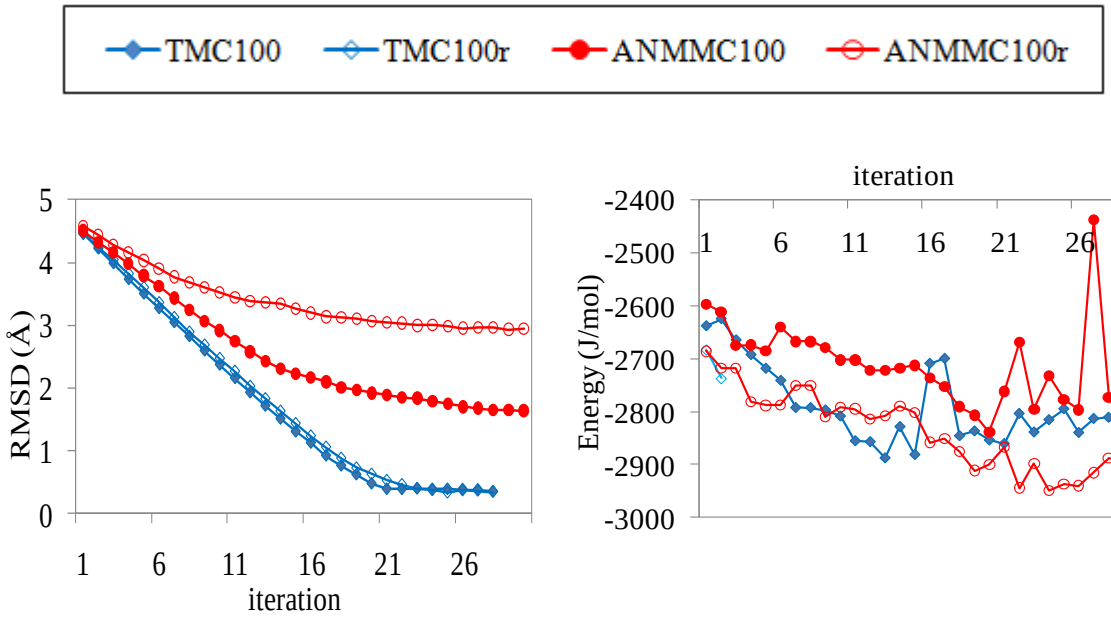
(b) Total potential energy



(c) overlap and selected modes in ANM-MC

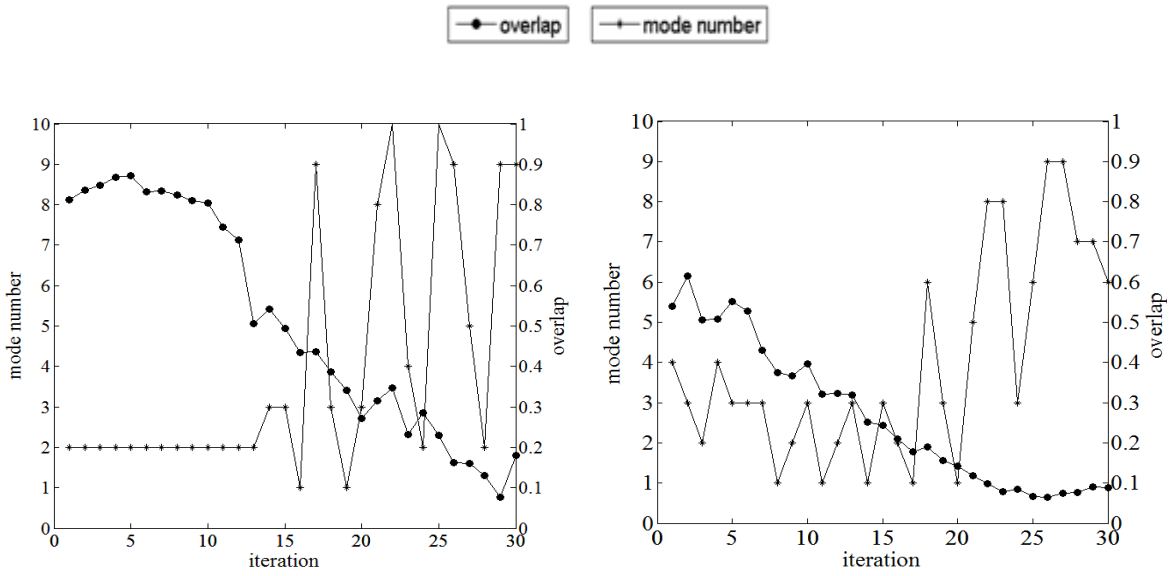
(d) overlap and selected modes in ANM-MCr

Figure 4.19. ANM-MC and TMC results (profiles) as a function of iteration number for lactoferrin (reverse denoted by "r").



(a) RMSD from target

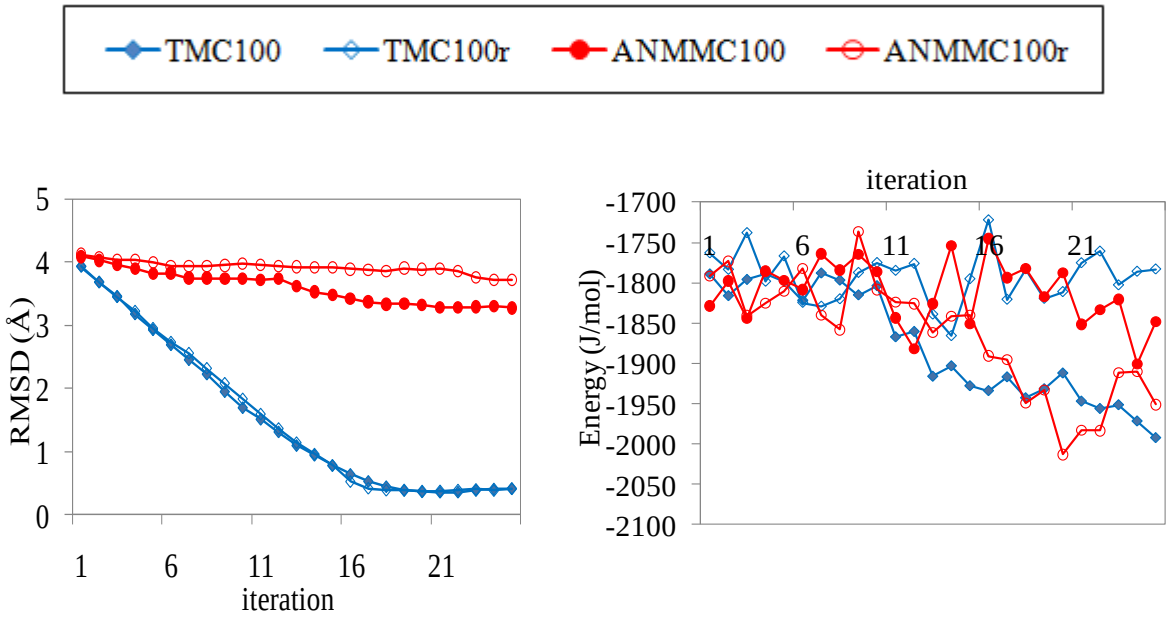
(b) Total potential energy



(c) overlap and selected modes in ANM-MC

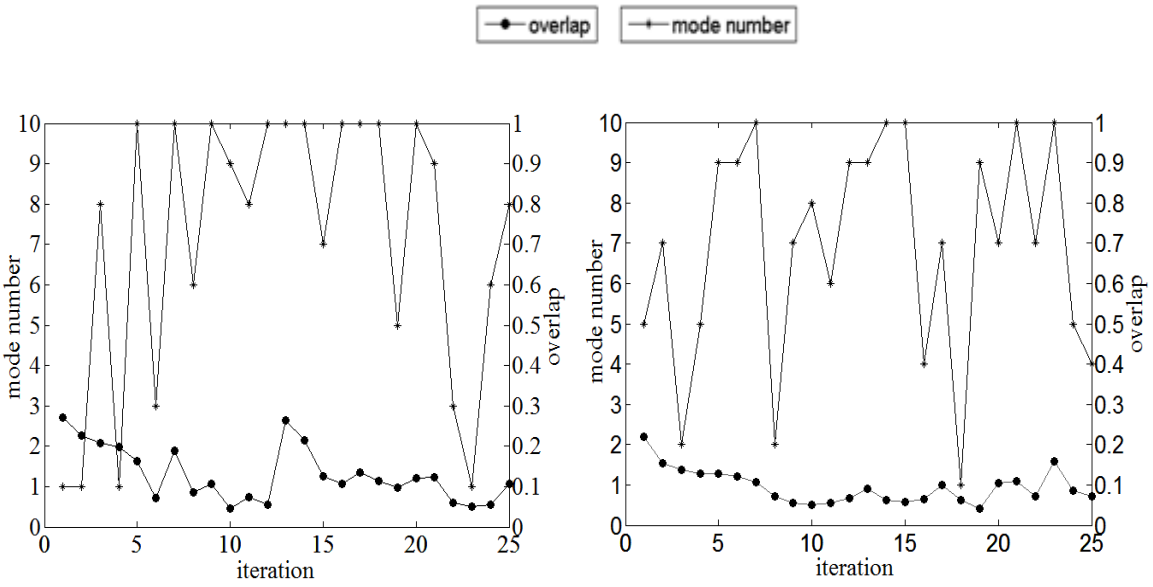
(d) overlap and selected modes in ANM-MCr

Figure 4.20. ANM-MC and TMC results (profiles) as a function of iteration number for LAO binding protein (reverse denoted by "r").



(a) RMSD from target

(b) Total potential energy



(c) overlap and selected modes in ANM-MC

(d) overlap and selected modes in ANM-MCr

Figure 4.21. ANM-MC and TMC results (profiles) as a function of iteration number for small G-protein Arf6 (reverse denoted by "r").

4.5. Transition Pathways

Most of the proteins studied in this thesis undergo hinge bending movement. When the transition pathways and the snapshots are investigated using ANM-MC, it is observed that open-to-close and close-to-open transition are observed through the hinge motion between the two domains. During the transition, several intermediate conformations are observed on the pathway between the open and closed conformational states. The domain movements and transition pathways are illustrated in Figures 4.22-57 and Appendix). Domains are colored according to dynamic domains, which consist of fixed (blue) and moving (red and yellow) domains (Hayward *et al*., 1997; Hayward and Berendsen, 1998).

The pathways between open and closed are characterized by the analysis of the contact maps during the transition. The evolution of the contact maps within a cutoff distance of 15 Å are followed and then marked in red on the target structure's contact map. (Figures 4.23-58). It is clearly seen that the contacts formed in the final conformational state obtained by the simulations are very similar to those of the target state at this resolution. Moreover, the forward and reverse runs, referring to the open-to-close and close-to-open transitions respectively, indicate that the forward and reverse pathways are similar. In both, similar conformational states are visited in the transition. On the other hand, the target structures are reached by TMC simulations, where the target's contact map is fully reproduced.

In this section; calmodulin and lactoferrin will be analyzed in detail including the intermediate snapshots, contact maps and the illustration of the new contacts on the protein structures in forward and reverse TMC and ANM-MC runs. The reason why these proteins are selected or investigated in detail is that calmodulin has a large conformational transition between initial and target states, and lactoferrin has three domains. The snapshots and the contact maps of the other proteins studied in this work are given in Appendix in detail, whereas only the initial, target and final state results are presented in this section.

4.5.1. Calmodulin (CaM)

Calmodulin has a large hinge bending motion with an RMSD of 15.0 Å between both open/closed structures. The open-to-close transition mechanism is studied through the intermediate snapshots generated by both ANM-MC and TMC simulations. Figures 4.22-4.29 displays the initial structure, the target structure and the conformations sampled during the open-to-close and close-to-open transition in ANM-MC simulations. The N-terminal and C-terminal domains are colored in red and blue, respectively. The N-terminal domain is known as inherently more flexible than C-terminal domain (Chattopadhyaya R. *et al.*, 1992, Vertessy B.G. *et al.*, 1998, Kuboniwa *et al.*, 1995). The corresponding contact maps are given in Figures 4.23-4.30. The main contact region on the contact maps are defined at the end helices between the residues 1^{st}-70^{th} and 90^{th}-143^{rd}.

Beginning from the initial RMSD of 15 Å, the final state approach the target state with an RMSD of 2.9 Å and 3.7 Å in the forward and reverse ANM-MC runs, respectively. In the first iteration of forward ANM-MC, initial contacts are formed at the end helices of both C- and N- terminal domains among 57 residues as shown in Figure 4.24. The main contact difference between initial and target states are marked in black (Figure 4.23). The contacts on the main contact region first appear at snapshot 80 and then increase during the iterations. They are shown in blue, orange and green in the formation order. No contacts are formed at residue 143 at the final state (Figure 4.23). On the other hand, in the reverse ANM-MC run, all contacts disappeared in snapshot 80. The missing residue regions are indicated by orange, green and blue in order. Blue region appears first in the forward ANM-MC and disappears last in the reverse ANM-MC, which is an expected case. There is a slight change in the order of contact formation between the forward and reverse ANM-MC runs (Orange and green contacts in Figures 4.23 and Figures 4.26)

In TMC simulations, the RMSD to the target reach a plateau at 0.3 Å. The final state obtained from TMC simulation shows very similar contacts to the target state and prove to generate more contacts when compared to ANM-MC due to the higher degree of approach to the target. The main contacts first appear in snapshot 40 on the red region and afterwards on the blue region followed by yellow, green and purple regions at the same time (Figure 4.28). In the reverse TMC case; the initial contacts disappear in the reverse

order of contact formation of the forward run. They fully disappear at snapshot 30. This shows that the forward and reverse TMC cases follow a similar pathway in terms of contacts (Figure 4.30).

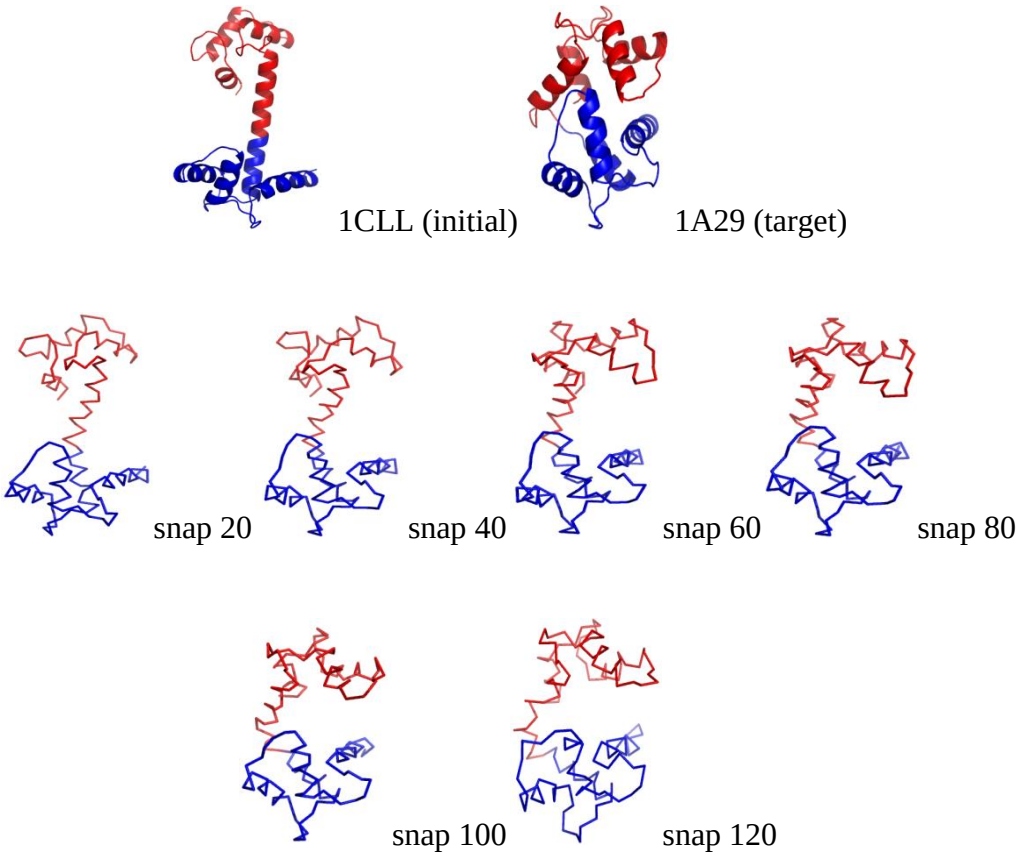


Figure 4.22. Several intermediate structures obtained during forward ANM-MC simulation of Calmodulin transition from initial to target conformation.

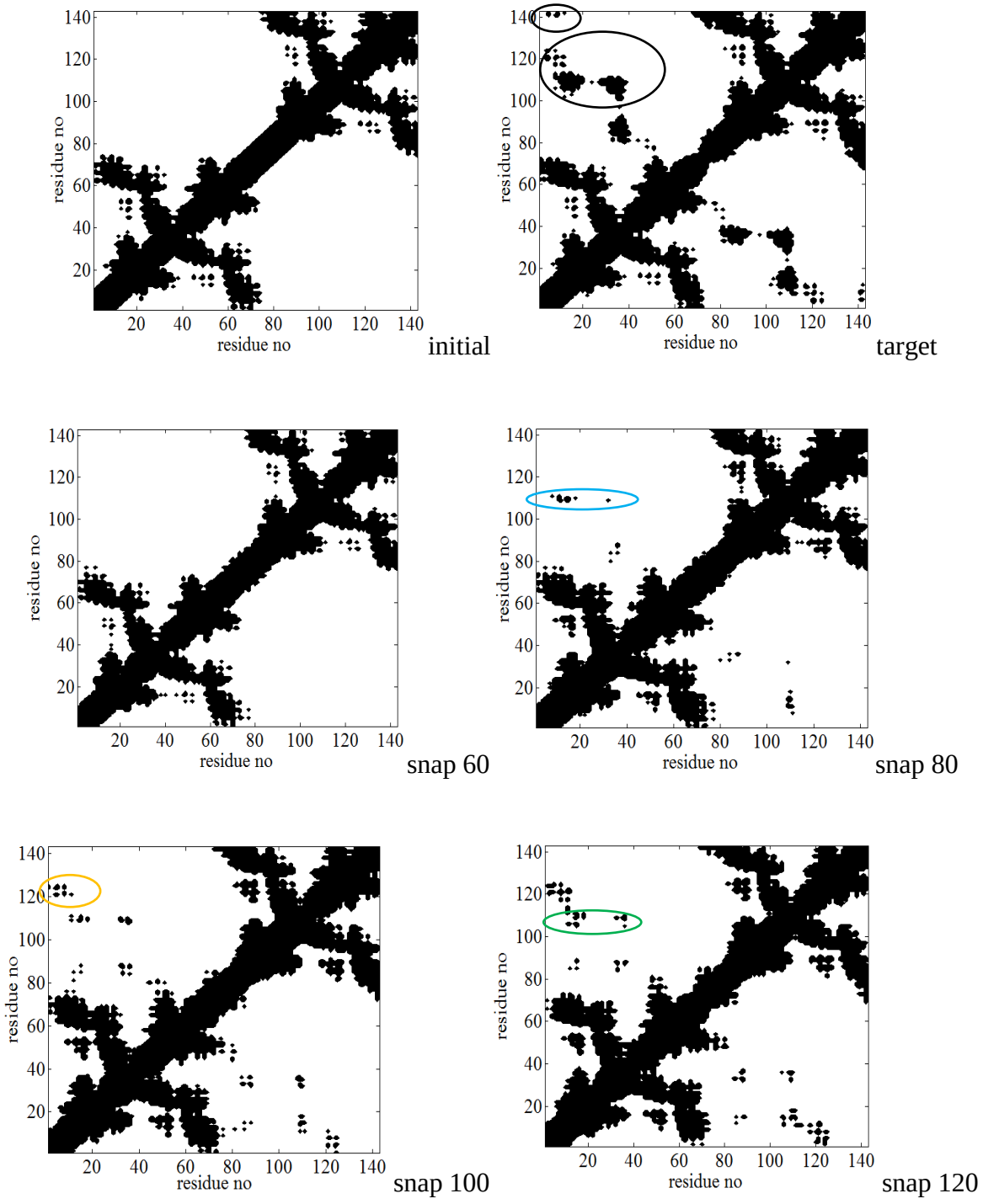


Figure 4.23. Contact maps for open, closed forms of Calmodulin and for snapshots 60-80-100-120 for forward ANM-MC run.

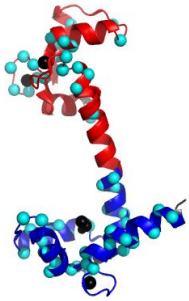


Figure 4.24. Initial contacts (cyan) on the open structure of calmodulin

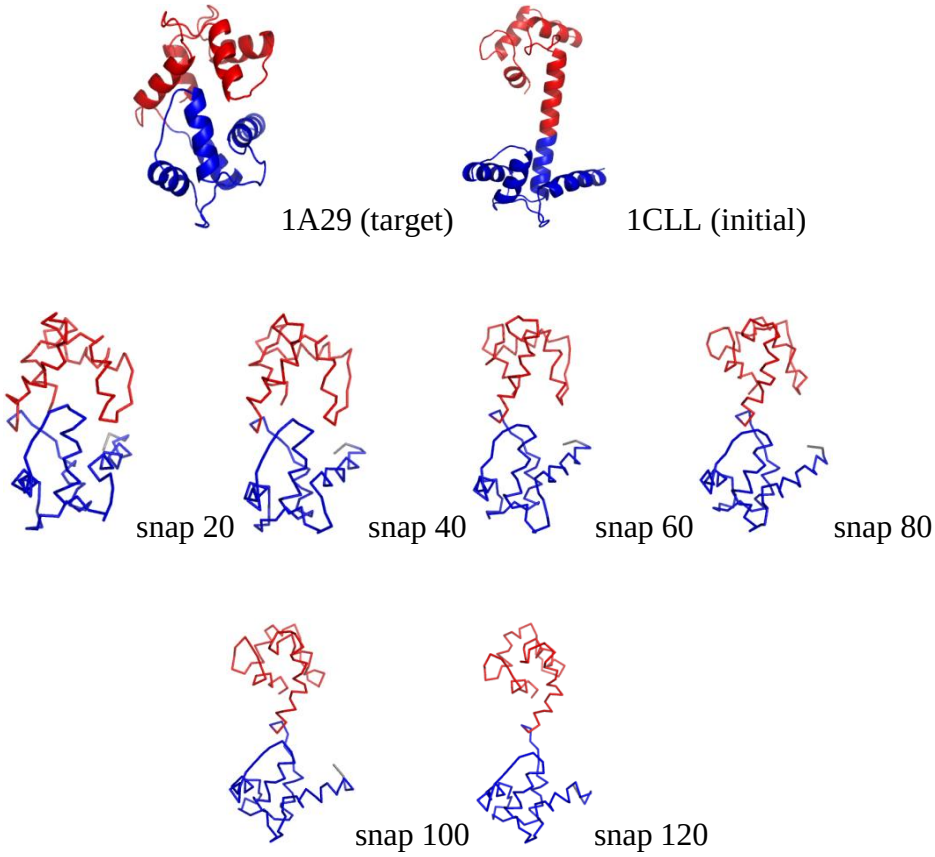


Figure 4.25. Several intermediate structures obtained during reverse ANM-MC simulation of Calmodulin transition from target to initial conformation.

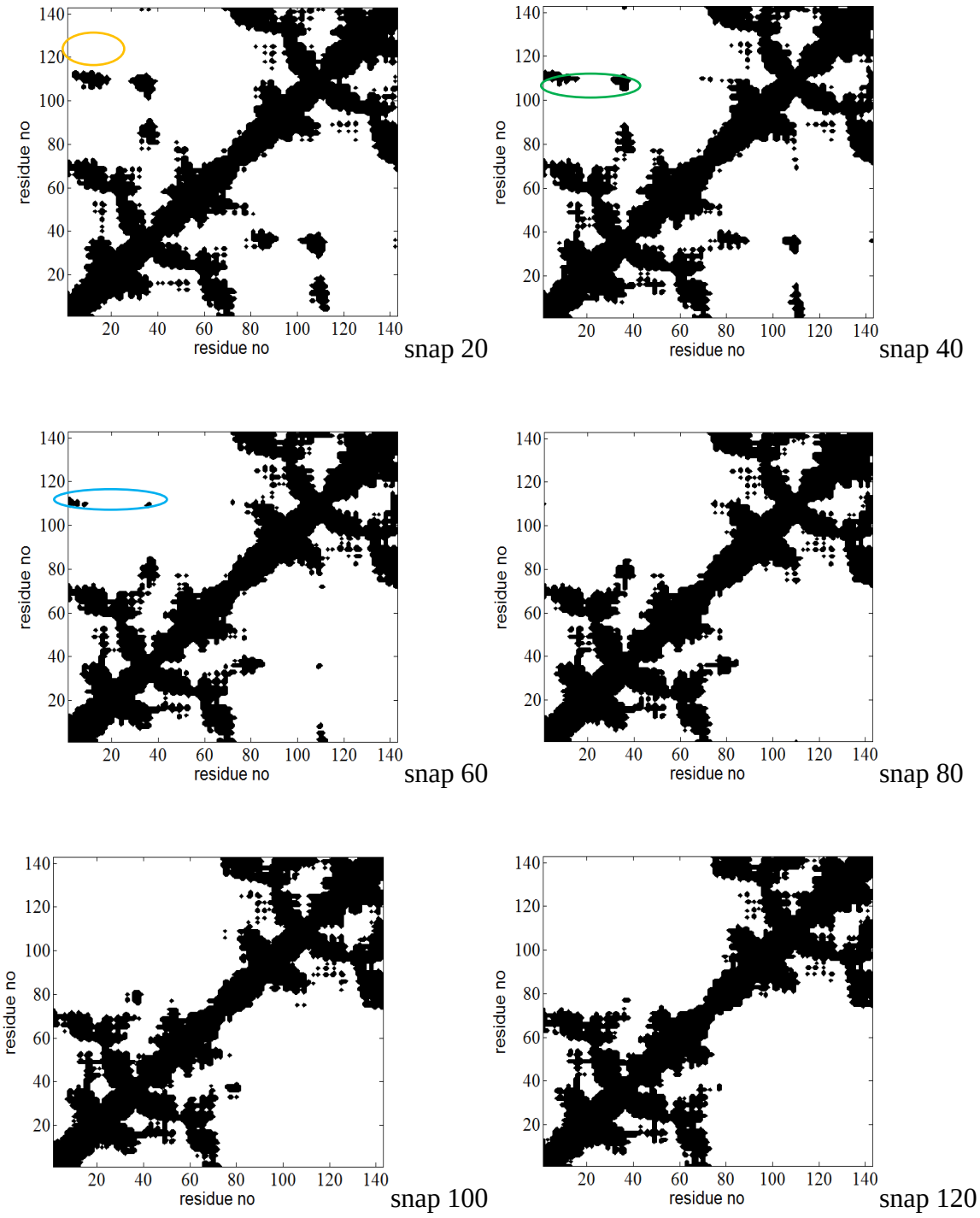


Figure 4.26. Contact maps for open, closed forms of Calmodulin and for snapshots 20-40-60-80-100-120 for forward ANM-MC run.

