ANALYSIS OF CRACKING PHENOMENA ALONG CONFORMATIONAL TRANSITION PATHWAYS OF PROTEINS

by

Çağlar Meriçer

B.S., Chemical Engineering, Istanbul Technical University, 2009

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

Graduate Program in Chemical Engineering

Boğaziçi University

2012

ACKNOWLEDGEMENTS

Firstly, I would like to thank my thesis supervisor, Prof. Pemra Doruker Turgut, for her very precious guidance and support through all periods of this research. It was a great pleasure and honor for me to work with her. I also owe thanks to the other members of my thesis committee, Prof. Türkan Haliloğlu and Assoc. Prof. Ebru Demet Akdoğan, who have read and commented on my thesis.

Special thanks to Selin Okur, Serhan Güner, İsa Çal, Pelin Dönümcü, Aybala Beste Türkay, Mehmet Ünal Güneş, Duygu Kocaman, Merve Eropak, Bahar Nalbantoğlu and Okan Yüzüak for their invaluable presence in my life which helped to provide motivation for this study. I am very grateful to all the members of Polymer Research Center for their suggestions and friendships, in particular to Arzu Uyar, who really helped me kindly with her knowledge and experience during the whole time.

I also want to express my gratitude to my colleagues Şefik Kerem Ovalı, Özer Özcan, Aybüke Leba, Murat Erol, Melek Selcen Başar, Caner Ülgüel, Nilay Aktürk, Ali Uzun, İpek Paksoy, Burcu Yoğurtçu, Aslıgül Doğan, Simay Yalaz, İhsan Ömür Akdağ, Mehmet Tarık Can, Gülsüm Kaya Ersoy, Mehmet İrfan Hösükoğlu, Vasfiye Çimenoğlu, Sıla Kumbasar and Ahmet Özcan for making the period of my master education more enjoyable and cheerful.

Finally, my heartfelt thanks go to all my family, especially İhsan Meriçer, Nezahat Meriçer and my brother Zafer Meriçer, for their endless support as always during my life. This thesis is dedicated to my family, since it would not have been possible without them.

ABSTRACT

ANALYSIS OF CRACKING PHENOMENA ALONG CONFORMATIONAL TRANSITION PATHWAYS OF PROTEINS

The weighted ensemble (WE) methodology was incorporated into an iterative ANM-MC (Anisotropic network model- Monte Carlo) simulation algorithm, which has been developed to generating targeted conformational transition pathways of proteins. Through the WE sampling, it is possible to produce unbiased pathways by increasing the number of ANM-MC simulations along the reaction coordinate, which is specified as the root-mean-squared distance to target structure. In this modified algorithm, the structures are deformed along randomly chosen (unbiased) ANM modes (among the slowest five modes) and then their energy is minimized by Monte Carlo (MC) method. Three proteins were studied using 18 Å and 10 Å cutoff distances and distinct successful pathways were obtained for a total of 13 systems (simulations). Various sequences of domain motions (LID and AMP domain closures) were observed for adenylate kinase (AdK) systems. Calmodulin (CaM) systems followed quite similar pathways, whereas glutamine binding protein (GBP) displayed high conformational freedom. Strain energies and dihedral angle changes of each pathway for all systems were calculated to detect possible cracking regions where high strain energy causes local unfolding. Strained regions well matched with other studies for AdK with five out of eight previously reported hinge regions having high strain energies. Two sites for AdK and CaM can be reported for cracking and there was no cracking in GBP motion.

ÖZET

PROTEİNLERİN KONFORMASYON GEÇİŞ YOLİZLERİ ÜZERİNDE ÇATLAMANIN İNCELENMESİ

Protein biliminde önemli bir konu olan yapı-işlev ilişkisini açıklamak için pek çok hesaplamalı yöntem geliştirilmiştir. Elastik ağyapı modelleri konformasyonel geçişleri incelemek için basit bir yöntem olup, oldukça başarılı sonuçlar vermektedir. Anizotropik ağyapı modeli (ANM) yalnızca C^{α} atomlarını kullanarak, ürettiği kollektif hareketlerin yavaş modları ile yapıyı deforme ederek çalışır. Bu yöntemlerin çoğunda hedef yapıya ulaştıracak şekilde zorlanmış deformasyon yapılır. Ancak büyük konformasyon değişimi geçiren moleküller farklı yolizlerini takip edebilirler. Bunun için örnekleme metodları kullanılarak istatistiksel olarak hedef yapıya ulaşma şansı artırılır. Ağırlıklı örnekleme metodu (WE) ilerleme koordinatını küçük parçalara ayırarak dinamikler sonucu her parçada sabit sayıda ve olasılıkta simülasyon kalmasını ayarlar. Bu çalışmada ANM metodunun ürettiği en yavaş beş moddan bir tanesi rastgele seçilerek yapı o yönde deforme edilir. Daha sonra MC metoduyla enerji minimizasyonu yapılır. Üç farklı protein için iki farklı etkileşim uzaklığı (18 Å ve 10 Å) kullanılarak 13 sistem çalıştırılmıştır. AdK sistemleri farklı bölgelerinin sırayla hareketlerini içeren yolizlerini ortaya çıkarmıştır. CaM yolizleri birbirine benzerken GBP yapısı gereği çok hareketlidir ve en başından itibaren farklı yolizlerini takip eder. Çalışmanın diğer bölümünde ise gerilme enerjisi ile dönme açılarının yüksek olduğu yerlerde çatlamanın (ikincil yapıda bozulma) gösterilmesi amaçlanmıştır. Buna gore AdK ve CaM için iki bölge bulunmuştur. GBP hareketinde ise çatlamaya rastlanmamıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

C^{α}	Alfa carbon atom
D_i	Strain energy of residue i
H	Hessian matrix
K	Kelvin
M	Number of simulations per bin
N_{bin}	Number of bins
N_{res}	Number of residues
N_{τ}	Number of iterations
r_c	Cutoff distance
$\mathbf{R}_{ij}$	Instantaneous position vector between residues i and j
$\mathbf{R}_{ij}^{0}$	Equilibrium position vector between residues i and j
T_i	Unfolding parameter of residue i
V	Harmonic potential of a spring
V_{tot}	Total potential energy of ANM
Å	Angstrom
γ	Force constant
θ_{AMP}	Inter-domain angle between AMP and CORE domains
θ_{LID}	Inter-domain angle between LID and CORE domains
Δ	Delta value between two distances
ΔD_{RMSD}	Difference between RMSDs to closed and open structure
$\Delta\mathbf{R}_i$	Instantaneous fluctuation vector of residue i
$\Delta\mathbf{R}_j$	Instantaneous fluctuation vector of residue j
$\Delta\mathbf{R}_{ij}^{0}$	Equilibrium fluctuation vector of residues i and j
ΔX_{ij}	X-component coordinate difference between residues i and j
ΔY_{ij}	Y-component coordinate difference between residues i and j
ΔZ_{ij}	Z-component coordinate difference between residues i and j

LIST OF ACRONYMS/ABBREVIATIONS

AdK	Adenylate Kinase
ADP	Adenosine Diphosphate
AMP	Adenosine Monophosphate
ANM	Anisotropic Network Model
ANM-MC	Anisotropic Network Model
ATP	Adenosine Triphosphate
Ca	Calcium
CaM	Calmodulin
Cryo-EM	Cryo-Electron Microscopy
DF	Deformation Factor
DIMS	Dynamic Importance Sampling
DNA	Deoxyribonucleic Acid
ENM	Elastic Network Model
FFS	Forward Flux Sampling
FRET	Fluorescence Resonance Energy Transfer
GBP	Glutamine Binding Protein
GNM	Gaussian Network Model
MC	Monte Carlo
MCS	Monte Carlo Step
MD	Molecular Dynamics
Mg	Magnesium
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
PNM	Plastic Network Model
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
RP-TMD	Restricted Perturbation Targeted Molecular Dynamics
TIS	Transition Interface Sampling

TMC	Targeted Monte Carlo
TMD	Targeted Molecular Dynamics
TPS	Transition Path Sampling
tRNA	Transfer RNA
WE	Weighted Ensemble
YCM	Yeast Chorismate Mutase

1. INTRODUCTION

Proteins are essential macromolecules for living organisms because of participating in specific functions such as catalysis and signal transduction. These functions are dependent on the different conformations that the protein can adopt around the native state due to its flexibility. Conformational transition for proteins is generally observed during the procedure of interactions with other proteins or ligand binding in order to perform their function. Common examples of proteins exhibiting conformational transitions are hemoglobin as a transport protein, enzymes such as adenylate kinase, signaling proteins such as calmodulin, and myosin from motor proteins (Berg *et al.*, 2002).

Fluctuations around the native state are suggested to be crucial for conformational changes and biological functions. Lowest frequency modes obtained from normal mode analysis (NMA) has been shown to correlate with the conformational change direction that occurs upon ligand binding for a number of proteins (Tama and Sanejouand, 2001; Dobbins *et al*., 2008). Coarse-grained (low-resolution) NMA such as elastic network models (ENM) provide computational high efficiency and have been used widely for determining structure-function relationship of proteins (Bahar and Rader, 2005; Ma, 2005).

Main experimental methods for identifying the three-dimensional (3D) structure of proteins are NMR spectroscopy, X-ray crystallography and cryo-electron microscopy. Although these methods may provide sufficient information about the major conformational states of proteins such as the apo (free) and the ligand-bound states, they have difficulties in representing less populated intermediates and dynamics at an atomistic level. As a result, both atomistic and coarse-grained computational models and simulations have gained importance for conformational search and transition pathway analysis (Lei *et al*., 2004; Maragakis and Karplus, 2005; Zhang *et al*., 2007; Liwo *et al*., 2008; Kantarci-Carsibasi *et al*., 2008).

Standard molecular dynamics (MD) and Monte Carlo (MC) simulations and elastic network models can be modified for producing large-scale conformational transitional pathways of biological macromolecules that are targeted or biased between the apo and bound (open and closed) states. MD and NMA methods mostly utilize all atomistic empirical potentials, while MC and ENM are featured as coarse-grained methods. One of the ENM methods, Anisotropic Network Model (ANM), is used here with the collaboration of MC simulations.

In this study, conformational transition pathways of different proteins were investigated by using simultaneous ANM-MC simulations with applying the principles of the Weighted Ensemble methodology. First (slowest) five modes of ANM are randomly chosen to obtain the next intermediate structure by giving a small perturbation along the chosen direction and the final structure is reached after energy minimization procedures by Monte Carlo (MC) simulations. Structures are not forced to move towards any direction (unbiased) but the Weighted Ensemble sampling algorithm creates many trajectories that some of them make their way to the target (desired) state. At the end of simultaneous simulations, strain energies of the successful ensemble of structures following each unbiased pathway were calculated for three different proteins. Accumulation of the strain energy in certain parts of the structures is claimed to cause partially unfolding (cracking); hence, strain energies along with torsion angles are analyzed to observe the possible result of cracking (due to high strain energy) on the structure and conformational transition.

2. CONFORMATIONAL TRANSITIONS OF PROTEINS AND CRACKING

2.1. Conformational Dynamics of Proteins

Scientific research about protein dynamics focuses on exploring the relationship between structure and function of proteins. The three-dimensional arrangement of atoms in a bio-macromolecule and the conformational space that is available to it are key determinants of biological function. Thus, experimental and computational data on protein structure and dynamics are crucial for interpreting physical, chemical and biological properties of a macromolecule (Leach, 2001). Even small variations in bond lengths and bond angles, more importantly rotations around bonds, result in conformational changes. There are also motions and transitions on a larger time scale, i.e. on the order of nanoseconds to milliseconds, which are called collective or domain motions.

Ligand binding and conformational changes are key determinants of protein function such as catalysis and signaling. In this respect, experimentally-determined protein structures have shown that protein-ligand binding is generally coupled to either shear or hinge-bending type domain motions and related conformational changes of proteins. Historically, three distinct models can be mentioned that try to explain binding mechanisms of proteins. In the lock-and-key model, protein and ligand interaction is resembled to the perfect fit of a lock and key, such that the key only fits and binds to the proper rigid conformation of the protein (Fischer, 1894). Induced-fit model introduces and revises the classical lock-and-key model such that the flexible protein structure can undergo conformational adjustments while binding to the ligand (Koshland, 1958). Current view of binding is based on the pre-existing equilibrium hypothesis suggesting that the protein can sample multiple conformations at equilibrium, among which a specific one is stabilized upon ligand binding (Weber, 1972).

There are currently more than 80,000 structures including protein, DNA, RNA and their complexes which are experimentally found and published in Protein Data Bank (PDB) (Berman *et al*., 2000). X-ray crystallography (87.5%) and nuclear magnetic resonance (NMR) (11.5%) are the leading techniques used for determining the 3D structures. X-ray is preferred to NMR for large molecules because of the size limitation in NMR technique. Electron microscopy (cryo-EM) is used for determining the 3D structures of supramolecular assemblies, such as the ribosome, together with X-ray data. Cryo-EM and fluorescence resonance energy transfer (FRET) have recently gained interest as important techniques for investigating conformational transition mechanisms of proteins and their complexes (Frank, 1996; Kasprzak *et al*., 1988; Lakowicz, 1999; Lesk, 2001). As experimental techniques gain interest, validation of the data becomes crucial, because of the noise in the variables and probable mistakes during data-collecting stages of the experiments (Kleywegt, 2009).

In the context of conformational transitions, molecular simulation techniques offer valuable insight to experiments in terms of scanning the conformational space and revealing intermediate structures. Conventional molecular dynamics (MD) simulations are generally inefficient for the investigation of large system sizes, comprising conformational transitions of large proteins and/or transitions with large RMSDs between the initial and final states. MD simulations that provide full atomistic details of the structure are still applicable for relatively small proteins or for describing the preliminary stages of the conformational transition.

Low-resolution models and simulations have gained interest due to their computational efficiency in simulating larger proteins and even supramolecular systems, such as the ribosome. Even though the original geometry and interaction potentials are coarse-grained that lead to inaccuracy or lower precision, these models still provide valuable information for conformational dynamics. One successful example is a coarse-grained Monte Carlo technique (Haliloglu and Bahar, 1998; Kurt and Haliloglu, 1999) that has been developed for the simulation of protein conformational dynamics. In this model, two coarse-grained sites represent each residue, which are located at the α-carbon atom of backbone and the side chain

interaction site for related amino acid. Randomly chosen sites are displaced by a small perturbation and the new structure is accepted if the coarse-grained energies of current and previous conformations satisfy the Metropolis criterion.

More simple models, such as the elastic network models, have been found successful in terms of describing fluctuation dynamics of proteins. In these models, the protein structure is represented as a network of coarse-grained nodes, which are located at the α-carbons of the backbone and inter-connected by harmonic springs. Residues fluctuate about their mean positions with a Gaussian distribution. Gaussian Network Model (GNM) results verify experimental temperature factors (Bahar *et al*., 1997; 1998; 1999; Haliloglu *et al*., 1997; Bahar and Jernigan, 1998; 1999). However, all fluctuations in the GNM are isotropic, where only magnitude of the fluctuation is known. In reality, three dimensional (3D), anisotropic fluctuations should be accounted for exploring functional dynamics and mechanisms. Anisotropic Network Model (ANM), a 3D version of GNM, has been applied for many proteins and their complexes (Doruker *et al*., 2000; Atilgan *et al*., 2001). Successful results have been obtained considering the comparison with experimental temperature factors and verification of global motion mechanism of the protein structure for ligand binding (Tama and Sanejouand, 2001; Bahar, 1999).

A comparison of analytical approaches GNM and ANM with MD simulations (Doruker *et al*., 2000) has been performed by applying these methods to α-amylase inhibitor. All approaches exhibit similar characteristics in terms of mean-square fluctuations of residues and cross-correlations between domain movements. Still, ANM is more precise in terms of conforming with MD results considering the contribution of low-frequency modes and magnitudes of motion.

2.2. Unbiased Conformational Transition Pathways

Computational methods used for the exploration of conformational transition pathways generally employ targeted (biased) simulations starting from the open (apo/free) structure to reach the closed (holo/complex) structure. For instance, Kong *et*

al. (2006) used targeted molecular dynamics (TMD), which has been proposed by Schlitter *et al*. (1993), to study the transition of yeast chorismate mutase (YCM) by forcing the intermediate structure towards the target state with a magnitude proportional to the distance between two structures at each MD step. TMD method was also applied to alanine dipeptide using the restricted perturbation TMD method (RP-TMD) in order to ensure minimum free energy pathway (van der Vaart and Karplus, 2005). RP-TMD both minimizes the root-mean square distance/displacement (RMSD) and free energy by giving a perturbation after each MD step which prevents crossing high energy barriers while moving towards target structure (van der Vaart and Karplus, 2005; 2007).

In another recent study (Kantarci-Carsibasi *et al*., 2008), ANM-MC and targeted MC (TMC) methodologies were applied to adenylate kinase (AdK) to sample biased conformational transition pathways. ANM-MC chooses among the slowest collective modes the mode that has the highest overlap towards the final structure and then deforms the structure along that this mode. with a weight of desired deformation factor in ANM part before energy minimization with MC iterations. In TMC method, instead of ANM calculations, slowest collective modes are not accounted and structures only follow the target direction while the rest of the algorithm remains the same. Both methods produce successful intermediates when compared to experimental crystal structures.

However, proteins that undergo large conformational changes may follow alternative paths between initial and final states. Thus, determination of unbiased alternative paths between apo and bound states of proteins may be crucial for describing structure-function relationships.

In CONTRA MD (Conformational Transitions by Molecular Dynamics with minimum biasing) method (Harvey and Gabb, 1993), short pieces of MD simulations have been used to investigate the slow conformational transitions of deoxyadenosine and phenylalanine tRNA. The algorithm keeps track of the last conformation and checks the value of chosen criteria in the progress coordinate at the end of each piece and accepts only if the last conformation has evolved towards the final target conformation

by satisfying the criteria, otherwise, it is deleted. Applying this procedure yields a trajectory/sequence of conformations that show a random path towards the target structure. Three pathways were found at 300 K for the transition of deoxyadenosine using a function of two different angles as progress coordinate. Same conditions were used for analyzing the hinge bending motion of phenylalanine tRNA with the bending angle between two arms of the molecule as the progress coordinate of the simulations. It is concluded that minimum biasing (no guiding force) principle of CONTRA MD method preserves van der Waals interactions and thus gives reasonable results compared toadiabatic mapping method for both opening and closure of the structure (Harvey and Gabb, 1993).

Huber and Kim (1996) have developed a new algorithm, named as the weighted ensemble (WE) method, to simulate infrequent events such as those that take place during the conformational transition of a protein. The algorithm avoids free energy barrier restrictions by dividing the reaction coordinate into small bins and running parallel simulations to enrich the sampling of rare events that progress towards the target state. The method has been applied together with Brownian dynamics simulation by assuming one-dimensional energy surface model with a barrier and a model for protein-protein interaction reactions (Huber and Kim, 1996). The decision of bin placement, simulation and probability adjustments during the transition are important for efficiency of the algorithm and their definitions are explained in Chapter 3.5 of this thesis in detail. .

The WE method has also been adopted to produce unbiased pathways for the conformational change of calmodulin, which has a root-mean square distance/displacement (RMSD) of more than 10 Å between the open and closed conformations (Zhang *et al.*, 2007). In this work, a double-Gö potential is used for mimicking the conformational transition between the two end structures of the 72-residue model of calmodulin. The computationally efficient WE method results have been in reasonable agreement with those of brute-force simulations, considering the heterogeneity of pathways, the duration of distinct pathways and the reaction rates.

The WE method is also applicable to non-Markovian processes, such as forward flux sampling (FFS) (Allen *et al*., 2005), dynamic importance sampling (DIMS) (Zuckerman and Woolf, 1999), transition path sampling (TPS) (Dellago *et al*., 1998), transition interface sampling (TIS) (van Erp *et al*., 2003), where the history information of the transition is required for simulations. Variable binning with respect to time can also be used to produce different but statistically correct samplings. It is shown that alternating the number of bins during simulation produces the same results with the static bins case and this ensures the correctness of WE method (Zhang *et al*., 2010).

In another recent study by Bhatt and Zuckerman (2010), the WE method has been applied to adenylate kinase for exploring path heterogeneity with three different semi-atomistic models in order to improve the chemical efficiency of coarse-grained simulation methods. The backbone is represented with fully atomistic detail in each model but there are differences in inter-residue interactions. First and third models use only symmetric double-Gö interactions, whereas potential energy terms due to hydrogen bonding, Ramachandran propensities, and residue-specific contact interactions are also included in the second model. Third model, however, is designed for describing more stable closed crystal structure of adenylate kinase. It is found that two pathways (Open-LID-AMP-Closed and Open-AMP-LID-Closed) are available in the ensemble for the transition from open to close state in first two models, where reverse transition is much slower and no symmetry is encountered (only Closed-LID-AMP-Open). Both pathways are observed in third model (Bhatt and Zuckerman, 2010).

2.3. Cracking Phenomena during Transitions

Local unfolding events, called cracking, have been proposed to take place due to the accumulation of strain energy during conformational transitions of proteins. The relatively recent studies on cracking mainly focus on adenylate kinase (AdK), which is an important enzyme that catalyzes the reversible reaction (2 ADP ↔ AMP + ATP) in living cells.

Miyashita *et al*. (2003) used linear and nonlinear elastic models for investigating the conformational transition of adenylate kinase and concluded that the strain energy is localized in specific regions of the structure. Residues 10-12 in the loop region along with residues 110-120 and 160-170 constituting alpha helix secondary structures have higher strain energies than any other parts of the structure. These regions correlate well with hinge regions of adenylate kinase and they are proposed to unfold partially (cracking) due to high strain energy in these residues.

It has been proposed that partial unfolding of secondary structural elements (mainly helices) reduces strain during conformational change and increases catalytic activity of the enzyme. Regions including residues 60-70, 120-125 have higher average strain energies while there are also less significant peaks on residues 10-20, 30-35, 80-90 and 170-180. It has been found out that similar regions are under high strain with the studies using normal mode analysis except residues 30, 60 and 80 which are believed to represent high strain due to AMP domain closure at the end of the transition. Moreover, unfolding was measured using dihedral angles of intermediates and crystal structures and then regions are compared with strain energies to prove the relation between unfolding and high strain energy (Whitford *et al*., 2007).

Studies for adenylate kinase have shown that intrinsic fluctuations cause LID domain motion and protein-ligand interactions with partial unfolding cause AMP domain motion in coarse-grained and nonlinear NMA models. LID domain motion always precedes the AMP domain motion in both opening and closing events. AMP motion is induced by ligand binding and high stress in that region is an evidence for local unfolding. After the LID is closed, the AMP domain begins closing due to enthalpic interactions in the LID-AMP interface and phosphoryl transfer occurs. In the reverse pathway, fluctuations cause the LID domain to open, while the instability in LID-AMP interface causes opening of the AMP domain (Whitford *et al*., 2008).

The transition of adenylate kinase was investigated with a new molecular dynamics method named as Dynamic Importance Sampling (DIMS). The method simulates macromolecules which have known stable transition endpoints (structures)

and enhances path sampling by applying a suitable bias. Bias selects and simulates only the successful trajectories that are able to overcome the energy barrier and complete the transition. Thus, computational time is decreased by preventing expendable time near the free-energy minimum of the stable state and only the transition is simulated. Transitions obtained from the simulation are compared with experimental (crystal) structures available for adenylate kinase. Simulation results show in atomic detail that hinge regions partially unfold during conformational change. Structure loss in helices for five hinges (1, 2, 3, 5 and 7) was observed out of experimentally determined eight hinge regions (Beckstein *et al*., 2009). These are listed in Table 2.1 below (Henzler-Wildman *et al*., 2007).

In another study, adenylate kinase is simulated using the Plastic Network Model (PNM) where intermediate structures are also known with two end structures in order to obtain the minimum energy pathway. Residues 119-123 and 167 are common high strained regions with other studies and the loop between first helix and sheet regions (residues 8-12) is also observed in the study. Dihedral angle change in the transition correlates very well with strain energy distribution (Maragakis and Karplus, 2005).

Table 2.1. Experimentally identified hinge regions of AdK.

Hinge	**Residues**
1	28-32
2	49-51
3	58-63
4	74-81
5	112-113
6	119-121
7	157-163
8	164-172

3. MATERIALS AND METHODS

3.1. Elastic Network Models

The elastic network model (ENM), originally proposed in atomistic form by Tirion (1996), is an important method for exploring collective dynamics and large-scale conformational changes of biological macromolecules. Since all atomistic approaches become computationally inefficient with the increasing size of the protein, low-resolution ENMs are considered as effective tools for describing the collective motions of large proteins and supramolecular structures.

In ENM, the 3D native structure of a protein is described by a set of nodes that are connected with elastic springs. Each node generally represents an alpha-carbon atom on the backbone and the node pairs that are apart by less than a cutoff distance are connected with harmonic springs in order to form the elastic network. The Gaussian network model (GNM) (Bahar *et al*., 1997; Haliloglu *et al*., 1997) is the simplest model that assumes isotropic and Gaussian fluctuations. As an extension, the anisotropic network model (ANM) (Doruker et al., 2000; Atilgan et al. 2001), considers that the node fluctuations are anisotropic in x, y and z directions.

In ANM, the harmonic potential of a single spring is described by the formula given below:

where $\Delta \boldsymbol{R}_i$ and $\Delta \boldsymbol{R}_j$ represent fluctuations of position vectors for nodes i and j respectively and is the initial position vector between sites i and j. Vectors can be expanded as the difference in three coordinates ΔX, ΔY and ΔZ. γ represents the force constant for the springs in the network. The total potential energy is, thus, a summation

over all harmonic interactions connecting close-neighboring node pairs that fall within the cutoff distance.

In Equation 3.2, h is zero unless distance of node pairs is less than the cutoff distance (r_c); otherwise, h=1. R_{ij} is the distance between nodes i and j. Second derivative of the potential with respect to each component gives the total force acting on each node. These terms are organized in a (3N×3N) matrix called Hessian matrix.

where each ijth element ($i \neq j$) has the form:

H_{ii} elements on and off the diagonal are respectively given by the following relationships:

Eigenvalue decomposition of the Hessian matrix provides ($3N_{res}$ – 6) non-zero eigenvalues and eigenvectors, which represent the respective frequencies and shapes of the individual modes of motion. In general, the low-frequency modes correspond to the biologically relevant collective deformations that may also correlate closely with the conformational transition direction. As a result, hybrid methodologies incorporating ANM have been developed to study conformational transition pathways, such as the so-called ANM-MC methodology that will be used in this thesis and introduced next.

3.2. ANM-MC Simulation Technique

Haliloglu and Bahar (1998) have developed a coarse-grained Monte Carlo (MC) simulation method for the conformational search of proteins. In this MC/Metropolis algorithm, each amino-acid residue (except Glycine) is represented with two nodes including α-carbon atom of the backbone and the mass center of side chain atoms. The coarse-grained energy function is a knowledge-based potential derived from crystal structures of proteins (Bahar and Jernigan, 1997). New conformations in the conformational space are generated by giving local perturbations to randomly chosen nodes (Haliloglu and Bahar, 1998; Kurt and Haliloglu, 1999; Haliloglu, 1999).

