AN INTEGRATED COMPUTATIONAL AND EXPERIMENTAL APPROACH TO ALLOSTERIC CONTROL MECHANISM OF BIOMOLECULAR PROCESSES

by

Fidan Sümbül

B.S., Department of Chemical Engineering, Boğaziçi University, 2006

M.S., Department of Chemical Engineering, Boğaziçi University, 2008

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

Graduate Program in Chemical Engineering
Boğaziçi University
2016

ACKNOWLEDGEMENTS

Ph.D. is always the result of years of dedication and hard work, but equally or more important, it is the result of the support of people around where they allow us to follow our path and successfully finalize our works. For this reason, I want to thank the people that participated and supported me to complete this Ph.D. dissertation.

There are people in my work environment that I want to thank: First and the foremost, I would like to offer my sincerest gratitude to my supervisor Prof. Türkan Haliloğlu whose experience, leadership, friendship, sometimes parenthood, encouragement and continuous support throughout my graduate studies enables me to complete this dissertation. I really want to thank her for all the scientific things I learned from her during whole graduate level studies. I appreciate her help as a huge impulse to complete this work, mainly when things look hard to complete.

I also owe thanks to the members of my thesis committee, Assoc. Prof. Demet Akten Akdoğan, Prof. Pemra Doruker Turgut, Assist. Prof. Mehmet Gönen, and Assoc. Prof. Hamdi Torun, who has read and commented on my thesis.

I want to give a special acknowledgement to Assoc. Prof. Hamdi Torun for the discussions and for introducing me to Atomic Force Microscopy (AFM). I would like to offer my gratitude for your time to teach me a way to explore with systematic detail the working principles of AFM and design of an experimental research. Thanks for the patience and understanding with my dreamy-impatient energetic way to work during my Ph.D.

I also want to give a special acknowledgement to Assist. Prof. Mehmet Gönen for his friendship and very precious guidance and encouragement during my whole research carrier. His working style and enterprising personality have always inspired me.

I want to specially thank to Dr. Felix Rico, who will be my future supervisor. It

will be an honor and privilege for me to study such a dedicated and talented scientist in my postdoctoral carrier. I offer my deep appreciation to Prof. Ruth Nussionov, for her endless positive energy, support and encouragement. It has been both an honor and a privilege for me to be with the super-hero of science.

I want to acknowledge Prof. Nesrin Özören and Ali Can Sahillioğlu, for their collaboration during my research on ASC protein. I am sure that our collaboration has improved my scientific knowledge. I am indebted to Dr. Saliha Ece Acuner Özbabacan for our pleasant and fruitful collaboration. I also thank to Melis Önel for her collaboration. It has been a very valuable and a unique experience for me to work with them.

I want to specially thank my colleagues, friends and office mates, Dr. Burcu Aykaç Fas, Burçin Acar, Dr. Saliha Ece Acuner Özbabacan, and Zeynep Erge Akbaş. All the conversations (scientific and not scientific), good moments, and supportive spirit in hard moments along the PhD have been very important and crucial company for me to make it to the final. I thank you all a lot for your friendship!.

I would like to thank Alper Başpınar for his friendship and encouragement in my good and bad times. Our friendship is not very old but very strong. I would like to also thank Dr. Giray Enkavi for his useful suggestions and patient help with this research.

I am very grateful to all the former and present members of Department of Chemical Engineering for their suggestions and friendships. To all my office mates and colleagues, especially; Dr. Pemra Özbek Sarıca, Dr. Nevra Özer, Dr. Özge Kürkçüoğlu Levitas, Dr. Zeynep Kürkçüoğlu Soner, Dr. Seren Soner, Dr. Arzu Uyar, Serdar Özsezen, Şefik Kerem Ovalı, Gökçehan Kara. It has been a luck for me to meet you guys as a scientist and as a friend. I really hope that we find each other in the future and I will remember these times with you as very special in my life. I wish you the best of luck in the next stages of your career and I admire you both as a person and as a scientist.

Special thanks to Canan Dedeoğlu who was like a mother and a trusted good friend throughout all my graduate life. And also thanks to Yüksel Gülten for his tireless affords four our well-being and his loud laugh.

It was a great opportunity for me that this work has been supported by the Boğaziçi University B.A.P. projects (6524/11A05M4 and 8424/14A05M6), TÜBITAK projects (110T088 and 112T569), and Betil Fund.

In my personal life outside university there are a quite large number of people who supported me a lot along my PhD. These people are the spiritual collaborators of this thesis. I hope I have the words to thank you all!

I thank the members of my family: my mother Pakize Sümbül, my father Nusrettin Sümbül, my brothers Ünal Sümbül, Güven Sümbül, Savaş Sümbül, my sisters Nursen Güngörmez, Sevda Sümbül, Biler Kimençe, Eda Demir, my nephews Gamze Güngörmez, Gökdeniz Sümbül, Esma Demir, Arda Güngörmez, Buse Sümbül, Berna Ece Sümbül, Alp Kerem Kimençe, Taha Demir, Azra Tuğçe Sümbül. With their unconditional support and love only an outstanding family could provide, I have been able to keep focused on my goals in life. My family has been a pillar of strength for me from the day I started education and has always been there for me.

I feel very lucky to have friends like Sevinç Şimşek, Esin Yiğit Balcı and Murat Balcı, Evrim Itır Karaç, Gülefşan Gönen, Tülin Keçeli Çubuk and Alkan Çubuk. Thanks to their generous spirit and guidance I would like to give my warmest thanks to Bülent Güngör who allowed me to stay cheerful and strong even at difficult times with his unwavering support.

ABSTRACT

AN INTEGRATED COMPUTATIONAL AND EXPERIMENTAL APPROACH TO ALLOSTERIC CONTROL MECHANISM OF BIOMOLECULAR PROCESSES

Proteins are flexible and dynamic in nature and can undergo structural rearrangements to perform their functions. The motivation of this thesis is to develop a novel integrated computational and experimental approach to understand the allosteric control of conformational transitions and molecular recognition. In the first part, a novel methodology called Collective Modes Bias Exchange Metadynamics (CM-BexMetaD) was developed and transition and allosteric control of transition of Adenylate kinase was investigated. In the second part, allosteric control of binding of proteins were studied through Elastic Network Models (ENM), Molecular Dynamics (MD), in-vitro imaging and Dynamic Force Spectroscopy (DFS) using AFM. A strong association between the hinge positions of global modes and allosteric mutations was shown by a large scale statistical analysis. The binding behavior of pyrin domain (PYD) and assembly formation of ASC protein was studied in-silico supported by in-vitro experiments and the results showed that, the ASC speck is an organized structure and the interaction of the domains are controlled via hinge residues. Further, the allosteric control of binding have been elucidated on the kinesin- aÃ-tubulin and Rac1-PAK1 interactions by prediction of hinge residues that would affect the binding behavior of proteins and then the resulting change in the binding energy landscape were investigated by DFS and last the mechanistic explanation of the alteration of the allosteric communication network was studied via MD studies. Both kinesin- $\alpha\beta$-tubulin and Rac1-PAK1 protein complexes have alternative dissociation pathways where hinge residues as mechanistically key sites that allosterically control the binding behavior of molecules.

ÖZET

BIYOMOLEKÜLER SÜREÇLERIN ALOSTERIK KONTROL MEKANIZMASINA HESAPLAMALI VE DENEYSEL ENTEGRE BIR YAKLASIM

Proteinler esnek ve dinamik yapılardır ve işlevsel yapısal değişimler gösterirler. Bu tezin ana motivasyonu, konformasyonel geçişlerin ve moleküllerin bağlanma süreçlerinin alosterik kontrolünü anlamaya yönelik hesaplamalı ve deneysel entegre bir yöntem geliştirmektir. Oncelikle, kararlı yapılar arasındaki geçişler ve serbest enerji yüzeylerini belirlemeye yönelik, Kolektif Modlar Bayas Değişimli Metadinamik (CM-BexMetaD) olarak adlandırılan özgün bir yöntem geliştirilmiştir. Adenilat kinaz enziminin birçok ara yapısı belirlenmiş ve menteşe bölgelerinin alosterik etkisi gösterilmiştir. Ikinci bölümde, ENM, moleküler dinamik (MD), in-vitro görüntüleme ve Atomik Kuvvet Mikroskopu-Dinamik Kuvvet Spektroskopisi (AFM-DFS) yöntemleri ile proteinlerin bağlanma mekanizmalarının alosterik kontrolu üzerine çalışılmıştır. SKEMPI veritabanindaki mutasyonlar ile menteşe bölgeleri arasındaki kuvvetli ilişki, geniş çaplı istatistiksel bir analiz ile gosterilmistir. In-silico ve in-vitro deneyler ile PYD ve ASC proteinlerindeki menteşe bölgelerindeki mutasyonlarin bağlanma davranışı üzerindeki etkisi çalışılmıştır. ASC zerreciklerinin en az iki seviyeli organize yapılar olduğu ve menteşe bölgeleri aracılığı ile kontrol edildiği belirlenmiştir. Bunlara ek olarak, AFM-DFS yöntemini kapsayan hibrit bir yaklaşım geliştirilerek, protein-protein bağlanma mekanizmalarinin menteşe bölgeleri aracılığı ile alosterik olarak kontrol edildiği, kinezin-$\alpha\beta$-tübülin ve Rac1-PAK1 etkileşimleri üzerinden ortaya konulmuştur. Alosterik haberleşme ağındaki değişimin mekanik izahı ise MD simülasyonları ile belirlenmiştir. Elde edilen sonuçlar ile kinezin-$\alpha\beta$-tübülin ve Rac1-PAK1 kompleks yapılarının alternatif ayrılma patikalarına sahip oldukları ve menteşe bölgelerinin bağlanma davranışlarını alosterik olarak kontrol eden anahtar bölgeler olduğu ortaya konulmuştur.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\tilde{A}$	Angstrom
B_i	Debye-Waller or temperature factor of site i
C_a	Alpha-carbon atom
D	Diagonal matrix
E	Internal Energy
E_{boost}	Boost Energy
$E_{treshold}$	Threshold Energy
f	Force
$f*$	Most probable rupture force
G	Free Energy
H,	Hessian matrix
H_{ij}	Super element matrix of the hessian matrix
H^{-1}	Inverse of the super element matrix
h	Planck's constant
I	Identity matrix
k	Spring constant
k_B	Boltzmann's constant
$kcal$	kilo calorie
k_{off}	Dissociation rate
l_i	Bond length connecting atoms i-1 and I
l_i^S	Bond length connecting backbone and sidechain
m	Number of eigenvectors and corresponding conformations
N	Total number of residues
N_{CV}	Number of collective variables
n	Number of residues
nm	Nanometer
pN	Pico Newton
t_c	Cutoff radius
Q	Conformational Reaction Coordinate

R	Rotation matrix
R_i	Position vector of site I
R_{ij}	Distance between site i and j
r	Atomic conformation
r_{cut}	Cutoff radius
r_f	Loading rate
S_{conf}	Conformational Entropy
s	Second
s	Sensitivity
T	Absolute temperature
t	Transpose
t	Time
tr	Trace
U	Eigenvector
V	Potential energy
V_b	Bias potential
w	Gaussian hill height
α	Acceleration
α	Prefactor
β	Beta strand
γ	Force constant
ζ	Collective variable
τ_G	Time interval
ΔE	Internal energy difference
ΔG	Free energy difference
ΔG^o	Activation energy
ΔS_{conf}	Entropy difference
ΔR_i	Fluctuation of vector of site i
Γ	Kirchoff matrix
$< \Delta R_i^2 >$	Mean-square fluctuation of site i

x_β	Barrier width
ν	Velocity

LIST OF ABBREVIATIONS

AdK	Adenylate Kinase
ADK	Adenosine Diphosphate
AFM	Atomic Force Microscopy
ASC	Apoptosis Associated Speck-like Protein Containing
AMP	Adenosine Monophosphate
ANM	Anisotropic Network Model
ATP	Adenosine Triphosphate
aMD	Accelerated Molecular Dynamics
avg	Average
BSA	Bovine Serum Albumin
C	Carboxyl Terminus
CARD	Caspase Activation and Recruitment Domain
CG	Coarse Grained
CM-BexMetaD	Collective Modes Bias Exchange Metadynamics
CMW	Cytoskeleton Motor Werks
COSMIC	Catalogue of Somatic Mutations in Cancer
Cryo-EM	Cryo-Electron Microscopy
CV	Collective Variable
cMD	Conventional Molecular Dynamics
coMD	Collective Molecular Dynamics
cog	Center of Geometry
DFS	Dynamic Force Spectroscopy
DI	Deionized
DNA	Deoxyribonucleic Acid
DTT	1,4-Dithiothreitol
d	Distance Vector
diag	Diagonal of Matrix
E	Energy
EDTA	Ethylenediaminetetraacetic Acid

EGTA	Ethylene Glycol Tetraacetic Acid
EM	Electron Microscopy
ENM	Elastic Network Model
FEL	Free Energy Landscape
FES	Free Energy Surface
FRET	Fluorescence Resonance Energy Transfer
GAP	GTPase Activating Protein
GDP	Guanosine Diphosphate
GEF	Guanine Nucleotide Exchange Factor
GNM	Gaussian Network Model
GSH	Glutathione
GST	Glutathione S-Transferase
GTB	General Tubulin Buffer
GTP	Guanosine Trihosphate
H-REMD	Hamiltonian Replica Exchange Molecular Dynamics
hGH	Human Growth Hormone
KNF	Koshland-Nemety-Filmer Model
MD	Molecular Dynamics
MC	Monte Carlo
MT	Microtubule
MWC	Monod-Wyman-Changeux Model
mRNA	Messenger RNA
MSF	Mean Square Fluctuations
N	Amino Terminus
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NAMD	Nanoscale Molecular Dynamics
NMA	Normal Mode Analysis
NMR	Nueclar Magnetic Resonance
NL	Neck Linker
NTA	Nitrilotriacetic Acid
O	Overlap Value

PAK1	P21-Activated Kinase 1
PCA	Principal Component Analysis
PBS	Phosphate-Buffered Saline
PEG	Polyethylene Glycol
PDB	Protein Data Bank
PIPES	Piperazine-N,N'-Bis(2-Ethanesulfonic Acid)
PMF	Potential of Mean Force
PYD	Pyrin Domain
R	Relaxed Form
Rac1	Ras-Related C3 Botulinum Toxin Substrate 1
REMD	Replica Exchange Molecular Dynamics
RMSD	Root-Mean-Square Deviation
RNA	Ribonucleic Acid
SAXS	Small-Angle X-Ray Scattering
SEM	Standard Error of Mean
SKEMPI	Structural Dataset of Kinetics and Energetics of Mutant Protein Interactions
SMD	Steered Molecular Dynamics
T	Tense Form
TMD	Targeted Molecular Dynamics
T-REMD	Temperature Replica Exchange Molecular Dynamics
TST	Transition State Theory
Tris	Tris(Hydroxymethyl)Aminomethane
US	Umbrella Sampling
WHAM	Weighted Histogram Analysis Method
WTE	Well Tempered Ensemble
WTmetaD	Well-Tempered Metadynamics
wt	Wild Type
VMD	Visual Molecular Dynamics

1. INTRODUCTION

Proteins are building blocks of essential biological processes and they interact with other proteins and/or molecules such as small peptides, DNA, RNA, etc. to perform their functions within the cell. The correct binding between these molecules requires specific interaction between two or more of them through non-covalent bonding, which is refereed as molecular recognition. In recent years, dynamic models of molecular recognition are recognized rather than classical rigid lock-and-key binding model. Induced fit and conformational selection have recently appeared as two major models [1]. It is however now understood that the molecular recognition involves both induced fit and conformational selection mechanism and it thus requires detailed understanding of conformational flexibility and dynamics of the biomolecules over large time scales.

Proteins undergo local to global conformational changes to perform their functions. The conformational ensemble of proteins, which is related to their free energy landscape, needs to be explored to decipher the mechanism underlying biological function. Excursions of stable minima and the barriers in between are associated with functional dynamics of proteins. Even though under ambient conditions most proteins fold into three-dimensional structures (their native conformation), dynamical transitions between different conformations are essential to perform specific tasks. To this, computational and experimental techniques have evolved to explore conformational ensemble of proteins. Both complementing each other, bridging of computational tools with experimental techniques is still an important challenge.

Computer simulations of biomolecules have an enormous growth since the first atomistic biomolecular simulation published in 1977 [2]. Atomistic molecular dynamics (MD) simulations utilizing classical mechanics have great contribution to understanding the dynamic nature of biomolecules. However, MD simulations have two main limitations; despite limited time scales it has high computational cost and force fields may not be inaccurate for large time scales [3]. The time resolution of interesting

biomolecular processes is significantly longer than conventional MD simulations [4]. These limitations of MD simulations has led to the development of either advanced methods to enhance the sampling of high probability metastable states and rare events which requires crossing of high free energy barriers [5] or coarse grained methods (such as Elastic Network Models, ENM) to reduce the computational cost and to study large systems [6-9].

Experimental techniques may provide direct and indirect evidences for conformations of proteins and assemblies and their dynamics, such as X-ray crystallography, Nuclear Magnetic Resonance (NMR), electron microscopy (EM) or small-angle X-ray scattering (SAXS). However, these techniques provide rather static images of populated conformational states or sometimes not feasible for large or highly dynamic systems [10]. Within the last quarter of century, another experimental technique, Atomic Force Microscopy (AFM) has come into use either to obtain topographic images of molecules or to study the mechanics of molecular recognition. Since AFM can operate in a fluid medium, which is very close to the physiological conditions of biomolecules, and does not require any special treatment of biological sample such as crystallization or fixation, it makes possible to simultaneously and directly observe both the structure and dynamics of molecules in action.

The motivation of this thesis is to develop a novel integrated computational and experimental approach to understand the allosteric control of large conformational transitions and molecular recognition in proteins. In the first part of this thesis, a novel ENM driven metadynamics methodology called Collective Modes Bias Exchange Metadynamics (CM-BexMetaD) is developed to explore the mechanism of conformational transitions between stable states by constructing the underlying free energy landscape. As a case study, the nature of conformational transition pathways and the determinants of the allosteric dissemination of any chemical and physical perturbation were explored on a well-studied protein Adenylate kinase (AdK) Chapter 7. In the second part, mutation induced allosteric control of binding mechanism of proteins were studied by a sequence of studies involving computational and experimental means. ENM analysis and molecular simulations provided dynamic information for prediction

of mechanistically key sites that have an allosteric affect on binding whereas molecular simulations provided clues to see the effect of the perturbations at these sites on the ensemble of conformations and functional implications. On the other hand, in-vitro imaging (via inverted fluorescent microscope and confocal microscopy), western blotting, fluorescence resonance energy transfer (FRET) and Dynamic Force Spectroscopy (DFS) using AFM experiments of both wild type and site-directed mutants required measurement and observation of the effect of these functional sites. Specifically; a strong association between the hinge positions of global modes and allosteric mutations that lead to high binding affinity changes determined by alanine scanning mutagenesis was shown by a large scale statistical analysis over the data present in the Structural Kinetic and Energetic Database of Mutant Protein Interactions (SKEMPI) database. The mechanism of allosteric dynamics was further analyzed by changes in local interactions and global dynamics of human growth hormone (hGH). Moreover, the effect of perturbation at a hinge site on binding behavior of pyrin domain (PYD) and assembly formation of apoptosis associated speck-like protein containing CARD (ASC) was studied in-silico and in-vitro experimentation Chapter 8. Functionally important hinge residues that would affect the binding behavior of protein pairs were predicted via ENM studies and the resulting change in the binding energy landscape were investigated by a dynamic force spectroscopy (DFS) approach using AFM. The mechanistic explanation of the alteration of the allosteric communication network was studied via molecular dynamics studies. This hybrid protocol was applied to kinesin-tubulin Chapter 9 and Rac1-PAK1 Chapter 10 interactions.

2. PROTEIN STRUCTURE, FUNCTION AND DYNAMICS

2.1. Proteins are Vital Components of Living Organisms

The estimated number of species living on Earth is in the millions with a huge diversity of organism that are all made of cells. Alongside of being able to produce energy, grow, divide, move, come together to build complex organizations, cells can create a different inner environment than the outer in terms of physical and chemical properties to confer uniqueness to the organisms they build, which makes them vital for maintaining life [11]. Although all living organism are made of same atoms, molecular compositions are different among biological cells. Cells contain large number of highly complex macromolecules. There are four types of macromolecules: proteins, carbohydrates, nucleic acids, and lipids, which are responsible for all life processes.

Proteins are the most abundant macromolecules within the cell and exist in large number. Proteins are responsible from many biological processes in living organisms such as catalysis of metabolic processes, energy transfer, gene expression, transport of molecules/solutes between and across the cell, cellular communication, molecular recognition, defense, etc. [11]. This numerous roles of proteins make them important in many metabolic diseases, cancer, immunology, apoptosis, etc. Therefore, proteins form % 80 of current pharmaceutical targets [12].

2.2. Protein Structure and Function

Proteins are polypeptide chains translated from mRNA into a linear chain of amino acids, which is called primary structure. There are 20 natural amino acids and their sequences are determined according to the information stored in the nucleotide sequence of the corresponding genes. The specific sequence of amino acids creates a unique pattern of interatomic forces that leads to folding into specific three-dimensional

structures to carry out their functions. This functional form of the proteins called their *nativeconformations* [13], i.e. the tertiary structure, where the second level of protein structure such as helices, sheets, loops, etc. is called *secondarystructures*. The packing of these local secondary structural elements into three-dimensional shapes stabilized by hydrophobic interactions and some local bonds are called *tertiarystructures*. The assembly of more than one tertiary structure that is able to function as a single protein complex [14] is the last level of protein structure named as *quaternarystructure*. Figure 2.1 shows the different level of protein structures.

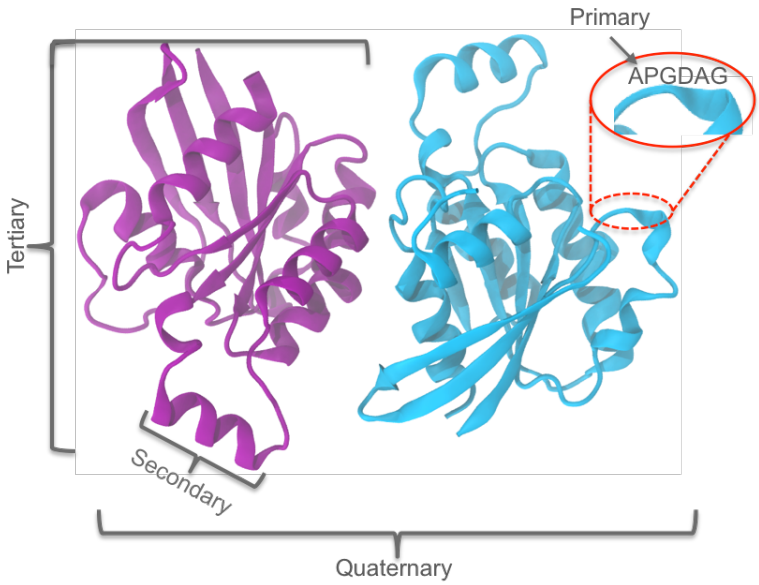


Figure 2.1. The Four Hierarchical Levels of Protein Structure. The Levels are Illustrated on the Structure of Rac1 Dimer Protein.

In addition to folding into its functional tertiary structures, proteins can undergo structural rearrangements-conformational transitions- in response to external stimuli and/or interaction with different binding partners such as ligand binding or allosteric mutations. Proteins are highly dynamical objects and the detailed molecular characterization of these structural changes assists further understanding the underlying mechanisms of their functions [15]. The structure-function relation is at the center of *structuralbiology* field.

2.3. Protein Structural Dynamics

Until the middle of the twentieth century, the scientific community thought of proteins as static objects, following Emil Fischer's lock-and-key model [16]. The awareness that proteins can undergo conformational changes to create better compliance between their structures and those of their substrates led to the development of induced-fit [17-18] and conformational selection[19,20]. Both induced-fit and conformational-selection mechanisms include exploration of energy landscape that often contains multiple stable conformational states in equilibrium. The induced-fit model describes a ligand-induced protein flexibility, whereas the conformational-selection model describes the preexistence of unbound form of the protein in an ensemble of conformations [21]. Since proteins are highly dynamic entities, the current view of protein structural dynamics is more complex than those suggested theories but appear as the combination of these suggested models [22, 23]. The protein dynamics are described by the motion of a single atom to group of several atoms within wide range of time scale (10^{-15}-10^{4} s), amplitude (1-100$\tilde{A}$) and energy level (0.1-100 kcal/mol), as depicted in Table 2.1 and Figure 2.2.