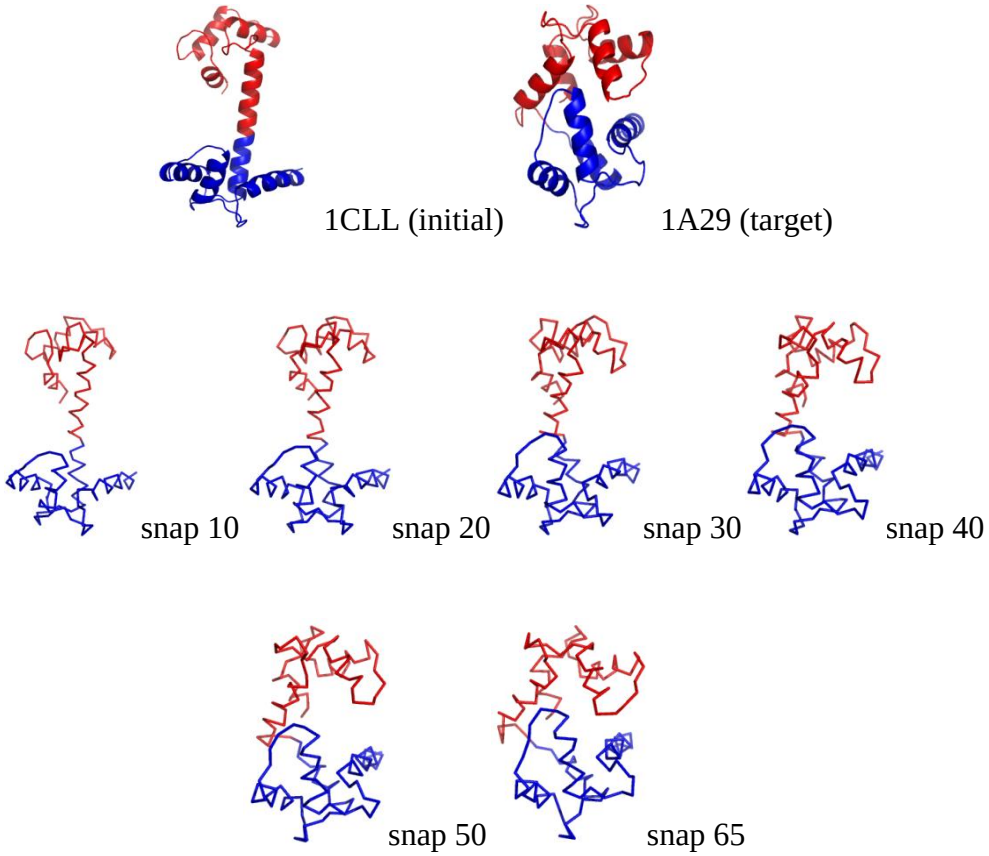


Figure 4.27. Several intermediate structures obtained during forward TMC simulation of Calmodulin transition from initial to target conformation.

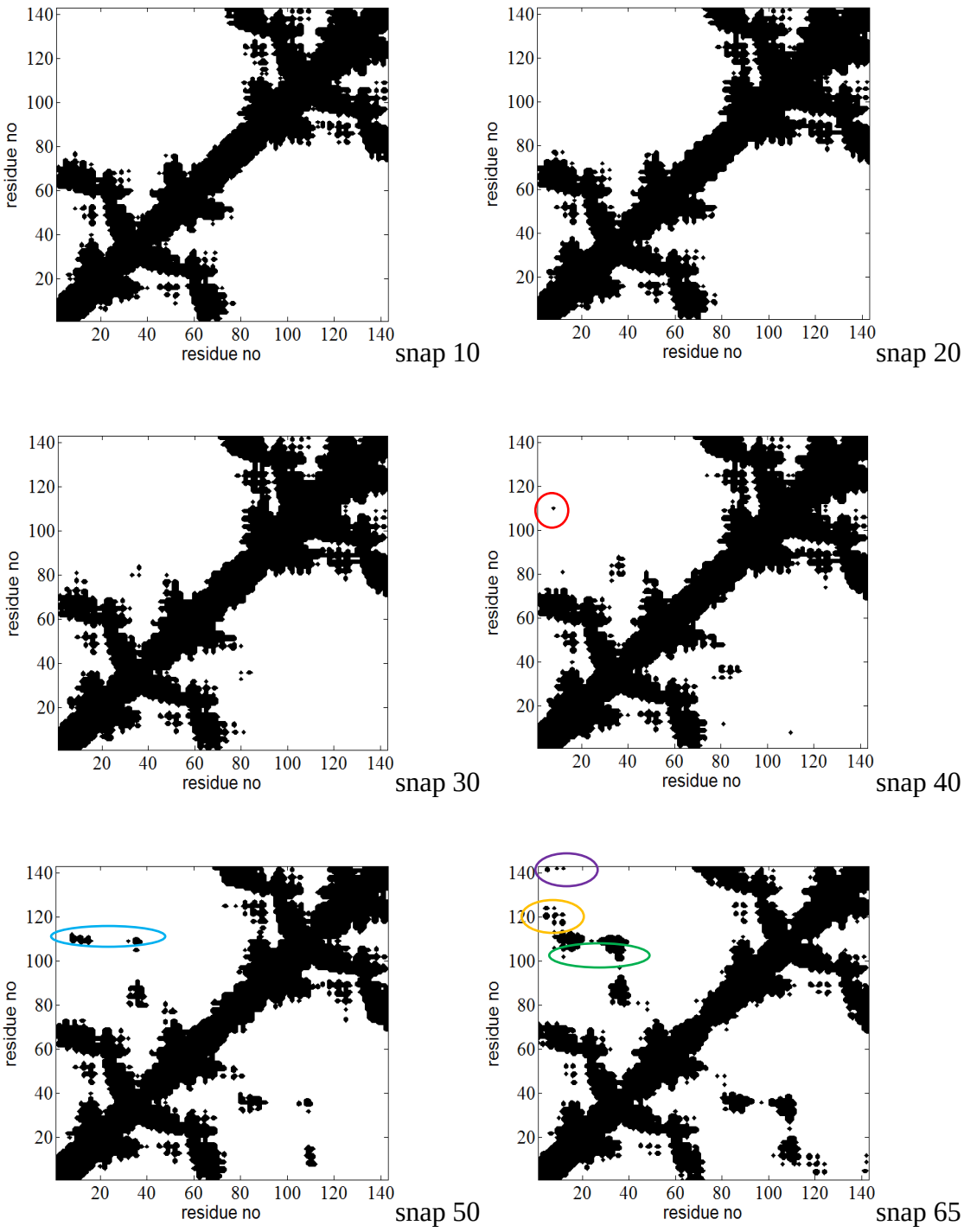


Figure 4.28. Contact maps for open, closed forms of Calmodulin and for snapshots 10-20-30-40-50-65 for forward TMC run.

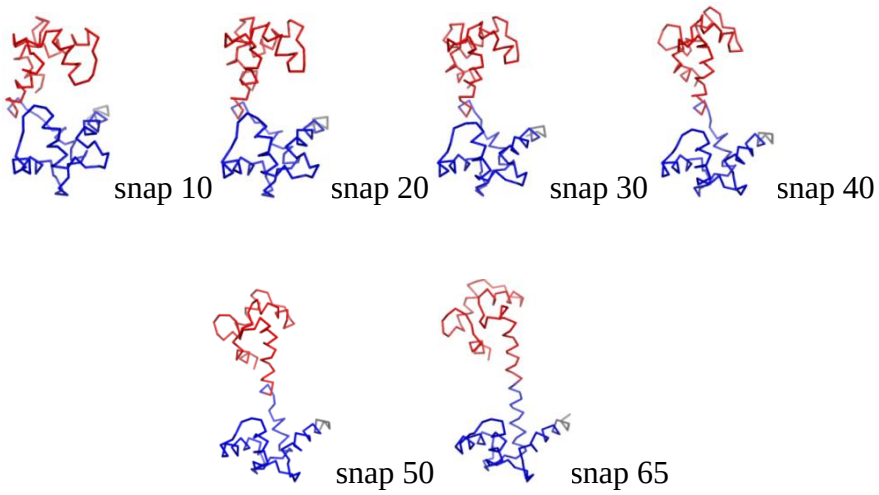


Figure 4.29. Several intermediate structures obtained during reverse TMC simulation of Calmodulin transition from target to initial conformation.

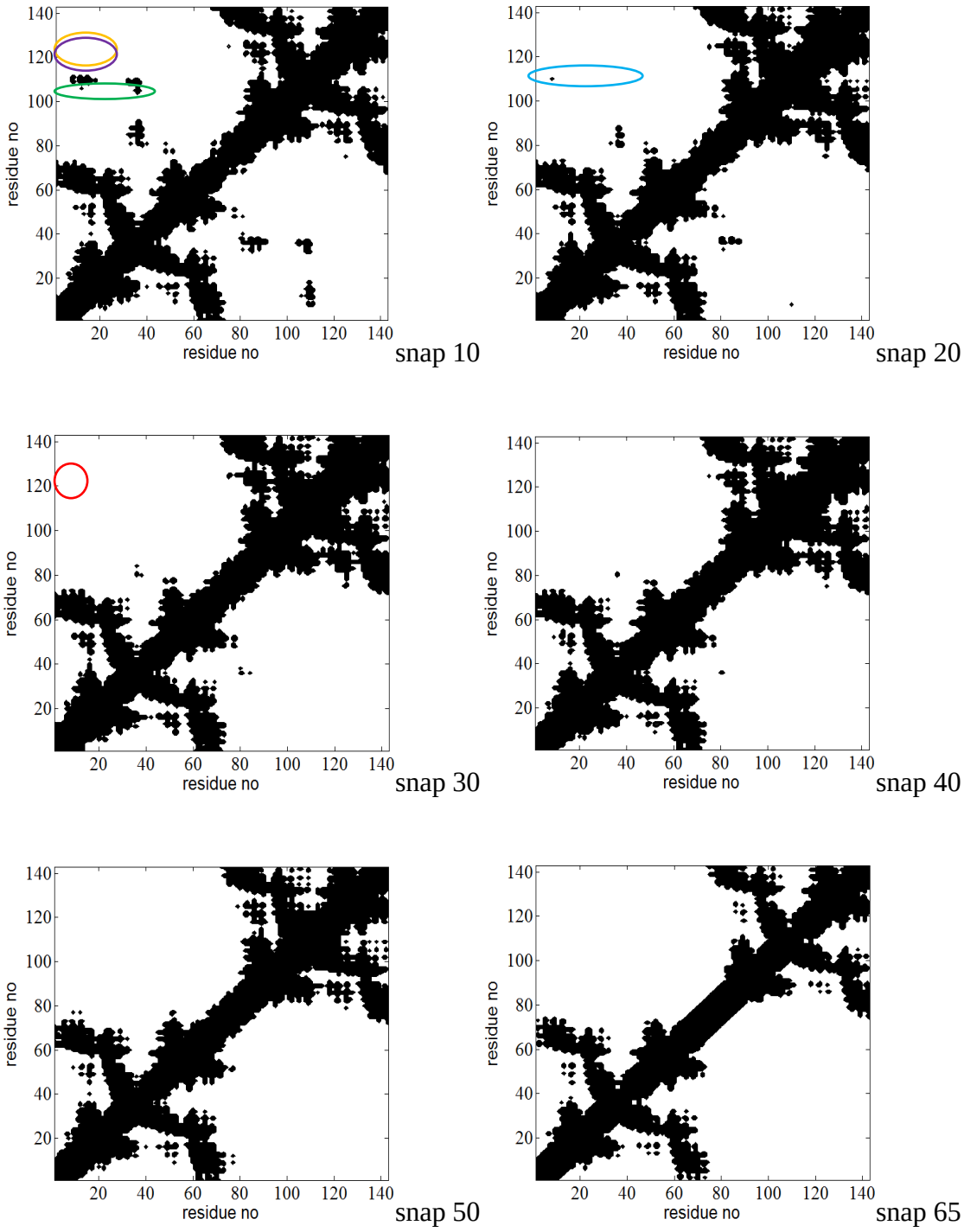


Figure 4.30. Contact maps for open, closed forms of Calmodulin and for snapshots 10-20-30-40-50-65 for reverse TMC run.

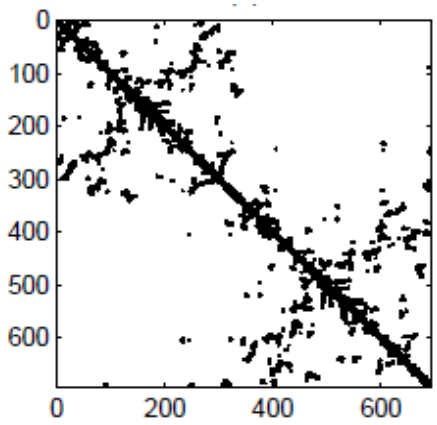
4.5.2. Lactoferrin

Lactoferrin is a large monomer containing 690 residues. Like dipeptide binding protein, the modified program for large proteins was applied to Lactoferrin. It has a conformational change of hinge motion with an RMSD of 6.4 Å between both open/closed structures. The transition mechanism between open and closed structures is studied through the intermediate snapshots generated in both TMC and ANM-MC simulations. The conformations sampled during the transition in the open-to-closed and closed to open directions are shown in Figures 4.32-4.38. The fixed domain is colored in blue and the remaining two moving domains are colored in red and yellow according to Dyndom classification (Hayward *et al.*, 1997; Hayward and Berendsen, 1998). The corresponding contact maps of the intermediate snapshots are given in Figures 4.33-4.39. The main contact region on the contact maps are defined at the hinge region between the residues 10^{th}-61^{st} and 118^{th}-210^{th}. Kim *et al.* (2002b, 2005) also studied the conformational transitions and contacts within a cutoff of 10 Å together with the transitional conformation from closed to open structures of lactoferrin by using coarse-grained ENM methodology. They concluded that some residues have poor connections at cutoff =10 Å, so that the resulting pathway cannot be realistic. Therefore, they recommended a larger cutoff value of 15 Å, which substantially increases the density of the linking matrix. This conclusion is supported by the final contact map of ANM-MC simulation as shown in Figure 4.33, which also indicates poor contacts. The contact matrix in the study of Kim and coworkers contains 2.91% contacts, whereas the final contact map of ANM-MC simulation contains 3.22% contacts. However, less contacts are formed at the end of ANM-MC simulation when compared to those in the study of Kim *et al.* (2002b, 2005) as given in Figure 4.31. The final states of ANM-MC and TMC simulations are shown to be similar to the target state in terms of structure and contacts. The forward and reverse pathways exhibit similar contacts and intermediate structures as shown in Figures 4.33-4.35.

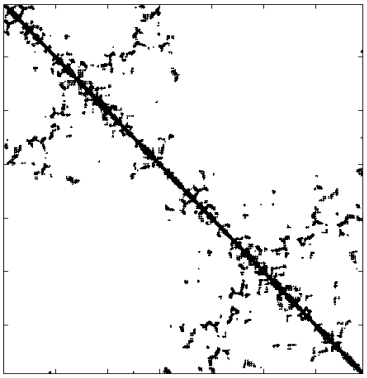
In the forward and reverse ANM-MC simulations, the final state approach the target with an RMSD of 2.2 Å and 3.1 Å, respectively beginning from the initial RMSD of 6.4 Å. The difference between the main contacts of the open and closed structures are marked in black in Figure 4.33. The first contact appear in snapshot 10 between residues 61 and 210 (red region) in ANM-MC simulation. New contacts formed in the following iterations

are shown in red, blue, yellow, green and purple in the formation order. The native contact between residues 10-161 is not formed at the final state in ANM-MC (Figure 4.33). On the other hand, in the reverse ANM-MC run, all main contacts except the contacts between residues 81-132 and 61-210 disappeared in snapshot 60. The missing residue regions are indicated by blue, yellow and green in order. Contacts in the yellow region decrease whereas others disappear at the final state. The formation orders are similar in the forward and reverse cases.

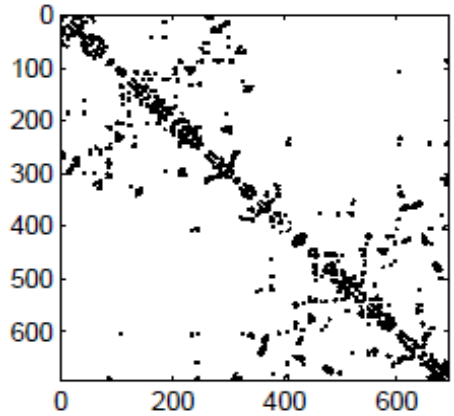
The RMSD to the target reach a plateau at 0.1 Å in TMC simulations. The final state obtained from TMC simulation shows very similar contacts to the target state and generate more contacts than ANM-MC due to the higher degree of approach to the target. The main contacts first appear in snapshot 10 on the red and blue regions. In snapshot 20, more native contacts appear on the yellow, green and purple region (Figure 4.37). Snapshot 35 shows new contacts between residues 10-118 and 10-161, which are not formed in ANM-MC. In the reverse TMC case; the initial contacts disappear in the reverse order of contact formation of the forward run. They fully disappear at snapshot 35. This shows that the forward and reverse TMC cases follow a similar pathway in terms of contacts (Figure 4.37 and 4.39).



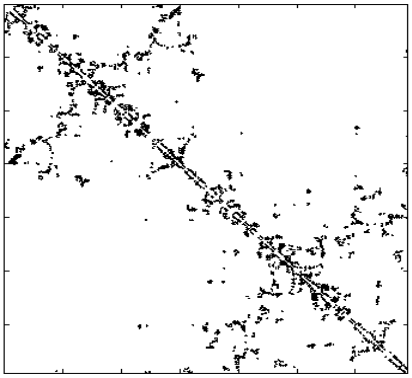
final state at cutoff=10
(Kim *et al.*, 2002b)



final state at cutoff=10
(this thesis)



new contacts formed at cutoff=10
(Kim *et al.*, 2002b)



new contacts formed at cutoff=10
(this thesis)

Figure 4.31. Comparison of the contact maps of forward ANM-MC with the study of Kim and coworkers (2002b).

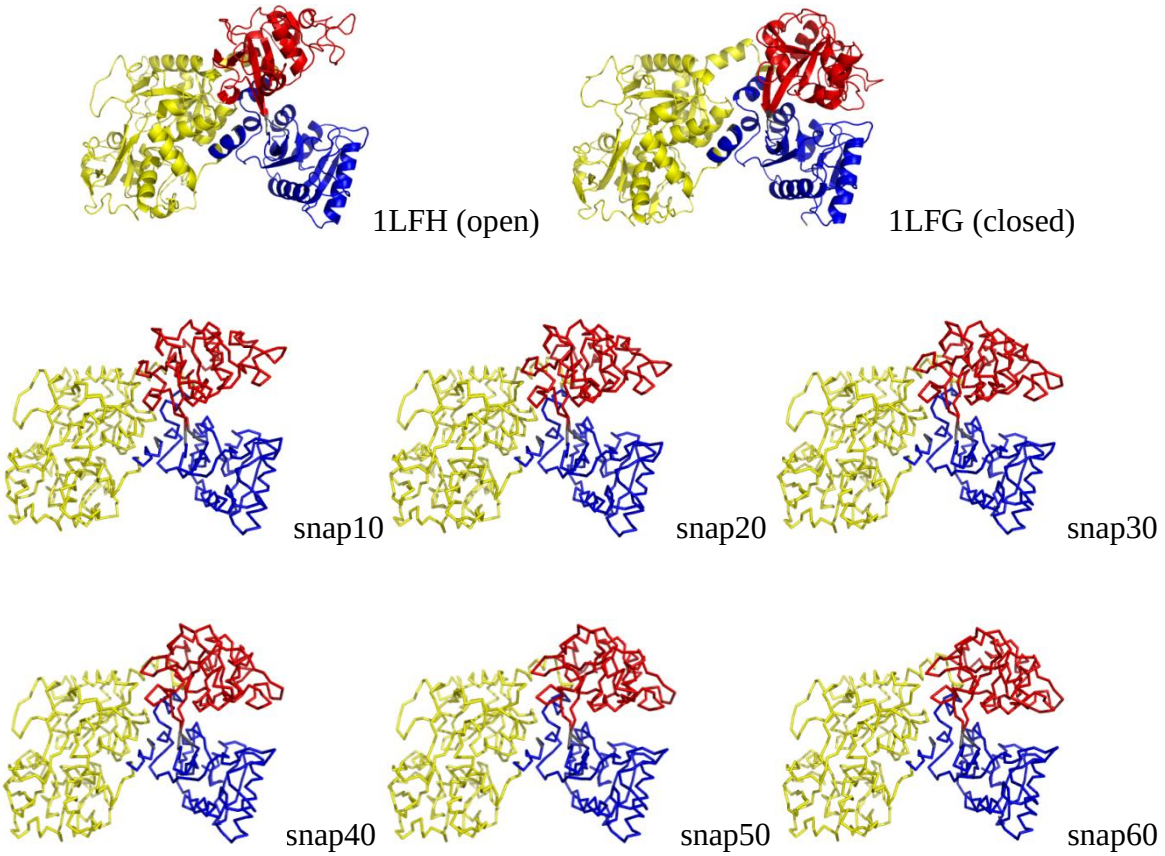


Figure 4.32. Several intermediate structures obtained during forward ANM-MC simulation of Lactoferrin transition from open to closed conformation.

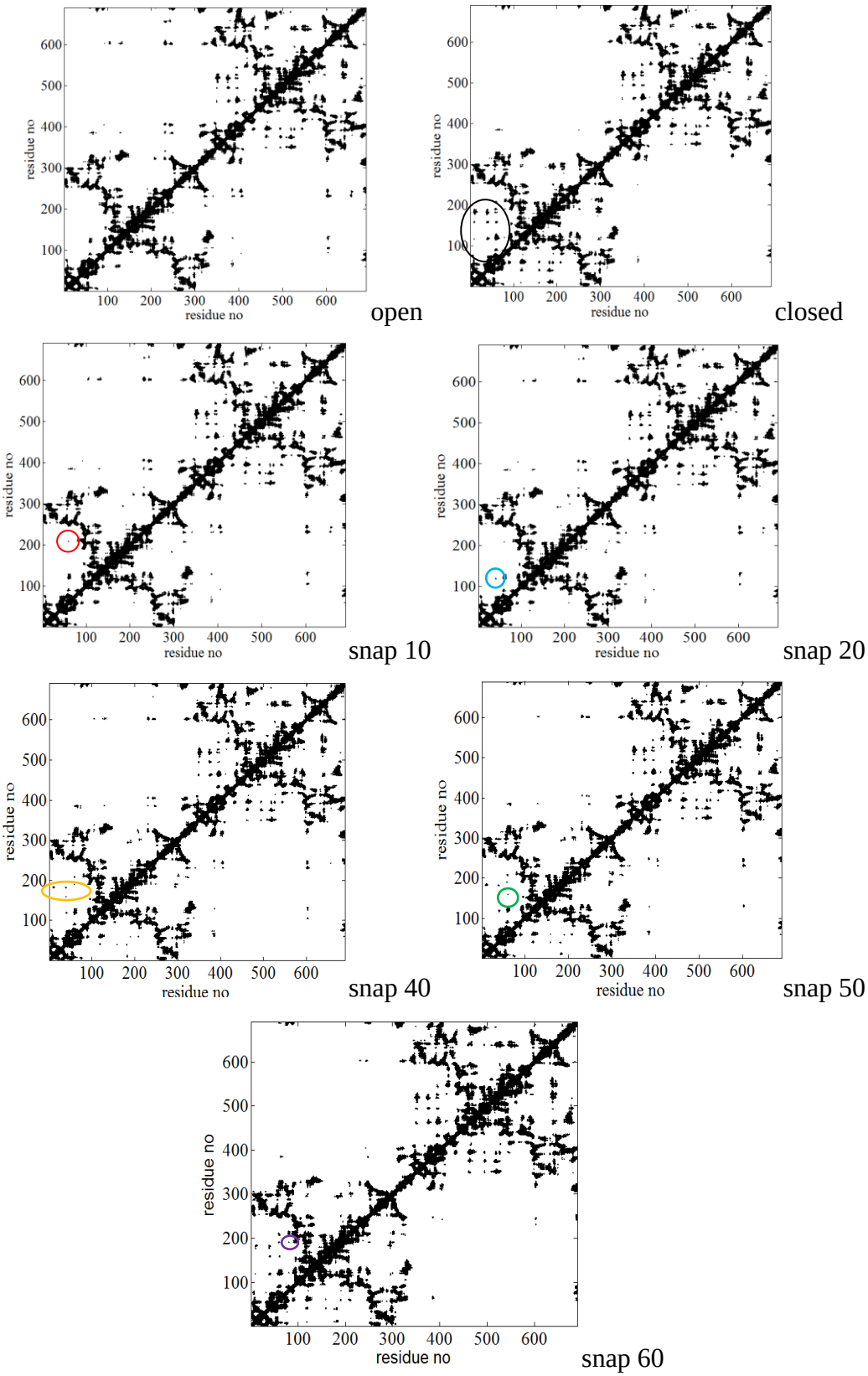


Figure 4.33. Contact maps for open, closed forms of Lactoferrin and for snapshots 10-20-30-40-50-60 for forward ANM-MC run.

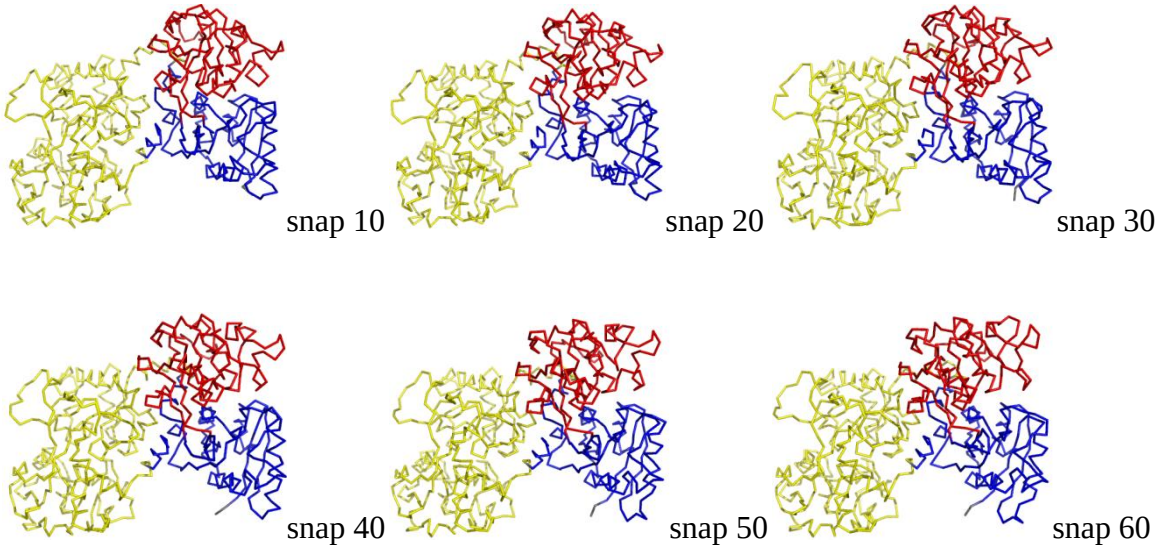


Figure 4.34. Several intermediate structures obtained during reverse ANM-MC simulation of Lactoferrin transition from closed to open conformation.