The ANM has been recently integrated with the coarse-grained MC simulation to explore conformational transition pathways by targeting the structure towards the closed conformation using collective modes from ANM. This computationally efficient technique utilizes an iterative procedure, where a new conformation is generated by deforming the structure along the slowest mode that correlates most (overlaps) with the experimentally determined transition direction between open and closed protein structures. The collective deformation (deformation factor, DF) performed in each cycle (iteration) is determined to be suitable using an overall RMSD of DF = 0.2 Å. Energy of the new structure is then minimized by MC simulation using random local moves and the resulting structure replaces the initial structure for the new iteration (Kantarci-Carsibasi *et al.*, 2008). The optimum number of local random moves (MC steps) performed for energy minimization in each cycle has been determined as MCS = 20 in previous studies and this number will be used in current work also.

ANM-MC simulations have been successful for generating biased conformational pathways between two conformational states (Kantarci-Carsibasi *et al*., 2008). The results have been consistent with the available pathway data in literature for adenylate kinase and hemoglobin. With this approach, it may be possible to efficiently simulate larger biological systems, such as supramolecular assemblies.

Figure 3.1 shows the flowchart of the original ANM-MC simulation technique (blue arrows) that produces biased (targeted) transition pathways, i.e. trajectories, between two predefined structures. ANM-MC iterations continue until a desired RMSD value to the target structure is reached. First, a steady decrease in RMSD is observed towards target state, then a plateau is reached, where the collective moves do not deform the structure further towards the target due to low overlap values.

In this thesis, the unbiased (random) version of the ANM-MC algorithm (red arrows) has been developed. As the name suggests, a collective mode (among the slowest five ANM modes) is randomly chosen for perturbation of the structure, instead of selecting the mode with the highest overlap to target direction. To perform efficient sampling of rare transition events between two states, the WE technique is integrated to the ANM-MC algorithm, as will be described in detail in the next section.

3.3. Weighted Ensemble (WE) Method

The WE method is used for exploring many rare events, such as the conformational transitions, the diffusion effects in binding and folding of proteins. The WE procedure stands out with different advantages among path sampling algorithms. The method is easy to apply and gives statistically correct, distinct and unbiased pathways. It also gives the reaction rate of transition from reactant to product state by calculating the probability of successful trajectory over unit time spent during the simulation. Furthermore, the WE method can analyze a wide area of stochastic dynamics in terms of prediction of future states from the trajectory history including both Markovian, where the next structure depends only on the last structure, and non-Markovian processes where the history is taken into account. Sampling can be achieved

even without the target information to explore various configurations in the evolution of the system (Zhang *et al*., 2007; Zhang *et al*., 2010).

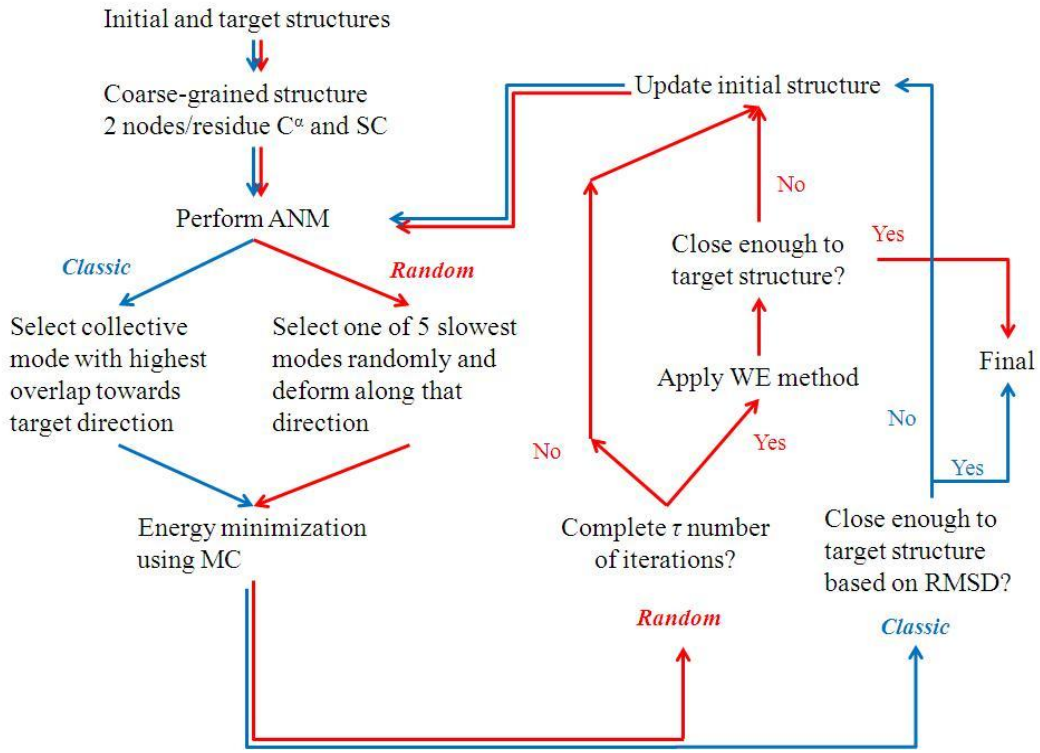


Figure 3.1. Flowchart of classic and random ANM-MC algorithms.

Many biochemical reactions are diffusion-limited, because of the infrequent reaction events due to lack of proper orientation between reactants. The WE method has been originally proposed and applied by Huber and Kim (1996) in conjunction with Brownian dynamics simulations to overcome the sampling problem of such rare events. The WE method has also been applied to give unbiased ensemble of paths for conformational transition of proteins (Bhatt and Zuckerman, 2010), where there is no guiding force or even any target information through the simulation in configurational space.

In the WE method, the reaction (progress) coordinate that shows the progress starting from the reactant state to the product state needs to be defined first. For many cases, progress coordinate with a single dimension is fairly sufficient, even though saddle points in free energy might be on the coordinate chosen and disrupt the sampling

procedure (Zhang *et al*., 2007). Then, the reaction coordinate is divided into small partitions, called bins, to determine the location of each simulation along the coordinate. Number of these bins may be static or changed with time during the simulation. It has been shown that varying binning does not affect the correctness of the WE method (Zhang *et al*., 2010).

In the case of static bins, where N_{bin} is the total number of bins along the reaction coordinate, each occupied bin should accommodate M trajectories that run parallel. The method starts with M independent trajectories with the initial conformation in the first bin. Each trajectory carrying a weight of 1/M is simulated for a fixed time interval composed of N_τ iterations or collective moves. At the end of each time interval N_τ, all bins are checked to determine the new distribution of trajectories along the progress coordinate. As expected, some particles might have moved forward or backwards during N_τ. Therefore, the trajectories with the lowest (or the highest) probabilities should be combined (or split) if there are more (or less) than M simulations in a specific bin (Huber and Kim, 1996; Zhang *et al*., 2007; Bhatt and Zuckerman, 2010). These adjustments can be applied by considering the rules explained in Huber and Kim's article (1996). The aim is to maintain the number of trajectories and probabilities roughly the same in each bin.

Schematic representation of the WE method is shown in Figure 3.2 for a sample system with two bins (N_{bin}=2) and two simulations per bin (M=2). There are two simulations carrying probabilities of one half in the first bin (at t = 0) that add up to a total probability of one. One of the simulations moves to second bin after a time period of dynamics (at t = τ), which represents the number of ANM-MC iterations or cycles (N_τ) in this study. Since there should be two simulations in each bin, each simulation and its probability is split into two to give a probability of one quarter each. This procedure ensures that at any stage the total probability of all simulations adds up to one. After the second step of dynamics (at t = 2τ), results show that one simulation moves back to the first bin. Three simulations in first bin and one simulation in the second bin require combination and splitting adjustments, respectively. Since there are more than two simulations (three) in the first bin, they should be combined so that the

combined simulation should have a probability equal to their sum (one half) and an average conformation based on their probabilities. In the second bin, one simulation has to split into two identical simulations, each carrying half of the main simulation probability (half of a quarter). Dynamics and rearrangements continue for all the simulations inside the bins. It is crucial to determine the total number of bins (N_{bin}) and simulations in each bin (M) and length of the cycles (N_{τ}), so as to keep the computational time at an acceptable level (Huber and Kim, 1996).

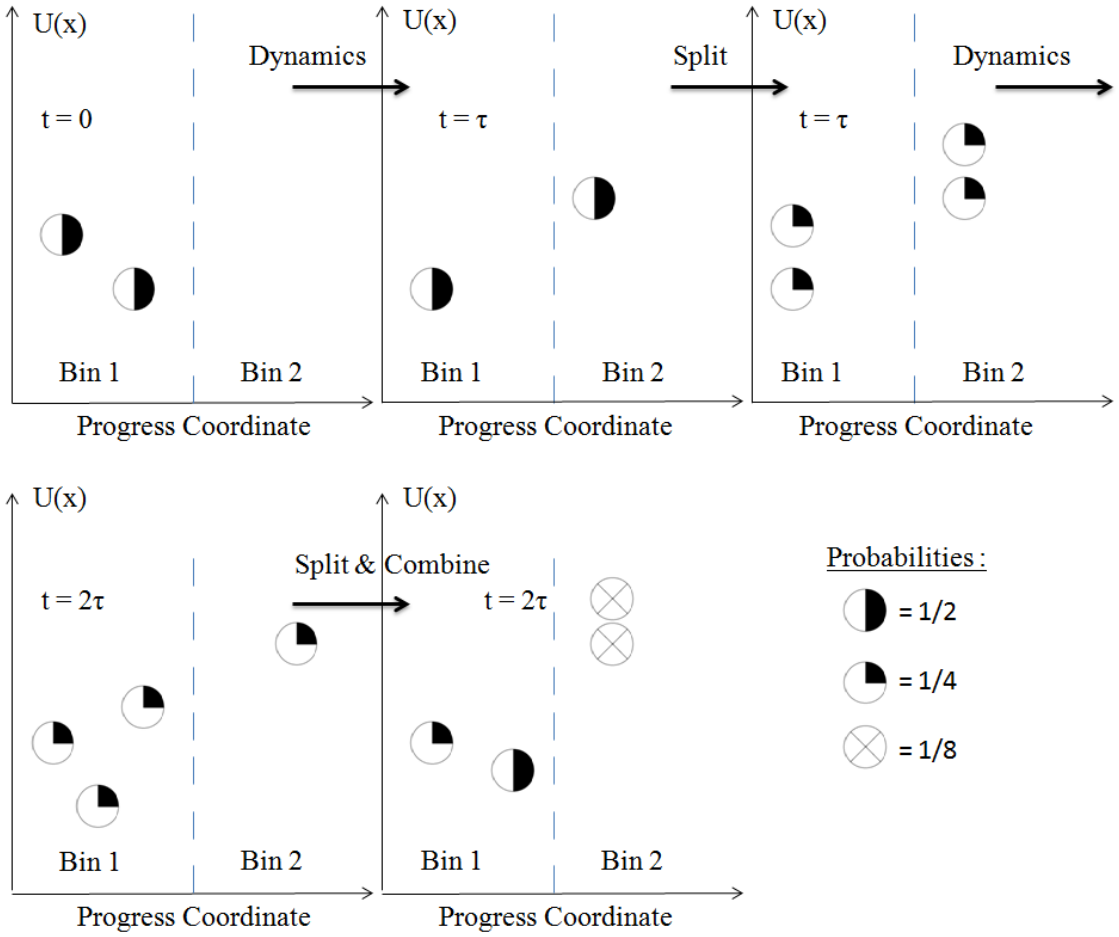


Figure 3.2. Schematic representation of the WE method.

3.4. Proteins Studied

The unbiased ANM-MC methodology using the WE algorithm is applied to three different proteins in this thesis, namely adenylate kinase (AdK), calmodulin (CaM) and glutamine binding protein (GBP). All three proteins exhibit hinge-bending type of

motion during their conformational transitions. Table 3.1 lists the Protein Data Bank (PDB) (Berman *et al*., 2000) codes of the initial and target crystal structures, the total numbers of residues, the root mean square deviation (RMSD) value between initial and target structures.

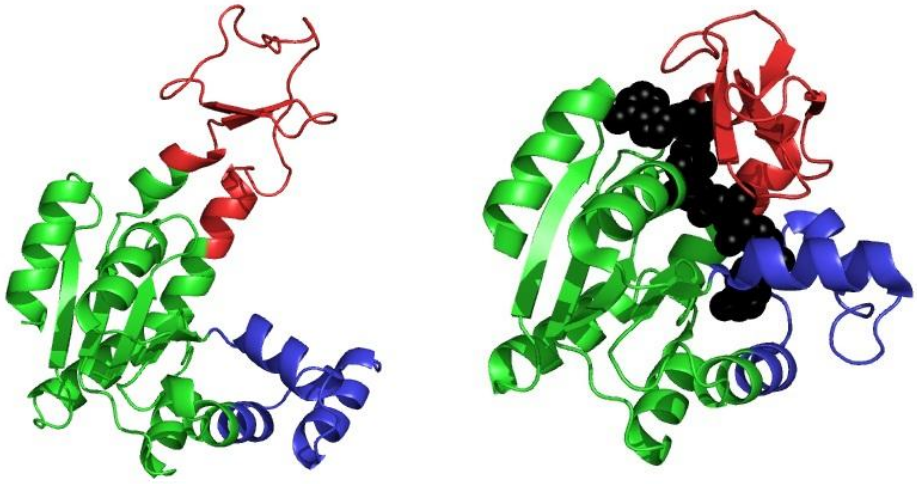
Table 3.1. Proteins studied by unbiased ANM-MC simulations.

Protein Name	**Initial PDB**	**Target PDB**	**Number of Residues (N_{res})**	**RMSD (Å)**
Adenylate Kinase	4AKE	1AKE	214	7.13
Calmodulin	1CLL	1A29	142	15.01
Glutamine Binding Protein	1GGG	1WDN	220	5.34

3.4.1. Adenylate Kinase (AdK)

AdK with 214 residues is composed of three domains, which are the ATP binding domain, commonly called the LID domain (residues 118–167), the AMP binding domain (residues 30–67), and the comparatively static CORE domain (residues 1–29, 68–117 and 161–214). The enzyme catalyzes the reversible reaction (Mg^{2+}·ADP + ADP ↔ Mg^{2+}·ATP + AMP) of phosphoryl transfer between ATP and AMP for regulating cellular energy. The apo/open (PDB code: 4AKE) and bound/closed (PDB code: 1AKE) crystal structures of AdK (only A chain in both cases) are shown in Figure 3.3 with three domains colored individually. The inhibitor diadenosine pentaphosphate (AP_5A) in the closed structure is shown with black spheres (Müller *et al.*, 1996; Müller and Schulz, 1992).

LID (red) and AMP-bind (blue) domains close over the static CORE domain (green) during the open-to-closed (apo-to-bound) transition and complete the hinge bending motion. The RMSD value between open and closed structures is 7.13 Å.



(a) (b)

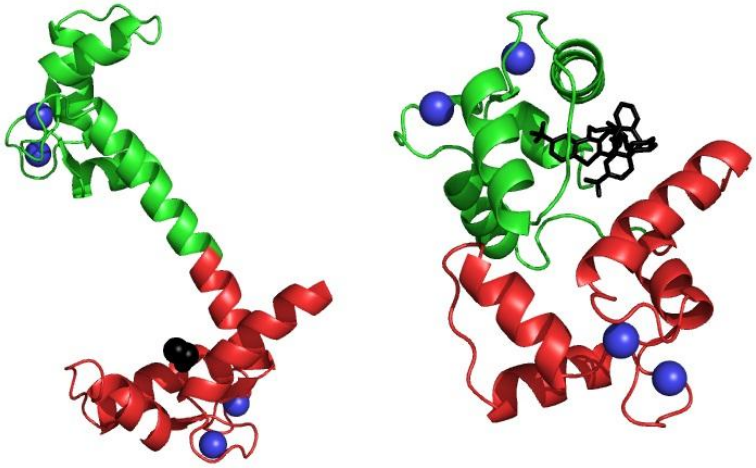
Figure 3.3. Crystal structures of adenylate kinase (a) Open, (b) Closed.

There are some studies about explaining the order of domain motions and its reasons in the literature. For instance, Whitford et al. (2007) pointed out that only LID domain closed intermediate structure is observed in contrast to elastic network models. Whitford *et al*. (2008) found out that LID domain motion always precedes AMP domain motion during both opening and closing procedures because of the higher intrinsic effects and lower energy barriers in the LID domain. An unbiased atomistic MD study showed no clear path starting from open conformation; however, for transition starting from the closed state, opening of LID and AMP domains happens after half opening of AMP domain first (Kubitzki and de Groot, 2008).

3.4.2. Calmodulin (CaM)

Calmodulin takes part in many important functions in the cell, such as transducing calcium signals to other organelles and storing calcium by binding to calcium ions. It is composed of 146 residues and two domains. Each domain consists of two helices with calcium binding loops. Central linker is a long helix that is unusually flexible. As shown in Figure 3.4, the large hinge-bending motion is observed with C-terminal domain (residues 78-146, green) closing over N-terminal domain (residues 5-77, red). The RMSD value between the open and closed structures is 15.01 Å.

The initial (open) structure (PDB code: 1CLL) has Ca^{+2} ions (blue spheres) and ethanol (EOH) (black spheres) bound, while trifluoperazine (black sticks) and Ca^{+2} ions (blue spheres) bind to form the closed structure (PDB code: 1A29) (Chattopadhyaya *et al*., 1992; Vertessy *et al*., 1998) First three residues are not available in the crystal structures and the fourth residue has no side chain atoms in the initial conformation. So, the remaining 142 residues of calmodulin structures are used in the ANM-MC simulations.



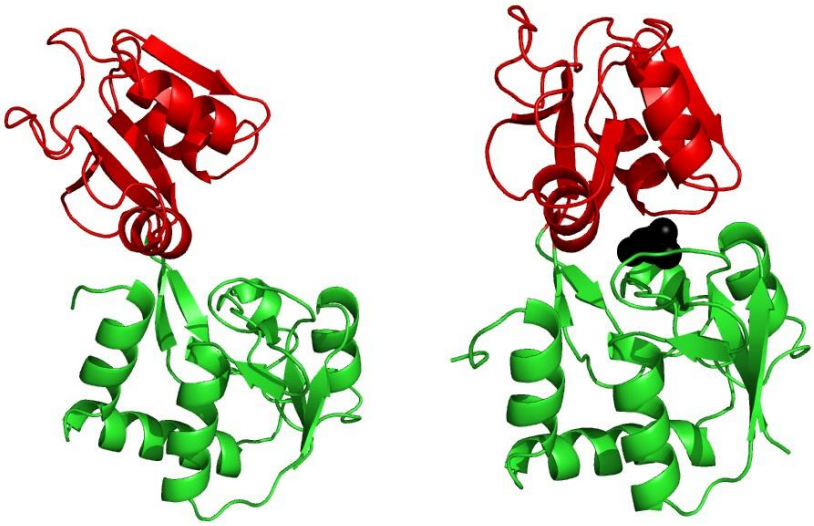
(a) (b)

Figure 3.4. Crystal structures of calmodulin (a) Open, (b) Closed.

3.4.3. Glutamine Binding Protein (GBP)

Being the smallest periplasmic binding protein determined by X-ray diffraction, GBP has two domains and a total number of 220 residues. Ligand-specific periplasmic binding proteins have two similar globular domains connected by two or three short flexible peptide hinges along with α/β secondary structural topologies. A ligand-induced conformational change is observed after the binding of the ligand with high affinity. Substrates are trapped between the two domains and transferred across the membrane through the interactions of protein-ligand complex and membrane-bound structures (Sun *et al*., 1998; Hsiao *et al*., 1996).

Figure 3.5 displays the large domain (green: residues 1-84, 186-226) and the small domains (red: residues 90-180) in unbound (open) (PDB code: 1GGG-A chain) and bound (closed) (PDB code: 1WDN-A chain) crystal structures of GBP. The ligand glutamine (Gln, black spheres) is also shown in the bound structure. Conformational change from open to closed states due to hinge bending motion of GBP has an initial RMSD value of 5.34 Å.



(a) (b)

Figure 3.5. Crystal structures of GBP (a) Open, (b) Closed.

4. RESULTS AND DISCUSSION

In this thesis, ANM-MC simulation with random collective moves is applied to three different proteins (AdK, CaM and GBP) by imposing the principles of the weighted ensemble method. Totally, 13 runs are simulated in order to investigate the effects of different parameters, such as the cutoff distance (r_c) in ANM, the number of collective ANM moves (or iterations) performed in each fixed time interval (N_τ), and the number of parallel simulations per bin (M). Table 4.1 gives a list of completed runs and the corresponding simulation parameters. Also the best (minimum) final RMSD to target structure among all the pathways generated. The number of local random perturbations given to the backbone or side chain nodes is taken as MCS = 20 in every run.

The conformational space (or the reaction coordinate) based on the RMSD between initial and final crystal structures is divided into (N_{bin}) 50, 75 and 30 bins for AdK, CaM and GBP, respectively, to give approximately the same bin sizes. The respective RMSDs between initial and target structures are 7.13, 15.01 and 5.34 Å for these proteins. In AdK_1 and AdK_2 runs among the nine runs held for AdK, the number of simulations per bin (M) is the only parameter changed (r_c = 18 Å). It will provide enrichment in sampling as the amount of total trajectories is doubled by increasing it from five to ten simulations. The next four runs (AdK_3, AdK_4, AdK_5 and AdK_6) employ a cutoff distance of 10 Å. In AdK_5 and AdK_6, the number of iterations used for the WE method (splitting and combining of parallel runs) is doubled (N_τ = 10), which is applied as N_τ = 5 in all other systems. AdK_R1, AdK_R2 and AdK_R3 search for the reverse conformational transition of AdK, where the cutoff distance is 18 Å in AdK_R1 and 10 Å for the rest. Other four runs are the forward simulations of CaM and GBP with two different cutoff values employed for each protein.

4.1. AdK Forward Runs with 18 Å Cutoff Distance

The results of the first two AdK runs, both of which use an ANM cutoff of 18 Å, will be analyzed in this section. AdK_1 is the only run with five parallel simulations per bin (M=5). This run approached the target structure with a final RMSD of 2.87 Å, corresponding to the 30^{th} bin. AdK_2 with the same parameters, but employing 10 runs per bin (M=10), reached the target down to an RMSD of 2.78 Å within the 31^{th} bin.

Figure 4.1 shows the RMSD profiles for a total number of 150 AdK_2 trajectories. As expected only some of these trajectories move towards the target structure, as a result of enhanced sampling due to the WE methodology. Figure 4.2a and 4.3a display all the 'successful' trajectories for AdK_1 and AdK_2, respectively. Only those trajectories with a final RMSD below 3.2 Å are thus displayed separately for clarity.

Among these 'successful' trajectories, some are chosen as distinct pathways for further analysis in terms of conformational transitions and cracking phenomena. In general for all the systems studied in this thesis, relatively early breakpoints in the RMSD profiles, where one path is divided into several pathways, are selected as distinct pathways. The pathway that has the minimum final RMSD value to target is always chosen to be one of the distinct pathways. Distinct pathways are then numbered from one to a maximum number of four.

Two such distinct pathways are determined for AdK_1, as indicated in Figure 4.2b. In the same manner, AdK_2 trajectories that have final RMSD lower than 3.2 Å are shown in Figure 4.3a, where four of them are chosen as significant pathways in Figure 4.3b. In both cases, Pathway 1 (red) corresponds to the minimum RMSD case.

Table 4.1. List of runs simulated with random ANM-MC method.

System	Cutoff Distance (r_c)	Total Bins (N_{bin})	Runs per Bin (M)	Iterations for WE method (N_τ)	Final RMSD to target (Å)
AdK_1	18	50	5	5	2.87
AdK_2	18	50	10	5	2.78
AdK_3	10	50	10	5	2.00
AdK_4	10	50	10	5	1.77
AdK_5	10	50	10	10	1.93
AdK_6	10	50	10	10	1.94
AdK_R1	18	50	10	5	5.48
AdK_R2	10	50	10	5	4.18
AdK_R3	10	50	10	5	4.32
CaM_1	18	75	10	5	3.44
CaM_2	10	75	10	5	4.62
GBP_1	18	30	10	5	2.02
GBP_2	10	30	10	5	1.92

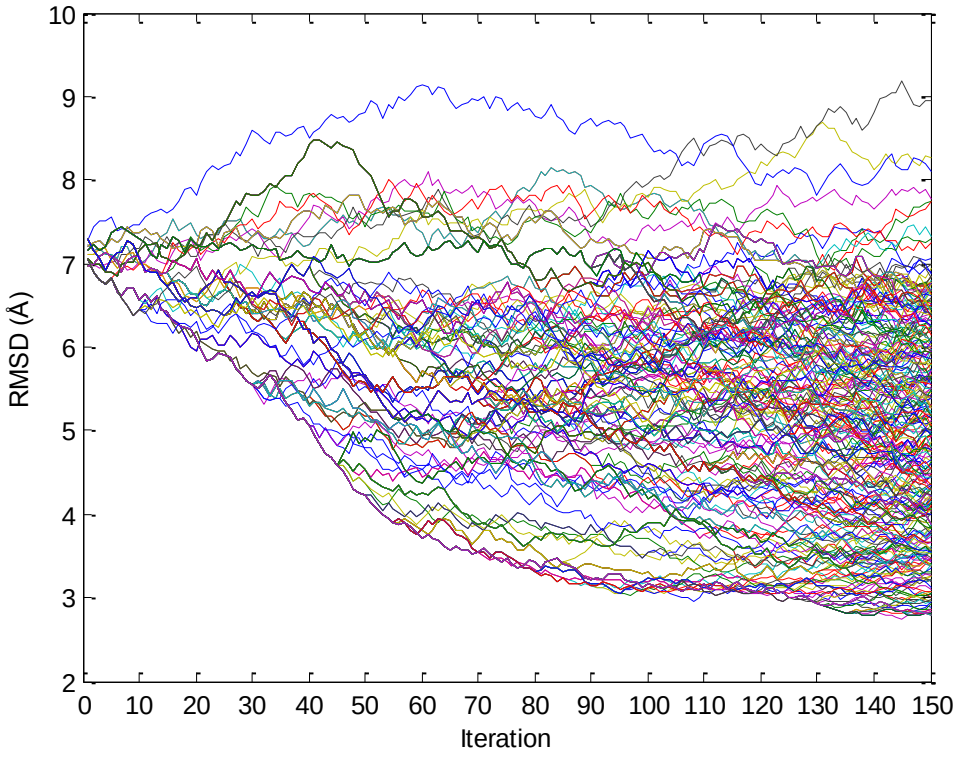


Figure 4.1. RMSD profiles of all trajectories for the AdK_2 system.

