CONFORMATIONAL TRANSITIONS OF PROTEINS EXPLORED BY MONTE CARLO SIMULATIONS INTEGRATED WITH COLLECTIVE MODES

by

Nigar Kantarcı Çarşıbaşı

B.S., Ch.E, Boğaziçi University, 2001

M.S., Ch.E, Boğaziçi University, 2003

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

Graduate Program in Chemical Engineering

Boğaziçi University

2009

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor, Prof. Pemra Doruker Turgut for her encouraging attitude and guidance. Her helpful comments and positive support is kindly appreciated all through my study. I also thank my thesis co-advisor Prof. Türkan Haliloğlu for her valuable suggestions.

Helpful comments of the thesis committee members: Prof. Viktorya Aviyente, Assist. Prof. Burak Alakent, Assist. Prof. Elif Özkırımlı Ölmez, and valuable time they have devoted reading and commenting on my thesis are appreciated as well. I would also like to acknowledge the financial supports from Bogazici University B.A.P. (06A508 and 04A502), DPT Project (03K120250), TUBITAK Projects (104M247,106M077), and COSBIOM FP6.

Deepest thanks are due to all my friends at Polymer Research Center (PRC) for their accompaniance and support in every way, whenever I needed. Special thanks are due to my collegues Özge Kürkçüoğlu Levitas, Sinem Özel, Burcu Aykaç Fas and Uğur Emekli and PRC staff Canan Dedeoğlu and Avni Ercan for their kind support especially during my pregnancy period. Special thanks are due to Arzu Uyar for her smiling face and encouraging support.

Heartfelt thanks are due to my family for their love, patience and understanding I felt all through my life. I wish to express my deepest gratitude to all of them especially to my mother Nermin Kantarcı and sister Didar Kantarcı. As always they were right there whenever I needed. Last but not least, I would like to express all my love to my husband Mustafa Nezih Çarşıbaşı for his never ending love, and patient support.

Finally, I would like to express my deepest love to my little princess Sena Nur Çarşıbaşı who is the most valuable gift given to me in the last periods of my thesis. This thesis was unforgettable with her participation, which put a new meaning to everything, and is indeed dedicated to her…

ABSTRACT

CONFORMATIONAL TRANSITIONS OF PROTEINS EXPLORED BY MONTE CARLO SIMULATIONS INTEGRATED WITH COLLECTIVE MODES

Elucidation of conformational transitions between open/closed or free/bound states of proteins sheds light on the mechanism of their action. In this thesis, a new technique (ANM-MC) is proposed, in which collective modes obtained from anisotropic network model (ANM) are used in conjunction with a Monte Carlo (MC) simulation approach to investigate conformational transition pathways and pathway intermediates of proteins. ANM-MC is applied to Adenylate Kinase (AK), hemoglobin, Human Serum Transferrin (HSTR), and Lysine/Arginine/Ornithine (LAO) binding protein. Target conformations are reached by root mean-square deviation (RMSD) of 2.27, 1.90, 1.81, and 1.40 Å for AK, hemoglobin, HSTR, and LAO-binding protein cases respectively. Intermediate snapshots seem as plausible pathway intermediates when compared with related x-ray structures. Targeted Monte Carlo (TMC) approach, which is a forcing algorithm towards the target, is utilized without the use of collective modes. Both ANM-MC and TMC can explore the sequence of events with an efficient yet realistic conformational search. ANM-MC is further improved to be applicable to proteins with unknown target conformations. In this technique, called RG-ANM-MC, starting from open conformation, transitional path is generated by selecting lowest energy conformations obtained by normal modes with decreasing radius of gyration (RG). Application of the method on AK, HSTR, and LAO-binding proteins reveal 3.18, 3.45, and 2.61 Å RMSD approach values to corresponding target states, respectively. RG-ANM-MC proves to be an efficient tool for proposing plausible closed states of proteins exhibiting hinge-like high amplitude collective motions. Conformational changes arising due to ligand binding are found to be intrinsic properties of binding protein, i.e. unliganded proteins possess a pre-existing fluctuation mechanism even in the absence of ligands. In both approaches (ANM-MC and RG-ANM-MC), lowest frequency modes are effective during transitions.

ÖZET

PROTEINLERDE KONFORMASYONEL GEÇİŞLERİN KOLEKTİF MODLAR VE MONTE CARLO SİMÜLASYON TEKNİKLERİ İLE İNCELENMESİ

Proteinlerde açık/kapalı veya serbest/bağlı yapı arası geçişlerin incelenmesi, fonksiyonel mekanizmaların aydınlanması açısından büyük önem taşımaktadır. Bu tezde, Anisotropik ağ modeli yardımıyla elde edilen kolektif modların Monte Carlo (MC) simülasyon tekniği ile birlikte kullanılmasıyla ANM-MC olarak adlandırılan bir yöntem geliştirilmiş ve bu yöntem konformasyonel geçiş yolizlerinin ve ara yapıların elde edilmesinde ve incelenmesinde kullanılmıştır. ANM-MC, Adenylate Kinase (AK), hemoglobin, human serum transferrin (HSTR) ve Lysine/Arginine/Ornithine (LAO) binding protein üzerinde uygulanmış ve bu proteinlerin kapalı (hedef) yapılarına sırasıyla 2.27, 1.90, 1.81, ve 1.40 Å kök ortalama kare sapması (RMSD) kadar yaklaşım sağlanabilmiştir. Simülasyon sonucu elde edilen ara yapılar, incelenen proteinlerle ilintili protein bilgi bankasından elde edilen kristal yapılarla karşılaştırılmış ve uygun ara yapı adayları oldukları gösterilmiştir. Kolektif hareketler kullanılmadan yapılan Hedeflendirilmiş Monte Carlo (TMC) tekniği de hedefe yaklaşabilmiş ve fiziksel olayların sıralarını göstermekte verimli bir yöntem teşkil etmiştir. Hedef yapıları bilinmeyen proteinlere de uygulanabilmesi adına geliştirilen RG-ANM-MC yöntemi ise düşük enerjili konformasyonlar ve proteinin kapanabilmesi için azalan jirasyon yarıçapı temel alınarak oluşturulmuştur. Bu yöntem ile, kapalı yapı ile ilgili hiçbir bilgi kullanılmadan AK, HSTR ve LAO-binding protein için hedef yapılara 3.18, 3.45 ve 2.61 Å RMSD kadar yaklaşım sağlanabilmiştir. RG-ANM-MC proteinlere, uygun kapalı yapılar önerebilecek bir yöntem teşkil etmektedir. Simülasyon sonuçları gösteriyor ki, bağlanma ile meydana gelen yapısal değişiklikler aslında proteinde var olan içsel bir özelliktir. Bir başka deyişle, bağlanan yapı yokluğunda da, protein serbest yapısıyla bu gibi konformasyonel geçişlere imkan verecek bir titreşimsel mekanizmaya sahiptir. Her iki yöntemde de (ANM-MC ve RG-ANM-MC) düşük titreşimli modlar ilgili geçişi elde etmekte etkili olmuştur.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABREVIATIONS

Å	Angstrom
A	Coordinate matrix of the initial structure
B	Coordinate matrix of the target structure
$\mathbf{B_{rot}}$	Rotated (superimposed) coordinates of target structure
B_i	Debye-Waller or temperature factor of site i
C^{α}	Alpha-carbon atom
D	Diagonal matrix
$\mathcal{H}$,	Hessian matrix
$\mathbf{H}_{ij}$	Super element matrix of the hessian matrix
$\mathbf{H}^{-1}$	Inverse of the super element matrix
h(x)	Heavyside step function
I	Identity matrix
k	Damping factor
k_B	Boltzmann's constant
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
Q	Target vector
r	random number
r_c	Cutoff radius
R	Rotation matrix
$\mathbf{R}_i$	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

$<\Delta \mathbf{R}_i^2>$	Mean-square fluctuation of site i
Δx	Strength of perturbation
T	Absolute temperature
t	Transpose
tr	Trace
U	Eigenvector
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
ACM	Amplified Collective Motion
AK	Adenylate Kinase
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
ANM	Anisotropic Network Model
Arg	Arginine
Asp	Aspargine

avg	Average
BB	Backbone-backbone
BLZPACK	Block Lanczos Package
C	Carboxyl terminus
com	Center of mass
Cryo-EM	Cryo-electron microscopy
det	determinant of a matrix
DF	Deformation factor
diag	diagonal of matirx
DNA	Deoxyribonucleic acid
E	Energy
ED	Essential dynamics
E.coli	Escherichia coli
ENM	Elastic Network Model
ENI	Elastic Network Interpolation
FRET	Fluorescence Resonance Energy Transfer
GNM	Gaussian Network Model
His	Histidine
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
MWC	Monod-Wyman-Changeux model
msf	mean square
N	Amino terminus
NMA	Normal Mode Analysis
NMP	Nucleoside monophosphate
NMR	Nueclar Magnetic Resonance

PDB	Protein Data Bank
PNM	Plastic Network Model
R2/R	Relaxed form
RG	Radius of gyration
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
SB	Sidechain-backbone
SC	Side-chain
SR	Short-range
SS	Sidechain-sidechain
T	Tense form
TMC	Targeted Monte Carlo
TMD	Targeted Molecular Dynamics
Tyr	Tyrosine

1. INTRODUCTION

Proteins carry out important biological functions in living organisms such as catalysis, regulation, and transportation. The function of a protein is closely related to the conformational ensemble accessible to its three dimensional (3-D) structure. Usually experimental techniques such as x-ray crystallography and nuclear magnetic resonance (NMR) are used to obtain detailed atomic structures of macromolecules. Recent developments in experimental approaches provide complementary information about protein conformations. Cryo-electron microscopy (cryo-EM), fluorescence resonance energy transfer (FRET), and single particle tracking provide low-resolution conformations and information about conformational transitions with no atomic details. However, they are useful for elucidating molecular mechanisms and thus biological functions of protein complexes (Kasprzak *et al.,* 1988; Frank, 1996; Branden and Tooze, 1999; Lakowicz, 1999).

Conformational changes between functional states of proteins are essential in biological activities such as opening and closing of ion channels or ligand binding to enzymes and receptors (Zheng *et al.,* 2007). Understanding these biological events relies on the elucidation of transitional complexes and pathways (Kim *et al.,* 2002; Lei *et al.,* 2004). NMR technique can be used to attain an average structure among an ensemble of conformations. The method explores highly populated conformations but fails to distinguish less populated intermediate conformations.

Although a large part of the current knowledge of conformational flexibility in proteins is derived from experimental data (especially x-ray crystallography and NMR), there is currently no experimental technique that allows monitoring of protein conformational changes at atomic resolution as a function of time. Hence, the use of computational molecular modeling techniques, namely molecular dynamics (MD) simulations, Monte Carlo (MC) sampling methods, normal mode analysis (NMA) or coarse-grained elastic network models (ENM) has gained considerable interest recently

(McCammon and Harvey, 1987; Kitao and Go, 1999; Lei *et al.,* 2004; Bahar and Rader, 2005; Ma, 2005).

MD and NMA techniques that make use of all-atom empirical potentials (Brooks and Karplus, 1983) are among the most popular techniques for simulating protein dynamics. These atomistic simulations can reach up to nanoseconds or at most a few microseconds, hence can not efficiently explore the protein conformational changes that are in the time range of microsecond-milliseconds to seconds. As a result, to overcome the time-scale limitations of atomistic approaches, coarse-grained approaches such as ENM (Brooks and Karplus, 1983; Bahar and Rader, 2005; Ma, 2005) and MC simulations (Metropolis and Ulam, 1949; Northrup and McCammon, 1980; Haliloglu and Bahar, 1998; Haliloglu, 1999; Kurt and Haliloglu, 1999) have become promising tools to model protein systems. Especially, recent studies rely on the success of ENM based approaches to describe 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.,* 2002, 2005; Xu *et al.,* 2003; Mouawad *et al,* 2002; Maragakis and Karplus, 2005; Krillova *et al.,* 2008).

The most widely adopted coarse-grained ENM approaches are the Gaussian network model (GNM) (Bahar *et al.*, 1997a, 1998, Haliloglu *et al.,* 1997; Bahar and Jernigan, 1998; Bahar, 1999) and the anisotropic network model (ANM) (Atilgan *et al.*, 2001). In both GNM and ANM, residues are assumed to undergo Gaussian-distributed fluctuations which are coupled by harmonic potentials. In GNM, these fluctuations are assumed to be isotropic, with no directional preferences, whereas in ANM they are anisotropic (Atilgan *et al.,* 2001). Low-frequency motions, i.e. collective motions can be described well even if there is no atomic detail. Hence, both GNM and ANM are shown to yield a successful description of a protein's internal motions. Latest computational studies (Kurkcuoglu *et al.*, 2004; Yildirim and Doruker, 2004; Kantarci *et al*., 2005) reported that GNM and ANM have been successful in reproducing collective modes and predicting atomic fluctuations for even large biological systems by using harmonic potentials between close-neighboring residues in the protein 3-D structure. Recent studies also revealed that hierarchical levels of coarse-graining considerably reduce the computational time,

moreover provide a realistic picture of harmonic motion of proteins and thus structure-function relationships (Doruker *et al.*, 2002). In this way, the analysis of even very large biological systems such as GroEL-GroES complex (Keskin *et al.*, 2002), ribosome complex (Wang *et al.*, 2004), RNA polymerase (Yildirim and Doruker, 2004), triosephosphate isomerase (Kurkcuoglu *et al.,* 2005; 2006) and restriction endonuclease *Eco*RI-DNA complex (Doruker *et al.*, 2006) could be accomplished. In most of the ENM based studies, the protein conformational transitions are explored by using the x-ray structures of the initial and final states. However, in many cases the initial state is known, whereas the end state, i.e. target structure is not available. Therefore, computational tools that predict the conformational changes even in the absence of final state information are of great interest (Zheng and Brooks, 2006). Moreover, most of these applications routinely perform successive deformations onto the initial structure (Mouawad and Perahia, 1996; Xu *et al.,* 2003) disregarding the fact that as the initial conformation is deformed, the eigenvectors of the initial structure become less accurate in representing the global motions of the new intermediate structures (Zheng and Brooks, 2006). Another point is that, ENM applications provide harmonic modes which illustrate the collective motions the protein exhibits, but does not give insight about the sequence of these physical events and the intermediate complexes and hence are insufficient in exploring transitional pathways.

The objective of this study is to develop a new approach by incorporating collective motions obtained from the elastic network model, ANM, into MC simulation technique so as to explore the conformational space of large biological systems and to investigate conformational transition pathways of proteins and their complexes in a computationally efficient way. In the present thesis, three methodologies are proposed, which are named as ANM-MC; RG-ANM-MC and Targeted Monte Carlo (TMC) simulation techniques for investigating the protein conformational transitions. As the names imply, the first two methods, ANM-MC and RG-ANM-MC (RG designates the radius of gyration consideration) are based on an iterative methodology composed of successive ANM and MC cycles and both techniques make use of collective modes. On the other hand, TMC does not use collective modes; instead it is based on a forcing algorithm towards the target state. The approaches developed in this thesis are aimed to prevail over the limitations of some previous applications by repeating NMA periodically throughout the simulation and

using the updated normal modes of the present state. Consequently, by repeated NMA followed by energy minimization using MC, the present methodologies could possibly generate a more feasible pathway between two conformations within short computational times. In ANM-MC methodology, first the normal modes of the initial structure are calculated by ANM. Then, the slow mode overlapping the desired transition is selected by calculating the overlap of each mode with the target direction which is found by superimposition of initial and final structures. The new structure is generated by deforming the initial conformation along the chosen mode. The energy of this new structure is then minimized by MC and the normal modes are recalculated. With this method, the transition between open-closed conformations of Adenylate kinase (AK), human serum transferrin (HSTR), and Lysine/Arginine/Ornithine (LAO) binding protein and tense (T)-relaxed (R2) forms of hemoglobin are investigated. In order to analyze the significance of using the normal modes, the ANM calculations are omitted from the ANM-MC algorithm and solely successive MC simulations are applied. This protocol, named as TMC forces the initial structure towards the target state along the desired direction.

Finally, ANM-MC technique is further improved to investigate the conformational transition pathways of protein systems with unavailable target structures as well. RG-ANM-MC is an extension of ANM-MC with an additional constraint introduced by the radius of gyration (RG). In RG-ANM-MC, similar iterative algorithm composed of ANM and MC cycles is also valid, but this time the new conformation is generated by making use of the energies of conformations generated by collective modes and their RG's. RG-ANM-MC is applied to AK, HSTR and (LAO)-binding protein. The plan of the present thesis is as follows: proteins, and their dynamics, protein conformational transitions and pathways together with the structural details of the protein systems studied in this thesis will be presented in the following chapter. Details regarding the elastic network model, i.e. ANM and MC simulation technique followed by the details about three developed protocols (ANM-MC, TMC, and RG-ANM-MC) can be found in the third chapter. The applications of the novel methodologies, namely ANM-MC, TMC, and RG-ANM-MC, on various protein systems are presented in chapters four and five and six, respectively. Finally, the conclusions and recommendations will follow.

2. PROTEINS: STRUCTURE, FUNCTION, DYNAMICS AND CONFORMATIONAL TRANSITIONS

2.1. Protein Structure and Function

Proteins are macromolecules constructed from linear sequences of amino acids. They perform many functions essential to life. There are twenty common naturally occurring amino acids, which are linked together via peptide (amide) bonds to form a polypeptide chain (Figure 2.1). All the naturally occurring amino acids have the same relative stereochemistry at the alpha-carbon. The side chains have different sizes, shapes, hydrogen bonding capabilities and charge distributions, which enable proteins to display the required biological functions (Leach, 2001).

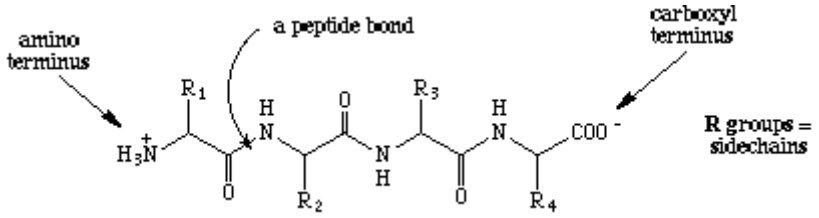


Figure 2.1. Aminoacids are linked via peptide bonds to give a polypeptide chain

Every function in the living cell depends on proteins. Understanding the function of a protein is fundamental for gaining insight into many biological processes. Among the important functions in which the proteins are involved several examples may be given such as: motion and locomotion of cells and organisms; the catalysis of all biochemical reactions; and the transport of materials in body fluids. Besides, the proteins constitute the structure of cells, and the extracellular matrix they are embedded in. Moreover, the receptors for hormones and other signaling molecules and the transcription factors that turn genes on and off to guide the differentiation of the cell are all made up of proteins (Stryer, 1988).

The details of a protein sequence are stored in the code of a gene. Through the processes of transcription and translation, a cell reads the genetic information and uses it to

construct the protein. In many cases, the resulting protein is then chemically altered (post-translational modification) before becoming functional. It is very common for proteins to work together to achieve a particular function, and often physically associate with one another to form a complex (Stryer, 1988).

The first x-ray structures revealed that proteins did not adopt regular structures but were more complex. Biochemists refer to four distinct aspects of a protein's structure. These are called the primary, secondary, tertiary and quaternary structures. The primary structure represents the amino acid sequence. The secondary structure designates the highly patterned sub-structures, i.e. helices and beta sheets that are formed by hydrogen bonding.

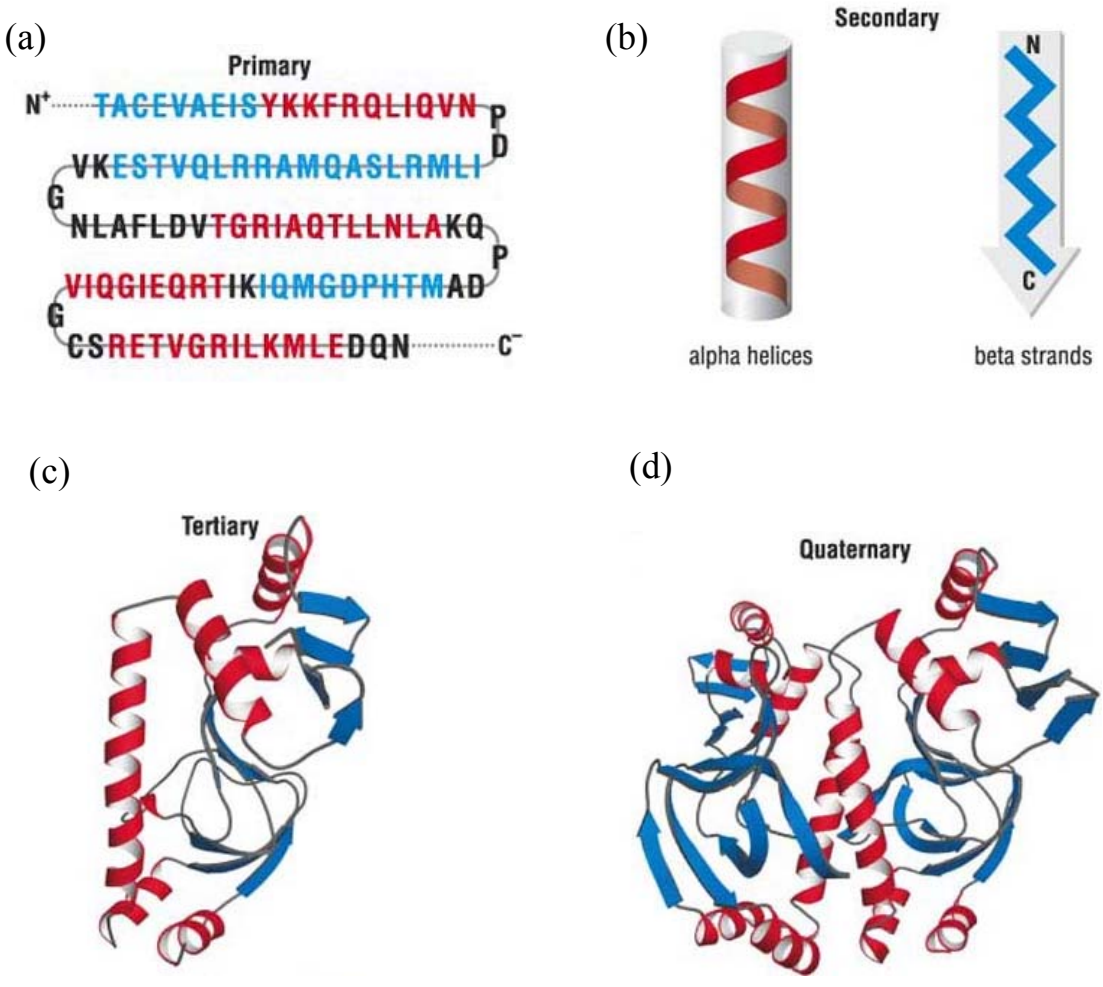


Figure 2.2. Primary, secondary, tertiary, and quaternary structures of proteins (Petsko and Ringe, 2004)

There can be many different secondary motifs present in one single protein molecule. The tertiary structure represents the overall shape of a single protein molecule, i.e. indicates the spatial relationship of the secondary structural motifs to one another. Tertiary structure is primarily formed by hydrophobic interactions, but hydrogen bonds, ionic interactions, and disulfide bonds are usually involved too. Finally, the Quaternary structure is associated with the shape or structure that results from the union of more than one protein molecule. It is usually called protein subunits in this context, which function as part of the larger assembly or protein complex (Alberts et al., 1994).

2.1.1. Experimental Techniques to Identify Protein Structure

X-ray crystallography and NMR are the most widely used experimental methods to obtain detailed (3-D) structural information about proteins. X-ray crystallography is a method for determining the arrangement of atoms within a crystal. As a beam of x-rays strikes a crystal and scatters into many different directions, a crystallographer can produce a 3-D picture of the density of electrons within the crystal. From this electron density, the mean positions of the atoms in the crystal can be determined, as well as their chemical bonds, their disorder and various other information. Crystal structures of proteins began to be solved in the late 1950s, beginning with the structure of sperm whale myoglobin (Kendrew et al., 1958) Since that date, nearly 40000 x-ray crystal structures of proteins, nucleic acids and other biological molecules have been determined. About 82% of all structures present in PDB were solved by x-ray crystallography. The technique is also used routinely by scientists to determine how a pharmaceutical interacts with its protein target and what changes might be advisable to improve it (Scapin et al., 2006). However, intrinsic membrane proteins remain challenging to crystallize because they require detergents or other means to solubilize them in isolation, and such detergents often interfere with crystallization. Such membrane proteins are a large component of the genome and include many proteins of great physiological importance, such as ion channels and receptors (Lundstrom, 2004; 2006).

NMR relaxation techniques can also be successfully employed to investigate the time-dependent conformational fluctuations of proteins (Palmer et al., 2001). This

technique makes use of the magnetic spin of atomic nuclei, such as 1H, 13C, 15N and 31P to provide information about the relative atom distances in a molecule, and thus its 3-D model. Moreover, NMR can also be employed for studying the flexibility and dynamics of proteins (Branden and Tooze, 1999). Although initial NMR relaxation experiments and analyses have concentrated on protein dynamics on the ps-ns timescale (Kay, 1998; Palmer, 1997) more recent methodologies focus on μs-ms motions (Palmer et al., 2001; Cavanagh and Venters, 2001). NMR has produced roughly 6000 structures (RCSB Protein Data Bank). While, crystallography can solve structures of large molecules, solution-state NMR is restricted to relatively small molecules (less than 70 kDa) (Scapin et al., 2006).

There exits several other experimental techniques for investigating the protein conformational dynamics and protein interactions at low resolution such as Cryo-electron microscopy (cryo-EM), fluorescence resonance energy transfer (FRET), and single particle tracking methods. These tools give structural transitions without atomic details. Nevertheless, they give insight about the molecular mechanisms and hence shed light to biological functions of proteins (Frank, 1996; Kasprzak et al., 1988; Lakowicz, 1999). Singe particle tracking is a straightforward method for measuring molecular motion. It is based on tagging the molecule with a label that is easily visualized, such as a small reporter particle or a fluorescent dye molecule and then using this label to track motion directly in an optical microscope (Kasprzak et al., 1988; Greenleaf et al., 2007). FRET is based on detection of the emission intensity when energy transition occurs between a donor and an acceptor fluorophores attached to the molecule of interest. It allows the measurement of nanometer-scale motions (Stryer, 1978; Lakowicz, 1999). Cryo-EM is an electron microscopic technique that involves freezing the biological sample in order to view the sample with the least possible distortion and the fewest possible artifacts. The sample is studied at cryogenic temperatures (generally liquid nitrogen temperatures) and a rather coarse-grained map of the protein structure is obtained as compared to x-ray crystallography (Frank, 1996).

2.1.2. Computational Techniques to Identify Protein Structure

Functional genomics is concerned with characterizing the proteins expressed by the genome i.e. assigning their biological function. To some extent, it is possible to determine the function of a protein just from a sequence analysis alone. However, linking the 3-D structure of a protein directly to its function is more appealing, which is known as structural genomics (Leach, 2001).

The number of protein sequences for which the experimental structural information is known is far less than the number of protein sequences discovered (Leach, 2001). The high-resolution images provide necessary information in the atomistic level related to atom rearrangements. Computational techniques such as ab-initio methods, homology modeling and threading are also helpful for survey of protein structures. However, according to PDB statistics, among nearly 40000 protein structures available, only 3% were obtained by computational methods. This once again calls attention to using experimental techniques in structure determination.

The ab-initio structure prediction technique attempts to determine protein structure from scratch by finding the global minimum of an energy function defined on the space of possible structural conformations of the protein. With present methods these techniques are extremely computationally costly and thus have been used only for very small proteins. Homology modeling is based on the reasonable assumption that two homologous proteins will share very similar structures. Given the amino acid sequence of an unknown structure and the solved structure of a homologous protein, each amino acid in the solved structure is mutated computationally into the corresponding amino acid in the unknown structure.

In the case of threading, with the amino acid sequence of a protein of interest, one attempts to align the sequence to each amino acid sequence in a library of template proteins of known structure in such a way that a quasi-energy score or other score is minimized. The score of an alignment is defined in such a way that the value of the score reflects the extent to which the given alignment predicts a structural similarity of the protein of interest to the template protein. Best structural alignment scores are computed for template protein and the template with the best score amongst all templates is returned. Threading relies on the fact that there are far more proteins than folds, so that a given

protein of unknown structure is likely to have structure quite similar to that of a protein of known structure (Leach, 2001).

2.2. Protein Dynamics and Conformational Transitions

A large diversity of processes in living organisms critically depends on protein activity. Virtually all biological processes that involve motion find their origin in protein dynamics. Muscle contraction, for instance, is based on the combined action of two proteins: actin and myosin. Dynamics also plays an important role in many proteins of which the primary function is not mobility itself, such as ability to change conformation (Stryer, 1988). In the following subsections the concepts of protein dynamics and conformational transitions in proteins will be reviewed.

2.2.1. Protein Dynamics

Proteins exhibit different types of collective domain motions such as shear, and hinge-bending motions, which are relevant to biological activity and function (Gerstein et al., 1994). Most of the collective motions of proteins are hinge-like motions occurring around one or multiple centers (Jernigan et al., 1999). The visual interpretation of shear and hinge-bending motions is presented in Figure 2.3.

Hinge motion usually occurs in proteins that have two domains (or fragments) connected by linkers (i.e. hinges) that are relatively unconstrained by packing. A few large torsion angle changes in the hinges are sufficient to produce almost the whole motion. The rest of the protein rotates essentially as a rigid body, with the axis of the overall rotation passing through the hinges. The overall motion is always perpendicular to the plane of the interface. Hinge motion involves a few large changes in main chain torsion angles at the hinge connecting two domains, constrained only by the Ramachandran allowance of torsional angles (Janin and Wodak, 1983; Perutz, 1989; Gerstein et al., 1994; Gerstein and Krebs, 1998).

On the other hand, shear-motion basically describes the special kind of sliding motion a protein must undergo if it wants to maintain a well-packed interface. Because of the constraints on interface structure, individual shear motions have to be very small. Shear motion is accommodated by small changes in side chain torsional angles with no significant deformation in main chain torsional configuration of the interface segments. The whole motion is parallel to the plane of the interface and is limited to total translations of ~2 Å and rotations of 15°. Since an individual shear motion is so small, a single one is not sufficient to produce a large overall motion, and a number of shear motions have to be concatenated to give a large effect. Consequently, proteins that undergo shear often have a layered architecture. Examples include citrate synthase, Trp repressor and aspartate amino transferase (Janin and Wodak, 1983; Perutz, 1989; Gerstein et al., 1994; Gerstein and Krebs, 1998).

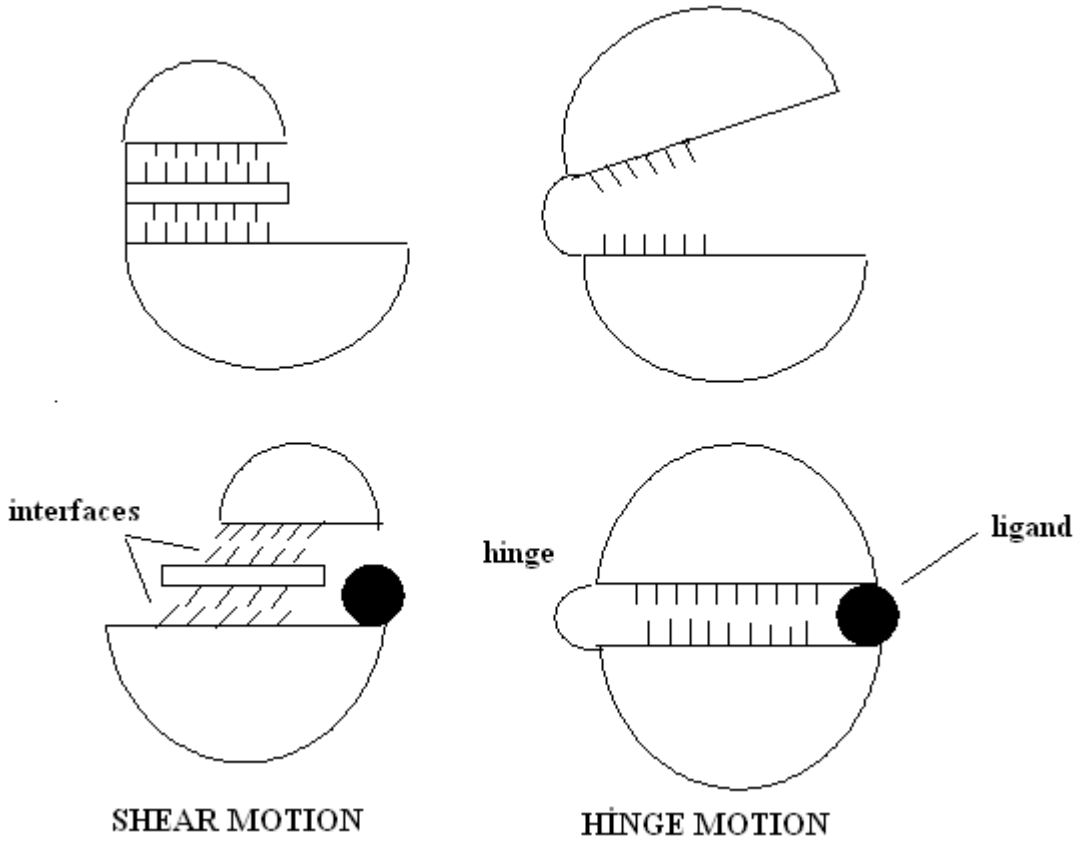


Figure 2.3. Interpretation of shear and hinge motions in proteins (Gerstein and Krebs, 1998)

2.2.2. Conformational Transitions in Proteins

Proteins may shift between several similar structures in performing their biological function. These structures are referred as conformations and transitions between them are called conformational changes. Thus any unique polypeptide may have more than one stable folded conformation and each conformation may have its own biological activity (Brooks et al., 1988). The function of a protein is mainly dependent upon its adopted conformation. Unlike synthetic polymers, which may adopt many different conformations, a protein prefers a unique conformation in its native state. This unique single structure adopted corresponds to the global minimum of the free energy under physiological conditions. The ability to change conformation of a protein is essential for the function of many transport proteins, proteins involved in signal transduction, proteins in the immune system, and numerous enzymes. The conformational changes involved range from very subtle, local changes, as in the case of e.g. myoglobin, to global conformational changes, involving motions of significant amplitude for large parts of a protein (e.g. hemoglobin). Dynamics plays an important role not only in the function of the protein, but also the mechanism by which a protein reaches that native conformation (Stryer, 1988).