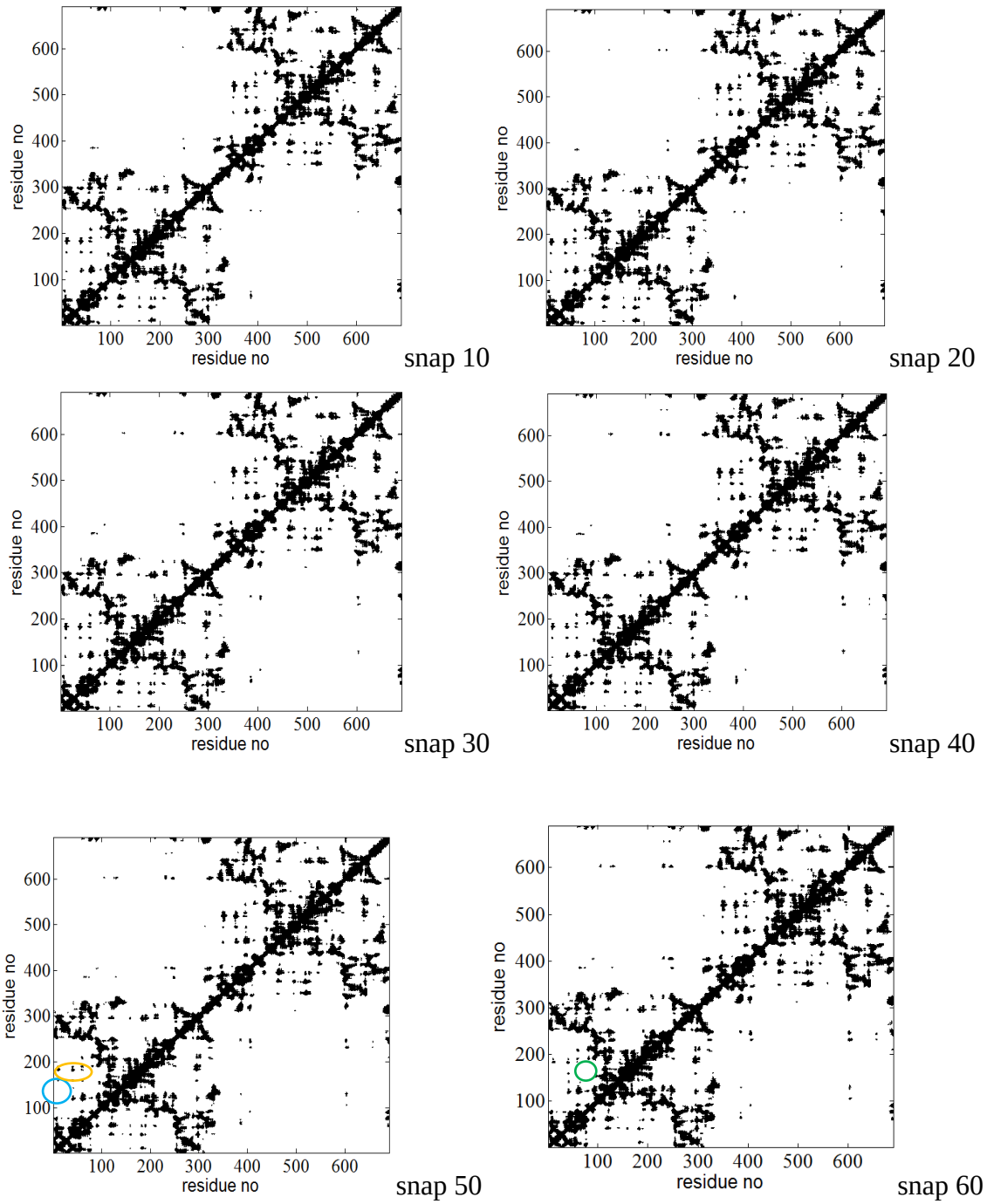


Figure 4.35. Contact maps for open, closed forms of Lactoferrin and for snapshots 10-20-30-40-50-60 for reverse ANM-MC run.

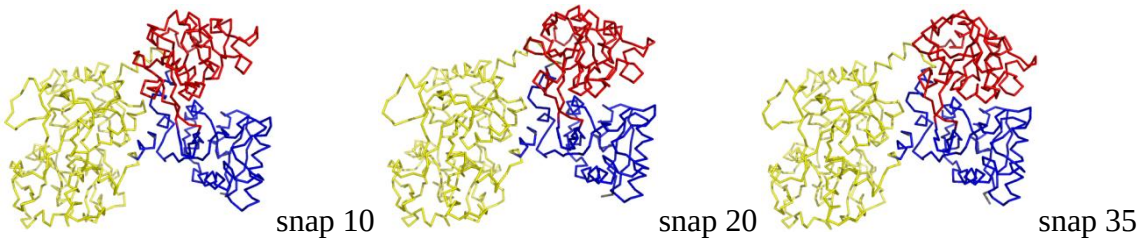


Figure 4.36. Several intermediate structures obtained during forward TMC simulation of Lactoferrin transition from open to closed conformation.

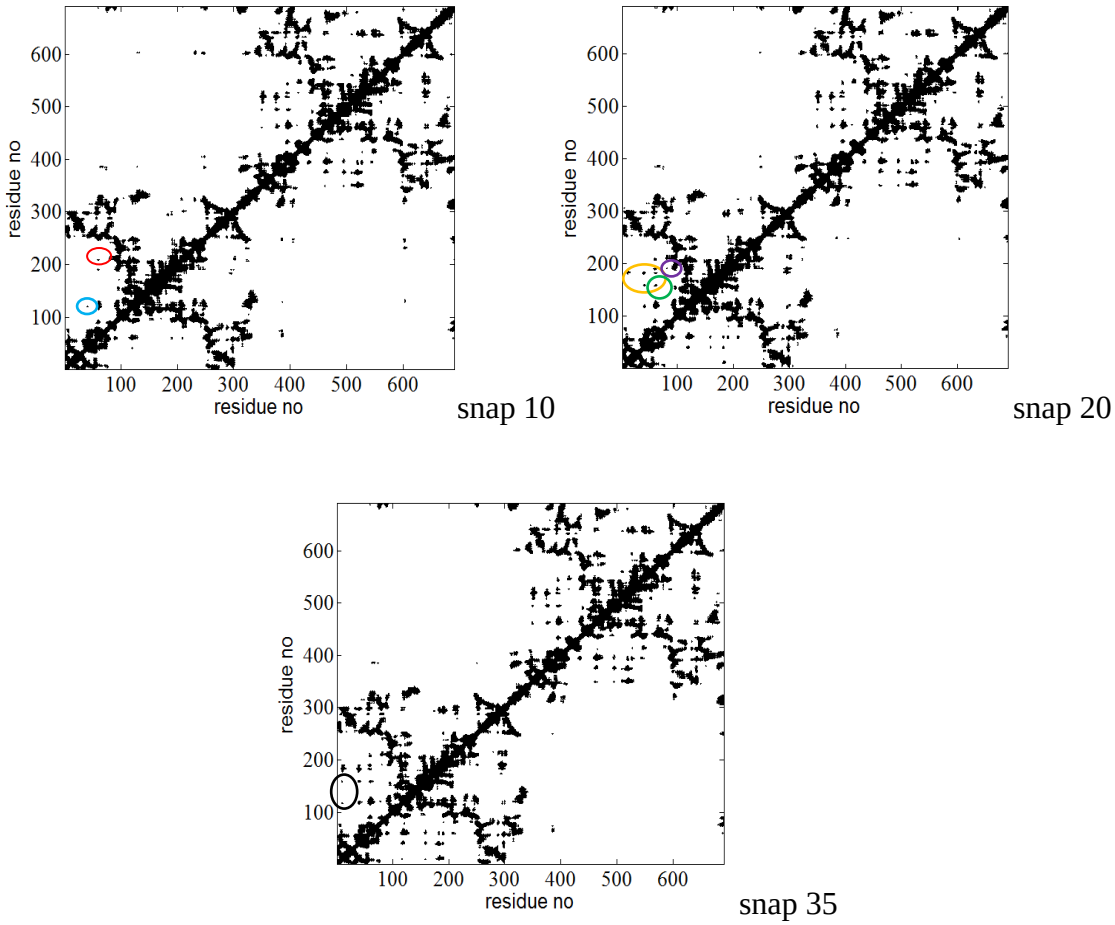


Figure 4.37. Contact maps for open, closed forms of Lactoferrin and for snapshots 10-20-35 for forward TMC run.

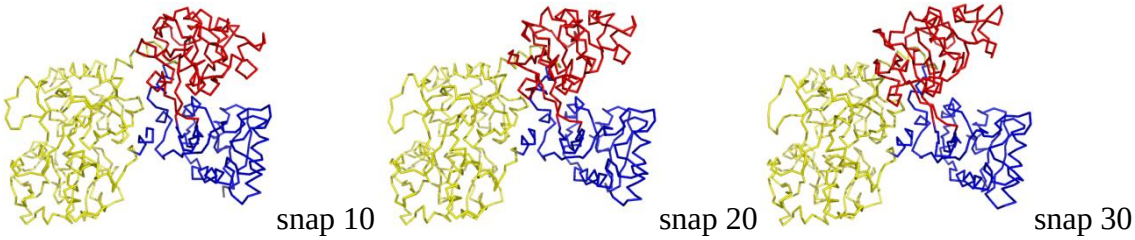


Figure 4.38. Several intermediate structures obtained during reverse TMC simulation of Lactoferrin transition from closed to open conformation.

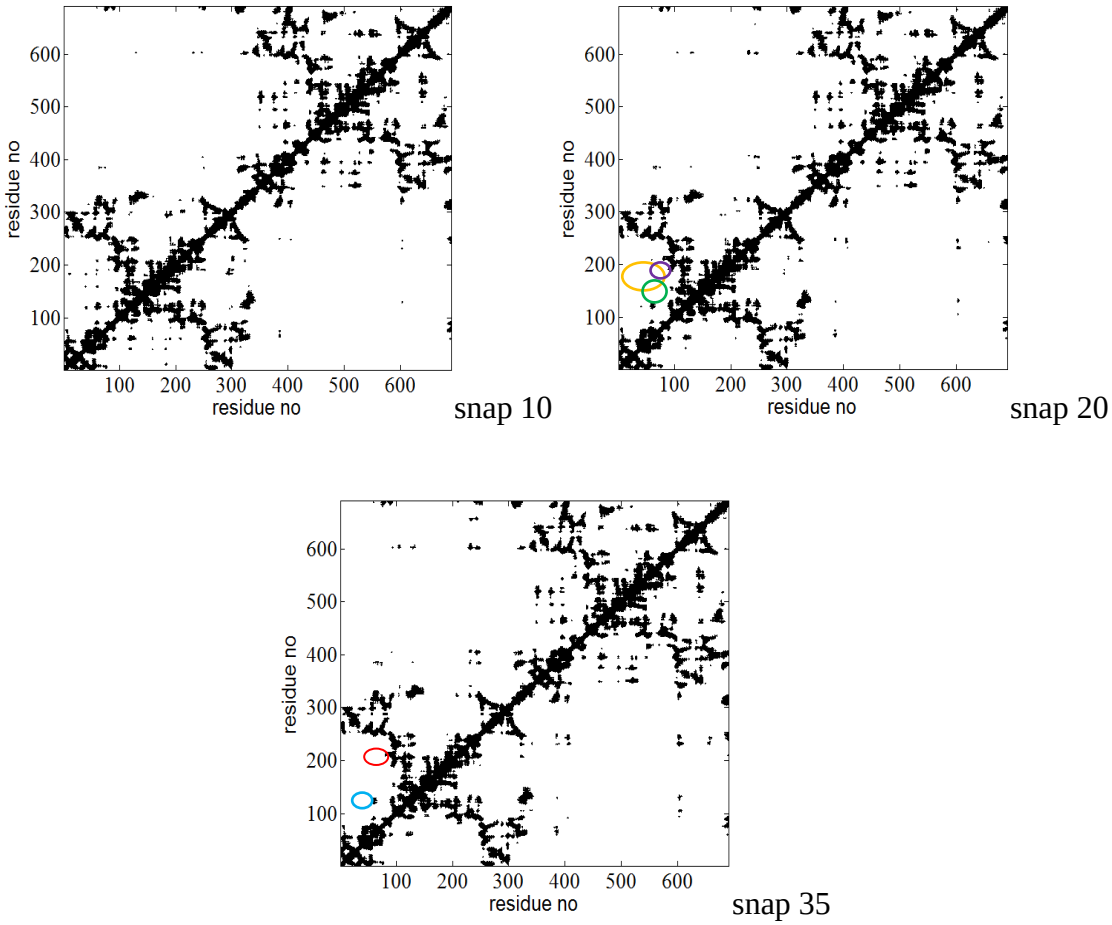
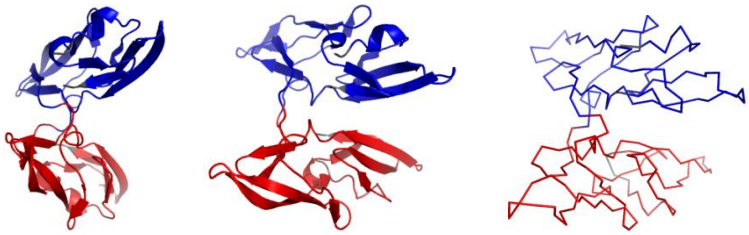


Figure 4.39. Contact maps for open, closed forms of Lactoferrin and for snapshots 10-20-35 for reverse TMC run.

4.5.3. Other Proteins

The final snapshot and the contact map of the other proteins except calmodulin, and lactoferrin are given in this subsection. According to the Figures 4.40-4.58; the final state, which is obtained from the simulation is very similar to the target state in both ANM-MC and TMC simulations. More contacts are observed on the contact maps of TMC when compared to ANM-MC simulation due to the fact that the final form obtained in TMC is more similar to the target state than that of ANM-MC. In other words, it shows the higher degree of approach to the target state in TMC.

The diphtheria toxin protein has a missing residue region, which is completed with the consecutive residues in the monomeric stucture. The protein is separated into two chains at the missing residue region to obtain a dimeric structure. Both monomeric and dimeric structures generate similar final structures at the end of the simulation and have a slight difference in terms of contact maps.



1L5E (open) 1L5B (closed) final (snap 50)

Figure 4.40. The final snapshot obtained from forward ANM-MC of Antiviral protein.

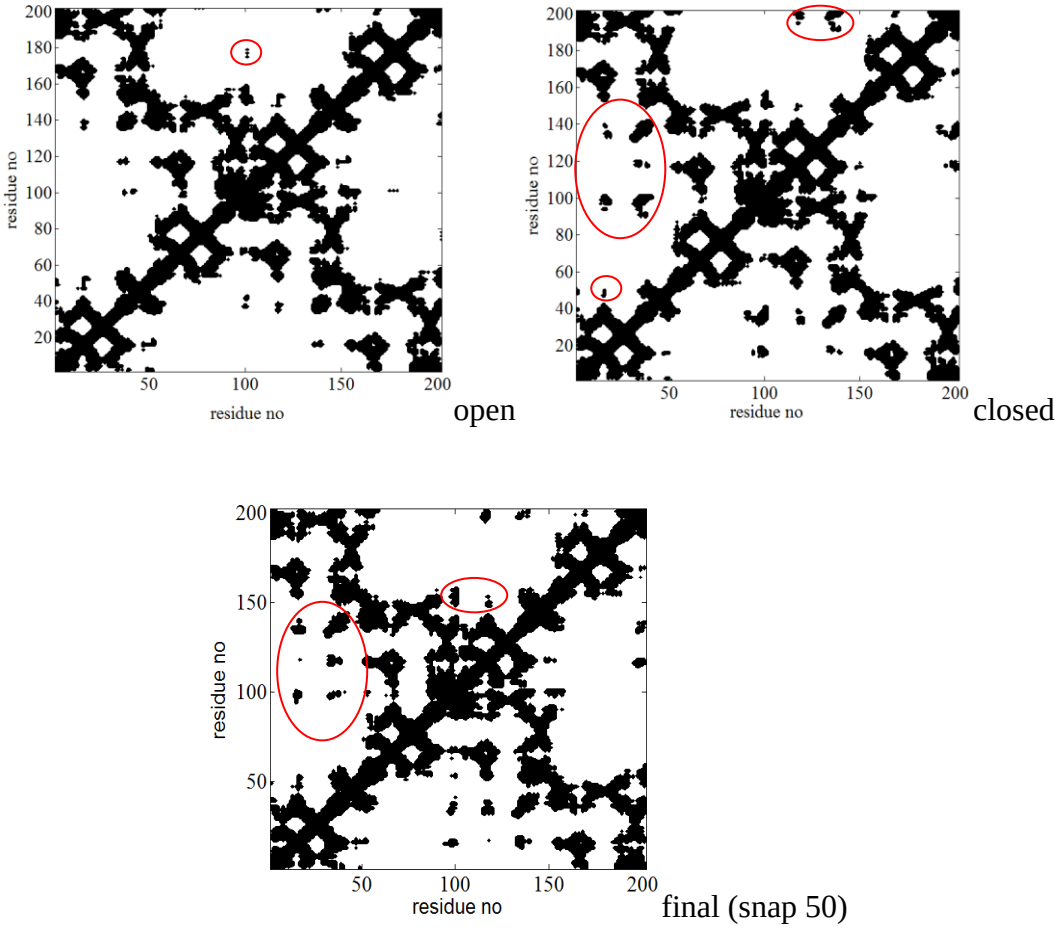
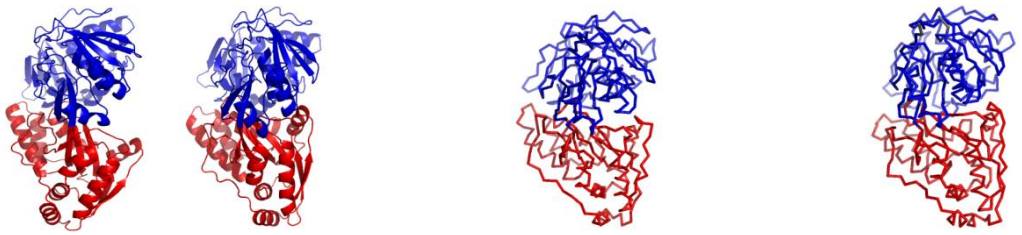


Figure 4.41. Contact maps for open, closed and final forms of Antiviral protein in forward ANM-MC run.



1DPE (open) 1DPP (closed) final (snap 65) in ANM-MC final (snap 30) in TMC

Figure 4.42. The final snapshots obtained from forward ANM-MC and TMC of Dipeptide binding protein.

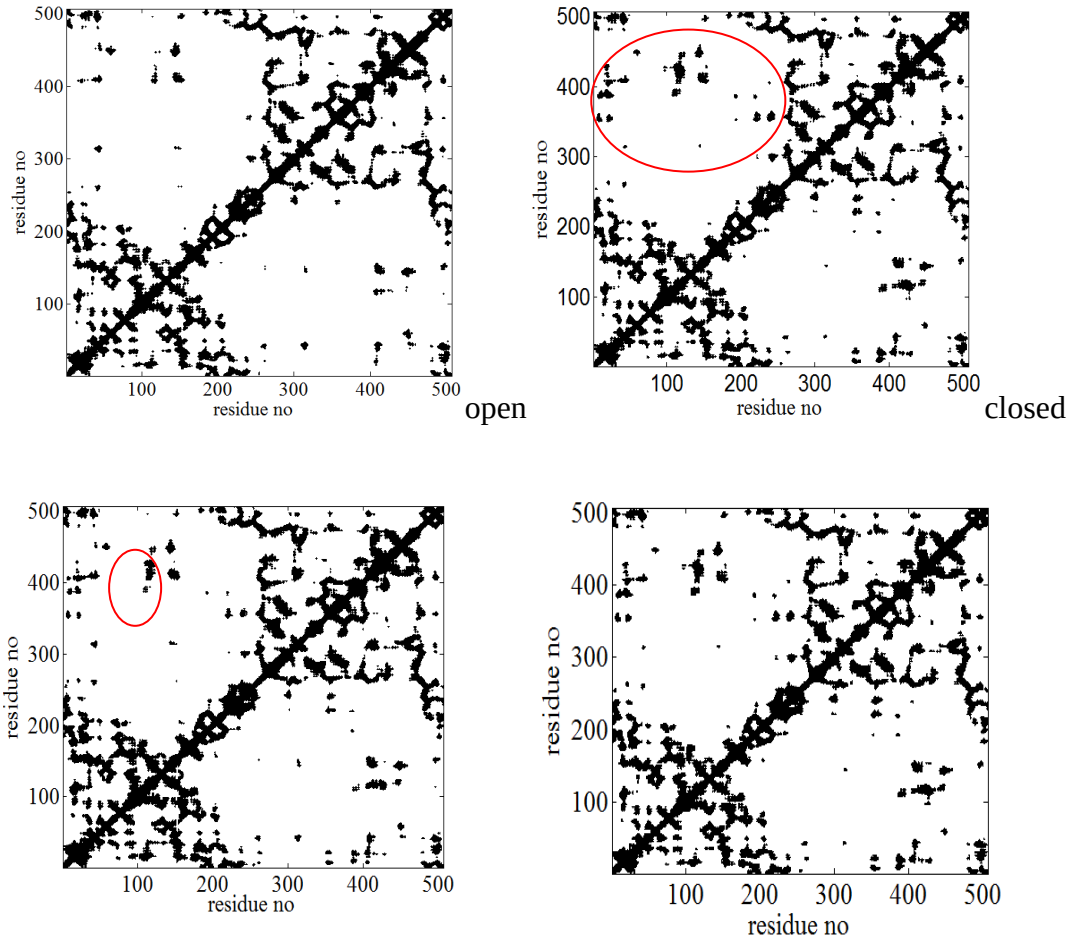


Figure 4.43. Contact maps for open, closed and final forms of Dipeptide binding protein. in forward ANM-MC and TMC runs.

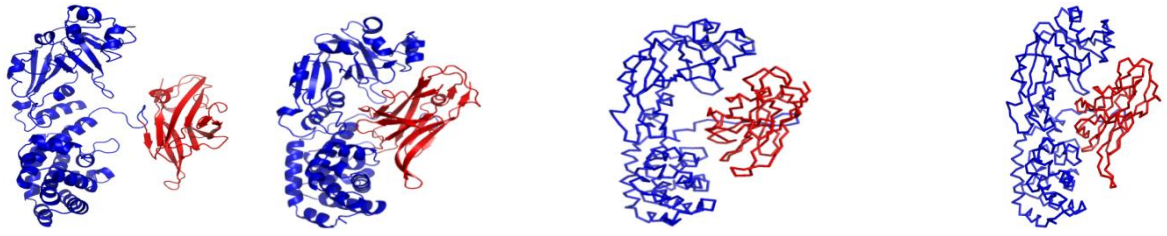


Figure 4.44. The final snapshots obtained from forward ANM-MC and TMC of DT.

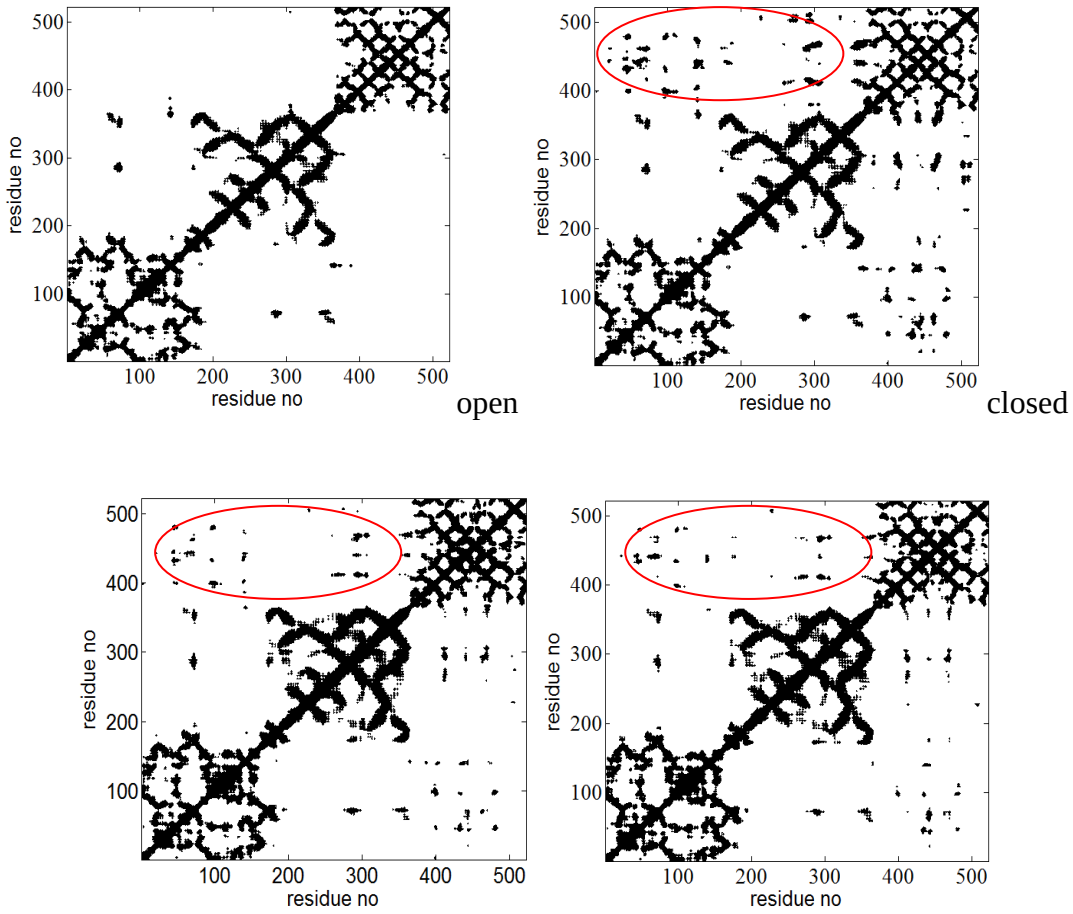


Figure 4.45. Contact maps for open, closed and final forms of DT in forward ANM-MC.

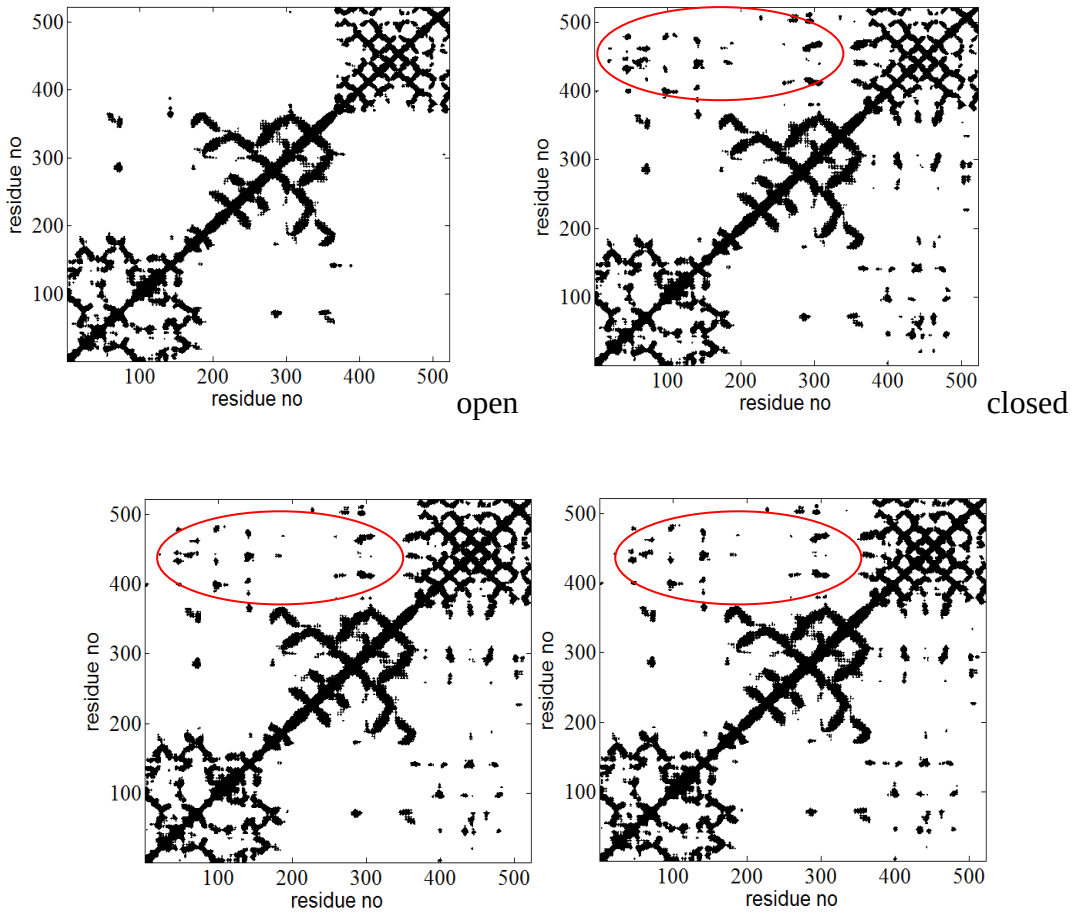
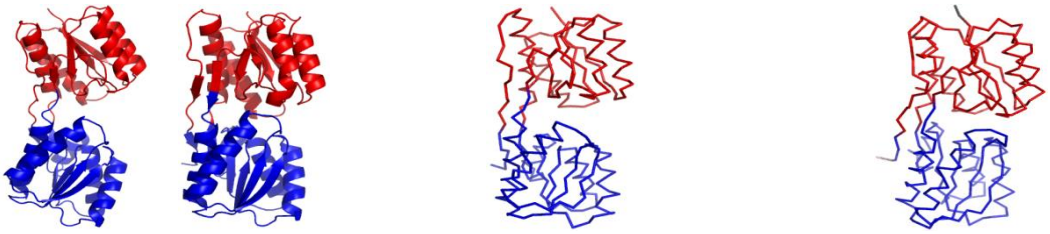


Figure 4.46. Contact maps for open, closed and final forms of DT in forward TMC run.