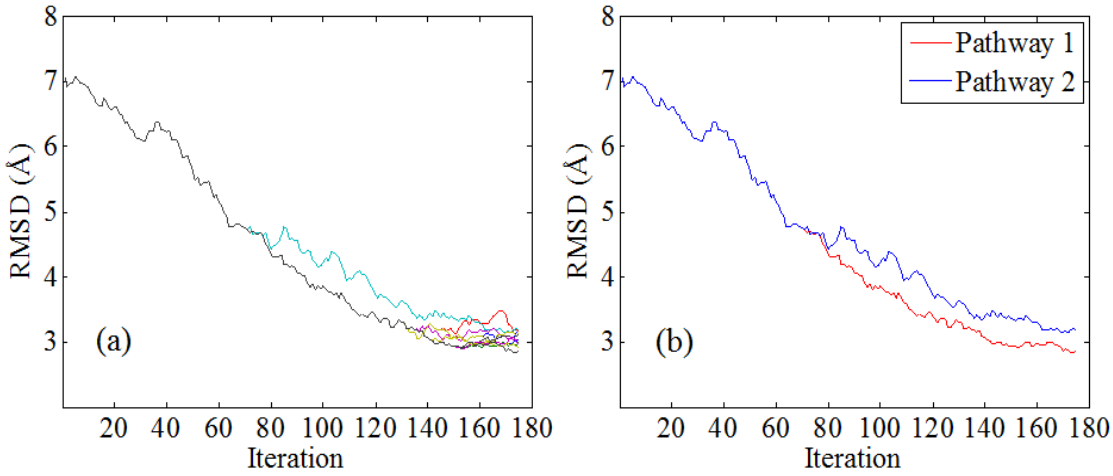


Figure 4.2. AdK_1 RMSD profiles in forward direction (a) All trajectories that reach a final RMSD below 3.2 Å, (b) Two main pathways for AdK_1 system that get close to target.

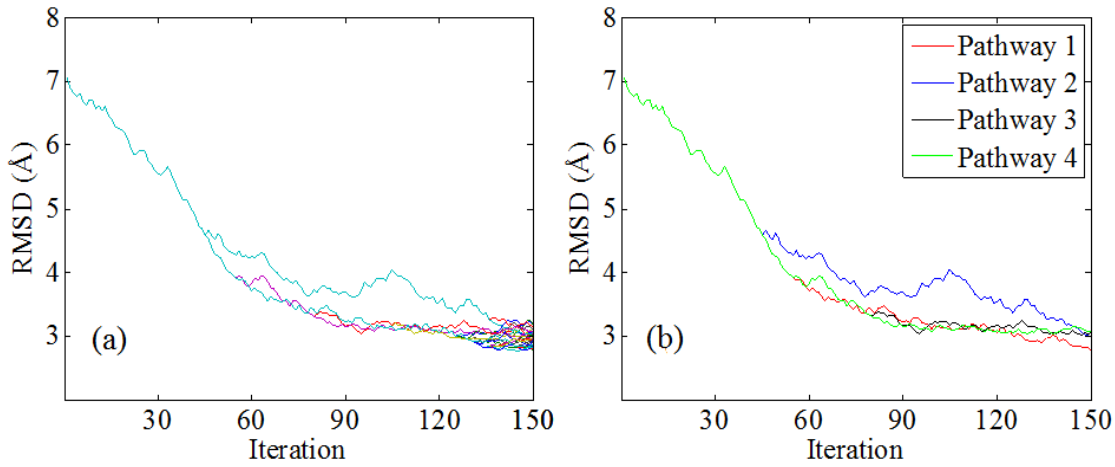


Figure 4.3. AdK_2 RMSD profiles in forward direction (a) All trajectories that reach a final RMSD below 3.2 Å, (b) Four main pathways for AdK_2 system that get close to target.

Further analysis is performed on the main pathways of AdK_1 and AdK_2 systems in terms of inter-domain distances and angles that were defined in previous studies to investigate the conformational transitions. Three residue pairs for AdK, called FRET pair distances, are considered as a measure of the movement of LID, CORE and AMP domains with respect to each other in experimental and computational studies (Beckstein *et al*., 2009). These LID-CORE (ALA127-ALA194), AMP-CORE (ALA55-VAL169) and LID-AMP (ILE52-LYS145) inter-domain (C^{α}-C^{α}) distances in the initial, open structure are measured as 38.77 Å, 29.50 Å and 43.56 Å, respectively, and shown in Figure 4.4. LID-CORE and AMP-CORE distances are tracked throughout the simulations to observe the closure of LID and AMP domains. Figure 4.5a and 4.5b show LID-CORE and AMP-CORE distance transitions for each pathway with respect to iteration number for AdK_1 and AdK_2 systems. Red line represents first pathway where blue is for second, black for third and green line for fourth pathways.

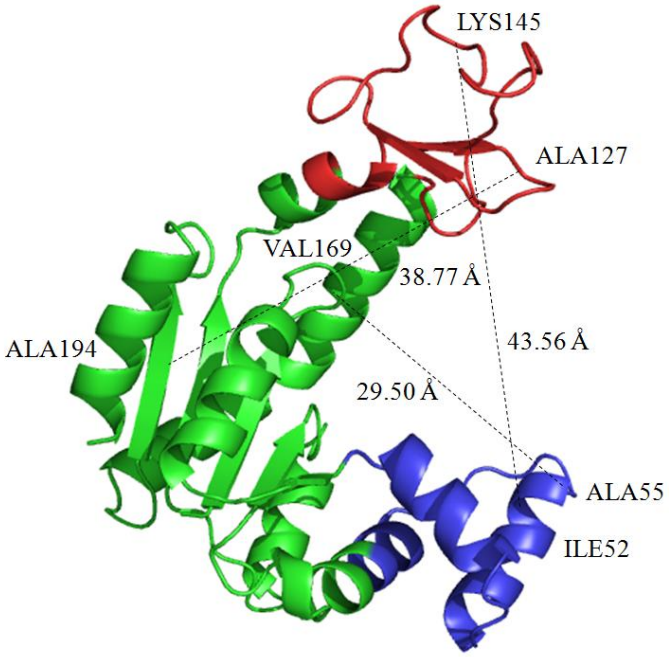


Figure 4.4. FRET residue pairs and initial distances shown on the open crystal structure of AdK.

The same FRET distances of LID-CORE and AMP-CORE domains are plotted versus each other in Figure 4.6a and 4.6b for AdK_1 and AdK_2 systems, respectively. The location of open (circle) and closed (triangle) crystal structures are also indicated on the plots. The sequence of closure motion of LID and AMP domains are observed to be quite different in these pathways. The LID and AMP domains move simultaneously towards the CORE in both pathways of AdK_1 system. In contrast, AdK_2 main pathways exhibit more diverse character in terms of the timing of domain closure. For example, LID domain closure is more dominant in Pathway 2 (blue), while Pathway 3 (black) takes more part in AMP domain closure with the LID-CORE distance staying nearly constant. However, none of the pathways seem to get close enough to the closed/target structure.

Another definition for measuring the transition of AdK towards the closed (bound) state is the domain angles. LID-CORE (θ_{LID}) and AMP-CORE (θ_{NMP}) angles are defined as the angles between relatively static CORE domain and the respective LID and AMP domains. Participating residues for the calculation of the θ_{NMP} angle is the centers of geometry of the backbone and C^{β} atoms located in three regions comprising

residues 115–125 (CORE–LID), 90–100 (CORE) and 35–55 (AMP) of AdK. Similarly, θ_{LID} is calculated by the residues between 179–185 (CORE), 115–125 (CORE–hinge–LID) and 125–153 (LID) (Beckstein *et al.*, 2009). The change in these two angles for each pathway with the angles of initial (shown with blue triangle) and target (shown with black triangle) structures using only C^{α} atoms are represented in Figure 4.7a and 4.7b. The angles for intermediate structures are calculated with using C^{α} atoms.

Figure 4.7a reveals that AdK_1 motion nearly completes the LID-CORE angle transition, while there is still 10°-15° difference to the closed (target) structure in AMP-CORE angle. Figure 4.7b shows that AdK_2 system has evolved slightly closer to the target structure and Pathway 3 (black) brings down the difference to nearly 5° in AMP-CORE angle.

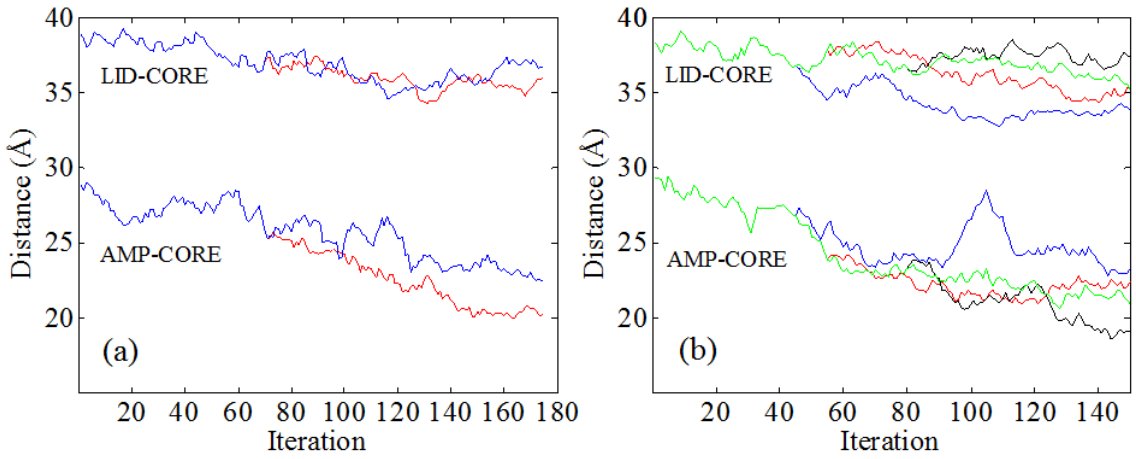


Figure 4.5. FRET distance profiles along the transition for the main pathways of (a) AdK_1 system, (b) AdK_2 system.

In summary, AdK_1 and AdK_2 systems have similar final RMSD values to target (2.87 and 2.78 Å). But AdK_2 reaches faster to the target (looking at RMSD profiles) and produces more distinct pathways by looking at the FRET distances in Figure 4.5 and 4.6. This may result from the higher number of parallel runs in AdK_2 (M=10) as compared to AdK_1 (M=5).

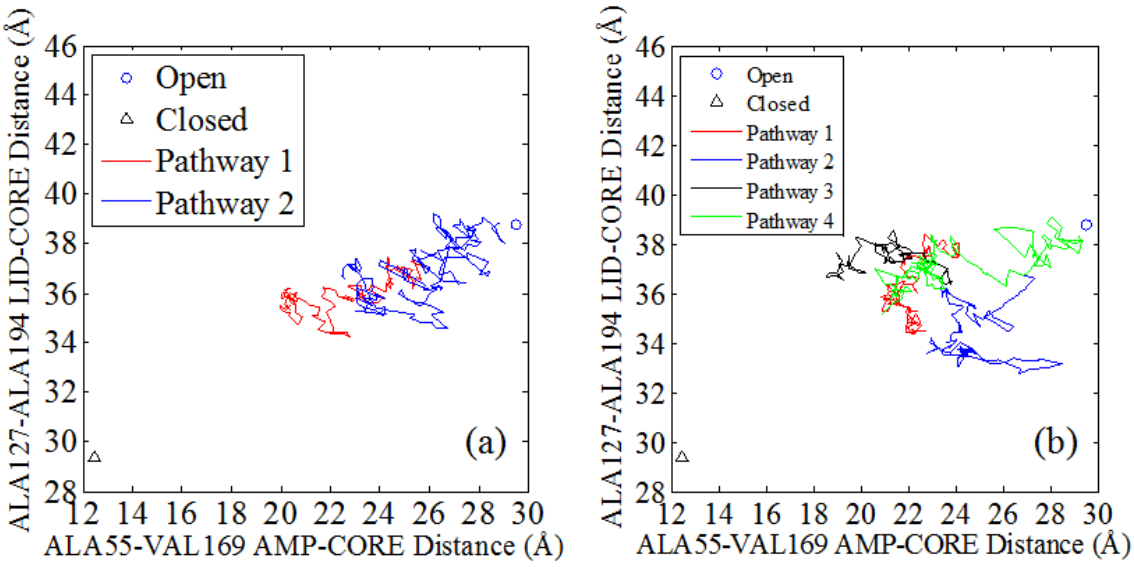


Figure 4.6. FRET pair distances for the main pathways of (a) AdK_1 system, (b) AdK_2 system.

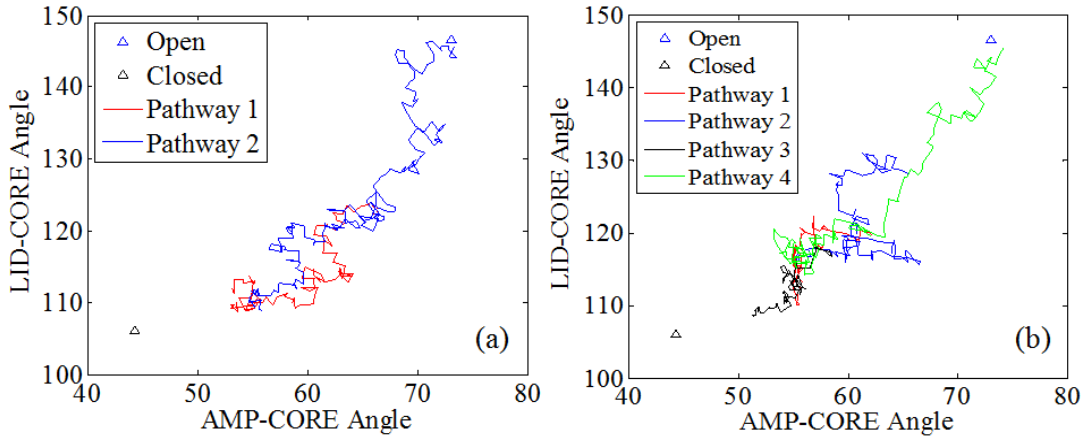


Figure 4.7. LID-CORE (θ_{LID}) and AMP-CORE (θ_{AMP}) angles for the main pathways of (a) AdK_1 system, (b) AdK_2 system.

Thus, based on these two AdK simulations, 10 simulations per bin (M=10) seem satisfactory for effective sampling. Moreover, AdK_2 displays earlier splitting (first and second splitting after 45 and 55 iterations) than AdK_1 system (first split after 70 iterations). Still, there is only half-way domain closure in all of these pathways.

4.2. AdK Forward Runs with 10 Å Cutoff Distance

In this section, random ANM_MC results based on a cutoff distance of 10 Å will be presented so as to observe whether the target may be more closely reached. All simulations with 10 Å cutoff distance employ the same number of simulations per bin (M=10). In AdK_3 and AdK_4, the number of iterations for WE sampling (N_τ) is taken as 5, whereas N_τ is increased to 10 in AdK_5 and AdK_6 systems.

Using a lower cutoff distance of 10 Å definitely produces better results in terms of reaching the target state, as the final RMSD for all runs is reported to be lower than 2 Å in Table 4.1. This is line with our unpublished results using targeted ANM_MC simulations. Best result is achieved in AdK_4 with 1.77 Å. In line with RMSD results, higher-indexed bins (36^{th} bin in AdK_3, 38^{th} bin in AdK_4, and 37^{th} bin in AdK_5 and AdK_6) become populated with 10 Å cutoff distance, in contrast to 30^{th} and 31^{th} bins reached with 18 Å cutoff.

In this case trajectories with a final RMSD lower than 2.5 Å to the target state are considered as successful pathways and shown in Figure 4.8. Distinct pathways are further defined in Figure 4.9. Successful pathways in AdK_3 and AdK_4 exhibit quite similar RMSD profiles, except Pathway 3 in AdK_3. AdK_5 and AdK_6 systems, for which N_τ is increased to 10, both follow similar routes in terms of RMSD profiles towards the target, but an earlier breakpoint is observed in AdK_5.

In the same manner, the inter-domain distances (LID-CORE and AMP-CORE) are calculated for the systems with 10 Å cutoff distances. In Figure 4.10 these distances are plotted with respect to iteration number for all four systems. The colorings of the pathways are consistent with the definitions in previous RMSD figures. LID-CORE distances are plotted versus ANM-CORE distances in Figure 4.11. Similarly, the inter-domain angles, as defined in previous section, are shown in Figure 4.12.

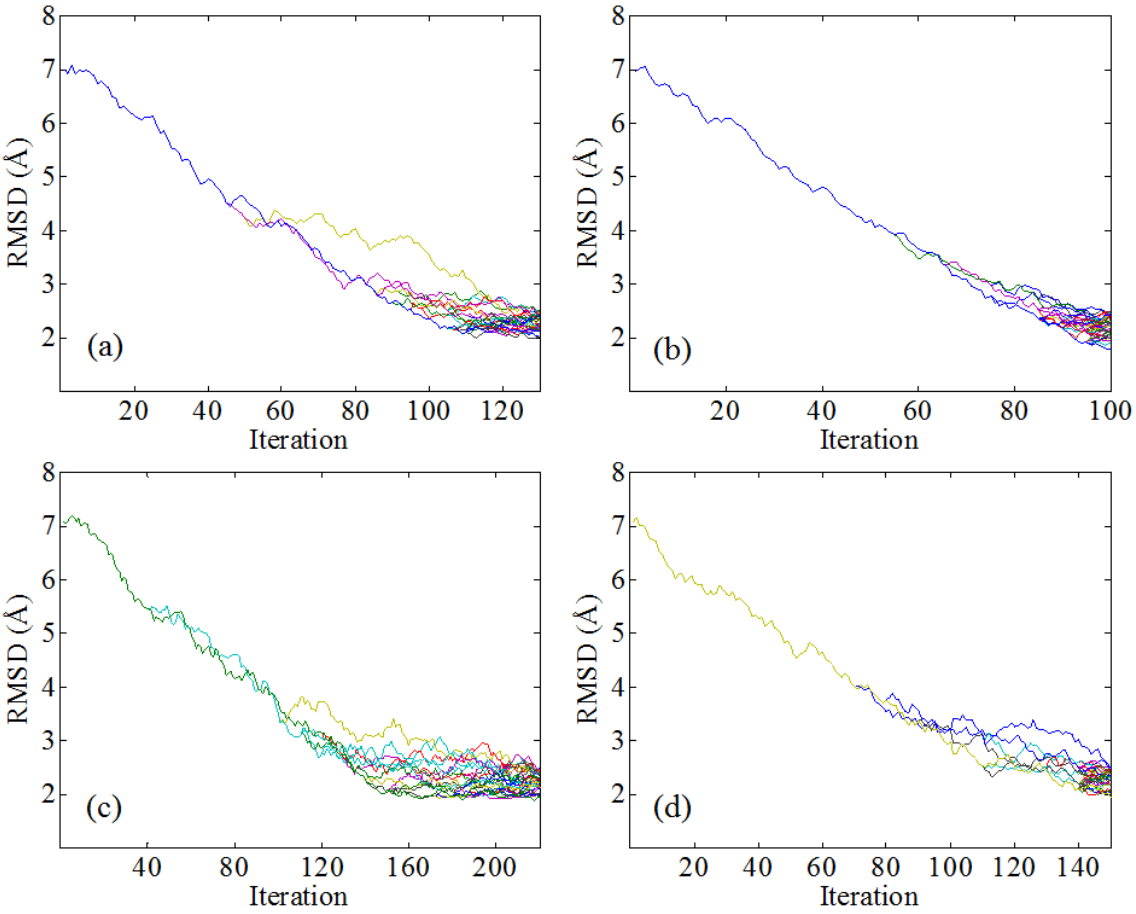


Figure 4.8. Trajectories that reach a final RMSD below 2.5 Å for AdK systems with 10 Å cutoff (a) AdK_3, (b) AdK_4, (c) AdK_5 and (d) AdK_6.

In AdK_3 transition pathways, the LID and AMP domains close together onto the CORE domain by following quite similar routes (see Figure 4.10a and 4.11a). AdK_4 system follows a narrow RMSD band and the pathways are not much different from each other. FRET distance plots in Figure 4.10b and 4.11b show the sequential motions of the two domains in the order of LID half-closure, AMP half-closure and then the rest of LID closure. In Figure 4.12b, the inter-domain angles show simultaneous closure of the two domains following a partial LID closure in the beginning.

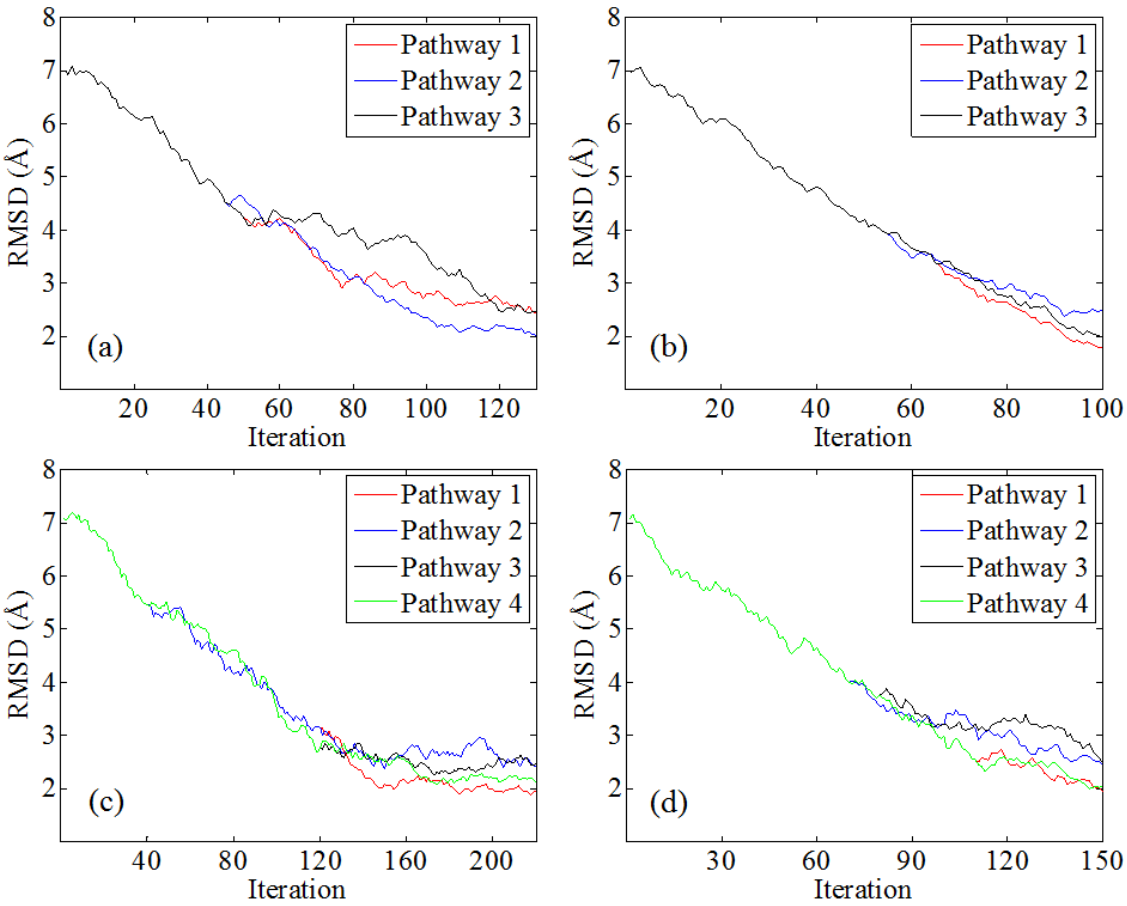


Figure 4.9. Main pathways for AdK systems with 10 Å cutoff (a) AdK_3, (b) AdK_4, (c) AdK_5 and (d) AdK_6.

AdK_5 transition is divided into two distinct routes, which can be seen clearly in Figure 4.11c and 4.12c. Pathways 1 and 2 form the lower route (LID closure first), while Pathways 3 and 4 lie on the upper route (AMP closure first). The two routes rejoin at some point close to the target state. AdK_6 system mainly describes an initial AMP domain closure, during which the LID-CORE distance stays almost constant, and closure of the LID domain in later stages, as shown in Figure 4.10d and 4.11d.

In summary, 10 Å cutoff distance in ANM provides more success in terms of approaching the target state. Still all four systems studied provide a different view for the sequence of domain closure events.

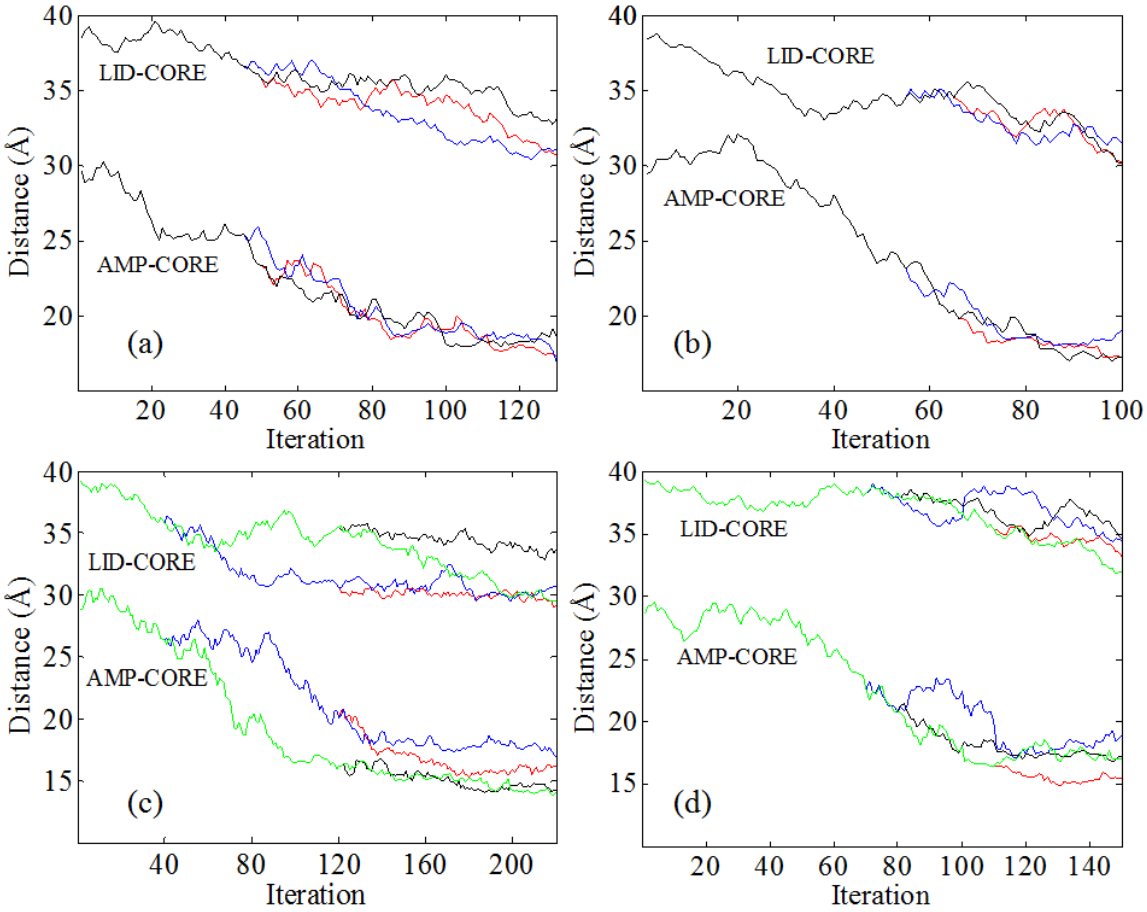


Figure 4.10. FRET distance profiles along the transition for the main pathways of systems (a) AdK_3, (b) AdK_4, (c) AdK_5 and (d) AdK_6.