Proteins are dynamic molecules and often undergo conformational change upon ligand binding. It is widely accepted that flexible loop regions have a critical functional role in enzyme proteins. In many enzymes, conformational changes serve to enclose the substrate, thereby preventing its release from the protein and ideally positioning it for the protein to perform its function, as in the case of lysozyme. Immunoglobulins are highly flexible in order to be able to deal with a large range of ligands. Another role of protein dynamics is found in G-proteins, binding of a hormone to its receptor triggers the dissociation of the α-domain from the rest of the protein after a GTP-mediated conformational change. A special class of conformational transitions is found in so-called allosteric proteins. Substrate binding to one subunit of these multimeric proteins triggers a conformational change that alters the substrate affinity of the other subunits, thereby sharpening the switching response of these proteins (Stryer, 1988). Antibody-antigen assemblies form another important class of protein complexes exhibiting conformational

changes (Keskin, 2007). Each conformation generates a distinct binding site topology, allowing the receptor to engage multiple ligands at the same region of the binding site.

Global motions of proteins control the actions induced by protein-ligand interactions such as substrate binding to enzymes and antibody-antigen interactions (Amadei et al., 1993; Hayward et al., 1997; Tama and Sanejound, 2001; Keskin et al., 2002; Tobi and Bahar, 2005; Keskin, 2007). Elucidation of the mechanisms by which the proteins bind to each other or to ligands is of great importance to control protein-protein, protein-ligand interactions. There are several different models proposed to explain protein binding mechanisms such as the lock and key model, induced-fit model, and pre-existing equilibrium model that are illustrated in Figure 2.4.

The enzyme-substrate interaction was first explained by the lock and key analogy, where the lock is the enzyme and the key is the substrate. The idea was that, only the correctly sized key (substrate) fits into the key hole (active site) of the lock (enzyme). In the light of experimental evidences another approach, the induced-fit theory, was proposed to explain the enzyme-substrate binding (Koshland, 1958). This approach assumes that the enzyme is partially flexible and its final shape is determined by the substrate.

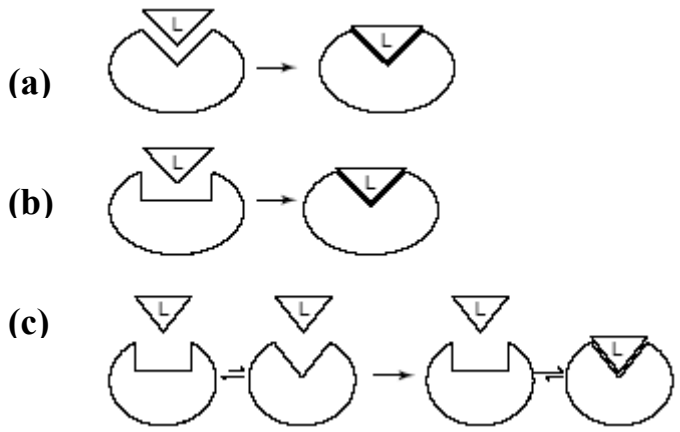


Figure 2.4. Models for protein binding mechanisms (a) Lock and key model, (b) Induced-fit model, (c) Pre-existing equilibrium model. L: Ligand (Goh *et al.,* 2004)

Pre-existing equilibrium is an alternative model to describe the mechanisms of protein interactions (Tsai *et al.,* 1999; Ma *et al.,* 2002; Li *et al.,* 2004). This hypothesis is based on protein folding theories of the funnel energy landscape (Frauenfelder, 1991; Karplus, 1997). In this model, the native state protein possesses an ensemble of conformations at its binding site. The ligand will bind selectively to an active conformation, thereby biasing the equilibrium toward the binding conformation. According to the preexisting equilibrium model, a protein possesses multiple structures and, hence, may have multiple active-sites and functions (Goh *et al.,* 2004; Keskin, 2007).

For proteins that exhibit allosteric behavior, the binding of a ligand to one area of the protein can affect the conformation of the protein at a distant region away from the binding site (Goh *et al.,* 2004). An illustrative example undergoing allosteric conformational transitions can be given as the enzyme adenylate kinase (AK). In AK, there exist two binding sites, i.e. for binding adenosine triphosphate (ATP) and adenosine monophosphate (AMP). Literature studies reveal different findings about domain closure sequence in AK. It is speculated that the ATP binding takes place prior to AMP; hence the ATP-binding domain closes first. Due to its allosteric behavior, the closure of the ATP-binding domain triggers the closure of the AMP-bind domain. Nevertheless, domain closure in AK is still open to debate.

Another important allosteric protein is human hemoglobin. Hemoglobin is a tetrameric protein that transports H^+ and CO_2 in addition to O_2. The oxygen binding properties of hemoglobin are regulated by interactions between its α and β subunits. In other words, the binding of O_2 enhances the binding affinity of another O_2 to the same hemoglobin molecule. This indicates that binding of O_2 is cooperative which indeed makes hemoglobin a more efficient O_2 transporter. In fact, this cooperative binding enables hemoglobin to deliver 1.83 times as much O_2 as it would if the binding sites were independent (Styer, 1988).

The allosteric mechanism of hemoglobin and other allosteric proteins is generally described by the Monod-Wyman-Changeux (MWC) model proposed in 1965. The model assumes equilibrium between the two interconvertable conformational states; the relaxed,

liganded (R) state and tense unliganded (T) state. It also assumes that the subunits of each form should be in the same conformational state, i.e. for a protein having two identical subunits for instance TT or RR states are allowed, whereas RT is not accessible. The R state has a high affinity for the substrate whereas the T state has a low affinity. The addition of substrate shifts the conformational equilibrium in the direction of the relaxed high-affinity R form, because substrate binds only to R form. Hence, propagation of enzyme molecules in the R form increases as more substrate is added which implies a cooperative binding behavior of substrate (Stryer, 1988; Tama and Sanejound, 2001; Goh *et al.,* 2004).

2.3. Molecular Modeling and Simulation Techniques

In this section, molecular modeling and simulation techniques used for proteins will be discussed in two subsections. First, various computational tools currently used for molecular modeling will be presented. In the next subsection, the use of these techniques, especially in studying conformational transitions, is discussed.

2.3.1. Overview and Current State of Art

Various computational techniques are adopted to explore molecular motions and protein dynamics. Molecular dynamics (MD) is one of the most popular atomistic simulation techniques. MD simulations at pico/nano-second time scales output one or more trajectory files, which describe the coordinates of each individual atom over time. In MD, an attempt is made to describe the time evolution of molecular systems as realistically as possible. In a typical simulation, a starting configuration is generated from an experimentally determined structure, and put in an environment that best mimics its natural environment. Then, Newton's equation of motion is solved for this configuration (Leach, 2001). The main problem with animating these trajectories is that, taking large time steps would destroy the impression of smooth motion, while small time steps may not screen interesting motions. Hence, the degree to which the simulations adequately sample the conformational space of the protein is an important task in MD applications. If a given property is poorly sampled over the MD simulations, the results obtained have a limited

usefulness. Although with improvements in computer power and algorithms the simulation times of MD have been progressed to several microseconds, still this timescale is too short to observe many important protein processes, such as slow conformational changes and protein folding/unfolding (Daggett, 2000). To improve the existing capability of the atomic simulations more efficient approaches needed to be introduced.

Collective coordinates can be used as a basis for efficient sampling. Collective modes obtained from short MD simulations are used to generate a coarse-grained description of conformational sub-state and construct a bias potential that lowers the free energy barriers of structural transitions (Muller et al., 2002). Berendsen and coworkers developed the so called “essential dynamics” (ED) which serves as an improvement over conventional MD simulation (Amadei et al., 1993, 1996; Berendsen and Hayward, 2000). In this approach, the protein motions are constrained to move along the essential collective modes, while the motions along the other degrees of freedom obey the usual equations of motion. As a result, an “essential subspace”, spanned by a small number of collective modes can be obtained. However, the basic limitation of these approaches is the fact that different short MD runs each having different initial starting conformations would give different collective mode directions. Biasing the move along that direction would only span the conformational space of the corresponding initial structure. That is, the essential subspace may vary when the protein conformation belongs to different local states. Hence, these techniques need to be further improved to serve as reliable tools for conformational search of large proteins.

In relatively large molecular systems (tens of thousands of particles) the combinatorial problem of calculating all pair-wise interactions makes the force calculations required for MD simulations extremely time-consuming. A clear gap exists between computer simulation times and the times required for most biological processes (Van Gunsteren and Berendsen 1990). With current state of the art methods and computers, simulation time scales generally range between nanoseconds-to several microseconds, whereas most biological processes take place at times ranging between microseconds to seconds (or even minutes). Especially in dealing with large biological systems, NMA,

especially coarse-grained approaches such as elastic network models (ENM) have gained considerable interest recently (Lei *et al.,* 2004; Bahar and Rader, 2005; Ma, 2005).

The use of collective coordinates has become a powerful tool for extracting functionally relevant modes of motion from simulation results (Kitao and Go, 1999). The most widely adopted atomistic technique is the Normal Mode Analysis (NMA) approach. NMA is based on the notion that most positional fluctuation of proteins occurs along collective degrees of freedom (Levitt *et al.,* 1983; Go *et al.,* 1983; Brooks and Karplus, 1983). In NMA approach, proteins are assumed to possess a spectrum of vibrations from slow to fast taking place due to the vibration of atoms around their covalent bonds. Vibrations occurring at high frequencies are referred as the fast motions (fast modes). On the other hand, the large domain motions occur at lower frequencies, referred as slow motions (slow modes). Residues active in the fastest modes possess a strong resistance to conformational changes which implies their important role in maintaining the structure. Residues active in the slowest modes, on the other hand, are susceptible to large scale (global) motions associated with the collective dynamics of the overall tertiary structure. Hence, slow motions are relevant to biological function (Bahar *et al.*, 1998).

In NMA, the potential energy surface is assumed to be harmonic. Collective variables are obtained by diagonalisation of the Hessian matrix (second derivative of the potential energy) in a local energy minimum. Quasi harmonic analysis (Karplus and Kushick, 1981; Levy *et al.,* 1984; Teeter and Case, 1990), principal component analysis (Kitao *et al.,* 1991; Amadei *et al.,* 1991; Garcia, 1992) and singular value decomposition (Romo *et al,* 1995) of MD trajectories of proteins have shown that even beyond the harmonic approximation, protein dynamics is dominated by a limited number of collective coordinates. These methods seek those collective degrees of freedom that best approximate the total amount of fluctuation. It was shown that low frequency collective motions from NMA correlate well with experimental data related to biological function (Tama and Sanejouand, 2001).

Another widely used tool used to carry out extensive search on the conformational space of especially larger proteins is the Monte Carlo (MC) simulation technique. MC

search methods are stochastic techniques based on the use of random numbers and probability statistics to sample conformational space. In general MC search consists of two steps: (1) generating a "trial conformation" and (2) deciding whether the new conformation will be accepted or rejected. Starting from an initial conformation, random numbers are used to generate the next trial conformation by constructing moves. Multiple torsion moves as well as Cartesian coordinate moves are among the many possible variations. Once a new "trial conformation" is created, it is necessary to determine whether this conformation will be accepted or rejected. If rejected, randomly creating new conformations is repeated until one of them is accepted. The new conformation is accepted or rejected according Metropolis criterion which is based on choosing the low energy conformation. If the energy of the new conformation is lower than the current conformation than the move is accepted, i.e. the trial conformation is now taken as the current conformation. However, even if the energy of the trial conformation is higher than the current energy, there is a certain probability, proportional to the Boltzmann factor, that it will be accepted (Becker, 2001). MC calculations using coarse grained protein models similar to those used for threading have shown that native state dynamics of proteins can successfully be simulated at a rate one order of magnitude faster than can be obtained by all-atom models (Haliloglu and Bahar, 1997; Bahar *et al.,* 1997c). This simplified yet realistic method allows efficient simulation of the dynamics with multiple independent trajectories at long time scales.

2.3.2. Simulation Techniques Used in Studying Conformational Transitions

The conformational transitions are usually too fast to be measured experimentally, or can be rarely observed with molecular dynamics simulations; which pose a difficult challenge (Noe *et al.,* 2003). Atomistic simulation techniques that make use of all-atom empirical potentials (Brooks and Karplus, 1983) such as MD and NMA generally become computationally inefficient for investigating conformational transitions of proteins, especially as the system size increases.

Currently, MD simulations that can reach up to nanoseconds or at most a few microseconds are not suitable for exploration of conformational changes in the time scale of microsecond/milliseconds to seconds. For this purpose, targeted MD simulations (TMD)

(Schlitter *et al.,* 1993; Kong *et al.,* 2006, van der Vaart and Karplus 2005, 2007) are adopted to simulate large conformational transitions by biasing the conventional MD technique to sample the conformational space in a predefined direction. It is based on knowledge of both the initial and end structures and performing a molecular dynamics simulation starting from one conformational state as initial structure and using the RMSD from the end state for a directing, i.e. biasing constraint (Schlitter *et al.*, 1993) The TMD proved successful in finding continuous pathways for the investigated transitions. However, it does not always yield reversible pathways and necessarily follow lowest energy pathway (van der Vaart and Karplus 2005).

Alternative to atomistic models, low resolution coarse-grained approaches, such as ENM and MC simulation techniques may appear as efficient tools for conformational analysis of large proteins and their complexes. Recently several ENM based approaches frequently employed in conformational transition studies have been reported (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.* 2002, 2005; Xu *et al.,* 2003; Mouawad *et al.,* 2002; Maragakis and Karplus, 2005; Krillova *et al.,* 2008). An NMA based computational study was reported in which a set of different proteins with different binding mechanisms were analyzed (Krebs *et al.,* 2002). The results showed that half of the proteins studied undergo conformational changes that are governed by the two or three lowest frequency modes. This suggests that, conformational transition of protein between unbound and bound (to a ligand) structures follow the lowest frequency normal modes of the protein.

Mouawad and Perahia (1996) aimed to reveal the transitional pathway of hemoglobin. The authors carried out single NMA on the T structure and along both positive and negative directions. T structure was continuously deformed along first 3 slowest modes consecutively followed by energy minimization. Four intermediate structures were proposed between T-R and maximum approach to R state was 1.82 Å. The study of Xu *et al.* (2003) is somewhat similar to Mouawad and Perahia (1996), in which the authors demonstrated a transition from T to R2 in hemoglobin. This study revealed that

the T to R2 transition is in accord with the most global (slowest) mode. The authors did not propose a transition pathway but reported an approach of 2.4 Å to R2 state.

Among various computational methodologies, based on ENM that are developed for investigating the conformational transitions, one such promising approach is the method developed by Zheng and Brooks in which they proposed an ENM based approach that makes use of the crystal structure of the initial state and several distance constraints for the end state (Zheng and Brooks, 2005a; 2005b; 2006). The method, which is utilized by iteratively minimizing the error of fitting the given distance constraints as well as the energy cost, proved successful in maintaining the associated transitions in a set of 16 protein structures by iteratively minimizing the error of fitting the given distance constraints as well as the energy cost. In their recent work, Zheng and coworkers developed a mixed elastic network model (MENM) to study the large scale conformational changes of motor proteins KIF1A kinasin and myosin II (Zheng *et al.,* 2007). This approach combines the elastic network potentials of the initial and end states from known x-ray structures by adding their respective partition functions. The MENM energy function is generated and transition paths are characterized by connecting the beginning and end structures which are retained as local minima on the MENM surface.

Maragakis and Karplus (2005) developed the plastic network model (PNM) which generates a minimum energy path between two end structures. With this methodology, the authors studied the transitional pathway of adenylate kinase (AK) from open to closed conformations and suggested a set of crystal structures that may possibly be present in the pathway. Kim and coworkers developed the rigid-cluster elastic network interpolation (ENI) algorithm based on uniformly interpolating the distances in two different conformations within the framework of elastic network model (Kim *et al.,* 2002; 2005). In another recent study, the authors combined geometric path planning algorithms originating from robotics research and ENM in order to study large-amplitude conformational changes with application to AK (Krillova *et al.,* 2008). By using path planning algorithm, the conformational exploration is performed and then guided with the directions of collective motions obtained by low-frequency modes.

Recently, there have been several hybrid strategies combining different approaches summarized so far in order to enhance the conventional techniques and overcome their limitations. One such approach, the so called Amplified Collective Motion (ACM), was proposed by Zhang and coworkers. The method uses ENM-derived normal modes for improving the simulation efficiency of MD simulations (Zhang *et al.,* 2003; He *et al.,* 2003). In this novel approach, the authors aim to accelerate the conformational sampling of proteins through MD simulations by making use of collective modes obtained from coarse-grained elastic network models. The authors applied their methodology to bacteriophage T4 lysozyme and villin headpiece subdomain (HP-36). Both in terms of sampling low energy conformations and the coverage of the conformational space by the sampled conformations, these new schemes are reported to be able to achieve the expected increase in sampling efficiency.

2.4. Proteins Studied

2.4.1. Adenylate Kinase (AK)

E.coli AK is a 214-residue allosteric protein belonging to nucleoside monophosphate kinases (NMP kinases), which catalyze the transfer of the terminal phosphoryl group from adenosine triophosphate (ATP) to adenosine monophosphate (AMP). AK involves three domains, namely a core domain, an ATP-binding or LID domain (residues 122-158) and an AMP-binding (AMP-bind) domain (residues 30-63) (Figure. 2.5). In apo/unligated structure, the ATP-bind and AMP-bind domains assume open conformations. During catalysis, the core domain of AK is mainly preserved unlike the LID and AMP-bind domains that undergo large conformational changes. AK is known to attain two unique conformations, namely open and closed states (Berry *et al.,* 1994; Müller and Schulz, 1992; Müller *et al.*, 1996). The corresponding RMSD between the open and closed conformations is 7.13 Å.

The ribbon diagrams presented in Figure 2.5 are generated in Pymol (De Lano, 2002). In the figure, the core, LID and AMP-bind domains are colored in blue, red and orange, respectively.

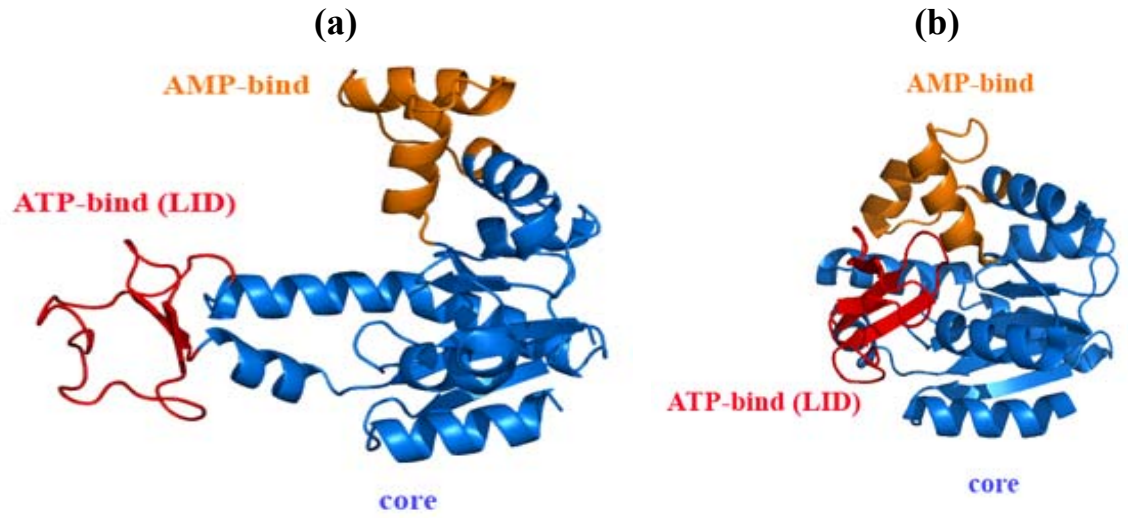


Figure 2.5. Ribbon diagrams for (a) apo/open conformation of AK (PDB code: 4AKE) (b) bound/closed conformation of AK (PDB code: 1AKE)

In Figure 2.5 (a), the open conformation is depicted with protein data bank (PDB) code: 4AKE (Müller *et al*., 1996)), and in panel (b), the closed conformation crystallized with the inhibitor P^1P^5-di(adenosine-5')pentaphosphate (AP_5A), available with PDB code: 1AKE (Müller and Schulz, 1992) is illustrated.

2.4.2. Hemoglobin

Hemoglobin is a common example for allosteric transitions in proteins. As an oxygen-binding tetrameric protein, hemoglobin adopts three conformations: a relaxed CO-bound (R2) state (PDB code: 1BBB) (Silva *et al.,* 1992), a relaxed O_2-bound (R) state (PDB code: 1HHO) (Shaanan, 1983), and a tense unliganded (T) state (PDB code: 1A3N) (Tame and Vallone, 2000).

Figure 2.6 depicts the ribbon diagrams of T (unliganded) and R2 (CO-bound) forms of tetrameric human hemoglobin. Each of the four monomers $\alpha 1$, $\alpha 2$, $\beta 1$, $\beta 2$ are colored with cyan, yellow, green, magenta. The four heme groups are indicated with red. The structure R2 was first thought to be an intermediate between T and R, however

computational studies suggested that it is instead R being probably an intermediate between T and R2 (Srinivasan and Rose, 1994).

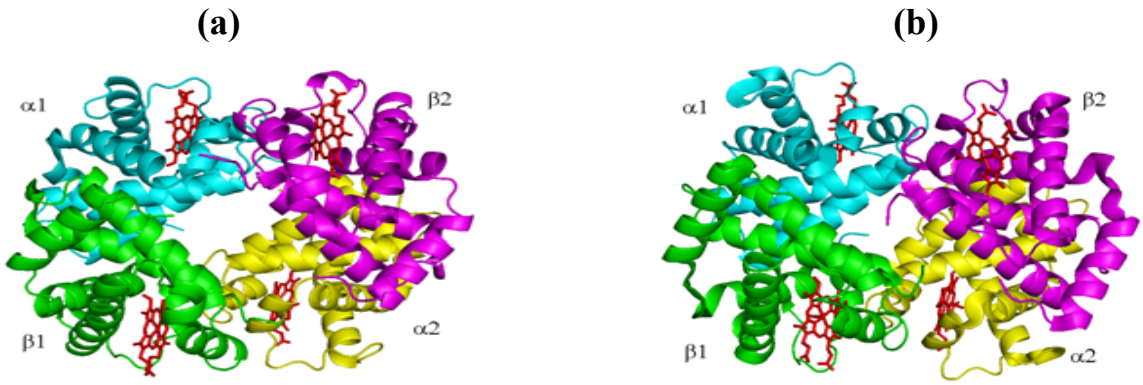


Figure 2.6. Tense (T) and relaxed (R2) conformations of tetrameric human hemoglobin (PDB entries: 1A3N and 1BBB, respectively).

2.4.3. Human Serum Transferrin (HSTR)

HSTR (Figure 2.7) is a 679 residue protein responsible from binding ferric ions in the bloodstream and transporting the bound irons to cells. Upon release of iron, the iron-free apo transferrin is returned to circulation without degradation (Jeffrey *et al*., 1998). The sequence of steps in iron release is not well known. However, kinetic and biophysical studies (El Hage Chahine *et al.,* 1995) reveal that the iron release mechanism is associated with a large conformational change that is similar in all transferrins. The protein is folded into two globular lobes: The N-lobe and C-lobe, each of which are able to bind one ferric ion. The two lobes are also divided into two subdomains, i.e N1 and N2; C1 and C2 lobes. The binding cleft is placed in between each subdomain to bind a single Fe^{3+} ion together with CO_3^{2-} ion. The iron ligand residues are Tyr 95 and Tyr 188 in N2-lobe; and Asp 63 and His 249 in N1-lobe (Baker *et al*., 2003).

Comparison of the apo (Jeffrey *et al.*, 1998) (PDB code: 1BP5-A chain) and holo (Macgillivray *et al*., 1998) (iron-bound; PDB code: 1A8E) forms of the protein depicted in Figures 2.7 (a) and (b), respectively, shows that a large rigid-body domain movement of

63^{o} in the N-lobe occurs in apo form as compared to holo form to enable an open binding cleft (Jeffrey *et al*., 1998). This large conformational change corresponds to 6.7 Å RMSD between the two conformations. This domain movement is also very similar in human lactoferrin protein (Anderson *et al*., 1990).

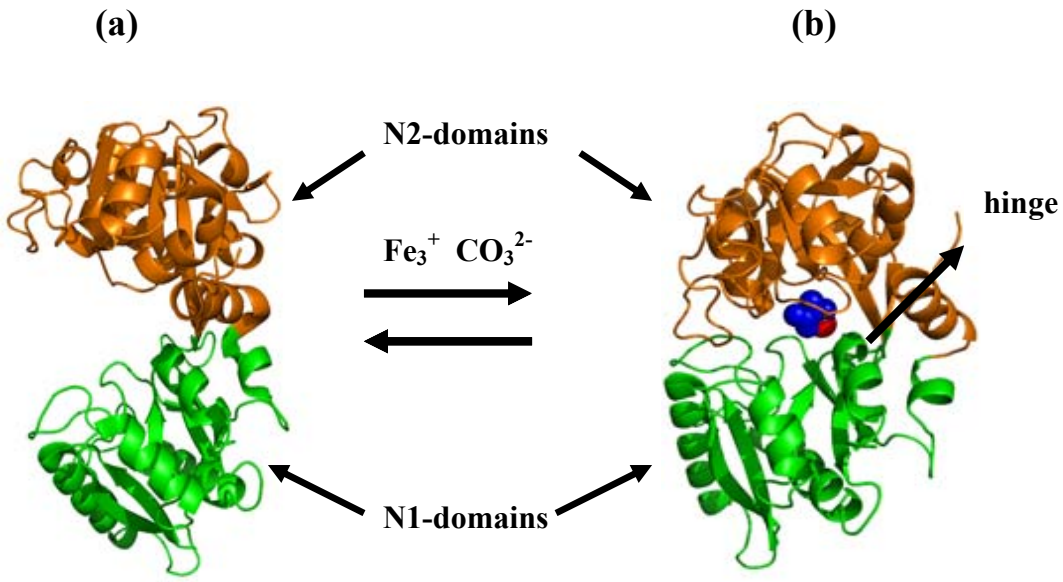


Figure 2.7. Ribbon diagrams for (a) apo/open conformation (PDB code: 1BP5-A chain) and (b) bound/closed conformation (PDB code: 1A8E) of HSTR.

2.4.4. Lysine/Arginine/Ornithine binding (LAO-binding) protein

LAO-binding protein (Figure 2.8) is part of bacterial periplasmic transport system which is consist of the substrate binding protein (receptor) and the membrane-bound complex proteins (Kang *et al*., 1991). LAO-binding protein is responsible for binding the Lysine, Arginine, or Ornithine amino acids in the cell. These amino acids are then translocated from the periplasm to cytoplasm and by interacting with the membrane proteins. As all other binding proteins, the LAO-binding protein undergoes a significant conformational change, characterized by an opening/closing motion of two domains, upon substrate binding (Oh *et al.*, 1993). In Figures 2.8 (a) and (b) the open (apo) (Kang *et al*., 1991) and closed (Oh *et al*., 1993) (liganded, Lysine bound) structures of the protein are

provided. The protein consists of 238 residues and the RMSD between the open and closed forms corresponds to 4.7 Å.

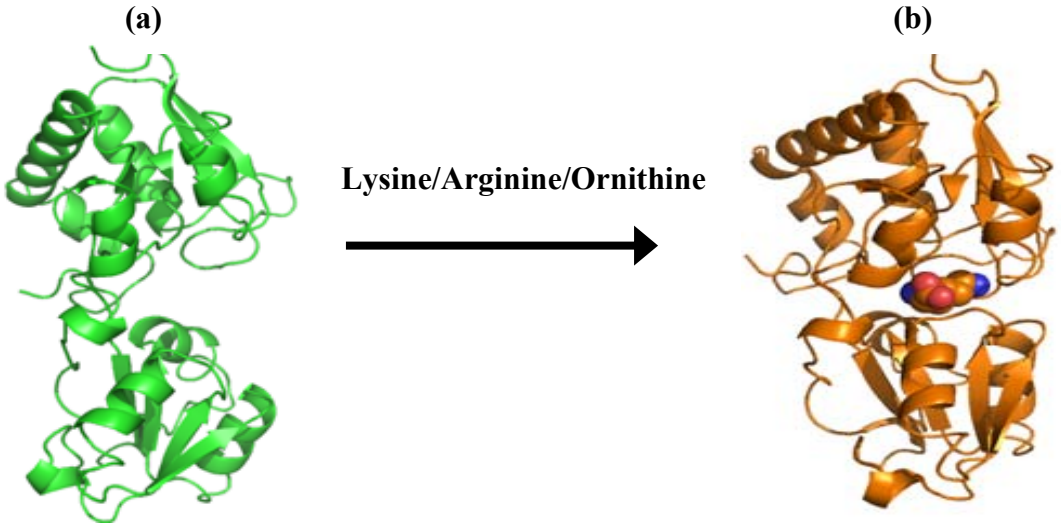


Figure 2.8. Ribbon diagrams for (a) apo/open conformation (PDB code: 2LAO) and (b) bound/closed conformation (PDB code: 1LST) of LAO-binding protein.

3. MATERIALS AND METHODS

3.1. Anisotropic Network Model (ANM)

ANM (Atilgan *et al.*, 2001) is a coarse-grained NMA tool, commonly used to determine the vibrational motions in proteins. By considering the 3D anisotropy of the residue fluctuations, ANM predicts the directions of collective motions, which provide information about the biological function of the protein and its mechanism of action. It estimates the magnitude and direction vectors of residue displacements using NMA for 3*N*-6 internal modes, for an *N*-residue structure. The coarse-grained representation of protein structures assumes C^{α} atoms of amino acids as interaction centers, i.e. nodes.

In the elastic network representation of a protein, all pairs of these coarse-grained sites/nodes that are closer than a cutoff distance, r_c (usually 13-18 Å), are connected by harmonic springs with a universal force constant γ. The corresponding total potential energy of the folded protein structure can be given as a summation over all harmonic interactions of (i, j) pairs as

$$V = (\gamma / 2)\sum_i \sum_j h(r_c - R_{ij})(\Delta \mathbf{R}_j - \Delta \mathbf{R}_i)^2 \qquad (3.1)$$

where *h(x)* is the heavy side step function [*h(x)* = 1 if $x \geq 0$, and zero otherwise]; γ is the universal force constant and R_{ij} is the distance between sites *i* and *j* in the native structure of protein. $\Delta \mathbf{R}_i$, is the fluctuation in the position vector $\mathbf{R}_i$ of site *i* ($1 \leq i \leq N$). The directional dependence of ANM incorporates the *X*, *Y*, and *Z* components of the position vector $\mathbf{R}_i$. Therefore, the overall potential calculation includes the fluctuations for all components. The potential energy of a structure with *N* interaction sites is expressed in matrix notation as

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

where, $\Delta\mathbf{R}$ is a 3N-dimensional vector of the fluctuations $\Delta\mathbf{R}_i$ in the position vectors $\mathbf{R}_i$ of all sites ($1 \le i \le N$), $\Delta\mathbf{R}^T$ being its transpose, and $\mathcal{H}$ is the Hessian matrix. In the general case of N residues, the second derivatives of the overall potential are organized in the (3Nx3N) Hessian matrix, $\mathcal{H}$, which is composed of (NxN) super-elements of size (3x3) and expressed as

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

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

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

with X_i, Y_i and Z_i being the components of $\mathbf{R}_i$. In order to calculate the normal modes of the elastic network, the symmetric Hessian matrix $\mathcal{H}$ is diagonalized into the form,

$$\mathbf{U}^t \mathcal{H} \mathbf{U} = \mathbf{\Lambda} \tag{3.5}$$

$\mathbf{\Lambda}$ is the (3N x 3N) diagonal matrix with diagonal elements being eigenvalues or squared normal mode frequencies. $\mathbf{U}$ is a (3N x 3N) orthogonal matrix ($\mathbf{U}^t\mathbf{U} = \mathbf{U}\mathbf{U}^t = \mathbf{I}$, $\mathbf{I}$ being the identity matrix), where the columns are the normalized eigenvectors giving the normal mode directions of motion. In normal mode calculations, the overall rotational and translational motion of the molecule is excluded corresponding to six zero eigenvalues and the overall motion is described over 3N-6 individual internal modes.

The correlation between $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ can be given by

$$<\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j> = (3k_BT/\gamma)\, tr\, [\mathbf{H}^{-1}]_{ij} \quad (3.6)$$

where k_B is the Boltzmann constant, T is the absolute temperature, and $tr[\mathbf{H}^{-1}]_{ij}$ is the trace of the ij^{th} submatrix $[\mathbf{H}^{-1}]_{ij}$ of $\mathcal{H}^{-1}$, that is the sum of the diagonal elements of this 3x3 matrix. It refers to three different components of $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$. The mean square fluctuations (msf) of the residues, $<\Delta\mathbf{R}_i^2>$, can also be found by the above equation when $i = j$.

The force constant γ can be determined by comparing theoretical and experimental residue fluctuations. Experimental msf can be obtained from the temperature factor or B-factor data of crystal structures with the following equation.

$$B_i = (8\pi^2 / 3)\, \langle \Delta\mathbf{R}_i^2 \rangle \quad (3.7)$$

Previous studies have demonstrated that essential fluctuation characteristics and important collective mode shapes could be successfully reproduced by coarse-grained ANM with high efficiency, i.e. the required computational time being several orders of magnitude less than that for atom-based simulation techniques like molecular dynamics (MD) simulations (Bahar *et al.,* 1997a, 1998; Doruker *et al.,* 2000, 2002). It has been demonstrated that the collective motions at low frequencies obtained by ANM and essential motions from MD agree with high overlaps (Doruker *et al.,* 2006).