Protein dynamics includes many types of motion such as hinge motions, rotations, translations, and even the folding/unfolding of some secondary structural elements. Hinge motions are the most common movements in proteins, which occur on the 10^{-9} s time scale and wide range of amplitude and energy levels. Hinge motion is usually responsible for a specific function, such as conversion of chemical energy to mechanical energy in motor proteins, transduction of allosteric changes and opening/closing of active sites in enzymes, opening/closing motion of channels, etc. [11].

Protein dynamics can be considered at two levels; folding kinetics and mechanisms and the dynamics of the folded state.

Table 2.1. Time Scales of Motions in Proteins. The data adapted from [11].

Type of Movement	Example	Time Scale (s)
1. Vibrations and local motions	Making and breaking of covalent bonds	10^{-15}-10^{-10}
	Formation of hydrogen bonds	10^{-14}-10^{-10}
	Transfer of ions between chemical groups	10^{-12}
2. Motions of side-chains, secondary elements and domains	Proton transport	10^{-9}-10^{-4}
	Electron tunneling	10^{-9}-10^{-4}
	Water structure reorganization	10^{-8}
	Ligand binding	10^{-8}-10
	Local denaturation	10^{-5}-10
	Allostery	10^{-5}-1
3. Protein Folding		10^{-6}-10^{4}

2.4. The Mechanism of Protein Folding

The folding of proteins into their native conformations is one of the most essential examples of conformational transitions. Understanding folding mechanisms of proteins provide unique insights.

2.4.1. Levinthal's Paradox and Energy Landscape Theory

Understanding molecular processes of protein folding is a challenging task for both in terms of experimentation and theory. Studies of denaturation and folding back to its native states by only changing the solution conditions [24, 25] provoke the Anfinsen's theory of a thermodynamic hypothesis [26]. These experiments proves that not only local structures but also native three-dimensional conformations of proteins can be obtained accurately by the information in the primary structure without any help of a biological machinery. However, going from the unfolded state to the folded state includes numerous intermediate states and it is unclear that how is it possible to fold into a unique functional conformation within a timescale that ranges from microseconds up to minutes. This suggests that there should be a bias towards the native state. This is named as $Levinthal's Paradox$ [27].

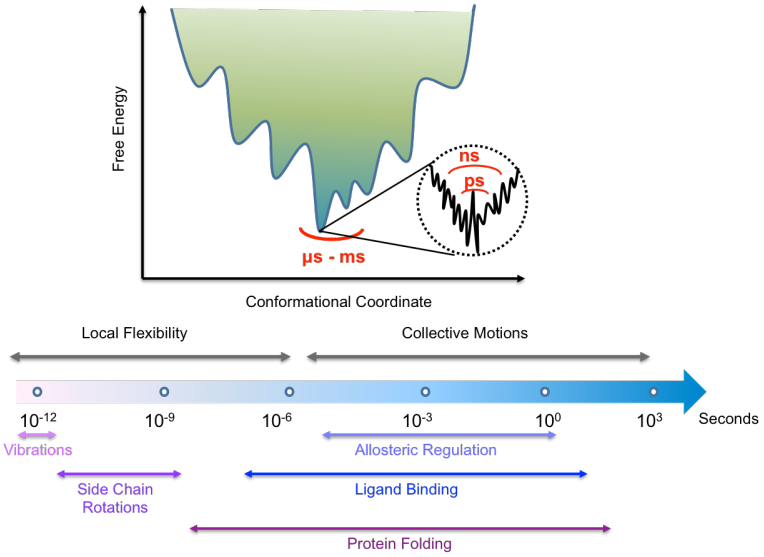


Figure 2.2. Some Examples of the Protein Conformational Dynamical Movements. These Conformational Events Can Span Times of Tens Order of Magnitudes.

The answer to this paradox relies on the idea that proteins do not sample all possible conformations to end up with the right one. There exists a specific pathway for folding composes of a well-defined sequence of events that follow one another [28]. However, statistical physics poses another view describing the folding process as a stochastic process of conformational changes where Brownian motion of protein is coupled with random forces exerted by water medium. Proteins fold in a cooperative manner and this cooperativity is interpreted by the *EnergyLandscapeTheory* [29-31]. In this theory, the folding process is described as a combined action of enthalpic and entropic contributions (Figure 2.3). The free energy is expressed as the sum of enthalpic and entropic effects as $G = E(Q) + TS_{conf}(Q)$, where E is the internal energy of the protein conformation (bonded and nonbonded interactions), S_{conf} is the conformational entropy described by hydrophobic interactions and Q is the conformational reaction coordinate. The ragged walls of the funnel describe the existence of many local energy minima separated by energy barriers that are crossed through intermediate structures to reach the global minimum. Statistical fluctuations thus give rise to the ensemble of pathways rather than only one towards the global minimum (folded state). From the unfolded state to the folded state, the protein does not have to follow a specific path, but multiple pathways, sampling various partially folded structures. Towards

the native state, the entropy decreases and the funnel shape of the landscape reduces possible accessible conformations and enables the protein to fold in a reasonable time.

The free energy difference between the folded and unfolded states (the difference between the energy and entropy of two states), $\Delta G = \Delta E - T\Delta S_{conf}$, is the measure of the protein stability. As the energy difference ΔE increases or the entropy difference ΔS_{conf} decreases between folded and unfolded states the folded conformation becomes more stable and it requires significant increase in the free energy to perturb the folded state.

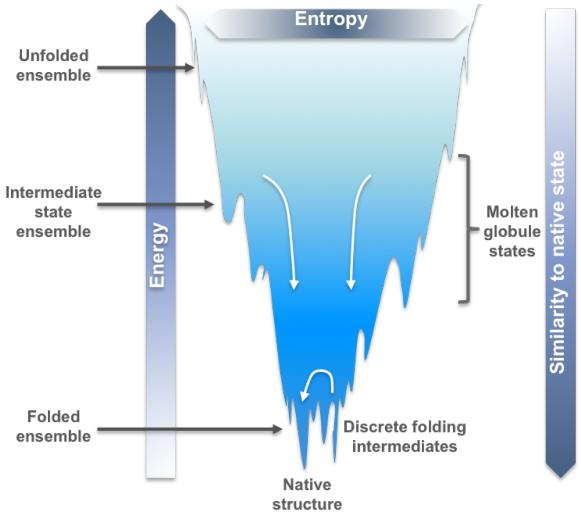


Figure 2.3. Schematic Representation of the Folding Funnel of a Protein. The Funnel Shape Reflects the Simultaneous Decrease in the Energy and Entropy as well as the Increase in Similarity to Native State of the Protein during its Folding. Adapted from [32].

2.5. Folded State Dynamics

Protein folding emphasizes the uniqueness of the native state and its functional importance. However the native state is not a single functional conformation but only the time averaged one [11] which could be influenced by an external stimuli such as ligand binding or changes in the environmental conditions.

2.5.1. Conformational Ensemble of Proteins

Koshland's induced-fit theory [17] changed the consideration of proteins as static entities that had been utilized to explain substrate/ligand induced conformational change of proteins. However, Linus Pauling proposed a model stating that antibodies might recognize different conformations with identical energies [33], which implied that proteins may undergo conformational transition in the absence of ligand or substrate. Following Pauling's theory, Monod and co-workers proposed pre-existing equilibrium theory [19] where it is suggested that proteins might exist in two different equilibrium conformations in the absence of ligands. The current model [34] combined these ideas and assumes that proteins exist in spontaneous equilibrium between many different conformations called substates and thus in an $ensembleofconformations$ [2, 35].

The variation between those substates might be just a different position of only a single atom or repositioning of whole region. The timescale of the shifts between the substates of the ensemble is different depending on the height of the energy barrier separating them. With the inverse relation between the exponential of the free energy ΔG and the probability to exist in a certain substate (P), the protein expected to stay longer at those substates with the least energy. Native conformation in crystallographic structures is thus the most stable (lowest energy) one. Since the spontaneous sampling of conformations is driven by the thermal energy, the protein stays most of the time at its native state or in other conformations that are within +0.6 kcal/mol (value of RT at room temperature) or less of the native state [36]. Conformations with very high energy have almost zero probability to be sampled within the lifetime of a protein, yet some high energy conformations could be sampled for a short period of time within long enough timescales, which are called rare events [11].

2.5.2. Conformational Changes in Proteins

The protein function is mainly driven by its intrinsic dynamics, which can be characterized by a multidimensional free energy landscape (FEL). Proteins can sample a large ensemble of conformations spontaneously with the thermal energy and fluctuate

around the average conformation in the most favorable equilibrium state. A change in any type of external conditions such as temperature, pressure, solvent conditions or binding of a ligand and mutation can influence the relative population of existing substates and the kinetics between them. Protein dynamics involves any kind of time-dependent change in atomic coordinates and the resulting formation or destruction of chemical bonding within the system. Fluctuations observed in the protein dynamics can be grouped into two categories: slow timescale and fast timescale motions. In slow timescale motions, substates are separated by high free energy barriers the transitions between which takes microseconds, milliseconds or even seconds. These slow timescales fluctuations are usually linked to large amplitude collective motions. In fast timescale motions, substates are separated by low energy barriers and the transition between these state correspond to fluctuations of picosecond or nanosecond timescales that are associated with small amplitude conformational changes [23]. An example of a well in the free energy landscape describing the amplitude and timescales of the protein motions is given in Figure 2.4 [37]. The shape of the landscape may be changed upon with an external or internal perturbation and the population of the substates are redistributed.

2.5.3. Protein Dynamics and Allostery

Allostery is the process of transmission of the effect of a perturbation (ligand binding, mutations, etc.) at one site of a protein to another distal site, assuming the regulatory control of a functional action. This has been referred to as "second secret of life" [38]. It has remained a central focus in biology because of two reasons; it is fundamental to an understanding of most biological processes and it is at play in targeted drug discovery [19, 38, 39]. A positive allosteric effector enhances the activity by increasing affinity while a negative allosteric effector will reduce the activity a the distant functional site of a protein [40].

There were two dominant models for allostery; the "concerted" or MWC (Monod-Wyman-Changeux) model [19] and the "sequential" or KNF (Koshland-Nemety-Filmer) model [41]. The MWC model postulated the pre-existence of two end struc-

tures (tensed (T) and relaxed (R)) in equilibrium where this equilibrium was shifted with ligand binding to an allosteric site [19], whereas the KNF model assumed that the binding of a substrate induces the conformational change from the T form to the R form. The latter type of behavior was later accepted by the induced-fit theory of substrate binding. The binding of the substrate to one subunit simplify the binding (of another substrate) to the other subunit that makes the phenomena "sequential" [41]. Both of these two early definitions were agreed on the requirement of a conformational change between two well defined structurally stable states [34].

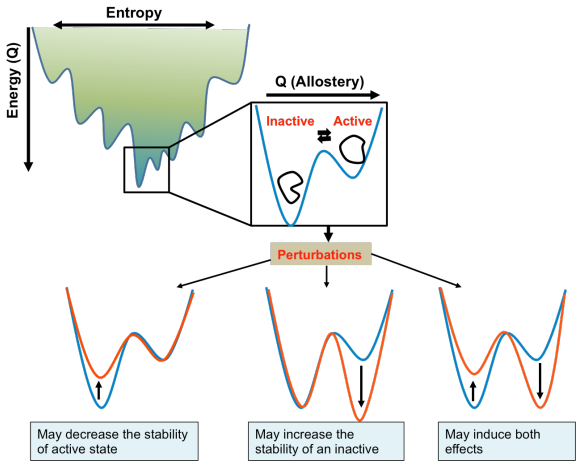


Figure 2.4. Schematic Representation of a Free Energy Landscape and the Possible Effects of Allosteric Stimuli. The Effects of any External Stimuli on the Shape of the FEL are Also Mapped in Amplified View. Adapted From [23].

The pioneering work of Frauenfelder and co-workers [42] describes protein dynamics and conformational ensemble in terms of the free energy landscape, which provided a means to explain allostery [43, 44]. Considering the equilibrium in the ensemble of conformations and the free energy landscape, an allosteric determinant acts through shifting the ensemble from a populated inactive state to a populated active state, or vice versa [39]. With this, an alternative allosteric mechanism has evolved without requiring structural switching and a ligand-induced conformational changes as encountered in the previous models [40, 45- 48]. For example, an allosteric effector (ligand binding, mutations, etc.) may not lead to a significant conformational change but affect the dynamics at the binding interface [1, 49]. An example of an allosteric effector induced population shift in the ensemble of conformations and alteration of the

free energy landscape without any major conformational change is illustrated in Figure 2.4. This alternative mechanism has been developed as consequences of the comparison of *apo* and *holo* crystal structures, NMR data, and molecular dynamic simulations [50-52].

3. COMPUTATIONAL APPROACHES FOR STUDYING CONFORMATIONAL TRANSITIONS AND FUNCTIONAL DYNAMICS OF PROTEINS

Time resolved crystallography and spectroscopy techniques provide intermediate protein conformations in the pathway of conformational change and dynamical information along the pathways. Nevertheless, up to date experimental techniques still have limitations such as requiring significant populations and sufficiently long time scales to detect the conformational changes. Therefore, it is necessary to complement experiments with computational techniques to obtain more detailed information on conformational flexibility and transitions. A wide community of scientific researchers uses computer simulations, so called "computer experiments". Computer simulations fill the gap between experimental and theoretical works.

There is a close connection between computer simulations and statistical mechanics. Microscopic properties (atomic coordinates, velocities, etc.) are determined via simulations and the corresponding properties measured by experiments are calculated by statistical averages over sampled microscopic conformations. Two critical points in computer simulations are:

(i) The identification of the suitable *statisticalensemble*
(ii) The issue of *ergodicity* (the equivalence of time average and ensemble average over the phase space) [53].

First atomistic biomolecular simulation was published in 1977 by McCammon and co-workers [2]. Since that time, computer simulations of biomolecular systems made an adorable progress to simulation of large protein complexes in solvated environment like tens of millions of atoms [54] and millisecond time scales [55].

3.1. Molecular Dynamics (MD) and Related Methods

MD simulation is a technique used for an atomistic explanation of a classical many-body system. The term classical defines the motion of the particles within the investigated system obey to the laws of classical mechanics. In MD simulation, numerical methods are used to solve Newton's equation of motion for a series of finite time step. It uses an empirically derived potential energy function describing all molecular interactions [56]. The generated atomistic coordinates of the system, so called trajectory, can be considered as deterministic and reproducible given the exact initial conditions and time.

However, even though it has very successful investigations, MD simulations have two main limitations; despite limited time scales it has high computational cost and the force fields are inaccurate for large time scales [3]. Unfortunately, most MD trajectories are far from being ergodic which is a crucial point in molecular simulations, because they leave unexplored regions in conformational space [57]. This is because of the existence of high-energy barriers between high probability "metastable" conformations and low-probability "transition" conformations. Recently, the development of dedicated machines such as Anton [58], distributed computing protocols (parallel computing) [59], and using GPUs for MD simulations [60], has eliminated these issues to some extent. However, time resolution of interesting biomolecular processes and many experimental techniques are significantly longer than conventional MD simulations [4].

The aforementioned limitations of MD simulations has led to the development of either advanced methods to enhance the sampling of high probability metastable states and rare events which requires crossing of high free energy barriers [5] or coarse grained methods (namely Elastic Network Models, ENM) to reduce the computational cost and to study large systems (6-9). Coarse-grained MD [61, 62], Accelerated MD [63], steered MD [64], milestoning [65], transition path sampling [66], umbrella sampling [67], replica exchange [68], metadynamics [69] and their combinations and derivatives are among the most widely used methods to enhance conformational sampling.

3.2. Enhanced Sampling Techniques for Studying Rare Events

MD simulations have been used for the atomistic simulation of conformational changes in biomolecules. Large conformational changes which requires long time collective behavior and crossing over relatively higher energy barriers in the energy landscape are usually important for the activity of proteins [3]. Although these conformational transitions observed too fast in high-resolution experimental methods, they occur too slowly in fully atomistic MD simulations within tolerable time of computation. These important transition events between highly populated and long-lasting states are called *rareevents*. Even with most sophisticated computational power, simulations can only access up to millisecond timescale [70]. Therefore, developing special computational methods that allow the crossing of high free energy barriers to enhance sampling of conformational space became crucial to study rare events in protein dynamics.

Several sophisticated methods have been developed to study rare events and applied to conformational transition of many proteins and complex systems, such as coarse graining the description of the system, replica exchange molecular dynamics (REMD), parallel tempering, metadynamics, umbrella sampling, simulated annealing, etc.

3.2.1. Finding the Mechanism of Rare Events

Sampling timescale of rare events depends on the energy barrier or entropic bottleneck that separates the ensemble into different stable states. A schematic representation of a rare event is given in Figure 3.1. This simple energy landscape has two global minimum separated by a high energy barrier. Starting from one stable state A, the system stays longer in state A before passing over the barrier to another stable state B. The system stays at state B for a shorter time relative to state A. The transition time from one stable state to another is much shorter than the time at states A or B. This transition can thus be classified as rare events and the time scale of crossing the energy barrier is mostly impossible to access via classical MD simulations. More sophisticated methods are needed to observe this transition.

Transition state theory (TST) is a baseline for the development of computational techniques to study barrier crossing, presenting a simple theoretical framework to the rare event sampling [71]. In TST, the system has to visit some intermediate configurations called transition states corresponding to the saddle points on the potential energy surface to overcome a barrier from the reactant to the product state. According to this theory, the identification of relevant degrees of freedom describing the conformational transition, so-called *reactioncoordinates*, is the elementary step on studying rare events.

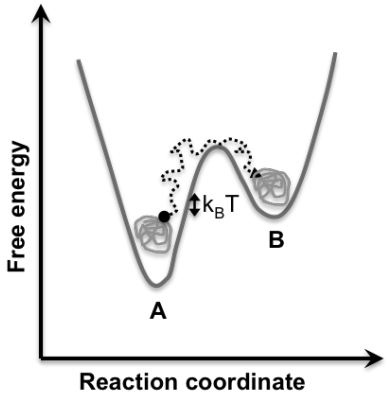


Figure 3.1. Schematic Representation of Two Stable States A and B on the FEL of a Molecular System.

3.2.2. Defining Reaction Coordinates to Describe a Transition

In chemistry, a reaction coordinate is a one-dimensional coordinate that represents progress along a reaction pathway. Describing a transition of a full atomistic molecular system containing N number of atoms in the $6N$-dimensions (translation and rotation) and momentum of the phase space has a very large number of degrees of freedom. The high dimensionality of large conformational change in molecules makes it difficult to interpret and analyze. It is thus necessary to reduce the degrees of freedom by projecting a high dimensional conformational space into a few easy to track geometric properties, which can be either analyzed individually or manipulated to control the dynamics. These properties are called collective variables (CVs), which refer to any multidimensional function ζ of $3N$-dimensional atomic conformation $r \equiv (r_i|i = 1, ...3N)$. The functions $\zeta_1, \zeta_2, ...\zeta(N_{CV})$ map configuration r onto N_{CV}-dimensional CV space

where N_{CV} is the number of reduced degrees of freedom and usually $N_{CV} \ll 3N$ [57].

CVs can be either simple geometric parameters, such as, bond angles, torsions, distance, etc., or relatively more complicated parameters, such as Root Mean Square Deviation (RMSD) of the whole or some part of the molecular complex, coordination number, secondary structural content, etc. [72]. CVs should not just measure the progress of the transition, but also should be useful to understand the dynamics.

3.2.3. Methodological Approaches to Enhance Conformational Sampling

3.2.3.1. Coarse-Graining. In order to simulate either a huge system or any systems at longer time scales, simplification of the description of the system is required such a coarse-grained (CG) modeling of the atomistic system [73]. There are three properties that one should decide during designing a CG model; the basic simulation unit, the effective CG potential function and the dynamical equations describing the time evolution of the CG system [61]. There are many coarse graining methods ranging from qualitative solvent free models to models that consider the chemical specificity with explicit water. Special force fields for coarse-grained simulations were developed for proteins [74-76], membrane [76, 77], nucleic acids [78], and also carbohydrates [79]. All these developed models may perform well where bulk properties dominate over atomic details, but the major drawback is the loss of atomic details.

3.2.3.2. Accelerated Molecular Dynamics (aMD). aMD is one of the enhanced sampling method where conformational sampling is accelerated by adding an energy term to the potential energy of the system that reduces the energy barriers between different substates [63]. This additional energy term is called boost energy (E_{boost}) and depends on the difference between the potential energy of the system and a pre-determined reference energy called threshold energy ($E_{treshold}$). However, the correct value of the E_{boost} can not be known a priory [80]. For example; if one selects a $E_{treshold}$ lower than the energy barrier, it will never cross it.

3.2.3.3. Targeted and Steered Molecular Dynamics (TMD and SMD). TMD simulation is a derivative of MD, developed to observe large conformational transition between two known conformations of a molecule [81-83]. An artificial restraint potential term is added to the energy function proportional to the square of the RMSD difference between the current conformation and the target conformation. This added potential term forces the system to evolve to the specified state of interest. Another type of TMD is steered molecular dynamics (SMD) where it is invented to mimic Atomic Force Microscopy (AFM) experiments [64]. It employs an external, time dependent and/or position dependent pulling force to a pre-determined position of a protein (to a specific atom or a fictitious point) [84]. During the simulation, all atoms in the system adjust to the forced change in the structure and conformations along a particular pathway are sampled. As the simulation progress, the potential energy of the system is calculated. This calculated potential of mean force (PMF) is the interpretation of the free energy change along the chosen reaction coordinates for the progress.

3.2.3.4. Replica-Exchange Molecular Dynamics (REMD). Systems at higher temperatures can overcome high energy barriers more easily compared to lower temperatures. Therefore, elevating temperature in simulations significantly improves conformational sampling [85]. REMD is an enhanced sampling method [86], which employs independent (non-interacting) parallel MD simulations carried out at different temperatures and periodically exchanging (with a specified transition probability) the coordinates of replicas between ensembles depending on the temperature and energy difference of the randomly chosen replicas see Figure 3.2. The replicas are swapped according to the Monte Carlo-Metropolis acceptance criterion. Exchanging coordinates of replicas can also be considered as exchanging temperatures. Since the selection of pair replicas to exchange is a random process, the resulting REMD algorithms becomes stochastic and this stochasticity enables crossing large energy barriers and therefore prevents stucking at a local minimum [87]. This method is also referred to as parallel tempering.

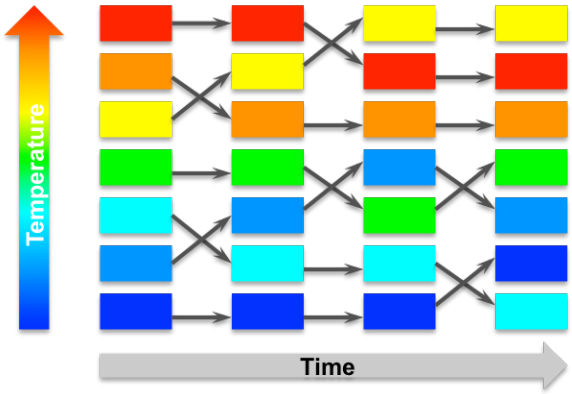


Figure 3.2. Schematic Representation of the REMD Method.

The efficiency of REMD to describe actual protein dynamics depends sensitively on maximum temperature selected for the simulation and the effectiveness is strongly depend on the activation energy of conformational transition [88]. Therefore, although temperature dependent REMD (T-REMD) has many successful applications in enhancing sampling in MD, it does not guarantee better convergence. A more prevailing form of REMD comprises exchanging Hamiltonians (sum of kinetic and potential energy) instead of temperatures (H-REMD), where the force field or the selected force field contributions are modified along the replicas [89]. H-REMD is now widely used and sometimes referred to as multidimensional replica exchange molecular dynamics (M-REMD) [80].