4.3. AdK Reverse Transition Runs with Cutoff Distances of 18 Å and 10 Å

AdK reverse transition is also investigated to assess how successful the randomly chosen modes are in producing unbiased pathways from the closed to the open state. Three closed-to-open runs are analyzed in this section, namely AdK_R1 with 18 Å cutoff, and AdK_R2 and AdK_R3 with 10 Å cutoff. Starting from the closed structure (PDB code: 1AKE), Figure 4.13 shows the evolution of all AdK_1 trajectories formed by WE method. AdK_R1 system has 5.48 Å final RMSD value to the open structure, while AdK_R2 is 4.18 Å and AdK_R3 is 4.32 Å away from the target, open state (PDBcode: 4AKE). Thus, none of these runs are successful in terms of reaching the target state, in contrast to the case of forward runs. Still distinct pathways are analyzed

if the final RMSD is closer than 5.7 Å for AdK_R1, 4.5 Å for AdK_R2 and AdK_R3 (Figure 4.14-4.16).

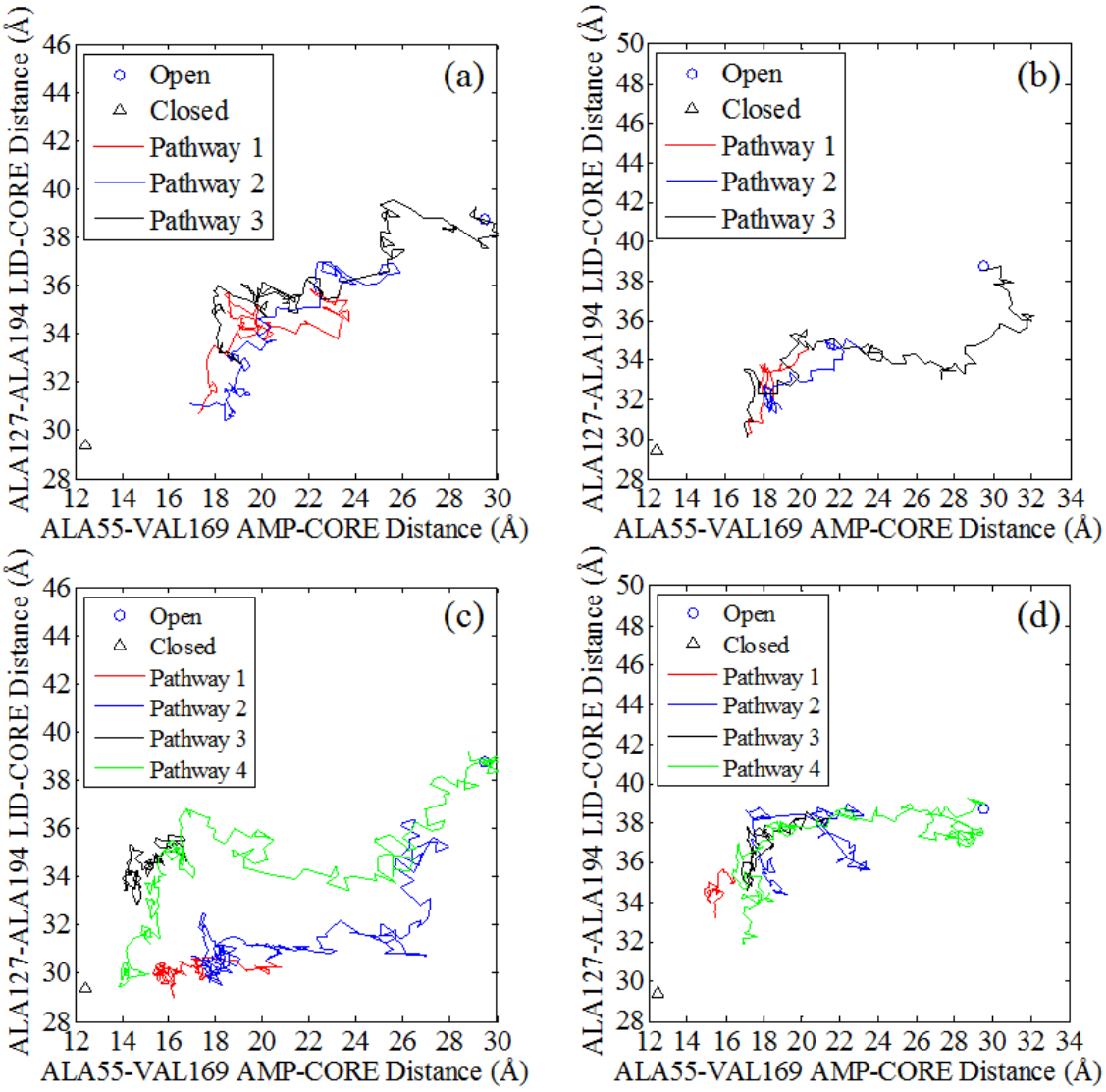


Figure 4.11. FRET distances for the main pathways of systems (a) AdK_3, (b) AdK_4, (c) AdK_5 and (d) AdK_6.

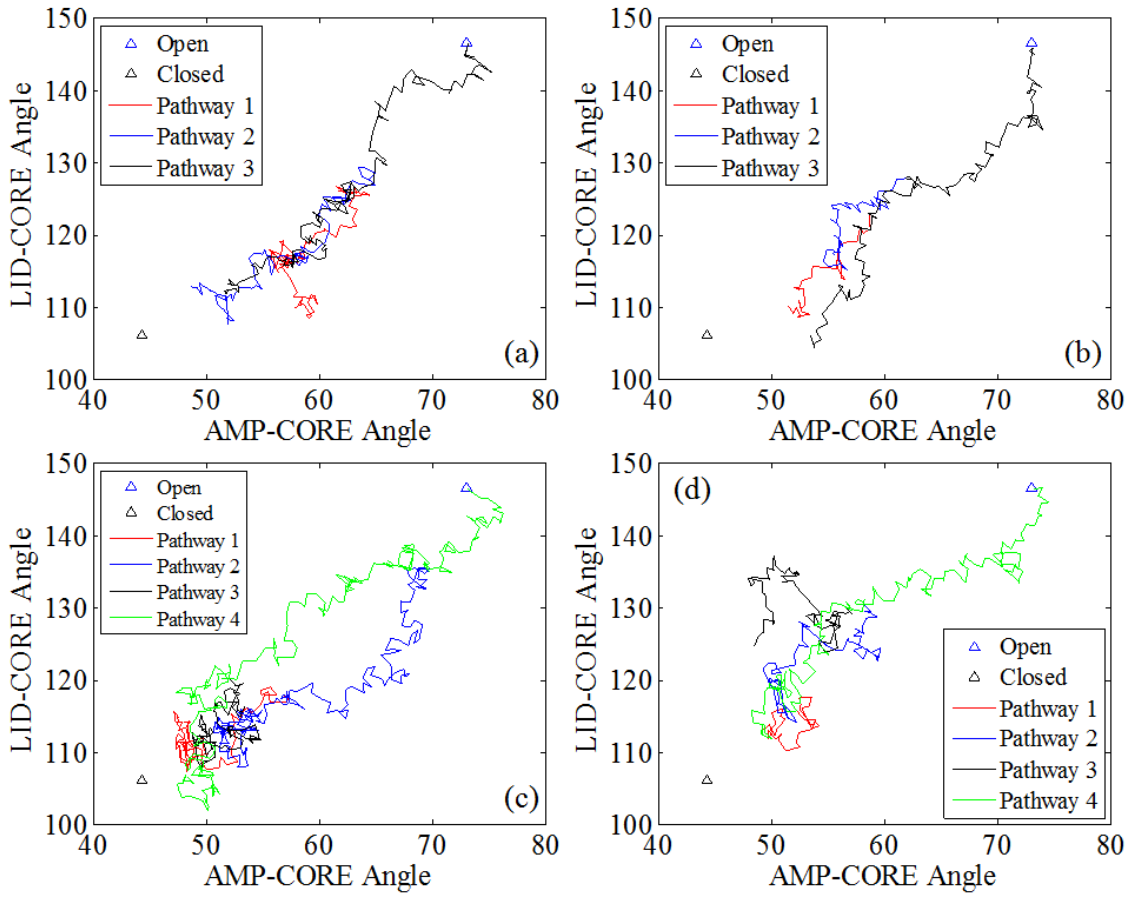


Figure 4.12. LID-CORE (θ_{LID}) and AMP-CORE (θ_{AMP}) angles for the main pathways of systems (a) AdK_3, (b) AdK_4, (c) AdK_5 and (d) AdK_6.

AdK_R1 system (with 18 Å cutoff) does not evolve close enough to target (open) structure within 200 iterations. Four distinct pathways are shown with an RMSD less than 5.7 Å to the target state (Figure 4.14b). Also, Figure 4.17a and 4.18a show that inter-domain distances do not change/decrease significantly throughout the simulation. AdK_R2 and AdK_R3 systems perform slightly better, but still the final RMSDs are 4.18 Å and 4.32 Å, respectively. In both systems, three main pathways are determined and given in Figure 4.15b and 4.16b. Looking at FRET distance pairs, the AMP-CORE distance stays almost constant for these two systems, while Figure 4.17b-4.18b and Figure 4.17c-4.18c show that LID-CORE distance increases by 4 Å in AdK_R2 and by 7 Å in AdK_R3, respectively. Thus, the reverse transition starts with the opening of the LID domain. Domain angles, shown in Figure 4.19, indicate that both LID-CORE (θ_{LID}) and AMP-CORE (θ_{AMP}) angles change during the transitions.

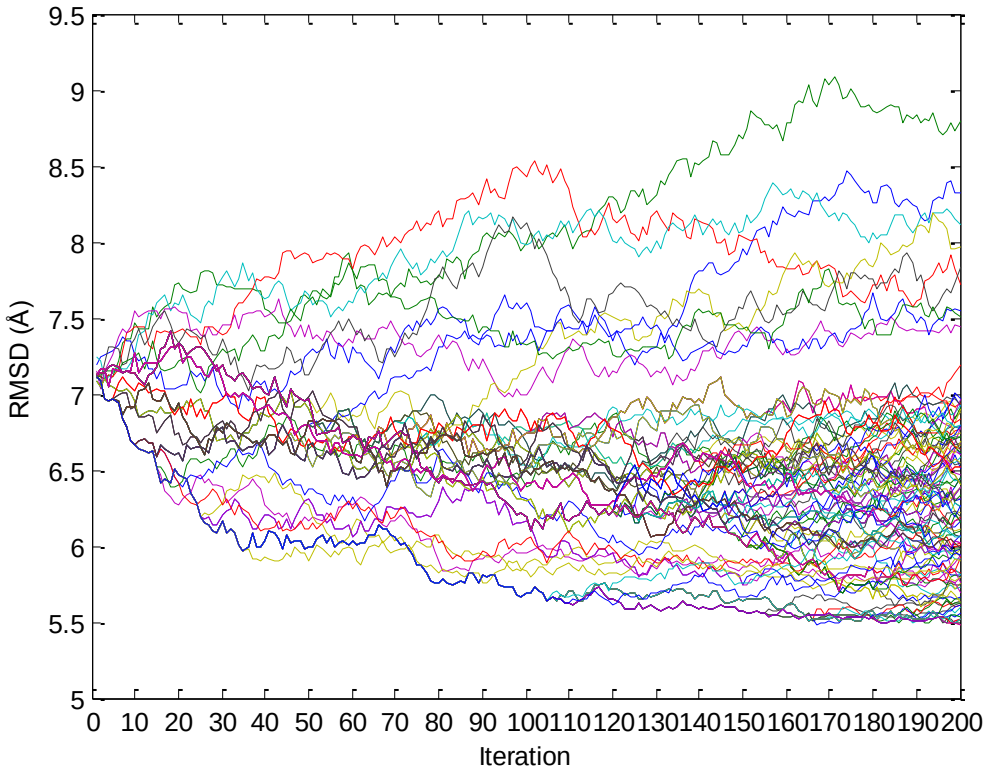


Figure 4.13. RMSD profiles of all trajectories for the AdK_R1 system.

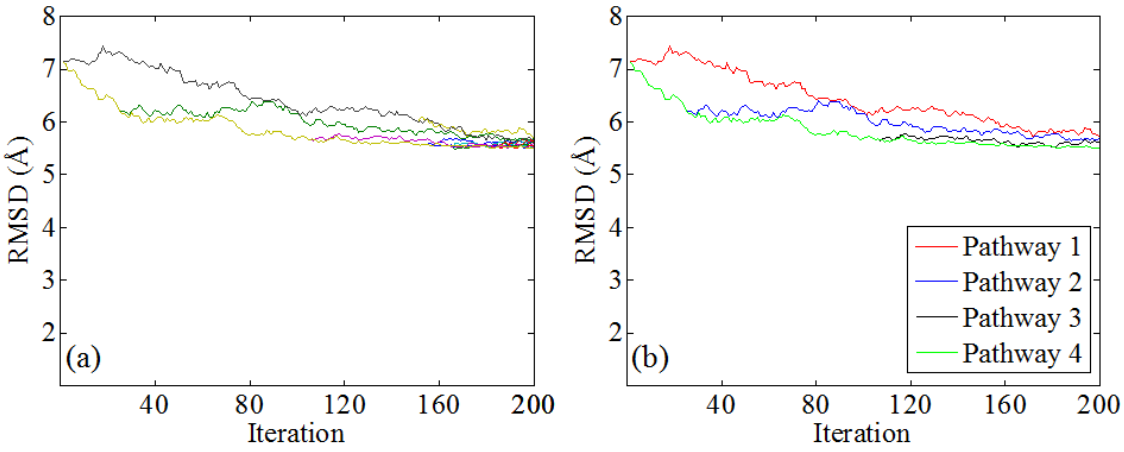


Figure 4.14. AdK_R1 RMSD profiles in reverse direction (a) All trajectories that reach a final RMSD below 5.7 Å, (b) Four main pathways for AdK_R1 system that get close to target.

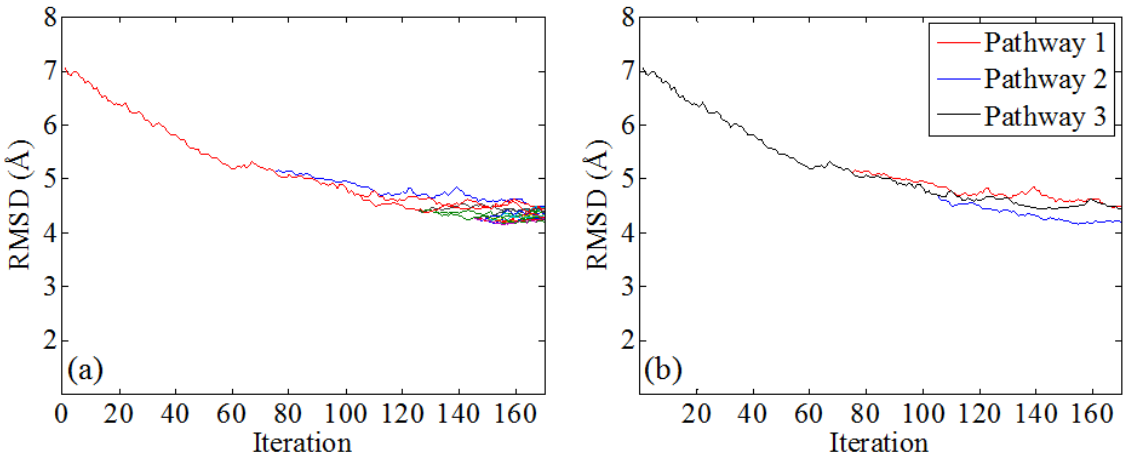


Figure 4.15. AdK_R2 RMSD profiles in reverse direction (a) All trajectories that reach a final RMSD below 4.5 Å, (b) Three main pathways for AdK_1 system that get close to target.

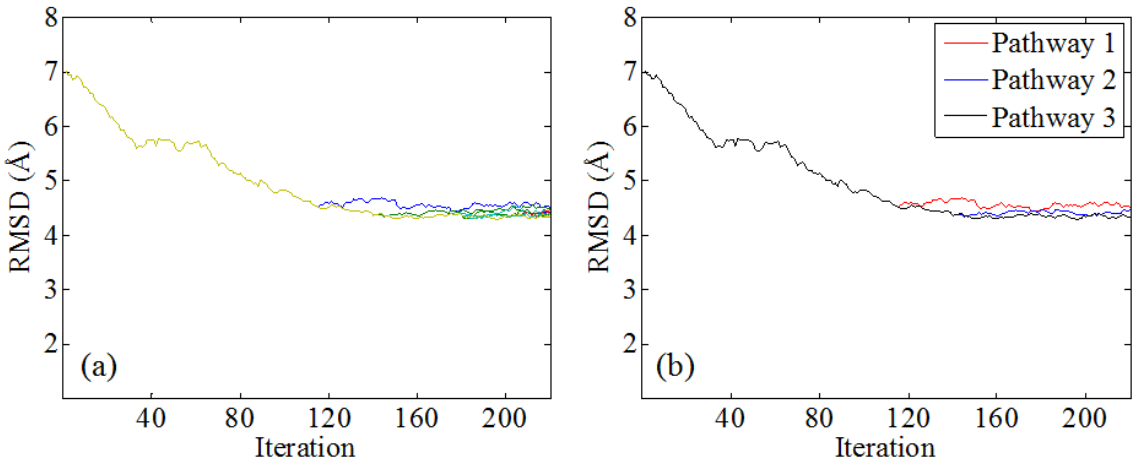


Figure 4.16. AdK_R3 RMSD profiles in reverse direction (a) All trajectories that reach a final RMSD below 4.5 Å, (b) Three main pathways for AdK_1 system that get close to target.

The changes in LID-CORE angle dominate in AdK_R1 and AdK_R2 systems, but in AdK_R3 system, the changes in both angles are comparable. It is still an open question why the open state cannot be reached from the closed crystal structure in AdK.

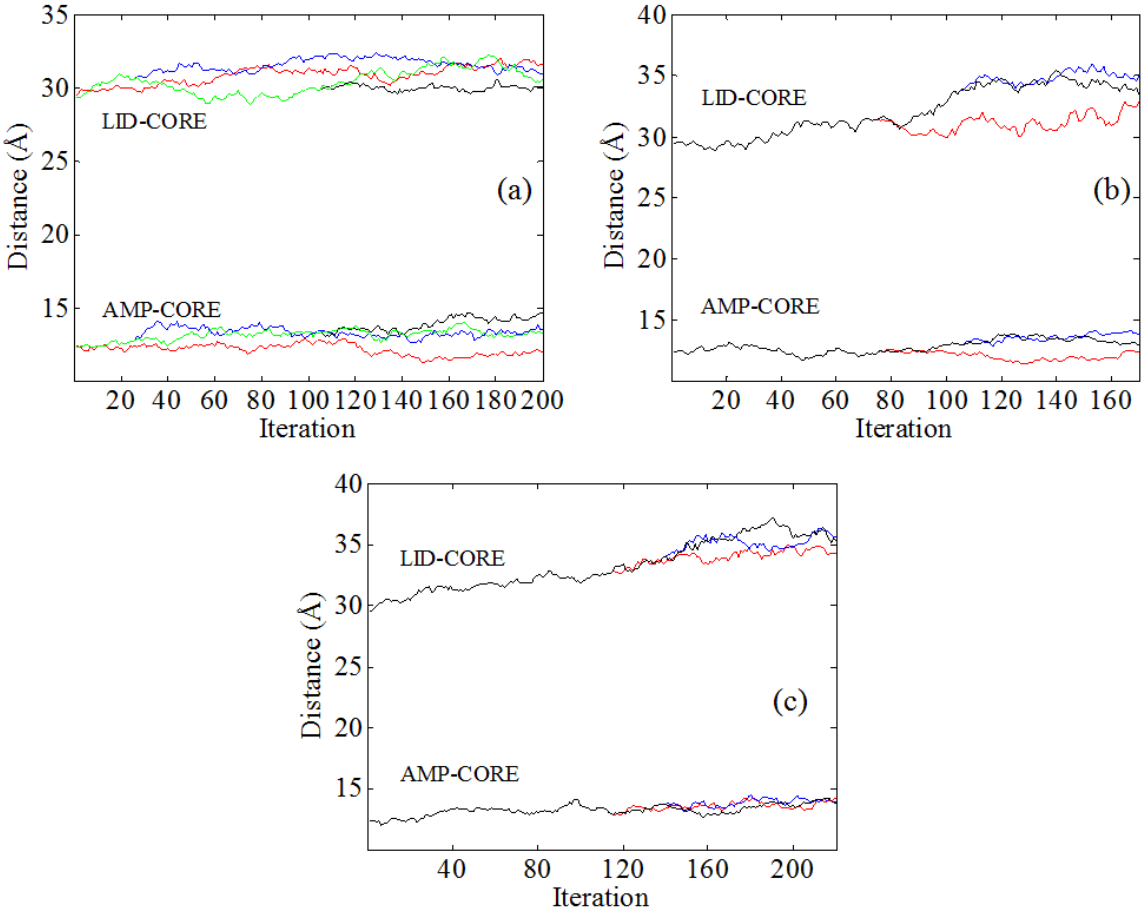


Figure 4.17. FRET distance profiles along the transition for the main pathways of systems (a) AdK_R1, (b) AdK_R2 and (c) AdK_R3.

4.4. Comparison of all AdK Systems

In this section, the free energy surfaces obtained from previous studies in literature will be used to compare the pathways of all AdK systems. Arora and Brooks (2007) defined the reaction coordinate ΔD_{RMSD} as the difference between the RMSDs of an intermediate structure to open and closed crystal structures. They calculated the free energy surface based on ΔD_{RMSD} and the distance between the mass centers of residues ILE52 and LYS145 that belong to the AMP and LID domains, respectively. Figure 4.20 shows ILE52-LYS45 distance scanned by all successful trajectories (total 20 pathways) including six AdK forward systems. In Figure 4.21, ΔD_{RMSD} and distance values are plotted for those pathways that have minimum final RMSD values in each of the six AdK forward systems. In this figure, the data are projected onto the free energy profile

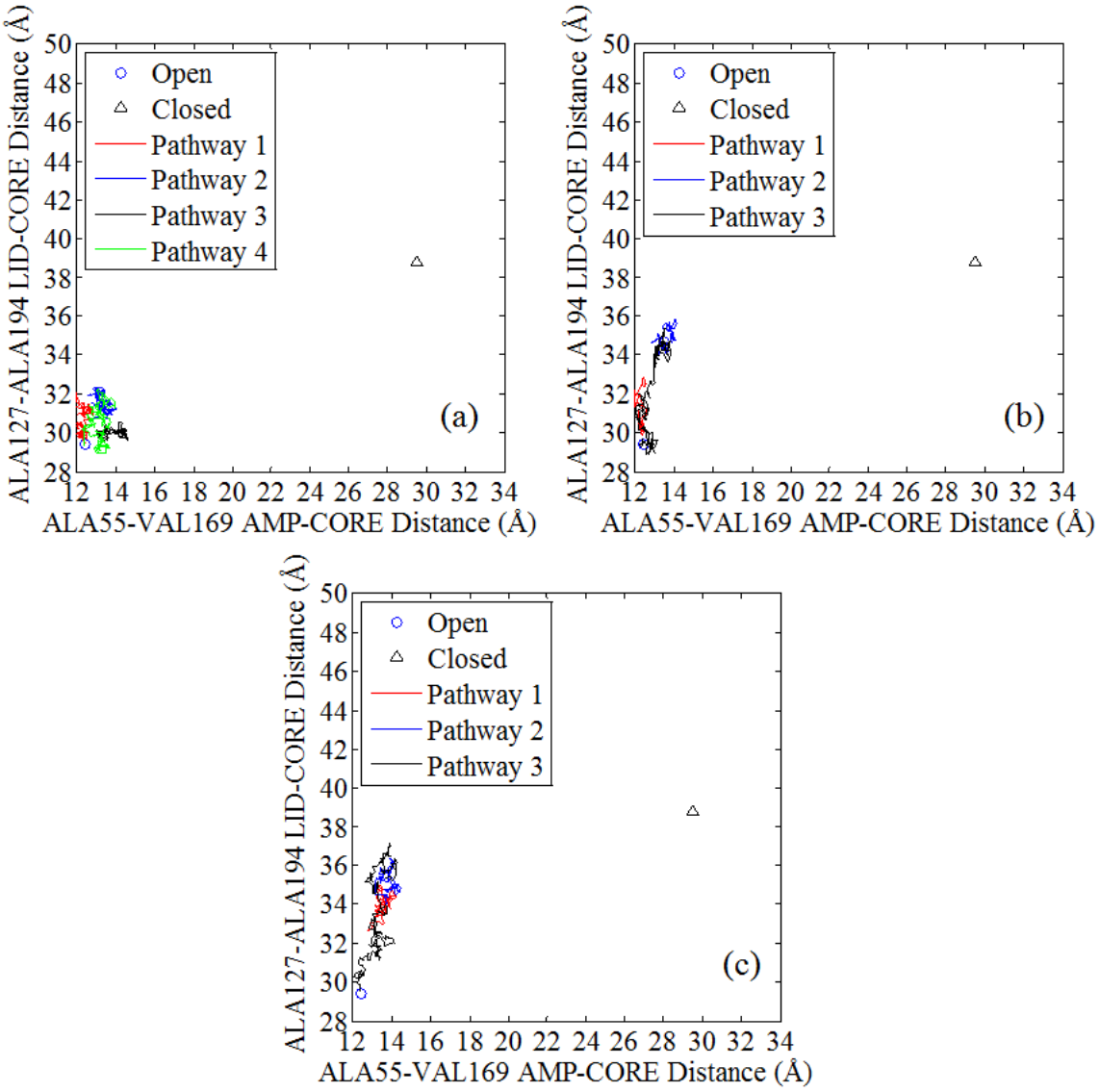


Figure 4.18. FRET distances for the main pathways of systems (a) AdK_R1, (b) AdK_R2 and (c) AdK_R3.

from Arora and Brooks (2007). This plot shows that most trajectories pass through the lower energy regions (shown in blue, color scale in kcal/mol) during the transition. Starting value for ILE52-LYS45 distance in the crystal structure is 44.73 Å, which lies slightly out of the low energy region. Moreover, the inter-domain center of mass distances (LID-CORE and AMP-CORE) are calculated for the best pathways of six AdK forward systems and projected onto the free energy landscapes (color scale in kcal/mol) in Figure 4.22 (Arora and Brooks, 2007).