1URP (open) 2DRI (closed) final (snap 30) in ANM-MC final (snap 20) in TMC

Figure 4.47. The final snapshots obtained from forward ANM-MC and TMC of d-ribose binding protein

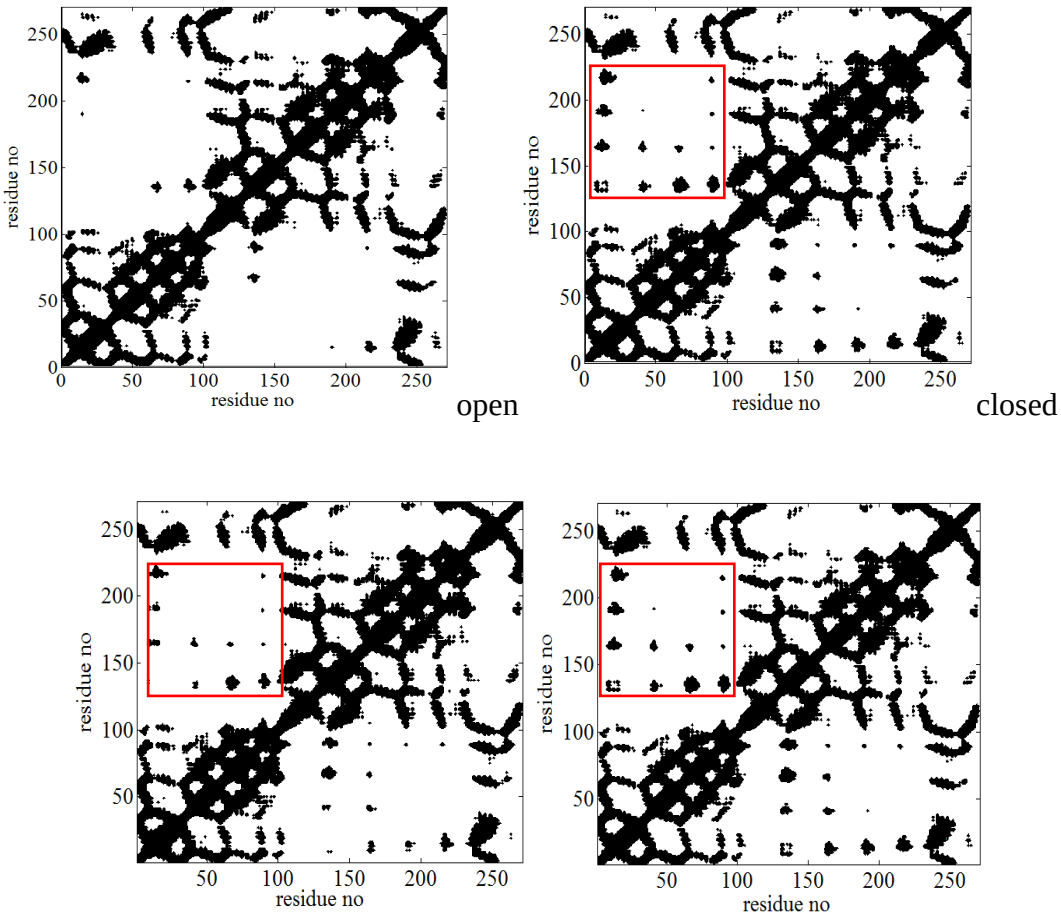


Figure 4.48. The final snapshots obtained from forward ANM-MC and TMC of d-ribose binding protein.

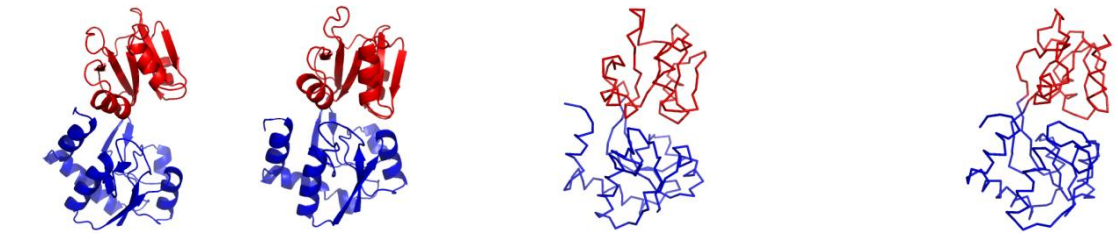


Figure 4.49. The final snapshots obtained from forward ANM-MC and TMC of GBP.

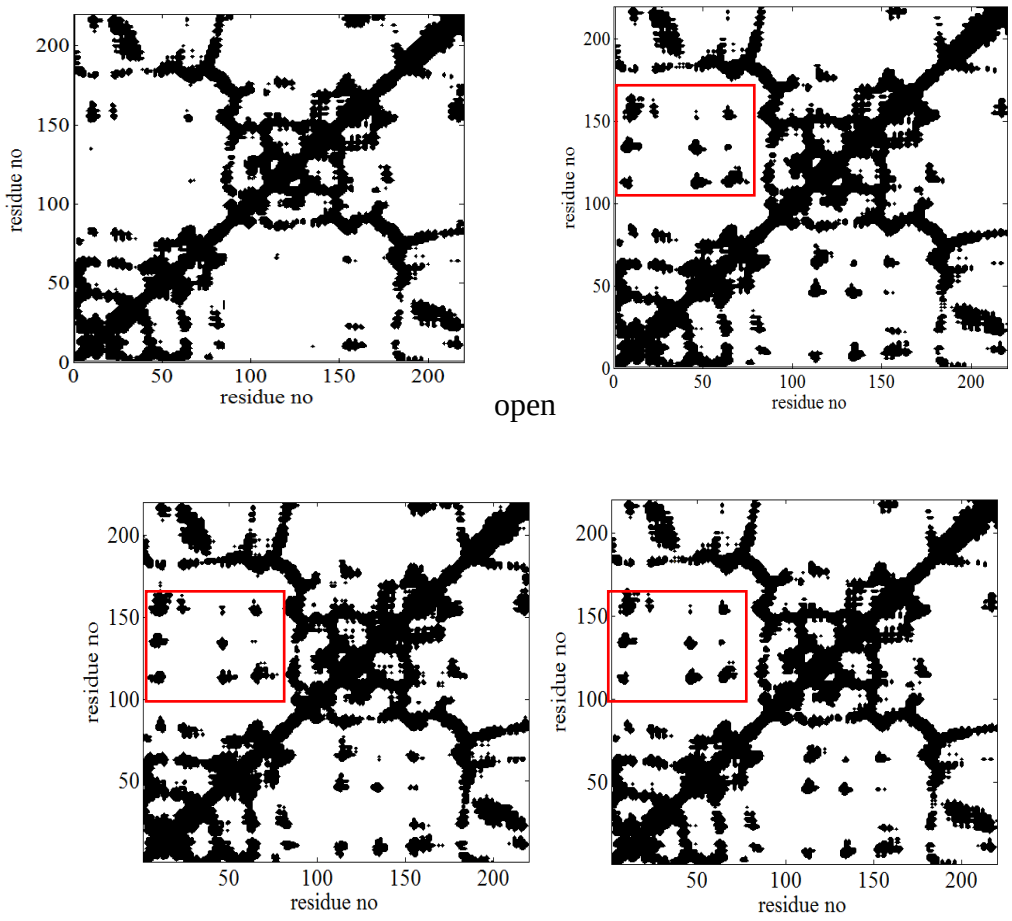


Figure 4.50. The final snapshots obtained from forward ANM-MC and TMC of GBP.

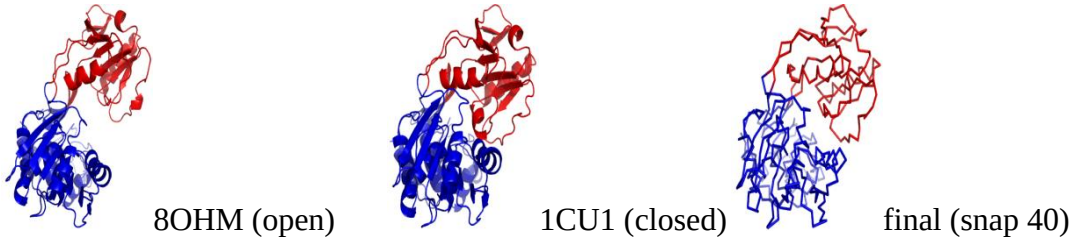
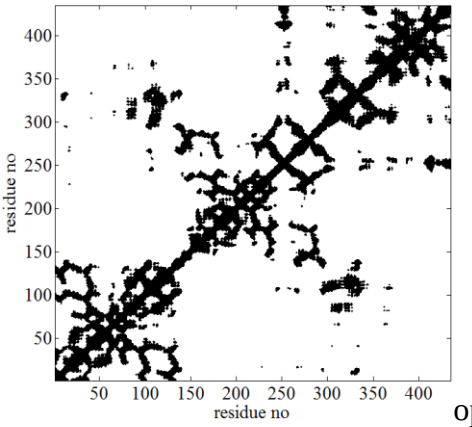
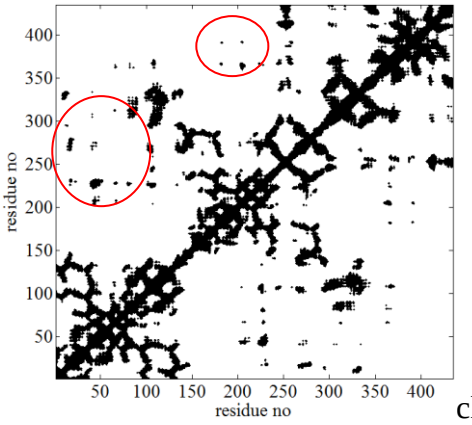


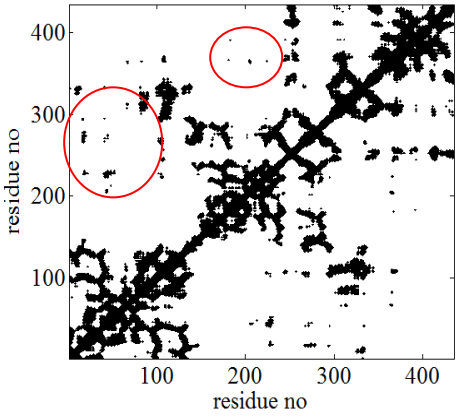
Figure 4.51. The final snapshot obtained from forward ANM-MC of helicase.



open

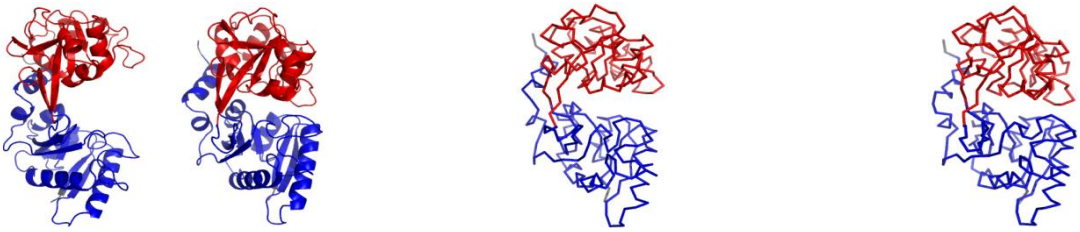


closed



final (snap 40)

Figure 4.52. The final snapshot obtained from forward ANM-MC of helicase



1BP5 (open) 1A8E (closed) final (snap 50) in ANM-MC final (snap 30) in TMC

Figure 4.53. The final snapshots obtained from forward ANM-MC and TMC of HSTR.

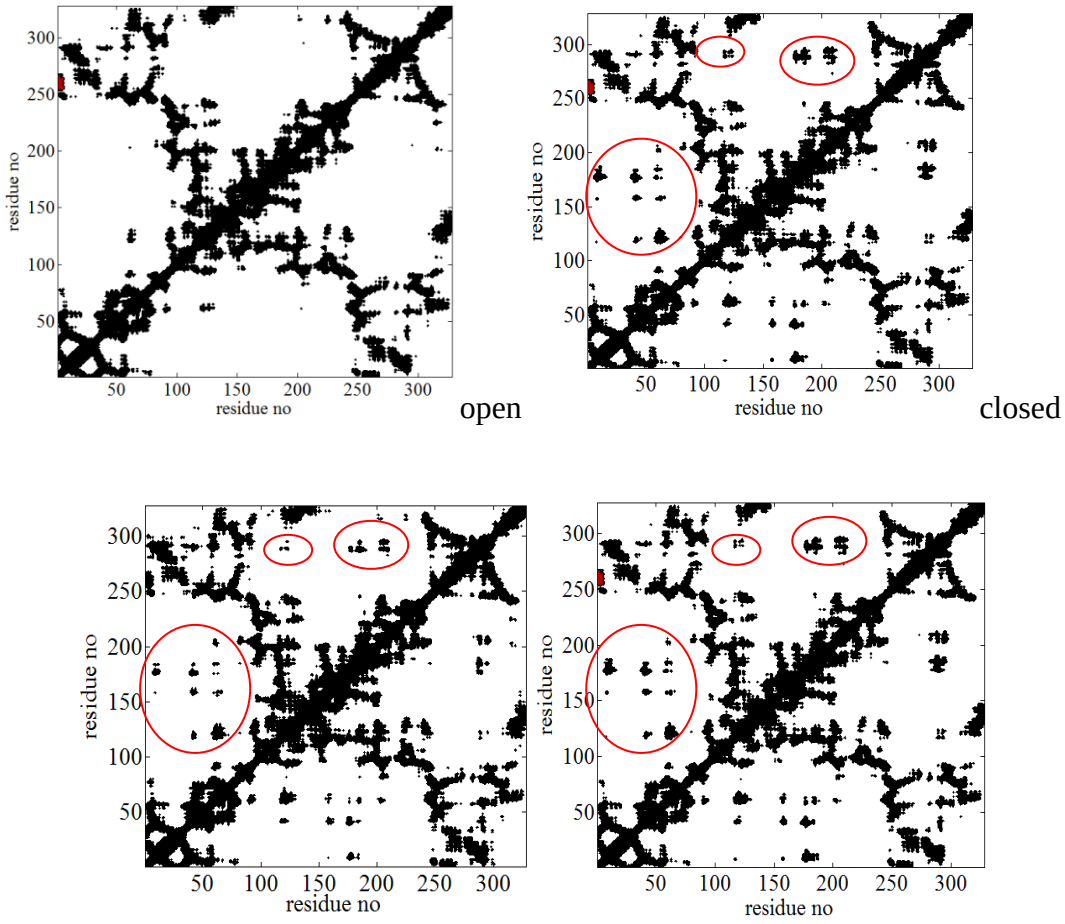


Figure 4.54. The final snapshots obtained from forward ANM-MC and TMC of HSTR.

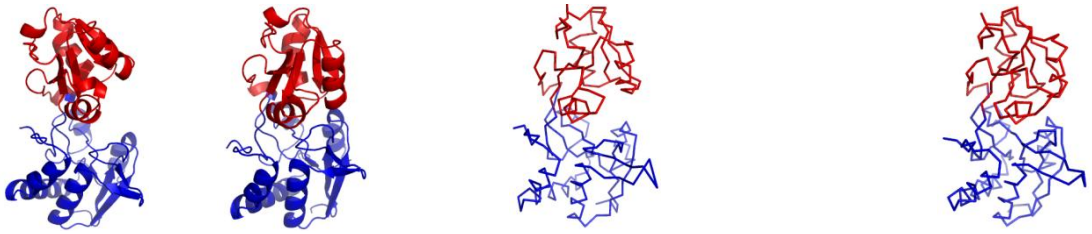
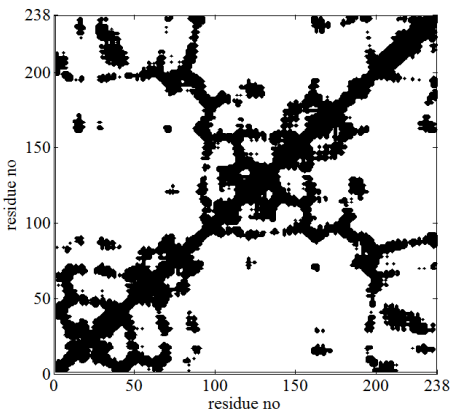
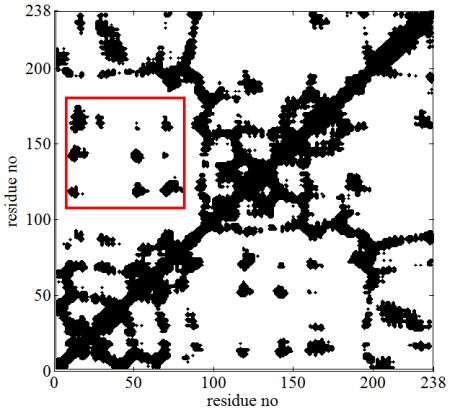


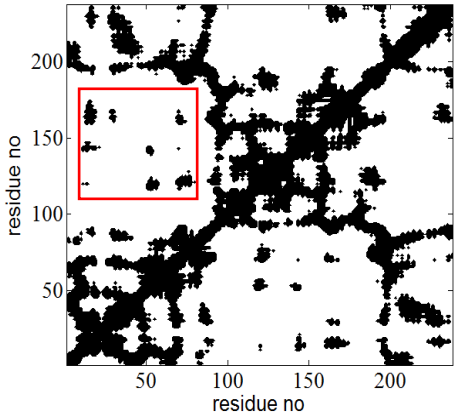
Figure 4.55. The final snapshots obtained from forward ANM-MC and TMC of LAO binding protein.



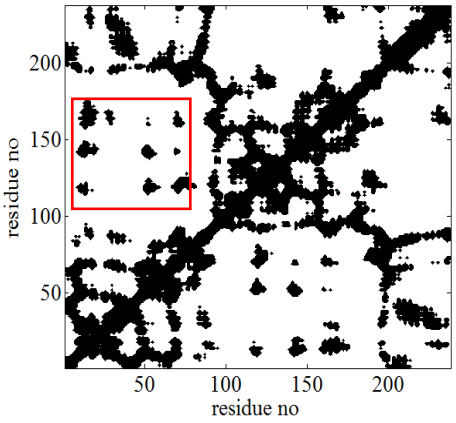
open



closed

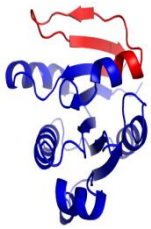


final (snap 30) in ANM-MC

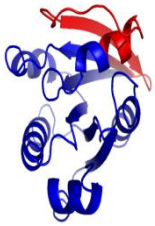


final (snap 20) in TMC

Figure 4.56. The final snapshot obtained from forward ANM-MC and TMC of LAO binding protein.



1E0S (open)



2J5X (closed)

final (snap 20) in TMC

Figure 4.57. The final snapshots obtained from forward TMC of small G-protein.

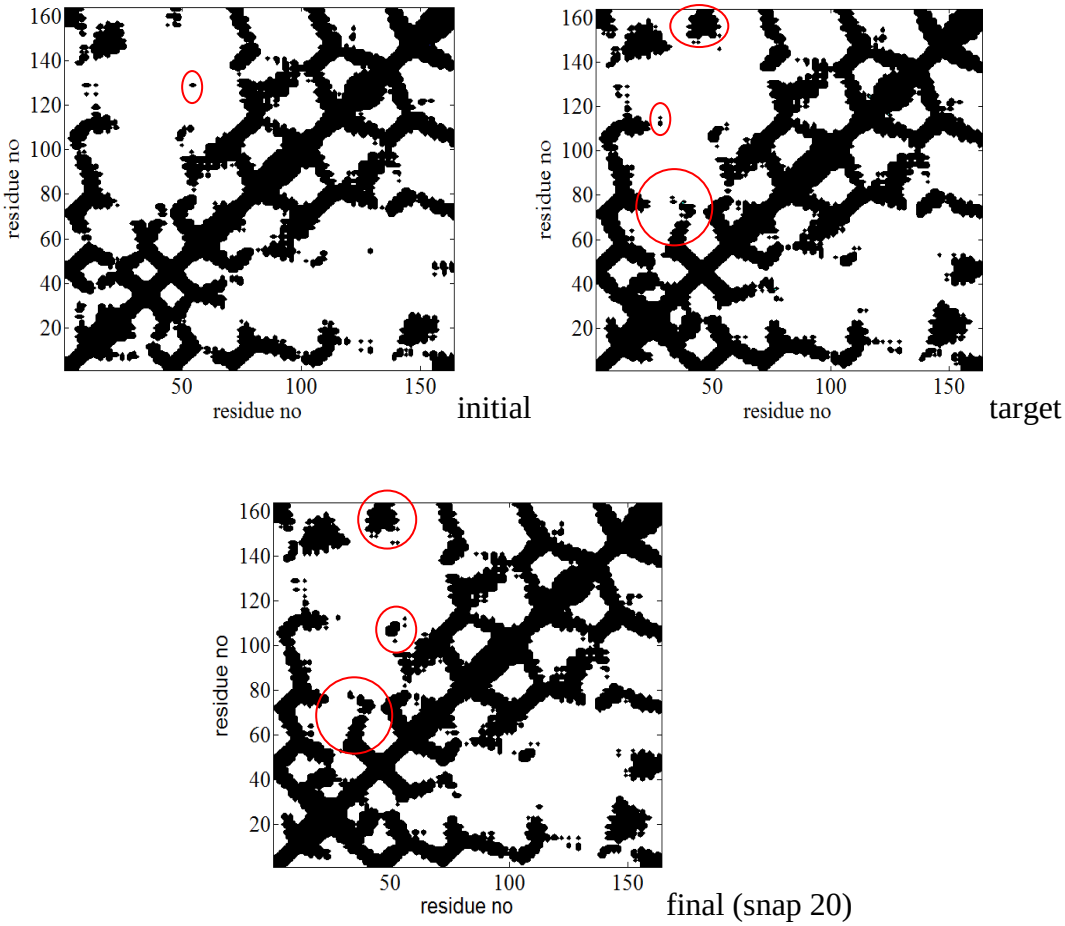


Figure 4.58. The final snapshot obtained from forward TMC of Small G-protein Arf6.

5. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

In this thesis, TMC and ANM-MC simulation techniques (Kantarci-Carsibasi *et al.*, 2008; Kantarci-Carsibasi, 2009) have been fully automated and improved in terms of simulation parameters in order to investigate conformational transition pathways of proteins that undergo large conformational changes (RMSD > 4 Å) between apo/initial and closed/target states. ANM-MC and TMC methodology have been applied to 12 proteins varying in size and domain structure to obtain transitions in the forward and reverse directions.

The simulation results for adenylate kinase, including the RMSD, total potential energy, and overlap value profiles, and the intermediate snapshots/contact maps, have been compared with the initial study to adjust the simulation parameters for performing better energy minimization and more native contacts in the final structures. A combination of DF = 0.2 Å, perturbation strength = 0.1 Å and MCS = 50-100 as simulation parameters has proven successful in maintaining satisfactory RMSD and energy profiles, thus achieving a close approach to the target state with reasonable pathway intermediates.

The RMSD values to the target state reach a plateau which indicates the end of the simulations. The final structures approach the target with an RMSD value within 0.4 Å in TMC simulations, whereas it varies in a range of 1.4 to 3.9 Å in the forward ANM-MC runs and between 1.8 Å and 4.7 Å in the reverse ANM-MC runs, except the reverse case of diphtheria toxin. The initial RMSD values of the proteins from the target fall between 4.1 and 15.6 Å. These initial RMSD values affect the number of iterations to reach the plateau region, i.e. the proteins with higher initial RMSD values reach the plateau region in more iterations. The most successful results in terms of final RMSDs are obtained for the proteins exhibiting hinge motion between two domains (both consisting of alpha/beta folds). The final structure obtained from the simulation approach the target more in forward runs when compared to reverse runs. In the case of small G-protein Arf6, which is the only case presenting a shear motion, the initial RMSD of 4.2 decreases to only 3.1 Å.

In general the total potential energy profiles follow a decreasing trend, which results from satisfactory energy minimization during MC iterations. However; diphtheria toxin with missing residues is an exception and exhibits a maximum energy value in TMC simulation.

The overlap values decrease during the iterations and fall below 0.3 at the plateau region. The higher the initial overlap values between the selected modes and the target direction, the closer the final structure approaches the target. The most selected modes at the first stages of the iterations are the first five modes, which are functionally the most important modes in driving the conformational changes (Tama and Sanejouand, 2001). There is a fluctuating trend in the selected modes after the plateau region.

Analysis of snapshots and the contact maps show that the final structure and the native contacts formed are very similar to those of the target structure. As expected, more contacts are formed in TMC simulations when compared to ANM-MC simulations. Both techniques are successful to illustrate the intermediate structures and the transition pathways. The detailed analysis of calmodulin and lactoferrin reveal that the new contacts appear on the hinge region, which approach each other during the conformational transition. The final contacts in the lactoferrin case are consistent with a previous study (Kim *et al*., 2002b).

5.2. Recommendations

In this thesis, the simulations have been applied to monomeric proteins and one dimeric protein up to maximum 690 residues. Further work is required for the methodology to cover a wide range of proteins composed of larger system sizes and multi-chains, such as GroEL-GroES and K-channel proteins. MCS can be taken as 20 or 50 to reduce the computational time for the simulations of larger systems.

Specifically for the proteins exhibiting low overlap values at the first stages of the simulation, the possibility of using more than one slowest mode (a combination of several slow modes) may be investigated to possibly obtain higher overlap values. When the overlap values between the selected mode and the target direction reduces below 0.3 and

the RMSD from the target reach a plateau and stabilize; the final state cannot approach the target state any more. After this point, the ANM-MC simulations could be switched to TMC methodology without the normal modes to obtain a closer approach to the target structure .

An automated way of determining if the predicted structures transition pathway fall close to any of the available crystal structures in the Protein Data Bank would provide a useful means of describing the transition intermediates. This has been carried out in previous studies on adenylate kinase (Maragakis and Karplus, Kantarci et al.)

RG-ANM-MC methodology (Kantarci thesis) has been described as an initial attempt at predicting closed structures from open ones without using target structure information. The method uses the radius of gyration of the protein as a constrain during ANM-MC simulations. This methodology could be improved and applied to the protein set used in this thesis in order assess its success.

APPENDIX A: SNAPSHOTS AND CONTACT MAPS OF OTHER PROTEINS

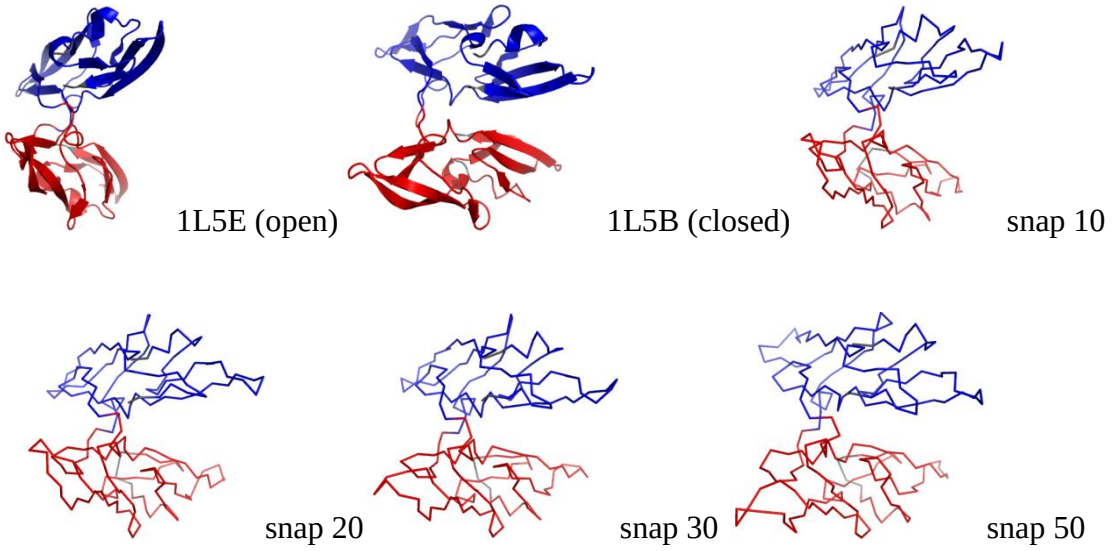


Figure A.1. Several intermediate structures obtained during forward ANM-MC simulation of Antiviral protein transition from open to closed conformation.