3.2.3.5. Umbrella Sampling (US). US is another enhanced sampling method [90], which is related to importance sampling in statistics. In this method, the simulated system coordinates are grouped into sets of collective variables (CVs) such as; distance between two atoms or group of atoms, torsional/rotational angles, etc. referred as reaction coordinates, where each set determining a separate umbrella window [91]. A bias potential, which is usually in quadratic or harmonic form, is applied to the reaction coordinate to restrain and pull to a targeted value of the reaction coordinate [92]. This restrain potential determines the weighting function of a given window, where it scales with the distance from the equilibrium state. As a result, the simulation will be biased away from the initial conformation [91]. Since US is done in a series of windows, the

information coming from each window is combined and converted to local probability histograms using weighted histogram analysis (WHAM) [93]. This methodology is readily applied to both MD and Monte Carlo (MC) simulations.

In order to improve the efficiency and convergence of this methodology, some experimental restrains such as end-to-end distances determined by fluorescent resonance energy transfer (FRET) are used to formulate the restrain potential, which is called "guided US" [94].

The main disadvantage of US is the requirement of apriori knowledge about the FEL of the system of interest such as either the position of the barrier with respect to reaction coordinate or the exact height of it.

3.2.3.6. Metadynamics. Metadynamics is yet another powerful enhanced sampling method and construction of relative free energy landscape (FEL) developed by Laio and Parrinello in 2002 where ergodicity is prevented by the shape of the system's energy landscape [69]. Although based on the concept of CVs, it deviates on many significant aspects from US by inserting memory in the sampling. Metadynamics involves adding a history dependent biasing potential to overcome high energy barriers between energetically favorable metastable states and hence to sample rare events [69]. Eventually, the FEL of the conformational ensemble is compensated by the deposited biasing potential and can be reconstructed by using this deposited biasing potential [95]. Darve and Phorille [96] describe this methodology as "filling the free energy wells with computational sand" (Figure 3.3). Since metadynamics can search through the entire FEL, it is a promising method to study many biological problems such as protein folding [97], molecular docking [98], and conformational changes [95, 99].

A small set of (usually two or three) collective variables should be selected to start a metadynamics simulation. The only difference between the classical MD and metadynamics simulations is that for the latter a bias potential (a Gaussian hill) is being added to the potential energy in regular intervals. These hills are the function of

the CVs and accumulate along the simulation where inverse of this accumulated bias potential hills gives the FEL. However, since there is no limitation on the accumulation of bias potential hills, one may end-up with undesired overfilling of the energy landscape of the system. In order to solve this issue, employing adaptive Gaussian hills was suggested in so-called Well-Tempered Metadynamics (WTmetaD) approach [100].

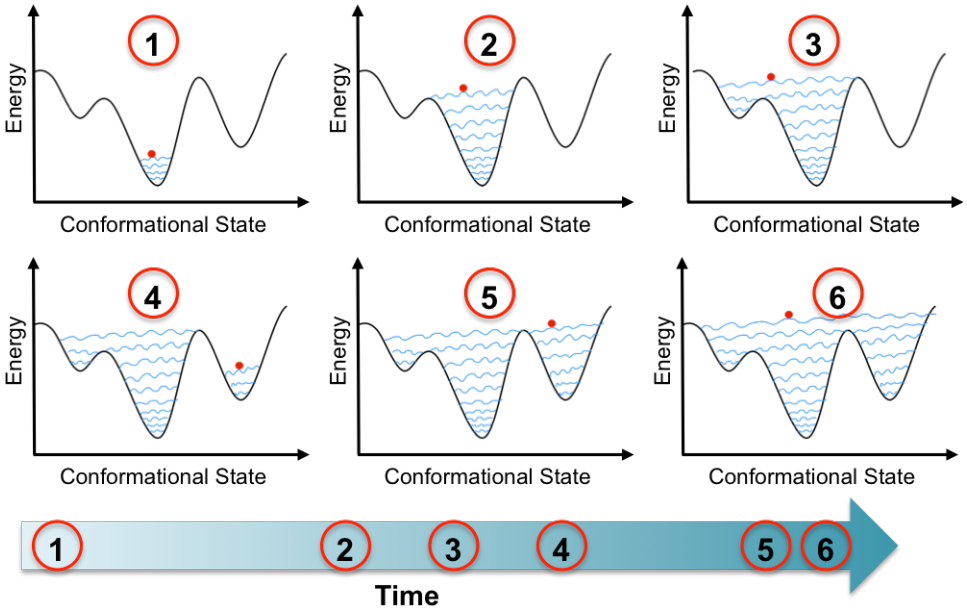


Figure 3.3. A Schematic Representation of Meta Dynamics Method. The Simulation Searches Energy well Avoiding Over Sampling Using its Memory. When the Bias Potential Fills the First Minima, the System Goes to a Lower Energy Configuration.

3.3. Combination of MD Related Enhanced Sampling Methods

The enhanced sampling methods described above have both pros and cons. For example, one can precisely model a certain process defined by a set of CVs and deduce the underlying free energy landscape. However, sampling of only few number of CVs can be actively enhanced by existing algorithms. On the other hand, one can enhance sampling of all degrees of freedom through elevated temperatures in parallel tempering without defining any CVs, but in parallel tempering, the number of replicas increases as the system size increase leading to computational inefficiency. The combination of these methods may give better results.

Parallel tempering and metadynamics methods were successfully combined by Bussi and coworkers in parallel tempering metadynamics [97]. In this combined methodology, the studied system is simulated in multiple replicas at different temperatures but the same bias potential according to defined CVs is applied to each replica. As a result, the conformational sampling is enhanced both in terms of defined reaction coordinates and also elevated temperatures. The free energy surfaces were thus calculated at different temperatures. Successful applications of parallel tempering metadynamics indicated that it is very efficient in simulating flexible ligand binding or conformational change of proteins. A recently introduced well tempered ensemble (WTE) approach [101] allows using less number of replicas in parallel tempering metadynamics. In WTE, the potential energy of the system is used as CV to obtain an efficient assessment of the probability distribution of energy at numerous temperatures [102].

Another successful combination of aforementioned enhanced sampling methods is called bias-exchange metadynamics, which amalgamates metadynamics and replica exchange introduced by Piana and Laio [103]. The system is simulated again multiple replicas where each replica is biased by a single CV at the same temperature and therefore subject to one-dimensional bias potential. The number of replicas is equivalent to the number of CVs; therefore one can use numerous CVs to bias the simulation. Replicas can periodically swap their coordinates allows the system biased along one CV to become biased along another CV. Exchanging coordinates is according to the Monte Carlo-Metropolis criteria and determined from the bias potentials.

4. NORMAL MODE ANALYSIS TO PREDICT THE CONFORMATIONAL ENSEMBLE AND FUNCTIONALLY IMPORTANT SITES AND MOTION

Although many enhanced sampling algorithms are developed to overcome the time and size limitations of the conventional MD in biological applications, there is still a room for more simplified and computationally efficient methods. Normal modes and related techniques have increasingly become popular to study protein dynamics [104, 105].

4.1. Conformational Ensemble from NMA

4.1.1. Standard NMA

Normal mode analysis (NMA) is a powerful method used for the analysis of collective motions in biomolecules. Classical NMA uses the same force field as used in MD simulations, which analyzes vibrational motions around a local energy minimum. This simplified approach assumes that the conformational energy surface around a local minimum can be approximated by a parabola over the range of thermal fluctuations [106]. However, this is in fact may not always be the case in proteins at physiological temperatures [107].

A standard NMA starts with an energy minimization of the folded protein structure. Then, the so-called "Hessian" matrix which the matrix of the second derivative of the potential energy function with respect to the mass-weighted atomic coordinates. The last step is the determination of the vibrational modes of the system, which is the diagonalization of the Hessian matrix and extraction of the eigenvalues and eigenvectors (the "normal modes") [106]. $3N$-6 non-zero normal modes and eigenvector are extracted form the $3Nx3N$ Hessian matrix for a system of N atoms [108]. In the high frequency modes the displacements (eigenvectors corresponding to highest eigenval-

ues) give the local fluctuations of covalently bonded atoms, whereas the low frequency modes provides the collective movements involving large parts of the structure [109]. The diagonalization of this $3Nx3N$ Hessian matrix is computationally demanding, the computational efficiency of standard NMA thus decreases as N becomes larger.

4.1.2. Elastic Network Models: ANM and GNM

Because of the computational difficulties of standard NMA, elastic network models (ENM), which is a form of coarse grained NMA have been developed [6-8]. The hypothesis is that the lowest frequency normal modes are a global property of the structure and shape of the proteins [105, 110- 114] and thus the elimination of interatomic interactions does not affect the global motion of proteins. This approach was first introduced by Tirion [115], where the Lennard-Jones and electrostatic interactions were replaced by Hookean springs for atom pairs within a cutoff distance. Later, Bahar and coworkers proposed the application of this model at the amino-acid level [7, 9]. This approach has two main advantages compared to standard NMA. There is no need for energy minimization as the elastic connections are taken to be at their minimum energy length and the use of residues instead of atoms that reduce the computational time for the diagonalization of the Hessian matrix [106]. In Gaussian Network Model (GNM), the residue fluctuations are assumed to be isotropic (same in all directions) [7, 9], whereas in the anisotropic network model (ANM) the anisotropic effect is taken into account [6]. Through ANM, one can have information about the three-dimensional motion of the system.

The relatively low computational cost and simplicity of ENM has provided opportunity to its wide application in analysis of conformational transition and dynamics of proteins. Application of ENM to the known open and/or holo forms of proteins have indicated that one or more low frequency normal modes highly overlap with the conformational transition of proteins from the “apo” to the “holo” forms [116-119]. Exploration of conformations along the displacement of normal modes of a given native structure has been used broadly for many cases.

4.1.3. Combining NMA with Molecular Simulations to Enhance Conformational Sampling

A significant portion of recent studies employs combination of NMA and molecular simulations to improve efficiency and accuracy of conformational sampling. There are successful applications of standard NMA combined with Monte Carlo (MC) simulations for predicting conformational transitions of gramicidin A [120], KscA K+ channel [121] and ApcT amino acid transporter [122] without providing the target structure. Perahia and coworkers [123] used multiple minima all-atom NMA to explore the functional motions of HIV-1 protease, where NMA on a set of energy-minimized structures from short MD simulations were utilized to obtain a consensus covariance matrix. Then, they concluded that collective modes correspond the biologically relevant motion of the protease including flap opening and closing. Another hybrid approach called ANM-MC [124, 125] integrates ANM collective modes to a form of knowledge-based Monte Carlo (MC) simulation techniques to generate conformational transition pathways. This methodology have both targeted and non-targeted versions. In the targeted adaptation, the structure was iteratively deformed along the direction of the collective mode that overlaps with the direction of conformational transition. In the non-targeted adaptation, only the radius of gyration of the target structure was used as input to predict the conformational transition pathway and the closed state of the structure.

Collective molecular dynamics (coMD) [126] is an enhanced MD sampling method where the conformational sampling is enhanced through collective dynamics of the molecular system guided by low frequency ANM modes [6]. In this methodology, the basic strategy is to deform both the initial and the target structures along the most dominant collective ANM modes of each structure. In this methodology, the modes overlapping with the direction of the initial/target transition are accepted upon a Monte Carlo/Metropolis strategy. This stochasticity allows the system to deviate from the shortest path from initial to target structure and to sample lower energy conformations. The ANMPathway is another approach for conformational sampling, which uses the collective ANM modes extracted of the two end structures of the conformational

transition [127]. Both coMD and ANMPathway methods were applied to predict the conformational transition pathway between open-close structures of adenylate kinase (AdK) [126, 127] and several membrane transporters [127] with a significantly lower computational cost than the other methods [128].

Combining collective modes with the enhanced sampling algorithms also provides computationally effective and efficient sampling of conformational space of proteins. Vashisth and Brooks combined temperature accelerated MD with low frequency normal modes obtained from Ca-based elastic network model (ENM) with a single parameter potential [115] to enhance conformational sampling of maltose binding protein and nucleotide binding domains of a maltose-transporter [129]. They concluded that single or a combination of two to three collective modes was able to describe the open to closed conformational transition in temperature-accelerated MD generated pathways. Martin Zacharias incorporated ENM with H-REMD to speed up the conformational transition of proteins [130]. In this approach, a distance-dependent penalty (flooding or biasing) potential that derives the structure away from the current conformation along the direction of the information extracted from ENM analysis added to the MD force field. Later, they improved their penalty potential to accelerate the large-scale domain motions again using the information obtained from ENM with H-REMD [131]. Wang and coworkers [132] combined NMA in internal coordinates of all heavy atoms [133] and umbrella sampling MD with both *apo* and *holo* form of the proteins to investigate transition pathways and free energy profiles of adenylate kinase, calmoduline and p38α kinase, which yielded transient conformations consistent with experimental and other computational studies.

Kurkcuoglu-Soner and coworkers developed an ENM-based iterative methodology called ClustENM for the sampling of entropically accessible states of a protein from a starting structure [10]. In this methodology, the global modes were extracted from an energetically minimized native structure using ENM, and then the native structure was deformed along the combination of these global modes that results in new structures. Next, these generated structures were clustered to obtain representative conformers. Minimization-deformation along ENM modes- clustering steps was applied iteratively

to obtain more conformers to represent conformational ensemble of a protein.

4.2. Functionally Important Sites from NMA

Interatomic interactions lead to local to global motion adopted to the functionality, or to the functional global motion. Hinges are the major key components of the dynamic infrastructure and any perturbation, physical and/or chemical, at a hinge region provide an aspect of allosteric control of both functional local fluctuations as well as global dynamics through an impact on relative probabilities of accessible conformational states, i.e. on the ensemble of the conformations [48, 52].

As suggested earlier, sequential variation in the network of residues related to the hinge bending motion of proteins results in alteration of dimerization properties as well as specificity towards ligand binding [134]. Perturbation on a hinge regions leads to a modification on the interatomic bonding network within the protein, which may results in a shift in the ensemble of conformations or alteration of functional dynamics of the protein. Importance of hinge residues on the functionally relevant collective motion and the allosteric network that affect binding affinities of proteins was studied on numerous numbers of proteins, such as adenylate kinase (AdK) [23], neurolysin [135], PDZ domains [136, 137] and HIV1 protease [138]. AdK is an enzyme that catalyzes the reversible nucleotide phosphoryl exchange within the cell. Two distinct functional forms of this protein are open and closed forms, and great number of intermediate structures has been captured via experimental studies. AdK undergoes a large conformational change even at the absence of substrate molecule [91]. Comparison of the protein dynamics of a hyperthermophilic and a mesophilic AdK revealed that the physical origin of catalytically important collective domain motions relies on a hinge motion that controls the conformational transition of the enzyme [23]. In addition, Hines and coworkers discovered a potent inhibitor that binds away from the catalytic site of neurolysin. The binding position of the inhibitor is a hinge site suggesting that the inhibition on the enzyme was achieved by preventing a hinge-like motion and thus reversing a substrate-associated conformational change [135].

In addition to low frequency fluctuating residues, the high frequency fluctuating ones are also shown to have structural and functional importance [9, 139-142]. Since binding sites are associated with high as well as low stability regions, the high frequency fluctuating residues at the surface of the proteins are associated with the binding core amino acids as they form folding core when they located at the interior of the proteins [143-145].

5. ATOMIC FORCE MICROSCOPY OF PROTEIN INTERACTIONS

The mechanics of biological systems is crucial for their function. Individual proteins require flexibility of secondary structures to allow microsecond conformational changes. In the living cell, shape maintenance and structural stability is assured by the mechanics of the cytoskeleton, adhesion complexes and the plasma membrane. Thus, the study of the mechanics of proteins, protein complexes and the membrane is important to understand biological processes.

Interaction between two biomolecules such as proteins DNA, RNA drug molecules is the key phenomena in many biological processes. They recognize and bind each other in the cell environment to conduct their function. Until recently, protein-protein interactions were characterized through binding affinities and/or rate constants determined by biochemical methods. Since these experimental methods are bulk measurements, they can only give averaged estimates of desired properties. However, proteins exist in an ensemble of conformations where their populations change with varying conditions, thus. the characterization of the interaction dynamics that expose various intermediate states or alternative reaction pathways is critical [146]. Single molecule approaches allow investigation of characteristics of individual interacting molecules in real time and at similar circumstances with physiological conditions. With these attributes, they have become essential tools for understanding of association / disassociation processes of biomolecular complexes together with intermediate states and the underlying energy landscape [147-151].

The quantification of forces holding protein complexes together using biophysical methods have evolved with the development of nanotechnologies such as optical and magnetic tweezers and the biomembrane force probe [152]. One of the most versatile nanotechniques is atomic force microscopy (AFM), which enables topographical imaging and force measurements at the nanometer scale with piconewton force reso-

lution. Importantly, AFM works under liquid conditions allowing characterization of biological samples. The flexible cantilever is sensitive to picometer deflection changes and has associated a spring constant that can be as low as 5 pN/nm. Thus, it is an excellent force sensor that allows manipulation of the sample and force application. In force spectroscopy mode, the cantilever is moved in the vertical direction (z) to apply and measure mechanical forces by pulling or poking the sample. The quantitative nature of force spectroscopy measurements has converted AFM into a valuable tool in biophysics. Force spectroscopy allowed determining the forces required to unfold protein domains and to disrupt individual receptor/ligand bonds as well as probing the mechanical properties of normal and cancerous cells [153-156].

5.1. Atomic Force Spectroscopy (AFM) Experimentation and Principles

A flexible cantilever, a tip mounted on the far end of the cantilever, a sample stage, a piezoelectric translator that moves the sample stage or cantilever depending on the design, and an optical deflection system composes of a laser, a photodetector and a signal processing unit, which records the changes in cantilever deflection are the core components of and AFM device as shown in Figure 5.1.

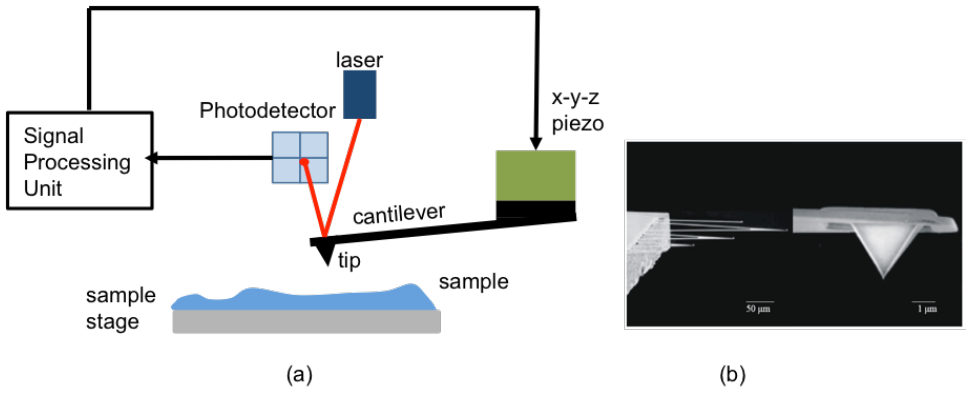


Figure 5.1. Schematic Illustration of the Fundamental Components of an AFM Device with Working Principles (a) and a Sample Cantilever (b).

AFM was developed as an imaging tool where the topographic images of the sample are obtained by scanning the surface in the x, y plane with a flexible cantilever [149]. A feedback controller programmed with the piezo continuously adjust the z-

position of the cantilever during scanning to keep the deflection constant as illustrated in Figure 5.2 [146].

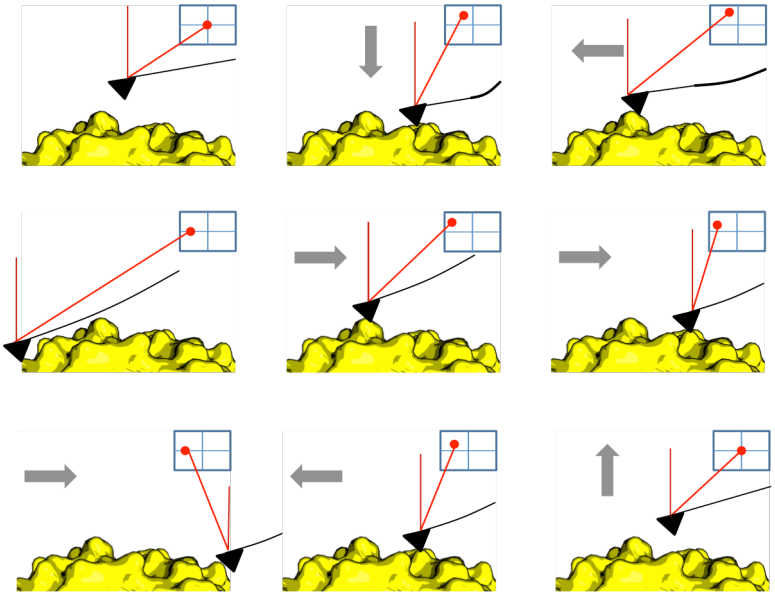


Figure 5.2. Illustration of the Topographic Imaging using AFM. The Deflection of the Laser on the Photodetector is used to Record the Position of the Cantilever and also as a Feedback to the Controller.

The difference of the position of the deflected laser beam on the photodetector gives the bending of the AFM cantilever due to the interaction forces between the tip and the sample. In order to convert the deflection amount of the signal into units of force, spring constant of the cantilever (k) and the deflection sensitivity (s) of the cantilever must be calibrated before any measurements [146]. Different techniques are used for the cantilever calibration, but the most popular one is Hutter and Bechhoefer method [157]. In the latter, the cantilever is considered as a simple harmonic oscillator and the spring constant is derived from the power spectrum of thermal fluctuations of the cantilever.

5.2. AFM Measurement of Single Molecule Interactions

One of the applications of AFM is force scan mode to measure interaction between two pair of surfaces at the single molecule level, which is sometimes referred as

biomolecular force spectroscopy. In a typical AFM-based force spectroscopy setup, a flexible cantilever (spring constant 10-100 mN/m [146]) functionalized with a molecule, is positioned close to the sample surface again functionalized with the other molecule of interest. The thin pin AFM tip, with an average apex diameter of 10-50 nm diameter, narrows the interacting surfaces to the limits of single molecule interactions [158].

The position of the cantilever and the corresponding photodetector signal recorded during a typical force spectroscopy experiment is schematically shown in Figure 5.3. In the recorded signal, y-axis represents the photodetector signal proportional to the bending of the cantilever, which will be converted to the force applied on the cantilever by using calibrated properties of the cantilever. The x-axis represents the time and the distance between the base of the piezo and the stationary surface can be calculated using the time and the velocity of the piezo. The experiment begins with positioning of the functionalized cantilever a few microns above the sample surface as in part I in Figure 5.3. The cantilever is then moved towards the stationary sample surface with the help of piezo and makes contact with the sample surface as in part II. The deflection signal is maintained at a stable baseline level via a feedback control mechanism corresponding to "zero force" until the molecules attached on the tip touch to the molecules functionalized on the stationary surface. After the tip touches to the sample surface, the cantilever is further pushed towards the sample surface until a preset deflection / force is reached to ensure the formation of the bonds between the molecules. This extra pushing causes the cantilever to bend upward that leads a positive change of the deflection/force signal as in part III. After the peak force is reached, the cantilever starts to move upward to its initial position. In the course of upward movement of the cantilever, the adhesive contacts formed during the tip is in contact with the sample surface are revealed causing a downward deflection of the cantilever (part IV). If specific interaction is formed between the molecule functionalized on the tip and the molecule attached on the stationary surface, the cantilever bends further downward as in part V. In this part, further retraction movement of the cantilever from the contact point results in a gradual increase of the tension applied to the intermolecular bond(s) until the bond(s) are ruptured. This will cause a sharp positive change in deflection/force signal. Then, the cantilever returns to its original constant baseline as in part VI.