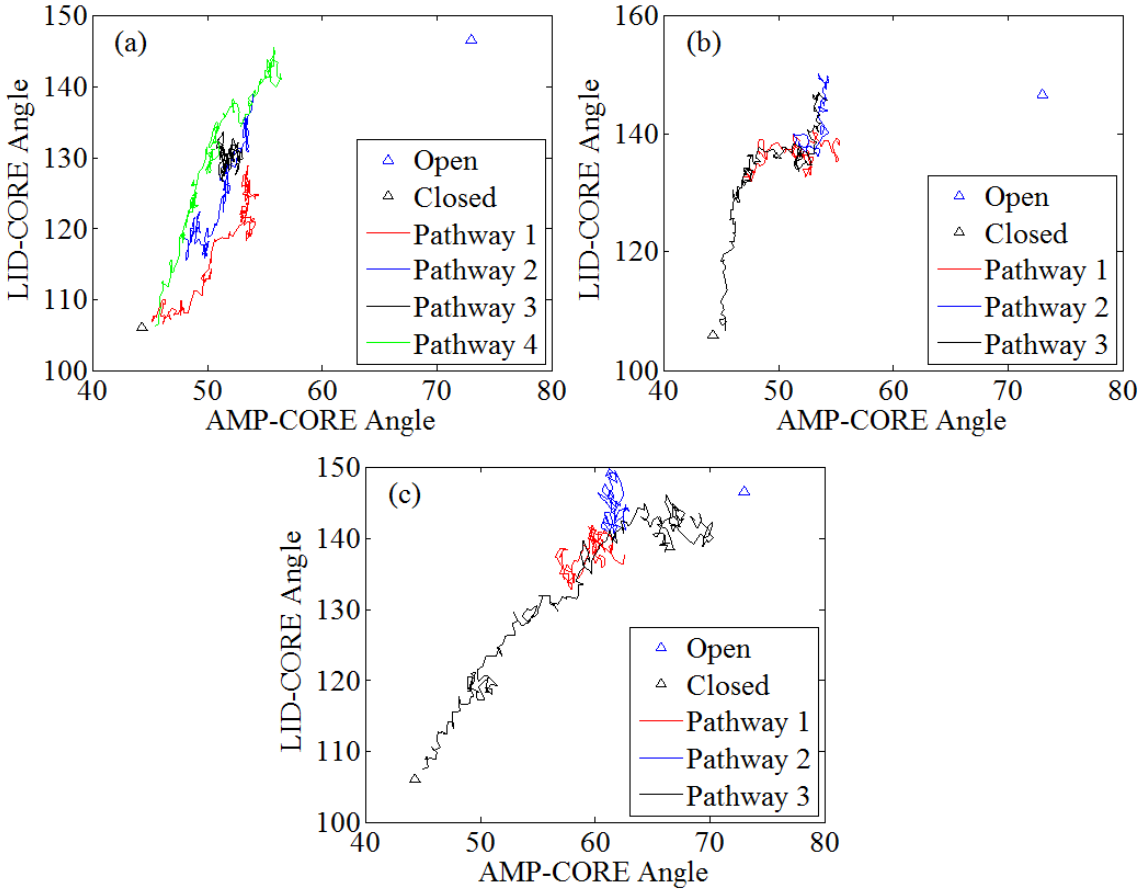


Figure 4.19. LID-CORE (θ_{LID}) and AMP-CORE (θ_{AMP}) angles for the main pathways of systems (a) AdK_R1, (b) AdK_R2 and (c) AdK_R3.

In Figure 4.23 domain angles, LID-CORE (θ_{LID}) and AMP-CORE (θ_{AMP}), for the best pathways of six forward AdK systems are projected onto the free energy landscape, calculated in the study of Beckstein *et al.* (2009). White continuous lines represent Dynamic Importance Sampling (DIMS) transition reverse pathways, while dashed lines represent forward pathways (Beckstein *et al.*, 2009).

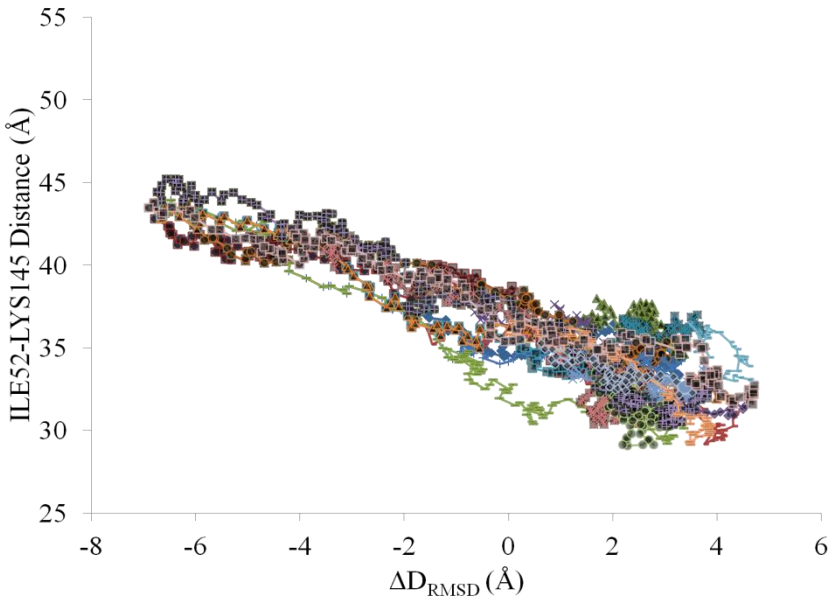


Figure 4.20. ILE52-LYS145 distance for all successful trajectories of six AdK systems.

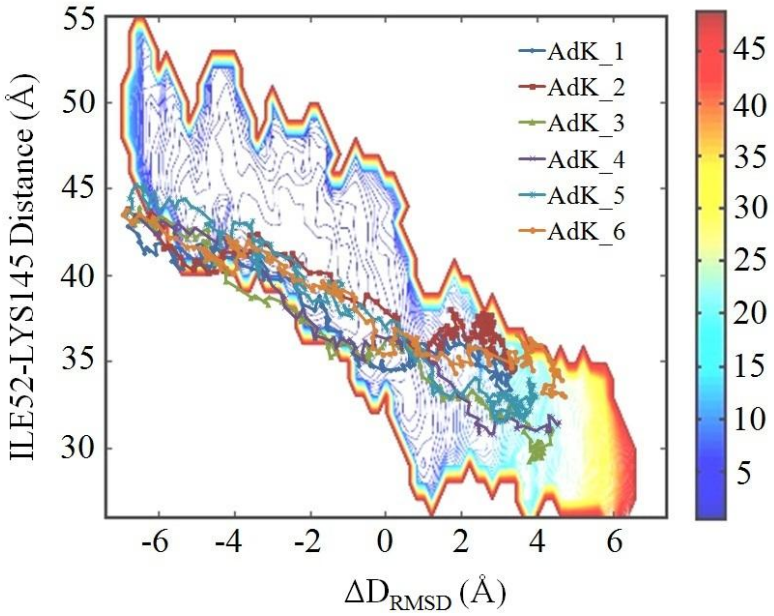


Figure 4.21. Free energy transition for the most successful pathways of six forward AdK systems.

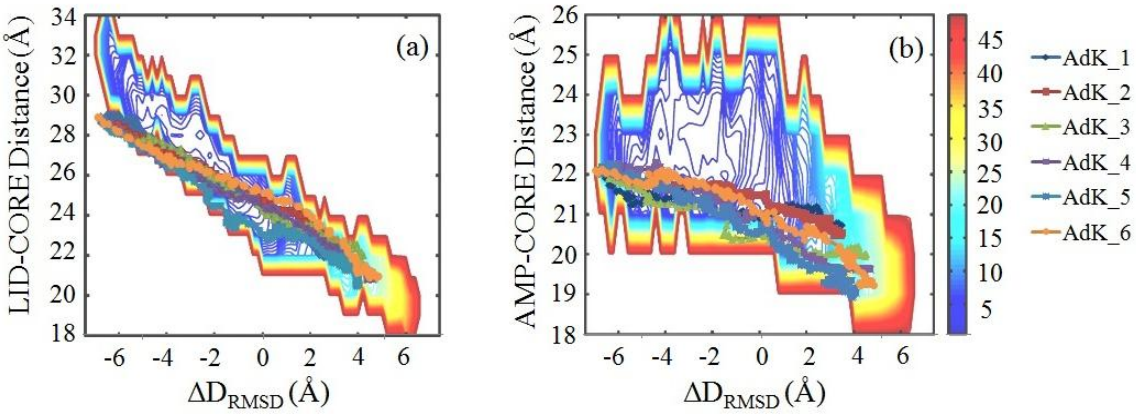


Figure 4.22. Mass center distances of (a) LID-CORE and (b) AMP-CORE along the transition for the best pathways of six AdK forward systems.

This landscape has shown that the change in θ_{LID} transition is energetically more favorable than the θ_{AMP} where there are higher energy barriers especially in the region $55° < \theta_{AMP} < 60°$. Starting from the open structure (shown with black box) on the top right corner, AdK systems do not pass through any region with the free energy over 5 kcal/mol during their transition to the closed structure (shown with black box on the bottom left corner). Figure 4.24 displays domain angles of the best pathways of three reverse AdK systems on the same free energy profile. AdK_R3 tends to follow the minimum free energy pathway as it has no more than 2 kcal/mol of free energy through the transition. Other two systems AdK_R1 and AdK_R2, however, go through different regions and the energy increases at the end of the transition. Three closed to open AdK simulations show that the transition is not reversible due to using closed crystal structure as starting configuration instead of starting from the closest structure obtained with forward simulations. Ligand induced interactions possibly form a barrier to reach the open structure.

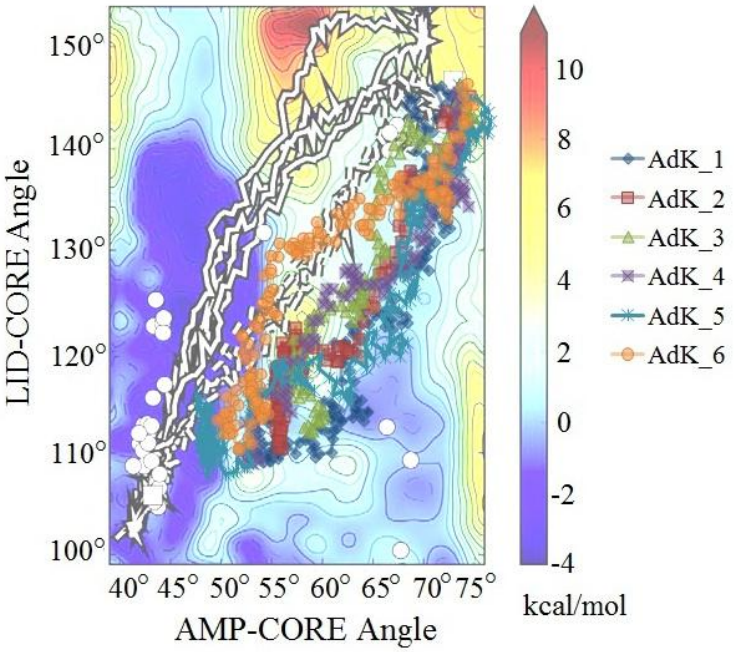


Figure 4.23. Projection of inter-domain angles θ_{LID} and θ_{NMP} onto free energy profile for the best pathways of six forward AdK systems.

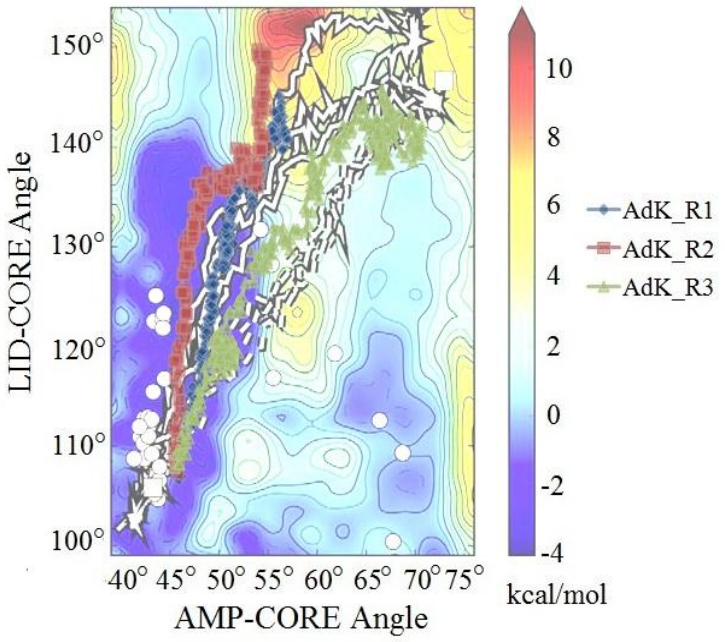


Figure 4.24. Projection of inter-domain angles θ_{LID} and θ_{NMP} onto free energy profile for the best pathways of three reverse AdK systems.

4.5. CaM Forward Runs with Cutoff Distances of 18 Å and 10 Å

Conformational transition of CaM is also investigated by applying the same procedure, namely the WE method with random ANM-MC iterations. Two systems, namely CaM_1 and CaM_2, are simulated with cutoff distances are taken as 18 Å and 10 Å, respectively. In the simulations, the reaction coordinate (RMSD space or conformational change) is split into 75 bins (N_{bin}=75) to give each bin an interval of 0.2 Å. Five collective ANM moves (or iterations) are performed in each fixed time interval (N_τ = 5) used for WE sampling and 10 parallel simulations exist per bin (M=10).

The initial RMSD of 15.01 Å decreases down to 3.44 Å and 4.62 Å in CaM_1 and CaM_2 systems, respectively. Figure 4.25 shows all the resulting trajectories of CaM_1 system. Successful trajectories that have final RMSD below 4.5 Å are plotted in Figure 4.26a together with the three main pathways (Figure 4.26b) used for further analysis. Similarly, distinct pathways of successful CaM_2 system trajectories (final RMSD less than 4.5 Å) are selected based on defining early (first few) breakpoints in RMSD transition as significant pathways and shown in Figure 4.27.

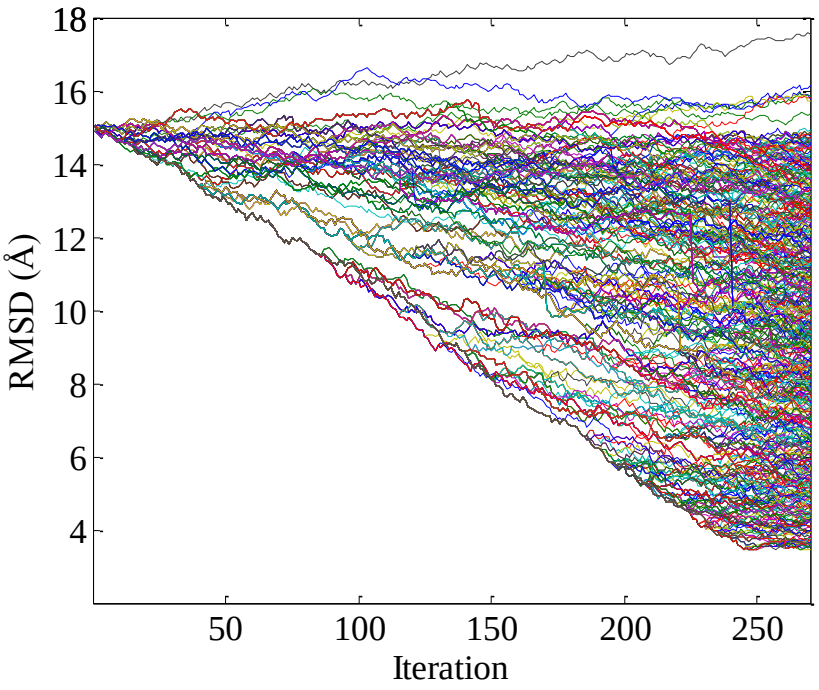


Figure 4.25. RMSD profiles of all trajectories for CaM_1 system.

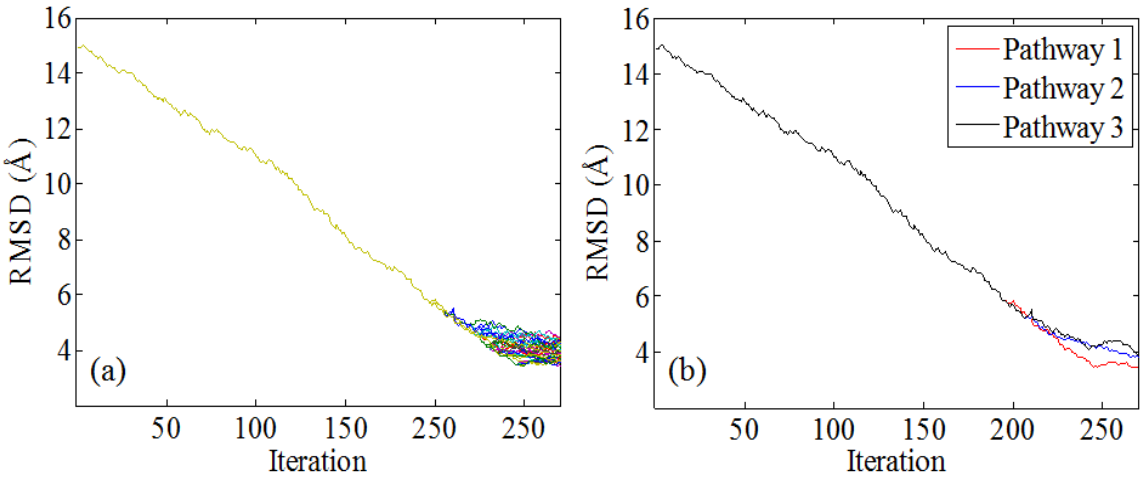


Figure 4.26. CaM_1 RMSD profiles in forward direction (a) All trajectories that reach a final RMSD below 4.5 Å, (b) Three main pathways for CaM_1 system that get close to target.

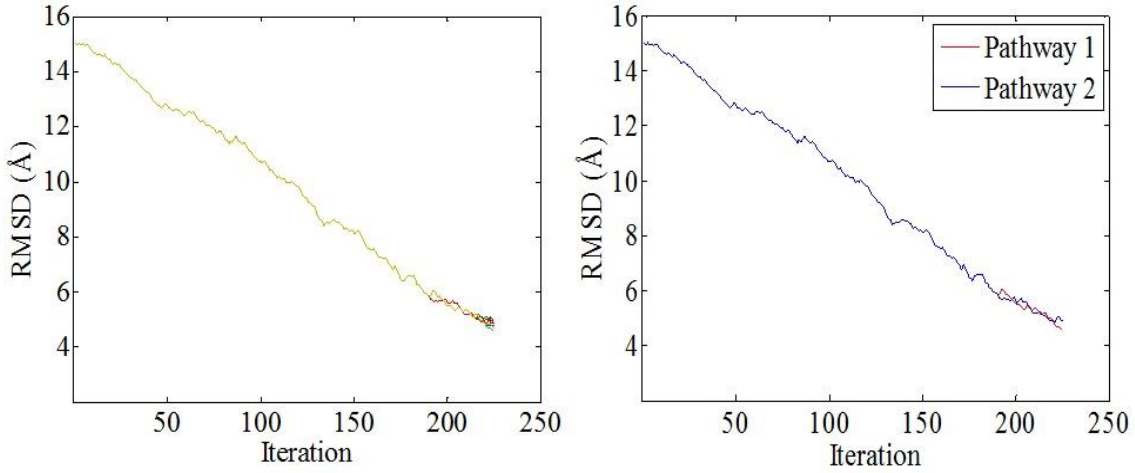


Figure 4.27. CaM_2 RMSD profiles in forward direction (a) All trajectories that reach a final RMSD below 5.0 Å, (b) Two main pathways for CaM_2 system that get close to target.

Center of mass distances between N-terminal (residues 5-77) and C-terminal (residues 78-146) domains for each pathway of CaM_1 system are shown in Figure 4.28. It starts from 36.9 Å to reach the minimum value of 22.8 Å for the first pathway.

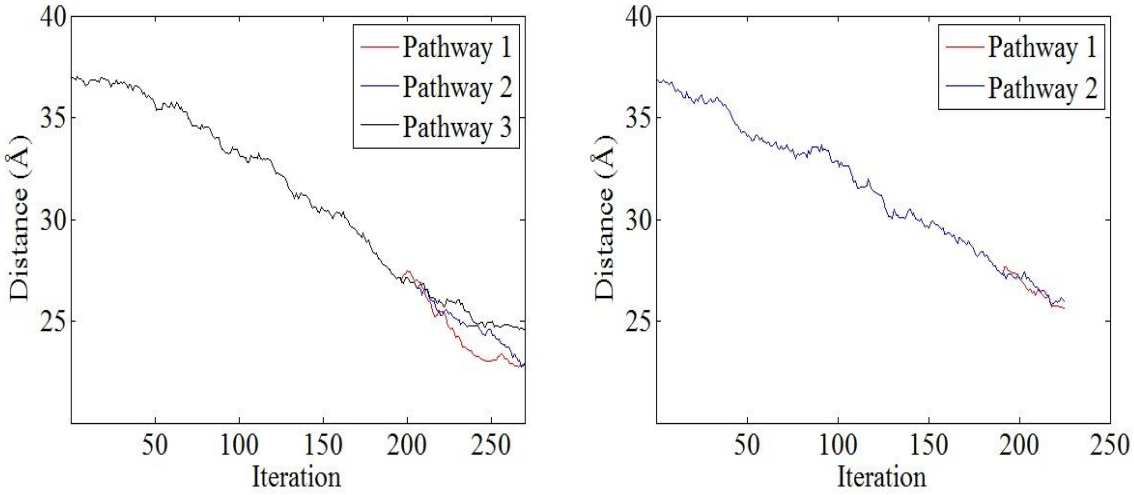


Figure 4.28. Center of mass distances between N-terminal and C-terminal domains of (a) CaM_1 system, (b) CaM_2 system.

4.6. GBP Forward Runs with Cutoff Distances of 18 Å and 10 Å

The same procedure is also applied to GBP using cutoff distances of 18 Å and 10 Å to observe distinct and unbiased pathways. RMSD space of 5.34 Å between open and closed crystal structures of GBP is divided into 30 small bins (N_{bin}=30). The other parameters are the same as previously explained CaM systems. At the end of N_τ = 5 iterations, combination and splitting are performed to retain 10 simulations (M=10) per bin by using the WE method principles. Figure 4.29 shows the RMSD profiles of all trajectories for GBP_2 system. Successful trajectories of GBP_1 and GBP_2 systems are analyzed using a 2.5 Å threshold of final RMSD value to target state. Figure 4.30 and 4.31 display all successful trajectories and four distinct pathways for GBP_1 and GBP_2 systems reaching minimum final RMSD values of 2.02 Å and 1.92 Å, respectively.

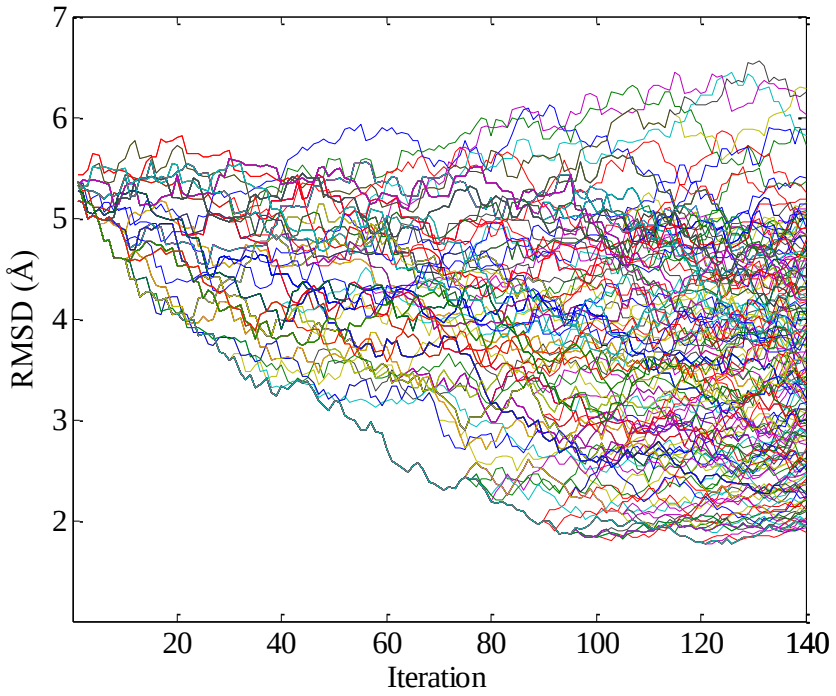


Figure 4.29. RMSD profiles of all trajectories for GBP_2 system.

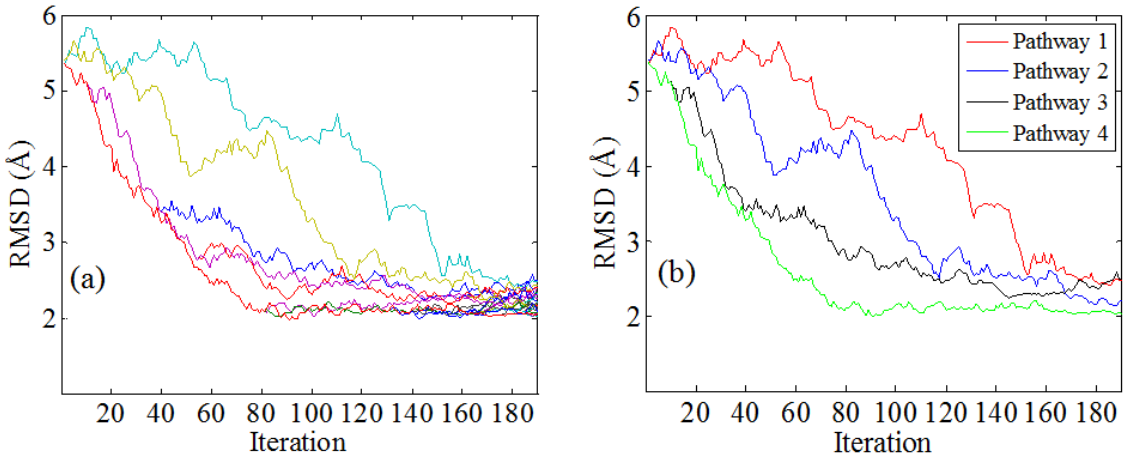


Figure 4.30. GBP_1 RMSD profiles in forward direction (a) All trajectories that reach a final RMSD below 2.5 Å, (b) Four main pathways for GBP_1 system that get close to target.