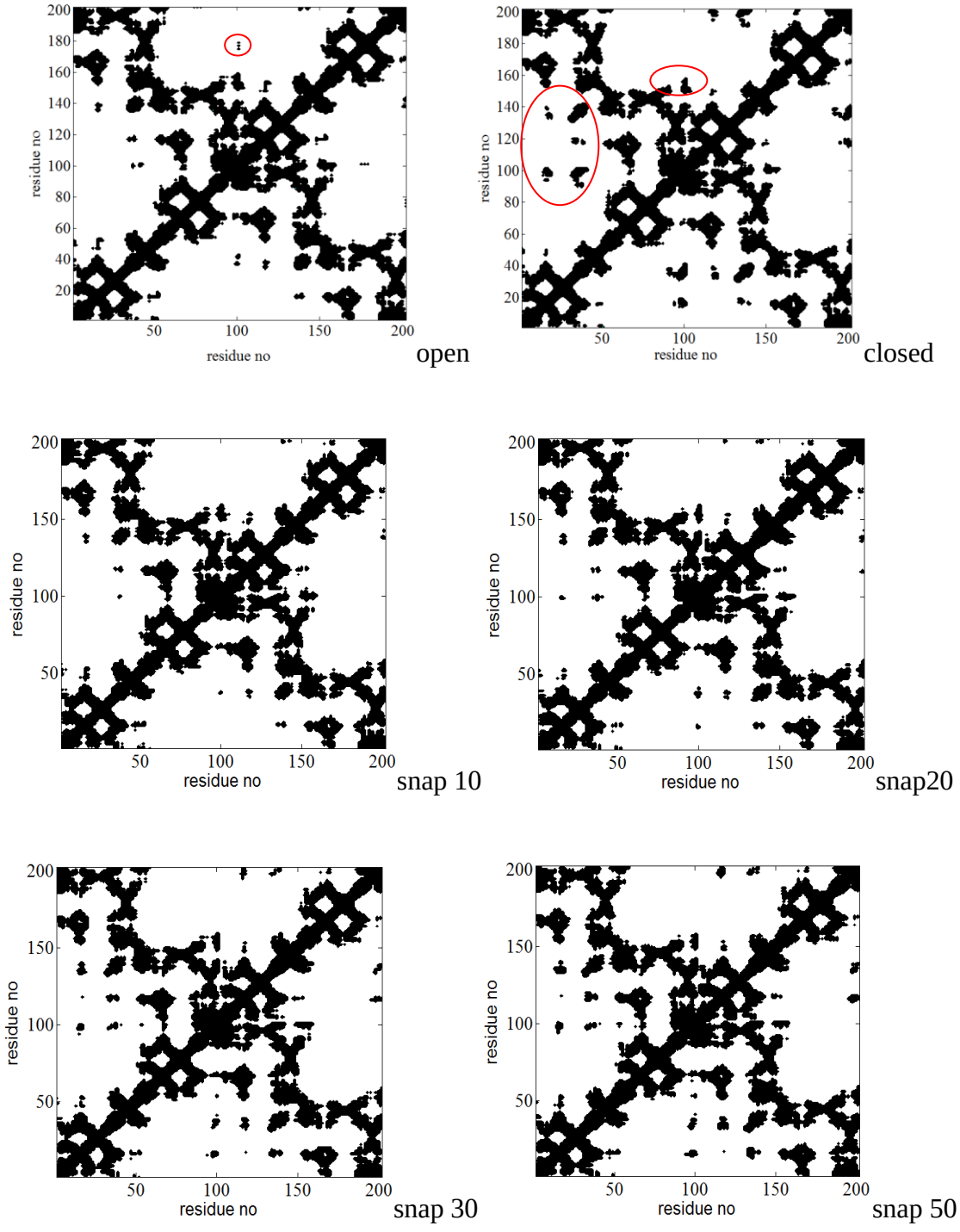


Figure A.2. Contact maps for open, closed forms of Antiviral protein and for snapshots 10-20-30-50 for forward ANM-MC run.

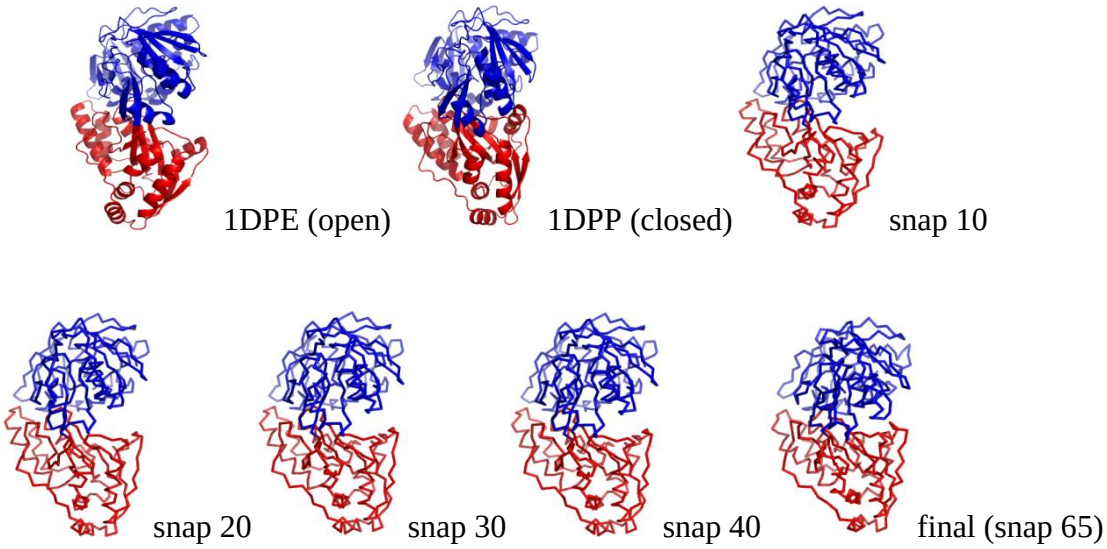


Figure A.3. Several intermediate structures obtained during forward ANM-MC simulation of Dipeptide Binding Protein transition from open to closed conformation.

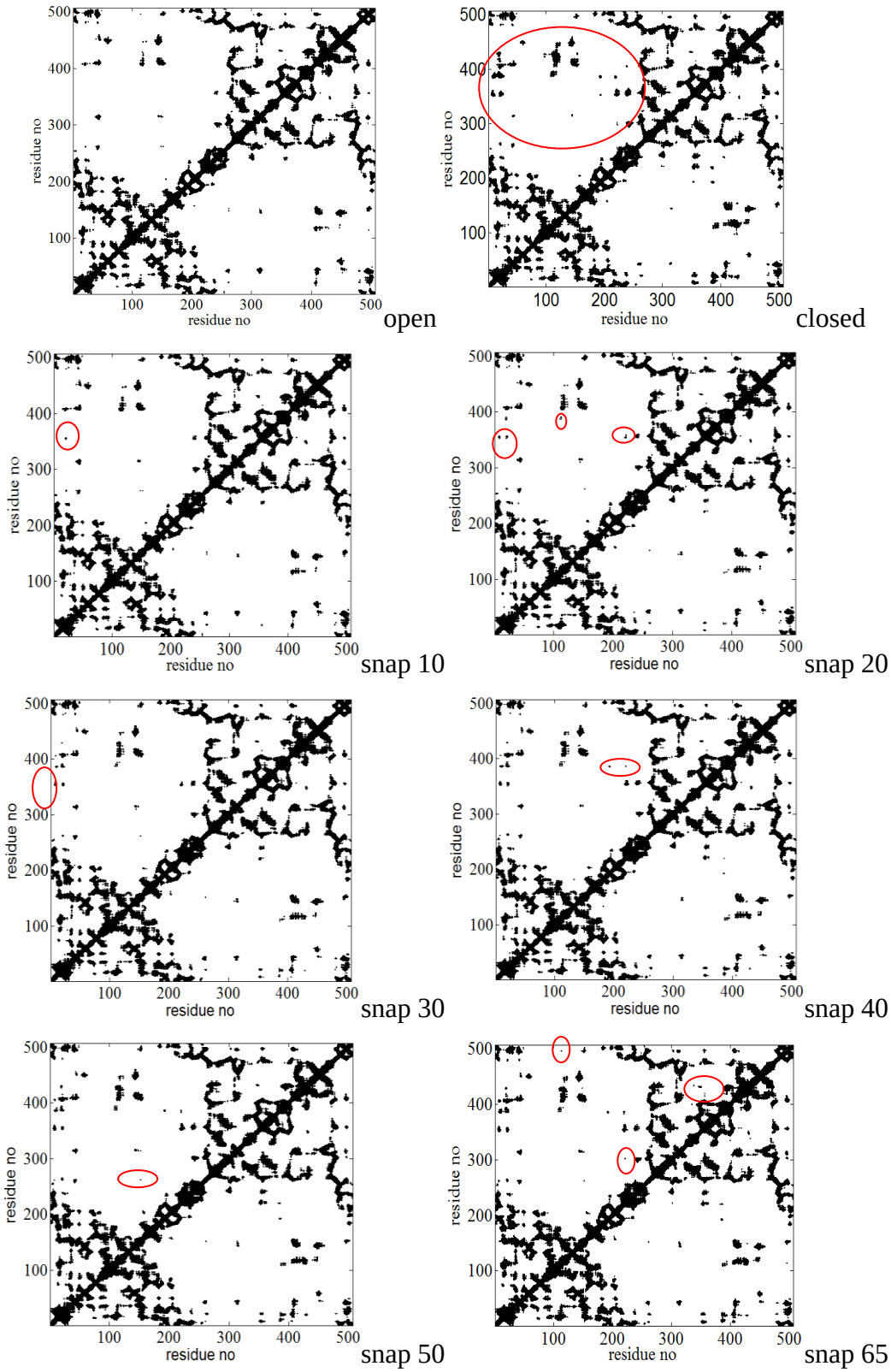


Figure A.4. Contact maps for open, closed forms of Dipeptide Binding protein and for snapshots 10-20-30-40-50-65 for forward ANM-MC run.

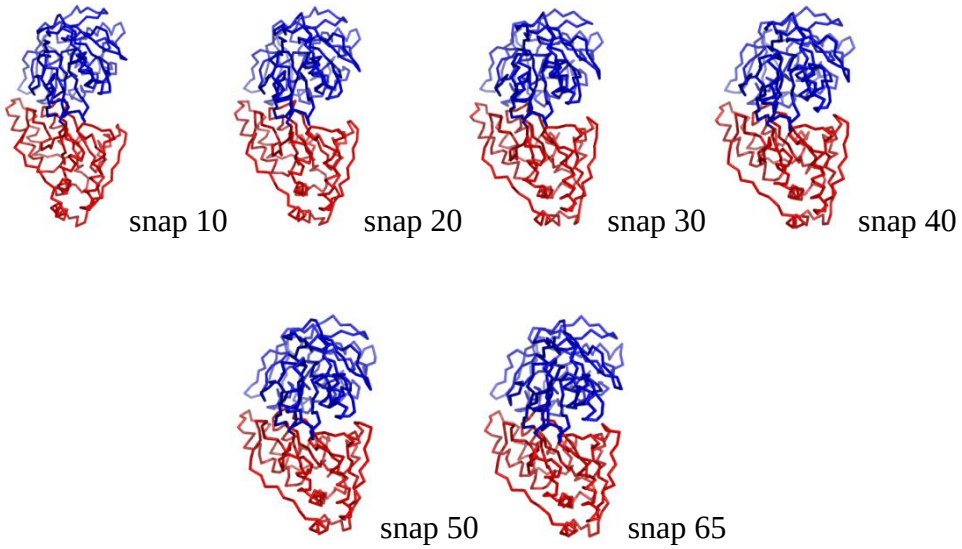


Figure A.5. Several intermediate structures obtained during reverse ANM-MC simulation of dipeptide binding protein transition from closed to open conformation.

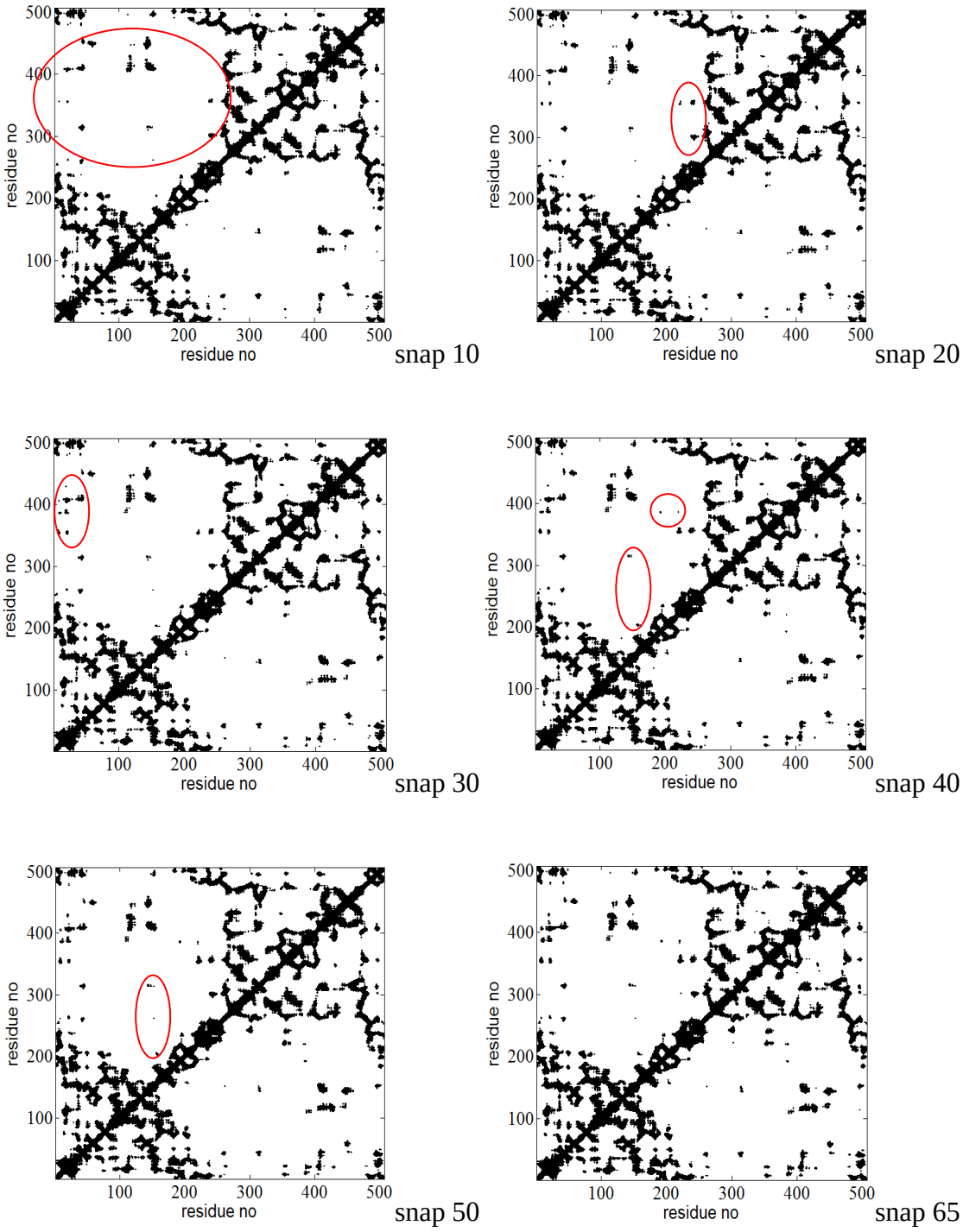


Figure A.6. Contact maps for open, closed forms of Dipeptide Binding protein and for snapshots 10-20-30-40-50-65 for reverse ANM-MC run.

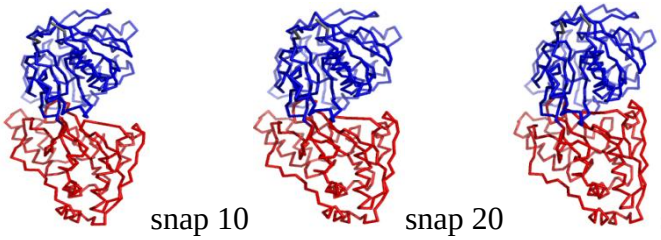


Figure A.7. Several intermediate structures obtained during forward TMC simulation of dipeptide binding protein transition from open to closed conformation.

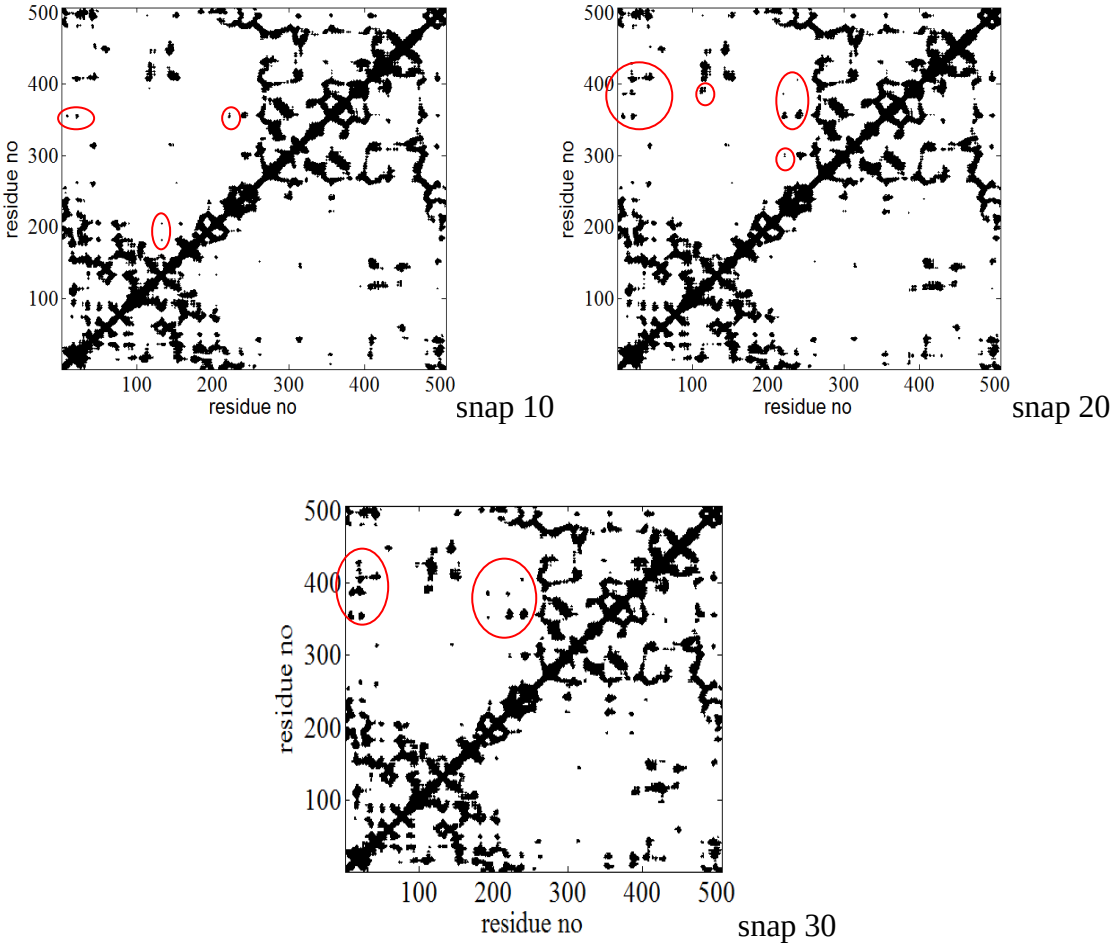


Figure A.8. Contact maps for Dipeptide Binding protein and for snapshots 10-20-30 for forward TMC run.

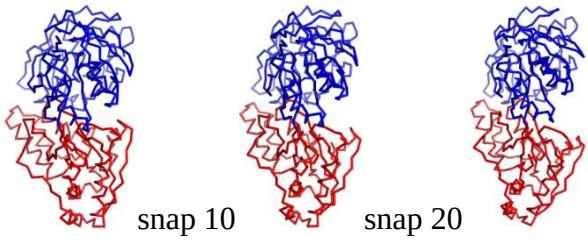
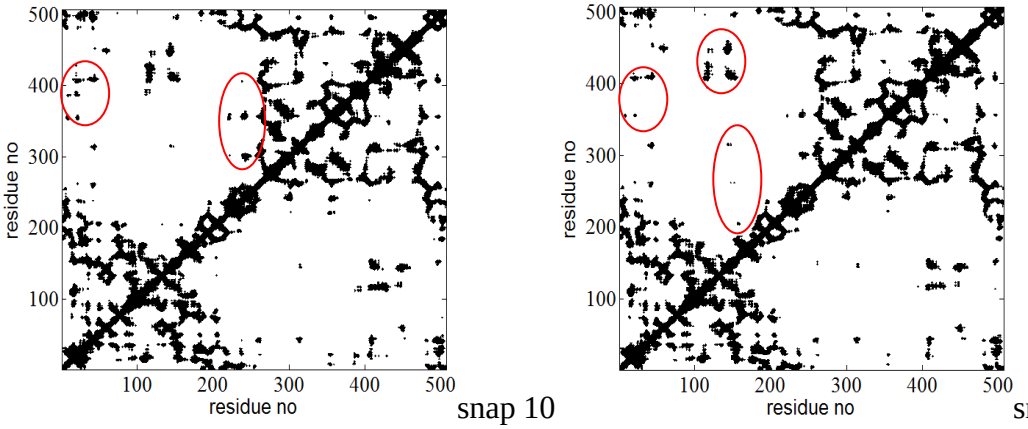


Figure A.9. Several intermediate structures obtained during reverse TMC simulation of dipeptide binding protein transition from closed to open conformation.



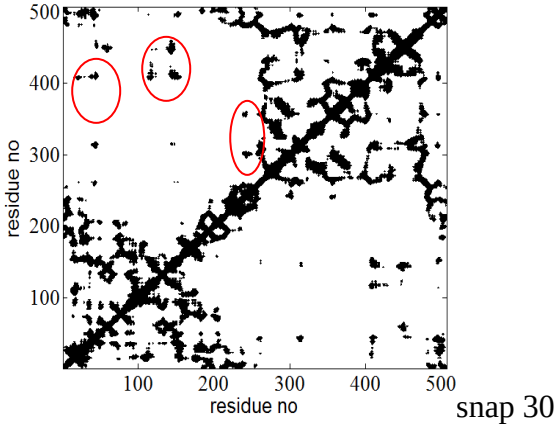


Figure A.10. Contact maps for Dipeptide Binding protein and for snapshots 10-20-30 for reverse TMC run.

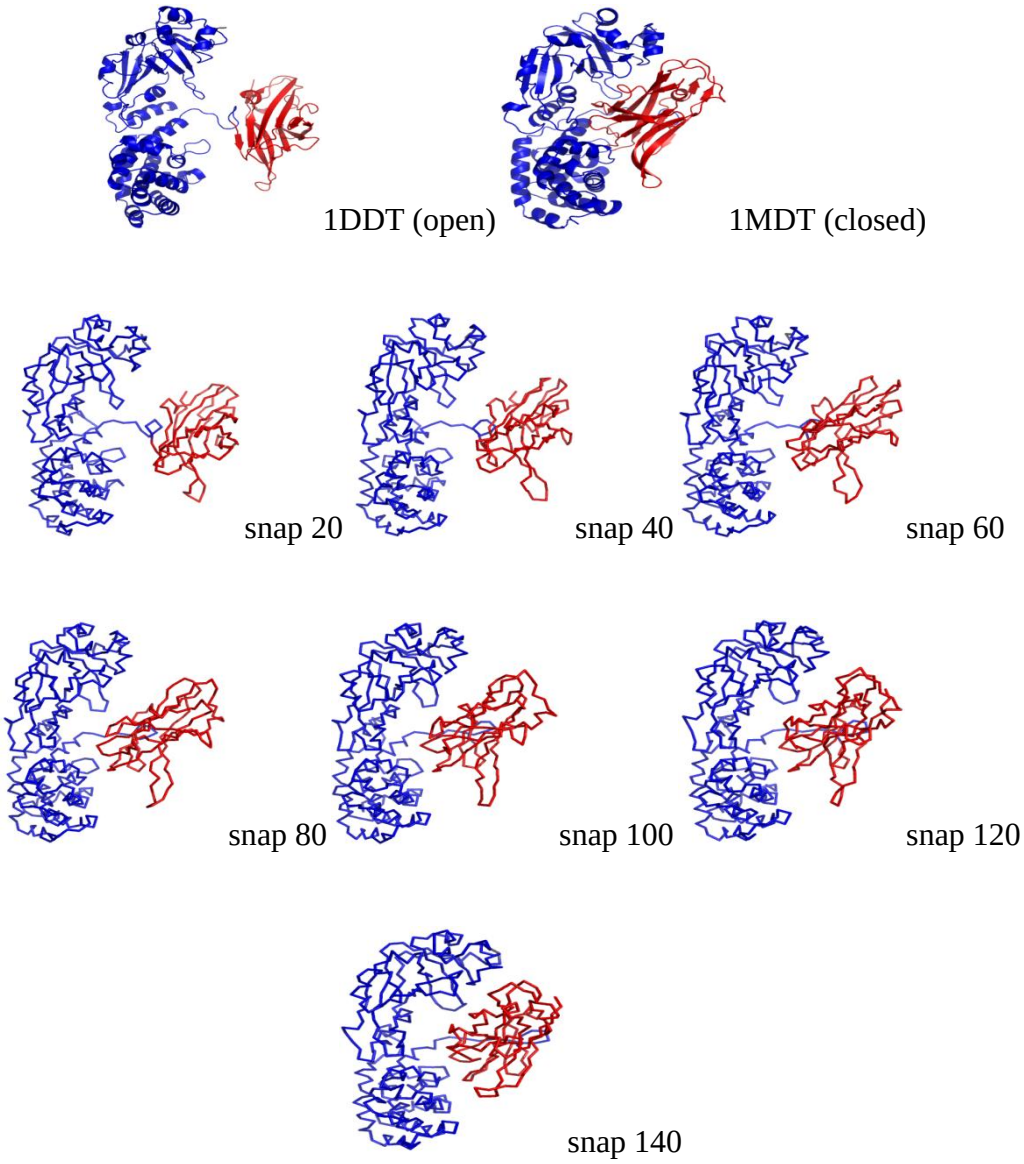


Figure A.11. Several intermediate structures obtained during forward ANM-MC simulation of monomeric diphtheria toxin transition from open to closed conformation.