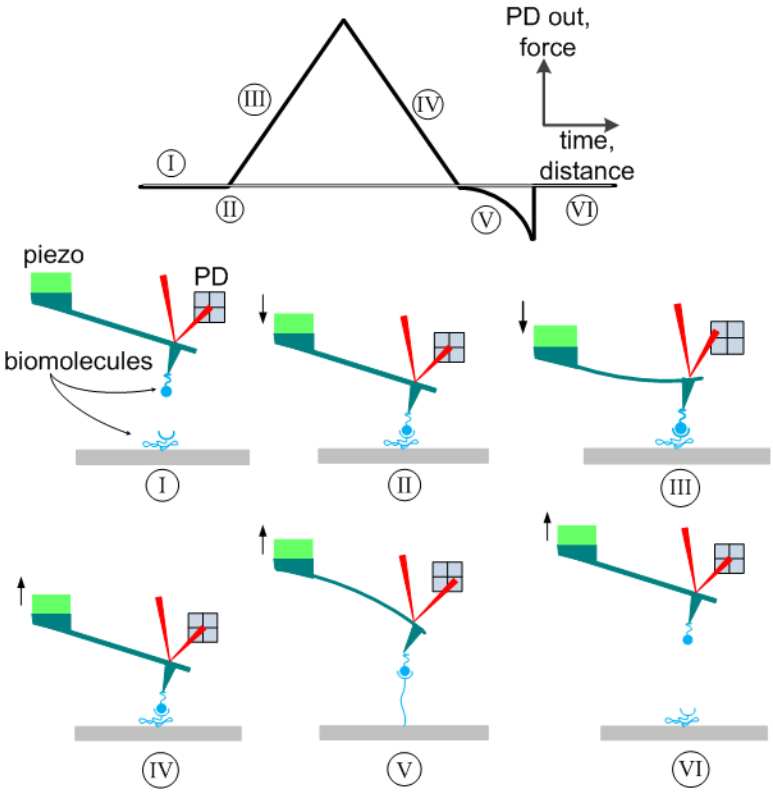


Figure 5.3. A Typical Photodetector Signal Recorded during a Biomolecular force Spectroscopy Experiment Together with the Corresponding Situation of Cantilever and Surface.

The liquid medium used in the AFM experimentation exerts a drag force to the cantilever that causes a constant upward bending proportional to the velocity of the piezo and the viscosity of the medium, which varies with temperature, during its approach/retract cycle. This hydrodynamic effect causes a constant bias added to the measured unbinding forces especially at high velocities [146]. Therefore, the measured unbinding forces should be corrected taking into account this measured drag force. Additionally, there are other factors affecting the drag force exerted on the cantilever, such as; geometry of the cantilever, the topography of the sample, etc. [159, 160]. Thus this viscous drag force should be determined and extracted from the measurements in AFM force spectroscopy experiments.

Lee *et al.* [161] and Florin *et al.* [158] first used force scan mode of AFM to measure the unbinding force of avidin-biotin interaction in 1994. Since then, many

protein-protein interactions have been studied. Some of the recent studies and the determined unbinding forces together with the reported Bell-Evans model parameters are listed in Table 5.1.

Table 5.1. A Partial list of Reported Atomic force Microscope Unbinding Studies.

Molecule Pair	Loading rate interval (pN/sec)	Measured Rupture forces (pN)	Barrier width ($\tilde{A}$)	Ref.
Aptamer/IgE	12E3-17E5	50-190	0.91, 2.54	[162]
Digoxigenin/antibody	3E1-63E3	20-80	0.35, 1.15	[163]
LFA-1/ICAM-2	5E1-6E4	20-120	0.31, 4.5	[164]
PDZ domain/peptide	38E2-14E4	40-220	0.4, 2.1	[165]
SfiI/DNA	21E2-63E4	25-100	1.8	[166]
Streptavidin-biotin	5E-2-6E4	5-170	1.2, 5	[167]
GTPase Ran/impÃ	3E2-8E4	40-90,75-160*	N/A	[168]
Transferrin(holo)/receptor	4E2-7E4	40-140	1.5, 9.3	[169]
Transferrin(apo)/receptor	5E2-4E4	25-40	8.1	[169]
VLA-4/VCAM-1 (WT)	3E1-2E5	15-130	1, 5.5	[170]
VLA-4/VCAM-1 (D40A)	3E2-1E5	25-70	5.9	[170]
VLA-4/VCAM-1 (Q38G)	2E2-2E5	25-100	1.7, 5.8	[170]
VLA-4/VCAM-1 (L43A)	2E2-2E5	20-100	1.5, 5.7	[170]
VLA-4/VCAM-1 (D143A)	3E2-2E5	25-140	0.95, 5.8	[170]

*Two populations of unbinding forces were reported, reflecting the existence of two conformational states in the Ran/impβ complexes.

ICAM-2, intercellular adhesion molecules 2; IgE, immunoglobulin E;
LFA-I, leukocyte function-associated antigen-I;
VCAM-l, vascular cell adhesion molecule-I; VLA-4, very late antigen-4.

5.3. Tip and Sample Functionalization Strategies

In biomolecular AFM force spectroscopy experiments, one of the molecules (protein or ligand) is immobilized on the surface of a tip mounted at very end of the cantilever while the other is attached to the surface. The two surfaces bring into contact to make intermolecular interactions. Then the unbinding forces of the interested molecule pair are experimentally determined from retraction curves, which is the deflection of the cantilever. The proper functionalization of the corresponding molecules to the tip and surface is thus crucial for the success of this type of study. It must guarantee that the measured forces are specific to the studied intermolecular bonding. The most commonly used techniques to functionalize the molecules to the cantilevers are

physisorption [158] and chemisorption [171]. In order to provide greater mobility and better rearrangement of the molecule functionalized on the tip to access the molecule on the surface, a linker molecule (usually a polyethylene glycol (PEG) linker) can be used between cantilever tip and the molecule of interest that will be functionalized on the tip.

Similar immobilization techniques can also be used to functionalize the corresponding binding partners to a suitable flat surface. Mica, polystyrene, and glass are most common substrates used for functionalization of proteins to the surface because they are atomically flat it is easy to obtain a perfectly clean surface by fresh cleaving.

Although both tip and surface functionalization protocols are well established, one should be aware of that strength of functionalization depends on the types of cantilevers or surfaces, alongside of molecules of interest. For example, if a biotin-streptavidin interaction is going to be used in a multi-functionalization protocol of the interested molecule to the AFM tip, one should keep in mind the range of biotin-streptavidin interaction forces (i.e. 20-400 pN). Thus, the quality of functionalization of both AFM tips and surfaces should be validated before performing actual force measurements.

5.4. Energy Landscape of Biomolecular Interactions via AFM Measurements

An AFM based biomolecular force spectroscopy measurements reveals the unbinding forces of an individual biomolecular complexes under a constant pulling velocity. The rate of change in the tension applied to break the intermolecular bond(s), $r_f = df/dt$, is defined as "loading rate". The loading rate applied to intermolecular interactions can be controlled experimentally by changing the retraction velocity and/or using a different cantilever with different spring constant. It has been shown that the unbinding of the intermolecular bond (s) is a stochastic process and the force required to break those bonds is proportional to the logarithm of the loading rate [172]. The relation between the measured unbinding forces and the applied loading rate can be

used to expose the binding energy landscape of the interested molecular complex.

The effect of external pulling force on the disassociation of a biomolecular complex was first formulated by Bell in 1978 [173]. Evans and Ritchie further improved this formulation in 1997 [172]. Their approach is called Bell-Evans model, which characterizes the relation between an external pulling force and dissociation rate of intermolecular bond (s), which will yield the binding energy landscape. The latter approach has been refined and generalized in recent studies[174-176]. The Bell-Evans model is based on the typical transition-state theory, where separation of two molecules requires overcoming an activation energy barrier. If there exists a single energy barrier on the binding energy landscape of the biomolecular complex, the dissociation rate of this complex can be expressed as

$$k_{off} = \alpha \frac{k_B T}{h} exp\left(\frac{-\Delta G^0}{k_B T}\right) \tag{5.1}$$

where α is a prefactor identifying the potential energy well, k_B is Boltzmann's constant, T is the absolute temperature, h is Planck's constant, and ΔG^o is the activation energy. When an external pulling force is applied to the complex, an energy term which is the work done to the system by the applied pulling force, -fx, is added to the potential of the system. This additional energy reduces the activation energy barrier by fx_β, where x_β the barrier width between the bound complex and the activated complex. Thus, the dissociation rate at the existence of an external pulling force becomes

$$k_{off}(f) = \alpha \frac{k_B T}{h} exp\left(\frac{-(\Delta G^0 - fx_\beta)}{k_B T}\right) = k_{off} exp\left(\frac{fx_\beta}{k_B T}\right) \tag{5.2}$$

where k_{off} is dissociation rate without any external pulling force and the k_{off} (f) is the dissociation rate when a pulling force is applied. As can be seen from Equation 5.2, the dissociation rate of the complex exponentially increases as the pulling force increases. The two parameters k_{off} and x_β characterize the binding energy landscape, where they often referred as Bell model parameters. The height of the activation energy barrier is characterized by k_{off} and the width of the barrier is x_β that corresponds to the force resistance of a molecular complex as illustrated in Figure 5.4. If a complex has small

x_β, which means a narrow activation energy barrier, then the effect of the pulling force on the force-dependent dissociation rate k_{off} is small; and if it is high, fx_β will be higher, which means the complex is more sensitive to an external force.

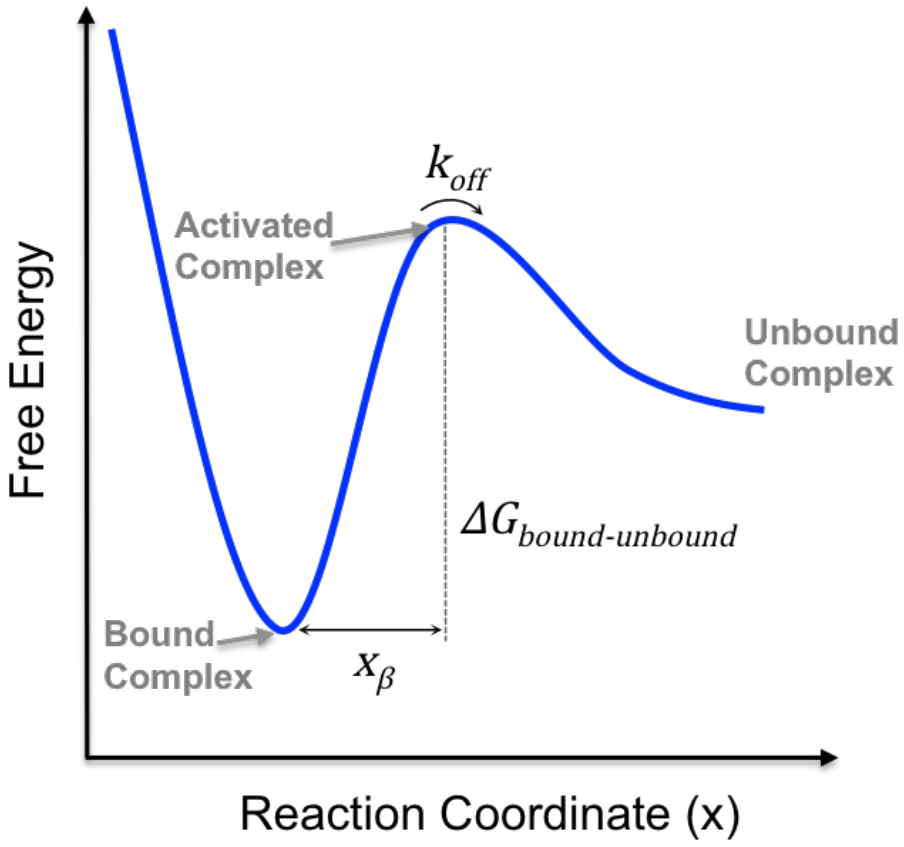


Figure 5.4. Free Energy Landscape of the Dissociation of a Biomolecular Complex, Together with the Associated Bell Model Parameters.

Equation 5.2 depicts the changes in the bond dissociation under a constant external pulling force. Since it is difficult to keep the pulling force constant in an AFM force spectroscopy experiment, a dynamic force approach is usually used to characterize the dissociation of a biomolecular complex under an external force [172, 177]. According to theoretical framework, the probability density function for the forced unbinding of a complex under constant loading rate r_f is,

$$P(f) = k_{off} exp\left(\frac{fx_\beta}{k_BT}\right) exp\left\{\frac{k_{off}k_BT}{r_f x_\beta}\left[1 - exp\left(\frac{fx_\beta}{k_BT}\right)\right]\right\} \tag{5.3}$$

Then, the most probable rupture force f* (i.e., the maximum of the probability distribution function $(\partial P(f))/\partial f = 0$) can be obtained from Equation 5.3 as

$$f* = \frac{k_B T}{x_\beta} \ln\left(\frac{x_\beta}{k_{off} k_B T}\right) + \frac{k_B T}{x_\beta} \ln\left(r_f\right) \tag{5.4}$$

Equation 5.4 shows that, there is a linear relation between the most probable rupture force $f*$and the natural logarithm of the loading rate r_f. Therefore, the Bell model parameters can be obtained from the best fitting straight line of f* versus $\ln(r_f)$ as shown in Figure 5.5, the dynamic force spectrum (DFS) of the complex [172].

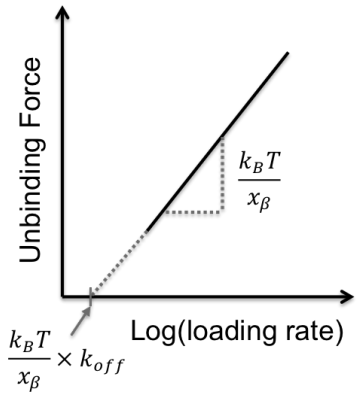


Figure 5.5. The Linear Relation between the Unbinding force and the Logarithm of the Loading Rate. Bell Parameters can be Extracted from the Slope and the Intercept of the Linear Relation [177].

The relation given in Figure 5.5 is shown to be correct for slip bonds (bonds that slip apart easily in the presence of an external force) only. For slip bonds the external pulling force tilts the energy surface and decreases the activation energy barrier. Therefore, increasing force decreases the lifetime of biological bonds. However, catch bonds (bonds act like molecular hooks that dissociate easily in the absence of external force but hold firm when stretched by external forces) have the opposite behavior. In this kind of bonds, it is observed that the applied external force increases the lifetime of the biological bonds [178, 179].

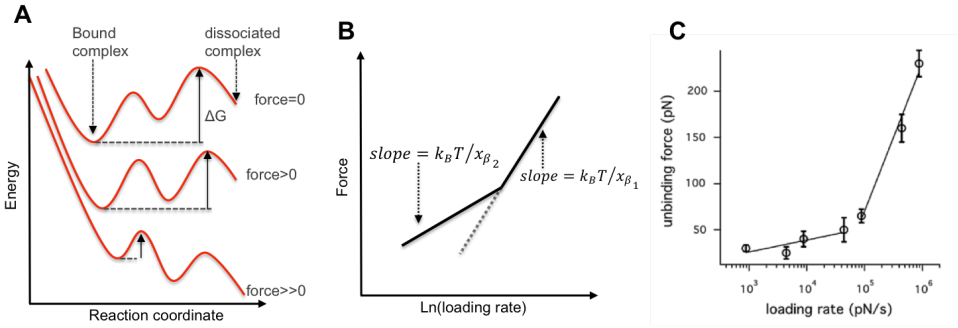


Figure 5.6. Energy Landscape of a Molecular Complex having two Transition States and the Corresponding DFS Results. a) Two-Barrier model in Energy Landscape b) Corresponding two Linear Regimes. c) Dynamic force Spectra of Biotin-Streptavidin Complex [156].

The binding energy landscape of a biomolecular complex may have more than one activation energy barriers from bound to unbound forms as depicted in Figure [19, 23]. Existence of multiple linear regimes with different slopes is associated with multiple energy barriers on the dissociation reaction of a molecular complex. The simplest case is the two-barrier model as shown in (Figure 5.6A). The energy landscape of the complex has two transition states along the reaction coordinate, which is the pulling direction in the absence of an external force (top). In this case, the outer energy barrier determines the dissociation kinetics of the complex. Small external force corresponding to the lower loading rates reduces the outer barrier and further increase in force results in switching the main barrier from outer transition state to the inner one. This switching in transition states results in two different linear regimes with ascending slopes in the relation of unbinding force and the logarithm of loading rate. The increase in slope indicates that the outer barrier has been restrained by force and the inner barrier dominates the dissociation reaction (Figure 5.6B). Multiple activation barriers in the energy landscape were found in a number of biomolecular complexes, including the (strep) avidin/biotin complexes as shown in (Figure 5.1C). The best-fitted curves (solid lines) were obtained using Equation 5.4 applied to each of the two loading regimes. Error bars are the standard error of the mean (SEM). Other examples of molecular pairs having multiples transition states are tabulated in Table 5.1.

5.4.1. Interpretation of Multiple Energy Barriers along with Biomolecular Disassociation

AFM is used to study many biomolecular pairs. Using DFS methodology, it has been found that the dissociation of many protein complexes involves one or multiple energy barriers where some of these studies are listed in Table 5.1. The molecular and structural constitutes that create these energy barriers are revealed by mutating residues and/or changing physiological conditions that considered to affect the interested biomolecular interactions.

Nevo and coworkers analyzed the interaction between the small GTP (guanosine triphosphate) bound GTPase Ran and its binding partner nuclear import receptor importin β1 (impβ) to answer the question whether the change in the stability of Ran-impβ interaction obeys induced fit or dynamic population shift models using DFS [168]. Ran regulates the association and dissociation of receptor-cargo complexes through interacting with the impβ where release of Ran from impβ is induced by small effector protein Ran-binding protein RanBP1. They found that GppNHp, which is a nonhydrolyzable analog of GTP, bound Ran and impβ complex has two activation energy barriers and thus two distinct bound states (ruptured at relatively low and high unbinding forces corresponding lower-strength and higher-strength states, respectively). Addition of the effector protein RanBP1 into the solution abolishes the higher-strength population and increases the population of lower-strength state. However, RanBP1 did not change the most probable unbinding forces meaning that the strength of interaction does not depend on the existence of the effector protein [180]. These observations indicate that the allosteric regulation of GTP bound Ran and impβ by RanBP1 is accomplished by a dynamic shift in the pre-existing conformational states in the ensemble of conformation of the complex where the effector protein selectively binds to the lower-strength state. Additionally, the low affinity GDP bound Ran and impβ forms a weak complex characterized by unimodal distribution of rupture forces in the absence of the effector protein. Introducing RanBP1 into the solution results in the measurement of higher rupture forces as well as increasing the probability of adhesion (the number of unbinding events with respect to the number of approach/retract cycle). These results

indicate that RanBP1 stimulates Ran-GDP and impβ complex formation as well as amplifying the binding strength by inducing a conformational change in the complex implying an induced-fit mechanism. The authors also studied the effect of the mutation on the Ran and impβ complex. RanQ69L mutant, where it is analogous to the well-studied and potentially oncogenic RasQ61L, considerably lowers the GTPase activity of Ran. According to DFS results of the GDP bound RanQ69L and impβ has similar force spectrum with the GDP bound wild-type Ran and impβ. However, the mutation increases the unimodal distribution of rupture forces in GTP-bound RanQ69L form compared to the wild-type Ran analog, suggesting that RanQ69L-GTP abolishes one bound state and makes only the lower-strength state accessible that is again implying a population-shift mechanism.

In another study, Ritco-Vonsovici *et al.*, [181]) elucidated the relation between the differentiation in conformational states and multifunctioning of β-catenin using DFS and MD simulations. To test idea that whether the two main function (cell adhesion and control of cell differentiation, development) of β-catenin originates from the two molecular form of the protein or not, the interaction of β-catenin with transcription factor Tcf4 was studied using AFM. They obtained a bimodal force distribution in the unbinding process of β-catenin/Tcf4 complex at low loading rates, implying two distinct populations in the ensemble of conformations of the complex where it shifts to a single population as loading rate increases (only lower-strength state was survived). Addition of selective inhibitor (ICAT) that prevents β-catenin/Tcf4 interaction reduced the probability of lower-strength state at low loading rates while the interaction totally disappeared at high loading rates (both lower and higher-strength states). These findings show that binding of the inhibitor either selectively inhibits one of the states or inhibits them both implying inhibitor adapts a conformational selection mechanism. The MD simulation of apo β-catenin results indicated that there are at least two conformers of the protein differentiating at the orientation of the flexible regions at its binding site. Therefore, the two different population of β-catenin/Tcf4 dissociation reaction may originate from the two distinct conformational isomers of β-catenin.

Existence of multiple energy barriers was also found in dissociation reaction of the iron transporter protein transferrin (Tf) and its receptor (TfR) [169]. Yersin et al. functionalized iron-loaded Tf (holo-Tf) or iron-free Tf (apo-Tf) to the AFM tip and TfR to the surface to study the differences in binding mechanism of both apo and holo Tf with TfR and the effect of pH on the complex formation where holo to apo transition is favored at low pH values. They found that the unbinding of holo-Tf from TfR requires higher forces compared to the apo-Tf. In addition, the dissociation of holo-Tf-TfR complex requires overcoming of two energy barriers, whereas apo-Tf-TfR complex includes only one. Moreover, the strength of apo-Tf-TfR interaction is pH dependent. The variation in the energy landscape of holo and apo form of the Tf-TfR interaction could be considered as supportive argument to the idea that the holo-Tf interacts with TfR through its two lobes, whereas apo Tf binds TfR by its one lobe only [182].

Zhang *et al.*, [170] employed DFS to understand the dissociation mechanism of the $\alpha 4\beta 1$ and VCAM-1 complex for both wild type and different mutants of VCAM-1. $\alpha 4\beta 1$/VCAM-1 complex can resists large shear forces induced by the blood stream. Their results revealed that the dissociation of the $\alpha 4\beta 1$/VCAM-1 complex involves at least two energy barriers, an inner barrier and an outer barrier corresponding to larger and smaller rupture forces, respectively. This two barriers kinetics can be considered as reflection of the biophysical basis of two different physiological functions (i.e., cell rolling and firm adhesion) of the $\alpha 4\beta 1$/VCAM-1 interaction. D40A mutation at the binding interface of VCAM-1 decreases the unbinding forces of $\alpha 4\beta 1$/VCAM-1 complex supporting the idea of stabilization of the complex by electrostatic interactions. Other mutational studies at different structural elements of VCAM-1 provide molecular explanation to the functional roles of these regions on $\alpha 4\beta 1$/VCAM-1 interaction.

6. MATERIALS AND METHODS

6.1. Gaussian Network Model

The fluctuation dynamics of proteins can be effectively described by coarse-grained normal mode analysis. In the elastic network model description, carbon alpha of each amino acid residue is taken as a node and these nodes are connected by springs with the "neigboring nodes" (see Figure 6.1). The "neigboring nodes" are defined here by a cut-off radius; the nodes that have distances less than the defined cut-off distance are thought to be connected by linear springs. In this model, neither connectivity nor residue types are taken into consideration [7]. This model is simply referred as Gaussian network model (GNM) if the fluctuations are assumed to be isotropic, i.e. with no directional preferences. The studies have shown that GNM is in accordance with X-ray crystallographic Debye-Waller factors [7, 9], the H/D exchange free energies [183] and the order parameters from NMR-relaxation measurements [184].

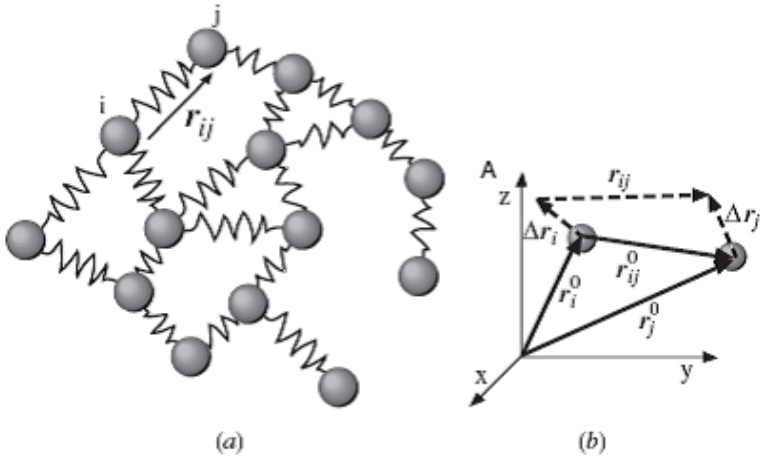


Figure 6.1. Description of the Gaussian Network model (GNM).