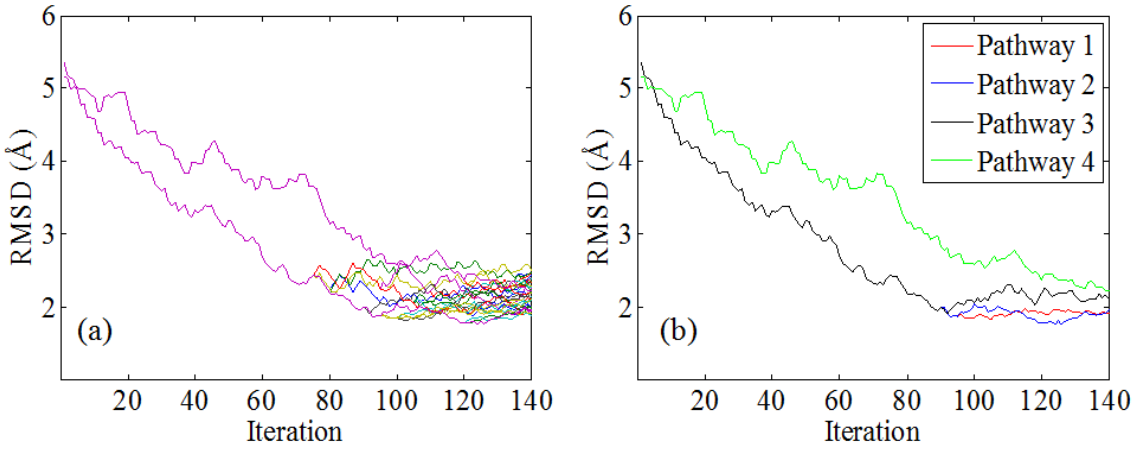


Figure 4.31. GBP_2 RMSD profiles in forward direction (a) All trajectories that reach a final RMSD below 2.5 Å, (b) Four main pathways for GBP_2 system that get close to target.

4.7. Cracking Analysis

Cracking phenomena, which has been a subject of debate lately in the literature, is investigated in this section for the three proteins. Accumulation of strain energy in certain parts of the structure has been proposed to cause cracking and local unfolding. In this thesis local strain energy is calculated based on ANM by using the equation given below (Wang *et al.*, 2004).

Here the strain energy of residue i for the k^{th} mode is calculated as a summation over n_i, the number of neighboring residues, which are closer than a specific cutoff distance to i^{th} residue. N_{res} represents the total number of residues and other variables have been described in Materials and Methods chapter. Eigenvalue for k^{th} mode is represented with λ and force constant γ is taken as 1 for all strain energy calculations. Residue-based local strain energies for each pathway are calculated along the transition using Equation 4.1.

Another key feature for cracking analysis is the calculation of pseudodihedral (torsion) angles. Coarse-grained intermediate structures are represented with two nodes located at the backbone (C_α atom) and a sidechain site for each residue (except Gly). Pseudodihedral angle φ_i for the backbone is calculated by using C_α^{i-1}, C_α^{i}, C_α^{i+1} and C_α^{i+2} coordinates according to the formula given below:

where $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$ are three vectors connecting four consecutive C_α atoms.

Dihedral angles are calculated for all the intermediates along the selected pathways of six AdK forward systems. Then, the change (difference) in dihedral angles of the current structure (iteration) and the previous one are calculated. For each residue an average dihedral angle change is calculated based on all iterations and plotted in Figure 4.32-37 together with the total strain energy of each residue, summed over the trajectory (over all iterations). Cutoff distance used for local strain energy calculation is taken as 18 Å in these graphs.

Secondary structures based on the open crystal structure of AdK are also shown at the top of these plots by light and dark green lines. Light green and dark green regions indicate α-helices and β-strands, respectively. And the rest represent loop regions. The aim is to observe whether high strain accumulation and any significant dihedral changes correspond to any of these secondary structural elements. In general, the different pathways in a specific AdK system exhibit quite similar strain energy and dihedral change profiles.

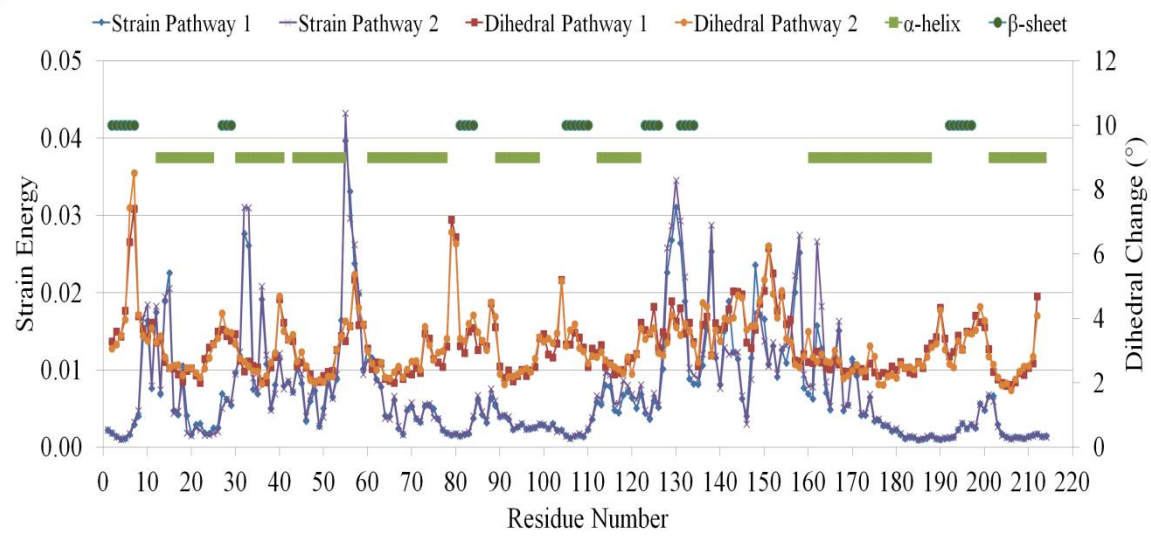


Figure 4.32. Strain energy and dihedral angle change for two pathways of AdK_1 system.

Strain energy is recalculated by changing the cutoff distance to 10 Å. In Figure 4.38, 4.39 and 4.40, the strain energy is plotted for the 'best' (most successful) pathway that ends with minimum RMSD value (Pathway 1) to target in each AdK system. In these figures, the residue-based local strain energy is calculated as an average over the whole trajectory, i.e. over all iterations, using a cutoff distance of either 18 Å or 10 Å, respectively.

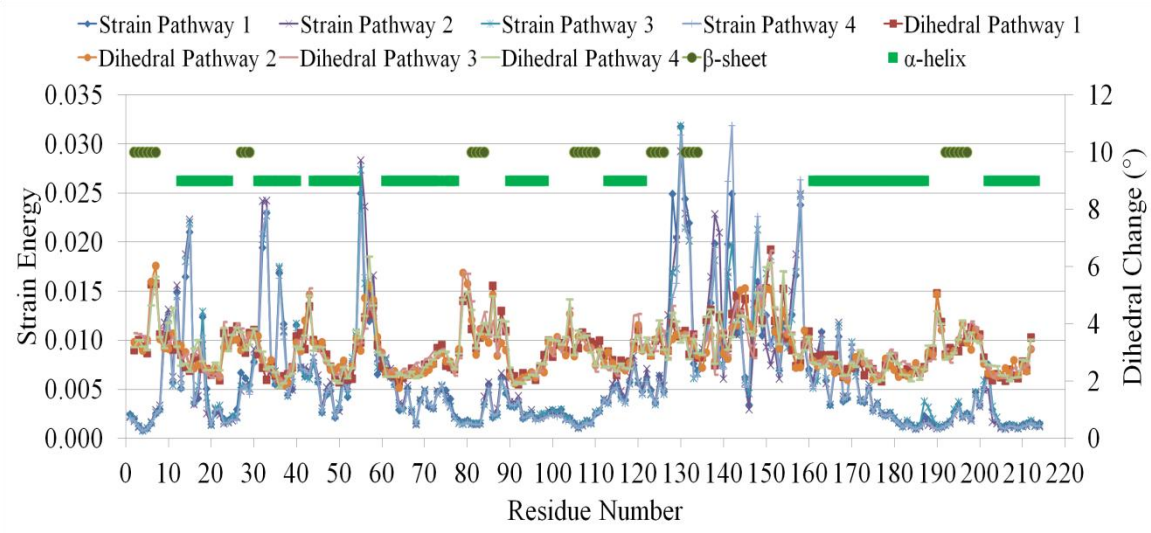


Figure 4.33. Strain energy and dihedral angle change for four pathways of AdK_2 system.

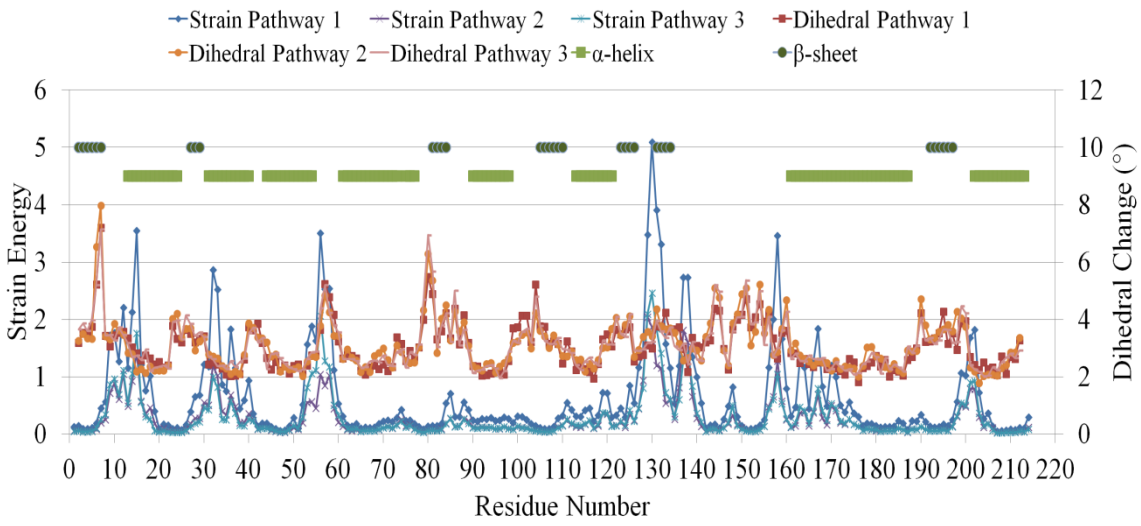


Figure 4.34. Strain energy and dihedral angle change for three pathways of AdK_3 system.

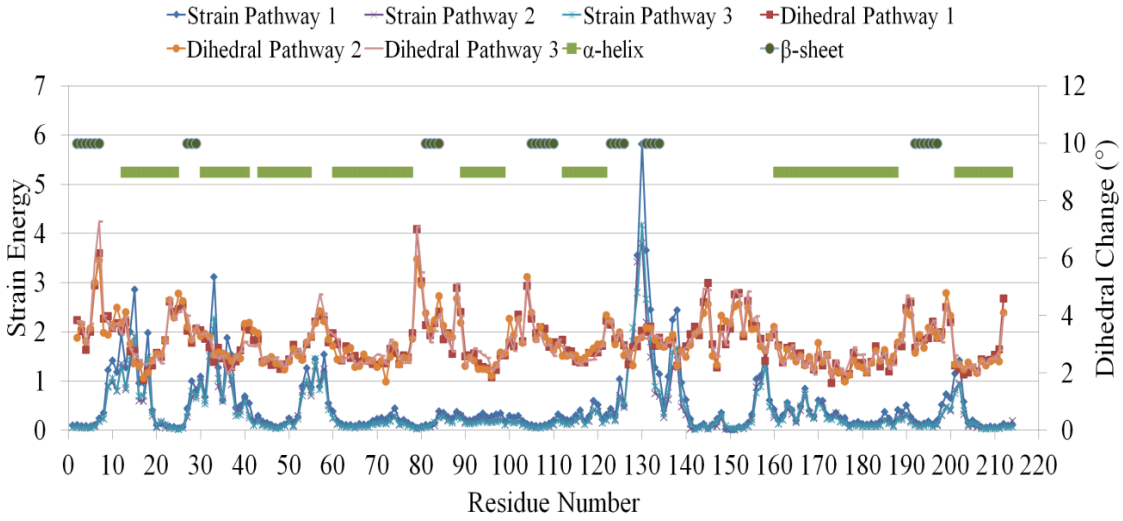


Figure 4.35. Strain energy and dihedral angle change for three pathways of AdK_4 system.

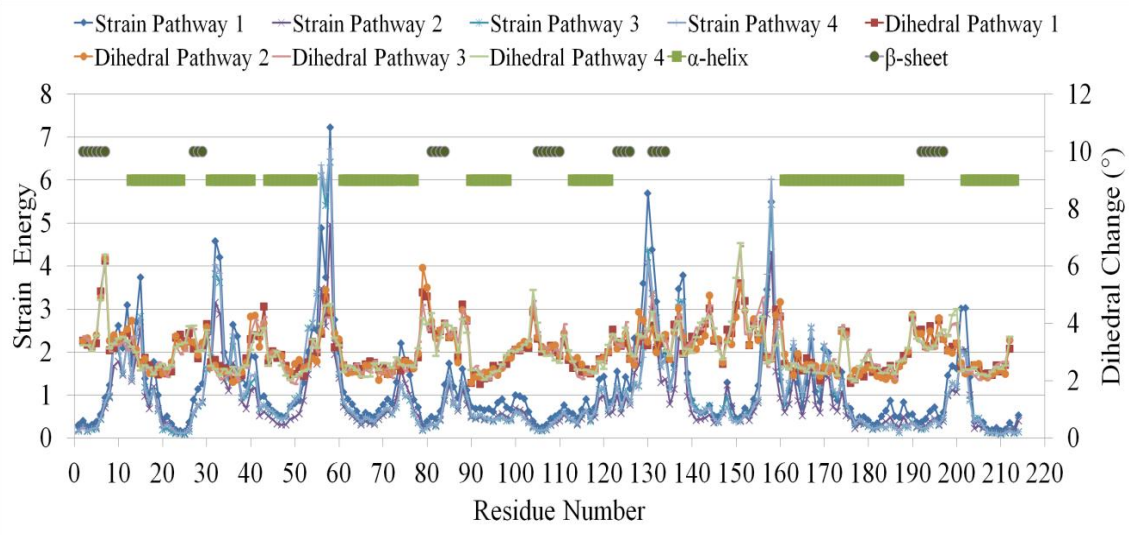


Figure 4.36. Strain energy and dihedral angle change for four pathways of AdK_5 system.

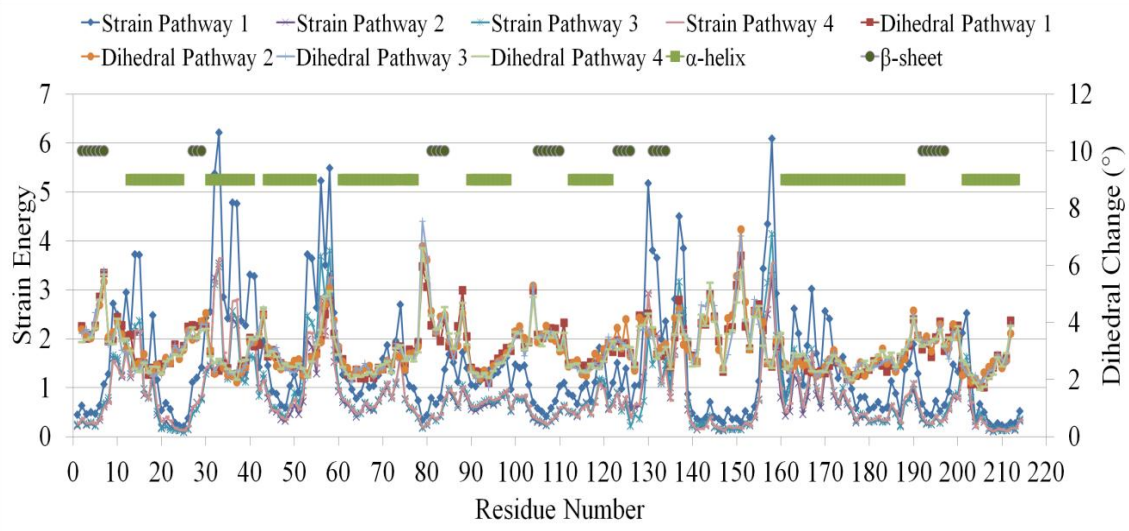


Figure 4.37. Strain energy and dihedral angle change for four pathways of AdK_6 system.

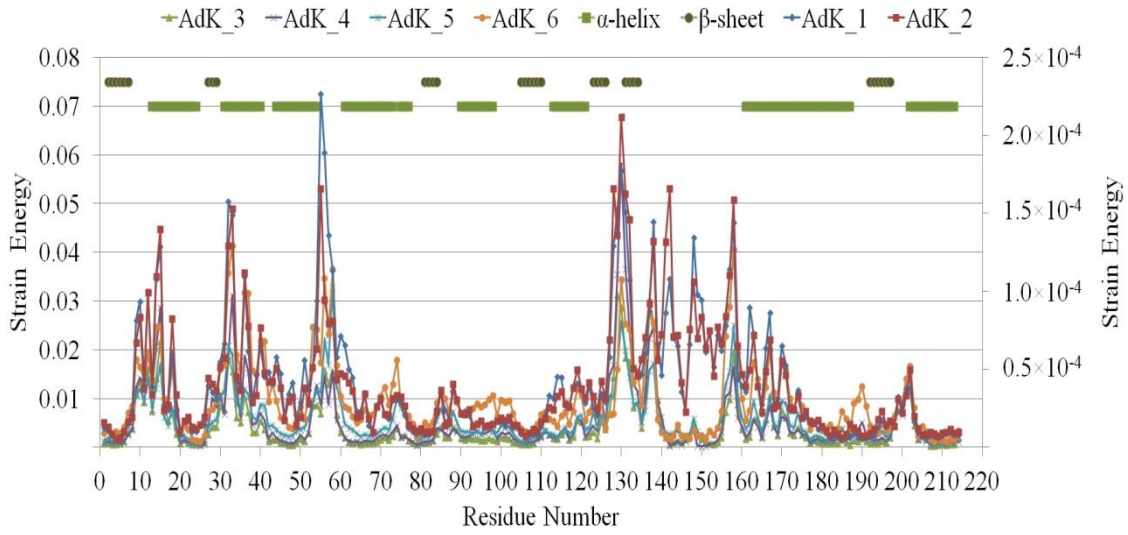


Figure 4.38. Strain energies for the best pathways of six forward AdK systems with a cutoff distance of 18 Å.

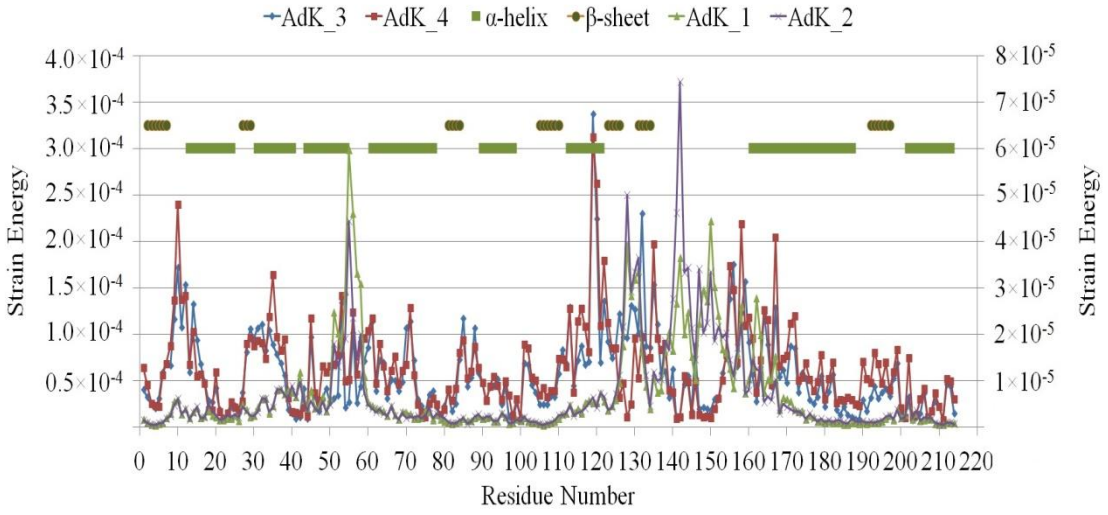


Figure 4.39. Strain energies for the best pathways of four forward AdK systems (AdK_1- AdK_4) with a cutoff distance of 10 Å.

Since the strain energy scales are different for AdK systems, systems with the strain energy cutoff distance of 10 Å are shown separately in Figure 4.39 and 4.40. In Figure 4.39, AdK_3 and AdK_4 systems are compared with AdK_1 and AdK_2 systems. Figure 4.40 shows the strain energy profiles for AdK_5 and AdK_6 systems against AdK_1 and AdK_2 systems.

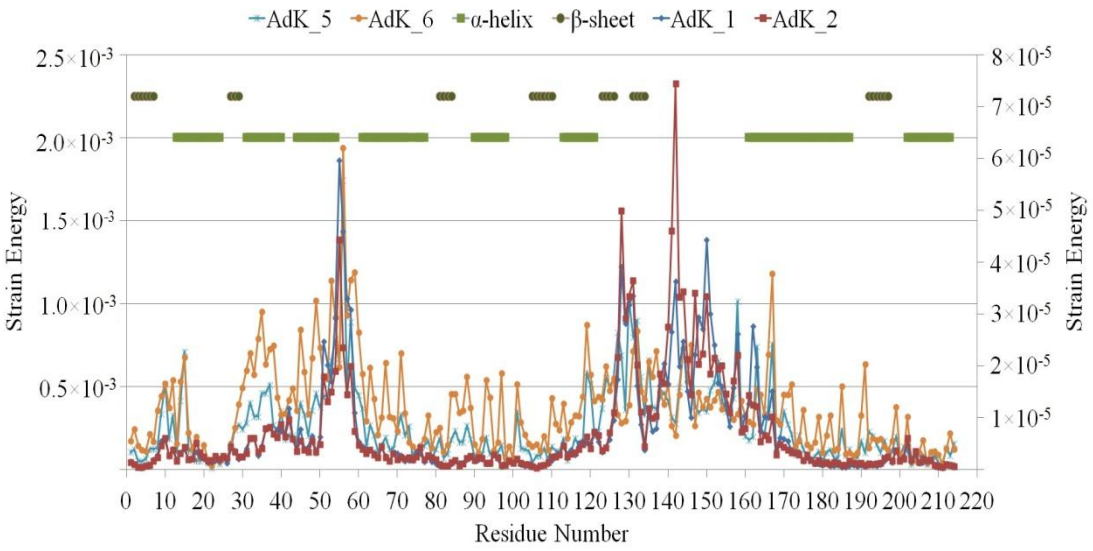


Figure 4.40. Strain energies for the best pathways of four forward AdK systems (AdK_1- AdK_2, AdK_5- AdK_6) with a cutoff distance of 10 Å.

Local unfolding or cracking is expected to take place at regions, where the torsion angle changes are significantly higher than any other parts of the structure. Another definition that has been used for local unfolding, T_i, is based on the deviations of pseudodihedral angle φ_i from its values calculated for open and closed structures (Whitford *et al*., 2007).

Since there are N-3 torsion angles for N residues in a structure, one needs to make an assumption to assign a value for each residue. First and last two residues were assumed to have the first and last torsion angle values, respectively. However, T_i for the other residues is calculated as an average of T_i and T_{i+1} that correspond to the torsion angle changes on the left and righthand side of i^{th} residue (Whitford *et al*., 2007). This definition can in principle detect whether the change in any residue's torsion angle follows a smooth transition from the open to the closed structure or goes through a significant jump that may be indicative of local unfolding. Dihedral angle change representing the unfolding parameter (T_i) is plotted against residues for the best pathways (Pathway 1) in six AdK forward systems in Figure 4.41.

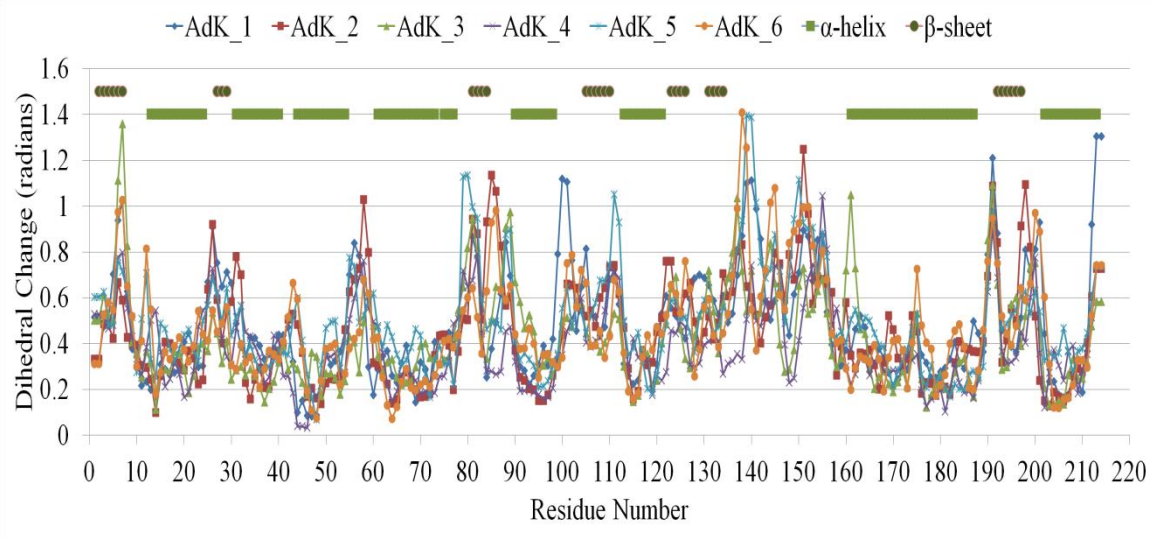


Figure 4.41. Dihedral angle change in radians for best pathways of six forward AdK systems.

Considering six forward AdK systems in Figure 4.32-37, torsion angle changes and total strain energies of distinct pathways do not differ from each other. They show peaks in similar regions of the structure. Cracking (local unfolding) has been claimed to take place at regions, where both torsion angle changes and strain energies display high values, i.e. peaks (Miyashita *et al.*, 2003). To detect the residues in α-helical or β-strand regions that possibly exhibit local unfolding or cracking, threshold values need to be determined first for dihedral changes and strain energies. For this purpose, the mean and the standard deviation of these properties are calculated based on all residues for each system independently. The threshold is set as the summation of mean and standard deviation value for each system. Residues with higher strain energy (red circles) and/or dihedral angle change (black circles) than the corresponding threshold values are shown for the best pathways of six forward AdK systems in Figure 4.42 and 4.43, for which the strain energies are calculated with cutoff distances of 18 Å and 10 Å, respectively.