3.2. Monte Carlo (MC) Simulation Technique

An off-lattice dynamic Monte Carlo (MC) / Metropolis simulation method (Haliloglu and Bahar, 1998) has been developed to simulate protein dynamics at different time scales. Simulation results have been shown to be in agreement with those from NMR measurements, such as order parameters (Haliloglu and Bahar, 1999) and hydrogen exchange data (Kurt and Haliloglu, 1999). This method uses a simplified yet realistic

model for representing the protein structure and allows efficient simulation of the dynamics with multiple independent trajectories at long time scales.

In this simplified model, the backbone of the protein structure is represented by the virtual bond model originally proposed by Flory (Flory, 1969). Each residue is represented by two interaction sites, its alpha-carbon atom and center of mass of side chain. A schematic representation of the model is given in Figure 1, where a protein segment between backbone sites C^{α}_{i-2} and C^{α}_{i+2} is shown. The conformation of the backbone is defined by 3*N*-6 variables: *N*-1 backbone virtual bonds l_i connecting 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, describing the torsional rotation of the bond l_i. The sidechain conformation, is expressed by the set of generalized variables $\{l_i^S, \theta_i^S, \phi_i^S\}$, l_i^S being the bond length connecting backbone and sidechain interaction site, θ_i^S is the bond angle between l_i and l_i^S, and ϕ_i^S the torsion angle defined by l_{i-1}, l_i and l_i^S.

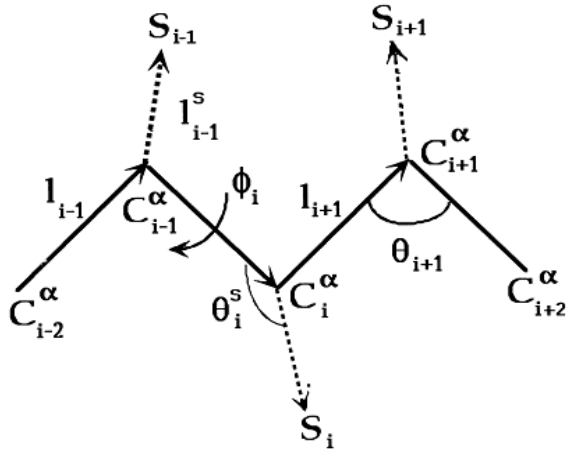


Figure 3.1. Schematic representation of the virtual bond model

To calculate the energy $E(\Phi)$ of a given protein conformation Φ, contributions of two types of interactions are added: long-range (LR) interactions between non-bonded residues that are close in space, and short-range (SR) interactions between covalently bonded units along the chain sequence in the form,

$$E(\Phi) = E_{LR}(\Phi) + E_{SR}(\Phi) \tag{3.8}$$

The long-range potentials given by Bahar and Jernigan (1997) are used to evaluate $E_{LR}(\Phi)$ according to the following expression:

$$E_{LR}(\Phi) = \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} E_{SS}(r_{ij}) + \sum_{i=1}^{N-4}\sum_{j=i+4}^{N} E_{SB}(r_{ij}) + \sum_{i=1}^{N-5}\sum_{j=i+5}^{N} E_{BB}(r_{ij}) \tag{3.9}$$

where r_{ij} is the distance between sites *i* and *j* in conformation Φ. The first term stands for 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 one between two backbone sites (BB) of residues *i* and *j,* respectively.

The short-range conformational energy of the backbone is evaluated using statistical potentials extracted form protein structures as based on virtual bond model and formulations reported by Bahar *et al.* (1997b) and Haliloglu and Bahar (1998) as expressed with the following equation.

$$E_{SR}(\Phi) = \sum_{i=2}^{N} E(l_i) + \sum_{i=2}^{N-1} E(\theta_i) + \sum_{i=3}^{N-1}\left[E(\phi_i^-)/2 + E(\phi_i^+)/2 + \Delta E(\phi_i^-, \phi_i^+)\right] \tag{3.10}$$
$$+ \sum_{i=3}^{N-1}\left[\Delta E(\theta_i, \phi_i^-) + \Delta E(\theta_i, \phi_i^+)\right]$$

In this equation, the first summation is the potential associated with the stretching of the virtual backbone bonds, approximated by a stiff harmonic potential. The second summation refers to the bending of backbone bond angles; the third stands for 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. Terms are also included to account for pairwise interdependence of the torsion and/or bond angle bending. For the short-range conformational energy of the side chains, the statistical potentials converted from the probability distributions for packing of side chains in low resolution models are used (Kurt

et al., 2003). For instance, the energy associated with a side chain bond angle at state θi for a residue type A is

$$E_A(\theta_i) = -RT\, ln\left[P_A(\theta)/\, P_A^0(\theta)\right] \tag{3.11}$$

where $P_A(\theta)$ is the statistical probability of finding that bond at angle θ and $P_A^0(\theta)$ is the background probability assuming a uniform distribution. Similar expressions are derived for side chain bond length and torsions and side chain conformation energy is summed over all side chains as

$$E_{SR}^S(\Phi) = \sum_{i=1}^{N} E\left(l_i^S\right) + \sum_{i=2}^{N} E\left(\theta_i^S\right) + \sum_{i=2}^{N} E\left(\phi_i^S\right) \tag{3.12}$$

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

In the low resolution MC/Metropolis simulation technique, a randomly chosen site, either an alpha-carbon or a sidechain site, is subjected to a differential perturbation using a uniformly distributed random number generator. The strength of the perturbation, Δx is controlled by the formula,

$$\Delta x = k(2r - 1) \tag{3.13}$$

Here, r is the random number variable in the range $0 \le r \le 1$, and k is a damping factor that may be adjusted to mimic the response at a given temperature. The acceptance of a move is based on the Metropolis criterion which is based on generating a new random number and accepting the new conformation Φ if the following condition is satisfied, where Φ_o represents the original conformation:

$$\exp\left\{-\left[E(\Phi) - E(\Phi)_0\right]/\, RT\right\} \le r \tag{3.14}$$

One MC step (MCS) is defined as N perturbations and may be viewed as the average time for all N residues of the protein to have a chance to move. An MC algorithm with only local moves in some cases could give insight about the real time of a process. However, there is no correspondence between MCS and real time in the algorithm due to the implementation of both global deformations on collective modes and local moves. The analyses would be reliable if performed over multiple independent runs that are long enough for efficient sampling of the conformational space.

3.3. ANM-MC Simulation

In the developed algorithm (Kantarci-Carsibasi *et al.,* 2008), ANM is successively employed to provide collective deformation directions and short MC runs to provide energy minimization on the deformed conformations. Schematic flowchart of the method developed is provided in Fig. 3.2. The algorithm requires the crystal structures of the initial and final/target states as inputs. The initial state may be the apo/open conformation of a protein and the final state may be the bound/closed conformation. Atomistic structures are first coarse-grained (STEP 1) by assigning two nodes for each residue: the alpha-carbon and the center of mass (com) of the side chain in accordance with the knowledge based potentials used. For glycine, only one node (alpha carbon) is taken into account. Next, the target structure is aligned on the initial structure based on alpha-carbon coordinates in order to define the target direction (STEP 2). The target direction ($\boldsymbol{Q}$), which is a $3N$ dimensional vector comprised of alpha-carbon coordinate deformations, is calculated by subtracting the initial coordinates from the aligned coordinates of the target. Details of the superimposition methodology can be found in the study by Umeyama (1991) which is reviewed in Appendix A.

ANM is then applied to the initial structure (STEP 3) to extract m lowest frequency eigenmodes (m = 10 and 20 is used in this thesis) using BLZPACK (Block Lanczos Package)- a Fortran 77 implementation of the block Lanczos algorithm (Grimes *et al.,* 1994; Marques, 1995, 2001).

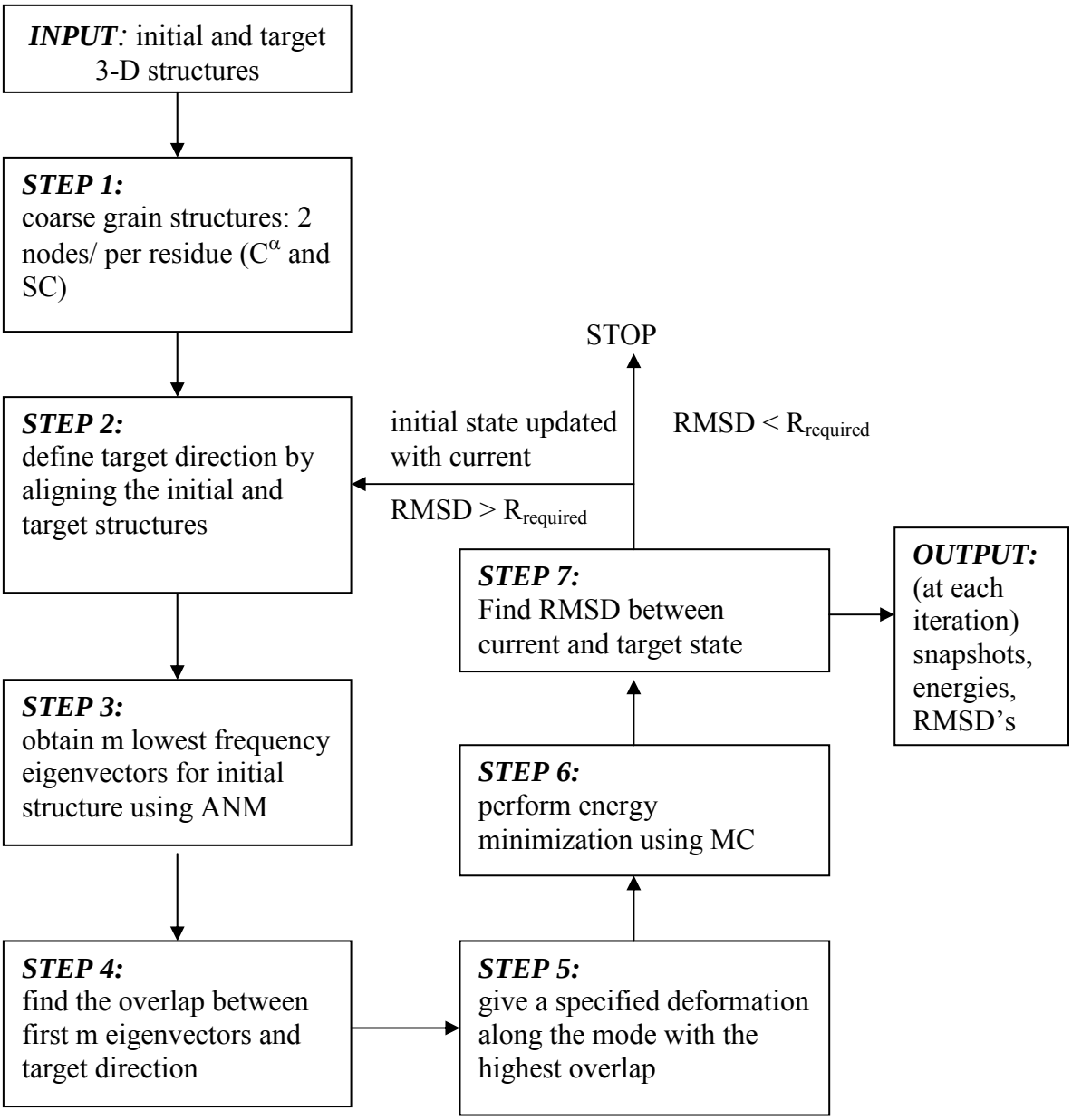


Figure 3.2. Flowchart of the ANM-MC algorithm

This algorithm facilitates calculating the generalized and standard eigenvalue problem and allows obtaining corresponding eigenvectors and eigenvalues of a very large matrix within a considerably short time. The cutoff distance r_c is taken to be 18 Å in ANM calculations using only alpha-carbon coordinates. The dot products of the m eigenvectors with the target direction are computed (considering both positive and negative directions) and the

eigenvector **U** (Eq. 3.5) giving the highest dot product- overlap- with the target direction ***Q*** is determined (STEP 4).

A new conformation is now generated by deforming the initial structure along this mode direction using a prespecified deformation factor (DF = 0.1–0.5 Å in this thesis) as expresses in the following notation: $\mathbf{R}_{new} = \mathbf{R} \pm \mathbf{U}$ (rescale factor). Here, $\mathbf{R}_{new}$ and **R** denote the coordinate matrix of the new conformation generated and the initial state for which normal mode analysis is carried out, respectively (STEP 5). **U** is the eigenvector and it is multiplied by a rescale factor such that the RMSD between the new and the old conformation corresponds to selected DF. In other words, for instance, DF=0.2 Å means that the deformation is given such that the newly generated conformation has an average RMSD of 0.2 Å from the previous conformation. Although the ANM is performed using the alpha-carbon atoms, the structure is deformed along the selected mode direction by applying the same deformation to the alpha-carbon and side chain nodes of each residue. The deformed conformation is allowed to relax by a specified number of MC steps (MCS = 100, 500 or 1000 employed here) by utilizing the MC algorithm described in Section 3.2 (STEP 6).

Finally, the RMSD between the energy-minimized new structure and the target structure is calculated (STEP 7). Once the two structures, i.e. the target structure **B** is superimposed on the initial structure **A**, the coordinates of B are rotated ($\mathbf{B_{rot}}$) and the corresponding RMSD based on alpha carbon coordinates of the initial structure **A** and superimposed target structure ($\mathbf{B_{rot}}$) can then be calculated by Equation 3.15.

$$\mathrm{RMSD} = \left[\frac{\sum_{i=1}^{N} [(x_{A,i} - x_{B_{rot},i})^2 + (y_{A,i} - y_{B_{rot},i})^2 + (z_{A,i} - z_{B_{rot},i})^2]}{N} \right]^{1/2} \qquad (3.15)$$

RMSD calculation is performed for each residue and summation over all residues is divided by the residue number and then the square root is taken, hence an average RMSD is calculated. If the RMSD is smaller than a desired value ($R_{required}$) the simulation stops; if

not the simulation enters a new iteration after the initial state is updated with the output from the previous MC step. Obviously, the target vector is also updated at each iteration based on the modification of the initial state.

The ANM-MC simulation is performed in an automated way after providing the x-ray structures of the initial and final states of a protein. The necessary input parameters are, the deformation factor (DF), the number of MC steps (MCS) to be employed at each iteration, and the cutoff distance for ANM calculations (r_c = 18 Å). Parameter adjustment is carried out in this thesis to determine suitable values for DF and MCS which are 0.2 Å and 1000, respectively. ANM-MC is applied first to adenylate kinase (AK) and then to another frequently studied allosteric protein, the human hemoglobin.

3.4. Targeted Monte Carlo (TMC) Simulation

TMC simulations (Kantarci-Carsibasi *et al.,* 2008) are carried out to assess the effect of incorporating normal mode directions in the ANM-MC methodology. Similarly, starting with the open conformer the protein is forced toward the closed conformer by deforming the coordinates along the target direction without using slow modes. Hence the new conformation can be obtained by $\mathbf{R}_{new} = \mathbf{R} \pm \boldsymbol{Q}$ (DF). The new conformation is again relaxed by MC and the rest of the algorithm is same (Fig. 3.2). In short, the MC conformational search is now targeted absolutely in the direction of desired conformational transition. This is similar to morphing algorithms (Krebs and Gerstein, 2000), where the path between two conformations is linearly interpolated together with energy minimization to conserve the chemical structure. The simulation parameters are DF and MCS in the TMC method.

3.5. RG-ANM-MC Simulation

RG-ANM-MC method (Kantarci-Carsibasi *et al.,* 2009) originates from ANM-MC method with additional advantage of its applicability to cases with unknown final state 3-D structure. The schematic flow diagram of the algorithm is presented in Fig.3.3, which can be summarized as follows: In STEP 1, the atomistic pdb structure of the initial

conformation is coarse-grained by assigning two nodes for each residue: the alpha-carbon (C^{α}) and the center of mass of the side chain (SC) in accordance with the knowledge based potentials used (For glycine, only one node, C^{α}, is taken into account).

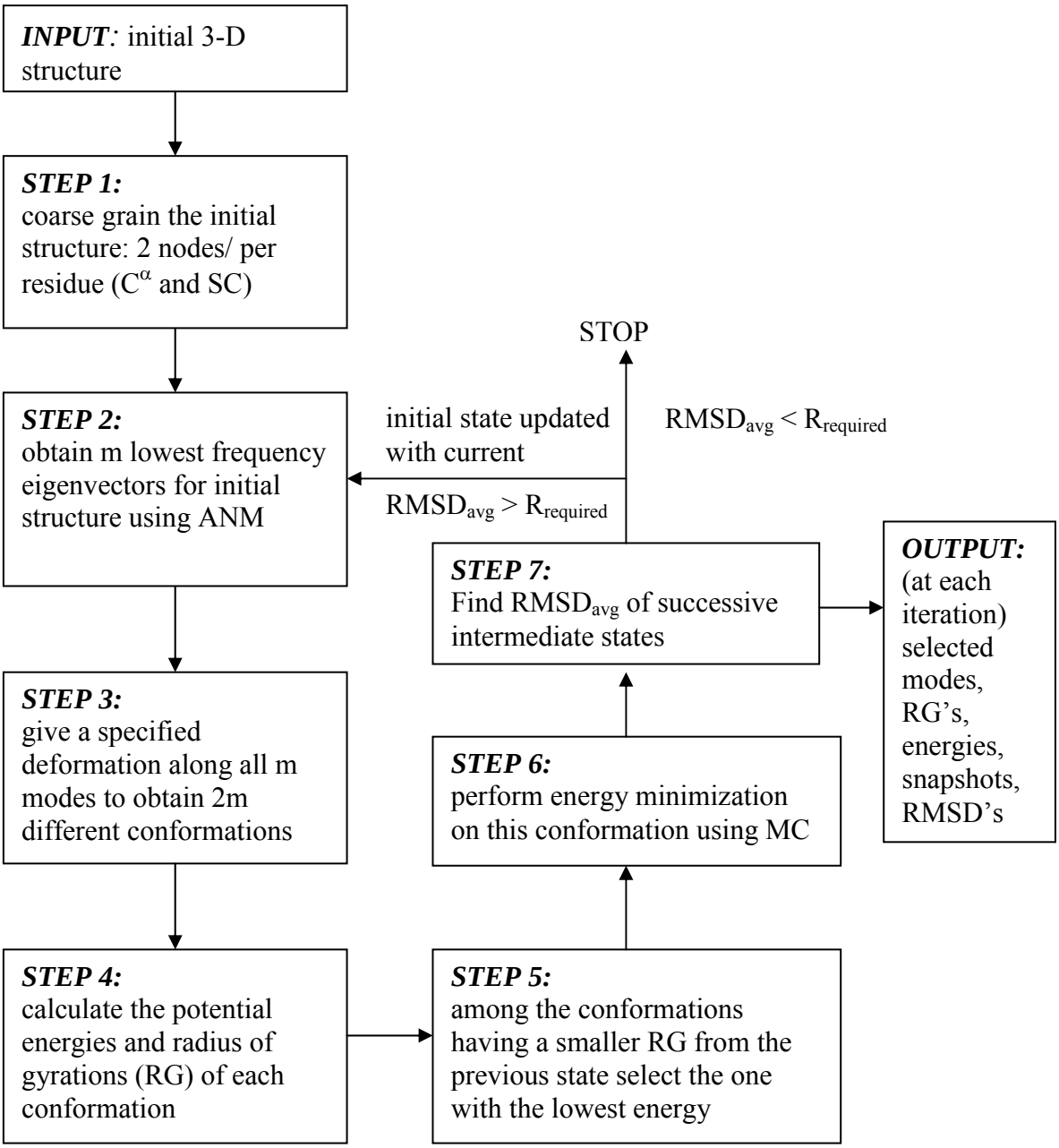


Figure 3.3. Flowchart of the RG-ANM-MC algorithm

Then, ANM is applied to this coarse-grained initial structure (STEP 2) to extract m lowest frequency eigenmodes using BLZPACK (m=10 here). The initial structure is deformed using a prespecified (DF) in all m directions which results in 2m different conformations (STEP 3) (positive-negative direction of each mode, i.e. taking m=10 results in 20 different conformations). Among these 2m different conformations, the "essential conformation" approaching to the closed conformer of the protein is selected based on two criteria: the potential energies and the radius of gyrations (RG) of the conformations. The definition of RG is presented by Eq. 3.16, where the distance of each residue from the center of mass (com) of the protein is summed over all residues.

$$\mathrm{RG} = \frac{\sum_{i=1}^{N} (x_i - x_{com})^2 + (y_i - y_{com})^2 + (z_i - z_{com})^2}{N} \tag{3.16}$$

In STEP 4, potential energies of all 2m conformers together with their RG's are calculated. Then the conformations having lower RG than the starting conformation are identified and among them a single conformation having the minimum energy as compared to others is selected (STEP 5). In this way, the transition is guided by a constraint put on RG. In other words, the open protein is guided towards a conformation having an RG close to the closed state through an energetically favorable pathway. However, it should be noticed that closed state 3-D structure is not used except that the RG of the open conformation should obviously decrease in order for the protein to close. Similarly, the obtained structure is energy minimized by MC simulation (STEP 6).

The point where the iterative algorithm should stop depends on the closeness of the successive snapshots obtained. Since the target structure information is assumed to be unknown in this case, there is no chance for comparing the intermediate structures with the target. Hence in order to end the program, the RMSD between the present and the previous intermediate structures, i.e. successive RMSD's are recorded at each iteration (STEP 7) and average RMSD value of, for instance, 10 successive intermediate structures is computed. If this average RMSD is less than a desired value the program can be ended since this shows that the successive snapshots resemble each other and hence maximum

approach is attained with a stabilized RMSD profile. Otherwise, the iteration continues by performing ANM and obtaining new eigenmodes for the present conformation. It should be noted that a sliding window concept is applied in computing the average RMSD of 10 successive snapshots.

4. ANM-MC SIMULATION RESULTS

In this chapter, the ANM-MC methodology is used to investigate the conformational transitions in two proteins, namely adenylate kinase (AK) and hemoglobin. The degree of approach to the target structure and the mechanism by which the transition takes place are analyzed by calculating the RMSD between the simulation snapshots and the target conformation. RMSD values are calculated based on C^{α} coordinates of the structures obtained. The potential energies of the simulation snapshots are compared to those of starting and end structures. Moreover, the conformational transition pathway is analyzed by comparing the simulation snapshots with the x-ray structures proposed as intermediates in literature.

4.1. Case study: Adenylate Kinase (AK)

AK undergoes a significant hinge-bending motion upon substrate binding (Figure 2.5). The motion is highly collective with an RMSD change of 7.13 Å and possesses an allosteric behavior. There is large amount of experimental and computational data on the AK conformational transition, including proposed candidate structures existing on the transitional pathway (Maragakis and Karplus, 2005). Hence, AK is a suitable system for the evaluation of ANM-MC technique and the adjustment of simulation parameters.

The transition of AK from the apo (4AKE) to the bound/closed state (1AKE) is studied by ANM-MC methodology. The first conformers (chain A) in both x-ray structures (reporting two chains) are chosen. The water molecules and the ligand atoms are not taken into consideration. First, the collective modes of apo AK are investigated by ANM using the HingeProt web server (Emekli *et al.,* 2007). The slowest two modes are associated with the motion in the LID domain, i.e. LID opening and closing. In 3^{rd} and 4^{th} eigen modes, LID closing is accompanied with AMP-bind domain closing. Figure 4.1 presents the alternative conformations for the first four modes. However, ANM itself does not provide information about the sequence of these events, i.e. which domain closes first still remains a question.

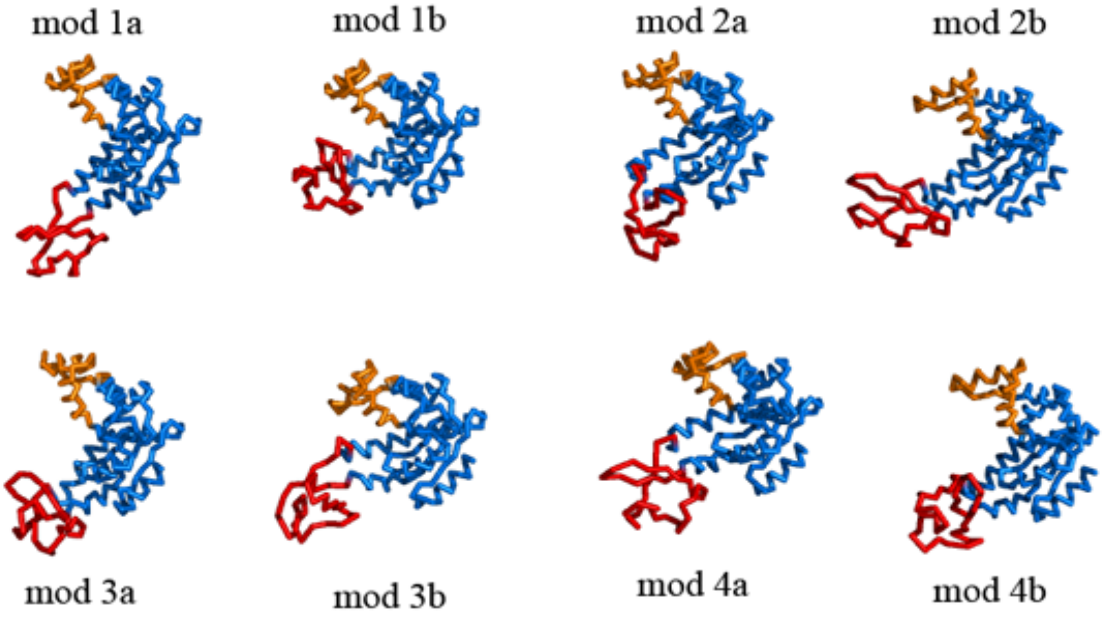


Figure 4.1. Four slowest collective modes of apo AK (PDB code: 4AKE)

4.1.1. Parameter Adjustment: RMSD and Energy Profiles in AK

As explained in Section 3, the simulation parameters of concern are the deformation factor (DF) applied along collective modes and the number of Monte Carlo Step (MCS) employed in each cycle. Hence, these two parameters are explored for parameter adjustment, using DF= 0.1, 0.2 and 0.5 Å and MCS = 100, 500 and 1000. Three independent runs are performed for each combination of DF and MCS, but results are shown for a single run because of the similarity of the different runs. Figure 4.2 presents the simulation results obtained for the transition from the open to the closed (target) conformation of AK for different MCS. In all cases, the x-axis represents the number of ANM-MC iterations or cycles. One iteration/cycle means an ANM calculation followed by deformation along a collective mode with the specified DF, and then followed by an MC simulation of certain steps (MCS = 100, 500, or 1000). In Figure 4.2 (a), the effect of MCS on the RMSD profiles is demonstrated for fixed DF=0.2 Å. These RMSD values are between the energy minimized intermediate structures obtained during the simulation and the target structure.

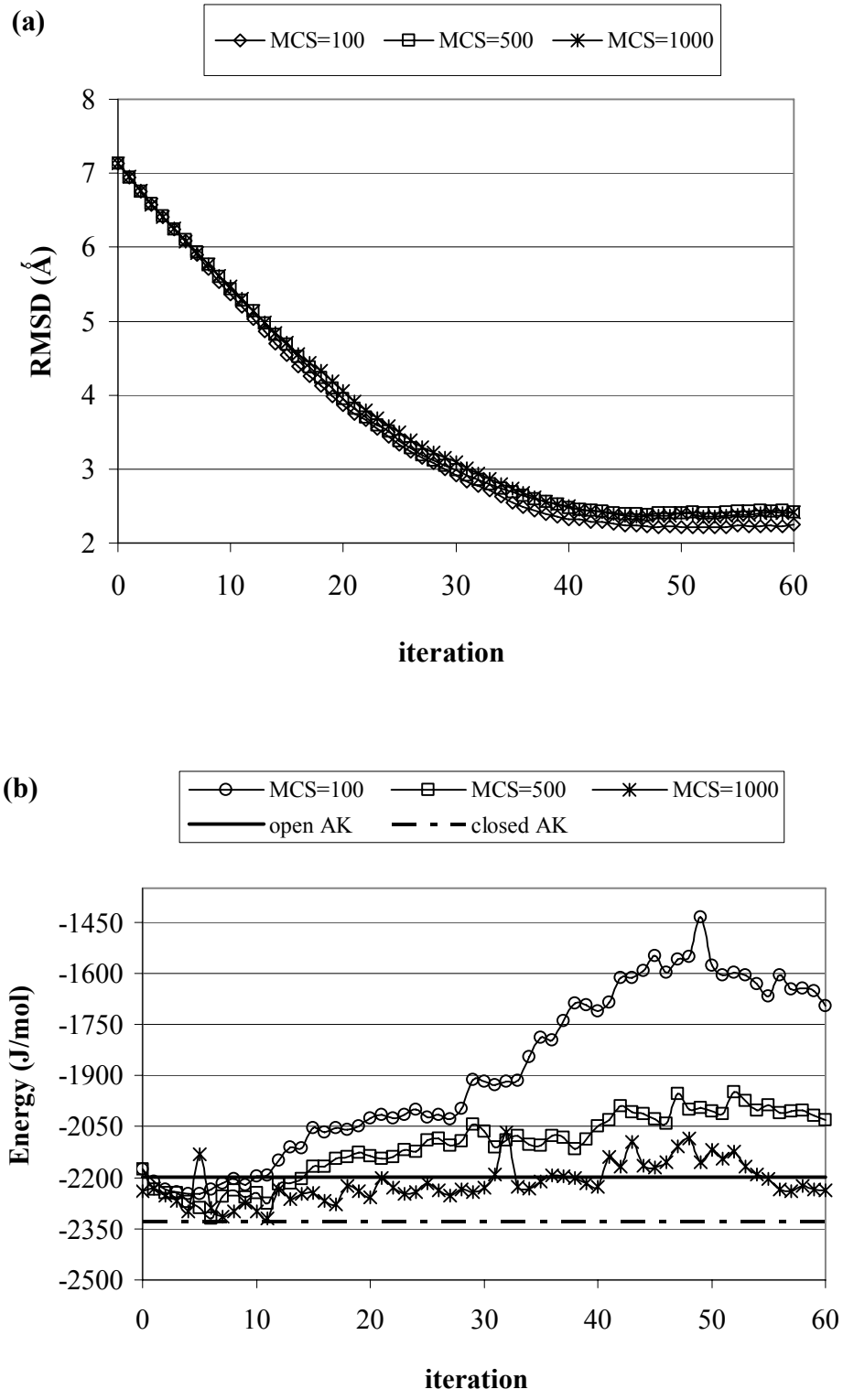


Figure 4.2. RMSD and energy values as a function of iteration/cycle number for various AK runs from open to closed states (a) Effect of MCS on RMSD values. (b) Effect of MCS on the total energies of the intermediate structures during the simulation (DF=0.2 Å)

DF is first fixed arbitrarily in order to determine MCS and then DF will also be adjusted by fixing the MCS. As observed from Figure 4.2 (a), RMSD value starts from 7.13 Å - the value between 4AKE and 1AKE- and smoothly decreases up to 2.27 Å (MCS=100); 2.29 Å (MCS=500); and 2.34 Å (MCS=1000) indicating a reasonable approach to the closed conformer. There is no significant difference in the RMSD profiles obtained for different MCS, i.e. RMSD to the target state is almost invariant with changing MCS. Thus it can be concluded that intermediate structures having similar backbone structures are obtained for this range of MCS values.

In general, longer MCS results in longer computational times hence shorter MCS would be preferred. However, the energies of the intermediates that are depicted in Figure 4.2 (b) should also be considered to determine a suitable MCS value. Figure 4.2 (b) demonstrates the effect of MCS on the total energy profiles (total energy=short-range + long-range energies, see Section 3, Eq. 3.8) of the intermediate structures throughout the simulation with DF = 0.2 Å. The x-ray structures of the open and closed conformers of AK are also relaxed by MC simulations and for comparison their average energies from the beginning to MCS = 1000 are depicted on the figure as horizontal lines. The energy of closed conformation (1AKE) is lower than that of open conformation (4AKE).

In contrast to the RMSD profiles, the total energies of these intermediates differ considerably with the MCS employed. Specifically, MCS=1000 in each cycle results in structures with lower energies that fall in the range of open and closed structure energies. As the backbone structures of these intermediates are somewhat similar, the reason for these energy differences should arise from side chain orientations. Indeed, as the components of the total energy is decomposed and examined, the total long-range and side chain short-range interactions (associated with side chain bond stretching, angle bending and bond torsion) obtained with longer energy minimizations are found to be lower than that of the shorter runs (not shown), hence leading to lower total energies. Thus, an MCS of 1000 is found to be suitable for energy minimization. Even longer MCS could be used in each cycle, however that would decrease the computational efficiency of the new protocol.

In Figure 4.3, the effect of deformation factor used in ANM calculations is demonstrated by fixing the MCS at 1000. In fact, many different DF values ranging from 0.1 to 3.0 Å are studied, but most appropriate results, in terms of RMSD and energy, are obtained for low DF values such as 0.1, 0.2, and 0.5 as depicted in Figure 4.3 (a). The variation in the RMSD to target is presented in this figure. With smaller deformation factors such as 0.1, 0.2 Å; smoothly decreasing RMSD profiles are obtained. The closest approaches to the closed conformer are 2.27 and 2.34 Å at the 104^{th} and 47^{th} cycles with DF = 0.1 and 0.2 Å, respectively.