In the GNM, in order to calculate the harmonic potential, firstly, the fluctuation vector should be calculated. The fluctuation vector, is simply the difference between the instantaneous position vector of i_{th} node, R_i, and the equilibrium position vector

of i_{th} node; R_i^o.

$$\Delta R_i = R_i - R_i^0 \tag{6.1}$$

The fluctuation vector, is the squareroot of sum of squares of the difference in X, Y and Z coordinates:

$$\Delta R_i = \sqrt{(x_i - x_i^o)^2 + (y_i - y_i^o)^2 + (z_i - z_i^o)^2} \tag{6.2}$$

The vector R_{ij} defines the position vector from i_{th} node to j_{th} node; whereas ΔR_{ij} defines the fluctuation of this vector:

$$\Delta R_{ij} = R_{ij} - R_{ij}^0 = \Delta R_j - \Delta R_i \tag{6.3}$$

If the fluctuations are assumed to be isotropic and Gaussian; the potential function can be found to be as the following;

$$V = \frac{\gamma}{2}\left[\sum_{ij}^{N} \Gamma_{ij}\left[(\Delta X_i - \Delta X_j)^2 + (\Delta Y_i - \Delta Y_j)^2 + (\Delta Z_i - \Delta Z_j)^2\right]\right] \tag{6.4}$$

In this equation, γ is the force constant for the spring; and G stands for the symmetric Kirchoff matrix; which represents the connectivity between nodes. In order to calculate the connectivity element of Kirchoff matrix between i_{th} and j_{th} residues; the following constraints are used:

$$\Gamma_{ij} = \begin{cases} -1 \text{if} \mathrm{i} \neq \mathrm{j} \text{and} \mathrm{R}_{ij} \leq \mathrm{r}_c \\ 0 \text{if} \mathrm{i} \neq \mathrm{j} \text{and} \mathrm{R}_{ij} \leq \mathrm{r}_c \\ -\sum\limits_{i,j \neq i} \Gamma_{ij} \text{if} \mathrm{i} = \mathrm{j} \end{cases} \tag{6.5}$$

By using the GNM, the goal is to determine either the mean square fluctuation of i^{th} node (residue), or the cross-correlation between i^{th} and j^{th} residues.

$$\left\langle \Delta R_i^2 \right\rangle = \langle \Delta R_i . \Delta R_i \rangle = \left\langle \Delta X_i^2 \right\rangle + \left\langle \Delta Y_i^2 \right\rangle + \left\langle \Delta Z_i^2 \right\rangle \tag{6.6}$$

$$\langle \Delta R_i . \Delta \mathrm{R}_j \rangle = \langle \Delta X_i . \Delta X_j \rangle + \langle \Delta Y_i . \Delta \mathrm{Y}_j \rangle + \langle \Delta Z_i . \Delta Z_j \rangle \tag{6.7}$$

Using the Kirchoff matrix, the cross-correlations are determined via diagonalization and inversion:

$$\langle \Delta R_i . \Delta \mathrm{R}_j \rangle = \frac{3k_B T}{\gamma} \left[\Gamma^{-1} \right]_{ij} = \frac{3k_B T}{\gamma} \left[U \Lambda^{-1} U^T \right]_{ij} \tag{6.8}$$

$$\langle \Delta R_i . \Delta \mathrm{R}_j \rangle = \frac{3k_B T}{\gamma} \sum_k \left[\lambda_k^{-1} u_k u_k^T \right]_{ij} \tag{6.9}$$

The matrix Λ, is the eigenvalue matrix; and represents the frequencies of the modes. Matrix U, is the eigenvector matrix corresponding to these modes. The eigenvalue matrix will have N-1 (residue number-1) nonzero elements, and the lowest eigenvalues will correspond to the slowest modes; which represent the collective and global motions. The highest eigenvalues however, will correspond the fastest modes. The eigenvalues are sorted in an ascending order for convenience.

An extension of the GNM, called anisotropic network model (ANM), has been proposed to incorporate the anisotropic effects on fluctuation dynamics [6, 7, 139]. The large scale collective motions obtained from both GNM and ANM have been shown to be closely related to those extracted from atomistic molecular dynamics studies for amylase inhibitor tendamistat [114]. This lends support to the effectiveness of these coarse-grained models in the analysis of the structure-function relationship of proteins

and their complexes.

6.2. Anisotropic Network Model

Elastic network models are constructed based on the three-dimensional (3D) folded structure of proteins, which have been determined by experimental techniques such as X-ray crystallography or nuclear magnetic resonance (NMR). The total potential of a protein structure with N interaction sites, i.e. residues, is given by,

$$V = (\gamma/2)\,\Delta R^T H \Delta R \tag{6.10}$$

where ΔR is 3N-dimensional vector of the fluctuations ΔR_i in the position vectors R_i of the individual sites, ΔR^T is its transpose, and H is the Hessian matrix composed of the second derivatives of the potential. The force constant γ is taken to be identical for all bonded and nonbonded interactions. Explicitly, the total potential energy can be written a summation over all harmonic interactions of (i, j) pairs that fall within a cutoff distance of r_c.

$$V = (\gamma/2) \sum_i \sum_j h\,(r_c - R_{ij})\,(\Delta R_j - \Delta R_i)^2 \tag{6.11}$$

Here, $h(x)$ is the heavyside step function $[h(x) = 1if \geq x0, andzerootherwise]$, R_{ij} is the distance between sites i and j.

Hessian matrix is expressed as a function of N^2 submatrices H_{ij} of the form

$$H_{ij} = \begin{bmatrix} \partial^2 V/\partial X_i \partial X_j \partial^2 \mathrm{V}/\partial \mathrm{X}_i \partial \mathrm{Y}_j \partial^2 V/\partial X_i \partial Z_j \\ \partial^2 V/\partial Y_i \partial X_j \partial^2 \mathrm{V}/\partial \mathrm{Y}_i \partial \mathrm{Y}_j \partial^2 V/\partial Y_i \partial Z_j \\ \partial^2 V/\partial Z_i \partial X_j \partial^2 \mathrm{V}/\partial \mathrm{Z}_i \partial \mathrm{Y}_j \partial^2 V/\partial Z_i \partial Z_j \end{bmatrix} \tag{6.12}$$

where X_i, Y_i and Z_i are the components of R_i.

The cross-correlations between the fluctuations of sites i and j are calculated from

$$\begin{aligned} < \Delta R_i.\Delta \mathrm{R}_j > &= (1/Z_n) \int \left(\Delta R_i.\Delta \mathrm{R}_j\right) \exp\left\{-V/k_b T\right\} \mathrm{d}\left\{\Delta R\right\} \\ &= (3k_B T/\gamma)\, \mathrm{tr} \lfloor H^{-1} \rfloor_{ij} \end{aligned} \tag{6.13}$$

including the mean-square fluctuation amplitudes $< (\Delta R_i)2 >$. Here, k_B is the Boltzmann constant, T is the absolute temperature in degrees Kelvin, Z_N is the conformational partition function, and $tr[H^{-1}]_{ij}$ is the trace of the $ijth$ submatrix $[H^{-1}]_{ij}$ of H^{-1}.

The dynamics of the protein around the minimum energy conformation can be represented as a collection of 3N-6 normal modes of motion. As a result, $< \Delta R_i.\Delta R_j >$ can be also expressed as $< \Delta R_i.\Delta R_j >= \Sigma[\Delta R_i.\Delta R_j]_k$ where the contribution of the k_{th} mode is

$$[\Delta R_i.\Delta \mathrm{R}_j]_k 0\, (3k_B T/\gamma)\, \mathrm{tr}\left\lfloor \lambda_k^{-1} u_k u_k^T \right\rfloor_{ij} \tag{6.14}$$

here λ_k is the k_{th} eigenvalue and u^k is the corresponding eigenvector. The eigenvalues represent the frequencies of the individual modes. The most important point is to analyze the slowest modes playing a dominating role in collective dynamics of the structure and in biological function.

6.3. Molecular Dynamics

The methodology of Molecular Dynamics Simulations is based on Newton's second law,

$$\begin{aligned} F_i &= m_i a_i \mathrm{or} \\ \tfrac{F_i}{m_i} &= \tfrac{d^2 x_i}{dt^2} \end{aligned} \tag{6.15}$$

where F_i is the force exerted on the particle i, m_i is its mass and α_i is its acceleration. Starting with the knowledge of the force on each atom, determination of the acceleration of each atom in the system is possible. Integration of the second form of the equation yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. The method is deterministic; meaning that knowing the positions and velocities of each atom, prediction of the state of the system at any time; future or past, is possible. Using this methodology, successive configurations of the system can be generated. In intermolecular interactions, the force applied on each particle changes whenever the particle itself of any other particle in interaction with it changes its position. This feature is implemented to the simulation with the use of continuous potential, in which the motions of all particles are coupled together [185]. At this stage, since molecular systems generally consist of a vast number of particles, it becomes impossible to find the properties of such a complex systems analytically. Therefore, MD simulation integrates the equations by using numerical methods.

The force exerted on each particle at any time during simulation, Fi is calculated from a forcefield in MD simulations. The forcefield defines the potential energy of a system as a function of the atomic positions/coordinates, and the Fi are obtained from derivatives of the potential function.

$$\begin{aligned}(R_1,\ldots,R_N) = &\sum_{bonds} \tfrac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \tfrac{k_i}{2}(l_i - l_{i,0})^2 + \\ &\sum_{torsions} \tfrac{V_n}{2}\left(1 + \cos\left(nw - \gamma\right)\right) + \\ &\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4_{ij}\left[\left(\tfrac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\tfrac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \tfrac{q_i q_j}{4\pi_0 r_{ij}}\right)\end{aligned} \tag{6.16}$$

The potential energy of particle i as a function of all R_j , is defined in terms of interactions between bonded atoms, bond angle and torsional angle potentials, and electrostatic and van der Waals interactions between non-bonded atoms. The first term in the equation describes the interaction of pairs of bonded atoms, where li is the bond length. The second term is similarly the summation over all the angles in the molecule modeled using a harmonic potential, where θ_i is the angle between the three successive atoms. Torsional potential describes the change in energy when a bond rotates, and is

depicted with the third term in the equation. The fourth contribution in the equation is for the non-bonded atoms, which are separated by at least three atoms. The non-bonded interactions are defined by two different potentials. The former one is the Lennard-Jones 12-6 potential function that accounts for van der Waals interactions, whereas the latter one is the Coulomb potential for electrostatic interactions [185]. From the potential function, the Fi are generated as:

$$F_i = -\nabla V_i \left(R_i, \ldots, R_N\right) = -\frac{\partial V \left(R_i, \ldots, R_N\right)}{\partial R_i} \tag{6.17}$$

Knowing the state of the structure at time t, the positions, velocities and accelerations are approximated by Taylor series, where α is the acceleration (2^{nd} derivative), b is the 3^{rd} derivative etc. (185).

$$\begin{aligned} r\left(t+\delta t\right) &= r\left(t\right) + \upsilon\left(t\right)\delta t + \tfrac{1}{2}\alpha\left(t\right)\delta t^2 + \ldots \\ \upsilon\left(t+\delta t\right) &= \upsilon\left(t\right) + \alpha\left(t\right)\delta t + \tfrac{1}{2}b\left(t\right)\delta t^2 + \ldots \\ \alpha\left(t+\delta t\right) &= \alpha\left(t\right) + b\left(t\right)\delta t + \ldots \end{aligned} \tag{6.18}$$

The most common algorithm used in integration is Verlet algorithm. Positions and accelerations at a time t and positions from time (t-δt) are used to calculate new positions at time (t + δt). Then the velocities can be obtained from the difference in positions. From generated values, new positions can be obtained successively [185]. Giving the appropriate equations:

$$\begin{aligned} & r\left(t+\delta t\right) = r\left(t\right) + \upsilon\left(t\right)\delta t + \tfrac{1}{2}\alpha\left(t\right)\delta t^2; r\left(t-\delta t\right) \\ & = r\left(t\right) - \upsilon\left(t\right)\delta t + \tfrac{1}{2}\alpha\left(t\right)\delta t^2 \\ & r\left(t+\delta t\right) = 2r\left(t\right) - r\left(t-\delta t\right) + \tfrac{1}{2}\alpha\left(t\right)\delta t^2 \\ & \upsilon\left(t\right) = \tfrac{r(t+\delta t)+r(t-\delta t)}{2\delta t} \end{aligned} \tag{6.19}$$

As can be seen from the structure of the algorithm, to be able to start a MD simulation and generate configurations of the system, the initial state should be defined.

Either experimental or theoretical inputs can be used at this stage. In addition to the positions of the elements of the network, the initial velocities should also be defined.

Experimental inputs are generally used for the atomistic coordinates of a structure, in the form of X-ray or NMR structure. The structure is subjected to energy minimization before the starting the simulation.

The initial velocities are generated by theoretical methods. A commonly used method is randomly selecting the initial velocities from a Maxwell-Boltzman distribution, for the specified temperature [185].

$$p\left(\upsilon_{ix}\right) = \left(\frac{m_i}{2\pi k_B T}\right)^{\frac{1}{2}} \exp\left(-\frac{m_i \upsilon_{ix}^2}{2k_B T}\right) \tag{6.20}$$

6.4. Metadynamics

Metadynamics is an enhanced sampling method developed by Laio and Parrinello [69, 186] to sample rare events and reconstruct free energy landscape, where a molecular system is simplified by means of reduced degrees of freedom. A set of variables referred as reaction coordinates, collective variables (CVs) or order parameters, were defined. The basic concepts of this methodology are as follows;

A system consisting of N number of particles interacts through a potential energy function $V(r)$ as a function of conformational/configurational coordinates r, which contributes to its free energy (E). The free energy difference (ΔE) is more significant than its absolute value (E). Calculating a difference implies the existence of at least two different states. Often these two different states can be distinguished along a few parameters, called reaction coordinates or collective variables (CVs). In molecular conformational space, these CVs, $\xi = (\xi_1, \xi_2, \ldots \xi_{N_{CV}})$, are often time dependent functions of configurational coordinates r and its values $\zeta \epsilon R^{(N_C V)}$ where $N_{CV} = 3N$ is the total number of CVs. They can be considered as degrees of freedom of the *phasespace* that represents the thermodynamic state of the system. The system evolving under

the action of a dynamics (i.e. Molecular Dynamics) may stuck in some local minima of V(r), if barriers are higher than the energy generated by thermal fluctuations (see Figure 2.3). The main idea of metadynamics is to add a history-dependent potential constructed as a sum of Gaussian distributions centered along the simulation in the CVs space to bias the system not to sample previously visited states in the conformational ensemble. The metadynamics bias potential expressed here is implemented as in [69, 100, 186],

$$V_B(\xi, t) = \sum_{\substack{t' = k\tau_G \\ k \in N}}^{t'<t} w \prod_{i=1}^{N_{CV}} e^{-\frac{[\xi_i - \xi_i(t')]^2}{2\sigma_i^2}} \tag{6.21}$$

where τ_G is the time interval of N_{CV}-dimensional Gaussian functions deposition. The height of the Gaussian "hills" is a user-defined energy constant w whose width is σ_i and the centers of the hills are previously sampled conformations $(\xi(\tau_G), \xi(2\tau_G), \ldots)$. A schematic representation of this deposition process is shown in Figure 6.2. The Gaussian height w, the Gaussian width σ_i, and the deposition time interval τ_G are the user-defined simulation parameters. Lower values of w and higher values of the hill frequency τ_G provide more accurate sampling but increase the simulation time to complete a potential of mean force (PMF) calculation.

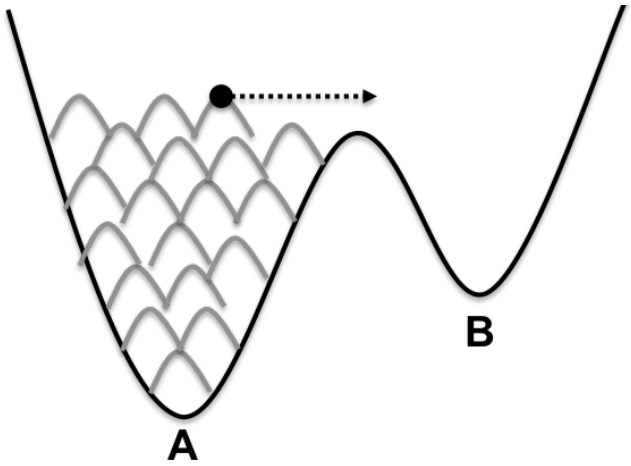


Figure 6.2. Schematic Representation of Gaussian Functions Deposition of the Potential energy of the system.

By adding the metadynamics bias potential, the Hamiltonian of the system becomes time dependent as,

$$\mathcal{H}_B\left(r,p,t\right) = K\left(p\right) + V\left(r\right) + V_B\left(\xi\left(r\right),t\right) \tag{6.22}$$

where, r is the conformational state (conformational/configuration coordinate) of the system, p is the momentum, t is time, $K(p)$ and $V(r)$ represent kinetic and unbiased potential energy of the system, $V_B(\zeta,t)$ is the biased potential.

And the force underlying this biased dynamics, where the potential energy is the sum of unbiased biased potentials, becomes

$$F_i^B = -\nabla_{r_i} V\left(r\right) - \nabla_{r_i} V_B\left(\xi,t\right) = F_i + F_i^{meta} \tag{6.23}$$

The explicit form of the metadynamics forces is obtained by inserting Equation 6.21 into Equation 6.22 as follows

$$\begin{aligned} F_i^{meta} &= -\nabla_{r_i} V_B\left(\xi,t\right) \\ &= -\sum_{\substack{t'=k\tau_G \\ k\in N}}^{t'<t} w\nabla_{r_i} \prod_{i=1}^{N_{CV}} e^{-\frac{\left[\xi_i-\xi_i(t')\right]^2}{2\sigma_i^2}} \\ &= \sum_{\substack{t'=k\tau_G \\ k\in N}}^{t'<t} w\left(\sum_{b=1}^{N_{CV}} \frac{\xi_b-\xi_b(t')}{\sigma_b^2}\nabla_{r_i}\xi_b\right) \prod_{i=1}^{N_{CV}} e^{-\frac{\left[\xi_i-\xi_i(t')\right]^2}{2\sigma_i^2}} \end{aligned} \tag{6.24}$$

As can be seen from Equation 6.24, the metadynamics forces are re-summations of the deposited Gaussians, scaled by the derivative $\nabla_{r_i}\xi = (\nabla_{r_i}\xi_1, \ldots, \nabla_{r_i}\xi_{N_{CV}})$ of the CV ζ with respect to the original coordinates r_i. The CV ζ should be chosen based on the criteria that $\nabla_{r_i}\xi$ should be calculated analytically.

Theoretically, the generated trajectory using metadynamics is an out-of-equilibrium dynamics. However, if very large τ_G is employed, the variation of $V_B(\xi, t)$ is very slow. Therefore, it can be considered as between Gaussian depositions, the trajectory of the system reveal equilibrium dynamics (classical molecular dynamics). The main assumption of standard metadynamics is that the bias potential is ultimately an unbiased estimator for the free energy as a function of the reaction coordinate. This assumption could be understood using the aforementioned slow-deposition argument as

$$\lim_{\tau \to +\infty} V_B(\xi, \tau) \quad - F(\xi) \tag{6.25}$$

It has been demonstrated that, after a transient time τ, which corresponds to the time required to fill all the relevant free energy minima of the system, $V_B(\zeta, t)$ reaches a stationary state in which it grows evenly around and average. However, this argument is only qualitative, and there is no proof of convergence that gives Equation 6.25.

6.4.1. Well-Tempered Metadynamics

Although metadynamics has many benefits to accurate sampling of the conformational ensemble, it has one main drawback. The bias potential does not converge to the free energy but oscillate around it in a single simulation [100, 187]. This handicap has two consequences; 1) Since the bias potential will overfill the underlying FES after sampling all minima, it will bring the system toward non-physical high energy regions in the CVs space. 2) It will be hard to decide when to stop the simulation. Well-tempered metadynamics [100] emerged as a solution to these problems. In well-tempered metadynamics, the height of the Gaussian function decreases as the simulation proceeds as

$$w(t') = w_0 e^{\left(-\frac{V_B(\xi,t)}{k_B T}\right)} \tag{6.26}$$

where w_0 is he initial Gaussian height, ΔT is a scaling parameter with the dimension of a temperature, and k_B is the Boltzmann constant. With this rescaling, the bias

potential does not fully compensate the underlying FES but it converges as in Equation 6.27 in the long time limit as

$$\lim_{\tau \to +\infty} V_B(\xi, \tau) = -\frac{\Delta T}{T + \Delta T} F(\xi) \tag{6.27}$$

This rescaling provides avoiding from overfilling of FES and also reduces the computational time when a large number of CVs are used [187].

6.4.2. Bias-Exchange Metadynamics

Accurate description of the free energy surface by metadynamics approach strongly depends on the choice of CVs. If an important vivid parameter is missing in controlling the functional free energy change of the system, the resulting free energy surface will have large fluctuations and errors. Moreover, the performance of the algorithms rapidly decreases as the dimensionality of the CV space increases. Unfortunately, complex reactions such as protein folding, conformational change, and biomolecular recognition acquire high degrees of freedom. It is thus often difficult to select a priori a limited number of variables accurately describing the underlying energy surface. The bias-exchange metadynamics (BexMetaD) [103] method allows to overcome the abovementioned difficulties. BexMetaD appears a combination of replica exchange and metadynamics. In this methodology, multiple metadynamics simulations at the same temperature are performed where each replica is biased with a time-dependent potential function acting on different CVs. Exchanges between the bias potentials (exchange of the configurations) acting on different CVs are periodically allowed according to the Metropolis scheme with the probability,

$$P_{ab} = min\left(1, e^{\beta\left(V_B^a(r^a,t) + V_B^b\left(r^b,t\right) - V_B^a\left(r^b,t\right) - V_B^b(r^a,t)\right)}\right) \tag{6.28}$$

If the attempt to exchange the biases is accepted, the trajectory that was previously biased in the direction of CV a continues its evolution biased by CV b, and vice

versa. A relatively large number of different CVs can be used to bias the system through this approach and a high-dimensional space can be explored after a sufficient number of exchanges. Adding more CVs increases the computational cost of the simulation linearly, although this happens exponentially in normal metadynamics approach.

6.4.3. Selection of Collective Variables

Effective application of metadynamics to correct sampling of conformational ensemble and construction of underlying FES strongly depends on the choice of CVs. Although there is no universal recipe for the selection of CVs, a good CV should satisfy the following three points [187];

(i) They should distinguish distinct states of the system, namely; initial, final, and all relevant intermediate states.
(ii) They should include all slow degrees of freedom related to the process of interest.
(iii) They should not exceed a limited quantity

As long as they fulfill the above criteria, every possible function of the spatial coordinates can be selected as CVs. There are many ready to use CVs implemented in available MD simulation tools such as distances, bond and dihedral angles, radius of gyration, coordination number, hydrogen bonds, RMSD, the projection of atomic coordinates on a vector, the composition of secondary structural elements, etc. For example, distance or dihedral angle biases can be introduced into the simulation as CV given in the following formulations. As stated by Rossetti *et al.*, $CV_{distance}$ can be in the form as

$$CV_{distance} = \sum_{ij} \frac{1 - (\mathrm{r}_{ij}/\mathrm{r}_0)^8}{1 - (\mathrm{r}_{ij}/\mathrm{r}_0)^{10}} \tag{6.29}$$

$CV_{distance}$ can be introduced as in the above form in which r_{ij} is distance between the selected atoms and r_0 is distance parameter (distance parameter is a pre-defined value depending on the system of interest, i.e. 6 $\tilde{A}$ for the typical carbon-carbon distance

(Fiorin *et al.*, 2006)).

$$CV_{dihedral} = \sum_{i} (1 - \cos(\Psi_i - 45^\circ))^2 \tag{6.30}$$

$CV_{dihedral}$ can be introduced as in the above form in which ψ_i is the backbone dihedral angle between consecutive N-Cα-C-N atoms (Rossetti *et al.*, 2011).

The difficulty of choosing the right CVs describing the process of interest before doing any calculations is not a defect of metadynamics or any CV-based methods. It is sometimes necessary to do some trials and errors before getting the correct definition of the reaction coordinate.

A great number of CVs have been used in the literature, but they usually depend on the nature of the specific process or specific to the molecule of interest. Using any global property defined only from intrinsic properties that will reflect large conformational transitions of the molecule in metadynamics simulations would be highly beneficial in the field of computational structural biology. Collective modes of motion obtained either from the principal component analysis (PCA) of conventional MD trajectories or from ENM are good candidates to fill this deficiency.