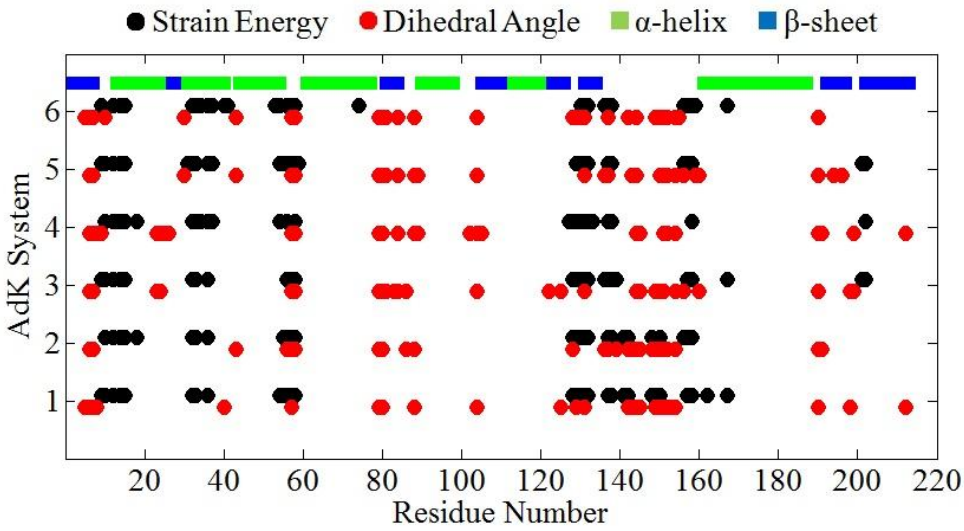


Figure 4.42. AdK forward system results for cracking with a strain energy cutoff distance of 18 Å.

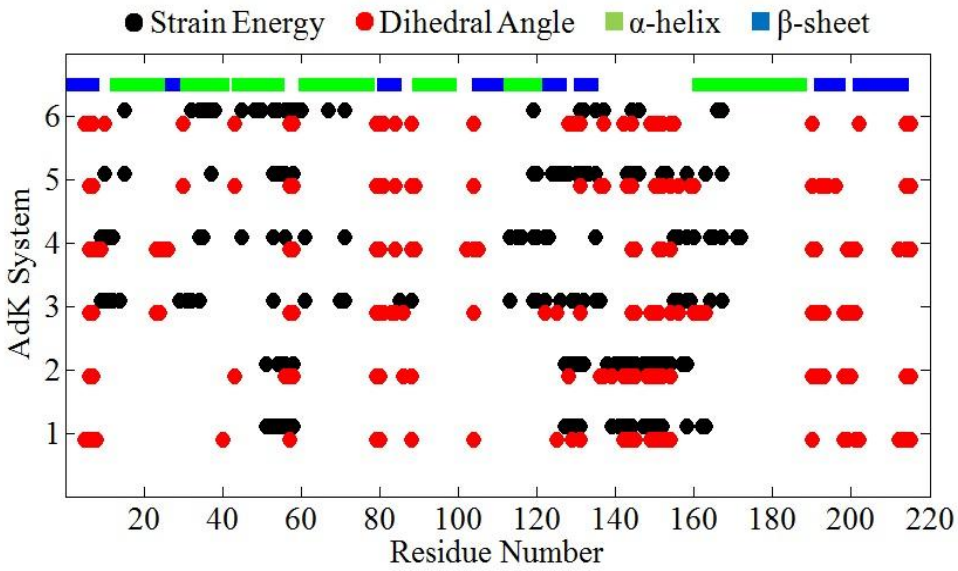


Figure 4.43. AdK forward system results for cracking with a strain energy cutoff distance of 10 Å.

Cracking analysis is also carried out using the unfolding parameter (T_i) from dihedral angles. The T_i threshold is defined in a similar manner for the best pathway of each AdK system. Residues showing significant dihedral angle changes are indicated in Figure 4.44 and 4.45, besides residues with high strain energies calculated with cutoff distances of 18 Å and 10 Å, respectively.

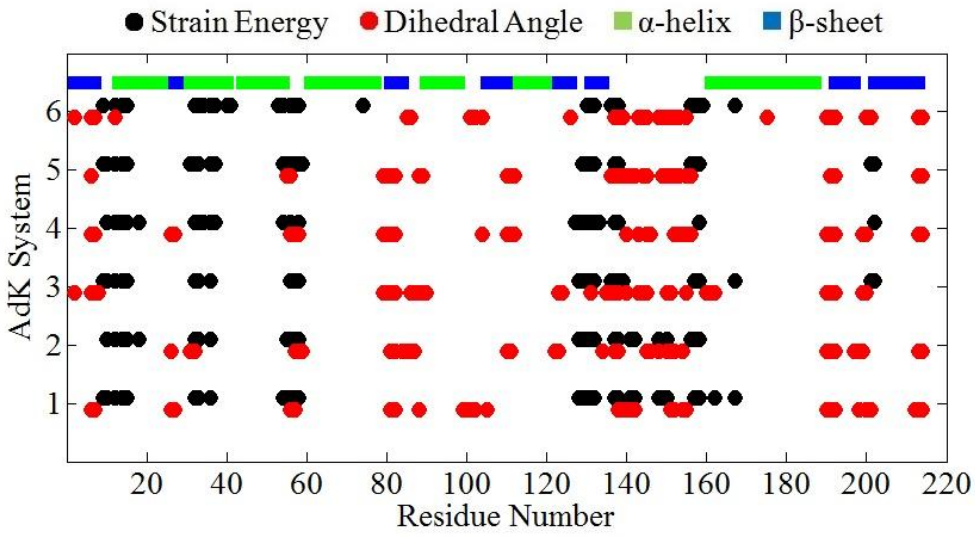


Figure 4.44. Local unfolding results for AdK forward systems with a strain energy cutoff distance of 18 Å.

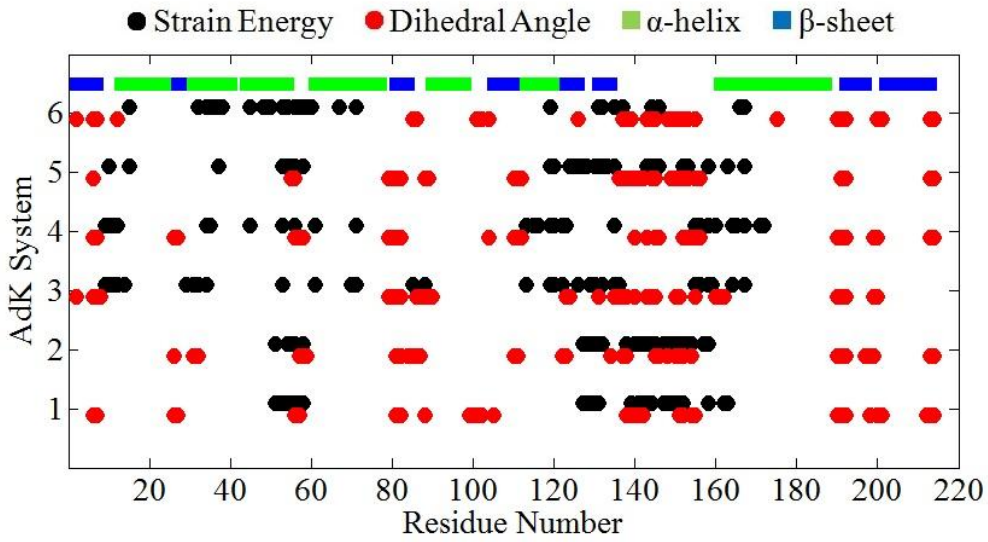


Figure 4.45. Local unfolding results for AdK forward systems with a strain energy cutoff distance of 10 Å.

By analyzing cracking figures, highly strained residues for six forward AdK systems can be summarized. Commonly observed regions with strain energy over the threshold value are listed in Table 4.2 below for cutoff distances of 18 Å and 10 Å used in strain energy calculations. These regions are listed separately for systems with ANM cutoff 18 Å (AdK_1 and AdK_2) and 10 Å (AdK_3-AdK_6). In the case of 18 Å,

residues are listed if high strain energy is observed in both AdK_1 and AdK_2 systems. For 10 Å cutoff systems, residues are reported if high strain energy is observed in at least three out of the four systems (AdK_3-AdK_6).

Significant dihedral angle changes are also listed following the same procedure in Table 4.3, using the two different definitions for dihedral angles for the systems with different ANM cutoffs. The first definition is the simple difference of dihedral angles in two consecutive iterations. The second refers to the unfolding parameter T_i calculation in Equation 4.3.

Table 4.2. Residues with high strain energy than threshold value for AdK forward systems.

Strain Cutoff (Å)	**ANM Cutoff (Å)**	**High Strain Regions**
18	18	10, 12, 14-15, 32, 33, 36, 55-58, 128-132, 137-138, 141-142, 148, 150, 157-158
18	10	9, 10, 12, 14-15, 32-33, 36-37, 56-58, 128-132, 137-138, 157-158, 202
10	18	51, 54-56, 58, 127-131, 141-144, 147-152, 158
10	10	10, 34, 53, 56, 71, 119-120, 132, 135, 158, 167

Our strain energy results for six forward AdK systems show similar regions with other previous studies in literature. Residues 10-12 are compatible with the study of Miyashita *et al*. (2003). Residues 30-35 are under high strain, mostly in AdK_1 and AdK_2 systems (cutoff 18 Å), also found in the study of Whitford *et al*. (2007). Five out of the eight hinge regions defined by Henzler-Wildman *et al*. (2007) are found to have high strain energies, which are α-helical regions of residues 30-32 (hinge 1), 50-54 (hinge 2), 164-172 (hinge 8) and residues 157-163 (hinge 7).

Table 4.3. Dihedral angle changes higher than threshold value for AdK forward systems.

Definition	ANM Cutoff (Å)	High Dihedral Angle Change
1	18	6-7, 57, 79-80, 88, 142-145, 149-152, 154, 190-191
	10	6-7, 57-58, 79-81, 84, 88, 104, 131, 144, 150-152, 154, 190
2	18	26, 57, 81-82, 151-152, 154, 190-192, 198, 213-214
	10	6-7, 79-82, 137-138, 140, 143-145, 150-153, 155, 190-192, 200, 213-214

The regions, where both dihedral angle change and strain energies display high values, correspond to residues 57-58, 130-132, 140-145 and 149-152. Table 4.4 summarizes results of cracking, where high dihedral change and strain energy are indicated for the eight hinge regions of AdK. Systems with 10 Å ANM cutoff distance (AdK_3- AdK_6) are considered here since they reached closer to the closed structure than systems with 18 Å (AdK_1- AdK_2).

Table 4.4. Identified hinge regions of AdK with the results of cracking.

Hinge No.	Residues	Strain Cutoff 18 Å	Strain Cutoff 10 Å	Dihedral Definition 1	Dihedral Definition 2
1	28-32	4/4	2/4	2/4	
2	49-51		1/4		
3	58-63	4/4	4/4	4/4	1/4
4	74-81	1/4		4/4	3/4
5	112-113		2/4		2/4
6	119-121		4/4		
7	157-163	4/4	3/4	2/4	1/4
8	164-172	2/4	4/4		

These hinge regions are shown on the crystal structure of AdK in Figure 4.46. Results show that LID domain hinge regions are under more strain than other parts of the structure. Hinge 3 and hinge 7 are the only regions in common in terms of high dihedral angle changes and strain energies. The reason for identifying less number of regions for cracking might be the energy minimization procedure in our simulations.

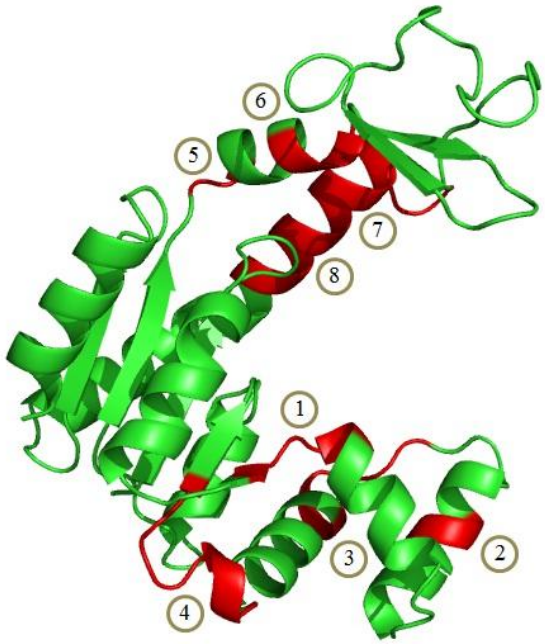


Figure 4.46. Hinge regions on the open crystal structure of AdK.

Cracking is also investigated for CaM system. Strain energies and dihedral angle changes are calculated for distinct pathways of the systems. Figure 4.47 shows the average strain energy and dihedral angle change for each residue in three pathways of CaM_1 for cutoff distance of 18 Å together with secondary structures.

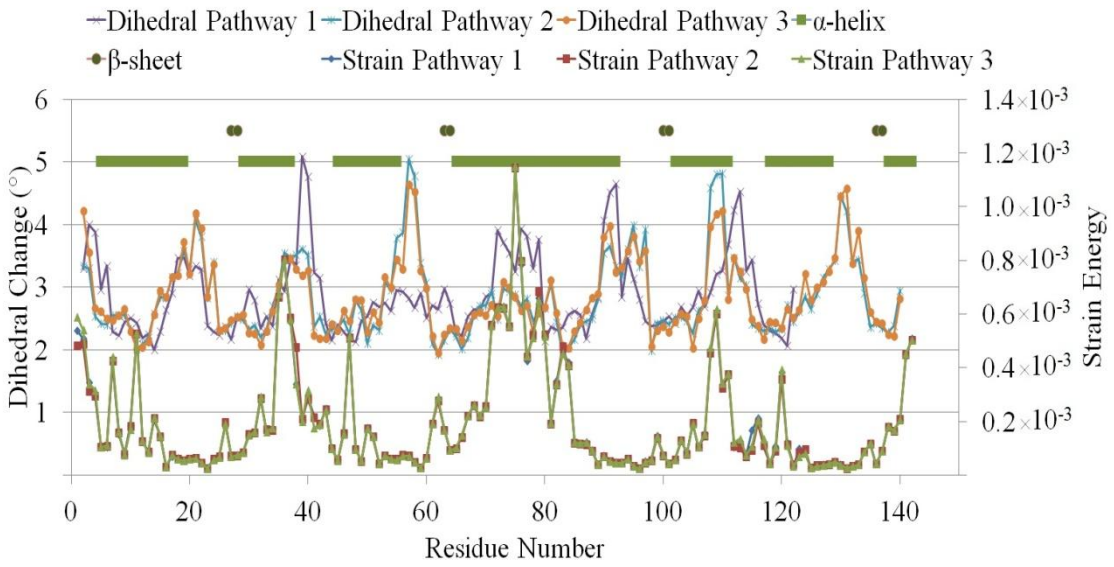


Figure 4.47. Strain energy and dihedral angle change for three pathways of CaM_1 system.

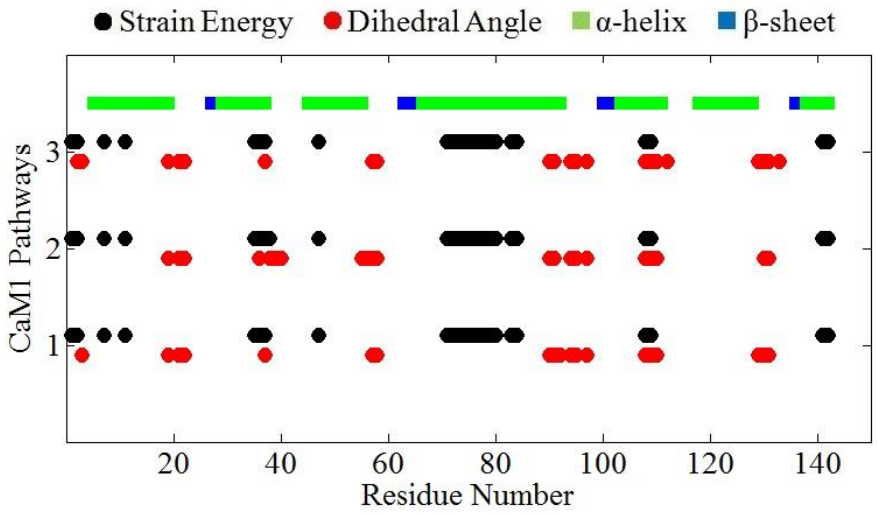


Figure 4.48. CaM_1 forward system results for cracking with a strain energy cutoff distance of 18 Å.

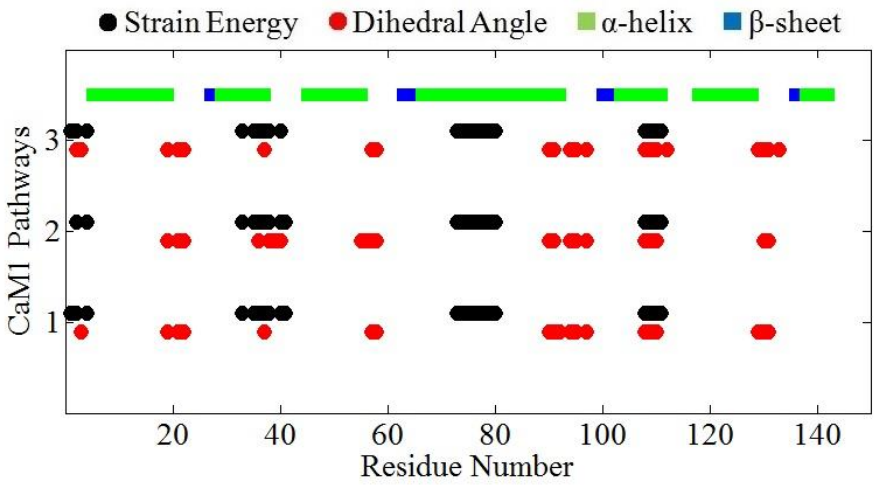


Figure 4.49. CaM_1 forward system results for cracking with a strain energy cutoff distance of 10 Å.

Figure 4.48 and 4.49 gives cracking results for CaM_1 system with strain energy cutoff distance of 18 Å and 10 Å, respectively, using the same threshold definition used for AdK systems. Same procedure is applied for CaM_2 system and shown in Figure 4.50-4.52.

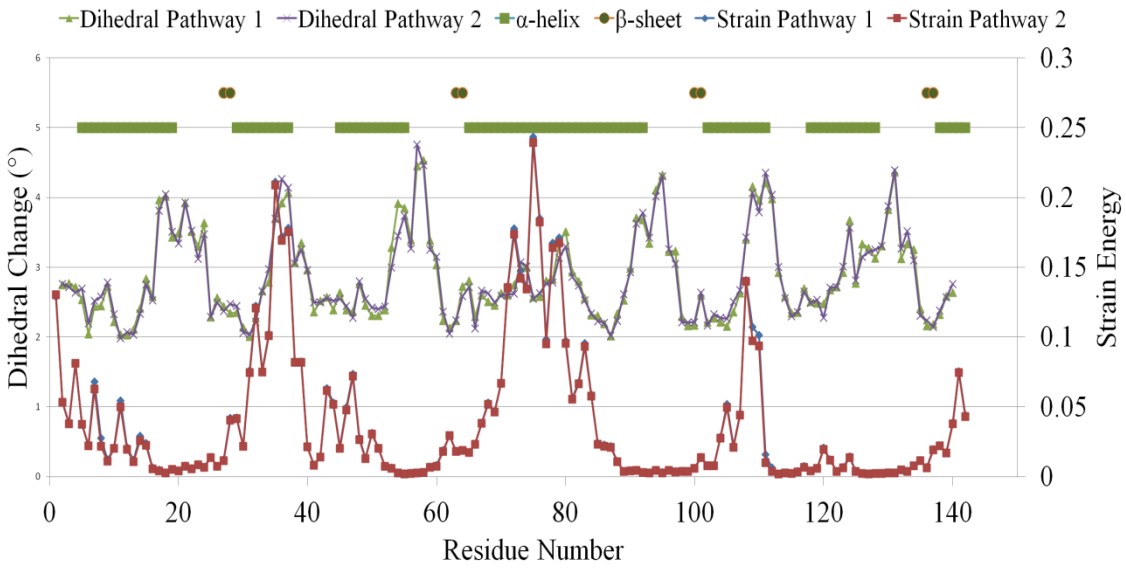


Figure 4.50. Strain energy and dihedral angle change for two pathways of CaM_2 system.

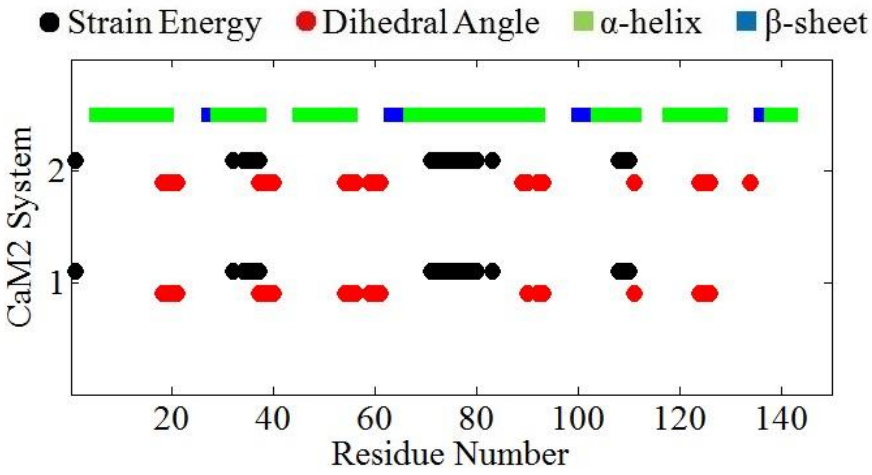


Figure 4.51. CaM_2 forward system results for cracking with a strain energy cutoff distance of 18 Å.

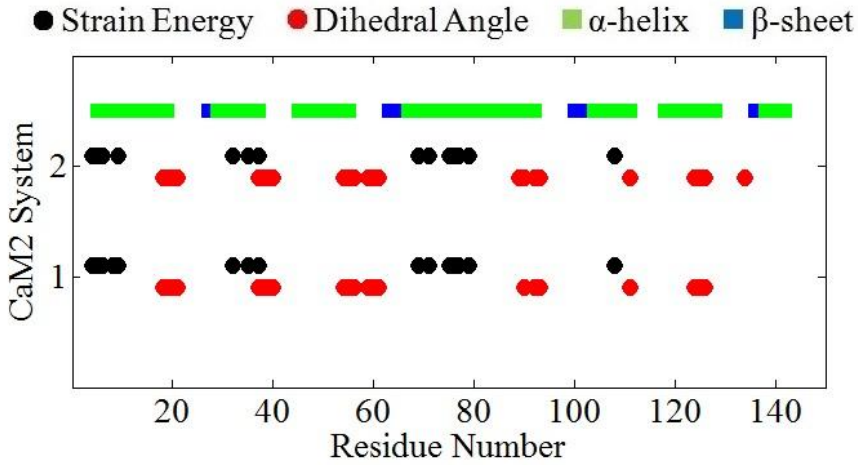


Figure 4.52. CaM_2 forward system results for cracking with a strain energy cutoff distance of 10 Å.

Residues over the threshold value are marked with points in the figures. Residues 37, 108 and 109 that match closely in terms of dihedral angle change and strain energy are possible regions for cracking with α-helical secondary structure. In Figure 4.53, common high strain and dihedral angle change regions observed in calculations these regions are shown in red. First region is residue 37 and the second represents residues

108 and 109. Central helix (residues 65-92), also called linker, is highly strained in calculations with both cutoff distances but it is not labeled as cracking region since dihedral angle change is limited.

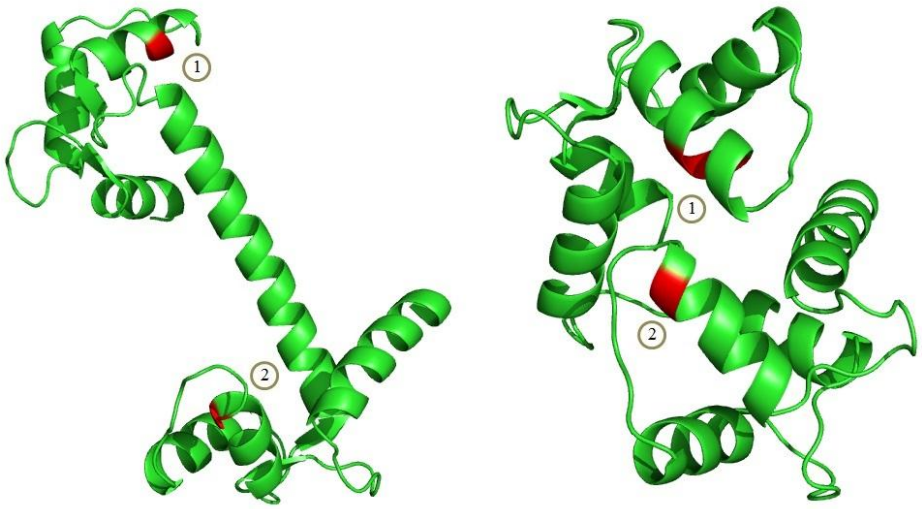


Figure 4.53. Identified cracking regions on the open crystal structure of CaM.

GBP_1 and GBP_2 systems with their four distinct pathways are also subject to cracking analysis in this study. Figure 4.54 gives residue profile for GBP_1 system and significant regions for cracking is determined in Figure 4.55 and 4.56. Residues 15-19 and 21 are shared regions in terms of strain energy and dihedral angles for GBP_1 system; however, they are part of a loop region.

Figure 4.57 shows strain energy and dihedral angle change profiles for each pathway of GBP_2 system. Figure 4.58 and 4.59 give residues with high strain energy and dihedral change considering the same threshold definition.

GBP_2 system results also show that residues 15-21 are under high strain and dihedral change both in two different cutoff distance values. Moreover, residues 81, 82, 84 and 216 are added in 10 Å cutoff distance case. GBP_2 system with cutoff distance of 18 Å includes residues 17-19 and 113 in common.