Higher deformation factors (data shown for DF = 0.5 Å) lead to trajectories that approach the target faster (minimum RMSD = 2.38 at the 21^{st} cycle) but show a subsequent increase in RMSD. This unstable behavior around the target is due to deforming the structure about slow mode directions that have in fact very low overlap values with the target direction after the initial decrease in RMSD, which will be discussed in the subsequent section. Hence, for closer approaches and stable RMSD profiles, smaller deformation factors, such as 0.1 and 0.2 Å, are chosen to be more appropriate. However, applying DF=0.5 until the closest approach to the target (minimum RMSD) and then using smaller DF's such as 0.1 (see inset of Figure 4.3 (a)) is another choice which results in smoothly decreasing and non-oscillatory RMSD profiles as well.

Figure 4.3 (b) points out the effect of deformation factor on the total energy profiles. The average energies of the initial (open) final (closed) states are depicted as horizontal lines as before. Smaller DF's (0.1 or 0.2 Å) provide energy values that fall in the range of initial and target structure energies. However, this does not mean that using DF=0.5 will always result in high energy values.

In fact, quite reasonable energy profiles may also be obtained with DF = 0.5 Å as well, if longer energy minimizations, such as for 3000 MCS, are performed, as demonstrated in Figure. 4.4. Among the different parameter combinations employed for AK, DF = 0.1, 0.2 Å can be regarded as appropriate choices with MCS = 1000. It is necessary to further investigate whether or not the choice of these parameters significantly alters the transitional pathways obtained.

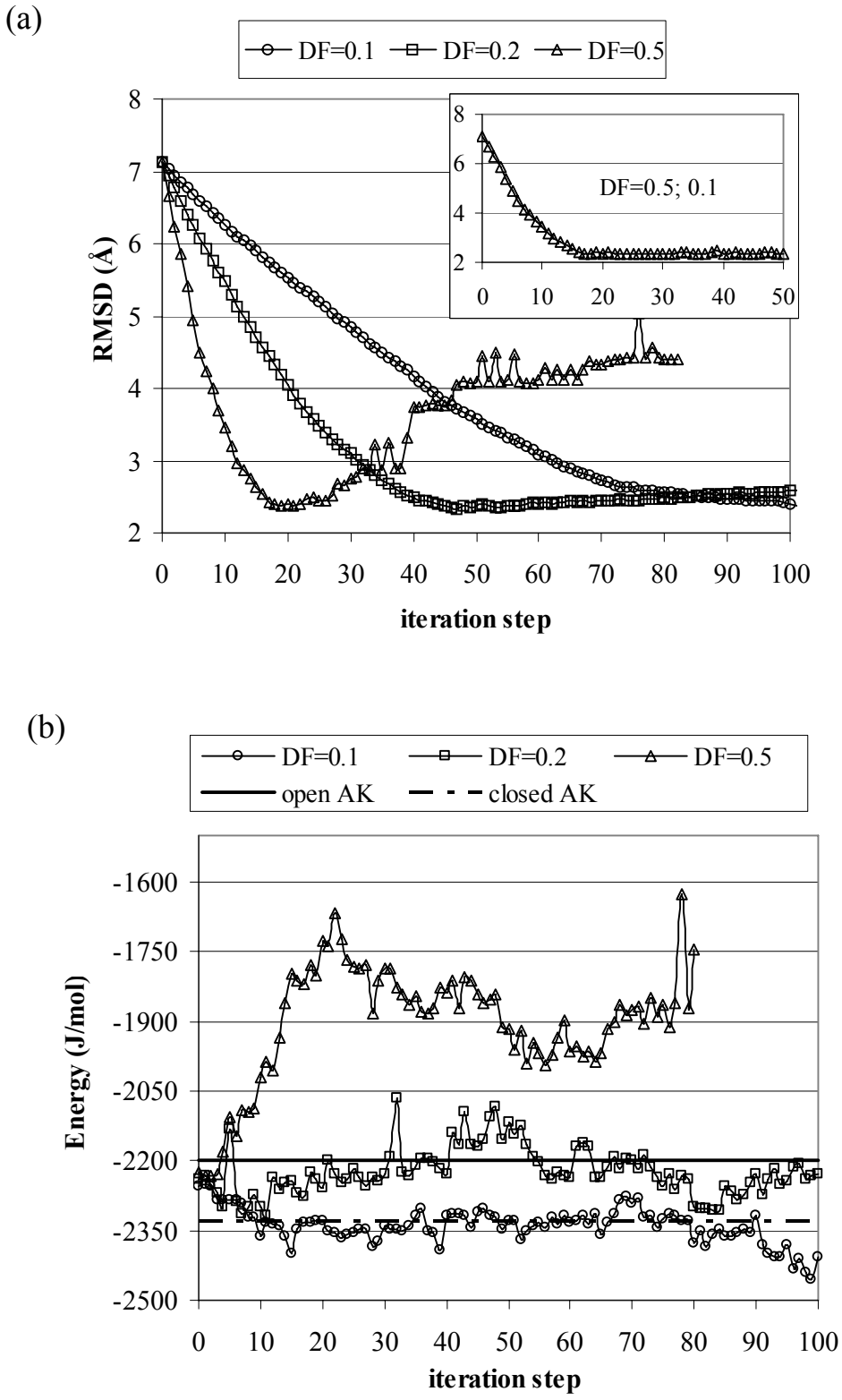


Figure 4.3. (a) RMSD of intermediate structures to target structure with various DF's (0.1, 0.2, 0.5 Å) with MCS= 1000. The inset presents results using DF =0.5 until maximum approach attained, then continuing with DF=0.1 (b) Energy profiles of AK

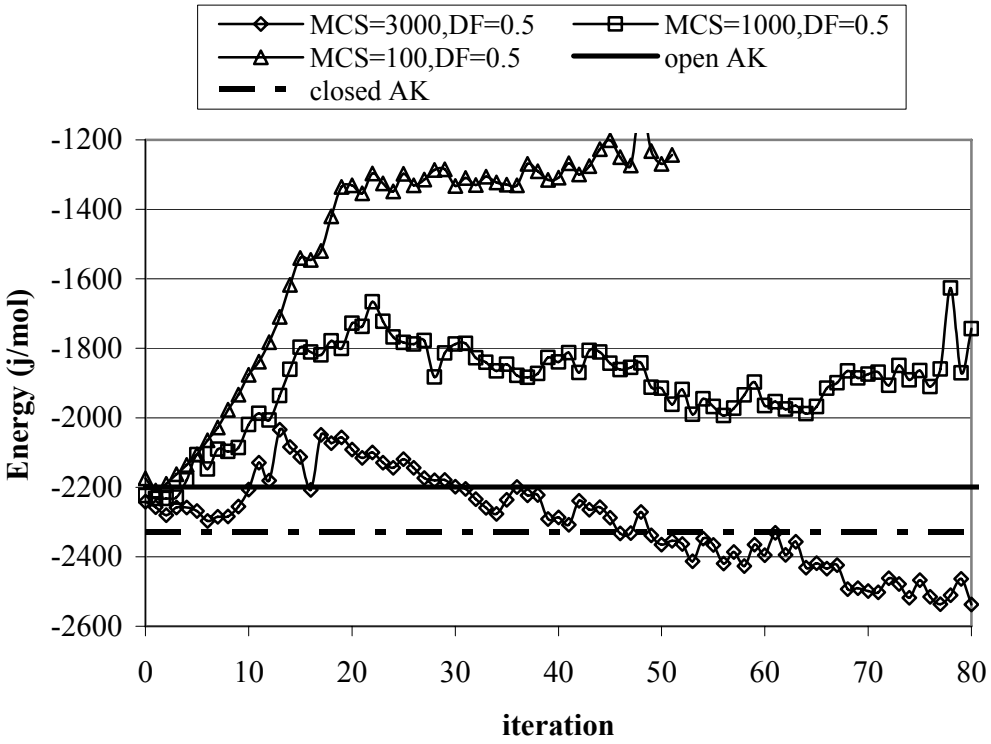


Figure 4.4. Effect of MCS on the energy profiles of AK intermediates with DF= 0.5

4.1.2. Transition Pathway and Pathway Intermediates of AK

The parameter combination of DF = 0.2 Å and MCS = 1000 will be used in this section for detailed analysis of the pathway intermediates. Figure 4.5 illustrates several snapshots up to 60th iteration taken from the ANM-MC simulation trajectory. During the initial cycles of the trajectory (up to snapshot ~30), the mobile LID domain (red) slowly closes over the core domain (blue). When the LID is almost closed, the AMP-bind domain (orange) begins to bend over the core. However, complete closure of the AMP-bind domain is not observed here. As a result, this visual inspection reveals that LID domain is more mobile as compared to AMP-bind domain and closure of the LID seems to precede that of AMP-bind domain.

The transition pathway of AK has been recently studied by a plastic network model study (Maragakis and Karplus, 2005), where several x-ray structures have been proposed

to be on the pathway. Table 4.1 tabulates the RMSD values between ANM-MC simulation snapshots depicted in Figure 4.5 and the x-ray structures proposed to be present on the transition pathway.

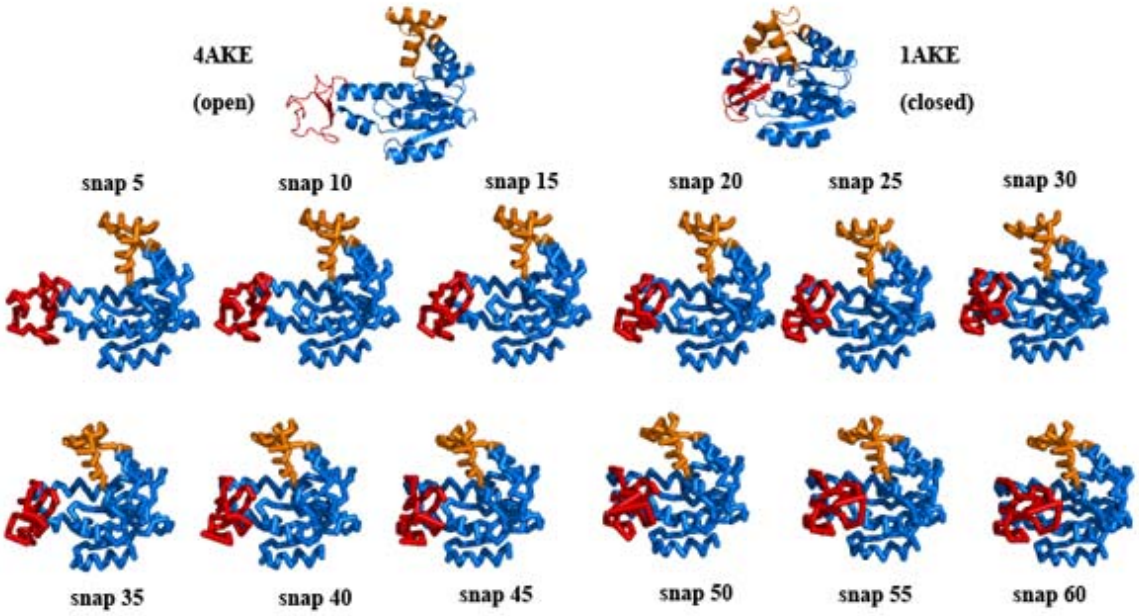


Figure 4.5. Several intermediate structures obtained during simulation of AK transition from open to closed conformation obtained for simulation with DF=0.2; MCS=1000.

RMSD values indicated in bold face represent the maximum approaches attained among the selected snapshots to the specific x-ray structures. The last two rows list the RMSD between the open (4AKE) and closed (1AKE) structures and the rest of the crystal structures studied. Earlier snapshots (iterations up to 30) exhibit maximum approach to the crystal structures: 1AK2, 1DVR (B chain) and 1DVR (A chain), sequentially. In 1AK2, the LID domain is about to close over the core, while the AMP-bind region is completely open. In 1DVR (A and B chains), the LID region is totally closed, while the AMP-bind is still open. Hence the overlap of these three x-ray structures with the earlier snapshots indicates the priority of LID closing.

Subsequent snapshots (iterations > 30) fall in proximity with the rest of four crystal structures, namely 1E4Y (A chain), 1E4V (A chain), 1ANK (A chain), and 2ECK (A chain). These crystal structures bear close resemblance to the closed conformer 1AKE, i.e. both the LID and AMP-bind domains are almost closed. Since the snapshots between 35th - 60th iterations fall within 2.3-2.8 Å RMSD range with these structures, the AMP-bind closing appears to take place- at least partially- in the late stages of the simulation.

Table 4.1. RMSD values between simulation snapshots (DF = 0.2 Å and MCS = 1000) and several AK crystal structures

Snapshot/ iteration	**4AKE**	**1AK2**[a]	**1DVR**[b] *B chain*	**1DVR**[b] *A chain*	**1E4Y**[c] *A chain*	**1E4V**[d] *A chain*	**1ANK**[e] *A chain*	**2ECK**[f] *A chain*	**1AKE**
5	2.09	4.09	4.21	4.10	5.93	6.23	6.25	6.31	6.25
10	2.71	3.55	3.51	3.35	5.19	5.46	5.44	5.53	5.47
15	3.63	**2.98**	2.86	2.67	4.49	4.70	4.65	4.75	4.71
20	4.46	3.14	**2.53**	2.38	3.92	4.05	3.99	4.08	4.05
25	5.04	3.45	2.54	**2.36**	3.40	3.49	3.44	3.52	3.49
30	5.30	3.64	2.73	2.60	2.98	3.09	3.06	3.14	3.09
35	5.73	3.98	2.75	2.63	2.66	2.74	2.71	2.76	2.74
40	6.23	4.17	2.92	2.81	2.52	2.50	2.47	2.51	2.49
45	6.56	4.54	3.26	3.12	**2.38**	2.42	2.38	2.40	**2.34**
50	6.94	4.84	3.55	3.42	2.46	**2.36**	2.37	2.38	2.37
55	6.95	4.96	3.54	3.41	2.43	2.39	**2.34**	**2.30**	2.38
60	7.02	4.95	3.51	3.40	2.50	2.42	2.40	2.30	2.41
4AKE	0.00	5.38	5.78	5.63	7.20	7.64	7.66	7.23	7.13
1AKE	7.13	5.57	4.02	3.92	0.93	0.65	0.45	0.28	0.00

[a] 1AK2 (Schlauderer and Schulz, 1996) belong to bovine mitochondria inter-membrane space AK.

[b] 1DVR (A and B chains) (Schlauderer *et al*., 1996) belong to the baker's yeast AK.

[c] 1E4Y (Müller and Schulz, 1993) is the Glysine-loop modified version of AK from *E.coli*.

[d] 1E4V (Müller and Schulz, 1993) is G10V mutant of AK from *E.coli*.

[e] 1ANK (Berry *et al*., 1994) is AMPPNP and AMP bound form of AK from *E.coli*,

[f] 2ECK (Berry *et al.,* 2006) is AMP and ADP bound form of AK from *E.coli*.

The present results are in conformity with recent studies pointing to the conformational changes and physical order of events during open to closed transition in AK. The plastic network model of Maragakis and Karplus (2005) has also indicated that LID closure precedes the bending motion of the AMP-bind region due to the highly flexible nature of the LID domain and its low elastic energy barrier cost. A coarse-grained structure-based Hamiltonian model proposed by Whitford *et al*. (2007) has revealed that the free energy barrier of LID domain closure is less than that of AMP-bind region, implying the AMP-bind closure as the rate-limiting step. An MD study by Lou and Cukier (2006) has indicated that the LID initially closes toward the core at high temperature.

Table 4.2 lists the closest approaches attained to x-ray structures in different AK runs carried out with various DF and MCS parameter combinations. The first value shows the minimum RMSD attained to each crystal structure and the value in parenthesis denotes the specific cycle, at which the maximum approach is attained. For instance, for DF=0.1 and MCS=1000, the closest approach attained to x-ray structure 1AK2 is 2.88 Å at 30^{th} iteration. Independent runs (1 and 2) performed using the same parameters (DF = 0.2 and MCS = 1000) provide similar intermediates. Moreover, different MCS (MCS = 100 or 1000) or DF (DF = 0.1, 0.2 or even 0.5, not shown) did not change significantly the minimum RMSD values attained to the x-ray structures. As a result, appropriate simulation parameters should be chosen according to the size of each system. For instance, in case of large proteins, the simulation time to reach the target would definitely increase due to the system size. In that case, choosing higher DF's such as 0.5 would be more efficient. Even though the MCS would also be preferred to be low for decreasing the computational time; it should be adjusted so as to attain comparable energy values with the initial/final state energies.

Performing long energy minimizations result in more appropriate intermediate structure energies with the cost of prolonged computational times. However, if the main topic of concern is the exploration of the transitional pathway, i.e. the pathway intermediate structures, relatively shorter energy minimizations and/or higher DF may be employed especially for large biological systems. In summary, the most important finding in ANM-MC simulations is that similar pathway intermediates are obtained for different combinations of parameters; even for those that lead to high energies compared to open and closed x-ray structures.

Table 4.2. Closest approaches attained to x-ray structures in different AK runs

ANM-MC Parameters	**1AK2**	**1DVR (B)**	**1DVR (A)**	**1E4Y (A)**	**1E4V (A)**	**2ECK (A)**	**1ANK (A)**	**1AKE**
DF=0.1, MCS=1000	2.88 30	2.52 40	2.30 40	2.35 100	2.32 100	2.30 100	2.30 100	2.29 105
DF=0.2, MCS=100	2.91 15	2.50 20	2.11 25	2.36 45	2.32 50	2.34 55	2.27 55	2.27 50
DF=0.2, MCS=1000 (1)	2.98 15	2.53 20	2.36 25	2.38 45	2.36 50	2.34 55	2.30 50	2.34 50
DF=0.2, MCS=1000 (2)	2.96 15	2.50 20	2.42 25	2.38 45	2.35 50	2.34 55	2.33 55	2.35 50

The mode directions chosen at each iteration with corresponding overlap values are plotted in Figure 4.6. As mentioned before, among the first ten slowest modes, the mode chosen at each step has the highest overlap with the target direction and the structure is deformed along this direction. Previous studies have reported that conformational changes are usually accompanied by the lowest frequency normal modes (Tama and Sanejouand, 2001). For DF = 0.2, MCS = 1000; initially the algorithm chooses the 1st and then 2nd slowest modes with high overlap values (0.5-0.7), which are associated with the LID closing. Up to 20-25 iterations, the conformational change is thus driven with the lowest frequency modes in conformity with previous studies and the RMSD to target structure

decreases from 7.13 Å to 4.44 Å. After 20-25 iterations, higher modes associated with AMP-bind domain come into picture.

The basic approach to the target state is accompanied by the lowest frequency two modes, but higher modes with lower overlap values provide for a more precise mapping of the structure on to target (down to 2.34 Å RMSD). This outcome is also supported by Patrone and Pande (2006), who has pointed out the relevance of using higher modes to map a conformational change. They reported that the low frequency modes typically bring the reference conformation about 50% closer to the target conformation based on their RMSD and further approach should be accompanied by higher modes. All the calculations presented so far are based on slowest ten eigen modes. It may be speculated whether there is any contribution or effect of including even higher modes. Hence, simulations are repeated by taking the first 20 slowest modes. Though the RMSD and energy profiles are quite similar, the maximum approach to target is attained as 2.17 Å for DF=0.2 and MCS=1000 (for ten mode case, this value was 2.34 Å). As an expected outcome, using higher modes provided a slightly more precise mapping to the target conformation.

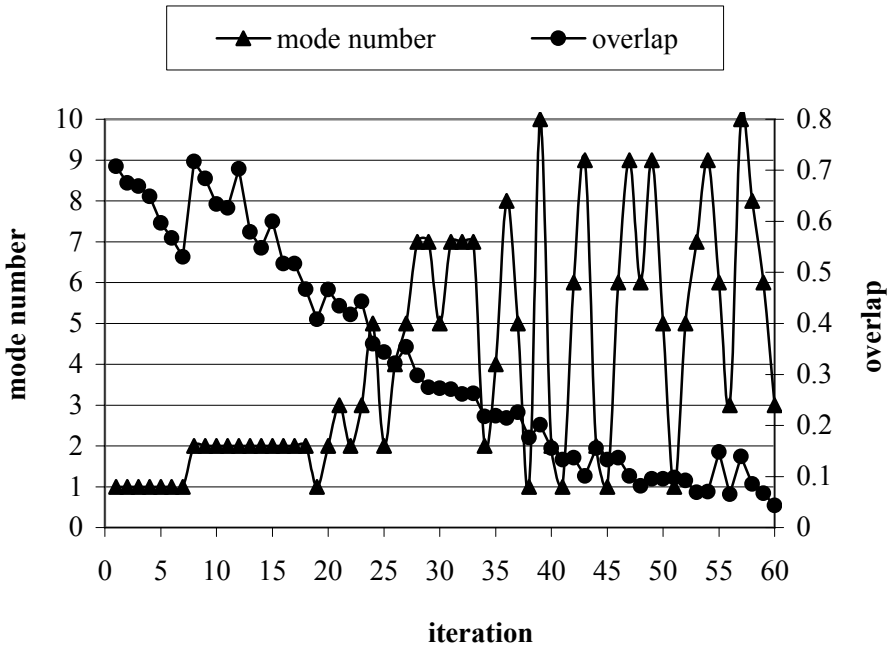


Figure 4.6. Mode directions preferred and corresponding overlap values at each iteration during simulation of AK from open-to-closed states (slowest 10 modes are included).

The mode directions preferred with corresponding overlap values are depicted in Figure 4.7 for the 20 mode case. When Figures 4.6 and 4.7 are compared, up to about iteration 20, the first two slowest modes are chosen, which drive the structure to undergo a large conformational change. Afterwards, higher modes are almost randomly chosen at each case for a closer approach. But the major conformational change is already driven by the slowest two modes with high overlap values, and hence similar intermediate structures are obtained throughout the trajectory. As a result, the first 10 modes seems to be a satisfactory choice.

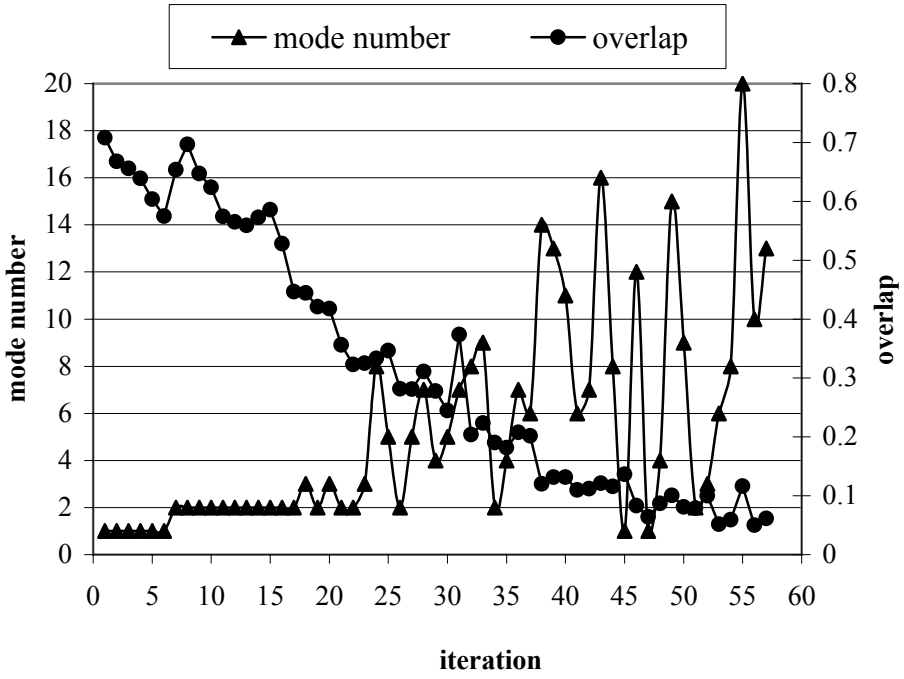


Figure 4.7. Mode directions preferred and corresponding overlap values at each iteration during simulation of AK from open-to-closed states (slowest 20 modes are included).

4.1.3. Contact Map Representations of AK Intermediates

The event of domain closure is explored more thoroughly at the residue level by plotting the residues contact maps. In Figure 4.8, the C^{α} atom pairs that are in contact within a cutoff distance of 15 Å are displayed for x-ray structures of open and closed conformers of AK in panels a and b, respectively. The contact maps are symmetric, hence only the upper diagonals are presented.

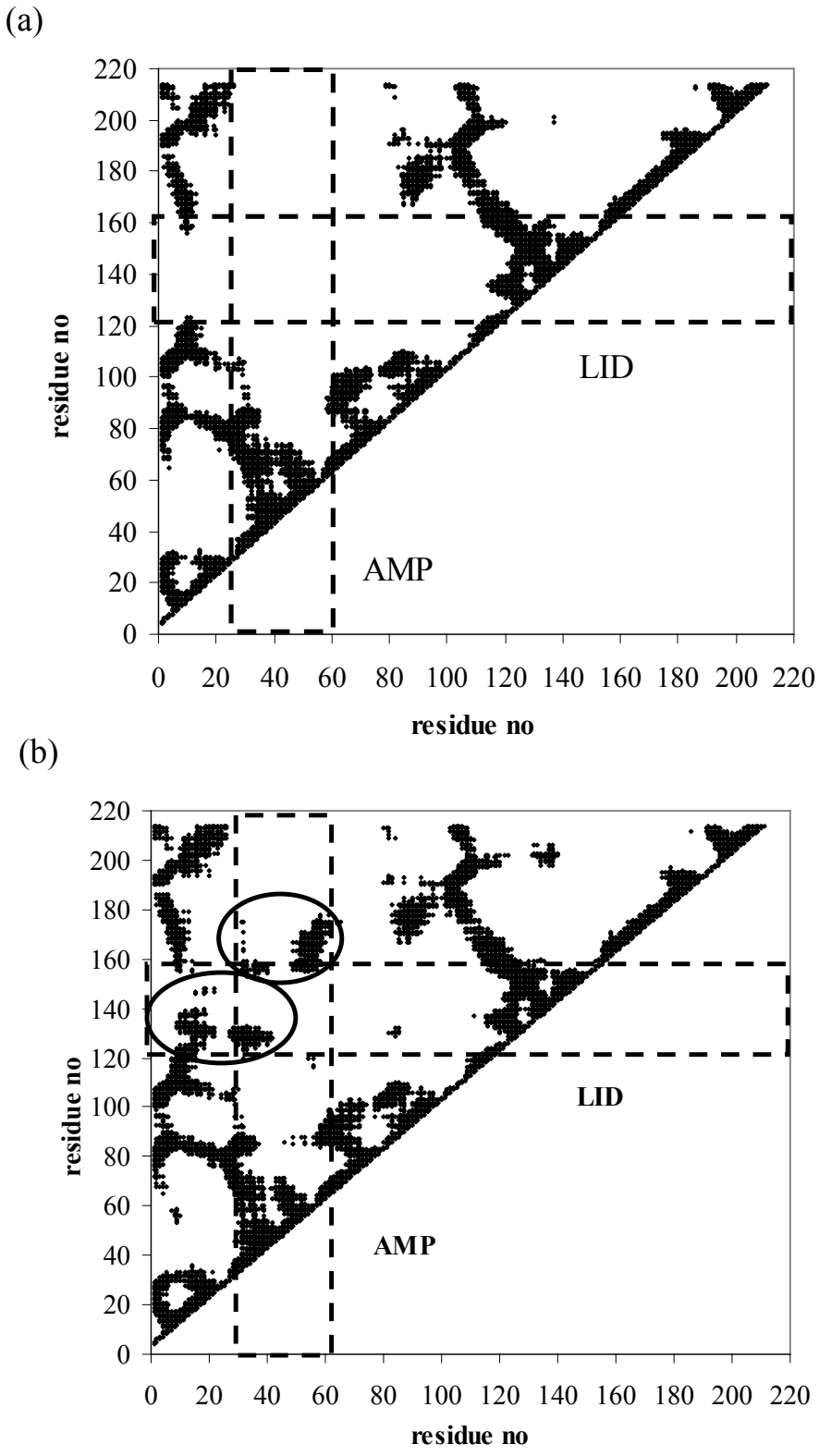
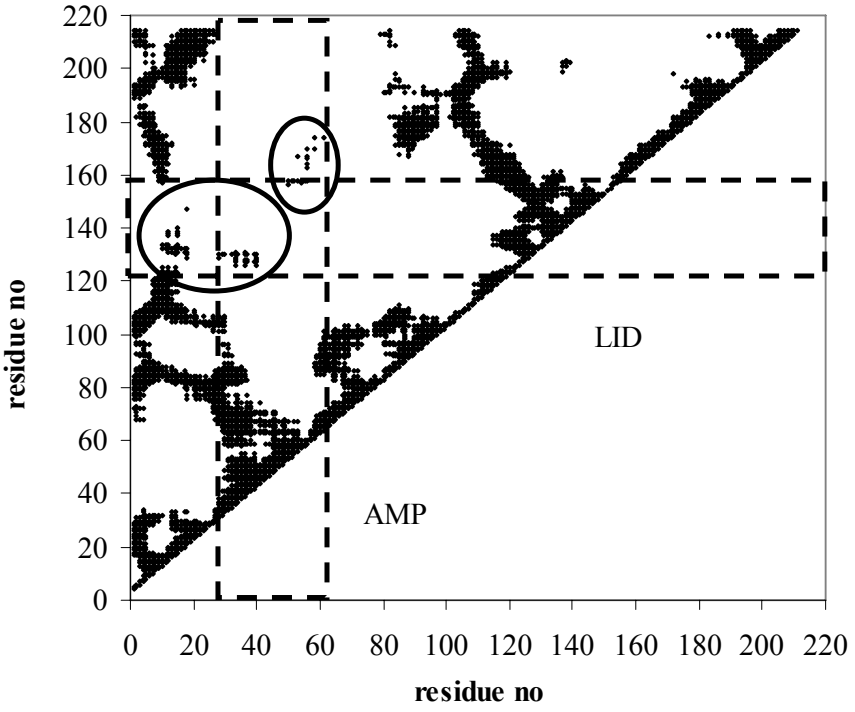


Figure 4.8. Contact maps for open (a), closed (b) forms of AK and for final snapshots (50th snapshot) of two independent simulations (c) and (d) with DF=0.2 Å and MCS=1000.

(c)



(d)

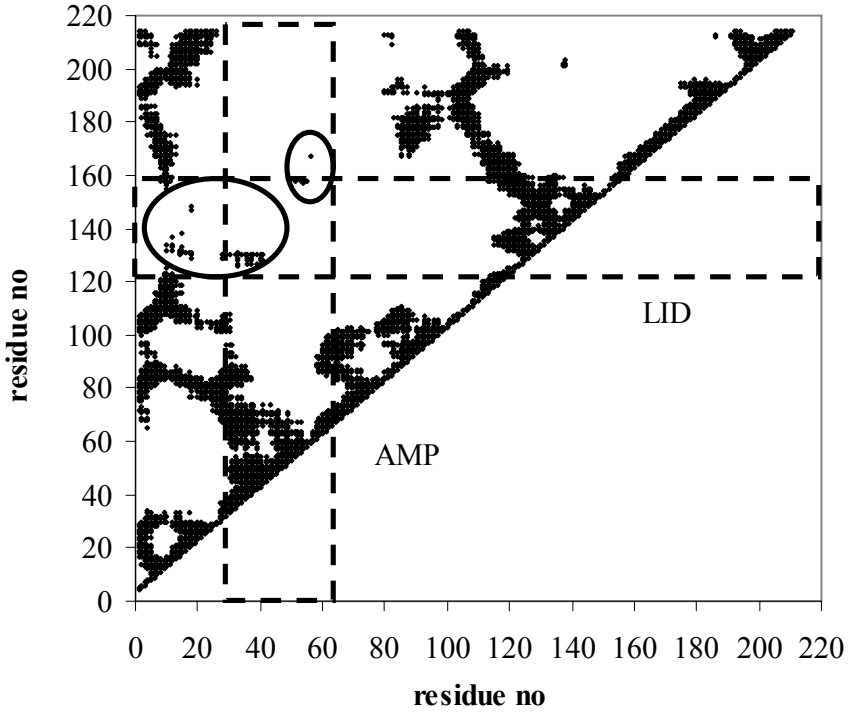


Figure 4.8. continued

The residues belonging to the LID and AMP-bind domains are highlighted within rectangles. The major differences in the residue contacts between the open and closed conformers are indicated within black circles in panel (b). Upon closure, the LID and the AMP-bind domains form close contact with each other and with the core domain. Similar analysis is carried out for the structures obtained at the end of simulation. Contact maps of final structures obtained from two independent simulations with DF = 0.2 Å and MCS = 1000 are generated and compared with that of the closed one. Figure 4.8 (c) and (d) show the plots for the 50th snapshot, which has attained maximum approach to the target state. Comparison with the target structure (Figure 4.8 (b)) indicates that in both runs an important part of the necessary contacts in black circles have been formed. The missing contacts are due to the fact that there is still a 2.34 and 2.41 Å RMSD of these snapshot with the target structure.

Table 4.3. Comparative analysis of new contact formation at each snapshot across two different runs of AK

Snapshots	**LID contacts (%)**		**AMP-bind contacts (%)**		**common residues in contact (%)**
	Run1	**Run2**	**Run1**	**Run2**	
5	14.1	13.0	0	0	80
10	15.8	15.8	0	0	75
15	18.1	18.1	0	0	78
20	18.1	18.6	1.0	1.0	75
25	18.6	22.0	1.0	1.4	69
30	22.6	22.6	3.1	4.1	65
35	25.4	25.4	5.1	5.8	65
40	26.0	26.0	6.1	6.5	60
45	28.2	29.4	8.5	9.9	67
50	31.1	29.9	11.0	10.2	57

The comparative quantitative analysis of new contact formation at each snapshot of two independent runs is tabulated in Table 4.3. The table shows cumulative percentage of LID (first column) and AMP-bind (second column) related contacts formed at each

snapshot and the overall percentage of common residues (third column) in contact at that particular snapshot across the two runs (both with DF = 0.2 Å and MCS = 1000). The percentages of LID and AMP-bind related contacts are calculated by taking the ratio of the number of contacts present at that snapshot to the corresponding number of contacts found in the target conformation (1AKE). The percentages of contacts belonging to LID and AMP-bind domains are very similar across the two runs. Both runs confirm the priority of LID closing which is then accompanied by AMP-closing. At the final snapshot (50) at most about 30% of LID contacts and 10 % of AMP-bind contacts that are to be present in the final closed structure are formed. The overall percentage of common residues is higher especially at initial stages of the simulation. Nevertheless, multiple independent runs result in a similar sequence of events (Tables 4.2 and 4.3) and MC serves for refining the conformations with the current implementation.