6.5. Collective Modes Driven well-Tempered Bias-Exchange Metadynamics (CM-BexMetaD)

Biomolecular simulations are regularly used in molecular biology researches, however they have a limited impact on pharmaceutical and biotechnology industry with a high computational power and time for a short incidence of the life of biomolecules. Although MD is an important and widely used research methodology, inadequate sampling limits its application. The reason of this limitation is the roughness of the energy landscape of the biomolecular motion containing many local minima and high energy barriers separating these minima. Hence, to understand protein dynamics and reveal whole conformational space through constructing the energy landscape of biomolecules, several approaches have been developed in terms of both hardware and methodological

means. Methodological approaches characterized by enhanced sampling techniques and coarse grained models [188]. Both of these two approaches have some strengths and weaknesses. It will thus be effective to combine these two approaches, where hybrid simulation methodologies are shown as successful [3, 189].

The hybrid methodology proposed in this study combines an ENM, namely Anisotropic Network Model (ANM), with explicit solvent bias-exchange well-tempered metadynamics simulations. This approach is named here as collective modes driven bias exchange metadynamics (CM-BexMetaD). It uses intrinsic functional dynamics of proteins to enhance conformational sampling and to collect parallel pathways between metastable states by constructing a free energy landscape of the transition.. Here, the initial conformation is biased to move along the collective ANM modes of motion, taking advantage of the explicit solvent atomistic molecular dynamics between the of insertion of Gaussian hills of metadynamics, which provide escaping from trapping local minima. We employed bias exchange well-tempered metadynamics simulations, which randomly select the replicas and exchange their biases and accept or refuse the exchange of biases according to the Metropolis criteria. This stochastic exchange of biases provides transition between two states via minimum energy pathway and the system is not directly enforced to the target state [126, 190]. The bias-exchange scheme of the approach is summarized in Figure 6.3.

The bias-exchange well-tempered metadynamics phase of the methodology begins with N number of replicas, Rep1,..., RepN (N=5 in this Figure 6.3) each biased with a different CVs, CV1,...,CVN. After a user-defined time step of well-tempered metadynamics (wt-MetaD) simulation, randomly selected two replicas (Rep2 and Rep3 in this Figure 6.3) exchange their biases (CVs) and in order to accept or reject this exchange of biases, four wt-MetaD simulations are conducted, where two replicas with exchanged biases and two replicas with unchanged biases. This exchange will be accepted or rejected according to Metropolis criteria (Equation 6.27). From the beginning of the N replica wt-MetaD to the end of Metropolis decision is called *biasexchangeattempt* within this methodology, which is a user-defined variable within the simulation.

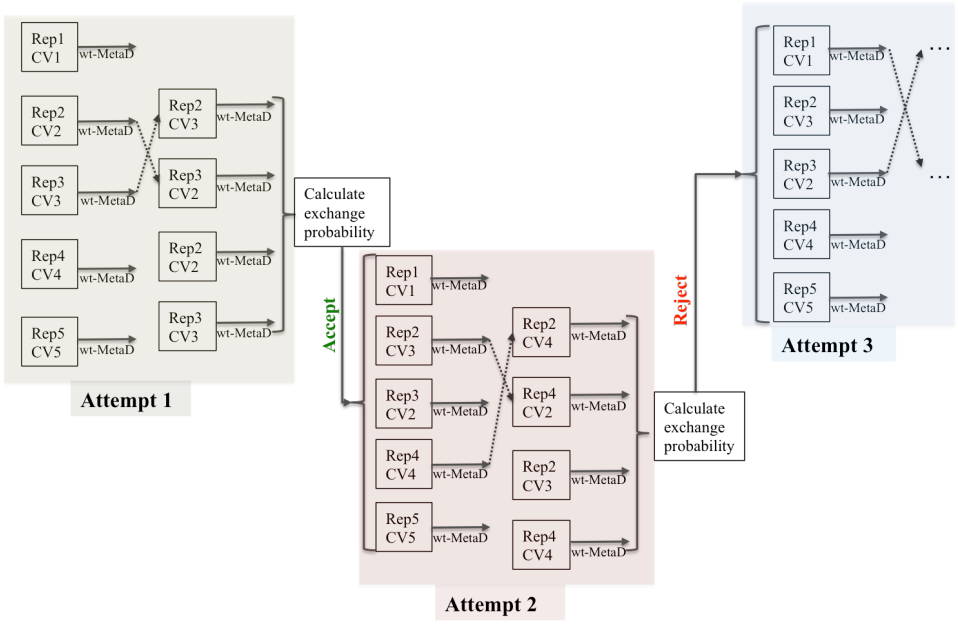


Figure 6.3. Flow Diagram of Bias-Exchange well-Tempered Metadynamics (wt-MetaD) Scheme. Here, a wt-MetaD Simulations with the Replicas (Rep1, ..., Rep5) Each Biased with a CV (CV1, ..., CV5) are Illustrated Schematically.

Since the methodology uses intrinsic properties of proteins, it can thus be applied to any system without a *priori* information about the protein specific, most valuable bias (collective variable), which will drive the simulation to any target state. The methodology is developed in three different ways with respect to how collective variables are utilized to bias metadynamics simulations; targeted (the highest overlapping slowest ANM modes with the initial to target difference vector are used), non-targeted (only a number of slowest ANM modes are used), and updating the collective modes used to bias the simulation at certain intervals in the targeted version of the methodology. The flowcharts of both approaches are given in Figure 6.4, Figure 6.5 and Figure 6.6.

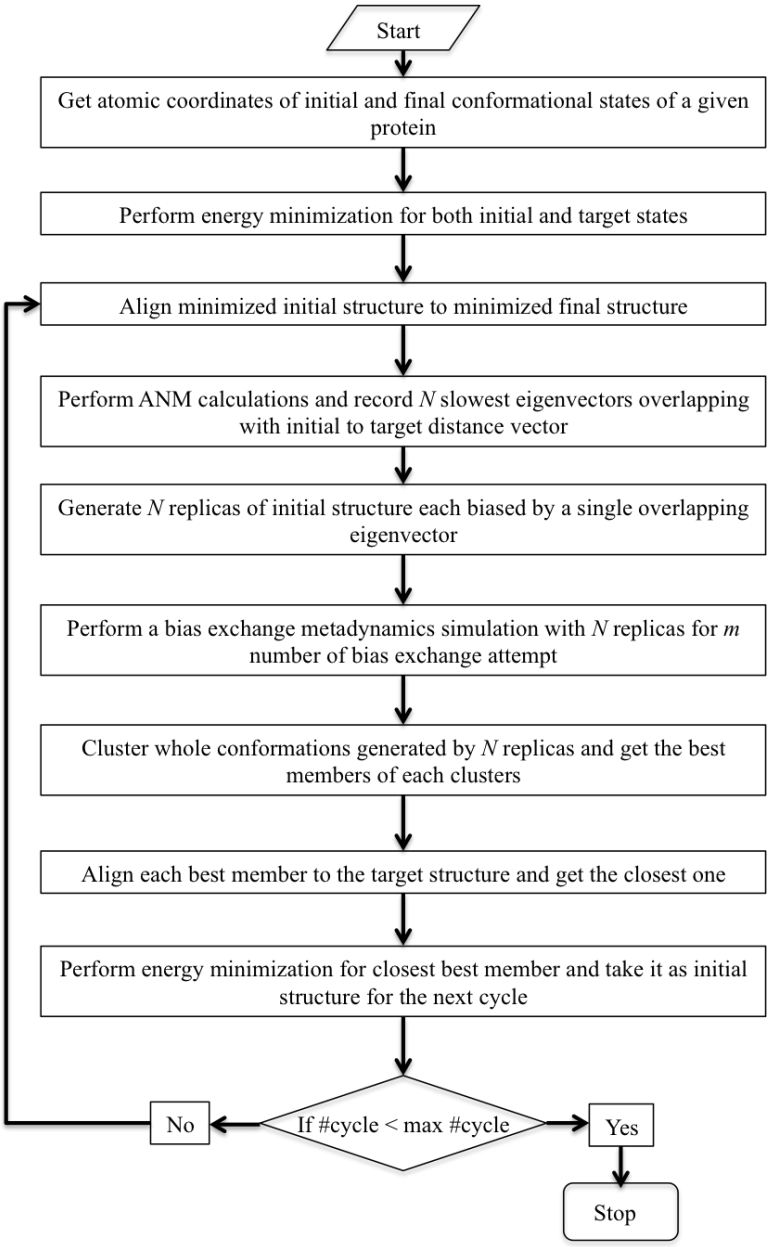


Figure 6.4. Flowchart of Targeted CM-BexMetaD-1 Approach.

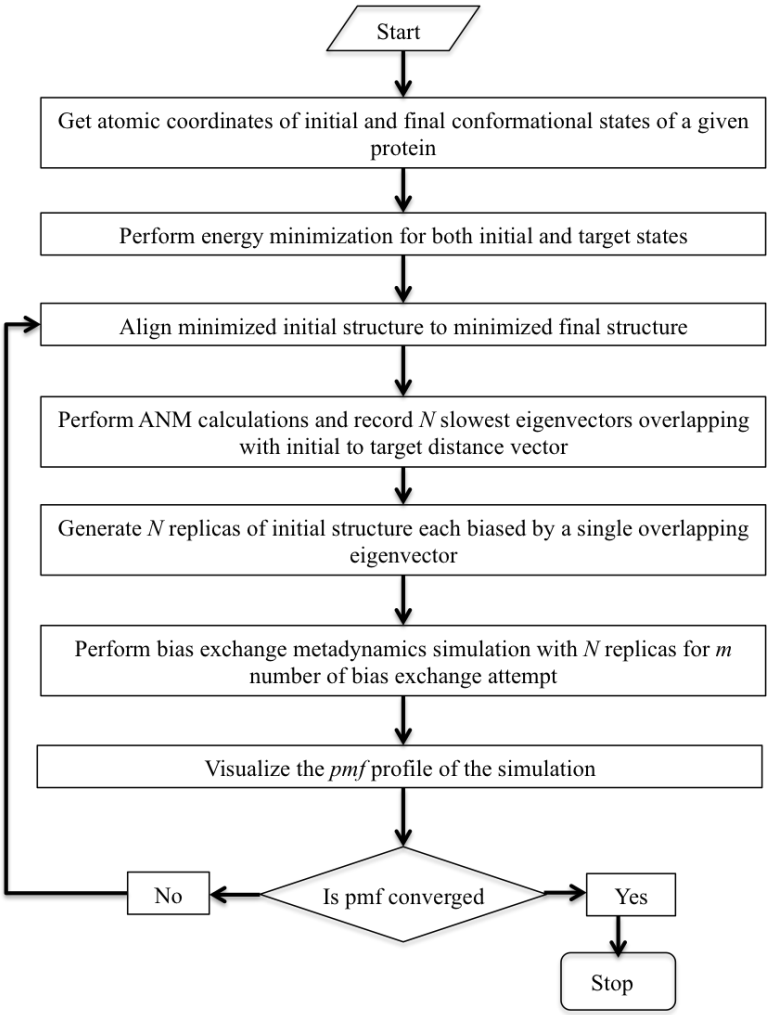


Figure 6.5. Flowchart of Targeted CM-BexMetaD-2 Approach.

The common (in both targeted and non-targeted versions) user-defined variables in this hybrid approach are the number of ANM modes (also equals to number of replicas), the cutoff radius (R_{cut}), Gaussian height w, the Gaussian widths σ_i, the hill deposition time intervalτ_G, the temperature "boost" ΔT, and the number of bias exchange attempt. RMSD value as similarity measure in clustering and the number of simulation/clustering cycle are additional user-defined variables for the CM-BexMetaD-1 approach.

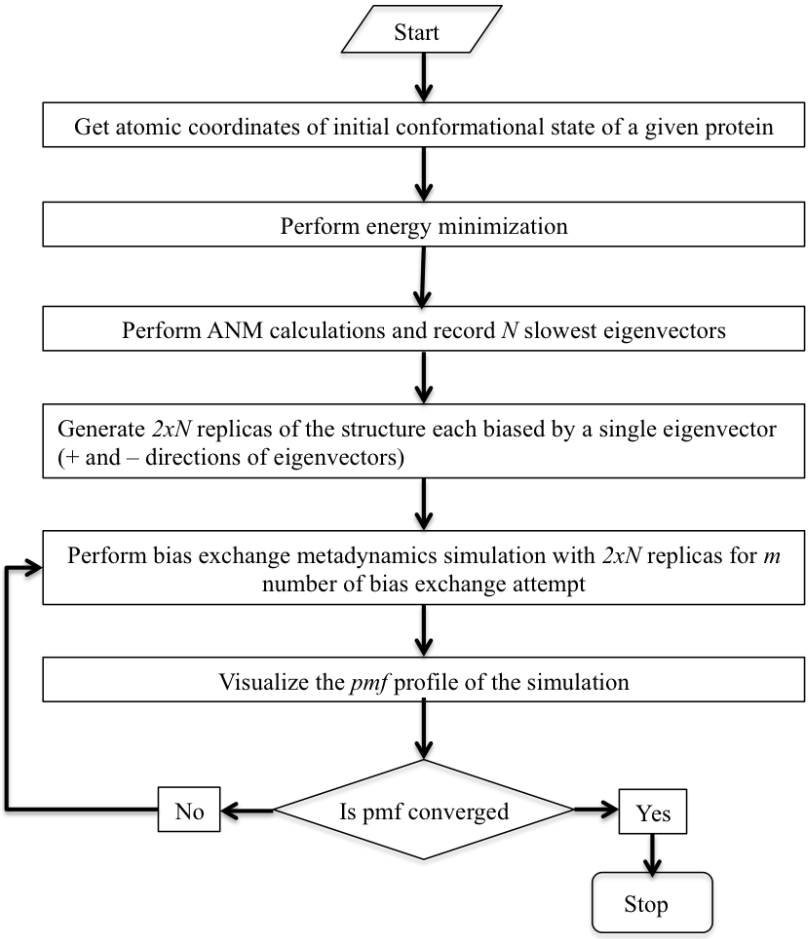


Figure 6.6. Flowchart of Non-Targeted CM-BexMetaD Approach.

6.5.1. Intrinsic Dynamic Modes as CVs in the Present Thesis

Eigenvector, which is practically the projection of the atomic coordinates onto a vector in R^{3N}, where N is the number of atoms, is a CV that is defined in ready-to-use molecular simulation tools. The projection of atomic coordinates r onto the specified vector is calculated using following form

$$\begin{aligned} & p(\{r_i(t)\}, \{r_i^{(ref)}\} \\ & = \left(\sum_{i=1}^{N} u_i^2\right)^{-1} \sum_{i=1}^{N} u_i . \left(U\left(r_i(t) - r_{cog}(t)\right)\right. \\ & \left. - \left(r_i^{(ref)} - r_{cog}^{(ref)}\right)\right) \end{aligned} \tag{6.31}$$

Here, U is optimal rotation matrix, $U^{\{r_i(t)\}\rightarrow\left\{r_i^{(ref)}\right\}}$ that best superimpose the coordinates $\{r_i\,(t)\}$, which is in our case transient configuration of protein, onto a set of reference coordinates $\left\{r_i^{(ref)}\right\}$, which is in our case the initial conformation. $r_{cog}\,(t)$ and $r_{cog}{}^{(ref)}$ are the centers of geometry of the current and the reference positions, respectively. u_i are the components of the vector for each atom, which is in our case the eigenvector corresponding to the slowest normal modes determined from ANM. This enhanced direction u_i can be eigenvectors of the covariance matrix (essential mode) (principal components of the motion which could be obtained from a relatively short MD simulations).

The bias potential expressed in Equation 6.20 can be easily calculated by replacing the ζ with the projection p

$$V_B\,(p,t) = \sum_{\substack{t' = k\tau_G \\ k \in N}}^{t'<t} w \prod_{i=1}^{N_{CV}} e^{-\frac{[p_i - p_i(t')]^2}{2\sigma_i^2}} \tag{6.32}$$

Essential coordinates derived from the PCA of a 500ns conventional MD simulation was used by Deriu and coworkers [191] as CVs to explore the conformational rearrangement of Josephin Domain and to construct the underlying free energy landscape characterizing the transition pathway from the open to closed conformation of the protein. However, the collective modes of the proteins obtained from ENM, which give the long time behavior of conformational transition have not been used as CVs to enhance the sampling in metadynamics yet. Calculating the collective modes of motion from ENMs reduces the required computational time to extract the essential coordinates.

In this proposed hybrid methodology, Anisotropic Network Model (ANM), which is a coarse grained ENM is used to extract collective modes of motion. ANM is carried

out for the initial conformational state of the molecule by decomposing the Hessian matrix as explained in Methods and Materials Section. For the targeted version of the proposed approach, the determination of collective modes and the corresponding 3N-dimensional eigenvectors u of motion is followed by the calculation of the overlap (O) between those eigenvectors and the initial to target difference vector (d) by simply taking the dot product of the vectors as;

$$O_i = \frac{|u_i.d|}{|u_i|\,.\,|d|}; i = 1:k \tag{6.33}$$

where, k is the maximum number of slowest ANM modes that is a user-defined variable. Before calculating the difference vector, initial and target structures are aligned to remove the irrelevant translation and rotation between the structures. N number of highest overlapping ANM modes are used as CVs, which is again a user-defined number.

For non-targeted version of the proposed approach, a pre-defined number of slowest ANM modes are used as CVs to derive the enhanced sampling along the highest collective modes of motion.

6.5.2. Computational Details of the Simulation

NAMD 2.10 [192] with CHARMM36 [194] force field was used for both minimization of structures and the well-tempered metadynamics simulation with 2 fs integration time step and periodic boundary conditions. The temperature was maintained at 310 K with a Langevin damping coefficient of 1 ps^{-1} [195]. The pressure was kept at 101.3 kPa by means of NoseÂ´-Hoover Langevin piston pressure control. The SHAKE algorithm was used to restrain the length of bonds involving hydrogen atoms for a time step of 2 fs. The initial crystal structures were immersed in a TIP3P-type water box [196] with at least $8\tilde{A}$ of padding between the solute and the edge of the box. The system was neutralized with Cl^- and Na^+ ions. A nonbonded cutoff of $12A\hat{A}^\circ$ was used for all Len- nard-Jones interactions with a switching function starting at $10\tilde{A}$, and the long-range electrostatics was treated according to the particle-mesh Ewald method [197]. The nonbonded pair list distance was $14\tilde{A}$. All systems were energetically minimized

by the conjugated gradient method for steric crush and crystal contact removal.

Amber Tools 14 package was used for both aligning the structures/trajectories and clustering the trajectories.

In-house developed script in MATLAB was used to accomplish bias-exchange procedure and ANM calculations.

7. CONFORMATIONAL SAMPLING AND CONSTRUCTION OF FREE ENERGY LANDSCAPE USING CM-BEXMETAD

7.1. Introduction

Biomolecular simulations are regularly used in molecular biology researches; however they have a limited impact pharmaceutical and biotechnology industry because of the fact that in order to simulate a short incidence from the life of biomolecules they require a high computational power and time. Although molecular dynamics (MD) simulation is an important and widely used research methodology, inadequate sampling limits its application. The reason of this limitation is the roughness of energy landscapes of biomolecular motion containing many local minima and high energy barriers separating these minima. Hence, to understand protein dynamics and reveal whole conformational space through constructing full energy landscape of biomolecules, several approaches have been developed in terms of both hardware and methodological means. Methodological approaches consists of coarse grained models and enhanced sampling techniques [188]. Both of these two approaches have some strengths and weaknesses; therefore it will be effective to combine the two approaches where hybrid simulation methodologies are shown as successful in recent literature [3, 189].

The methodology proposed in this study combines Anisotropic Network Model (ANM) with explicit solvent bias exchange metadynamics simulations called collective modes driven bias exchange metadynamics (CM-BexMetaD) to enhance conformational sampling using intrinsic dynamics of proteins and to collect parallel pathways between functional states by constructing the underlying free energy landscape of the transition between these states. The simulation protocol is designed not to sample only the shortest paths between stable states of the protein, instead some energetically unfavorable transitions which are not easy to sample via conventional molecular simulation techniques are biased to favor overcoming large energy barriers, as well as giving

flexibility to sample a short equilibrium dynamics between addition of bias potentials.

The developed CM-BexMetaD simulations were performed for the well-characterized exemplary case adenylate kinase (AdK), which has known open and closed conformations as well as many intermediate states determined by experimental methods. The energetics and the structural determinants underlying its large-scale conformational transition are also studied with many enhanced sampling methodologies, which provides comparative information to the results obtained with the present novel approach.

Using the developed protocol, it has been possible to sample some of the experimentally determined structures as intermediate states in the functional transition of the AdK enzyme. The one-dimensional free-energy landscape along the stochastic combination of collective modes of motion was also calculated. In addition, the effect of a perturbation on the conformational ensemble and energetics through mutating a hinge site that plausibly controls the functional transition of the enzyme was also investigated.

7.2. Materials and Methods

7.2.1. CM-BexMetaD Simulations

We have performed three different simulation protocols (CM-BexMetaD-1, CM-BexMetaD-2 and non-targeted CM-BexMetaD) between functionally stable conformation of AdK, taking the open conformation as initial structure to study open to closed conformation of the protein. The details of each simulation protocols were given in section X of this thesis and the summary of parameters used are listed in the following tables.

7.2.2. Targeted with Slowest Ten Modes

This is the first version developed as CM-BexMetaD-1. In this version, only slowest ten modes were used as CVs, but the directions of the eigenvectors were updated

according to the direction of the difference vector between the initial and target structure. The user-defined parameters and simulation lengths are summarized in Table 7.1.

Table 7.1. Summary of Simulation Parameters of CM-BexMetaD-1.

Gaussian Hill Height (kcal/mol)	0.5
Rcut for ANM calculations ($\tilde{A}$)	18
Hill Addition Frequency (ps)	1
MetaD length before Bias exchange attempt (ps)	10
# Bias exchange attempt before clustering	100
# of Clustering Cycle	8
Total Simulation Length (ns)	8

7.2.2.1. Targeted with Highest Overlapping Slowest Modes. In the second version of the methodology, the highest overlapping collective modes were used as CVs and the directions of the eigenvectors were updated according to the direction of the difference vector between the initial and target structure. The user-defined parameters and simulation lengths are summarized in Table 7.2.

Table 7.2. Summary of Simulation Parameters CM-BexMetaD-1.

Gaussian Hill Height (kcal/mol)	0.5
Rcut for ANM calculations ($\tilde{A}$)	18
Overlap cutoff	0.5
Number of ANM modes used as CVs	10
Hill Addition Frequency (ps)	1
MetaD length before Bias exchange attempt (ps)	10
# Bias exchange attempt before clustering	200
# of Clustering Cycle	8
Total Simulation Length (ns)	16

7.2.2.2. Targeted with Highest Overlapping Slowest Modes without Updating Modes. In this version of the developed method, the highest overlapping collective modes were used as CVs and the directions of the eigenvectors were updated according to the direction of the difference vector between initial and target structure. The user-defined parameters and simulation lengths are summarized in Table 7.3 and Table 7.4.

Table 7.3. Summary of Simulation Parameters of CM-BexMetaD-2 for wt-run1 and Y171W Mutant.

	wt-1	Y171W
Gaussian Hill Height (kcal/mol)	2	2
Rcut for ANM calculations ($\tilde{A}$)	18	18
Overlap cutoff	0.5	0.5
Number of ANM modes used as CVs	10	10
Hill Addition Frequency (ps)	0.2	0.2
MetaD length before Bias exchange attempt (ps)	10	10
# Bias exchange attempt	560	560
Total Simulation Length (ns)	5.6	5.6

Table 7.4. Summary of Simulation Parameters of CM - BexMetaD - 2 for wt - run2.

Gaussian Hill Height (kcal/mol)	10
Rcut for ANM calculations ($\tilde{A}$)	18
Number of ANM modes used as CVs	10
Hill Addition Frequency (ps)	0.2
MetaD length before Bias exchange attempt (ps)	10
# Bias exchange attempt	330
Total Simulation Length (ns)	3.3

7.2.2.3. Non-Targeted with Slowest Modes without Updating Modes. In this version of the developed method, only the slowest five collective modes were taken and the eigenvectors with both directions of each mode were also generated (simply multiplying with -1). Thus, ten modes were used as CVs. The user-defined parameters and

simulation lengths are summarized in Table 7.5.