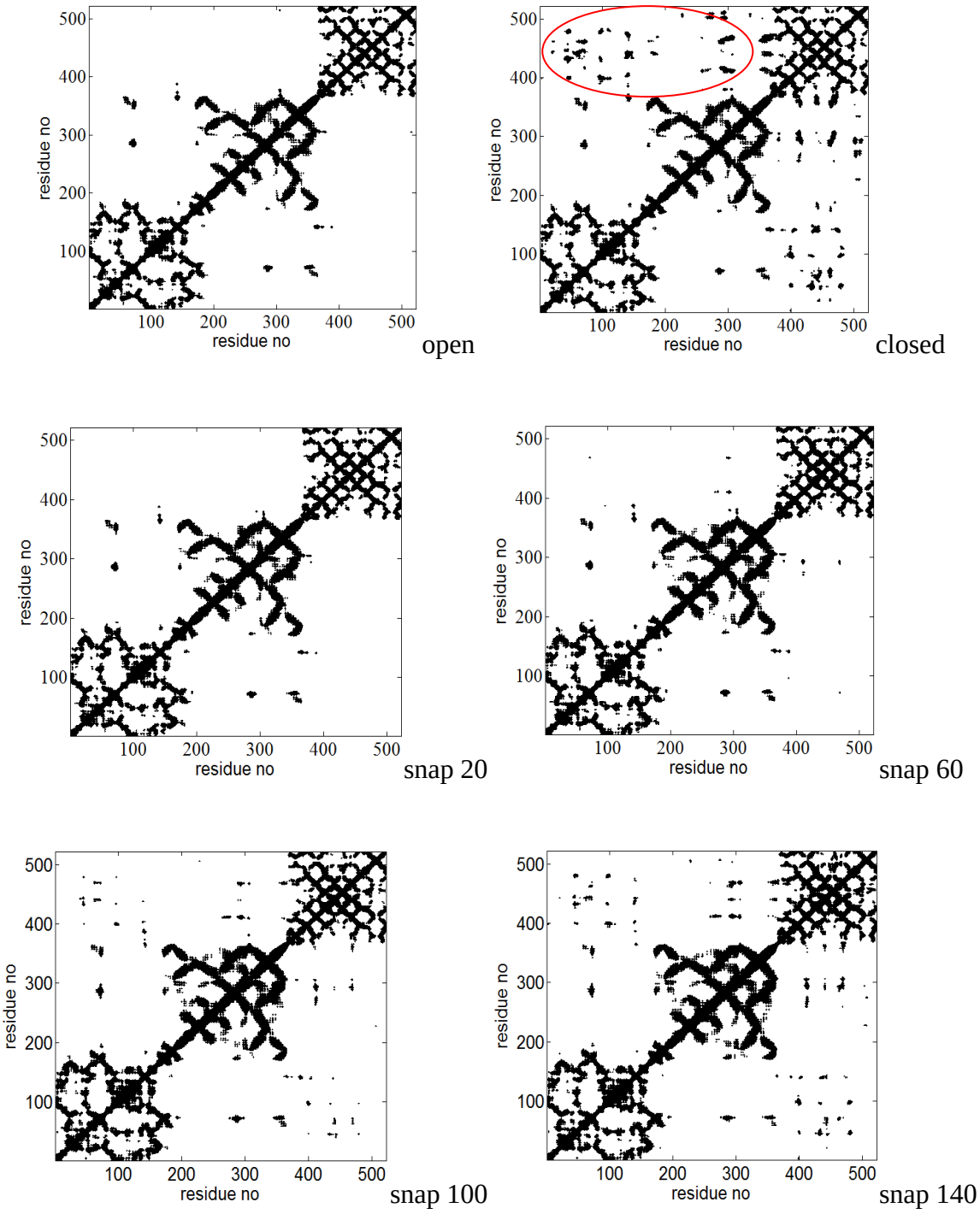


Figure A.12. Contact maps for open, closed forms of monomeric diphtheria toxin and for snapshots 20-60-100-140 for forward ANM-MC run.

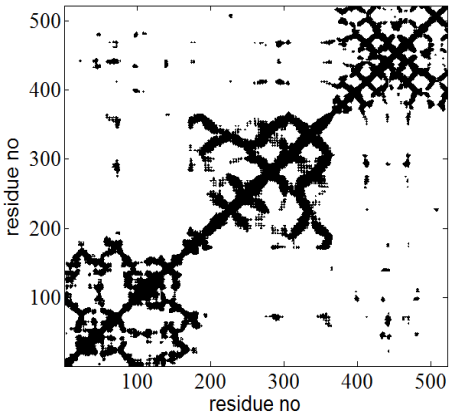


Figure A.13. Contact maps for dimeric diphtheria toxin and for snapshot 140 for forward ANM-MC run.

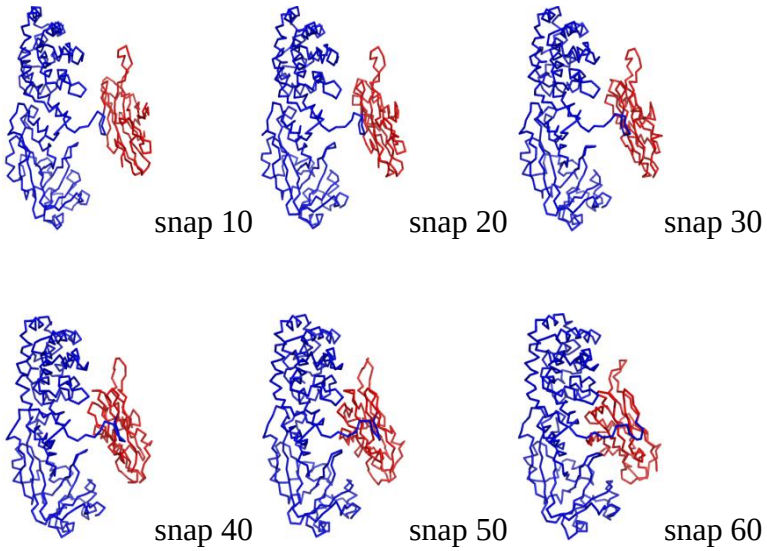


Figure A.14. Several intermediate structures obtained during forward TMC simulation of monomeric diphtheria toxin transition from open to closed conformation.

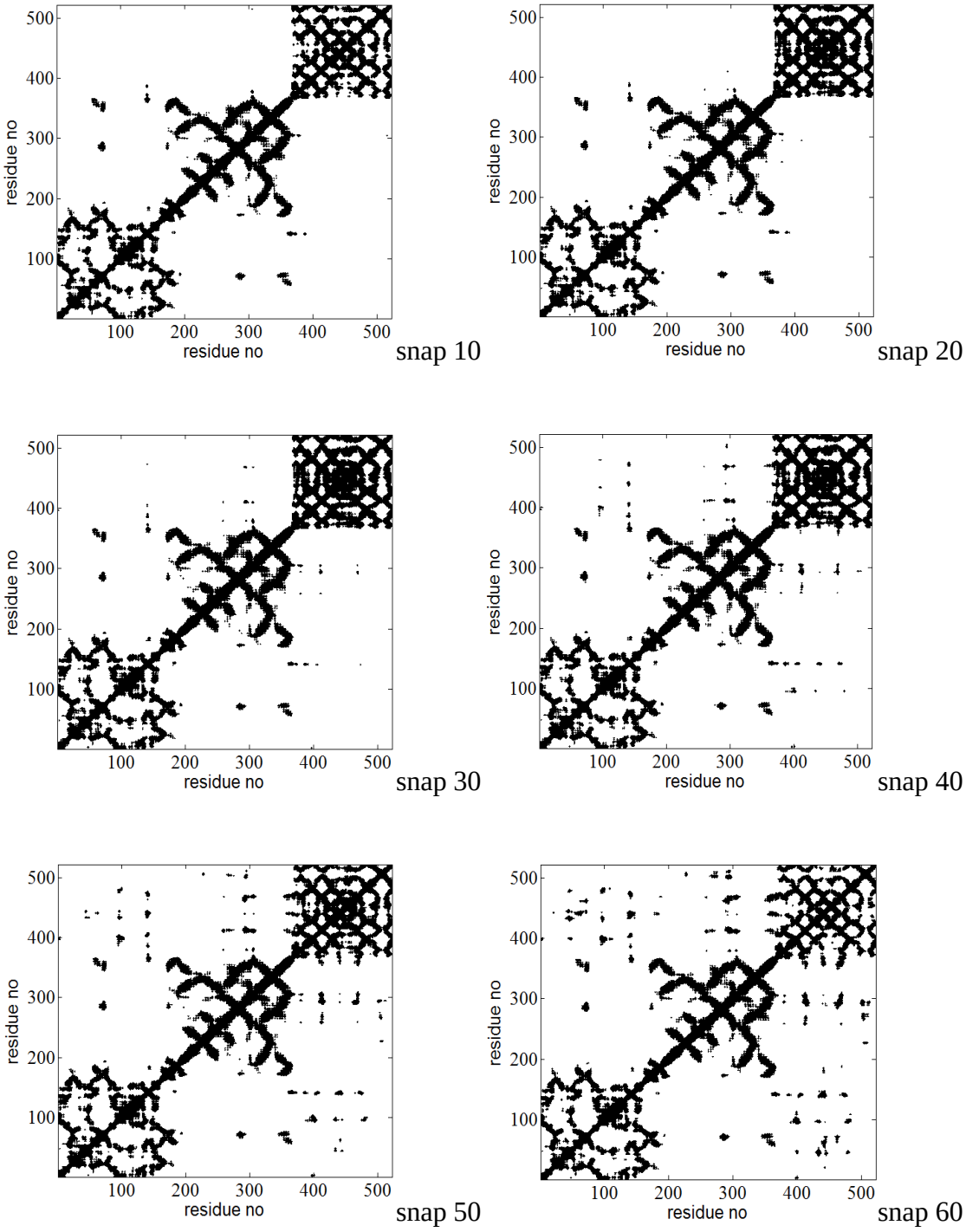


Figure A.15. Contact maps for monomeric diphtheria toxin and for snapshots 10-20-30-40-50-60 for forward TMC run.

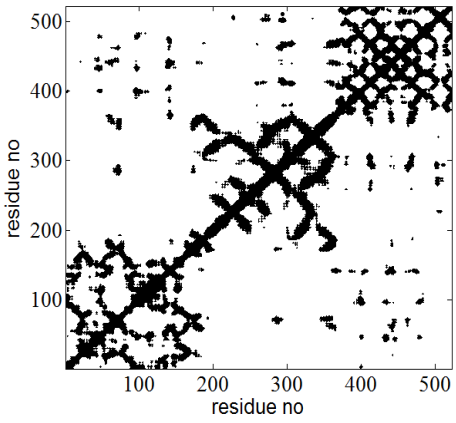


Figure A.16. Contact maps for dimeric diphtheria toxin and for snapshot 60 for forward TMC run.

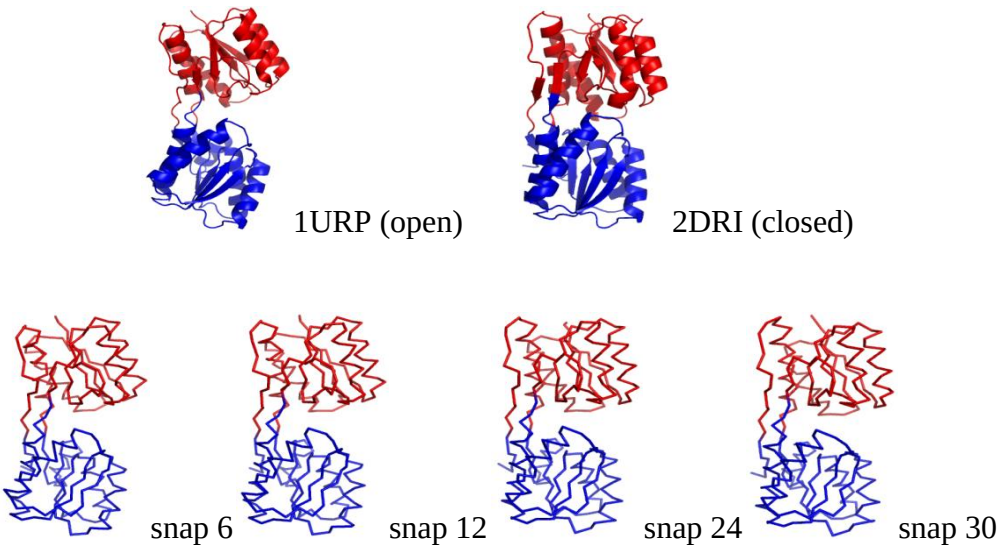


Figure A.17. Several intermediate structures obtained during forward ANM-MC simulation of D-Ribose Binding Protein transition from open to closed conformation.

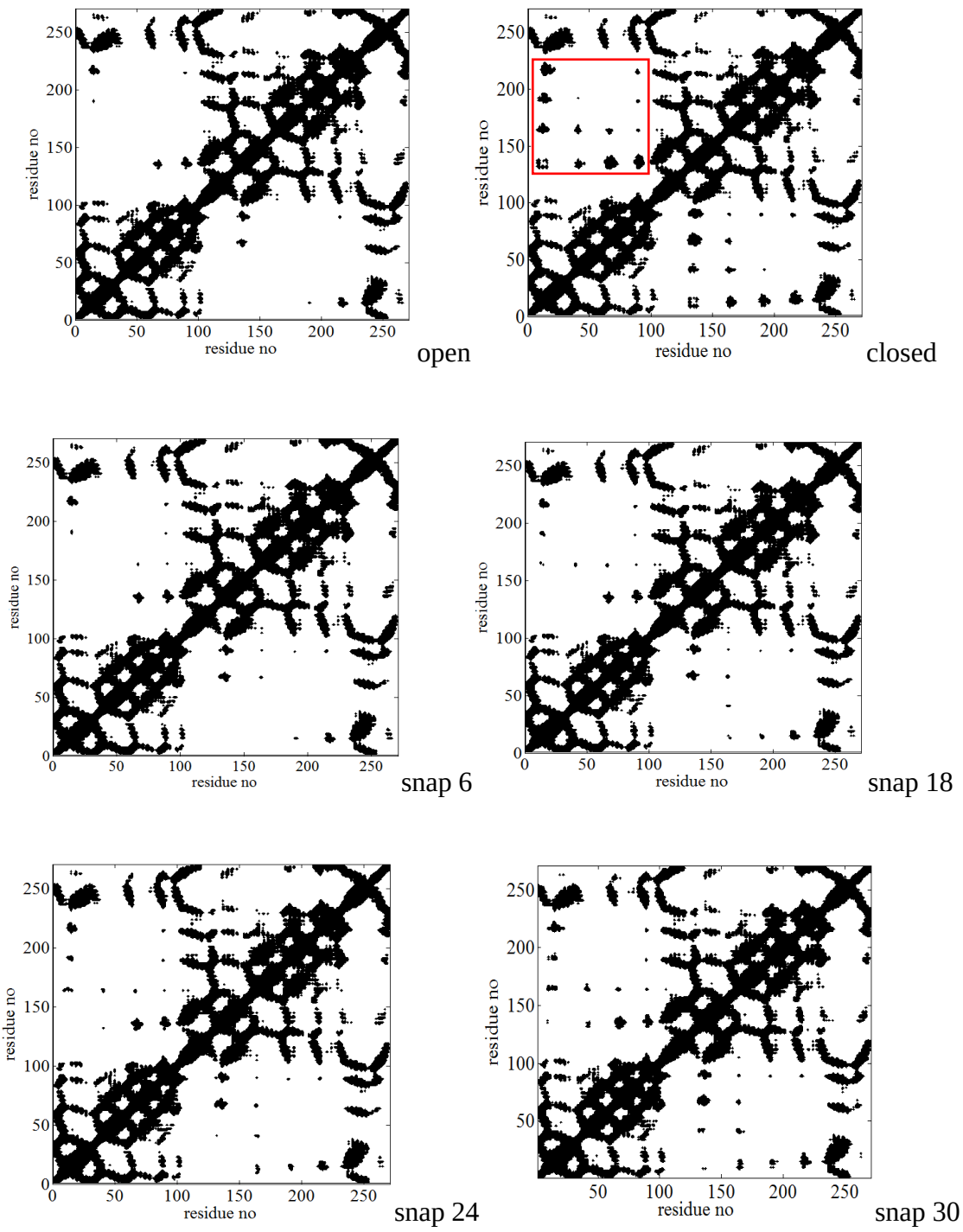


Figure A.18. Contact maps for open, closed forms of D-ribose Binding protein and for snapshots 6-18-24-30 for forward ANM-MC run.

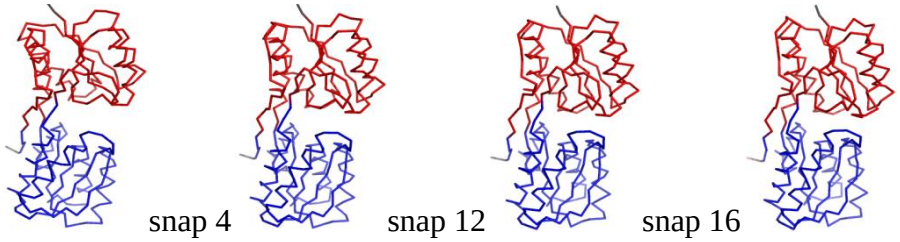


Figure A.19. Several intermediate structures obtained during forward TMC simulation of D-Ribose Binding Protein transition from open to closed conformation.

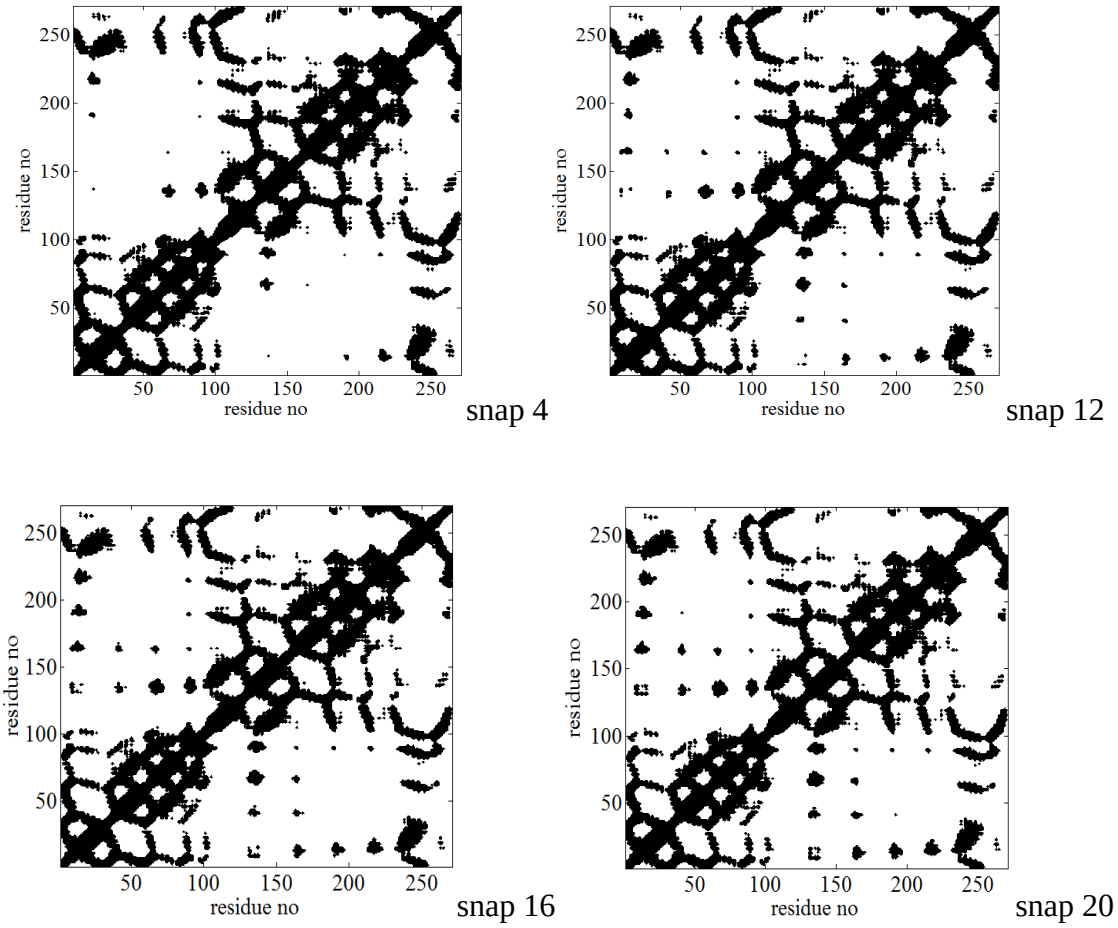


Figure A.20. Contact maps for D-ribose Binding protein and for snapshots 4-12-16-20 for forward TMC run.

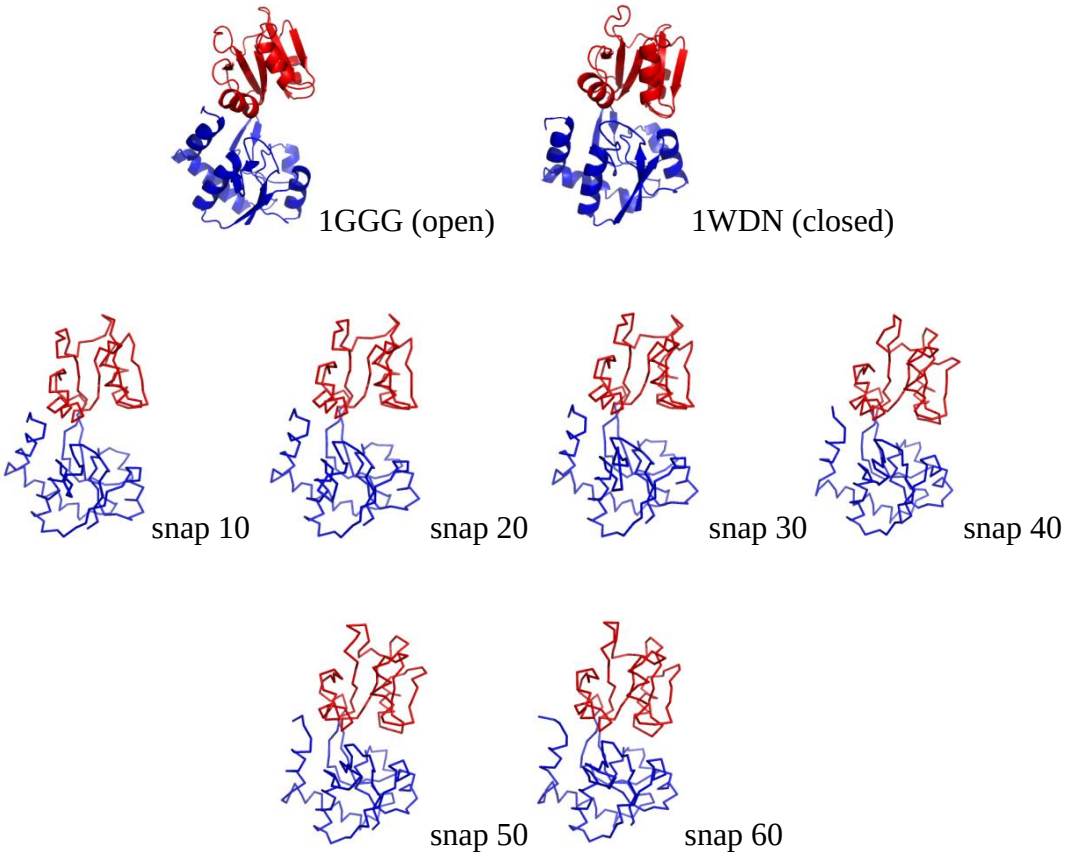


Figure A.21. Several intermediate structures obtained during forward ANM-MC simulation of GBP transition from open to closed conformation.

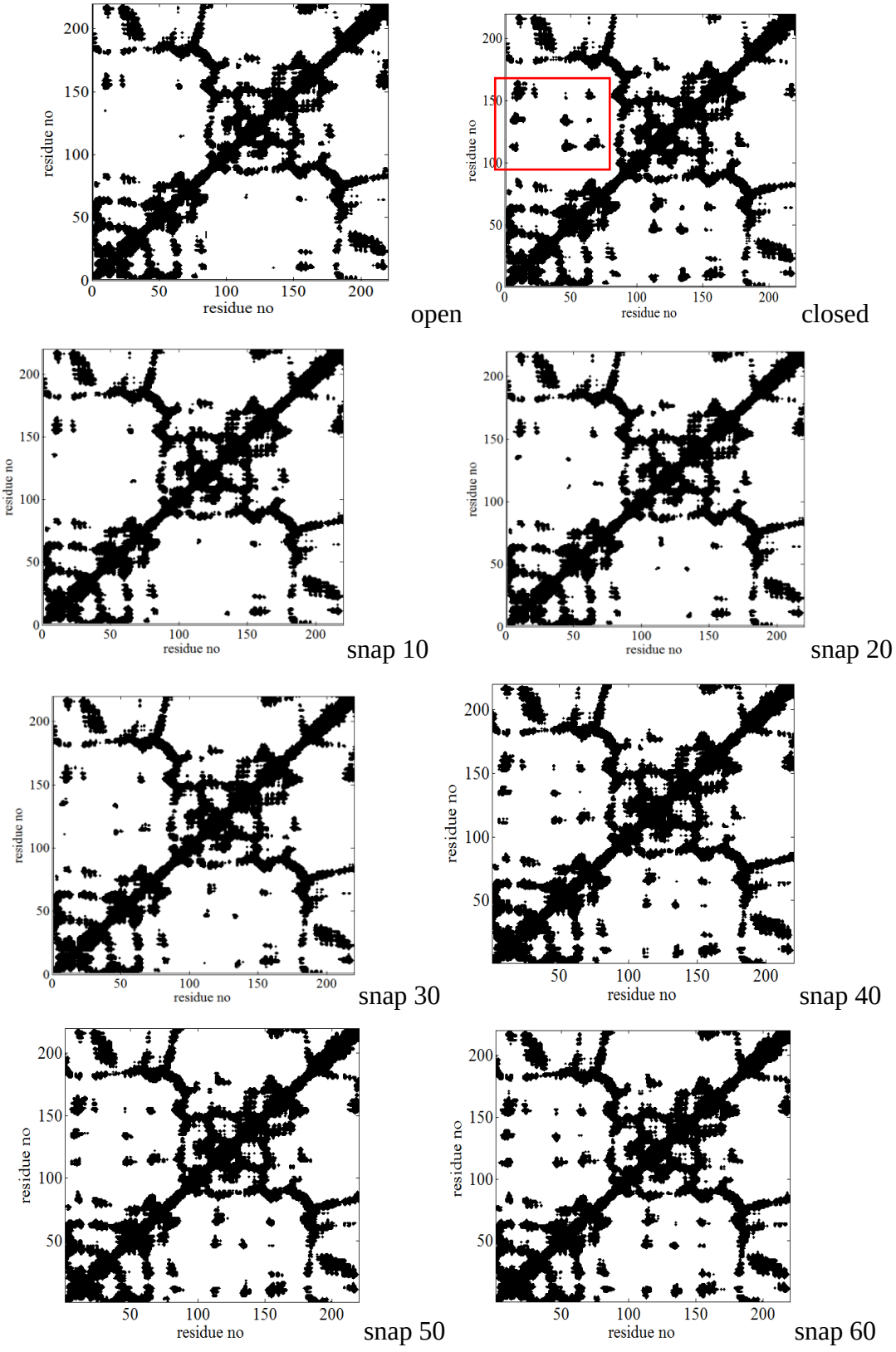


Figure A.22. Contact maps for open, closed forms of GBP and for snapshots 10-20-30-40-50-60 for forward ANM-MC run.

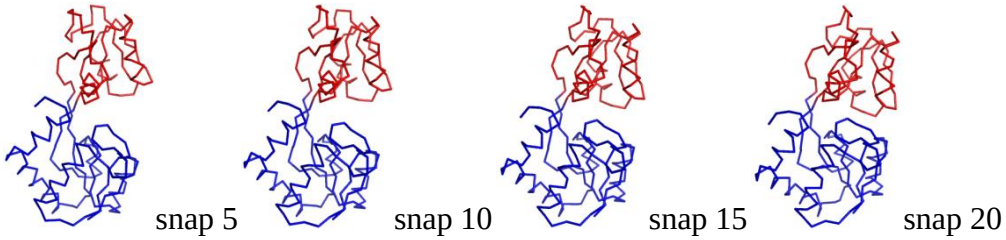


Figure A.23. Several intermediate structures obtained during forward TMC simulation of GBP transition from open to closed conformation.