It should be noted that reverse transition from the closed to the open state of AK could not be accomplished with the current ANM-MC methodology as effectively as the forward (open to closed) transition. The RMSD to the open structure, which is the target in this case, just decreases to about 5 Å, i.e. the final structure obtained could not attain a close mapping to the target, i.e. close mapping to the target structure could not be attained. This will be discussed in the next chapter in comparison with the TMC simulations.

4.1.4. Stability of end structures

Longer MC simulations are performed on the structures obtained at the end of ANM-MC, in order to validate their stability. Figure 4.9 illustrates the RMSD and energy profiles of the final snapshot (2.34 Å RMSD to target) of an ANM-MC simulation during prolonged MC simulation (10,000 MCS). In panel a, the RMSD of MC simulation snapshots from the target is depicted. The RMSD values oscillate between 2.2-2.6 Å around the initial value (2.34 Å).

In Figure 4.9 (b), the energy profile is compared with the average energies of open and closed states (horizontal solid lines). The energy trend attains a stable profile near the initial and final structure energies.

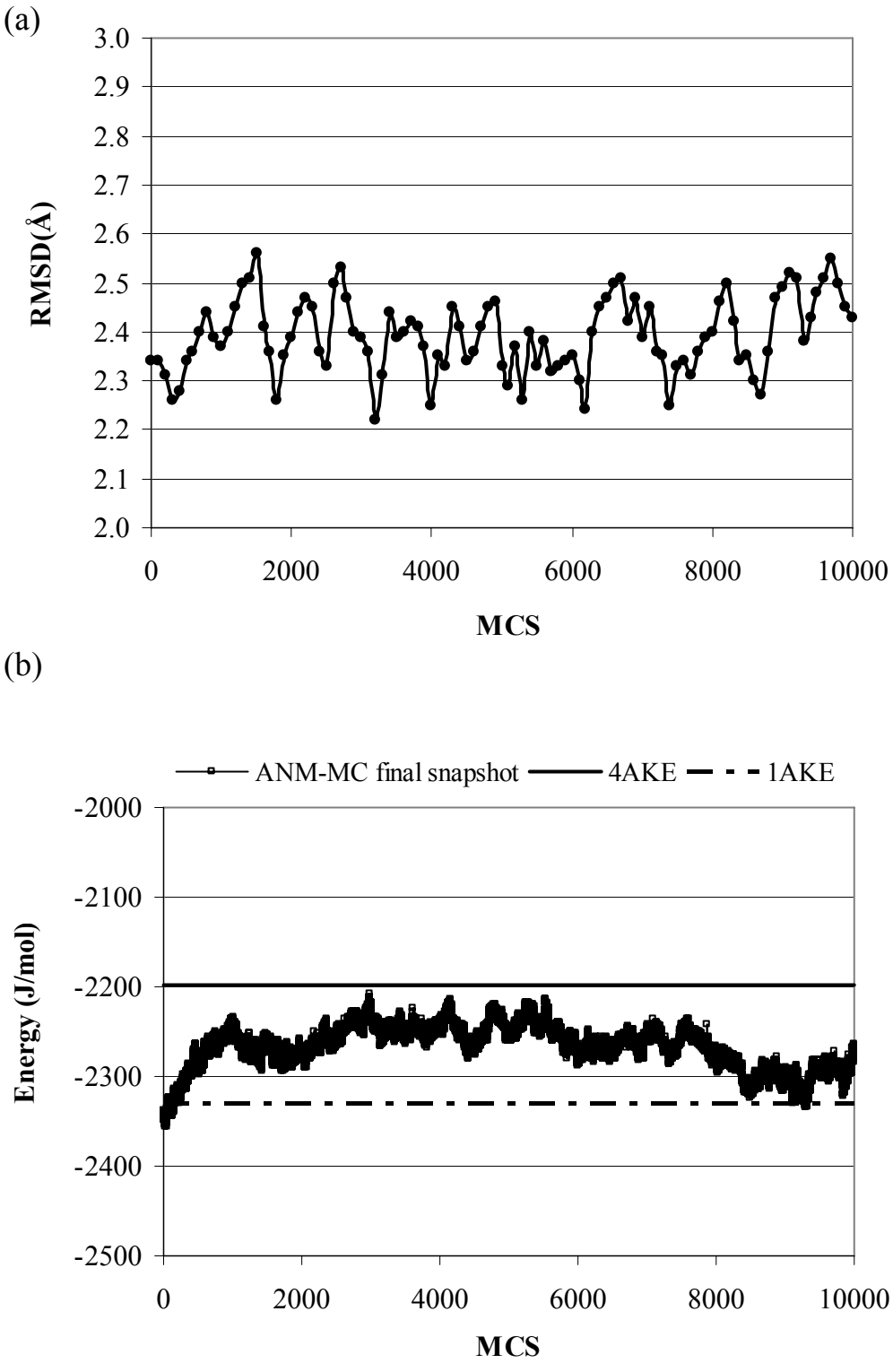


Figure 4.9. RMSD(a) and Energy (b) profiles of the final snapshot of an ANM-MC simulation during prolonged MC simulation

4.2. Case study: Hemoglobin

An important and widely investigated allosteric protein is hemoglobin, which consists of 4 chains with a total number of 572 residues. The system size is more than twice compared to AK resulting in longer computational times. ANM-MC methodology will also be tested for the case of hemoglobin, which exhibits a conformational change of 3.5 Å between T and R2 states. The RMSD between T-R and R-R2 states are 2.4 and 1.8 Å, respectively, where the R state has been proposed as an intermediate (Srinivasan and Rose, 1994). T-to-R2 transition will be studied and the transitional pathway will be analyzed from the aspect of approach of the simulation trajectory to the suggested pathway intermediate R-state. In R2, there are two additional valine residues in the beginning of chains B and D, which are discarded here to match the size of the initial and target structures.

In the study of Xu et al. (2003) the global dynamics of both T and R2 forms were investigated by ENM. It was reported that both of the conformations exhibit similar behavior. Moreover, the two α-chains exhibit similar dynamics as well as the two β-chains (Fig. 2.4). ANM modes indicated that the slowest mode is associated with the motion of the dimer α1-β1 relative to α2-β2, while the second slowest mode controls the relative movements of the dimers α1-β2 and α2-β1.

4.2.1. RMSD and Energy Profiles of Hemoglobin

As in the case of AK, simulation parameters (DF and MCS) are investigated first for hemoglobin. Figure 4.10 presents the ANM-MC simulation results obtained for hemoglobin with different DF values by fixing MCS at 1000. Similar to the case of AK, deformation factors of 0.1, 0.2 and 0.5 provide close approaches to target with stable profiles (panel a). Minimum RMSD values to target structure (R2) are 1.9 Å with DF = 0.1 and 0.2, and 1.95 Å with DF = 0.5. Figure 4.10 (b) demonstrates the energy values of the snapshots obtained.

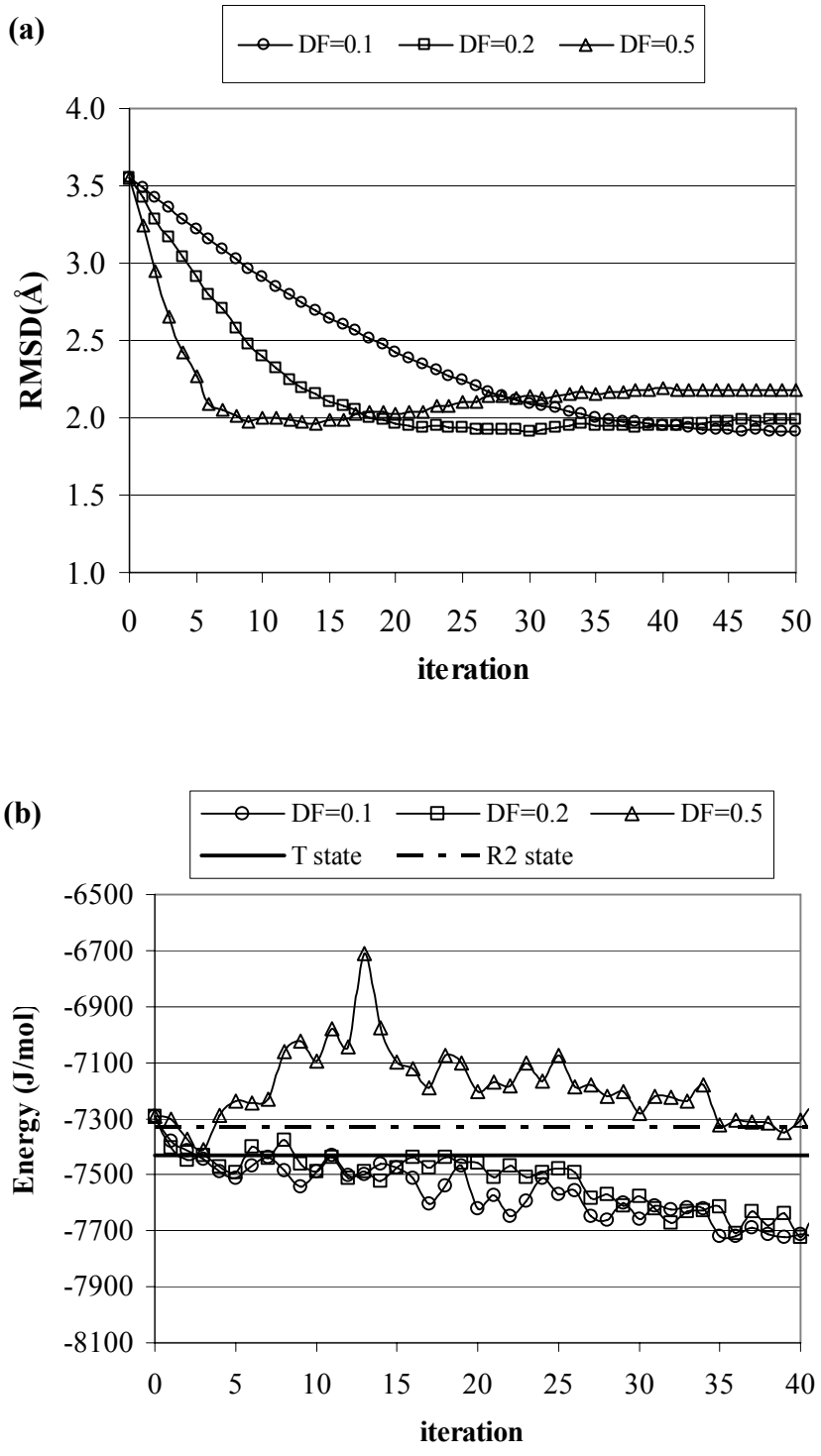


Figure 4.10. RMSD values of human hemoglobin (a) and corresponding energy profiles during simulation of transition from T-to-R2 form with different DF's (b).

All three values of DF lead to reasonable RMSD and energy profiles. Indeed, DF=0.5 Å seem to be a more appropriate choice in this case due to similar RMSD profiles with DF=0.1 and 0.2; and closer energy values to target structure. Moreover, the simulation time is almost half of that of the ones with DF=0.1 and 0.2.

4.2.2. Transition Pathway and Pathway Intermediates of Hemoglobin

The RMSD values of snapshots with T, R and R2 crystal structures are provided in Table 4.4 for the run performed with DF = 0.2 and MCS = 1000. The pdb file of 1HHO includes two monomers, hence the tetrameric structure is generated symmetrically using Pymol (DeLano, 2002). Srinivasan and Rose (1994) had suggested R state to be an intermediate state between T and R2. In accordance, intermediate snapshots of the simulation pass through R on the pathway from T to R2. As the snapshots begin to deviate from the T state, the R state is reached with a minimum RMSD = 1.89 Å (at 15th iteration). Subsequently, the R2 state is approached with an RMSD = 1.91 Å (at 30th iteration).

Table 4.4. RMSD values of the simulation snapshots with crystal structures of hemoglobin (T, R and R2 forms) with DF=0.2; MCS=1000

Snapshots	1A3N(T)	1HHO (R)	1BBB (R2)
5	1.04	2.08	2.91
10	1.99	2.04	2.40
15	2.54	**1.89**	2.10
20	2.89	1.94	1.97
25	3.06	2.04	1.93
30	3.07	1.98	**1.91**
35	3.20	2.14	1.95
40	3.24	2.12	1.94
45	3.30	2.19	1.97

Similar intermediate structures are obtained with DF = 0.5 and MCS = 1000 as presented in Table 4.5. Thus for relatively large proteins like hemoglobin, DF = 0.5 seems more appropriate in terms of computational efficiency, since the RMSD profiles and the intermediates are quite similar with the case of lower deformation factors. An important outcome of ANM-MC simulations is that although the simulation was targeted in the direction of T-R2 transition, on the path another state, the R state is visited first. As a result the snapshots reveal that R like structures exists on the hemoglobin transitional pathway.

Table 4.5. RMSD values of the simulation snapshots with crystal structures of hemoglobin (T, R and R2 forms) with DF=0.5; MCS=1000.

Snapshots	**1A3N(T)**	**1HHO (R)**	**1BBB (R2)**
2	0.82	1.88	2.95
5	2.48	2.24	2.27
7	2.72	**1.79**	2.05
10	2.94	1.81	2.01
13	2.95	1.90	1.99
15	3.27	2.26	**1.97**
18	3.30	2.03	2.04
20	3.33	2.35	2.05
23	3.35	2.34	2.08

The directional preferences, i.e. the modes preferred at each iteration with corresponding overlaps during the simulation are depicted in Figure 4.11. In the initial stage (up to 16th iteration), the first two global modes guide the major part of the transition. After that point higher modes with low overlap values (around 0.1) slightly decrease the RMSD with target from 2.1 to 1.9.

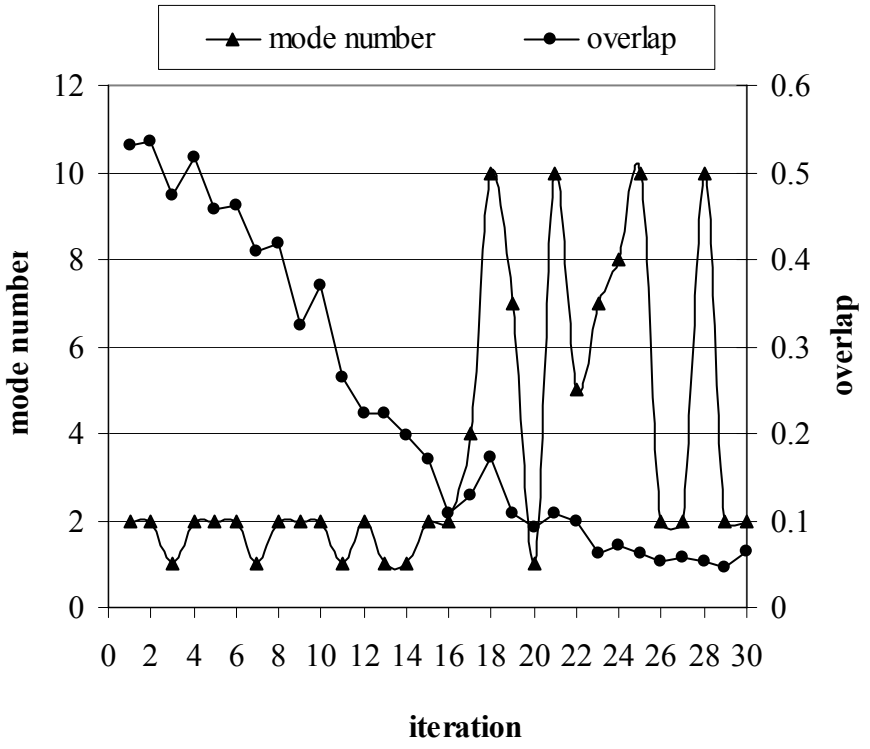


Figure 4.11. Mode directions preferred and corresponding overlap values with the target direction obtained for hemoglobin from T to R2 state (DF=0.2; MCS=1000 case).

Computational efficiency

Besides its simplicity, the ANM-MC algorithm requires reasonably short CPU times for completion. For instance, a single iteration for AK (214 residues), which includes the calculation of 10 slowest modes by ANM, followed by energy minimization of MCS = 100 (MCS = 1000) and an RMSD check, takes about 1 minute and 17 seconds (6 min and 10 seconds) on a 1.5 GHz Itanium2 processor with 2 GB RAM. A complete run that drives AK from the open to closed conformation by attaining a maximum approach lasts approximately 1 hour (5 hours) for MCS = 100 (MCS = 1000) and DF = 0.2 Å. In the case of hemoglobin (572 residue), a complete run with DF = 0.2 lasts about 5 days with MCS = 1000 and less than one day with MCS = 100, providing similar intermediates. Thus, a rough comparison of the CPU times necessary to simulate such conformational transitions with targeted MD and ANM-MC method indicates that the latter is much more efficient requiring couple of days, whereas the former may last on the order of weeks to months, depending on the system size and simulation parameters.

5. TARGETED MONTE CARLO SIMULATION RESULTS

TMC simulations are carried out to assess the effect of incorporating normal mode directions in the ANM-MC methodology. Similarly, starting with an initial conformation (e.g. open state), the protein is forced toward the target (e.g. closed state) conformation by deforming the coordinates along the target direction without the aid of slow modes. This targeted protocol is applied to AK and hemoglobin and the results are compared to those of the ANM-MC method.

5.1. Case Study: Adenylate Kinase (AK)

5.1.1. Open-to-closed transition

The RMSD and energy profiles for TMC simulations of AK are presented in Figures 5.1 (a) and (b), respectively. In these figures the results obtained in TMC and ANM-MC methodologies are compared. The TMC method reaches the target structure faster with a final RMSD ~ 0 Å, indicating almost a complete match of the final snapshot with the target conformation due to the forcing mechanism towards target. The energy profiles of two methodologies are quite similar. The trends are sensible as compared to the average energies of the open (4AKE) and closed (1AKE) states, which are indicated as horizontal lines.

Comparisons of TMC snapshots with the AK related x-ray structures are provided in Table 5.1. The RMSD (2.5 - 3.4 Å) of the TMC snapshots with the first four x-ray structures associated with LID closure (1AK2, 2AK2, 1DVRB, 1DVRA) are higher than the ones obtained with ANM-MC (2.3 – 2.9 Å, in Table 4.1). In contrast, TMC approaches closer to the rest of the structures (1E4YA, 1E4VA, 2ECKA, 1ANKA), that have very similar conformations to the target (RMSD < 1 Å). Hence, as compared to ANM-MC method, TMC succeeds in attaining close proximity with the final four x-ray structures that mainly resemble the target.

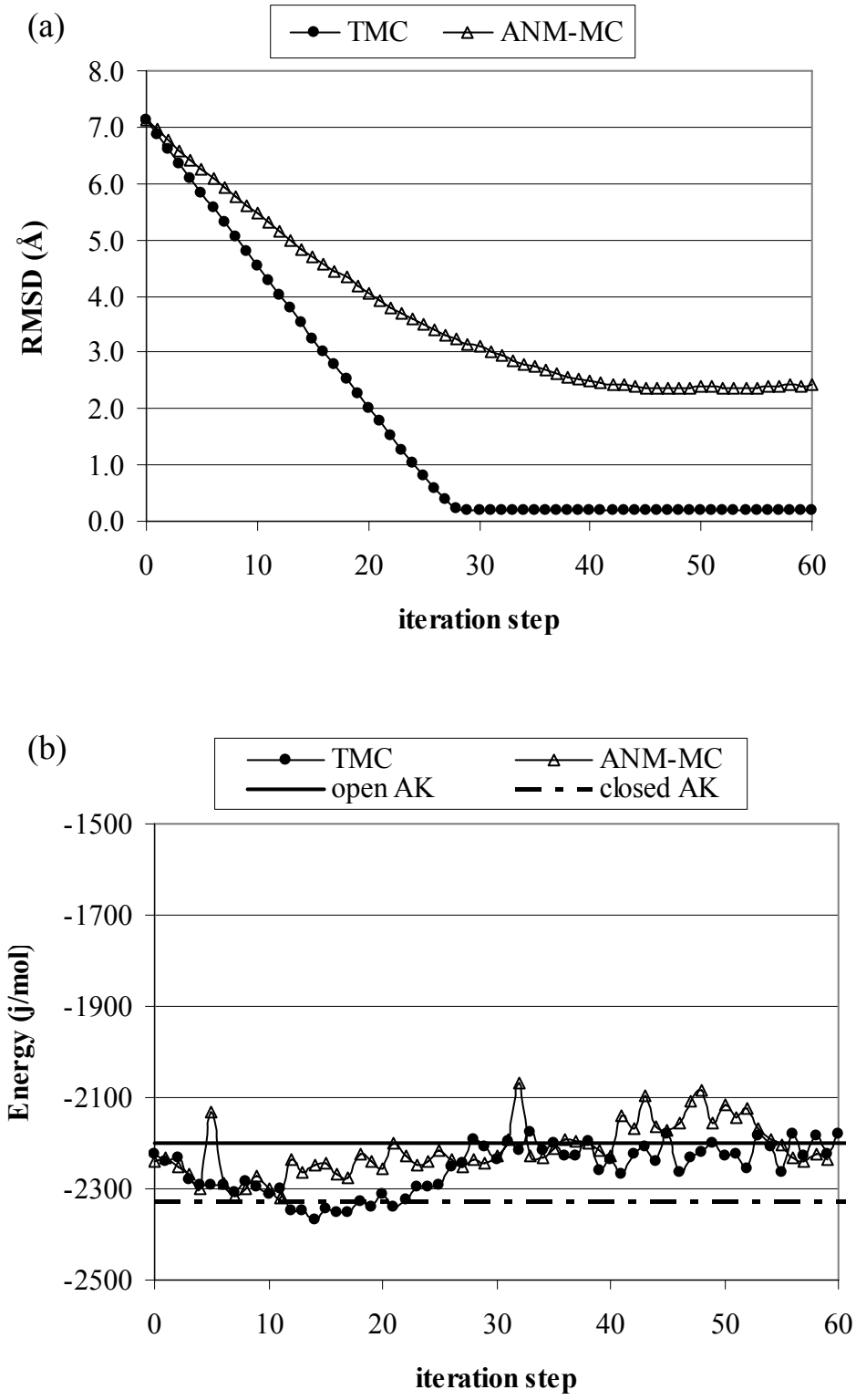


Figure 5.1. TMC simulation results as compared with ANM-MC method with application to AK (DF=0.2 Å; MCS=1000) (a) RMSD between the intermediate snapshots and target structure (b) Energy profiles of intermediates.

Table 5.1. RMSD values between the AK related x-ray structures and intermediate structures obtained with TMC simulations of AK from open-to-closed conformation.

snapshots	4AKE	1AK2	1DVR B	1DVR A	1E4Y A	1E4V A	2ECK A	1ANK A	1AKE
5	2.20	3.68	3.92	3.80	5.52	5.82	5.91	5.81	5.83
10	3.26	**3.43**	3.09	2.95	4.24	4.51	4.62	4.55	4.53
15	4.44	3.53	**2.70**	**2.55**	2.99	3.24	3.33	3.26	3.25
20	5.62	4.14	2.88	2.74	1.84	2.00	2.04	2.02	2.01
25	6.80	4.99	3.52	3.40	**0.96**	0.83	0.95	0.81	0.81
30	7.55	5.59	4.08	3.96	0.96	**0.28**	**0.33**	**0.48**	**0.15**

5.1.2. Reverse Transition

Even though most studies in literature focus on the open-to-closed conformational transitions, the reverse transition was also studied in AK by ANM-MC and TMC methods. The results pointed out that TMC succeeds in generating the reverse (closed-to-open) pathway with similar pathway intermediates obtained in the forward (open-to-closed) transition simulations.

On the other hand, the ANM-MC protocol can not reach the open conformation to the extent that TMC achieves. In the case of ANM-MC simulation, when the overlap values of the forward and reverse transitions are compared, it is observed that in the forward case, the overlap values of modes are quite high, whereas in the reverse transition case they are low.

In a comprehensive study (Tama and Sanejouand, 2001) performed over 20 different proteins exhibiting open/closed transitions both forward and reverse transitions were investigated by NMA. The authors concluded that when studying an open conformation the normal modes better described the conformational change, the corresponding overlap being significantly higher. In other words, the reverse transitions from closed-to-open states do not usually correspond to the slow modes obtained from normal mode analysis. This was attributed to the possibility that the property captured by

NMA may be related to the shape of the protein. The domains of the protein in the open form are more separated and better defined; hence NMA performs better with open forms as far as the shape is concerned. Since ANM-MC is a normal mode based protocol, the reverse transition of AK could not be accomplished. Due to the forcing algorithm of the TMC method, both the forward and the reverse transitions could be successfully simulated.

In Figure 5.2, RMSD and energy profiles of AK obtained by ANM-MC and TMC simulations of transition from the closed-to-open conformation, i.e. for reverse transition are given. In the ANM-MC case, the RMSD to the open conformation has just decreased from 7.13 to 5.02 Å at most (Figure 5.2(a)) and the energy profile (Figure 5.2 (b)) of the simulation trajectory is observed to be quite high as compared to the open and closed conformer average energies (indicated by horizontal lines). There seems to be large energy barriers to be overcome in order for the closed structure to open, which is not possible by the aid of normal modes.

In the case of TMC, almost a complete mapping to target with RMSD ~ 0 Å is attained. Energy values of both methods are comparable with the initial and target state average energies (panel b). As a matter of fact, TMC method proves more successful for accomplishing the reverse transition.

Table 5.2 tabulates the RMSD between the TMC snapshots of reverse transition simulation and AK related x-ray structures. Snapshots are between 2.2-2.9 Å proximity with the three open-like x-ray structures (1DVRA, 1DVRB, 1AK2). The TMC snapshots are initially close to the first four x-ray structures (2ECK, 1ANK, 1E4V, 1E4Y) which resemble the closed structure, and continuously move away as the simulation proceeds.

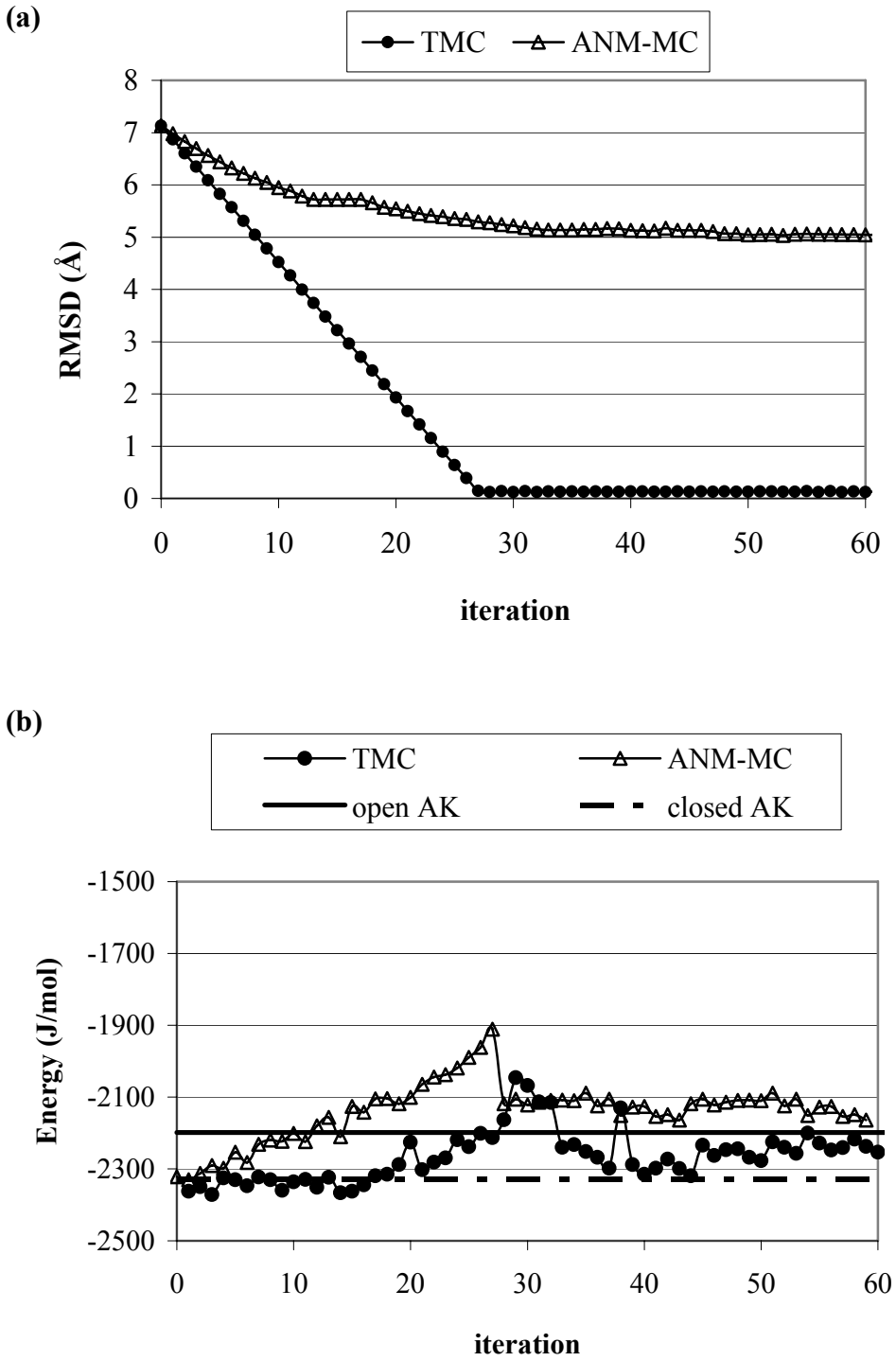


Figure 5.2. RMSD (a) and energy (b) values as a function of iteration/cycle number for reverse transition simulation of AK by ANM-MC and TMC methods (DF=0.2, MCS=1000)

Table 5.2. RMSD values between the AK related x-ray structures and intermediate structures obtained for reverse transition of AK with TMC simulation

snapshots	1AKE	2ECK A	1ANK A	1E4V A	1E4Y A	1DVR B	1DVR A	1AK2	4AKE
5	3.25	1.26	1.26	1.23	1.09	2.54	2.43	4.05	5.82
10	4.06	2.32	2.31	2.27	2.01	2.21	2.51	3.27	4.52
15	5.39	3.37	3.37	3.37	2.99	**2.38**	**2.23**	2.99	3.22
20	5.87	4.41	4.42	5.21	4.24	2.60	2.61	**2.93**	1.92
25	6.22	5.33	5.44	5.43	5.14	3.19	3.11	3.48	0.63
30	7.10	5.89	6.10	5.93	5.77	3.76	3.53	3.87	0.12

The mode directions chosen at each iteration with corresponding overlap values are plotted in Figure 5.3 for ANM-MC reverse transition of AK. The comparison of this figure with the one obtained in the forward transition case (Figure 4.6) reveals that the overlap values of modes with the target direction are considerably lower in the reverse transition case.

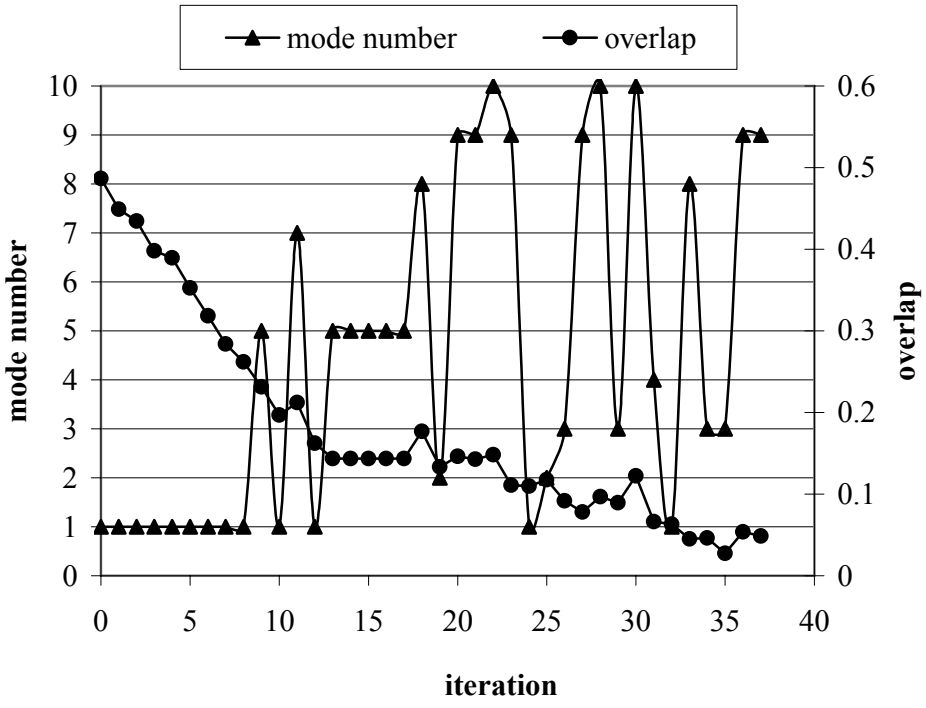


Figure 5.3. Modes chosen and corresponding overlap values at each iteration for ANM-MC reverse transition simulation of AK (DF=0.2; MCS=1000)

This implies that normal modes better describe the conformational changes taking place in the forward transition case. In the reverse transition case, at 10^{th} iteration, the overlap value has already fall down to 0.2. However, in the forward case, such a small overlap value was attained after about 40^{th} iteration until which almost maximum approach was attained.

5.2. Case Study: Hemoglobin

5.2.1. T-to-R2 transition

TMC simulations are also carried out for hemoglobin. Figure 5.4 illustrates the comparative results obtained by TMC and ANM-MC methods. TMC is examined to approach the target state R2 starting T state with an RMSD of almost 0 Å approach as in the case of AK. On the other hand, the maximum approach value attained by ANM-MC was 1.91 Å for DF=0.2; MCS=1000.