Table 7.5. Summary of Simulation Parameters for Non-Targeted Approach.

Gaussian Hill Height (kcal/mol)	2
R_{cut} for ANM calculations ($\tilde{A}$)	18
Number of ANM modes used as CVs	10
Number of ANM modes used as CVs	5x2
MetaD length before Bias exchange attempt (ps)	10
# Bias exchange attempt	460
Total Simulation Length (ns)	4.6

7.2.2.4. Simulation Details. Metadynamics simulations were performed in explicit solvent using CHARMM36 [194, 198] force field, TIP3P [196] water model and periodic boundary conditions in NAMD 2.10 [192]. The temperature was maintained at 310 K with a Langevin damping coefficient of 1 ps^{-1} [195]. The pressure was kept at 101.3 kPa by means of NosÃ©-Hoover Langevin piston pressure control. The SHAKE algorithm was used to restrain the length of bonds involving hydrogen atoms for a time step of 2 fs (Ryckaert *et al.*, 1977). The system was neutralized with Cl- and Na+ ions. A nonbonded cutoff of 12$\tilde{A}$ was used for all Lennard-Jones interactions with a switching function starting at 10$\tilde{A}$, and the long-range electrostatics was treated according to the particle-mesh Ewald method [197]. The nonbonded pair list distance was 14$\tilde{A}$. Gaussian potentials are deposited every 2 picoseconds and gradually decreased on the basis of the adaptive bias with 1500K.

7.3. Results and Discussion

7.3.1. Test Case: Adenylate Kinase (AdK)

Adenylate kinase is an enzyme coordinating different signaling pathways connected to several human diseases, such as heart failure, metabolic disorders, cancer and many neurodegenerative diseases [199, 200]. AdK enzyme regulates the ratio of

AMP and ADP/ATP and therefore controls the cellular energy supply [201]. In order to carry out its function, AdK requires a structural reorganization and collective domain motion on the $\mu s - ms$ timescale [202]. AdK consist of 214 residues and has three domains, called CORE, LID and NMP and two distinct nucleotide binding sites (Figure 7.1-A). In this figure, LID and NMP domains are colored in blue (residues 118-160) and red (residues 30-67), respectively. The CORE domain of the protein is colored in light gray (residues 1-29, 68-117, 161-214) B) Eigenvectors of the slowest two modes obtained from ANM calculations for fully-open structure are illustrated on the fully-open conformation of the AdK. ATP binding site is between the CORE and LID domains and AMP binding site is between CORE and NMP domains. LID and NMP domains achieves a large conformational transition between open and closed states where CORE domain is conformationally stable [203, 204]. Full or partial closing of LID and NMP domains upon binding of the nucleotides have been identified via experimental studies [205-207].

In addition to experimental studies, the structural and dynamic basis of the functional conformational transition of AdK was also studied via computational studies via classical MD or other atomistic simulations [204, 208-211]. However, these simulations are in limited timescales, where domain movements are rare event. Several $\mu s - ms$ range enhanced sampling studies in atomistic MD simulations have been also conducted for both apo state and bound to several ligands and inhibitors [95- 132], [212-216]. Through these experimental and computational studies it has been investigated that, there is a dynamic equilibrium between open and closed conformational states of the apo AdK. Several mutational analysis that affects the dynamics equilibrium between these stable states have been demonstrated [210, 217, 218]. Furthermore, kinetic experiments suggested that, the conformational change of the enzyme is the rate-limiting step of the catalytic reaction, thus its motion has functional role [202, 205, 219]. Although the high global flexibility of the protein increases the possibility to sample closed-like conformations, the existence of a significant free-energy barrier for accessing fully-closed conformation in which the fully-open conformation is the energetically most favorable state have been reported via enhanced sampling simulations [95, 216]. Besides, several computational studies revealed the sequence of events in both opening

and closing pathway of the enzyme; such that NMP domain opens first followed by opening of LID domain [95, 216] in the holo form and closing of LID domain precedes closing of NMP domain in apo form [208, 220].

Y171 is in the hinge site vicinity of AdK that allosterically regulates the conformational transition of the enzyme 208, 218]. The altered dynamics of the enzyme has been investigated via Y171W mutation in a recent NMR spectroscopy experiment, and both open and closed states of the mutated enzyme was determined by X-ray crystallography [218]. The kinetic stopped-flow experimental data answered the long-time standing question, whether the enzyme obeys induced-fit or conformational selection models of allosteric regulation. The protein obeys both of these models depending on the existence of substrate. Since the protein is able to sample both open and closed conformations even in the absence of substrate, the domain closure of the enzyme obeys conformational selection model. However, in the presence of the substrate, since the domain closure is achieved following the initial nucleotide binding, induced fit model is more appropriate to explain conformational transition of the enzyme [216, 218].

Here we used RMSD values and LID-CORE angle (the angle between the pseudo axes connecting the centers of geometry of the backbone and CÃ atoms of residues 179-185 (CORE), 115-125 (CORE-LID), 125-153 (LID)) and NMP-CORE angle (calculated similarly using residues 115-125 (CORE-LID), 90-100 (CORE) and 35-55 (NMP)) defined in Beckstein *et al.*, to represent to conformational transition of AdK [221].

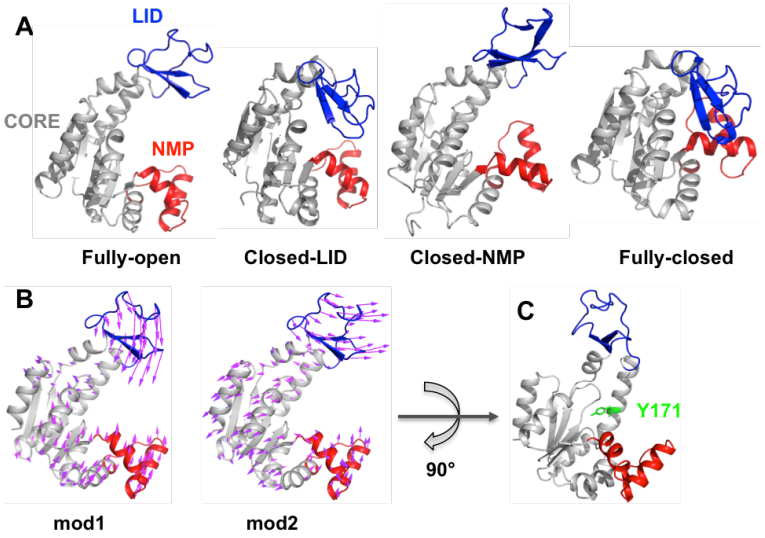


Figure 7.1. Structures and Slowest ANM Modes of AdK. a) Fully-open (PDB ID: 4AKE), Fully-closed (PDB ID: 1AKE), Closed-LID (PDB ID: 1DVR) and Closed NMP (PDB ID: 2AK3) Structures of AdK. c) Position of Y171 is Labeled on the Fully-open Conformation.

7.3.2. Investigation Conformational Transition of AdK via CM BexMetaD

First, CM-BexMetaD-1 approach is applied to study the conformational transition of the enzyme. In this methodology, only slowest ten global modes were used to bias the transition and the generated conformers are clustered after a short bias-exchange attempt period to obtain a representative closest conformation to the fully-closed state and the collective modes were recalculated from this regenerated conformer (see Figure 6.4). The details of the simulation parameters are given in Table 7.1. As an illustration, the RMSD values of each replica for the cycle 1 of the simulation are given in Figure 7.2. Since the simulation protocol is designed not to sample only the shortest paths between stable states of the protein, the RMSD values fluctuates during the simulation windows. The minimum RMSD values of each replica at each cycle form the target conformation are listed in Table 7.6. Through this methodology, we were able to approach the fully-closed conformation of the protein with a minimum RMSD values around $3\tilde{A}$ (no outlier rejection during alignment) for all replicas within very short time window. Continuing the simulation after 5th cycle seems not to achieve

further closing transition of the conformation. Stucking around this RMSD indicates the existence of a high-energy barrier that the protein must overcome separating the fully-closed conformation from other transition conformations.

In the second step, the same simulation protocol except using not only the slowest ten modes of motion but highest overlapping ten modes of motion with the open-closed difference vector were used as CVs. The parameters used in this simulation protocol are given in Table 7.2. Here, the number of bias-exchange attempt before clustering the trajectory is increased to ensure better sampling of the ensemble. As an illustration, the RMSD values of each replica for the cycle 1 of the simulation are given in Figure 7.3. The minimum RMSD values of each replica at each cycle from the target conformation are listed in Table 7.7. Through this methodology, we were able to approach the fully-closed conformation of the protein with a minimum RMSD values less than $3\tilde{A}$ (minimum at $2.5\tilde{A}$ again with no outlier rejection during alignment) at cycle 8 for all replicas within very short time window. Continuing the simulation after 8^{th} cycle might decrease the RMSD but it seems not to achieve further significant closing to the target conformation because RMSD profile of the replicas seems to reach a plateau. The high-energy barrier separating the fully-closed conformation from other transition conformations prevents further closing of the structure and clustering the generated conformers draws back the initial conformation of the simulation at each cycle to the nearest before sampled minima.

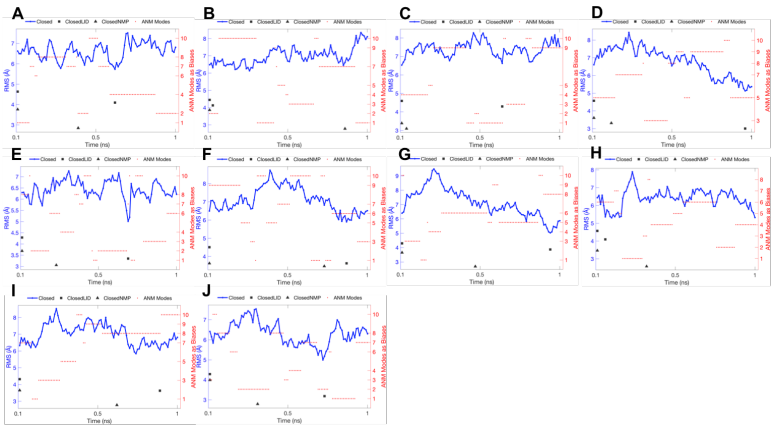


Figure 7.2. RMSD and the Biased CV Versus Simulation time of Each Replica at Cycle 1 by CM-BexMetaD-1. Replicas 1 to 10 are given in A to J, Respectively.

Table 7.6. Minimum RMSD Table of Each Replicas at Each Cycles to Fully blosed, Closed LID and Closed NMP. Closest RMSD Values of Each Replica at Each Cycle are written in bold and the Lowest RMSD is colored in red.

Cycle	Cases	Replicas									
		1	2	3	4	5	6	7	8	9	10
1	Fully Closed	5.73	6.14	6.53	5.13	5.01	5.9	5.03	5.3	5.84	4.99
	Closed LID	4.09	4.13	4.29	3	3.35	3.62	3.87	4.09	3.63	3.18
	Closed NMP	2.85	2.77	3.1	3.31	3.05	3.46	2.68	2.58	2.77	2.79
2	Fully Closed	5.1	5.38	4.95	4.91	5.01	5.28	5.3	5.29	4.5	5.2
	Closed LID	4.02	3.81	3.66	4	3.92	4.29	4.17	4.14	3.34	4.98
	Closed NMP	2.65	2.82	2.64	1.97	2.59	2.82	2.71	2.75	2.55	2.48
3	Fully Closed	5.26	3.57	4.45	4.69	4.5	4.61	4.4	4.44	4.02	4.99
	Closed LID	3.76	1.97	3.28	3.33	2.97	3.16	3.11	2.96	3.1	3.85
	Closed NMP	2.56	3.18	2.56	2.88	2.53	2.53	2.86	2.82	2.83	1.92
4	Fully Closed	3.82	3.76	3.76	3.65	3.18	3.27	3.77	3.24	3.65	3.62
	Closed LID	2	2.17	1.85	1.87	1.87	2.2	2.08	2.03	1.93	2.26
	Closed NMP	2.97	3.47	3.29	4.09	3.31	3.89	3.42	4.08	3.9	3.56
5	Fully Closed	3.81	3.3	3.9	3.37	3.68	3.54	3.12	3.33	3.92	3.57
	Closed LID	2.19	1.85	2.23	1.89	1.76	2.11	1.9	1.77	2.27	2.11
	Closed NMP	3.77	4.22	3.71	3.76	3.47	2.67	4.27	4.51	3.28	4.12
6	Fully Closed	3.7	4.03	3.95	3.58	4.05	3.64	3.3	3.81	3.33	3.29
	Closed LID	2.24	2.58	2.06	2.18	2.35	1.99	2.15	2.36	1.98	2.03
	Closed NMP	3.16	3.71	4.08	3.95	3.34	4.42	3.67	3.82	3.98	3.99
7	Fully Closed	3.69	3.46	4.06	4.03	4.33	3.95	3.51	3.93	3.85	4.13
	Closed LID	1.98	2.27	2.41	2.34	2.47	2.65	1.99	2.24	2.54	2.48
	Closed NMP	3.98	4.16	3.89	4.11	3.4	3.59	2.9	3.48	3.28	3.13
8	Fully Closed	3.76	3.46	3.98	3.86	4.1	4.14	4.19	3.8	3.99	3.53
	Closed LID	2.25	2.15	2.4	2.01	2.52	2.36	2.78	2.3	2.48	1.87
	Closed NMP	4.35	4.26	3.12	3.25	3.57	3.38	2.99	3.81	4.24	4.17

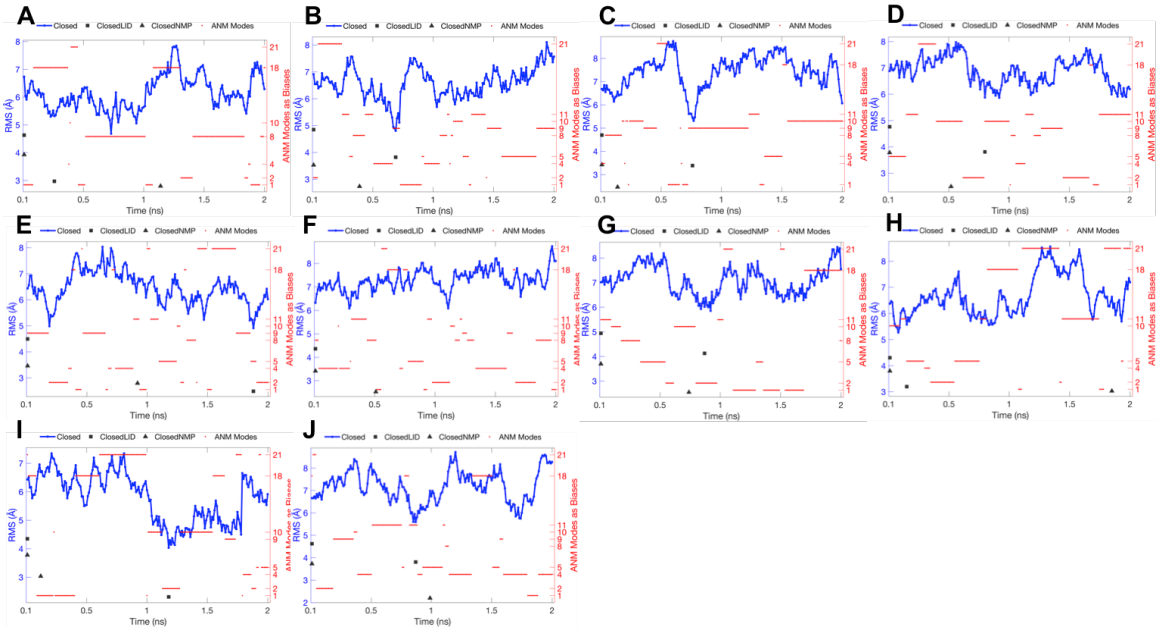


Figure 7.3. RMSD and the Biased CV Versus Simulation time of Each Replica at Cycle 1. Replicas 1 to 10 are given in A to J, respectively.

Table 7.7. Minimum RMSD Table of Each Replicas at Each Cycles to Fully Closed, Closed LID and Closed NMP. Closest RMSD values of Each Replica at Each Cycle are written in Bold and the Lowest RMSD is colored in red.

Cycle	Cases	Replicas									
		1	2	3	4	5	6	7	8	9	10
1	Fully Closed	4.68	4.8	5.31	5.86	4.91	6.08	5.85	5.27	4.04	5.59
	Closed LID	2.97	3.82	3.38	3.81	2.48	4.36	4.12	3.2	2.31	3.81
	Closed NMP	2.8	2.73	2.47	2.49	2.79	2.52	2.54	3.03	3.03	2.2
2	Fully Closed	4.01	4.09	4.23	3.86	3.83	4.48	4.56	3.62	4.4	3.83
	Closed LID	2.2	2.52	2.45	2.12	3.01	3.07	2.91	2.36	2.64	2.72
	Closed NMP	4.62	3.58	3.2	4.97	3.13	3.27	3.29	4.07	3.14	3.21
3	Fully Closed	3.56	3.31	3.76	3.43	3.63	3.59	3.6	3.15	3.05	3.17
	Closed LID	2.2	1.95	2.28	2.03	2.09	2.23	2.32	2.08	2.18	1.99
	Closed NMP	4.23	3.22	4.57	4.5	3.74	3.85	4.06	4.22	3.91	5.12
4	Fully Closed	3.4	3.15	3.3	3.01	3.16	3.18	3.97	3	3.15	3.24
	Closed LID	2.13	2.26	2.15	2.43	2.45	2.13	2.54	2.58	2.09	1.96
	Closed NMP	3.51	4.75	4.73	4.14	4.25	4.08	3.93	3.54	4.05	4.32
5	Fully Closed	2.89	3.11	3.06	3.01	3.01	3.17	3.24	3.42	2.94	2.93
	Closed LID	2.26	1.76	2.3	1.94	2.28	2.08	2.3	2.58	1.86	2.34
	Closed NMP	4.53	4.06	4.67	4.35	3.87	6.15	3.85	3.76	4.48	4.82
6	Fully Closed	3.15	3.12	3.2	3.2	3.04	3.13	3.05	2.79	3.15	3.23
	Closed LID	1.86	2.22	1.95	2.01	1.8	1.97	1.88	2.36	1.83	2.01
	Closed NMP	4.31	4.97	6.01	4.94	4.57	4.73	4.59	4.51	4.56	5.28
7	Fully Closed	3.05	4.43	2.79	2.75	2.93	3.02	2.59	2.87	2.76	2.62
	Closed LID	2.52	3.44	2.64	2.55	2.69	2.66	2.24	2.45	2.7	2.76
	Closed NMP	3.24	2.71	3.02	2.77	4.01	2.48	3.39	3.46	2.79	2.95
8	Fully Closed	2.96	3.1	4.3	2.9	2.81	2.94	3.04	2.86	2.88	2.49
	Closed LID	2.77	2.84	2.68	2.56	2.61	2.03	2.72	2.8	2.33	2.9
	Closed NMP	2.31	2.85	2.53	2.77	3.33	3.76	2.6	2.64	2.61	3.81

The second version of the developed approach is CM-BexMetaD-2 where it differs from CM-BexMetaD-1 as the trajectories generated at each replica do not clustered to obtain a metastable state and continue the simulation with updated ANM modes. Instead, the number of bias-exchange attempt kept high during the simulations. Three different simulations were carried out (two wild type AdK; one with a lower Gaussian hill height (wild type-1), and one with higher Gaussian hill height (wild type-2), one Y171W AdK with same lower Gaussian hill height) with this methodology. The parameters used in the simulations are listed in Table 7.4 and Table 7.4. RMSD values over the trajectory of each replica during the simulations are given in Figure 7.4 and Figure 7.5 for wild type-1 and Y171W, respectively and in Figure 7.6 for wild type-2. The minimum RMSD values of each replica from the target conformation are listed in Table 7.8 and Table 7.9 for lower and higher Gaussian hill height, respectively. Through this methodology, we were able to approach the fully-closed conformation of the protein with a minimum RMSD values $3.3\tilde{A}$ and 2.9 $\tilde{A}$ at wild type-1 and Y171W simulations, respectively within very short time window, where a lower Gaussian hill height (2kcal/mol) is used.

Table 7.8. Minimum RMSD Table of Each Replicas from the Fully Closed, Closed LID and Closed NMP of wild Type-1 and Y171W Simulations. Closest RMSD Values are colored in red.

Cycle	Cases	Replicas									
		1	2	3	4	5	6	7	8	9	10
wt-1	Fully Closed	3.29	4.4	3.79	4.85	4.3	5.05	4.1	3.45	4.81	4.15
	Closed LID	2	3.02	2.35	2.45	3.39	3.62	3.4	1.92	3.2	2.65
	Closed NMP	2.72	2.7	2.76	2.83	2.57	3	2.51	3.25	2.93	2.95
Y171W	Fully Closed	3.21	3.2	6.27	4.97	4.12	3.38	5.1	4.65	3.62	2.92
	Closed LID	2.1	1.95	4.39	2.88	2.26	2.09	2.75	2.92	2.09	1.89
	Closed NMP	2.85	4.05	2.64	3.01	3.1	3.36	3.08	3.17	3	2.81

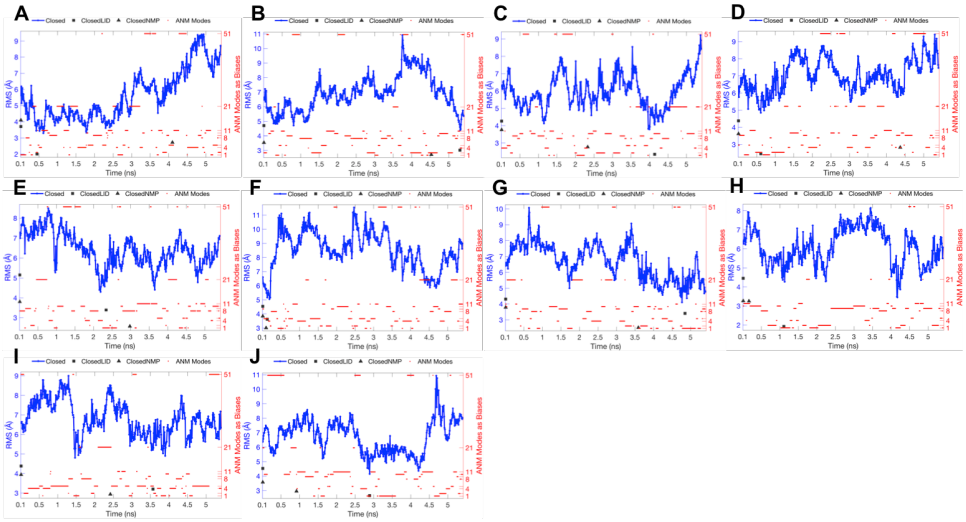


Figure 7.4. RMSD and the Biased CV versus Simulation Time of Each Replica for wild Type-1. Replicas 1 to 10 are given in A-J, respectively.

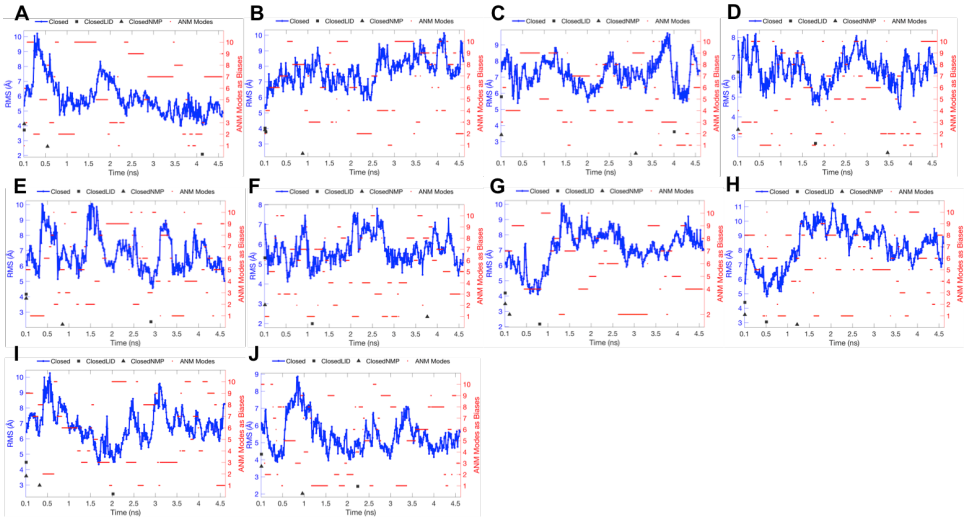


Figure 7.5. RMSD and the Biased CV versus Simulation Time of Each Replica for Y171W. Replicas 1 to 10 are given in A-J, respectively.