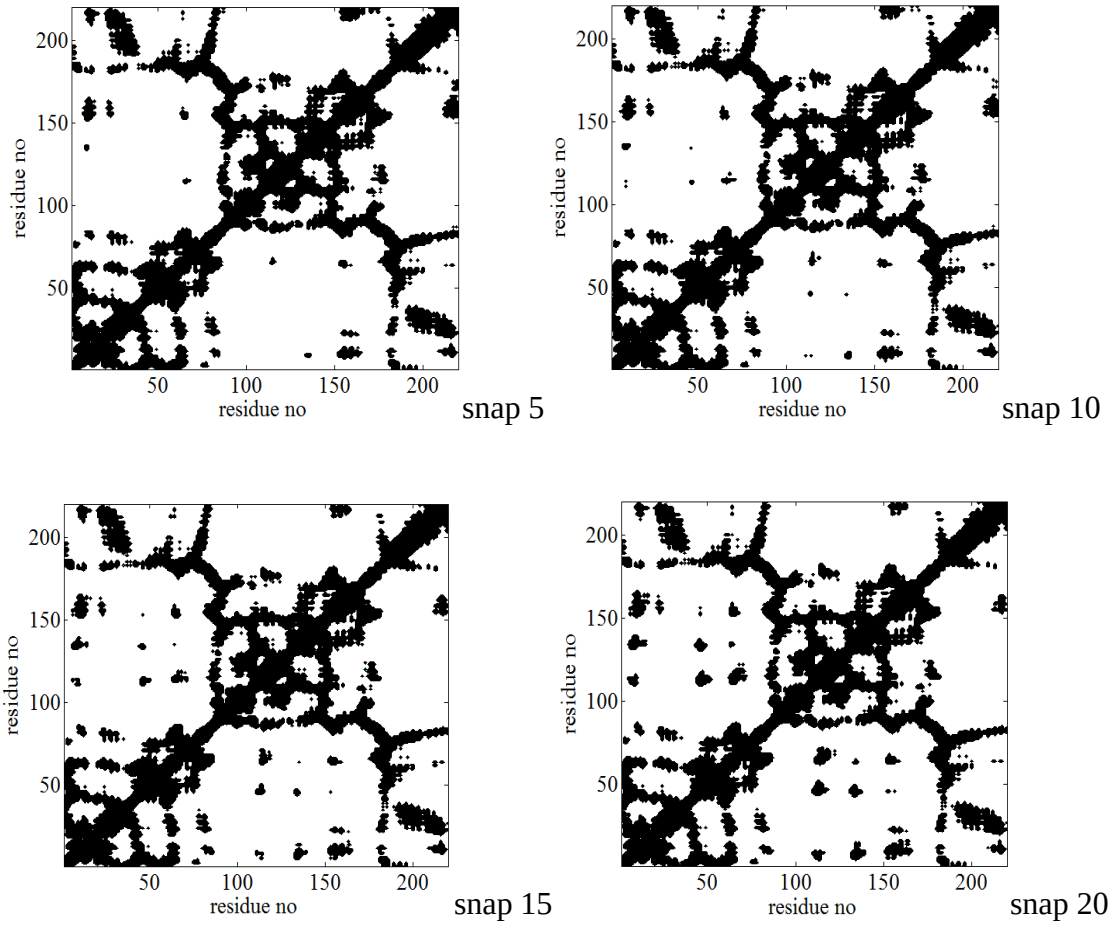


Figure A.24. Contact maps for GBP and for snapshots 5-10-15-20 for forward TMC run.

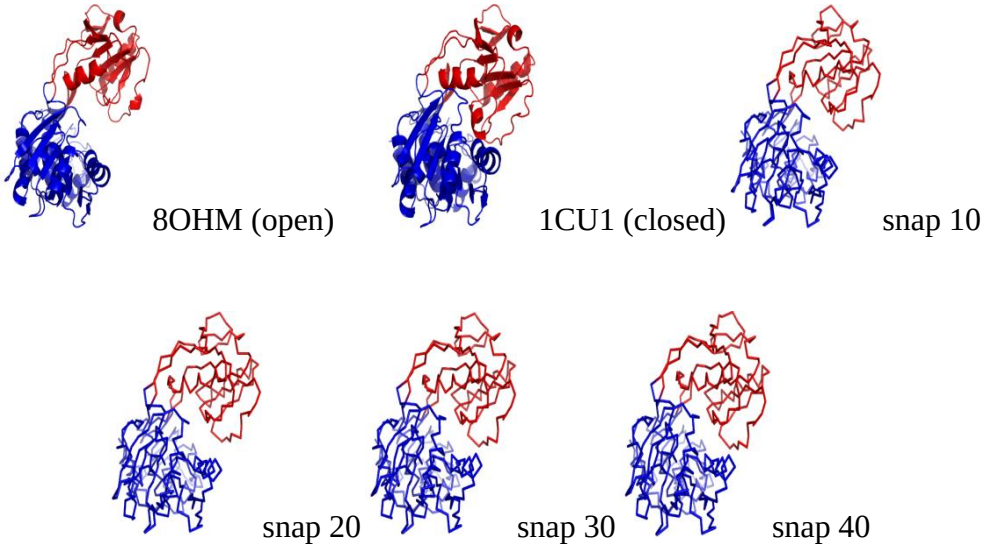


Figure A.25. Several intermediate structures obtained during forward ANM-MC simulation of Helicase transition from open to closed conformation.

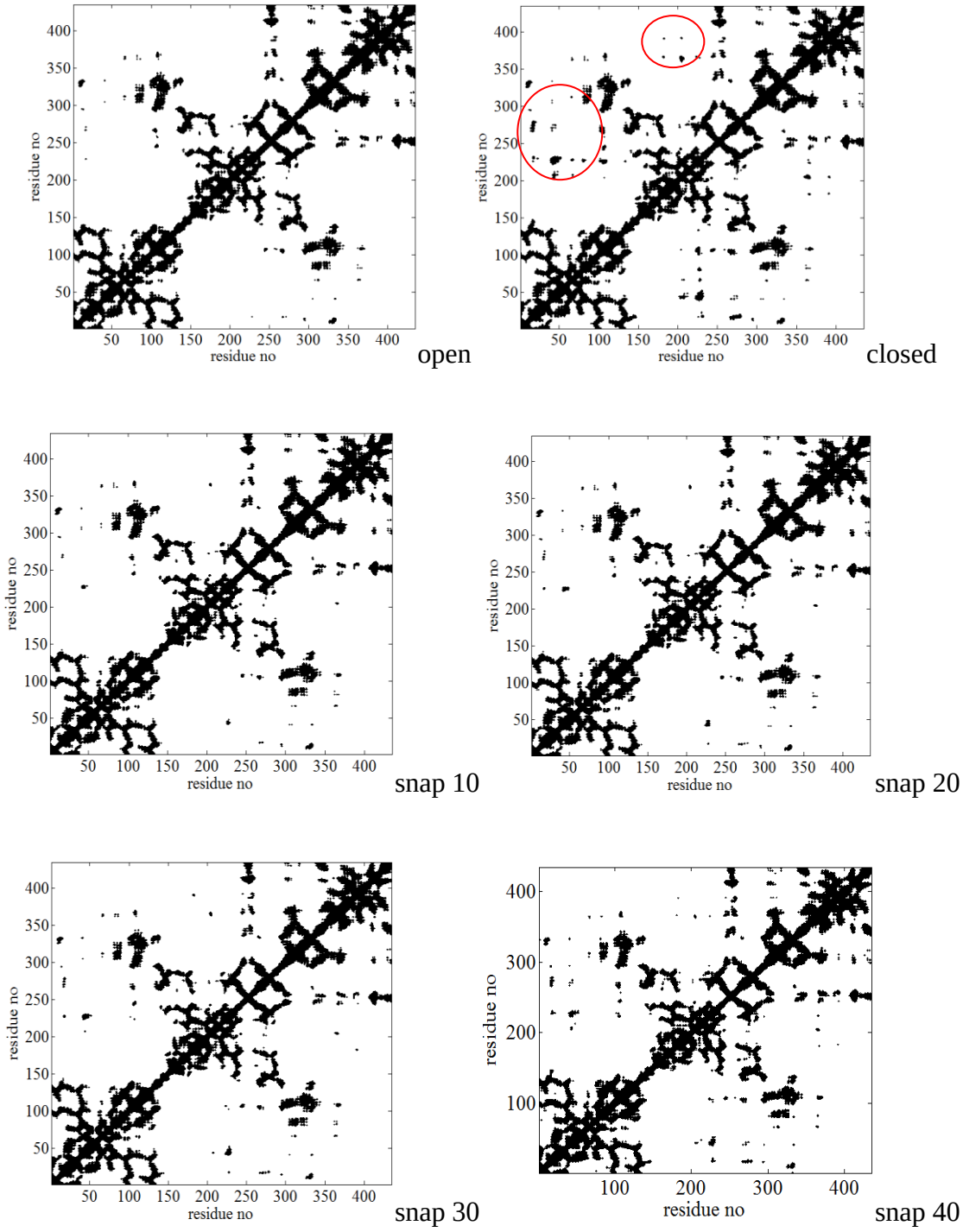


Figure A.26. Contact maps for open, closed forms of Helicase and for snapshots 10-20-30-40 for forward ANM-MC run.

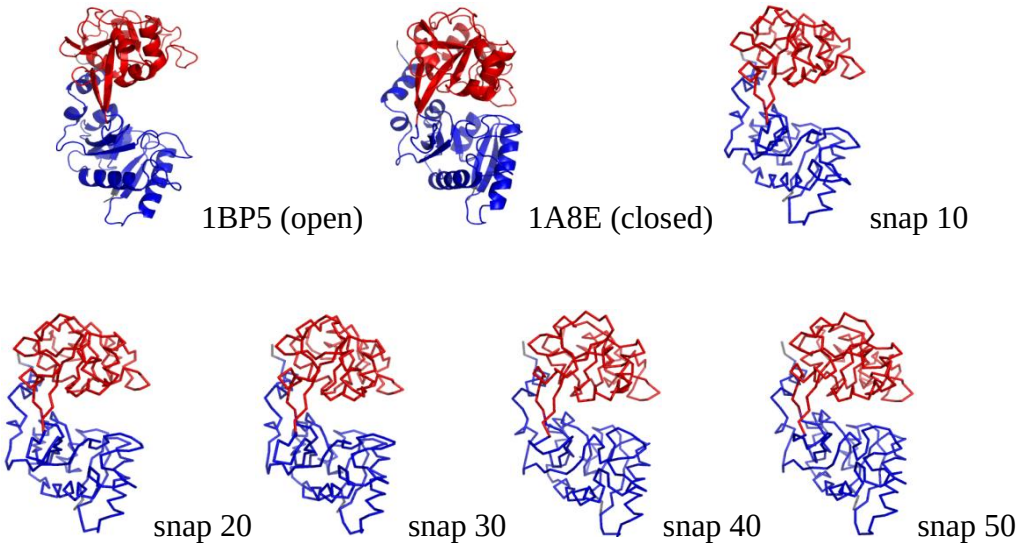


Figure A.27. Several intermediate structures obtained during forward ANM-MC simulation of HSTR transition from open to closed conformation.

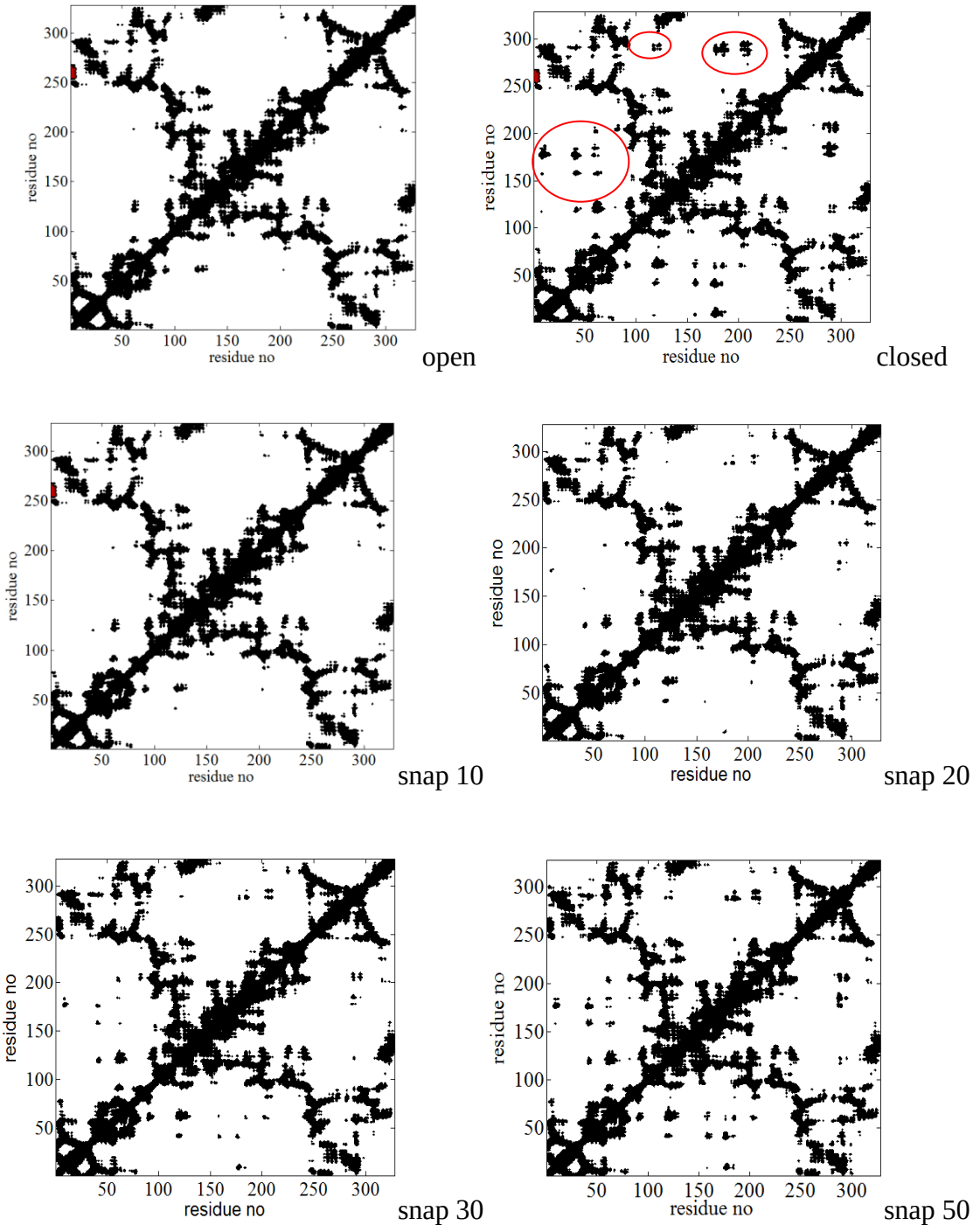


Figure A.28. Contact maps for open, closed forms of HSTR and for snapshots 10-20-30-40-50 for forward ANM-MC run.

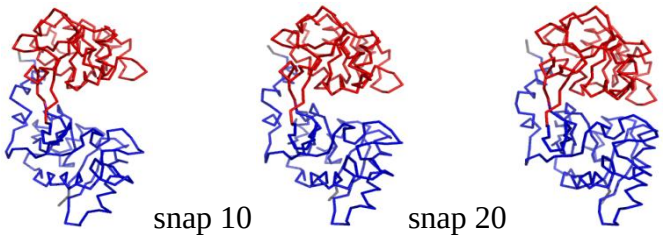


Figure A.29. Several intermediate structures obtained during forward TMC simulation of HSTR transition from open to closed conformation.

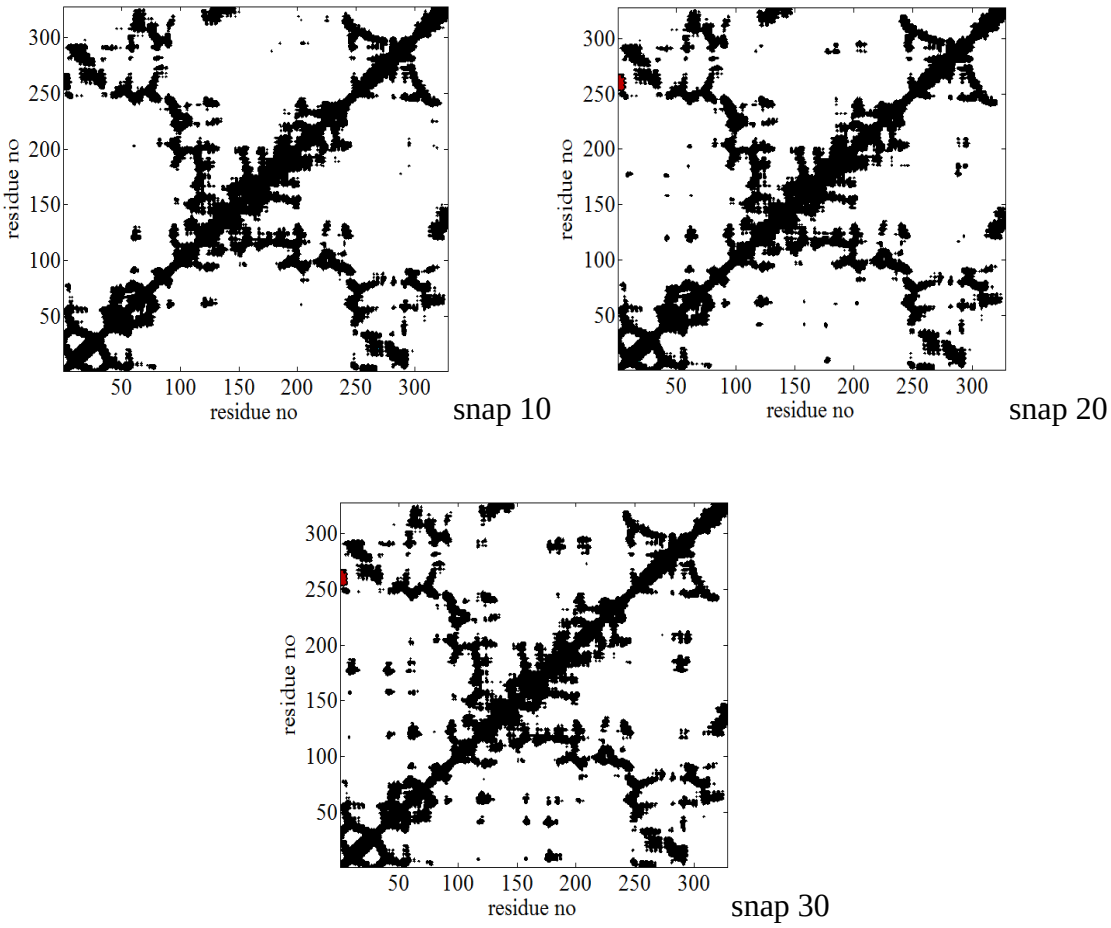


Figure A.30. Contact maps for HSTR and for snapshots 10-20-30 for forward TMC run.

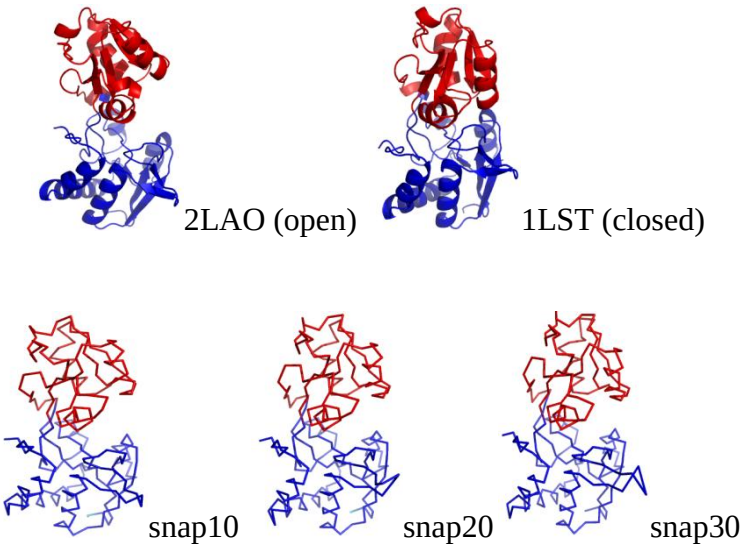


Figure A.31. Several intermediate structures obtained during forward ANM-MC simulation of LAO binding protein transition from open to closed conformation.

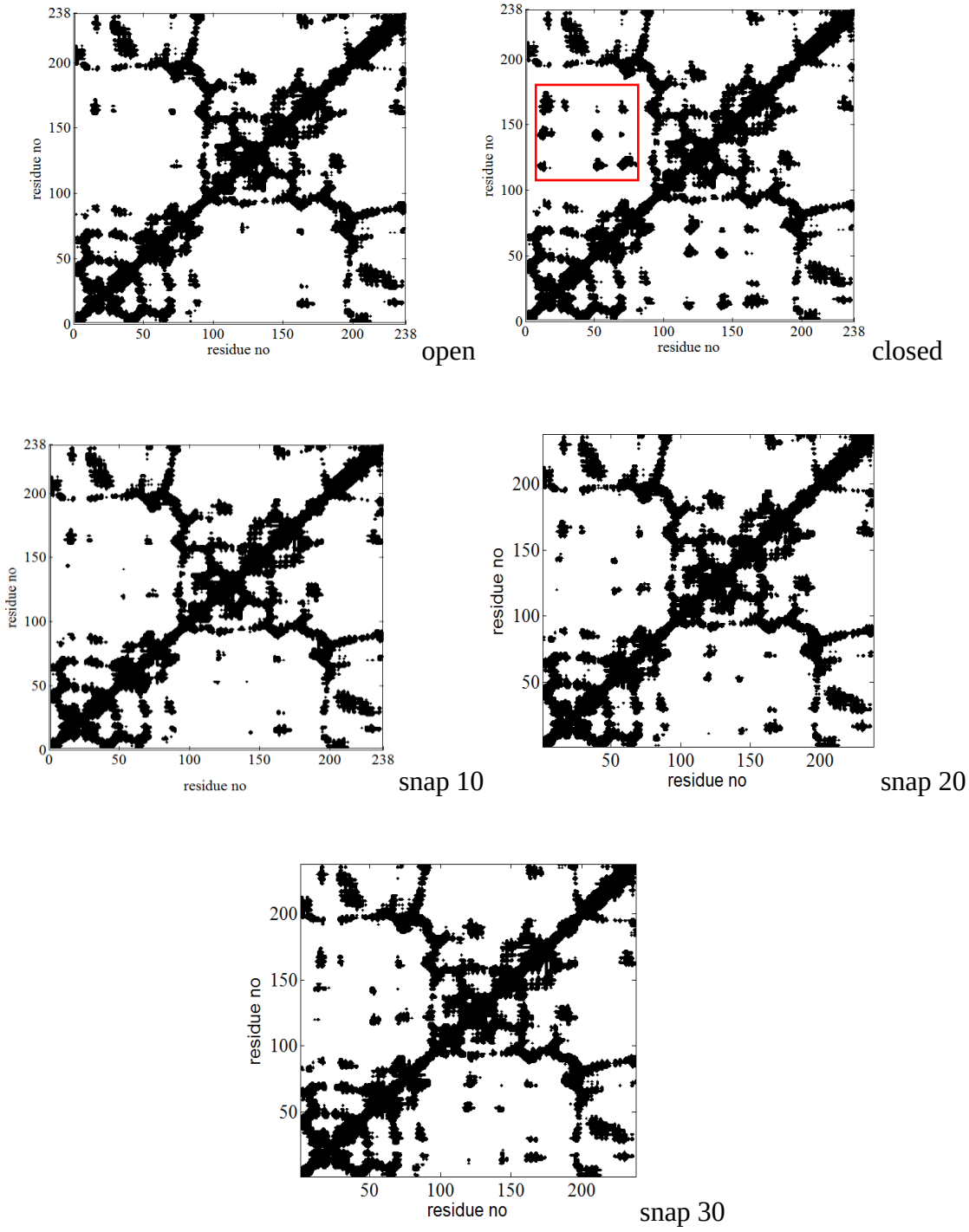


Figure A.32. Contact maps for LAO binding protein and for snapshots 10-20-30 for forward ANM-MC run.

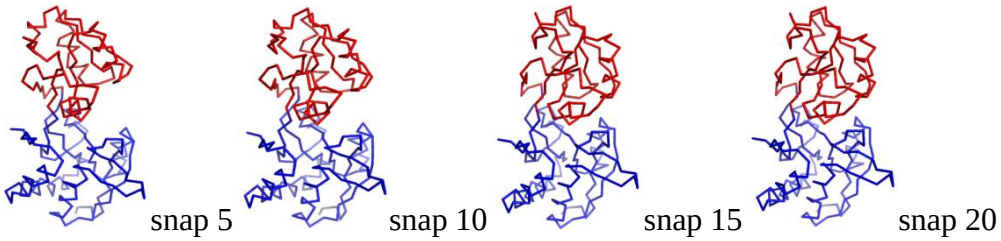


Figure A.33. Several intermediate structures obtained during forward TMC simulation of LAO binding protein transition from open to closed conformation.

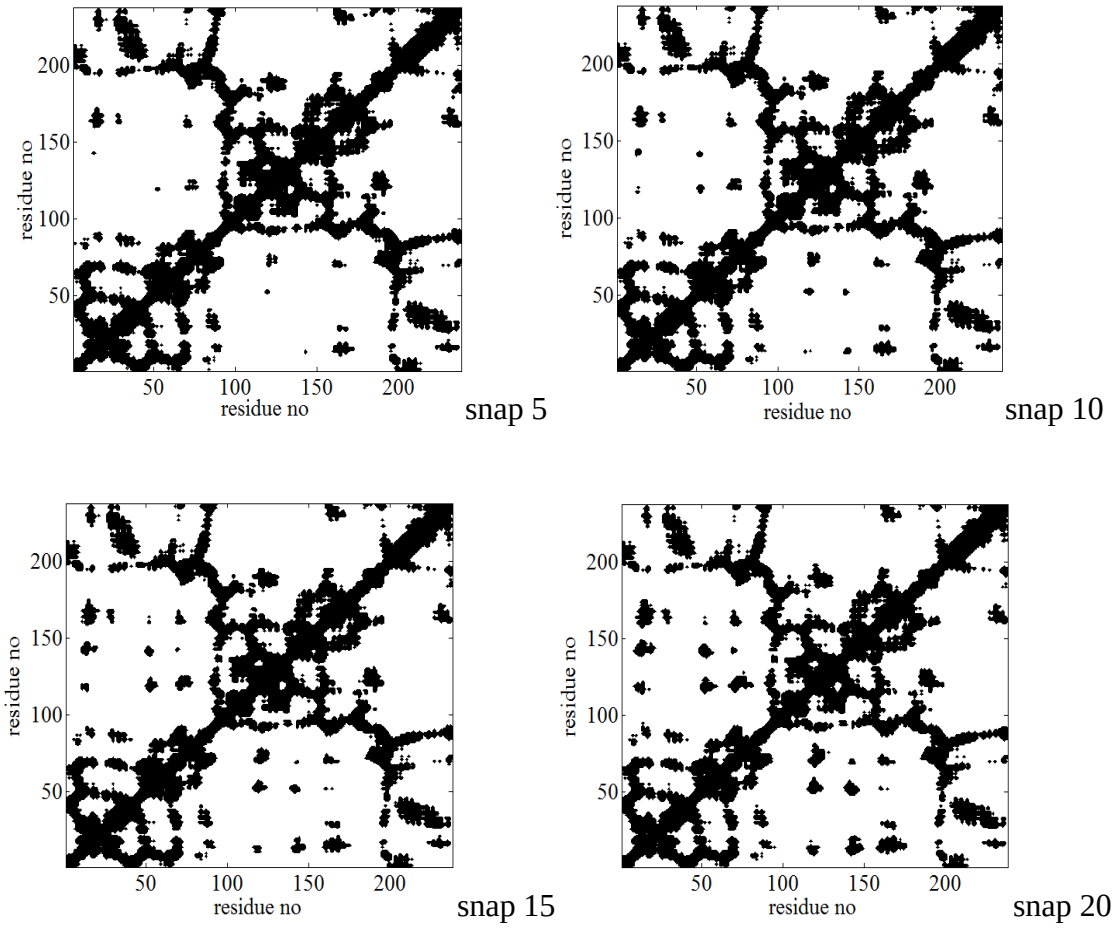


Figure A.34. Contact maps for LAO binding protein and for snapshots 5-10-15-20 for forward TMC run.