The RMSD and energy profiles for TMC simulations of hemoglobin are depicted in the Figures 5.4 (a) and (b), respectively, for comparison of TMC and ANM-MC methodologies. It may be inferred that TMC exhibits a sharp decrease of RMSD towards the target with energy profiles lower than the one obtained with ANM-MC method. However, in the later snapshots the energy profile seems to be higher than ANM-MC case. Nevertheless, the energy values are comparable with that of the target state.

Further analysis of the simulation trajectory is demonstrated in Table 5.3. The interpretation of the simulation snapshots with the initial (T state), final (R2 state) and the known intermediate (R state) of the simulation trajectory revealed that the intermediate R state is approached by 1.28 Å at 13^{th} snapshot and beyond 30^{th} snapshot, the structure does not change much as the RMSD from the target is around 0.1 Å.

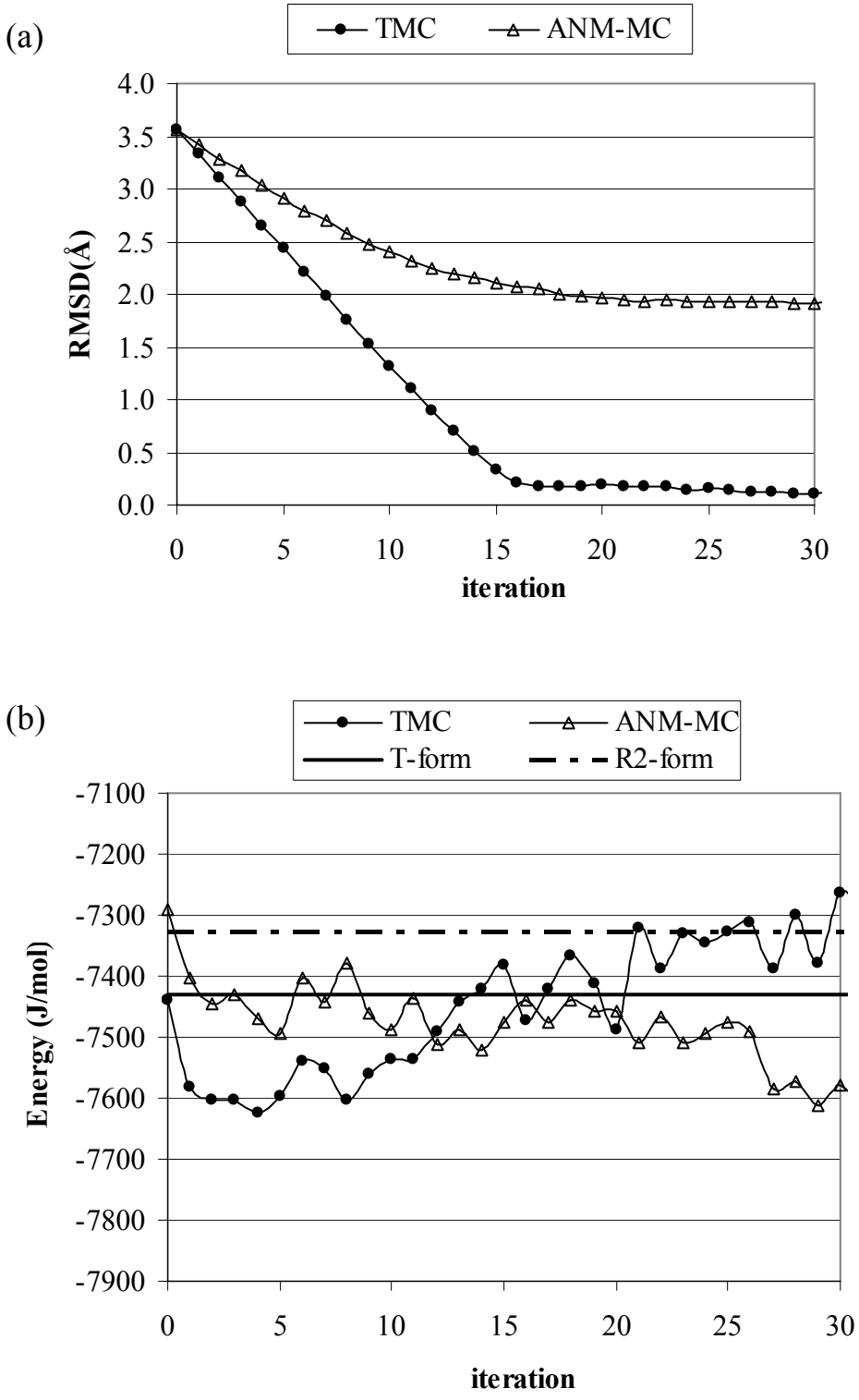


Figure 5.4. TMC simulation results as compared with ANM-MC method with application to hemoglobin for T-to-R2 transition (DF=0.2 Å; MCS=1000 for both cases). (a) RMSD between the intermediate snapshots and target structure. (b) Energy profiles of intermediates.

Table 5.3. RMSD values between TMC simulation snapshots and T, R2, and R form x-ray structures of hemoglobin (DF=0.2, MCS=1000)

snapshots	1A3N(T)	1HHO (R)	1BBB (R2)
3	0.93	1.65	2.87
5	1.16	1.49	2.43
7	1.85	1.38	1.98
10	2.3	1.31	1.32
13	2.76	**1.28**	0.69
15	2.91	1.32	0.32
20	3.21	1.57	0.18
25	3.47	1.78	0.15
30	3.52	1.8	**0.10**

5.2.2. Reverse Transition

As in case of AK, the reverse transition pathway study (R2-to-T) is also carried out for hemoglobin. Previous studies investigating hemoglobin transitional pathway reported that slowest modes successfully bring the T state close to the R and R2 states, however the backward passages did not prove to be accessible (Mouawad and Perahia, 1996; Xu *et al.,* 2003). In Figure 5.5, the RMSD and energy profiles of hemoglobin obtained for reverse transition simulations by ANM-MC and TMC methods are given. With TMC, the RMSD to the T state, which is the target in this case, has successfully decreased from 3.55 to around 0.1 Å (Figure 5.5(a)) and the energy profile (Figure 5.5 (b)) of the simulation trajectory is observed to be quite sensible as compared to the open and closed conformer average energies (indicated by horizontal lines).

In the case of ANM-MC, the target state could be approached an RMSD of 1.91 Å which not too far from the approach attained in the forward transition case (1.89 Å). The energy profile is observed to be similar to the TMC case and comparable with the initial and final state average energy values.

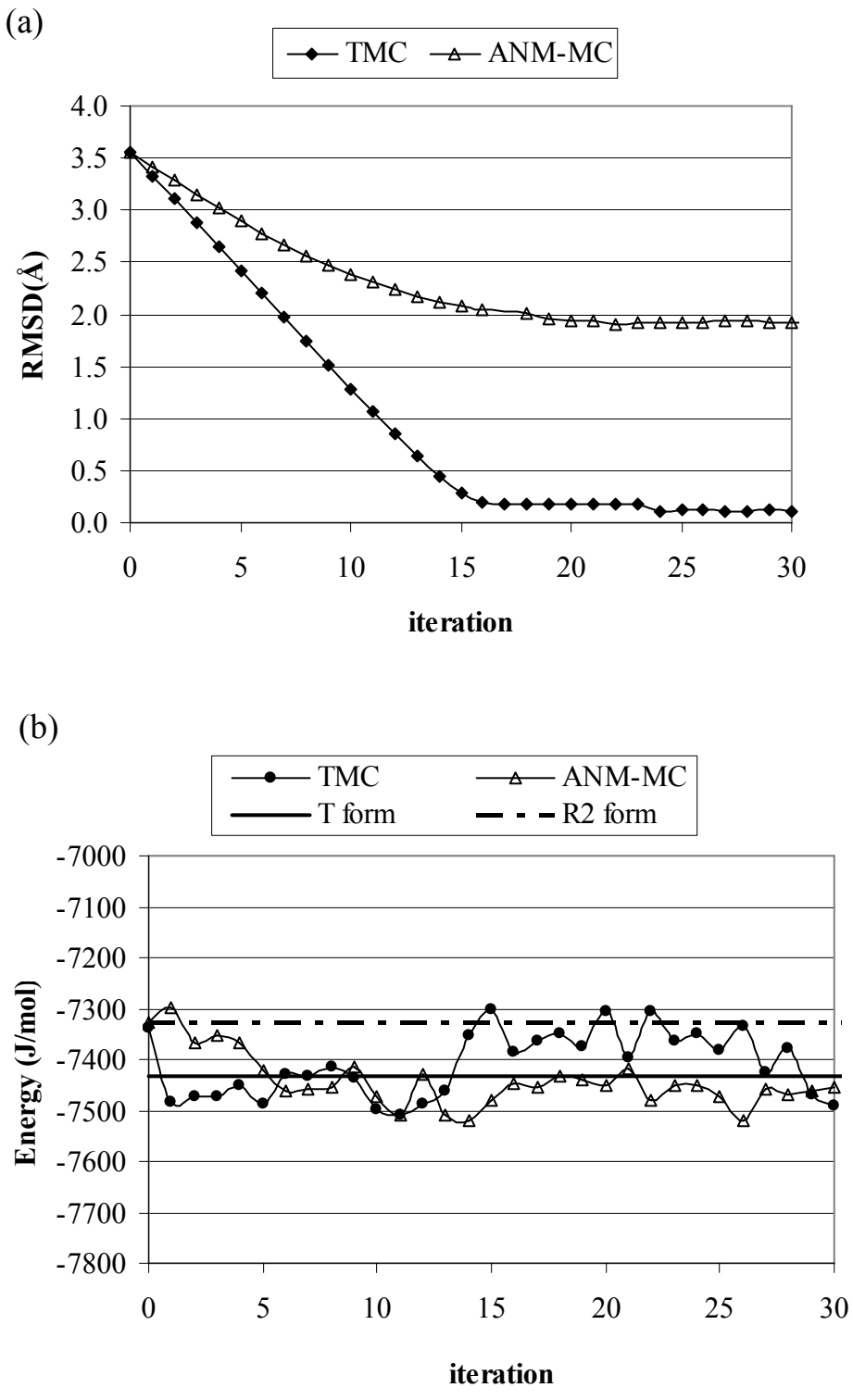


Figure 5.5. TMC simulation results as compared with ANM-MC method with application to hemoglobin for reverse transition (DF=0.2 Å; MCS=1000 for both cases). (a) RMSD between the intermediate snapshots and target structure. (b) Energy profiles of intermediates.

As a matter of fact, in the case of hemoglobin both TMC and ANM-MC methods can be successfully employed for accomplishing the reverse transition. The details regarding the comparison of the simulation trajectory with the initial (T state), final (R2 state) and the known intermediate (R state) are tabulated on Table 5.4. The simulation results revealed that the intermediate R state is approached by 1.43 Å at 10^{th} snapshot and beyond 30^{th} snapshot, RMSD from the target is observed to be around 0.1 Å.

Table 5.4. RMSD values between reverse TMC simulation snapshots and T, R2, and R form x-ray structures of hemoglobin

snapshots	**1BBB(R2)**	**1HHO (R)**	**1A3N (T)**
3	0.70	1.77	2.86
5	1.16	1.59	2.42
7	1.62	1.47	1.96
10	2.31	**1.43**	1.28
13	2.99	1.92	0.64
15	3.44	2.30	0.28
17	3.48	2.33	0.17
20	3.54	2.39	0.18
25	3.58	2.44	0.12
30	3.60	2.46	**0.10**

To summarize, both the ANM-MC and the TMC methods proposed in this work succeed in exhibiting transitions from open-to-closed forms of AK and free-to-bound states of hemoglobin, with reasonable pathway intermediates. TMC method is observed to reach the target conformation faster and with a more precise approach. In both methods, simulation snapshots were compared with the known x-ray structures that may be candidate pathway intermediates and the RMSD values indicated satisfactory approaches to the target conformations. In the reverse transition simulations, ANM-MC could not maintain very close mapping to the target state in the case of AK. This may be attributed to a high energy barrier that could not be overcome in the reverse transition pathway, hence preventing the molecule to open completely. Moreover, literature findings report that

generally reverse transitions might not be accompanied by the lowest frequency normal modes (Tama and Sanjound, 2001). However, TMC succeeded to attain almost complete mapping towards the target state with ~ 0 Å RMSD. In the case of hemoglobin, both ANM-MC and TMC enabled satisfactory mapping to the target conformation.

It is important to notice that the intermediate structures obtained in the case of forward and reverse transition of AK by TMC possess similarity with the x-ray structures proposed in literature. Nevertheless, the simulation snapshots reveal that both in the forward and reverse transitions the domain closure and/or opening is not in a sequential manner, i.e. the two domains close or open together in TMC unlike in ANM-MC, which clearly reflects the sequential order of domain closure.

Especially, for large systems TMC method could be efficiently used. However, in the case of unavailable target information, both ANM-MC and TMC methods would fail to predict and investigate the transitional pathway. At this point, RG-ANM-MC method, which will be presented in the following chapter, can successfully be used to simulate systems without target information by incorporating collective normal modes and tracking the radius of gyration.

6. RG-ANM-MC SIMULATION RESULTS

The ANM-MC methodology has been successful for analyzing the conformational transitions of AK and hemoglobin, for which the target conformations, i.e. closed state x-ray structures are available. In this chapter, a more generalized version of ANM-MC, the RG-ANM-MC methodology is presented. The origination of this method stems from the need for prediction of plausible closed states for proteins. Various alternative methods has been studied before finalizing RG-ANM-MC technique. Initially, only the conformations deformed along modes which result in lower energies were selected in search from the open to a plausible closed state. However, this protocol was unable to guide the structure towards its closed state, instead enabled a conformational search around the initial state. Furthermore, it was also not possible to reach to the closed state by selecting the conformations with decreasing RG's only. As a matter of fact, a combined approach utilizing both the energies and introducing RG's of the conformations as a constraint accomplished the desired transition.

The RG-ANM-MC methodology is applied to adenylate kinase (AK), human serum transferrin (HSTR) and Lysine/Arginine/Ornithine binding (LAO-binding) proteins that undergo large-amplitude conformational transitions. The RMSD of conformational changes are as follows: AK (7.13 Å); HSTR (6.70 Å); and LAO-binding protein (4.70 Å) with corresponding system sizes, i.e. residue numbers: 214, 238, and 328, respectively. Even though, for these proteins the x-ray structures of closed conformers exist, the algorithm does not use any structural information about target during simulation except that the radius of gyration of the simulation intermediates should decrease in order for the protein to close (see methods for details). Though not needed in the simulations, the closed state structures can still be utilized to test the success of the methodology. Hemoglobin was not taken as a test case protein here since distinct and large-amplitude conformational changes, such as open-to-close transitions have been rather preferred.

The adjustment of parameters, namely the deformation factor in ANM calculations and the Monte Carlo step used in ANM-MC simulations, was discussed in Section 4.

Among various combinations of DF and MCS, DF=0.2 Å gave satisfactory results within short CPU times for various protein systems. Hence, this value of DF=0.2 Å will be used here and MCS will be set by testing several alternative values (MCS=100, 500, 1000).

6.1. Case study: Adenylate kinase (AK)

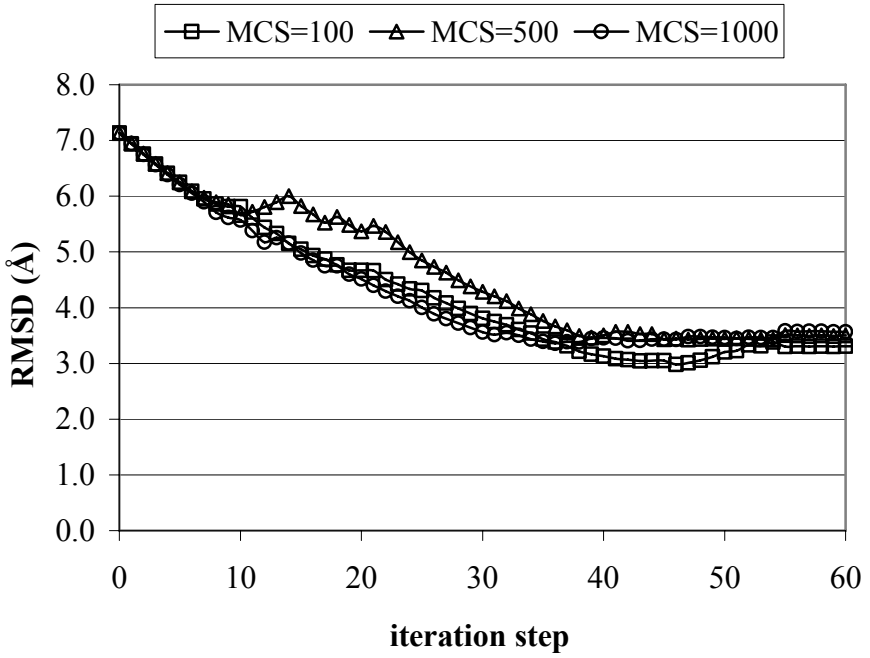
For the evaluation of ANM-MC technique, AK was chosen primarily due to its large conformational change (RMSD between open and closed structures= 7.13 Å) and the availability of large amount of experimental and computational data. In this section, the RG-ANM-MC simulation results of AK are presented in comparison with previous ANM-MC results. As before, the simulations are performed for the first chain disregarding the water and the ligand molecules.

6.1.1. RMSD, Energy and RG Profiles of AK

Figure 6.1 represents the simulation results obtained for the transition from the open to the closed conformation of AK for several MCS values with DF=0.2 Å. In Figure 6.1 (a) the RMSD between the simulation snapshots and the closed (target) conformation calculated at each iteration is demonstrated. Although the x-ray structure of the closed state as a target is not used in simulations, this figure illustrates the approach attained to the target. For the RG-ANM-MC simulation of AK, it can be observed that there is no significant difference between the RMSD profiles obtained for different MCS. But to decide on an appropriate value, the energy profiles should also be checked.

The corresponding energy profiles of the intermediate structures obtained during the simulations are given in Figure 6.1 (b). The initial and the target, i.e. the open and closed conformations are also relaxed by MC and average values from beginning to MCS=1000 are plotted as horizontal lines on the figure. As can be deduced from the RMSD and energy profiles, MCS=500 seems to be an appropriate choice with maximum approach to target state being 3.18 Å (Figure 6.1 (a)) and the potential energy being very close to that of the target conformation.

(a)



(b)

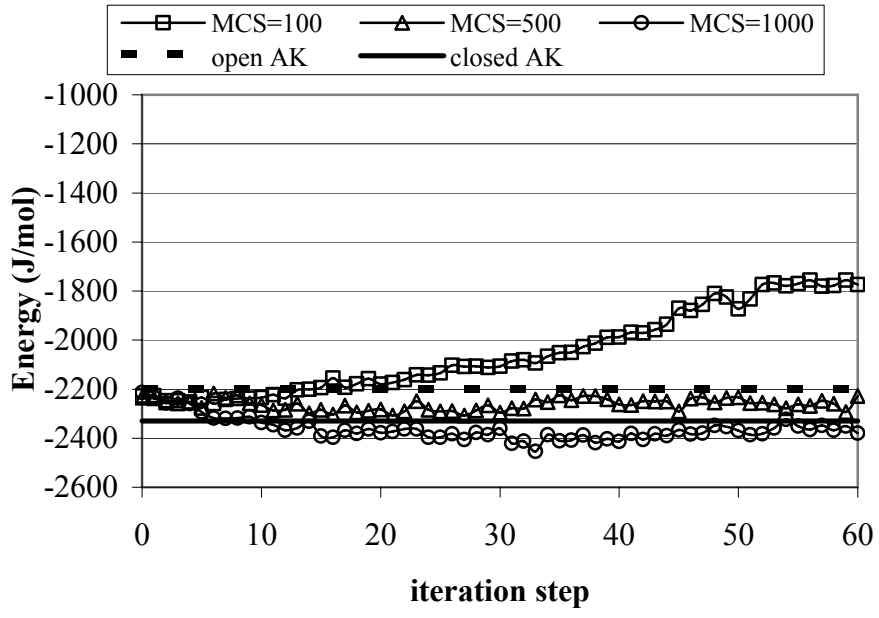


Figure 6.1. (a) Effect of Monte Carlo step (MCS) on the RMSD values between the intermediate structures and target (closed) structure as a function of iteration/cycle number for DF=0.2 (b) Corresponding energy profiles of AK.

In fact, a closer mapping to the target state with a lower RMSD profile is possible with MCS=100 (maximum approach=2.98 Å), however as far as the energy profile is concerned, MCS=500 seems to be a better choice. Therefore, the rest of the calculations on AK are based on the results of simulation with parameter values: DF=0.2 Å, and MCS=500.

Figure 6.2 illustrates further details of RG-ANM-MC simulation. Figure 6.2 (a) compares the RMSD profile of ANM-MC methodology, that targets the closed conformation using collective modes; with that obtained by RG-ANM-MC (for both methods DF=0.2 and MCS=500). The RMSD profiles exhibit similar trends with maximum approach values to the target state being 2.38 Å (at 48^{th} step) and 3.18 Å (at 44^{th} step) in ANM-MC and RG-ANM-MC methods, respectively. Though closer mapping to target is achieved by the previous ANM-MC methodology, RG-ANM-MC simulation without any target structure information proves successful with a 3.18 Å RMSD.

Previous studies reported that the conformational transitions are usually driven by the lowest frequency collective modes (Tama and Sanejouand, 2001; Krebs *et al.,* 2002; Xu *et al.,* 2003; Ma, 2005; Patrone and Pande, 2006). The results obtained by both ANM-MC and RG-ANM-MC simulations agree with literature. The specific collective mode chosen as the deformation direction with the corresponding radius of gyration value at each iteration are provided in Figure 6.2 (b). Initially, the algorithm chooses the 1^{st} and 2^{nd} slowest modes, which are associated with the LID closing. Up to 28^{th} iteration, the conformational change is driven with the two lowest frequency modes in conformity with previous studies. As a result, the RMSD to target structure decreases from 7.13 Å to 4.14 Å. In the following iterations, higher modes associated with AMP-bind domain motion enable a closer mapping to the target by decreasing the RMSD from 4.14 to 3.18 Å.

On the same figure (6.2 (b)), the RG values of the intermediates are also plotted. As a constraint of the RG-ANM-MC algorithm, the intermediate conformations are selected according to decreasing RG values, which is reflected by the RG profile in the figure. The RG values of the open and closed AK x-ray conformations are 376.36 Å and 268.06 Å, respectively. The RG of the final intermediate (60^{th} snapshot) is attained as 299.30 Å.

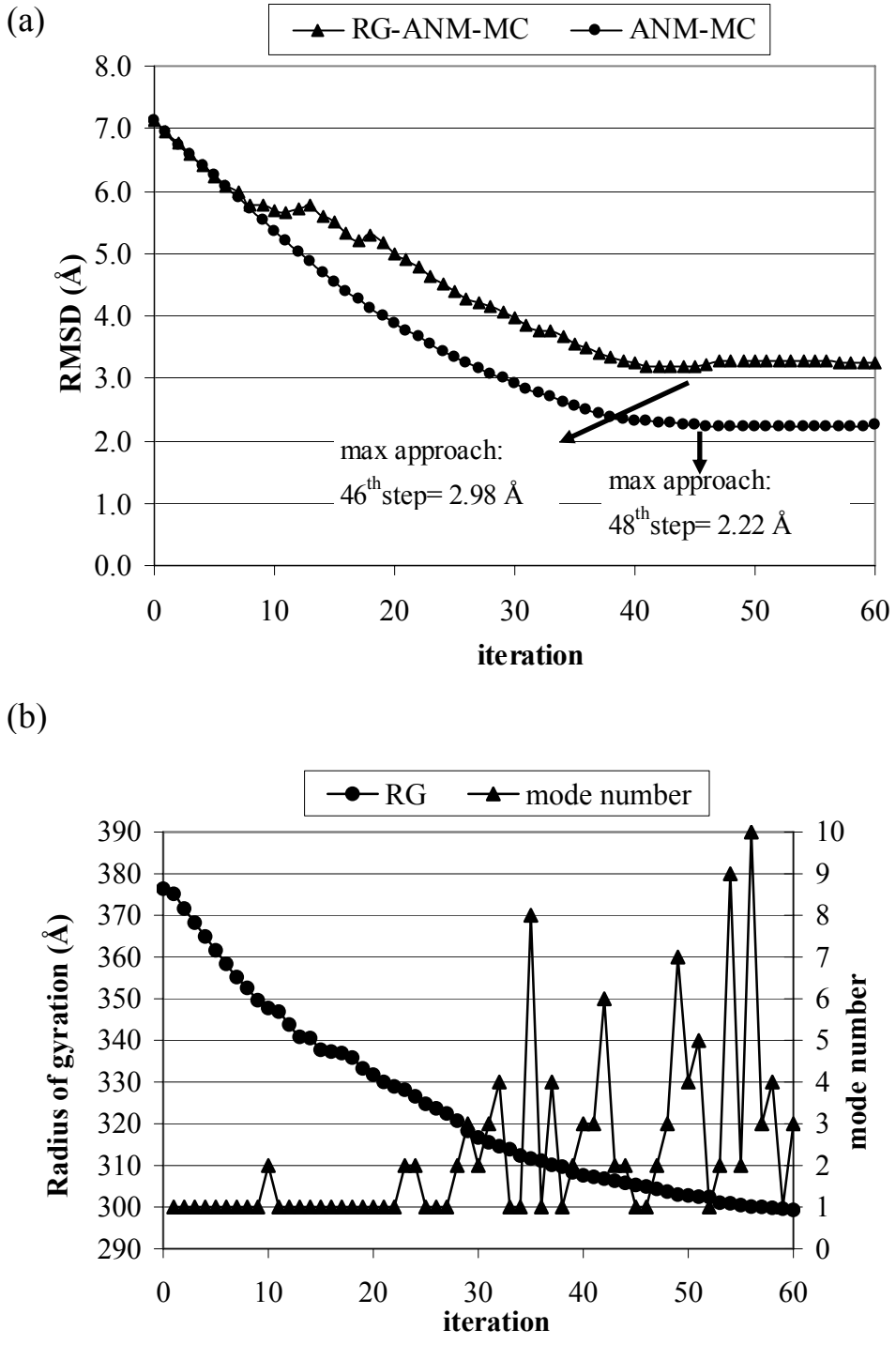


Figure 6.2. Comparison of the RMSD profiles of ANM-MC and RG-ANM-MC simulations of AK (DF=0.2; MCS=500) (a) and specific modes chosen and the corresponding RG values at each iteration (b).

In the previous ANM-MC method, the simulation was continued until a desired approach to the target state was maintained with a steady profile. In the case of RG-ANM-MC methodology, the duration of the simulation should be based on another factor since the target structure is unavailable. Hence, considering average RMSD of successive intermediates (see section 3) would be an option.

In Figures 6.3 (a) and (b), the simulation period required to attain maximum approach with a steady RMSD profile is discussed. In panel a, average RMSD of 10 successive intermediates ($RMSD10_{avg}$) is plotted. It can be inferred from the figure that for attaining $RMSD10_{avg} \leq 0.1$ the program should be continued more than 100 steps (until 80^{th} step shown) and for attaining $RMSD10_{avg} \leq 0.2$, the program should be continued up to about 60^{th} step. However, as far as the approach to the target structure is concerned (Figure 6.3 (b)), maximum approach has been already attained at 55^{th} step. Hence $RMSD10_{avg} \leq 0.2$ seems to be a sufficient constraint for ending the algorithm within a shorter computational time.

6.1.2. Transition Pathway and Pathway Intermediates of AK

Figure 6.4 illustrates several snapshots (up to 60^{th} iteration) selected from the RG-ANM-MC simulation trajectory. Earlier studies report that the LID domain (red) is much more mobile than AMP-bind domain (orange); and that the closure of the LID precedes the bending motion in the AMP-bind domain. (Maragakis and Karplus, 2005; Lou and Cukier, 2006; Withford *et al.,* 2007; Kantarci-Carsibasi *et al.*, 2008). This outcome is also supported by inspection of the simulation snapshots and their contact map representations presented in Section 4. Similar observation is valid in the case of RG-ANM-MC simulation results in Figure 6.4. In the first half of the simulation, the motion of the LID domain is more significant and motion of the AMP-bind region seems to be observable after $\sim 40^{th}$ iteration.

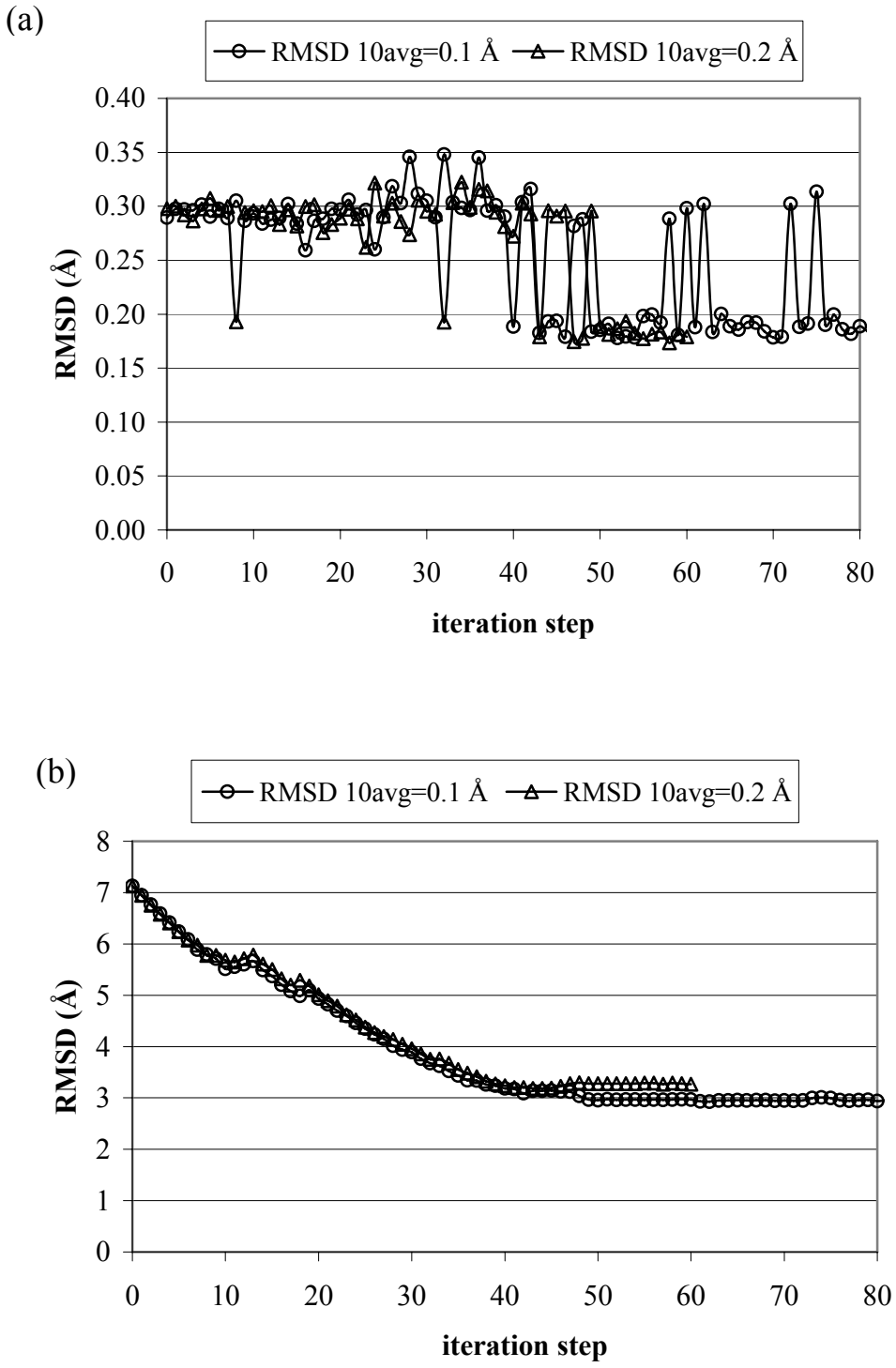


Figure 6.3. RMSD between the successive intermediate structures of RG-ANM-MC simulation of AK for attaining $RMSD10_{avg} \leq 0.1$ and 0.2 Å (a) and RMSD of simulation intermediates from the closed conformation of AK for $RMSD10_{avg} \leq 0.1$ and 0.2 Å (b).

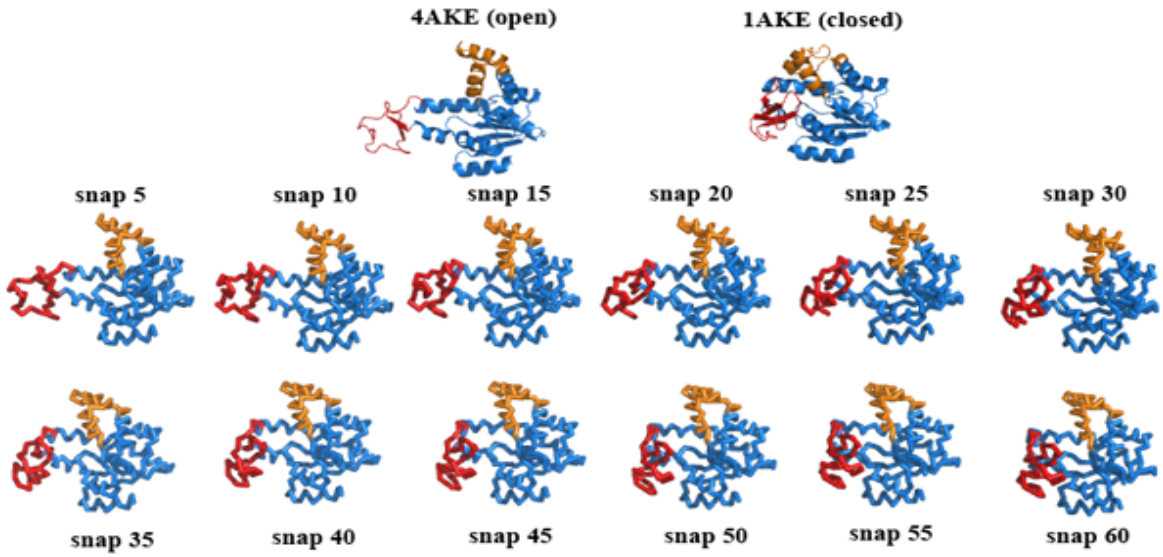


Figure 6.4. Several intermediate structures obtained during simulation of AK in open to closed transition obtained with RG-ANM-MC simulation (DF=0.2; MCS=500)

The RMSD between the RG-ANM-MC simulation snapshots presented in Figure 6.4, with the proposed AK intermediates is tabulated in Table 6.1. RMSD values indicated in bold face represent maximum approaches attained among the selected snapshots to the specific x-ray structures. The maximum approaches are observed to be between 2.7- 3.5 Å. With the previous ANM-MC methodology, the simulation snapshots were maintained between 2.3- 3.0 Å RMSD with these x-ray structures.