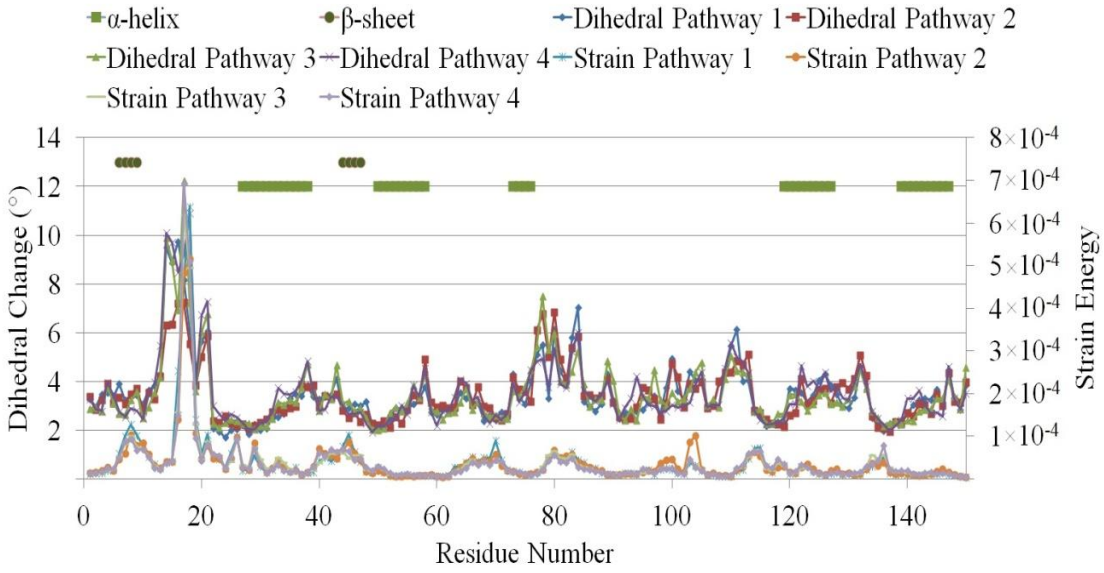


Figure 4.54. Strain energy and dihedral angle change for four pathways of GBP_1 system.

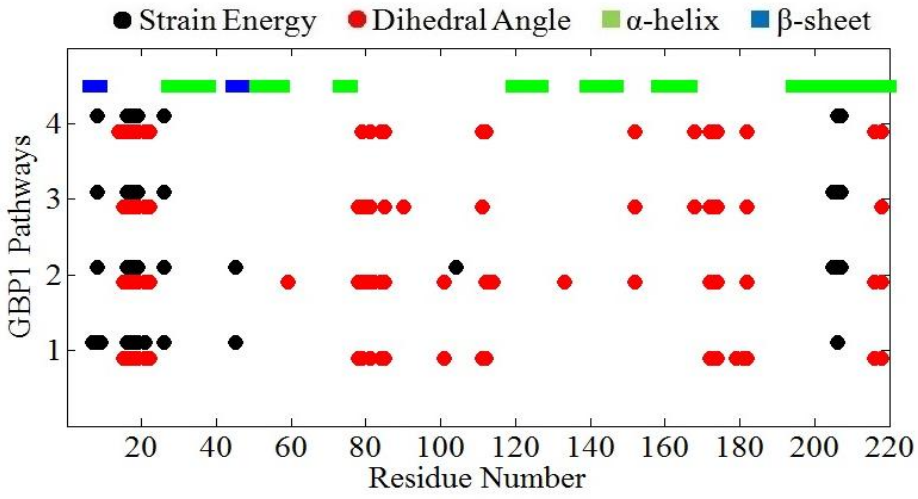


Figure 4.55. GBP_1 forward system results for cracking with a strain energy cutoff distance of 18 Å.

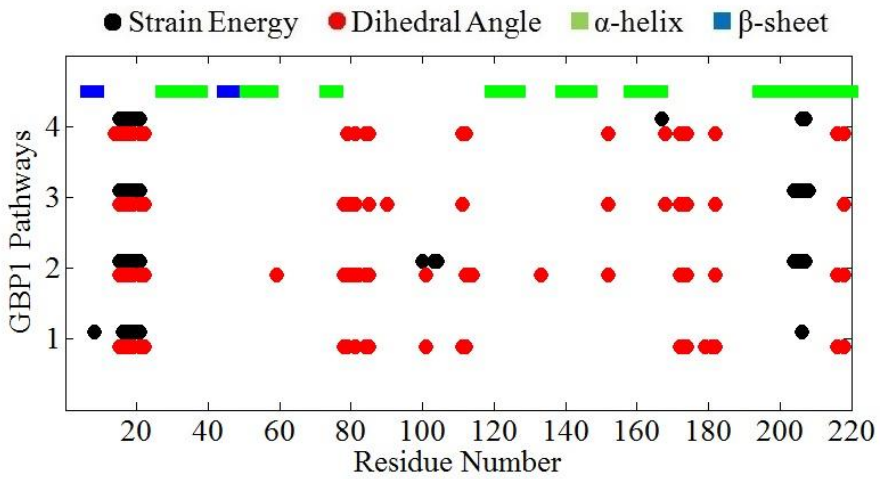


Figure 4.56. GBP_1 forward system results for cracking with a strain energy cutoff distance of 10 Å.

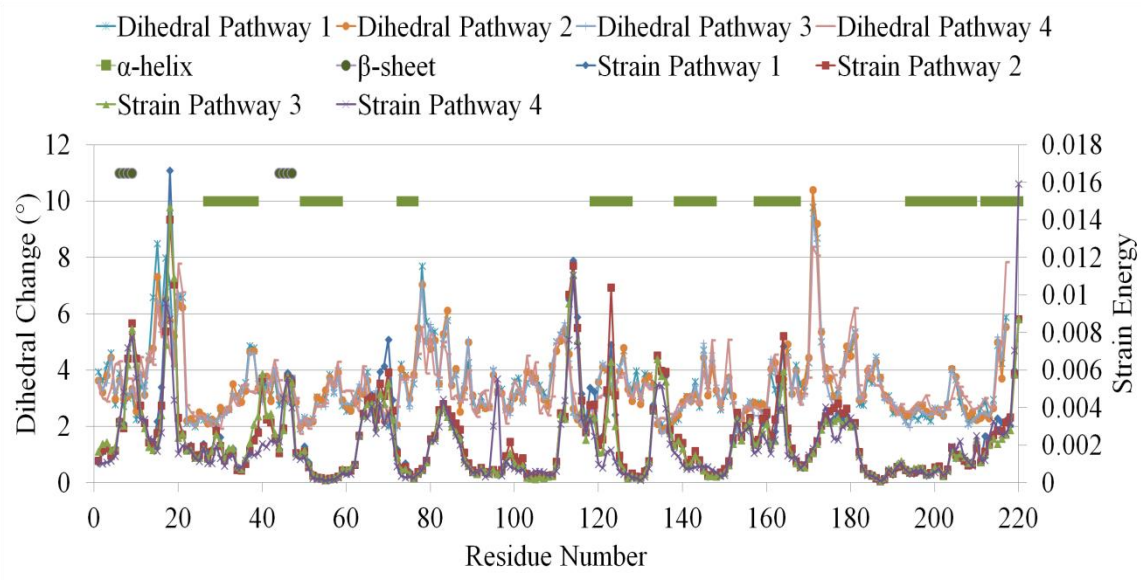


Figure 4.57. Strain energy and dihedral angle change for four pathways of GBP_2 system.

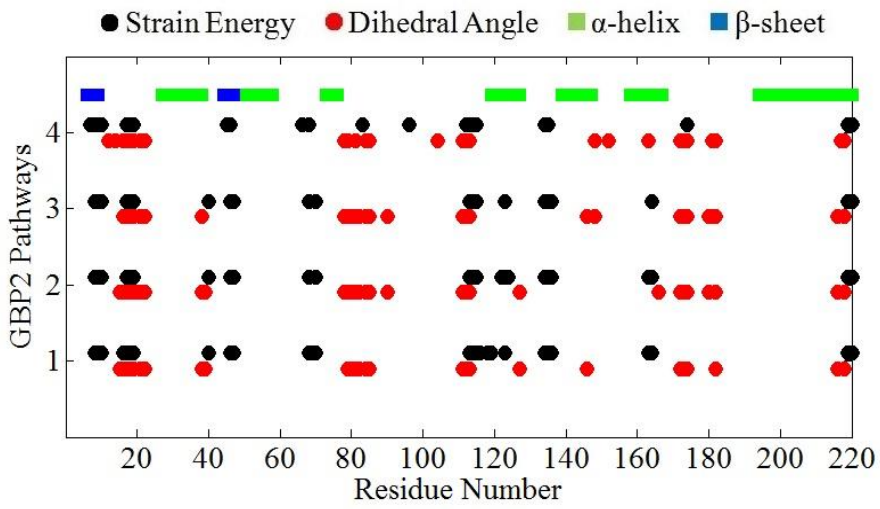


Figure 4.58. GBP_2 forward system results for cracking with a strain energy cutoff distance of 18 Å.

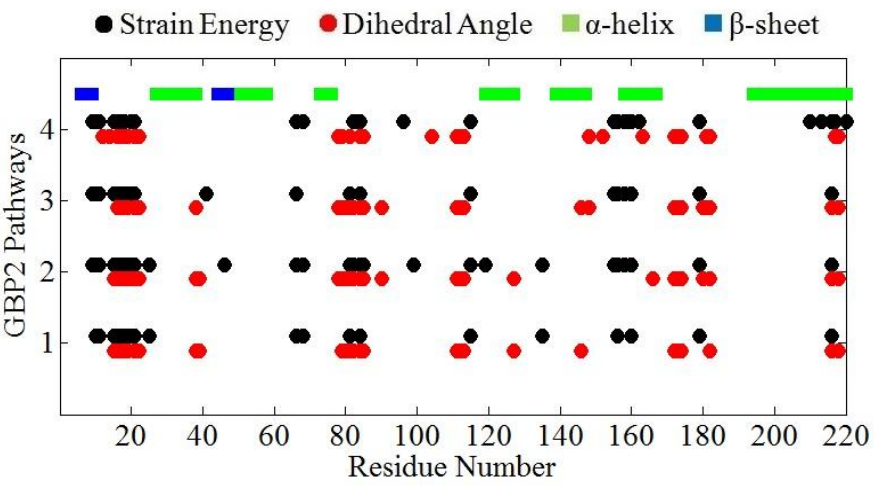


Figure 4.59. GBP_2 forward system results for cracking with a strain energy cutoff distance of 10 Å.

5. CONCLUSIONS AND RECOMMENDATIONS

Elastic network models are simple and efficient methods for studying conformational transitions of proteins. Anisotropic Network Model (ANM) forms a coarse-grained elastic network by connecting close-neighboring C^{α} atom pairs by harmonic springs. ANM provides slowest collective modes of motion relevant for defining function and conformational transitions. ANM_MC technique, which combines ANM collective moves with short MC simulations for energy minimization, provides biased dynamics to reach the target (closed) structures from the open state (Kantarci-Carsibasi *et al*., 2008). In fact, proteins displaying large conformational changes between two states may follow multiple pathways during the transition and computational techniques that provide unbiased pathways are necessary to visualize the realistic transition phenomena.

In this thesis, the WE sampling algorithm has been implemented to the ANM-MC simulation method for producing unbiased conformational transitions of proteins. The iterative methodology is described in detail in Chapter 3. At each iteration, the structures is deformed along the direction of a randomly selected mode among the slowest five modes that are produced by ANM. The procedure is then followed by energy minimization using local moves of the MC method. Sampling by WE algorithm enhances the observation of rare transitions and yields distinct and unbiased successful trajectories. This method has been applied to observe the conformational transitions of three proteins, namely AdK, CaM and GBP.

Two systems (AdK_1 and AdK_2) that use 18 Å cutoff distance in ANM calculations have final RMSD values 2.87 Å and 2.78 Å, respectively. Four other systems (AdK_3-AdK_6) with 10 Å ANM cutoff distance provide comparatively better results, such as AdK_4 reaching the target with an RMSD of 1.77 Å. Previously defined parameters such as FRET distances and inter-domain angles are used for comparison of transitions in these four AdK systems. Best pathways yield quite different domain movements during their transition. LID and AMP domains close together in AdK_3

system, while AdK_4 system follows the sequence of LID-AMP-LID partial closures during the transition. In AdK_5 and AdK_6 systems, the LID and AMP domains close first, respectively. In general, LID domain displays more dominant conformational changes in terms of closure compared to the AMP domain.

The parameters of the WE method have also been studied in thesis. Number of simulations per bin (M) for the WE method is doubled (M=10) in AdK_2, which helps to reach faster to the target state than AdK_1 system (M=5). Number of iterations (N_τ) for the WE method is increased to 10 in AdK_5 and AdK_6 systems, which increases the length of simulations and possibly provides more distinct pathways.

Reverse transition of AdK is also studied using three independent systems, in which 10 Å ANM cutoff distance is used in AdK_R2 and AdK_R3 and one uses 18 Å cutoff distance (AdK_R1). None of them displays a satisfactory approach to the target (open) structure (minimum RMSD is 4.18 Å). This may result from the fact that the closed crystal structure that is highly compact is used as the starting point. The approach may have been closer if one of the end structures of the forward runs have been used.

CaM and GBP systems are also simulated using two different cutoff distance values. CaM pathways are quite similar, while in GBP distinct pathways are formed at very early stages of the conformational transition. Minimum RMSDs for CaM and GBP systems are 3.44 Å and 1.92 Å, respectively.

Another aim of the study is to investigate cracking phenomena along the conformational transition of proteins. Previous studies claim that accumulation of strain energy in certain parts of the structure cause local unfolding of specific α-helical or β-strand regions. Local strain energies using two different cutoff distances and dihedral angle changes considering two alternative definitions are calculated for all pathways of AdK, CaM and GBP systems. The pathways in a specific simulation system do not display significant differences in terms of strain energy profiles and dihedral angle changes. Five out of eight previously reported hinge regions are highly strained for AdK. In terms of cracking, there are two regions common for highly strained regions

and dihedral angle changes, which corresponds to the third hinge of residues 58-63 and a slight peak around seventh hinge of residues 157-163. In CaM_1 system, two distinct regions display cracking in α-helix regions of residues 37 and 108-109. In contrast, no evidence is available for cracking in GBP systems, where loop regions only display high strain energy and dihedral angle change. The comparison of AdK, CaM and GBP transition provide significant differences in terms of structure and conformational transitions.

For future work, WE sampling algorithm can be applied to various proteins to investigate distinct pathways. WE parameters N_{bin}, M, N_{τ} may be further adjusted to improve the sampling enrichment. Other criteria can be selected as the progress coordinate, such as radius of gyration or FRET distances, rather than the RMSD in this study to determine the transition to target state. Even though coarse-grained ANM provides good results, defining atomistic details on protein structure would enhance the reliability of the results and their applications in terms of docking and drug design.

REFERENCES

Allen, R. J., P. B. Warren, P. R. Ten Wolde, 2005, "Sampling Rare Switching Events In Biochemical Networks", *Physical Review Letters*, Vol. 94, No. 1, pp. 018104 (1-4).

Arora, K., C. L. Brooks, 2007, "Large-scale Allosteric Conformational Transitions of Adenylate Kinase Appear to Involve a Population-shift Mechanism", *Proceedings of the National Academy of Sciences*, Vol. 104, No. 47, pp. 18496–18501.

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

Atilgan, A. R., A. O. Aykut, C. Atilgan, 2011, "Subtle pH Differences Trigger Single Residue Motions for Moderating Conformations of Calmodulin", *The Journal of Chemical Physics*, Vol. 135, No. 15, pp. 155102 (1–12).

Bahar, I., 1999, "Dynamics of Proteins and Biomolecular Complexes: Inferring Functional Motions from Structure", *Reviews in Chemical Engineering*, Vol. 15, No. 4, pp. 319–349.

Bahar, I., A. R. Atilgan, M. C. Demirel, B. Erman, 1998, "Vibrational Dynamics of Proteins: Significance of Slow and Fast Modes in Relation to Function and Stability", *Physical Review Letters*, Vol. 80, No. 12, pp. 2733–2736.

Bahar, I., A. R. Atilgan, B. Erman, 1997, "Direct Evaluation of Thermal Fluctuations in Protein Using a Single Parameter Harmonic Potential", *Folding & Design*, Vol. 2, No. 3, pp. 173–181.

Bahar, I., B. Erman, R. L. Jernigan, A. R. Atilgan, D. Covell, 1999, "Collective Motions of HIV-1 Reverse Transcriptase: Examination of Flexibility and Enzyme Function", *Journal of Molecular Biology*, Vol. 285, No. 3, pp. 1023–1037.

Bahar, I., R. L. Jernigan, 1997, "Inter-Residue Potentials in Globular Proteins and the Dominance of Highly Specific Hydrophilic Interactions at Close Separations", *Journal of Molecular Biology*, Vol. 266, No. 1, pp. 195–214.

Bahar, I., R. L. Jernigan, 1998, "Vibrational Dynamics of Transfer RNAs. Comparison of the Free and Enzyme-Bound Forms", *Journal of Molecular Biology*, Vol. 281, No. 5, pp. 871–884.

Bahar, I., R. L. Jernigan, 1999, "Cooperative Fluctuations and Subunit Communication in Tryptophan Synthase", *Biochemistry*, Vol. 38, No. 12, pp. 3478–3490.

Bahar, I., A. J. Rader, 2005, "Coarse-Grained Normal Mode Analysis in Structural Biology", *Current Opinion in Structural Biology*, Vol. 15, No. 5, pp. 586–592.

Beckstein, O., E. J. Denning, J. R. Perilla, T. B. Woolf, 2009, "Zipping and Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open↔Closed Transitions", *Journal of Molecular Biology*, Vol. 394, No. 1, pp. 160–176.

Berg, J. M., J. L. Tymoczko, L. Stryer, 2002, *Biochemistry*, 5th edition, Freeman, New York.

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

Bhatt, D., D. M. Zuckerman, 2010, "Heterogeneous Path Ensembles for Conformational Transitions in Semiatomistic Models of Adenylate Kinase", *Journal of Chemical Theory and Computation*, Vol. 6, No. 11, pp. 3527–3539.

Chattopadhyaya, R., W. E. Meador, A. R. Means, F. A. Qiocho, 1992, "Calmodulin Structure Refined at 1.7 Å Resolution", *Journal of Molecular Biology*, Vol. 228, No. 4, pp. 1177–1192.

Dellago, C., P. G. Bolhuis, F. S. Csajka, D. Chandler, "Transition Path Sampling and the Calculation of Rate Constants", 1998, *Journal of Chemical Physics*, Vol. 108, No. 5, pp. 1964–1977.

Dobbins, S. E., V. I. Lesk, J. E. Sternberg, 2008, "Insights into Protein Flexibility: The Relationship Between Normal Modes and Conformational Change upon Protein-Protein Docking", *Proceedings of the National Academy of Sciences*, Vol. 105, No. 30, pp. 10390–10395.

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

Fischer, E., 1894, "The Key and Lock Principle", *Berichte der Deutschen Chemischen Gesellschaft*, Vol. 27, No. 5, pp. 2986–3222.

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

Haliloglu, T., I. Bahar, B. Erman, 1997, "Gaussian Dynamics of Folded Proteins", *Physical Review Letters*, Vol. 79, No. 16, pp. 3090–3093.

Haliloglu, T., I. Bahar 1998, "Coarse-Grained Simulations of Conformational Dynamics of Proteins: Application to Apo-Myoglobin", *Proteins,* Vol. 31, No. 3, pp. 271–281.

Haliloglu, T., 1999, “Coarse-Grained Simulations of the Conformational Dynamics of Proteins”, *Computational and Theoretical Polymer Science*, Vol. 9, No. 3-4, pp. 255–260.

Harvey, S. C., H. A. Gabb, 1993, “Conformational Transitions Using Molecular Dynamics with Minimum Biasing”, *Biopolymers*, Vol. 33, No. 8, pp. 1167–1172.

Henzler-Wildman, K. A., M. Lei, V. Thai, S. J. Kerns, M. Karplus, D. Kern, 2007, “A Hierarchy of Timescales in Protein Dynamics is Linked to Enzyme Catalysis”, *Nature*, Vol. 450, No. 7171, pp. 913–916.

Hsiao, C. D., Y. J. Sun, J. Rose, B. C. Wang, 1996, “The Crystal Structure of Glutamine-binding Protein from Escherichia Coli”, *Journal of Molecular Biology*, Vol. 262, No. 2, pp. 225–242.

Huber, G. A., S. Kim, 1996, “Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions”, *Biophysical Journal*, Vol. 70, No. 1, pp. 97–110.

Kalman, M., N. Ben-Tal, 2010, “Quality Assessment of Protein Model-Structures Using Evolutionary Conservation”, *Bioinformatics*, Vol. 26, No. 10, pp. 1299–1307.

Kantarci-Carsibasi, N., T. Haliloglu, P. Doruker, 2008, “Conformational Transition Pathways Explored by Monte Carlo Simulation Integrated with Collective Modes”, *Biophysical Journal*, Vol. 95, No. 12, pp. 5862–5873.

Kasprzak, A. A., R. Takashi, 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, No. 12, pp. 4512–4522.

Kleywegt, G. J., 2009, "On Vital Aid: the Why, What and How of Validation", *Acta Crystallographica Section D Biological Crystallography*, Vol. 65, No. 2, pp. 134–139.

Koshland, D. E., 1958, "Application of a Theory of Enzyme Specificity to Protein Synthesis", *Proceedings of National Academy of Science*, Vol. 44, No. 2, pp. 98–104.

Kubitzki, M. B., B. L. de Groot, 2008, "The Atomistic Mechanism of Conformational Transition in Adenylate Kinase: A TEE-REX Molecular Dynamics Study", *Structure*, Vol. 16, No. 8, pp. 1175–1182.

Kurt, N., T. Haliloglu, 1999, "Conformational Dynamics of Chymotrypsin Inhibitor 2 by Coarse-Grained Simulations", *Proteins: Structure, Function, and Genetics*, Vol. 37, No. 3, pp. 454–464.

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

Leach, A. R., 2001, *Molecular Modelling Principles and Applications*, 2nd edition, Prentice Hall, Essex.

Lei, M., M. I. Zavodsky, L. A. Kuhn, M. F. Thope, 2004, "Sampling Protein Conformations and Pathways", *Journal of Computational Chemistry*, Vol. 25, No. 9, pp. 1133–1148.

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

Liwo, A., C. Czaplewski, S. Oldziej, H.A. Scheraga, 2008, "Computational Techniques for Efficient Conformational Sampling of Proteins", *Current Opinion in Structural Biology*, Vol. 18, No. 2, pp. 134–139.

Ma, J., 2005, "Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes", *Structure*, Vol. 13, No. 3, pp. 373–380.

Maragakis, P., M. Karplus, 2005, "Large Amplitude Conformational Change in Proteins Explored with a Plastic Network Model: Adenylate Kinase", *Journal of Molecular Biology*, Vol. 352, No. 4, pp. 807–822.

Miyashita, O., J. N. Onuchic, P. G. Wolynes, 2003, "Nonlinear Elasticity, Proteinquakes, and the Energy Landscapes of Functional Transitions in Proteins", *Proceedings of National Academy of Science*, Vol. 100, No. 22, pp. 12570–12575.

Müller, C. W., 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, No. 1, pp. 159–177.

Müller, C. W., G. J. Schlauderer, J. Reinstein, G. E. Shulz., 1996, "Adenylate Kinase Motions During Catalysis: an Energetic Counterweight Balancing Substrate Binding", *Structure*, Vol. 4, No. 2, pp. 147–156.

Rader, A. J., C. Chennubhotla, L. W. Yang, I. Bahar, 2006, "The Gaussian Network Model: Theory and Applications", *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, Chapter 3, Chapman & Hall/CRC, Florida.

Schlitter, J., M. Engels, P. Krüger, 1994, "Targeted Molecular Dynamics: A New Approach for Searching Pathways of Conformational Transitions", *Journal of Molecular Graphics*, Vol. 12, No. 2, pp. 84–89.

Sun, Y. J., J. Rose, B. C. Wang, C. D. Hsiao, "The Structure of Glutamine-binding Protein Complexed with Glutamine at 1.94 AÊ Resolution: Comparisons with other Amino Acid Binding Proteins", *Journal of Molecular Biology*, Vol. 278, No. 1, pp. 219–229.

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

Tirion, M. M., 1996, "Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis", *Physical Review Letters*, Vol. 77, No. 9, pp. 1905–1908.

Van der Spoel, D., B. L. de Groot, S. Hayward, H. J. C. Berendsen, H. J. Vogel, 1996, "Bending of the Calmodulin Central Helix: A Theoretical Study", *Protein Science*, Vol. 5, No. 10, pp. 2044–2053.

Van der Vaart, A., M. Karplus, 2005, "Simulation of Conformational Transitions by the Restricted Perturbation–Targeted Molecular Dynamics Method", *Journal of Chemical Physics*, Vol. 122, No. 11, pp. 114903–114909.

Van der Vaart, A., M. Karplus, 2007, "Minimum free energy pathways and free energy profiles for conformational transitions based on atomistic molecular dynamics simulations", *Journal of Chemical Physics*, Vol. 126, No. 16, pp. 164106–164123.

Van Erp, T. S., D. Moroni, P. G. Bolhuis, 2003, A Novel Path Sampling Method for the Calculation of Rate Constants, *Journal of Chemical Physics*, Vol. 118, No. 17, pp. 7762–7774.

Vertessy, B. G., V. Harmat, Z. Bocskei, G. Naray-Szabo, F. Orosz, J. Ovadi, 1998, "Simultaneous Binding of Drugs with Different Chemical Structures to Ca^{2+}-Calmodulin: Crystallographic and Spectroscopic Studies", *Biochemistry*, Vol. 37, No. 44, pp. 15300–15310.

Weber, G., 1972, “Ligand Binding and Internal Equilibria in Proteins”, *Biochemistry*, Vol. 11, No. 5, pp. 864–878.

Whitford, P. C., O. Miyashita, Y. Levy, J. N. Onuchic, 2007, “Conformational Transitions of Adenylate Kinase: Switching by Cracking”, *Journal of Molecular Biology*, Vol. 366, No. 5, pp. 1661–1671.

Whitford, P. C., S. Gosavi, J. N. Onuchic, 2008, “Conformational Transitions in Adenylate Kinase Allosteric Communication Reduces Misligation”, *Journal of Biological Chemistry*, Vol. 283, No. 4, pp. 2042–2048.

Yang, C., G. S. Jas, K. Kuczera, 2004, “Structure, Dynamics and Interaction with Kinase Targets: Computer Simulations of Calmodulin”, *Biochimica et Biophysica Acta*, Vol. 1697, No. 1-2, pp. 289–300.

Zhang, B. W., D. Jasnow, D. M. Zuckerman, 2007, “Efficient and Verified Simulation of a Path Ensemble for Conformational Change in a United-Residue Model of Calmodulin”, *Proceedings of National Academy of Science*, Vol. 104, No. 46, pp. 18043–18048.

Zhang, B. W., D. Jasnow, D. M. Zuckerman, 2010, “The ‘Weighted Ensemble’ Path Sampling Method is Statistically Exact for a Broad Class of Stochastic Processes and Binning Procedures”, *The Journal of Chemical Physics*, Vol. 132, No. 5, pp. 54–107.

Zuckerman, D. M., T. B. Woolf, 1999, “Dynamic Reaction Paths and Rates Through Importance-Sampled Stochastic Dynamics”, *The Journal of Chemical Physics*, Vol. 111, No. 21, pp. 9475–9484.