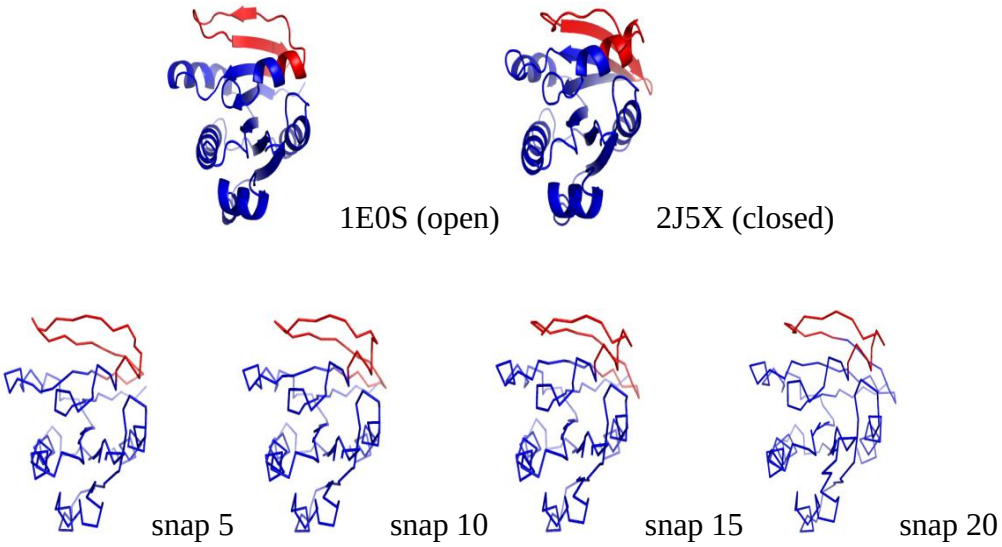


Figure A.35. Several intermediate structures obtained during forward ANM-MC simulation of Small G-protein Arf6 transition from open to closed conformation.

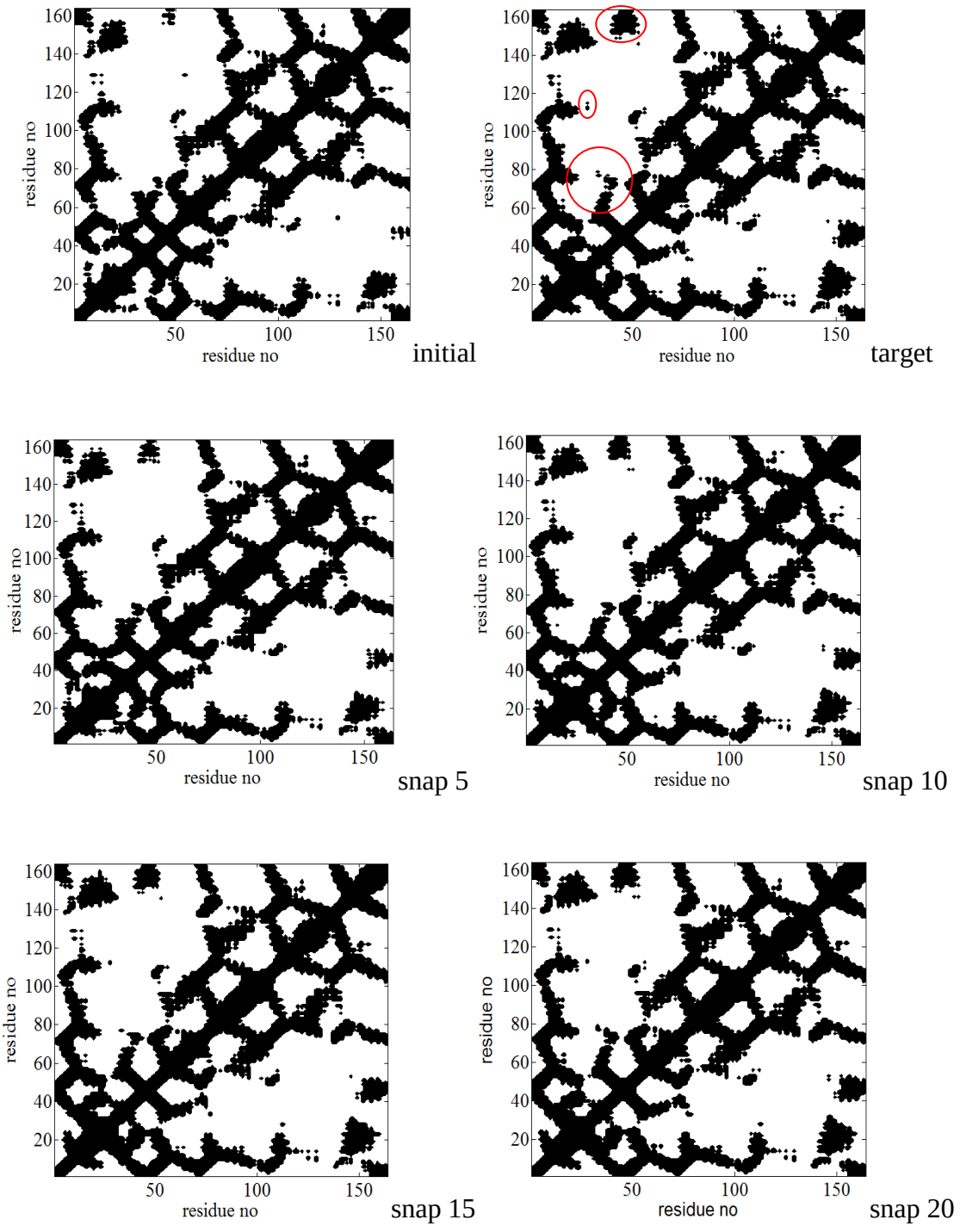


Figure A.36. Contact maps for initial, final forms of Small G-protein Arf6 and for snapshots 5-10-15-20 for forward TMC run.

REFERENCES

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

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

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

Bahar, I., M. Kaplan and R. L. Jernigan, 1997b, "Short-range conformational energies, secondary structure propensities, and recognition of correct sequence-structure matches", *Proteins*, Vol. 29, pp. 292-308.

Bahar, I., B. Erman, T. Haliloglu and R. L. Jernigan, 1997c, "Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations", *Biochemistry*, Vol. 36, pp. 13512-13523.

Bahar I. and R. L. Jernigan, 1998, "Vibrational Dynamics of Transfer RNAs: Comparison of the Free and Synthease Bound Forms", *Journal of Molecular Biology,* Vol. 281, pp. 871-885.

Bahar, I., 1999, "Dynamics of Proteins and Biomolecular Complexes", *Reviews in Chemical Engineering,* Vol. 15, pp. 319-347.

Bahar, I. and A. J. Rader, 2005, "Coarse-grained Normal Mode Analysis in Structural Biology", *Current Opinion in Structural Biology,* Vol. 15, pp. 586-592.

Barrientos, L., J. M. Louis, I. Botos, T. Mori, Z. Han, B. R. O'Keefe, M. R. Boyd, A. Wlodawer and A. M. Gronenborn, 2002, "The Domain-Swapped Dimer of Cyanovirin-N is in a Metastable Folded State: Reconciliation of X-Ray and NMR Structures", *Structure*, Vol. 10, pp. 673–686.

Bennett M. J., S. Choe and D. Eisenberg, 1994a, "Refined Structure of Dimeric Diphtheria Toxin At 2.0 A Resolution", *Protein Structure*, Vol. 3 (9), pp. 1444-1463.

Bennett M. J. and D. Eisenberg, 1994b, "Refined Structure of Monomeric Diphtheria Toxin At 2.3 A Resolution", *Protein Structure*, Vol. 3 (9), pp. 1464-1475.

Berteotti, A., A. Cavalli, D. Branduardi, F. L. Gervasio, M. Recanatini and M. Parrinello, 2008, "Protein Conformational Transitions: The Closure Mechanism of A Kinase Explored by Atomistic Simulations", *Journal of The American Chemical Society,* Vol. 131 (1), pp. 244-250.

Bhagavan, N. V., 2002, *Medical Biochemistry*, 4rd ed., Scientific American Books, New York.

Bjorkman A. J., R. A. Binnie, H. Zhang, L. B. Cole, M. A. Hermodson and S. L. J. Mowbray, 1994, "Probing protein-protein interactions. The ribose-binding protein in bacterial transport and chemotaxis", *Journal of Biological Chemistry*, Vol. 269 (48), pp. 30206- 11.

Bjorkman A. J. and S. L. Mowbray, 1998, "Multiple Open Forms of Ribose-binding Protein Trace the Path of its Conformational Change", *Journal of Molecular Biology,* Vol. 279, pp. 651-664.

Brooks, B. and M. Karplus, 1983, "Harmonic Dynamics of Proteins: Normal Modes and Fluctuations in Bovine Pancreatic Trypsin Inhibitor", *Proceedings of National Academy of Science USA,* Vol. 80, pp. 6571-6575.

Brooks, C. L., M. Karplus and B. M. Pettitt, 1988, "Proteins", John Wiley and Sons.

Cansu, S. and P. Doruker, 2008, “Dimerization affects collective dynamics of triosephosphate isomerase”, *Biochemistry*, Vol. 47, pp. 1358-1368.

Chattopadhyaya R., W. E. Meador , A. R. Means and F. A. Quiocho, 1992, "Calmodulin Structure Refined at 1.7 A Resolution", *Journal of Molecular Biology*, Vol. 228 (4), pp. 1177-92.

Cho H.S., N. C. Ha, L. W. Kang, K. M. Chung, S. H. Back, S. K. Jang and B. H. Oh, 1998, "Crystal structure of RNA Helicase from Genotype 1b Hepatitis C virus. A feasible mechanism of unwinding duplex RNA", *Journal of Molecular Biology*, Vol. 273 (24), pp. 15045-52.

Daggett, V., 2000, “Long timescale simulations”, *Current Opinion in Structural Biology*, Vol. 10, pp. 160–164.

DeLano, W. L., 2002, *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA, USA.

Delarue, M. and Y. H. Sanejouand, 2002, “Simplified normal mode analysis of conformational changes in DNA-dependent polymerases: the elastic network model”, *Journal of Molecular Biology*, Vol. 320, pp. 1011-1024.

Doruker, P., A. R. Atilgan and I. Bahar, 2000, “Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to α-Amylase Inhibitor”, *Proteins*, Vol. 40, pp. 512-524.

Doruker, P., R. L. Jernigan and I. Bahar, 2002, “Dynamics of large proteins through hierarchical levels of coarse-grained structures”, *Journal of Computational Chemistry* Vol. 23, pp. 119-127.

Doruker, P., L. Nilsson and O. Kurkcuoglu, 2006, “Collective Dynamics of *Eco*RI-DNA Complex by Elastic Network Model and Molecular Dynamics Simulations”, *Journal of Biomolecular Structure & Dynamics,* Vol. 24, pp. 1-15.

Dunten, P. and S. Mowbray, 1995, “Crystal structure of the dipeptide binding protein from Escherichia coli involved in active transport and chemotaxis”, *Protein Science,* Vol. 4, pp. 2327-2334.

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

Flory, P. J., 1969, “Statistical mechanics of chain molecules”, Interscience, New York.

Frank J., 1996, “Three-Dimensional Electron Microscopy of Macromolecular Assemblies”, Academic Press, New York.

Frauenfelder, H., S. G. Sligar and P. G. Wolynes, 1991, “The energy landscapes and motions of proteins”, *Science,* Vol. 254, pp. 1598-1603.

Gerstein, M., A. M. Lesk, and C. Chothia, 1994, “Structural Mechanisms for Domain Movements in Proteins”, *Biochemistry,* Vol. 33, pp. 6739-6749.

Gerstein, M. and W. Krebs, 1998, “A database of macromolecular motions”, *Nucleic Acids Research*, Vol. 26 (18), pp. 4280-4290.

Go, N., T. Noguti and T. Nishikawa, 1983, “Dynamics of a small globular protein in terms of low-frequency vibrational modes”, *Proceedings of National Acadademy of Science U.S.A.*, Vol. 80, pp. 3696-3700.

Goh, C. S., D. Milburn and M. Gerstein, 2004, “Conformational changes associated with protein-protein interactions”, *Current Opinion in Structural Biology*, Vol. 14, pp. 1-6.

Grant, B. J., A. A. Gorfe and J. A. McCammon, 2009, “Ras Conformational Switching: Simulating Nucleotide-Dependent Conformational Transitions with Accelerated Molecular Dynamics", *PLoS Computational Biology*, Vol. 5 (3): e1000325.

Greene, L.H., T. E. Lewis, S. Addou, A. Cuff, T. Dallman, M. Dibley, O. Redfern, F. Pearl, R. Nambudiry, A. Reid, I. Sillitoe, C. Yeats, J. M. Thornton and C. A. Orengo, 2007, "The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution*", Nucleic Acids Research*, Vol. 35, pp. D291-D297 (ISSN: 0305-1048).

Grimes, R. G, J. G. Lewis and H. D. Simon, 1994, “A shifted block Lanczos algorithm for solving sparse sysmetric eigenvalues problems”, *SIAM Journal on Matrix Analysis and Applications,* Vol. 15, pp. 228-272.

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

Haliloglu, T. and I. Bahar, 1998, “Coarse-grained simulations of conformational dynamics of proteins: Application to Apomyoglobin”, *Proteins*, Vol. 31, pp. 271-281.

Haliloglu, T., 1999, “Coarse-grained simulations of the conformational dynamics of proteins”, *Computational Theory of Polymer Science*, Vol. 9, pp. 255-260.

Hamelberg D., J. Mongan and J. A. McCammon, 2004, "Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules", *J. Chem. Phys.*, Vol. 120, pp. 11919–11929.

Hayward, S., A. Kitao and H. J. Berendsen, 1997, “Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme”, *Proteins,* Vol. 27 (3), pp. 425-437.

Hayward, S. and H. J. Berendsen, 1998, "Systematic Analysis of Domain Motions in Proteins from Conformational Change; New Results on Citrate Synthase and T4 Lysozyme", *Proteins, Structure, Function and Genetics*, Vol. 30, pp. 144.

He, J., Z. Zhang, Y. Shi and H. Liu, 2003, "Efficiently Explore the Energy Landscape of Proteins in Molecular Dynamics Simulations by Amplifying Collective Motions", *Journal of Chemical Physics*, Vol. 119, pp. 4005-40017.

Hsiao C. D., Y. J. Sun, J. Rose and B. C. Wang, 1996, "The crystal structure of glutamine-binding protein from Escherichia coli", *Journal of Molecular Biology*, Vol. 262 (2), pp. 225-242.

Janin, J. and S. J. Wodak, 1983, "Structural Domains in Proteins and their Role in the Dynamics of Protein Function", *Progress in Biophysics and Molecular Biology*, Vol. 42, pp. 21-78.

Jeffrey, P. D., M. C. Bewley, R. T. A. McGillivray, A. B. Mason, R. C. Woodworth and E. N. Baker, 1998, "Ligand-induced conformational change in transferrins: Crystal structure of the open form of N-terminal half-molecule of human transferrin", *Biochemistry*, Vol. 37 (40), pp. 13978-13986.

Kantarci-Carsibasi, N., T. Haliloglu and P. Doruker, 2008, "Conformational transition pathways explored by monte carlo simulation integrated with collective modes", *Biophysical Journal*, Vol. 95 (12), pp. 5862-5873.

Kantarci-Carsibasi, N., 2009, *Conformational Transitions of Proteins Explored by Monte Carlo Simulations Integrated with Collective Modes*, Ph.D. Thesis, Bogazici University.

Karplus, M., 1997, "The Levinthal paradox: yesterday and today", *Folding and Design*, Vol. 2, pp. 69-75.

Kasprzak, A. A., R. Takashi and M. F. Morales, 1988, "Orientation of actin monomer in the F-actin filament: Radial coordinate of glutamine-41 and effect of myosin

subfragment 1 binding on the monomer orientation", *Biochemistry,* Vol. 27, pp. 4512-4522.

Keskin, O., I. Bahar, D. Flatow, D. G. Covell and R. L. Jernigan, 2002, "Molecular Mechanisms of Chaperonin GroEL-GroES Function", *Biochemistry,* Vol. 41, pp. 491-501.

Keskin, O., 2007, "Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies", *BMC Structural Biology*, Vol. 7 (1), pp. 31-41.

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

Kim, M. K., G. S. Chirikjian and R. L. Jernigan, 2002b, " Elastic models of conformational transitions in macromolecules", *Journal of Molecular Graphics and Modelling*, Vol. 21, pp. 151-160.

Kim, M. K., R. L. Jernigan, and G. S. Chirikjian, 2005, "Rigid-cluster models of conformational transitions in macromolecular machines and assemblies", *Biophysical Journal,* Vol. 89, pp. 43-55.

Kitao, A. and N. Go, 1999, "Investigating Protein Dynamics in Collective Coordinate Space", *Current Opinion in Structural Biology*, Vol. 9, pp. 164-169.

Kong, Y., J. Ma, M. Karplus and W. N. Lipscomb, 2006, "The Allosteric Mechanism of Yeast Chorismate Mutase: A Dynamic Analysis", *Journal of Molecular Biology,* Vol. 356, pp. 237-247.

Koshland, D. E., 1958, "Application of a theory of enzyme specificity to protein synthesis", *Proceedings of National Academy of Science U. S. A.*, Vol. 44 (2), pp. 98-104.

Krebs, W. G. and M. Gerstein, 2000, “The morph server: The standardized system for analyzing and visualizing macromolecular motions in a database framework”, *Nucleic Acids Research*, Vol. 28 (8), pp. 1665-1675.

Krebs, W. G., V. Alexandrov, C. A. Wilson, N. Echols, H. Yu and M. Gerstein, 2002, “Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic”, *Proteins,* Vol. 48 (4), pp. 682-695.

Krillova, S., J. Cortes, A. Stefaniu and T. Simeon, 2008, “An NMA-guided path planning approach for computing large-amplitude conformational changes in proteins” *Proteins*, Vol. 70, pp. 131-143.

Kuboniwa H., N. Tjandra, S. Grzesiek, H. Ren, C. B. Klee and A. Bax, 1995, "Solution structure of calcium-free calmodulin", Nat. Struct. Biol., Vol. 2 (9), pp. 768-76.

Kurkcuoglu, O., P. Doruker and R. L. Jernigan, 2004, “Mixed Levels of Coarse-Graining of Large Proteins Using Elastic Network Model Succeeds in Extracting the Slowest Motions”, *Polymer*, Vol. 45, pp. 649-657.

Kurkcuoglu O, P. Doruker, T. Z. Sen, A. Kloczkowski and R. L. Jernigan, 2008, " The ribosome structure controls and directs mRNA entry, translocation and exit dynamics", *Phys. Biol.*, Vol. 5 (4), pp. 46005.

Kurkcuoglu O., Z. Kurkcuoglu, P. Doruker and R. L. Jernigan, 2009, "Collective dynamics of the ribosomal tunnel revealed by elastic network modeling", *Proteins*, Vol. 75 (4), pp. 837-45.

Kurt, N. and T. Haliloglu, 1999, “Conformational dynamics of Chymotrpsin inhibitor 2 by coarse-grained simulations”, *Proteins*, Vol. 37, pp. 454-464.

Kurt, N., T. Haliloglu and C. A. Schiffer, 2003, “Structure based prediction of potential binding and non-binding peptides to HIV-1 protease”, *Biophysical Journal*, Vol. 85, pp. 853-863.

Lakowicz, J. R., 1999, "Principles of Fluorescence Spectroscopy", Plenum Publishing Corporation, 2nd edition.

Leach, A. R., 2001, Molecular Modeling: Principles and Applications, Prentice Hall, 2nd edition.

Lei, M., M. I. Zavodszky, L. A. Kuhn and M. F. Thope, 2004, “Sampling Protein Conformations and Pathways”, *Journal of Computational Chemistry*, Vol. 25, pp. 1133-1148.

Lesk, A. M., 2001, *Introduction to Protein Architecture*, Oxford University Press, Oxford.

Levitt, M., C. Sander and P. S. Stern, 1983, “The normal modes of a protein: Native bovine pancreatic trypsin inhibitor”, *International Journal of Quantum Chemistry: Quantum Biology Symposium*, Vol. 10, pp. 181-199.

Li, X., O. Keskin, B. Ma, R. Nussinov and J. Liang, 2004, “Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking”, *Journal of Molecular Biology*, Vol. 344 (3), pp. 781-795.

Ma, B., M. Shatsky, H. J. Wolfson and R. Nussinov, 2002, “Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations”, *Protein Science*, Vol. 11 (2), pp. 184-197.

Ma, J., 2005, “Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes”, *Structure*, Vol. 13, pp. 373-380.

Macgillivray, R. T. A., S. A. Moore, J. Chen, B. F. Anderson, H. Baker, Y. Lou, M. Bewkley, C. A. Smith, M. E. P. Murphy, Y. Wang, A. B. Mason, R. C. Woodworth, G. D. Brayer and E. N. Baker, 1998, "Two high-resolution crystal structures of the recombinant N-lobe of human transferrin reveal a structural change implicated in iron release", *Biochemistry*, Vol. 37, pp. 7919-7928.

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

Marques, O. A., 1995, BLZPACK: Description and user's guide. Technical report TR/PA/95/30 CERFACS, Toulouse, France.

Marques, O. A., 2001, BLZPACK : User's guide. NERSC, Lawrence Berkely National Laboratory (http://crd.lbl.gov/~osni/marques.html#BLZPACK).

McCammon, A. and S. C. Harvey, 1987, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge.

Mouawad, L. and D. Perahia, 1996, "Motions in Hemoglobin Studied by Normal Mode Analysis and Energy Minimization: Evidence for the Existence of Tertiary T-Like, Quaternary R-Like Intermediate Structures", *Journal of Molecular Biology*, Vol. 258, pp. 393-410.

Mouawad, L., D. Perahia, C. H. Robert and C. Guilbert, 2002, "New insights into the Allosteric Mechanism of Human Hemoglobin from Molecular Dynamics Simulations", *Biophysical Journal*, Vol. 82, pp. 3224-3245.

Murzin, A. G., S. E. Brenner, T. Hubbard and C. Chothia, 1995, " SCOP: Structural Classification of Proteins", *J. Mol. Biol.*, Vol. 247, pp. 536-540.

Müller C. W. and G. E. Schulz, 1992, " Structure of the complex between adenylate kinase from Escherichia coli and the inhibitor Ap5A refined at 1.9 A resolution. A model for a catalytic transition state", *Journal of Molecular Biology*, Vol. 224 (1), pp. 159-177.

Müller C. W., G. J. Schlauderer, J. Reinstein and G. E. Schulz, 1996, "Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding", *Structure*, Vol. 4 (2), pp. 147-156.

Nickitenko A. V., S. Trakhanov and F. A. Quiocho, 1995, "2 Å resolution structure of DppA, a periplasmic dipeptide transport/chemosensory receptor", *Biochemistry*, Vol. 34 (51), pp. 16585-95.

Noe, F., S. J. Schwarzl, S. Fischer and J. C. Smith, 2003, "Computational tools for analyzing structural changes in proteins in solution", *Applied Bioinformatics*, Vol. 2 (3), pp. 11-17.

Norris G. E., B. F. Anderson and E. N. Baker, 1991, "Molecular replacement solution of the structure of apolactoferrin, a protein displaying large-scale conformational change", *Acta Cryst.*, Vol. 47, pp. 998-1004.

Oh, B., J. Pandit, C. Kangl, K. Nikaido, S. Gokcen, G. Ferro-Luzzi Ames and S. Kim, 1993, "Three-dimensional Structures of the Periplasmic Lysine/Arginine/Ornithine-binding Protein with and without a Ligand", *Journal of Biological Chemistry*, Vol. 268 (15), pp. 11348-11355.

Orengo, C. A., A. D. Michie, D. T. Jones, M. B. Swindells and J. M. Thornton, 1997, "CATH: A Hierarchic Classification of Protein Domain Structures.", *Structure*, Vol. 5, pp. 1093-1108.

Pasqualato, S., J. Menetrey, M. Franco and J. Cherfils, 2001, "The Structural GDP/GTP Cycle of Human Arf6", *EMBO reports*, Vol. 2, No. 3, pp. 234-238.

Perutz, M. F., 1989, “Mechanisms of cooperativity and allosteric regulation in proteins”, *Quarterly Review of Biophysics*, Vol. 22, pp. 139-236.

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

RCSB Protein Data Bank (pdb statistics http://pdbbeta.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/index.html).

Schlitter, J., M. Engels, P. Kruger and E. Jacoby, 1993, “Targeted Molecular Dynamics Simulation of Conformational Change-Application to T-R Transition in Insulin”, *Molecular Simulation*, Vol. 10, pp. 291-308.

Stryer, L., 1988, Biochemistry, New York: W. H. Freeman and Co., 3rd edition.

Sun, Y., J. Rose, B. Wang and C. Hsiao, 1998, "The Structure of Glutamine-binding Protein Complexed with Glutamine at 1.94 Å Resolution: Comparisons with other Amino Acid Binding Proteins", *Journal of Molecular Biology*, Vol. 278, pp. 219-229.

Tama, F. and Y. H. Sanejouand, 2001, “Conformational Change of Proteins Arising From Normal Mode Calculations”, *Protein Engineering*, Vol. 14, pp. 1-6.

Tama, F. and C. L. Brooks, 2002, “The Mechanism and Pathway of pH Induced Swelling in Cowpea Chlorotic Mottle Virus”, *Journal of Molecular Biology*, Vol. 318, pp. 733-747.

Tama, F. and C. L. Brooks, 2006, “Symmetry, form and shape: guiding principles for robustness in macromolecular machines”, *Annual Review of Biophysics and Biomolecular Strucure*, Vol. 35, pp. 115-133.

Tobi D. and I. Bahar, 2005, “Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state”, *Proceedings of National Academy of Science U. S. A.*, Vol. 102 (52), pp. 18908-18913.

Tsai, C. J., S. Kumar, B. Ma and R. Nussinov, 1999, “Folding funnels, binding funnels, and protein function”, *Protein Science*, Vol. 8, pp. 1181-1190.

Van der Vaart A. and M. Karplus, 2005, “Simulation of Conformational Transitions by the Restricted Perturbation-Targeted Molecular Dynamics Method”, *Journal of Chemical Physics*, Vol. 122, pp. 114903-114909.

Van der Vaart A. and M. Karplus, 2007, “Minimum Free Energy Pathways and Free Energy Profiles of Conformational Transitions Based on Atomistic Molecular Dynamics Simulations”, *Journal of Chemical Physics*, Vol. 126, pp. 164106-164123.

Vertessy B. G., V. Harmat, Z. Böcskei, G. Náray-Szabó, F. Orosz and J. Ovádi, 1998, "Simultaneous Binding Of Drugs With Different Chemical Structures To Ca2+-Calmodulin: Crystallographic And Spectroscopic Studies", *Biochemistry*, Vol. 37 (44), pp. 15300-10

Vitagliano, L., A. Merlino, A. Zagari and L. Mazzarella, 2000, "Productive and nonproductive binding to ribonuclease A: X-ray structure of two complexes with uridylyl (2',5') guanosine", *Protein Science*, Vol. 9, pp. 1217-1225.

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

Xu, C., D. Tobi and I. Bahar, 2003, “Allosteric Changes in Protein Structure Computed by a Simple Mechanical Model: Hemoglobin T↔R2 Transition”, *Journal of Molecular Biology*, Vol. 333, pp. 153-168.

Yao N., P. Reichert, S. S. Taremi, W. W. Prosise and P. C. Weber, 1999, "Molecular Views Of Viral Polyprotein Processing Revealed By The Crystal Structure Of The Hepatitis C Virus Bifunctional Protease-Helicase", *Structure*, Vol. 7 (11), pp. 1353-63.

Yildirim, Y. and P. Doruker, 2004, “Collective Motions of RNA Polymerases. Analysis of Core Enzyme, Elongation Complex and Holoenzyme”, *Journal of Biomolecular Structure & Dynamics*, Vol. 22 (3), pp. 267-280.

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

Zhang, J., C. Li, K. Chen, W. Zhu, X. Shen and H. Jiang, 2006, “Conformational Transition Pathway in the Allosteric Process of Human Glucokinase”, *PNAS*, Vol. 103, pp. 13368-13373.

Zheng, W. and B. R. Brooks, 2005a, “Normal-modes-based Prediction of Protein Conformational Changes Guided by Distance Constraints”, *Biophysical Journal,* Vol. 88, pp. 3109-3117.

Zheng, W. and B. R. Brooks, 2005b, “Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of elastic network model”, *Journal of Molecular Biology*, Vol. 346, pp. 745-759.

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

Zheng, W., B. R. Brooks and G. Hummer, 2007, “Protein conformational transitions explored by mixed elastic network models”, *Proteins*, Vol. 69, pp. 43-57.