Earlier snapshots (iterations up to 40) exhibit maximum approach to the crystal structures: 1AK2, 1DVR (B chain) and 1DVR (A chain), sequentially. In 1AK2, the LID domain is about to close over the core, while the AMP-bind region is completely open. In 1DVR (A and B chains), the LID region is totally closed, while the AMP-bind is still open. The overlap of these three x-ray structures with earlier snapshots indicates the priority of LID closing. Subsequent snapshots (iterations > 40) fall in proximity with the rest of four crystal structures, namely 1E4Y(A chain), 1E4V (A chain), 1ANK (A chain), and 2ECK (A chain). These crystal structures bear close resemblance to the closed conformer 1AKE, i.e. both the LID and AMP-bind domains are almost closed. The similarity of final

snapshots with these structures shows that the AMP-bind closing appears to take place more significantly in the late stages of the simulation.

Table 6.1. RMSD values between RG-ANM-MC simulation snapshots and several AK crystal structures (DF = 0.2 Å and MCS = 500)

Snapshot/ iteration	**4AKE**	**1AK2[a]**	**1DVR[b]** ***B chain***	**1DVR[b]** ***A chain***	**1E4Y[c]** ***A chain***	**1E4V[d]** ***A chain***	**1ANK[e]** ***A chain***	**2ECK[f]** ***A chain***	**1AKE**
5	2.35	4.47	4.22	4.10	5.93	6.23	6.25	6.31	6.25
10	3.13	4.10	4.13	4.02	5.50	5.80	5.84	5.87	5.81
15	3.59	4.08	3.52	3.38	4.76	5.03	5.06	5.10	5.66
20	3.80	3.73	3.15	2.94	4.40	4.66	4.64	4.71	5.36
25	4.13	3.65	2.95	2.84	4.06	4.30	4.28	4.35	4.84
30	4.73	**3.54**	2.83	2.70	3.63	4.10	3.78	3.84	4.28
35	5.28	3.75	2.81	2.68	3.52	3.63	3.53	3.66	3.76
40	5.74	3.80	**2.71**	**2.66**	3.50	3.53	3.48	3.54	3.50
45	6.44	4.03	2.83	2.76	**3.46**	**3.45**	**3.40**	**3.45**	**3.43**
50	7.17	4.64	3.44	3.40	3.53	3.49	3.56	3.59	3.48
55	7.54	5.16	3.91	3.88	3.55	3.50	3.58	3.58	3.50
60	7.54	5.17	3.91	3.89	3.55	3.53	3.58	3.59	3.51
4AKE	0.00	5.38	5.78	5.63	7.20	7.64	7.66	7.23	7.13
1AKE	7.13	5.57	4.02	3.92	0.93	0.65	0.45	0.28	0.00

[a] 1AK2 (Schlauderer and Schulz, 1996) belong to bovine mitochondria inter-membrane space AK.

[b] 1DVR (A and B chains) (Schlauderer *et al*.,1996) belong to the baker's yeast AK.

[c] 1E4Y (Müller and Schulz, 1993) is the Glysine-loop modified version of AK from *E.coli*.

[d] 1E4V (Müller and Schulz, 1993) is G10V mutant of AK from *E.coli*.

[e] 1ANK (Berry *et al*., 1994) is AMPPNP and AMP bound form of AK from *E.coli*.

[f] 2ECK (Berry *et al*., 2006) is AMP and ADP bound form of AK from *E.coli*.

RG-ANM-MC simulations are also performed for accomplishing reverse transition, beginning with the closed form of AK. Same procedure is valid, except that RG should increase during transition from the closed-to-open state. However, the desired transition

could not be managed by RG-ANM-MC method as was also not possible with ANM-MC method. In the study of Tama and Sanejouand (2001), the authors have also pointed out that the reverse transitions from closed-to-open states do not usually correspond to the slow modes obtained from normal mode analysis.

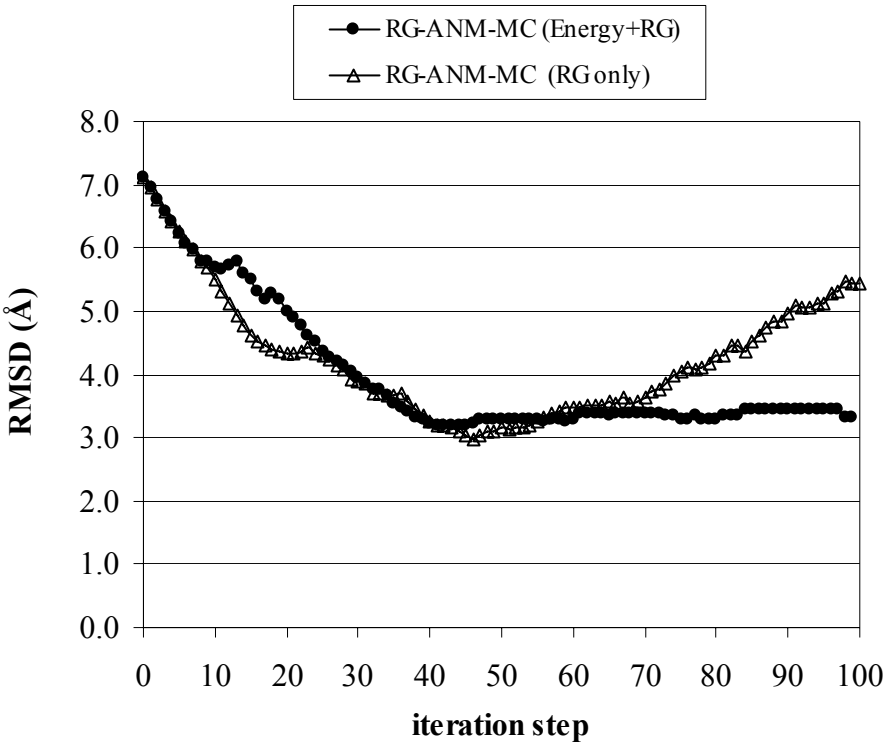
6.1.3. Significance of energy consideration of collective modes for the conformational transition

While developing the RG-ANM-MC method, several alternative procedures were tested. Initially, the transition was tried to be accomplished by using only normal modes and corresponding conformation energies. However, this protocol could not prove successful in approaching the target. Instead, the conformational space could be scanned around the initial structure. Hence the requirement for a constraint condition came into picture. RG is then used as a constraint that would lead the protein to closure.

It is also studied whether the transition could be accomplished by disregarding the conformation energies and choosing only the conformations having lower RG's. The results are presented in Figure 6.5. In panel a, the RMSD profiles obtained by both approaches are compared. In the case of disregarding the energies, the RMSD decreases approximately to step 45^{th}, attaining a maximum approach to target with 3.03 Å RMSD; and then starts to deviate with increasing RMSD. In the case of energy and RG consideration, the RMSD trend attains a stable profile around the maximum approach value. Moreover, it is indicated in Figure 6.5 (b), that the energy profile is significantly far from target state when only RG is considered.

When the simulation snapshots obtained in both cases are compared the case where RG criterion is used only, can not capture the x-ray structures such as 1AK2, 1DVRB, and 1DVRA. As a matter of fact, using a combined scheme proved more successful in obtaining a stable RMSD and energy profile. Furthermore, the combined scheme revealed the sequence of steps more clearly during the transition.

(a)



(b)

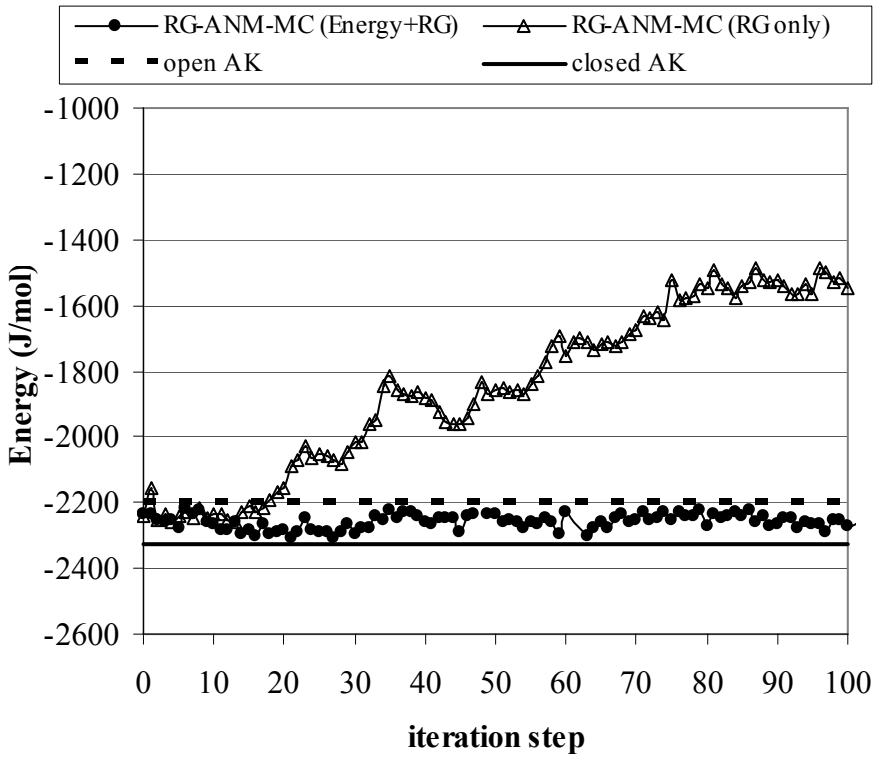


Figure 6.5. Comparison of RG-ANM-MC simulations with/without consideration of normal mode energies (a) RMSD profiles (b) Energy profiles

6.2. Case study: Human Serum Transferrin (HSTR)

In this section, the RG-ANM-MC simulation results will be presented for the transition from open to closed conformations of HSTR. N-lobe half-molecules (328 residue) with apo and holo form pdb codes: 1BP5-Achain and 1A8E are used and ligand and water molecules are disregarded in the simulation.

6.2.1. RMSD, Energy and RG Profiles in HSTR

First the parameters are adjusted as in the previous cases. By fixing DF=0.2 Å, a suitable MCS is determined. Figure 6.6 (a) and (b) represent the RG-ANM-MC simulation results obtained for HSTR by using MCS= 100, 500, and 1000. In Figure 6.6 (a), the RMSD between the simulation snapshots and the closed (target) conformation calculated at each iteration is demonstrated. The corresponding energy profiles obtained by this combination of parameters are depicted in Figure 6.6 (b). The initial and target structures are also relaxed by 1000 MCS and average energies are indicated by horizontal lines on the figure.

The simulation intermediates are desired to get as close as possible to the target state with a stable RMSD trend and comparable energy profiles with that of the target. As Figure 6.6 (a) implies, unlike the AK case, the MCS seems to alter the RMSD profiles considerably. For MCS = 100, 500, and 1000, the closest approach values to the target conformation are obtained as 2.71, 3.45, and 4.60 Å RMSD, respectively. Adjustment of the two parameters suggests MCS=500 is an appropriate choice. Hence, results will be presented for DF=0.2, and MCS=500.

Figure 6.7 illustrates further details of the RG-ANM-MC simulation of HSTR. The comparison of the RMSD profile obtained by RG-ANM-MC and ANM-MC simulations are demonstrated in Figure 6.7 (a) (for both methods DF=0.2, and MCS=500). The maximum approach values attained to the target state are 1.81 Å at 50^{th} iteration step and 3.45 Å at 55^{th} iteration steps by ANM-MC and RG-ANM-MC methodologies, respectively.

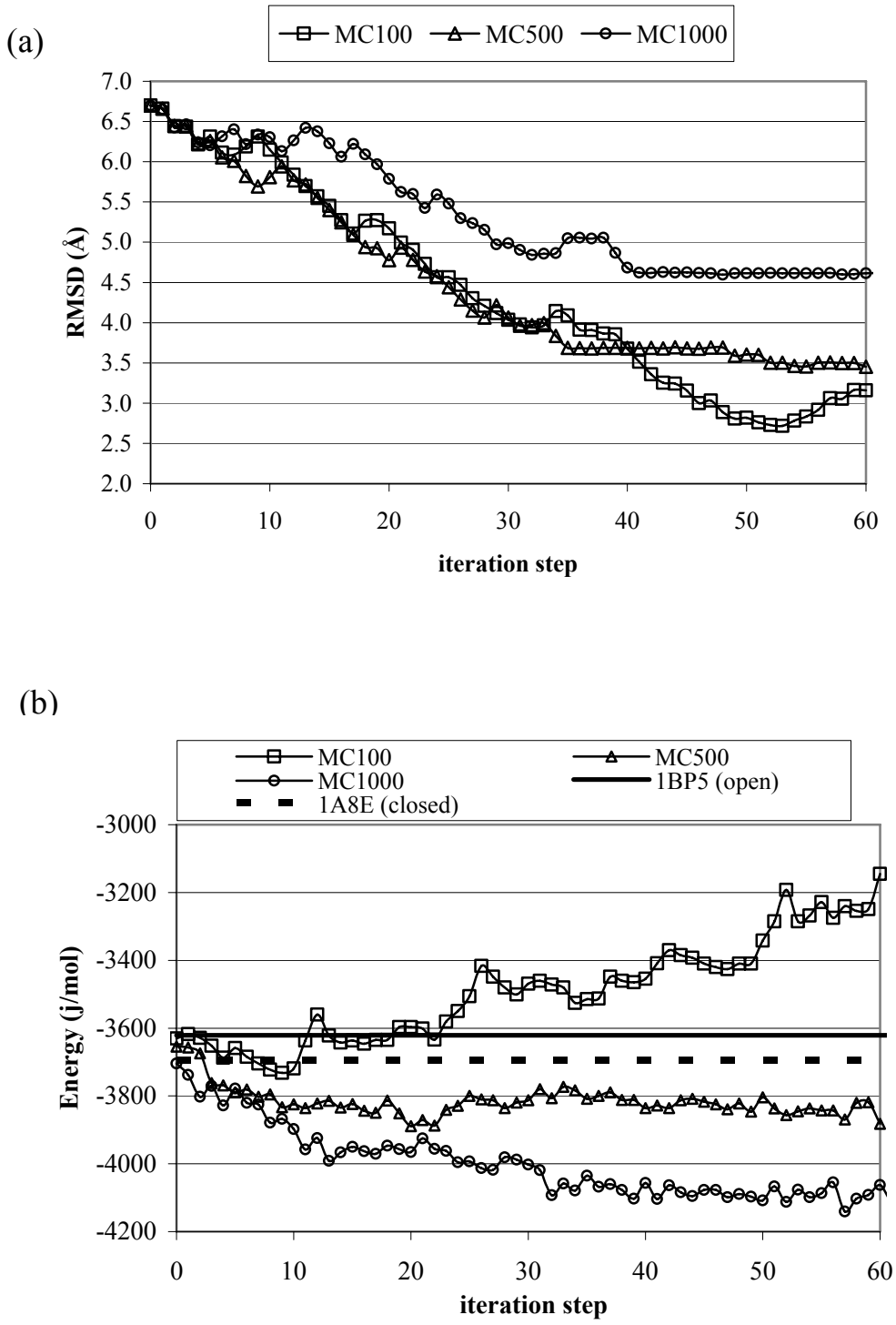


Figure 6.6. RMSD (a) and energy (b) profiles of intermediate structures of HSTR obtained by RG-ANM-MC with different MCS values (DF=0.2 Å)

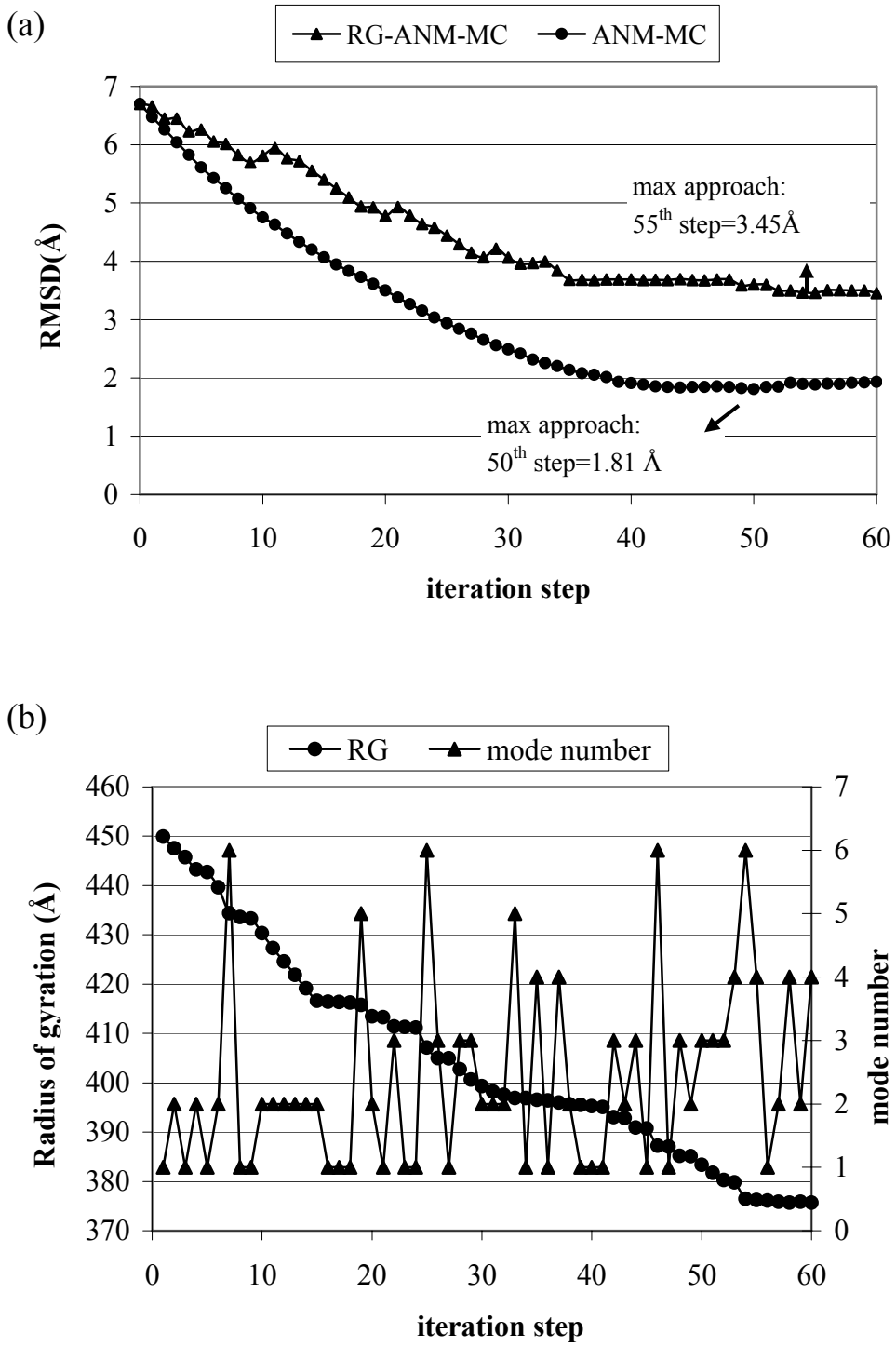


Figure 6.7. Comparison of the RMSD profiles of ANM-MC and RG-ANM-MC simulations of HSTR (DF=0.2; MCS=500) (a) and modes with corresponding RG's chosen at each iteration (b).

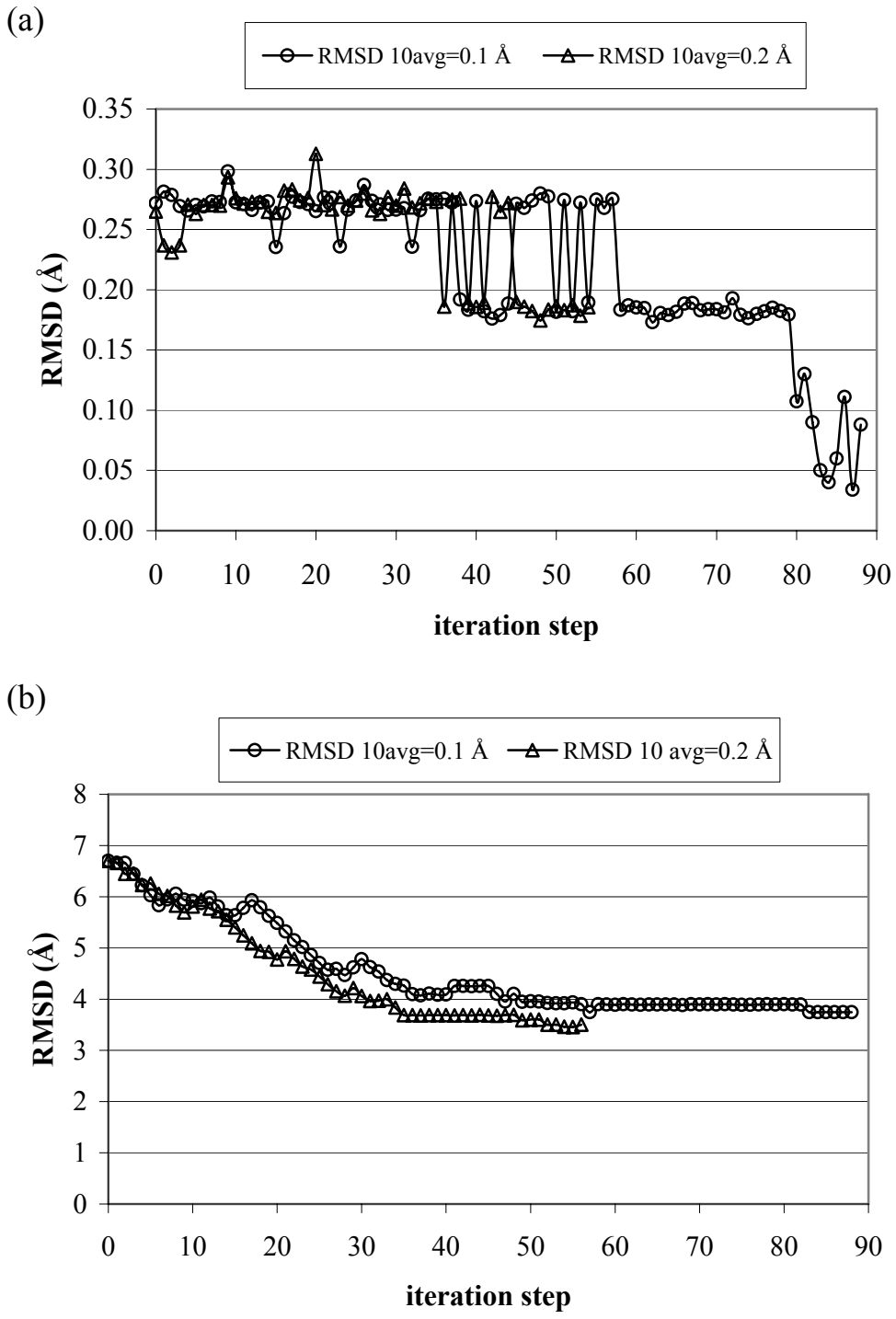


Figure 6.8. RMSD between the successive intermediate structures of RG-ANM-MC simulation of HSTR for attaining $RMSD10_{avg} \leq 0.1$ and 0.2 Å (a) and RMSD of simulation intermediates from closed conformation of HSTR for $RMSD10_{avg} \leq 0.1$ and 0.2 Å (b).

ANM-MC methodology provides a closer mapping to target state, however, without any target information RG-ANM-MC technique yields close proximity with the ANM-MC profile. In Figure 6.7 (b), the variations of RG and the mode direction throughout the simulation are presented. The direction of transition is guided by the lowest frequency normal modes initially which constitute the minimum energy direction, and then higher modes accompany as in the previous cases. The RG values of the open and closed conformations are 449.88 and 371.80, respectively. The RG of the final intermediate (60^{th} snapshot) is attained as 375.68 which is quite close to that of the target structure (Figure 6.7 (b)).

Similarly, in Figure 6.8 (a) and (b) the decision of the period of simulation is illustrated based on $RMSD10_{avg}$ values. As in the case of AK, $RMSD10_{avg}$= 0.2 is a sufficient choice for capturing the maximum approach to the target within reasonable computational time.

6.2.2. Transition Pathway and Pathway Intermediates of HSTR

In Figure 6.9, several intermediate structures (up to 60^{th} iteration) from the RG-ANM-MC simulation of HSTR are presented (DF= 0.2 Å; MCS= 500). All snapshots are displayed by superimposing on the initial (open) conformation. Visual interpretation reveals the closure of the two domains (N1 and N2) about the hinge point.

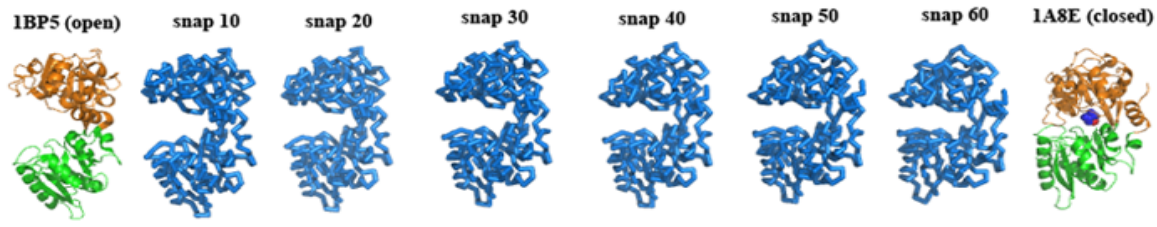


Figure 6.9. Several intermediate structures obtained during simulation of HSTR in transition from open to closed conformation obtained with RG-ANM-MC simulation (DF=0.2; MCS=500).

Based on the simulation snapshots a plausible transitional pathway is suggested by extensive x-ray structure search through Protein Data Bank. There is not yet a published transitional pathway suggested for HSTR. The database is scanned for protein structures in which the name "transferrin" appears. Among over 20 x-ray structures, several candidates are proposed to be present in the HSTR transitional pathway. Table 6.2 tabulates these x-ray structures and their RMSD from the simulation snapshots RMSD values indicated in bold face represent the maximum approach values attained among the selected snapshots to the specific x-ray structures.

Table 6.2. RMSD values between RG-ANM-MC simulation snapshots and several transferrin related crystal structures (DF = 0.2 Å and MCS = 500)

Snapshot/ iteration	**1BP5**	**1IQ7[a]** ***A chain***	**1FCK[b]** ***A chain***	**1BLF[c]** ***A chain***	**1TFD[d]** ***A chain***	**1A8E**
10	1.51	4.23	5.60	5.93	6.05	6.15
20	2.79	3.72	4.70	4.99	5.08	5.17
30	3.51	**3.59**	3.71	3.88	4.01	4.03
40	3.99	3.71	3.67	3.73	3.61	3.68
50	4.80	3.83	**3.49**	3.66	3.51	3.60
60	5.49	4.42	3.53	**3.55**	**3.40**	**3.45**
1BP5	0.00	4.58	6.16	6.52	6.63	6.70
1A8E	6.70	5.31	1.66	1.48	1.00	0.00

[a] Apo form of C-terminal ovotransferrin from chicken (Mizutani *et al.*, 2001)

[b] Diferric bovine lactoferrin (Moore *et al.,* 1997)

[c] Diceric human lactoferrin (Baker *et al.,* 2000)

[d] Rabbit serum transferrin (Sara *et al.,* 1990)

The maximum approach values are observed to be between 3.4- 3.6 Å. Thus, the transitional pathway intermediates proposed in this thesis includes apo form of C-terminal ovotransferrin from chicken, diferric bovine lactoferrin, diceric human lactoferrin, and

rabbit serum transferrin in a sequential manner. Especially it is suggested that apo form of C-terminal ovotransferrin (1IQ7) and diferric bovine lactoferrin (1FCK) are significant candidate intermediates that lie in the middle of the pathway. The other two x-ray structures are closer to the closed form; hence lie towards the end of the pathway.

The suggested candidate intermediates are validated by previous ANM-MC methodology. The RMSD values between the ANM-MC simulation intermediates and suggested transferrin related x-ray structures are tabulated on Table 6.3. Even closer approach values are obtained (RMSD between 1.95-2.85 Å) to the x-ray structures which confirm the RG-ANM-MC results.

Table 6.3. RMSD values between ANM-MC simulation snapshots and transferrin related crystal structures (DF = 0.2 Å and MCS = 500)

Snapshot/ iteration	**1BP5**	**1IQ7[a]** ***A chain***	**1FCK[c]** ***A chain***	**1BLF[b]** ***A chain***	**1TFD[d]** ***A chain***	**1A8E**
10	2.30	3.49	4.28	4.59	4.66	4.72
20	3.69	3.22	3.14	3.40	3.24	3.43
30	4.91	**2.85**	2.38	2.54	2.40	2.36
40	5.80	3.11	**2.25**	**2.32**	2.35	2.25
50	6.28	3.45	2.43	2.38	**1.95**	**1.81**
60	6.41	3.87	2.51	2.42	1.98	1.83
1BP5	0.00	4.58	6.16	6.52	6.63	6.70
1A8E	6.70	5.31	1.66	1.48	1.00	0.00

6.3. Case study: Lysine/Arginine/Ornithine Binding (LAO-binding) Protein

In this section, the RG-ANM-MC simulation results are presented for LAO-binding protein. This protein consists of 238 residues with apo and holo form PDB codes: 2LAO and 1LST, respectively. RMSD of the corresponding conformational change is 4.7 Å and the water and ligand molecules are disregarded in the simulation.

6.3.1. RMSD, Energy and RG Profiles in LAO-binding protein

The RMSD and energy profile results obtained in the simulation of LAO-binding protein are provided in Figure 6.10 (a) and (b), respectively for the same combination of simulation parameters. RMSD varies significantly with MCS, as in the case of HSTR. The maximum approaches to the target structure with MCS=100, 500 and 1000 are 2.10, 2.52, and 3.13 Å, respectively. Based on the energy plot, MCS=500 seems to be the best choice. Hence, as in the case of AK and HSTR for DF=0.2, MCS=500 seems to be suitable for the simulations.

Figure 6.11 demonstrates further details on the RG-ANM-MC simulation of LAO-protein. The comparative analysis of results obtained by the ANM-MC and RG-ANM-MC methods is given in Figure 6.11 (a). As in previous cases, similar RMSD profiles are observed in both cases with maximum approach values as 1.40 Å (27^{th} step) and 2.61 Å (26^{th} step) in the ANM-MC and RG-ANM-MC methods, respectively.

Figure 6.11 (b) reveals the mode directions selected with the corresponding RG values attained at each iteration. Up to 21^{st} step, first two slowest modes are selected as was observed in AK and HSTR. At the 21^{st} iteration the RMSD has already decreased from 4.7 to 2.7 Å; then higher modes come into for decreasing the RMSD from 2.7 to 2.6 Å. The RG values of the open and closed conformations are 362.78 and 313.29, respectively. The RG of the final intermediate is attained as 340.33 (Figure 6.11 (b)).

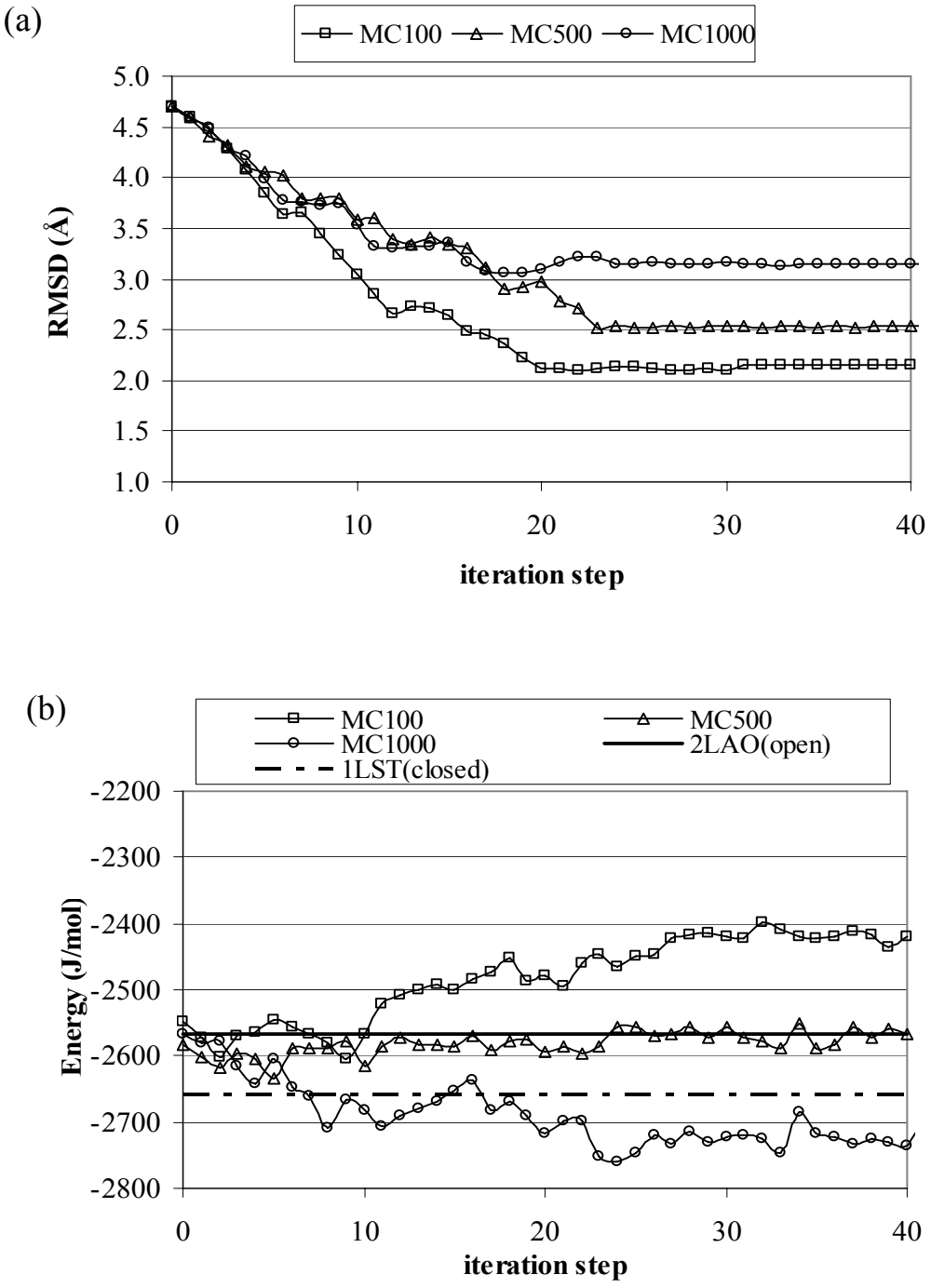


Figure 6.10. RMSD (a) and energy (b) profiles of intermediate structures of LAO-binding protein obtained by RG-ANM-MC with different MCS values (DF=0.2)

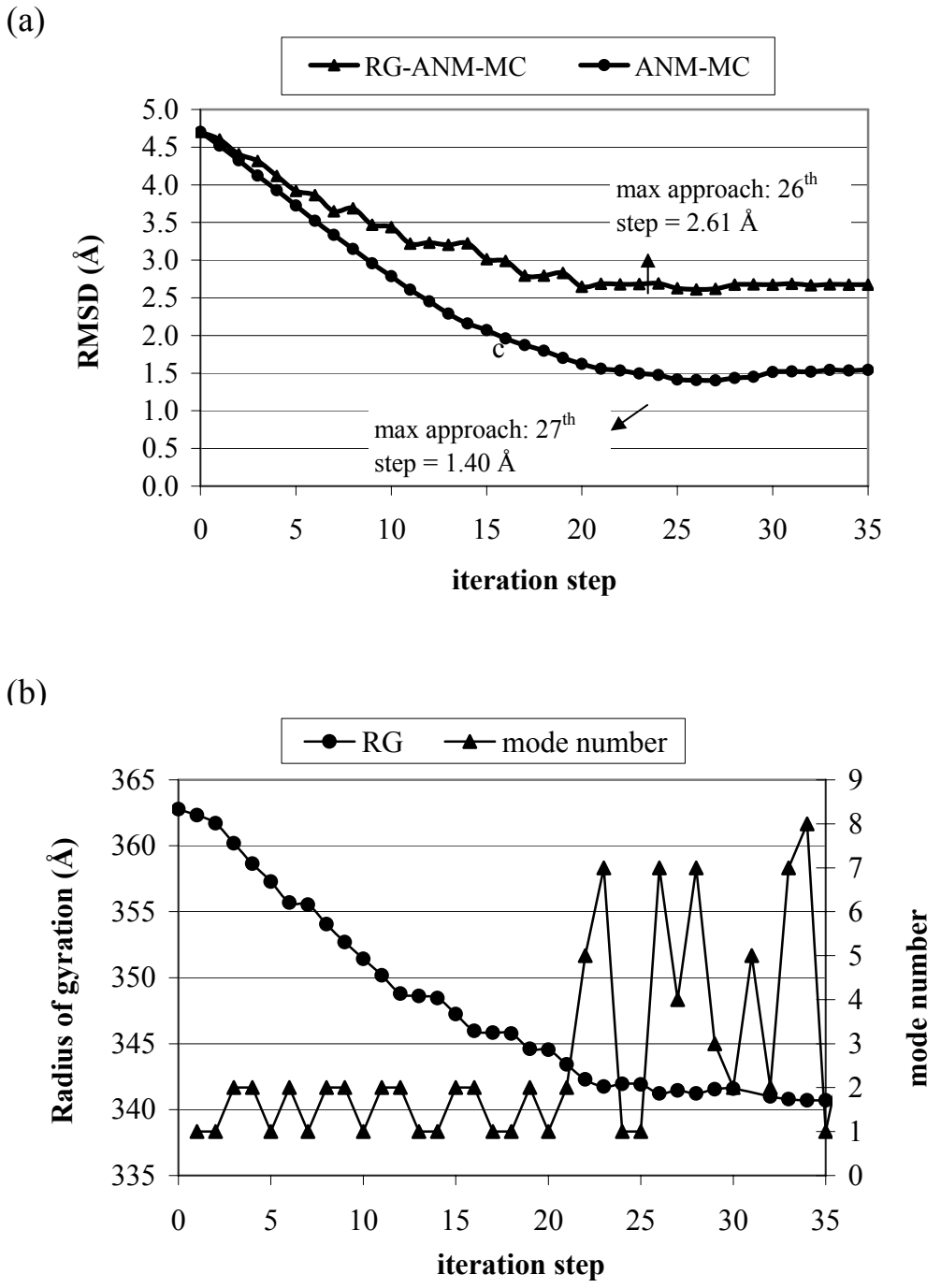


Figure 6.11. Comparison of the RMSD profiles of ANM-MC and RG-ANM-MC simulations of LAO-binding protein (DF=0.2; MCS=500) (a) and modes with corresponding RG's chosen at each iteration (b).

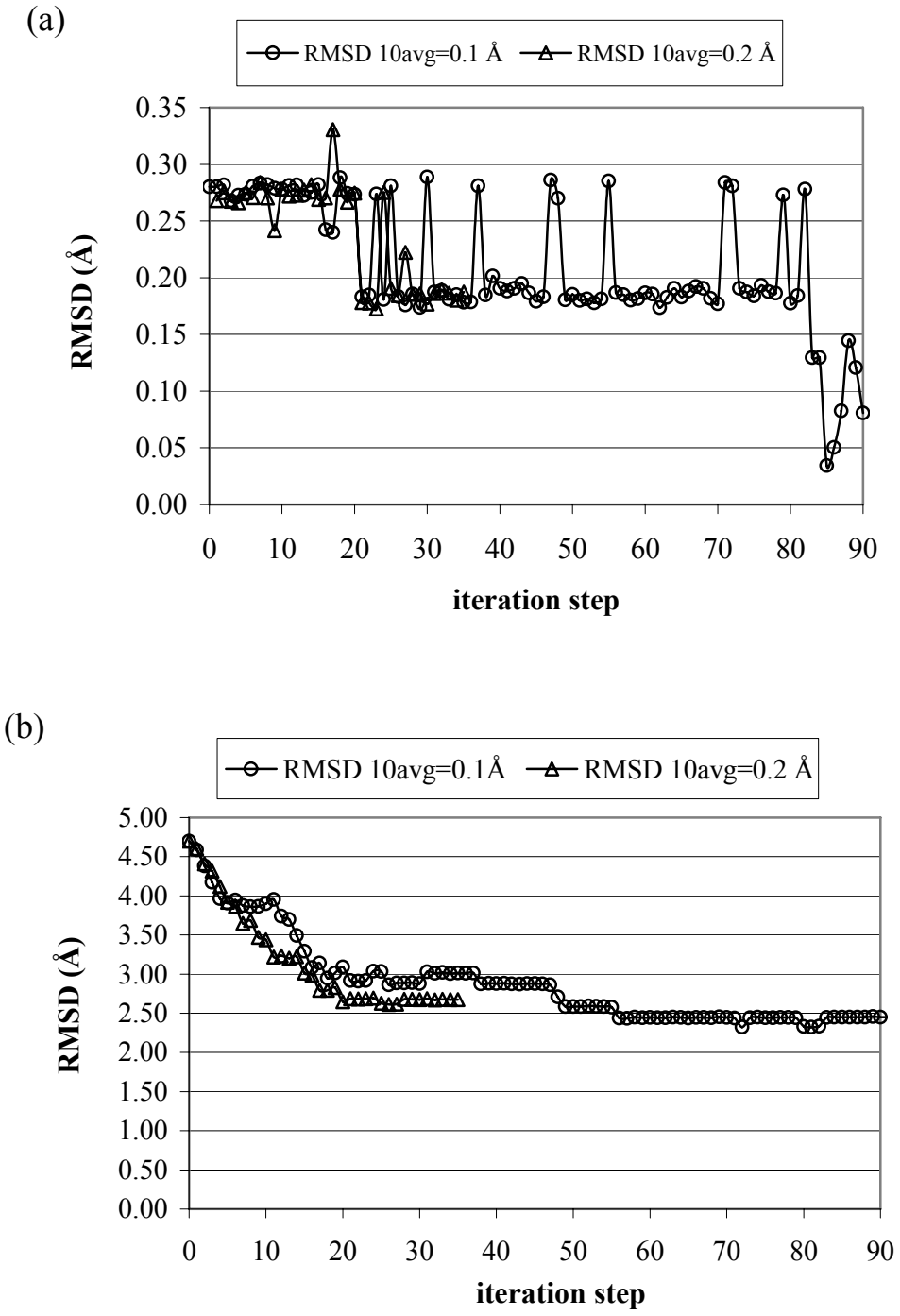


Figure 6.12. RMSD between the successive intermediate structures of RG-ANM-MC simulation of LAO-binding protein for attaining $RMSD10_{avg} \leq 0.1$ and 0.2 Å (a) and RMSD of simulation intermediates from closed conformation of LAO-binding protein for $RMSD10_{avg} \leq 0.1$ and 0.2 Å (b).

Analysis on the simulation duration for attaining $RMSD10_{avg}$ value of 0.1 and 0.2 is depicted in Figure 6.12 (a) and (b). Since similar to the previous test cases with $RMSD10_{avg}$ = 0.2, maximum approach is already attained at much earlier steps; hence $RMSD10_{avg}$ = 0.2 is once again regarded as the sufficient constraint for ending the algorithm.

6.3.2. Transition pathway and pathway intermediates of LAO-binding protein

Figure 6.13 presents the RG-ANM-MC simulation intermediates of LAO-binding protein. It is revealed that the two domains approach each other by bending about a hinge. There is still an RMSD of 2.6 Å between the final snapshot (Figure 6.10 (a), MCS=500) and the closed structure hence closure of the domains is not complete.

Recently, Keskin (2007) presented an ANM based study investigating the conformational changes that antibody proteins undergo. The author simulated 8 different proteins including LAO-binding protein and proposed a transition pathway by adding the eigenvector leading to the desired transition to the original coordinates using various rescale, i.e. deformation factors. It is observed that the two domains close as the ligand is bound to protein and the maximum approach value to the target is maintained as 2.13 Å.

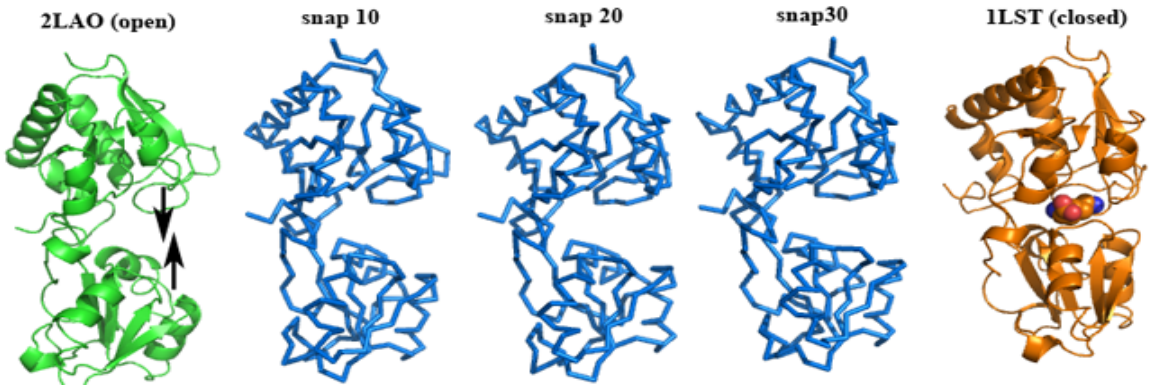


Figure 6.13. Several intermediate structures obtained during RG-ANM-MC simulation of LAO-binding protein in transition from open to closed conformation (DF=0.2; MCS=500).

7. CONCLUSIONS AND RECOMMENDATIONS

7.1. Conclusions

In this study, new approaches are developed to investigate conformational transition pathways of proteins which undergo large conformational changes upon ligand binding. In the first protocol, namely ANM-MC methodology, the elastic network model ANM is coupled with an MC algorithm that utilizes knowledge-based potentials of proteins. ANM-MC algorithm is shown to be useful for the investigation of protein conformational transition pathways. An important implication of this method is that it improves ANM as well as MC in the sense of providing information about the sequence of events as well as more efficient conformational search.

Investigation of parameters has indicated that small deformation factors such as 0.1, or 0.2 Å prove successful in maintaining satisfactory RMSD and energy profiles, thereby achieving a close approach to the target state with reasonable pathway intermediates. However, when faster sampling of the transitional pathways in larger systems like the hemoglobin in this thesis, one may choose larger DF's followed by shorter energy minimizations if the elucidation of intermediate conformations and the sequence of events is of main concern since effect of MCS on RMSD profiles and intermediate structures is not significant.

Application of the ANM-MC algorithm proves successful in achieving the transition between the open and closed conformations of AK (*E.coli*). Specific snapshots along the trajectory fall within acceptable RMSD ranges with the crystal structures proposed as intermediates on the AK pathway previously suggested in literature (Maragakis and Karplus, 2005). The closing of the LID domain precedes that of AMP bind domain, which is also validated with contact map representations. Similarly, the method accomplishes the transition from T to R2 states of hemoglobin, passing through the R state proposed to be an intermediate state between T and R2 (Srinivasan and Rose, 1994). Both applications on AK and hemoglobin reveal that the two lowest frequency modes are foremost in driving

the conformational changes. Nevertheless, higher modes are needed in later stages for a closer mapping towards the target structure.

In RG-ANM-MC method, which is an extension of ANM-MC, ANM and MC algorithms are used to simulate the conformational transitions of three proteins, namely AK, HSTR, and LAO-binding protein. As compared to the ANM-MC methodology, RG-ANM-MC possesses the very important advantage of being applicable to systems where only the initial conformation is available. In this method, the conformation having the minimum energy with a lower radius of gyration is selected as the intermediate in going from the open to closed conformation. Hence, previous methodology has somewhat been generalized for application to proteins exhibiting hinge-bending motions without the aid of target information.

Application of the RG-ANM-MC algorithm proves successful in achieving the transition between the open and closed conformations of AK, HSTR, and LAO-binding protein. Specific snapshots along the trajectory fall within acceptable RMSD ranges with the crystal structures proposed as intermediates on the AK pathway (Maragakis and Karplus, 2005). Moreover, in the case of AK, the closing of the LID domain precedes that of AMP bind domain. For the case of HSTR, several HSTR related x-ray structures that might be candidate pathway intermediates are proposed in this thesis. Especially, apo form of C-terminal ovotransferrin (1IQ7) and diferric bovine lactoferrin (1FCK) are suggested as significant candidate intermediates.

In all three test cases, the corresponding conformational changes, i.e. the transitional path is guided by the lowest frequency normal modes. Nevertheless, higher modes are needed in later stages for a closer mapping towards the target structure. This finding is in agreement with previous studies (Brooks and Karplus, 1985; Tama and Sanejouand, 2001; Zheng and Doniach, 2003; Emekli *et al*., 2007) which reported that usually the conformational transition is guided by the slowest mode, or sometimes by the two lowest frequency normal modes. It should also be noticed that, both techniques were successful for proteins exhibiting hinge-bending motion which is actually a common

property for open-to-closed transitions. Further work is necessary for the algorithm to cover a wider range of protein systems.

Both ANM-MC and RG-ANM-MC results are in conformity with literature studies reporting that the conformational change arising due to ligand binding actually is an intrinsic property of the binding-protein (Tsai *et al.,* 1999; Ma *et al.,* 2002; Lei *et al.,* 2004; Keskin, 2007). In all the three protein systems studied, unliganded proteins possess a pre-existing fluctuation mechanism of transition between the open/closed conformations in the unliganded form, i.e. it is in fact not the ligand that induces the conformational change, but instead the unliganded protein that we simulate exhibits transitions between the open and closed states. The transition is derived by the lowest frequency normal-modes even in the absence of ligand molecules. Nevertheless, ligands have a stabilizing effect on the closed forms (Oh *et al.,* 1993; Keskin, 2007).

The reverse transitions were also studied with ANM-MC, RG-ANM-MC and TMC methods with application to AK and hemoglobin. Although the target state is approached to some extent, the reverse transitions could not be completely achieved by the normal mode based methodologies, namely ANM-MC and RG-ANM-MC in the case of AK. In the case of hemoglobin both reverse and forward transitions could be better accomplished by both techniques. On the other hand, by TMC, the reverse transitions could be accomplished with almost complete match to the target state in both test cases. The intermediate structures are observed to be similar in the case of forward and reverse transitions.

As compared to previous ENM based conformational transition studies, (Xu *et al.,* 2003; Maragakis and Karplus, 2005; Zheng *et al.,* 2007) the methodologies developed in this thesis provide more feasible pathways by updating the normal modes at each step followed by energy minimizations both in the presence and absence of final state structure information. The protocols developed in this thesis are also advantageous due to short CPU times required for completion. The ANM based techniques (ANM-MC, RG-ANM-MC) that perform NMA almost take twice the time that of TMC takes, for a complete run for the protein to accomplish transition from open to closed conformation. For instance, with

MCS=1000, a complete run of AK from the open to closed state lasts about 5 hour for attaining maximum approach by ANM-MC. Same simulation by TMC requires about 2 hours and 50 minutes. In the case of RG-ANM-MC the search for a plausible closed state requires approximately 5 hours and 15 minutes for completion.

7.2. Recommendations

In the present thesis, three different methodologies, namely ANM-MC, TMC, and RG-ANM-MC gave satisfactory results in proteins with varying sizes, structures, and amplitude of transitions such as AK, hemoglobin, HSTR, and LAO-binding protein. ANM-MC and TMC are developed for analyzing the conformational transitions between two known states of a protein, i.e. open-to-closed, or vice versa. However, in cases when a target structure is unavailable TMC method can not be used. Instead, the extension of ANM-MC, RG-ANM-MC can be used to suggest plausible closed states using only the initial structure information. Hence, this method can be applied to cases with or without target state information. However it is worth noting that, the methodologies prove successful with proteins undergoing global and large amplitude conformational changes such as opening and closing motion. The protocols should be further adjusted to simulate proteins for which more localized motions are of interest such as an opening closing of a small loop section.

TMC simulations, which are performed without an ANM analysis, provided much closer approaches in shorter times to the target state due to its forcing mechanism towards the final structure. However, the algorithm is totally inapplicable in cases where the target information is not known. In cases where the target structure is available, an improved technique could be a combination of ANM-MC and TMC methods. Initially, ANM-MC can be employed and after RMSD values level off, TMC may be applied to provide a closer mapping to the target state and intermediates that lie close to the target.

These methodologies can also be adjusted for simulating protein-DNA or protein-RNA systems to investigate the corresponding changes upon nucleic acid binding. This requires the adjustment of energy parameters used in MC simulation for DNA and RNA.

For instance, p53-DNA or ribosome-RNA complexes, which are important protein-nucleic acid complexes, could be explored.

In the present thesis, the similarity of the simulation intermediates with the x-ray structures related to the protein family is analyzed by superimposing the structures and computing the corresponding RMSD values by using Pymol. This analysis could be performed in a more automated way by incorporating pairwise sequence alignment algorithms into the present protocol. By this way, the search for the x-ray structures that may be present on the transition pathway through the protein data bank could be much more efficient. Finally, all the three approaches presented may be used successfully to investigate the conformational transitions in protein-ligand, protein-protein or protein-nucleic acid systems. ANM-MC and RG-ANM-MC methodologies can be used in modeling the conformational transitions of proteins in cases where the target state structure is available or not, respectively.

APPENDIX A: SUPERIMPOSITION PROCEDURE

The steps encountered during superimposition of initial and final structures are listed below.

1) Make transformation: equate the center of mass of both initial and final structures to zero.
2) Calculate $\mathbf{AB^t}$ (**A**:coordinate matrix of the initial structure; **B**: coordinate matrix of the target structure. Both matrices have the dimensions *N*x3, *N* being the residue number and 3 stands for x, y, and z coordiantes)
3) Perform singular value decomposition of $\mathbf{AB^t}$ to obtain $\mathbf{LDV^t}$. **D** is the diagonal matrix and matrices **L** (3x3) and $\mathbf{V^t}$ (3x3) are computed.
4) Rotation matrix, $\boldsymbol{R}$ (3x3) is then calculated from: $\boldsymbol{R} = \mathbf{L\ S\ V^t}$ where matrix **S** (3x3) is identitiy matrix diag (1,1,1) if det(**L**).det(**V**)=1; or diag(1,1,-1) if det(**L**).det(**V**)=-1.
5) Multiplication of the matrix **B** (*N*x3) with rotation matrix **R** (3x3) gives the new rotated coordinates of matrix **B**, $\mathbf{B_{rot}}$ (*N*x3).
6) The target direction ($\boldsymbol{Q}$ matrix with dimensions *N*x3) is then defined by subtracting the initial state coordinates from the rotated new coordinates of target with following notation: $\boldsymbol{Q} = \mathbf{B_{rot} - A}$.

REFERENCES

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, 1994, "Molecular Biology of the Cell", 3rd edition, Garland Publishing, New York.

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.

Amadei, A., B. M. Linssen, B. L. de Groot, van Aalten, D. M. F., and H. J. C. Berendsen, 1996, "An efficient method for sampling the essential subspace of proteins", *Journal of Biomolecular Stucture and Dynamics*, Vol. 13, pp. 615-625.

Anderson, B. F., H. M. Baker, G. E. Norris, S. V. Rumball, and E. N. Baker, 1990, "Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins", *Nature (London),* Vol. 344, pp. 787-790.

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

Bahar, I., A. R. Atilgan, M. C. Demirel, and B. Erman. 1998, "Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability". *Physical Review Letters*, Vol 80 (12), pp. 2733-2736.

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

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., and R. L. Jernigan, 1997, "Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation", *Journal of Molecular Biology,* Vol.1, pp. 195-214.

Baker, H. M., C. J. Baker, C. A. Smith, E. N. Baker, 2000, "Metal substitution in transferrins: specific binding of cerium(iv) revealed by the crystal structure of cerium-substituted human lactoferrin". *Journal of Biology and Inorganic Chemisrty* Vol.5, pp. 692-698.

Baker, H. M., B. F. Anderson, and E. N. Baker, 2003, "Dealing with iron: Common structural principles in proteins that transport iron and heme", *Proceedings of National Academy of Science U.S.A.,* Vol. 7, pp. 3579-3583.

Berendsen, H. J. C., and S. Hayward, 2000, "Collective protein dynamics in relation to function", *Current Opinion in Structural Biology,* Vol. 10, pp.165–169.

Berry, M.B., B. Maedor, T. Bilderback, P. Liang, M. Glaser, and G. N. Phillips, 1994, "The closed conformation of a highly flexible protein: the structure of the *E. coli* adenylate kinase with bound AMP and AMPPNP". *Proteins,* Vol. 19, pp. 183-198.

Berry, M. B., E. Y. Bae, T. R. Bilderback, M. Glaser, and G. N. Phillips, 2006, "Crystal structure of ADP/AMP complex of Escherichia coli adenylate kinase". *Proteins* Vol. 62(2), pp. 555-556.

Branden, C. and J. Tooze, 1999, *Introduction to Protein Structure*, 2nd ed., Garland Publishing, Inc., NewYork.

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, B., and M. Karplus, 1985, "Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme", *Proceedings of National Academy of Science USA,* Vol. 82, pp. 4995–4999.

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

Cavanagh, J., and R.A. Venters, 2001, "Protein dynamic studies move to a new time slot", *Nature Structural Biology* Vol. 8, pp. 912-914.

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., 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., 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., 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.

El Hage Chahine, J.-M., and R. Pakdaman, 1995, "Transferrin, a mechanism of iron release", *European Journal of Biochemistry,* Vol. 230, pp. 1102-1110.

Emekli, U., D. Schneidman-Duhovny, H. J. Wolfson, R. Nussinov, and T. Haliloglu, 2007, "HingeProt: Automated Prediction of Hinges in Protein Structures". *Proteins,* Vol.70 (4), pp. 1219-1227.

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, P. G. Wolynes, 1991, "The energy landscapes and motions of proteins", *Science*, Vol. 254, pp. 1598-1603.

Garcia, A. E., 1992, “Large-amplitude nonlinear motions in proteins”. *Physical Review Letters*, Vol. 68, pp. 2696-2699.

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, T., 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.

Greenleaf, W. J., M. T. Woodside, and S. M. Block, 2007, “High Resolution Single Molecule Measurements of Biomolecular Motion”, *Annual Review of Biophysics and Biomolecular Structure*, Vol. 36, pp. 171-190.

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.

Ha, T., 2001, “Single-Molecule Fluorescence Resonance Energy Transfer”, *Methods,* Vol. 25, pp. 78–86

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.

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.

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.

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

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.

Jernigan, R. L., M. C. Demirel and I. Bahar, 1999, "Relating Structure to Function through the Dominant Modes of Motion of DNA Topoisomerase II", *International Journal of Quantum Chemistry*, Vol. 75, pp. 301-312.

Kang, C., W. Shine, Y. Yamagataj, S. Gokcen, G. Ferro-Luzzi Ames, S. Kim, 1991, "Crystal Structure of the Lysine-, Arginine-, Ornithine-binding Protein (LAO) from *Salmonella typhimurium* at 2.7-Å Resolution", *Journal of Biological Chemistry,* Vol.266 (35), pp. 23893-23899.

Kantarci, N., P. Doruker, and T. Haliloglu, 2006, “Cooperative fluctuations point to the dimerization interface of p53 core domain”, *Biophysical Journal*, Vol. 91(2), pp. 421-432.

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., T. Haliloglu, and P. Doruker, 2009, “Prediction of a plausible closed conformation from open state: explored by elastic network model with control of radius of gyration” (in preparation).

Karplus, M., and J. N. Kushick, 1981, “Method for estimating the configurational entropy of macromolecules”, *Macromolecules* Vol. 14, pp. 325-332.

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.

Kay, L.E., 1998, “ Protein dynamics from NMR”, *Nature Structural Biology Vol.* **5**, pp. 513–517.

Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, 1958, "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis". *Nature* Vol. 181 (4610), pp. 662–666.

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, 2005, "Rigid-cluster models of conformational transitions in macromolecular machines and assemblies", *Biophysical Journal,* Vol.89:, pp. 43-55.

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

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

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.

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 and R. L. Jernigan, 2005, "Collective Dynamics of Large Proteins from Mixed Coarse-Grained Elastic Network Model", *QSAR and Combinatorial Science*, Vol. 24, pp. 443-448.

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

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.

Levinthal, C, 1969, "How to Fold Graciously. Mossbauer Spectroscopy in Biological Systems", Proceedings of a meeting held at Allerton House, Monticello, Illinois. J.T.P. DeBrunner and E. Munck eds., University of Illinois Pres, pp. 22-24

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.

Levy, R. M., A. R. Srinivasan, W. K. Olson, and J. A. McCammon, 1984, "Quasiharmonic method for studying very low frequency modes in proteins", *Biopolymers* Vol. 23, pp. 1099-1112.

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.

Lou, H., and R. I. Cukier, 2006, "Molecular dynamics of apo-adenylate kinase: a principal component analysis", *Journal of Physical Chemistry B,* Vol. 110, pp. 12796-12808.

Lundstrom, K, 2006, "Structural genomics for membrane proteins". *Cellular and Molecular Life Sciences,* Vol. 63 (22), pp. 2597–607.

Lundstrom, K, 2004, "Structural genomics on membrane proteins: mini review". *Combinatorial Chemistry and High Throughput Screening,* Vol. 7 (5), pp. 431–439.

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

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.

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.

Metropolis, N., and S. Ulam, "The Monte Carlo Method", 1949, *Journal of the American Statistical Association*, Vol. 44, No. 247, pp. 335-341.

Mizutani, K., B. K. Muralidhara, H.Yamashita, S. Tabata, B. Mikami, and M. Hirose, 2001, "Anion-Mediated Fe^{3+} release mechanism in ovotransferrin C-lobe: A structurally identified SO_4^{2-} binding site and its implications for the kinetic pathway", *Journal of Biological Chemistry*, Vol. 276, pp. 35940-35946.

Moore, S. A., B. F. Anderson,C. R. Groom, M. Haridas, and E. N. Baker, 1997, "Three-dimensional structure of diferric bovine lactoferrin at 2.8 A resolution", *Journal of Molecular Biology,* Vol. 274, pp. 222-236.

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.

Muller, E. M., A. de Maijere, and H. Grubmuller, 2002, "Predicting unimolecular chemical reactions: chemical flooding", *Journal of Chemical Physics*, Vol. 116, pp. 897-905.

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

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

Müller, C. W., and G. Schulz, 1993, "Crystal structures of two mutants of adenylate kinase from *Escherichia coli* that modify the gly-loop", *Proteins: Structure Function and Genetics,* Vol. 15, pp. 42-49.

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.

Northrup, S.H., and J. A. McCammon, 1980, "simulation methods for protein structure fluctuations", *Biopolymers*, Vol. 19 (5), pp. 1001-1016.

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.

Palmer, A.G., C. D. Kroenke, and J. P. Loria, 2001, "NMR methods for quantifying microsecond-millisecond motions in biological macromolecules", *Methods in Enzymology,* Vol. 339, pp. 204–238.

Palmer, A.G, 1997, "Probing molecular motion by NMR", *Current Opinion in Structural Biology,* Vol. 7, pp. 732–737.

Patrone, P., and V. S. Pande, 2006, "Can conformational change be described by only a few normal modes?", *Biophysical Journal* Vol. 90 (5), pp. 1583-1593.

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)

Romo, T. D., J. B. Clarage, D. C. Sorensen, G. N. Phillips Jr, 1995, "Automatic identification of discrete substates in proteins: singular value decomposition analysis of

time-averaged crystallographic refinements", *Proteins: Structure Function and Genetics*, Vol. 22, pp. 311-321.

Sara, R., R. Garratt, B.Gorinsky, H. Jhoti, P. Lindley, 1990, "High-resolution X-ray studies on rabbit serum transferrin: preliminary structure analysis of the N-Terminal half-molecule at 2.3 angstroms resolution", *Acta Crystallographica Sect.B*, Vol. 46, pp.763-771.

Scapin, G, 2006, "Structural biology and drug discovery", *Current Pharmeceutical Design,* Vol. 12 (17), pp. 2087–97

Schlauderer, G., and G. Schulz, 1996, "The structure of bovine mitochondrial adenylate kinase: comparison with isoenzymes in other compartments", *Protein Science* Vol. 5, pp. 434-441.

Schlauderer, G., K. Proba, and G. Schulz, 1996, "Structure of a mutant adenylate kinase ligated with an ATP-analogue showing domain closure over ATP", *Journal of Molecular Biology,* Vol. 256, pp.223-227.

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.

Shaanan, B, 1983, "Structure of human oxyhemoglobin at 2.1 angstroms", *Journal of Molecular Biology,* Vol. 171 (1), pp. 31-59.

Silva, M. M., P. H. Rogers, and A. Arnone, 1992, "A third quaternary structure of human hemoglobin A at 1.7 Å resolution", *Journal of Biological Chemistry,* Vol. 267, pp. 17248-17256.

Srinivasan, R., and G. D. Rose, 1994, "The T-to-R transformation in hemoglobin: a reevaluation", *Proceedings of National Academy of Sciences U.S.A.* Vol. 91, pp. 11113-11117.

Stryer, L, 1988, Biochemistry. 3d Ed. New York: W. H. Freeman and co.

Stryer, L., 1978, "Fluorescence Energy Transfer As a Spectroscopic Ruler", *Annual Review of Biochemistry*, Vol. 47, pp. 819-846.

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.

Tame, J. R., and B. Vallone, 2000, "The structures of deoxy human hemoglobin and the mutant Hb Tyralpha42His at 120 K", *Acta Crystallographica D. Biological Crystallography* Vol. 56 (7), pp. 805-811.

Teeter, M. M., and D. A. Case, 1990, "Harmonic and quasi harmonic descriptions of crambin", *Journal of Physical Chemistry*, Vol. 94, pp. 8091-8097.

Tobi D, and I. Bahar I, 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.

Umeyama, S., 1991, “Least-squares estimation of transformation parameters between two points patterns”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13 (4), pp. 376-380.

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.

Van Gunsteren, W. F., and H. J. C. Berendsen, 1990, “Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry”, Angewandte Chemie International Edition, Vol. 29, pp. 992-1023.

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.

Whitford, P. C., O. Miyashita, Y. Levy, and J. N. Onuchic, 2007, “Conformational transitions of adenylate kinase: switching by cracking”, *Journal of Molecular Biology,* Vol. 366, pp. 1661-1671.

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.

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.

Zheng, W., and S. Doniach, 2003, "A comparative study of motor protein motions by using a simple elastic-network model", *Proceedings of National Academy of Sciences U.S.A.*, Vol. 100, pp. 13253–13258.

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, G. Hummer, 2007, "Protein conformational transitions explored by mixed elastic network models", *Proteins*, Vol. 69. pp. 43-57.