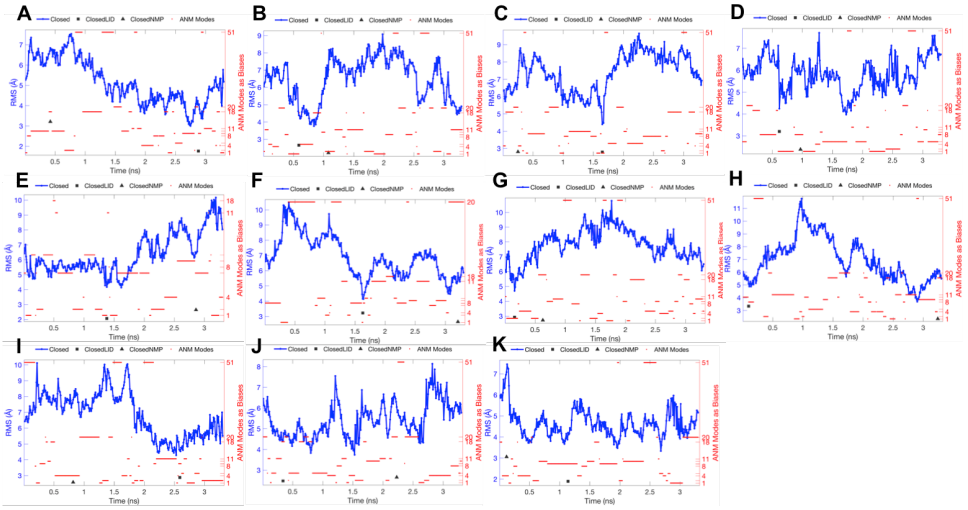


Figure 7.6. RMSD and the Biased CV versus Simulation Time of Each Replica for wild Type-2. Replicas 1 to 11 are given in A-K, respectively.

Since the minimum RMSD obtained is more than 3 $\tilde{A}$ from the fully-closed conformation, meaning that the high energy barrier separating the fully-closed ensemble basin from other minimum energy states is not achieved yet, another simulations was carried out with a higher initial Gaussian hill height. Even at a shorter simulation time, the minimum RMSD value to the fully-closed and closed-LID conformations decreased to 3 $\tilde{A}$ and 1.8 $\tilde{A}\tilde{A}$, respectively.

Table 7.9. Minimum RMSD Table of Each Replica from the Fully Closed, Closed LID and Closed NMP of wild Type-2 Simulation. Closest RMSD Value is colored in red.

Cases	Replicas										
	1	2	3	4	5	6	7	8	9	10	11
Fully Closed	3	3.77	4.4	3.98	4.13	4.13	4.7	3.69	4.29	3.74	3.34
Closed LID	1.76	2.66	2.79	3.2	2.07	3.23	2.93	3.34	2.88	2.47	1.9
Closed NMP	3.21	2.21	2.81	2.38	2.65	2.65	2.73	2.35	2.58	2.65	3.06

7.3.2.1. Domain Closures are Asymmetric Processes in apo AdK. In addition to RMSD values, the LID-CORE and NMP-CORE angles were calculated to better capturing the transition behavior explored via CM-BexMetaD simulations. The trajectory of the an-

gles of interest calculated over the wild type-1 simulation with the are given in Figure 7.7, the Y171W simulation in Figure 7.8, and the wild type-2 simulation are given in Figure 7.9. The values of each angles calculated for the experimentally determined fully-open, fully-closed and intermediate transition conformations are also labeled on the figures [221]. Although none of the simulations sampled the fully-closed conformations yet, some of the intermediate conformations corresponding to partial closing of LID and/or NMP are sampled during the simulations. Two very recent enhanced sampling approaches revealed that, the fully-open conformation is the global minimum of the apo AdK [95, 216]. The results obtained in our simulation within the limited time intervals, supports this idea as replicas spends longer time in the vicinity of the fully-open conformation.

The generated conformers throughout the simulations cover a broader range of calculated angles even in a few ns time-scales, which indicates the global conformational flexibility of the enzyme. This high flexibility increases the possibility of domain closures and observation of close-like conformations even in the absence of the nucleotides. Although none of the simulations converged to the fully-closed conformation within sampling time, closure of individual domains (closed LID and closed NMP) as well as some partially closed intermediate conformations are observed within the simulations. In the wild type-1 simulations, none of the replicas converged to the closed-LID conformation, but replica 5 converged to and replica 7 and 8 partially sampled to the closed-NMP conformations (see Figure 7.7). In the Y171W simulation, the conformational ensemble covers only the closed LID conformations in replicas 1, 2, 5, 6, 7, 9, and 10, but no NMP closure is observed. Increasing the sampling capacity of the simulation by employing higher Gaussian hills in the wild type-2 simulation, closed-LID conformation is sampled in replica 1, 2, 5, 9, 10 and 11 (see Figure 7.9). Closed NMP conformation was also sampled in this simulation in replicas 2, 6, 8, and 10, but it is not populated. Although the transition states in the vicinity of fully-open conformation are successfully sampled, no cooperative closure (closure of both domains simultaneously as time evolves) of the domains are observed (see the evolution of data points from blue to red). These results indicates that, the fully closure of the domains is not a cooperative event, it is rather an asymmetric process (no cooperativity in the

closure motion of domains).

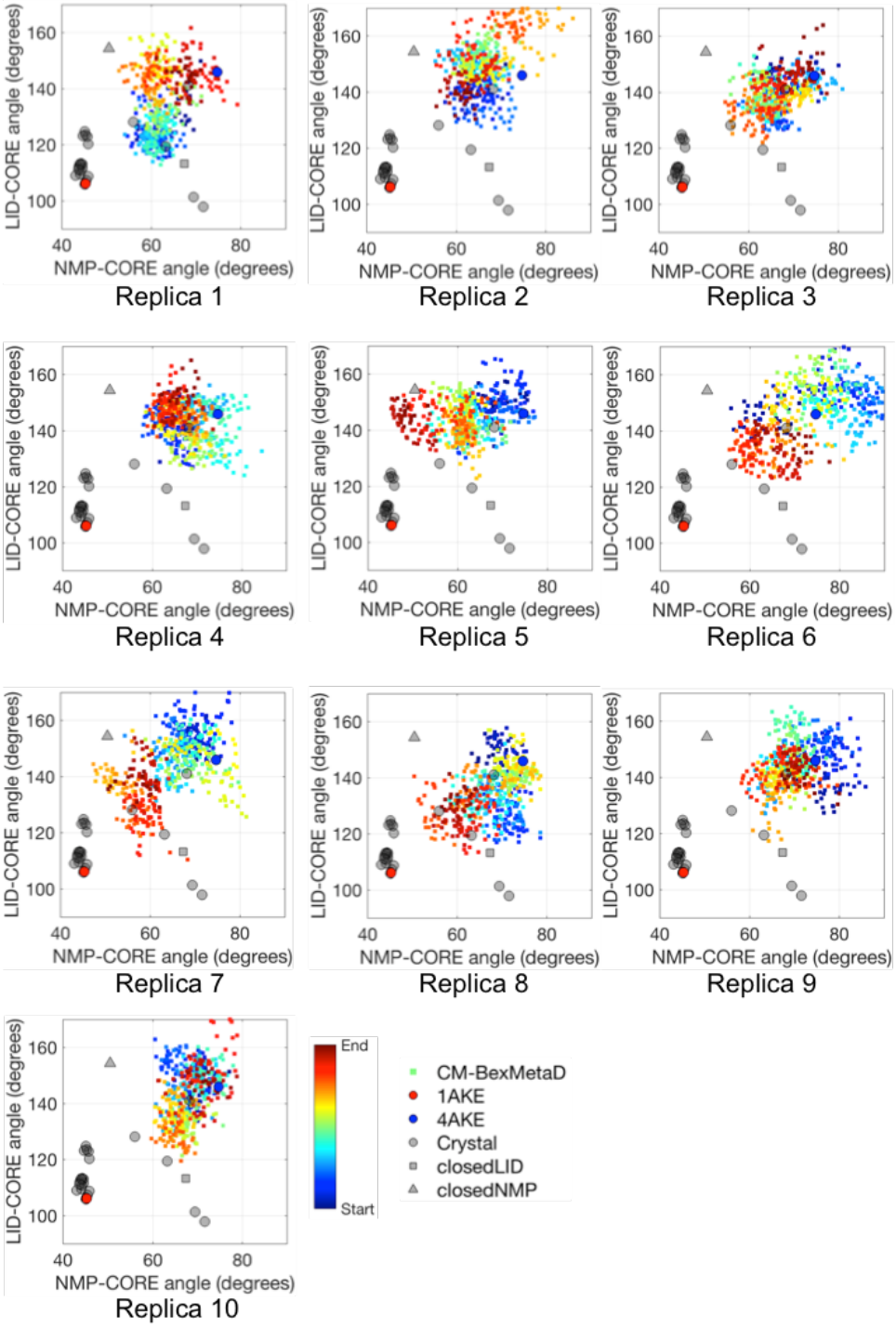


Figure 7.7. Conformational Sampling of AdK in Terms of LID - CORE and NMP - CORE Angles for the wild - type - 1 Simulation Via CM - BexMetaD - 2 Protocol Together with experimentally determined Structures. The Angles are Color - Coded According to the Simulation Time.

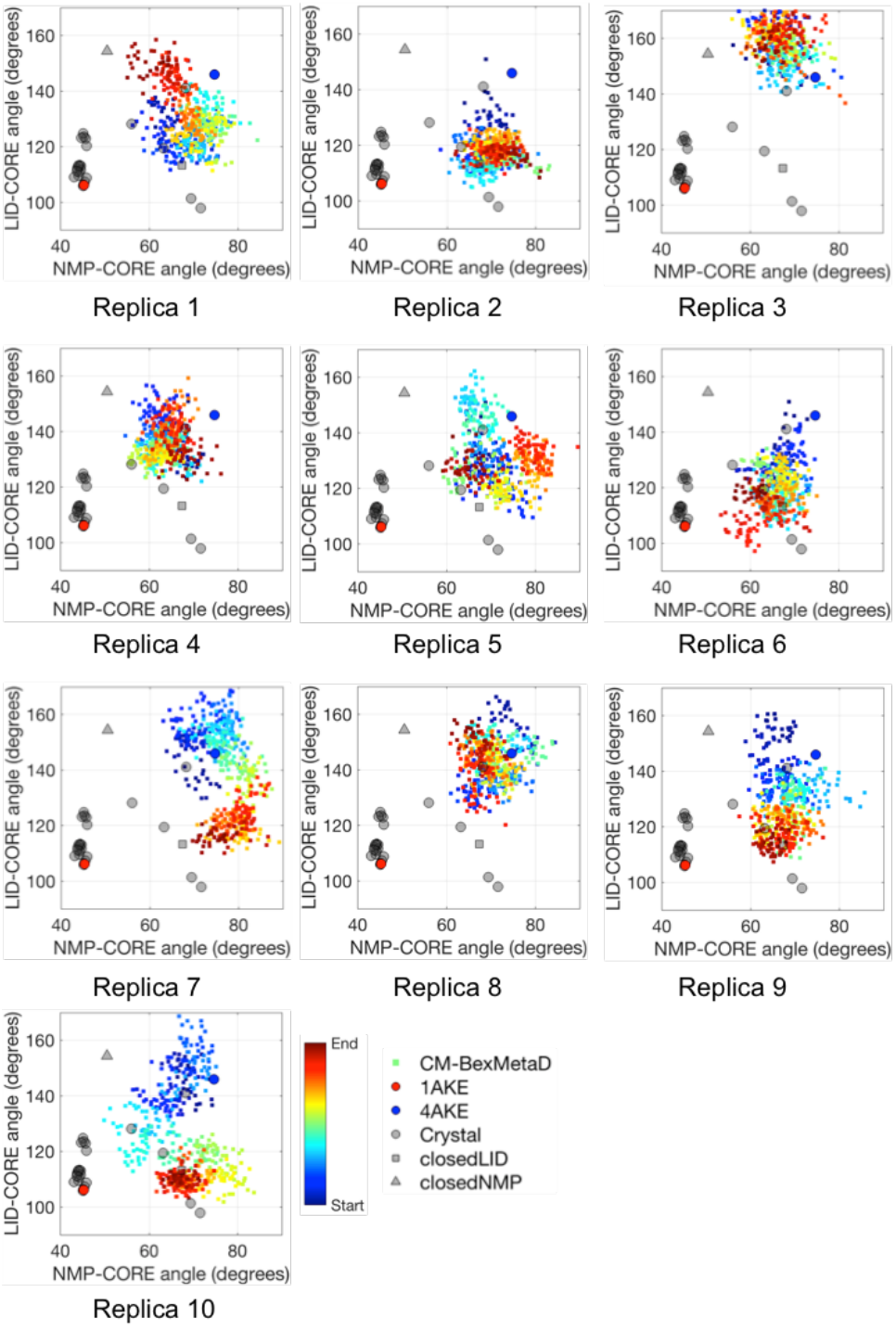


Figure 7.8. Conformational Sampling of AdK in Terms of LID - CORE and NMP - CORE Angles for Y171W Simulation via CM - BexMetaD - 2 Protocol Together with experimentally determined Structures. The Angles are Color - Coded According to the Simulation Time.

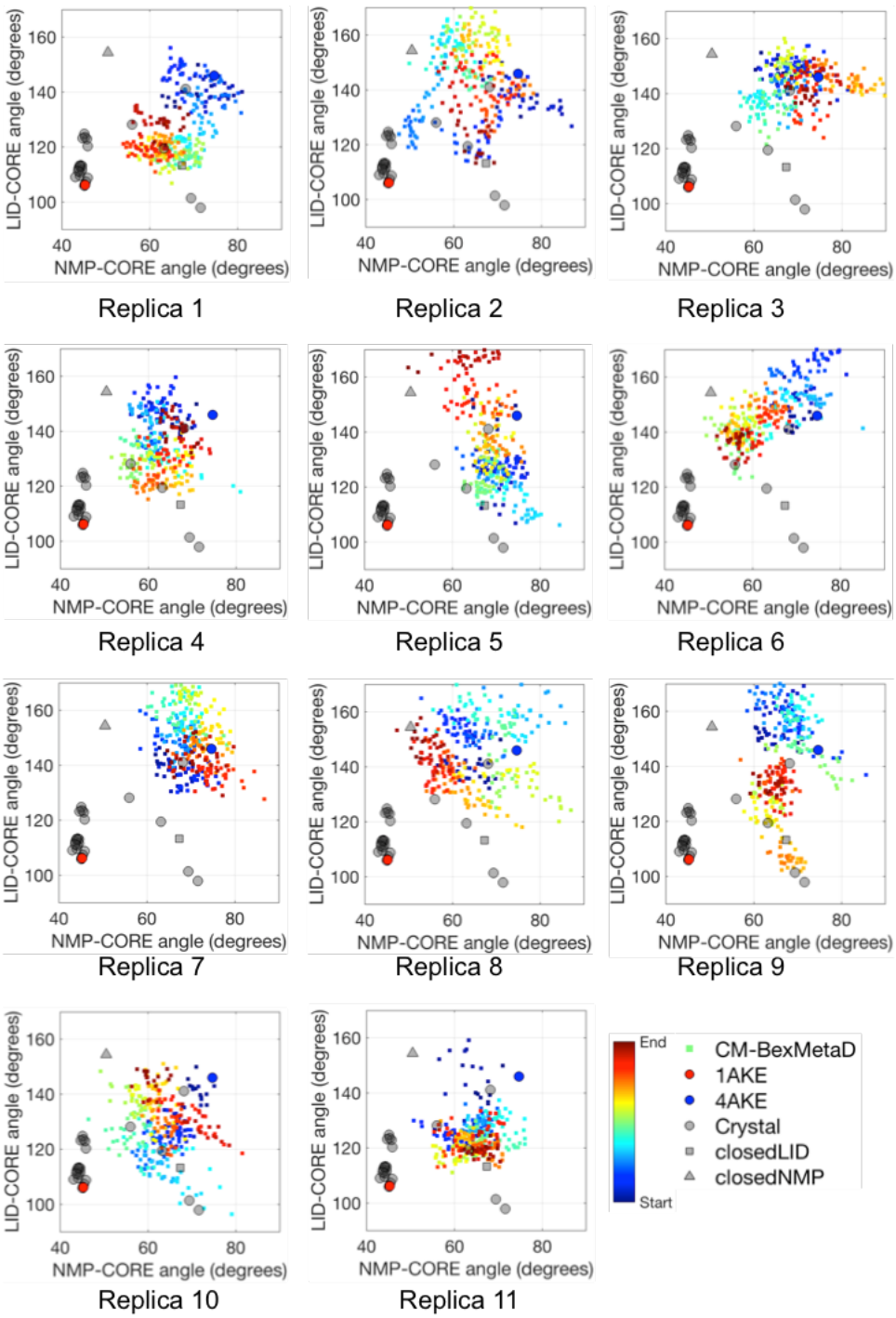


Figure 7.9. Conformational Sampling of AdK in Terms of LID - CORE and NMP - CORE Angles for the Wild - Type - 2 Simulation via CM - BexMetaD - 2 Protocol Together with experimentally determined Structures. The Angles are Color - Coded According to the Simulation Time.

7.3.3. Free Energy Landscape of Conformational Transition of apo AdK

We have calculated the hyper-dimensional PMF as free-energy landscape (FEL) of conformational transition of apo AdK, projected on the combination of collective modes as given in Figure 7.10. There are significant quantitative differences in the reported FEL underlying the conformational transition of AdK [95, 132, 216, 218]. The obtained FEL of wild type apo AdK well agrees qualitatively with the computational and experimental studies [95, 205, 216]. The vicinity of zero position in the reaction coordinate corresponds the initial conformation of the simulation, which is the open conformation of the enzyme in the present simulation. In accordance with the literature, the energy basin of the open conformation is the global minimum of the conformational ensemble of apo AdK.

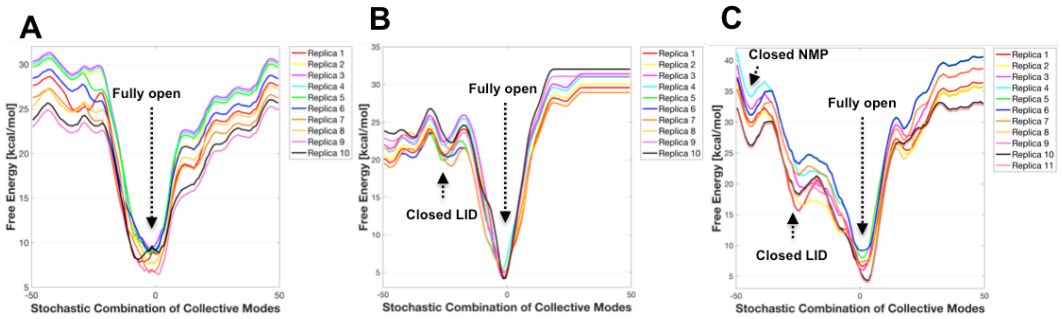


Figure 7.10. FEL Along the Conformational Transition of Wild Type and Mutant apo AdK. a) The Wild Type-1, b) Y171W, c) the Wild Type - 2.

Since none of the simulations are converged to the fully-closed state, it can be concluded that, the energy barrier enclosing the fully-closed state is higher than the other populated conformational states. Since simulations seem not to converged yet, the absolute value of the transition energy can not be calculated yet.

As explained in the previous section, closed LID conformation is observed in the wild type-2 and Y171W simulations, whereas closed NMP conformation is observed in the wild type-1 and wild type-2 simulations. Projecting the sampling capacity of closed LID and closed NMP conformations onto the constructed FEL along the conformational transition of the enzyme, it can be easily observed that, the closest energy basin to

the fully-open conformation is the closed LID state and the second basin observed in the wild type-2 and partially in the wild type-1 belongs to the closed NMP state as illustrated in Figure 7.10. This observation implies that along the transition of the open to closed states, closure of the LID domain precedes the closure of the NMP domain in accordance with the literature [208, 220].

7.3.3.1. Hinge Mutation Induces a Shift in the Conformational Ensemble of apo AdK. Y171 is a hinge site that controls the conformational transition of the enzyme, where the protein can still sample both conformational states, but the interconversion between these states is altered via Y171W mutation. Therefore it can be concluded that, this mutations affects the conformational transition via redistributing the FEL. The alteration effect of Y171A mutation is observed even at the conformational transition from open to semi-closed conformations. Effects of Y171W should be considered in both entropic and enthalpic point of view. According to the Y171W CM-BexMetaD simulation results, Y171W mutation increases the energy barrier between fully-open to closed-LID conformational states as seen in Figure 7.10. This is the energetic effect of the Y171W mutation. Furthermore, a steeper free-energy basin is observed via mutation where the entropic effects control the width of the free-energy landscape as explained in section X (tezin teori section). Therefore, it can be speculated that, the Y171W mutation redistributes the FEL making the conformational transition of the enzyme from fully-open to semi-closed states an entropically favorable but energetically unfavorable process. Therefore the interconversion of open to closed conformation of the mutated enzyme depends both energy and entropy of the system. Moreover, the transition state observed on the right hand side of the fully-open conformation (global minimum) in the FEL observed at the wild type simulation is distinguished with Y171W mutation. This result indicates the restricting effect of hinge site mutation on the conformational transition of the enzyme where at least one rare, locally stable conformation could not be sampled via mutation.

7.3.3.2. Slowest Modes are Sufficient to Sample the Transition of apo AdK. In a very recent study, Kurkcuoglu and Doruker successfully sampled the conformational transi-

tion of apo AdK using only slowest three modes of ANM through an unbiased approach [222]. Therefore, we designed a non-targeted CM-BexMetaD-2 approach to sample the conformational transition of AdK with a biased but non-targeted methodology.

Non-targeted CM-BexMetaD simulations were performed to understand whether the slowest modes provide sufficient degree of freedom to sample the conformational transition of AdK. Simulation parameters are summarized in Table 7.5. RMSD values over the trajectory of each replica during the simulations are given in Figure 7.11. The minimum RMSD values of each replica from the closed conformation are listed in Table 7.10. Through this methodology, we were able to approach the fully-closed conformation of the protein with a minimum RMSD values $3.9\tilde{A}$ within very short time window, where a lower Gaussian hill height (2 kcal/mol) is used.

Although the conformational transition to the fully-closed state can not be sampled with only the slowest five modes of motion within the simulation time window, closed-LID conformation is partially sampled in replica 1, 6, 7, 9, and 10 with a low probability (see Figure 7.12-A). In addition, the closed-NMP state is partially sampled in replica 10 with a low probability. The corresponding energy landscape basins were also observed with a very low probability (see Figure 7.12-B). It can be concluded that the conformational transition of AdK can be sampled via slowest five modes of motion although not the exact holo conformation was sampled yet. Indeed, we got conformers closer to the holo conformations with targeted approach. The energy basins of the semi-closed conformations (closed-LID and closed-NMP) could not be resolved as in the targeted protocol. These observations indicates the importance of local, intradomain bending motions in the conformational transition of AdK as suggested by Matsunaga *et al.*, [223] or requirement of substrate for further closing of the domains [222].

Table 7.10. Minimum RMSD Table of Each Replicas from The Fully Closed, Closed LID and Closed NMP of Non-Targeted Simulation of Apo AdK. Closest RMSD Value is colored in red.

Cases	Replicas									
	1	2	3	4	5	6	7	8	9	10
Fully Closed	3.99	5.29	5.39	4.39	4.56	4.11	4.1	4.8	4.33	3.87
Closed LID	2.09	3.77	3.61	2.67	2.38	1.99	2.17	3.05	2.42	2.44
Closed NMP	2.61	2.38	2.26	2.2	2.21	2.34	2.8	2.87	2.98	2.02

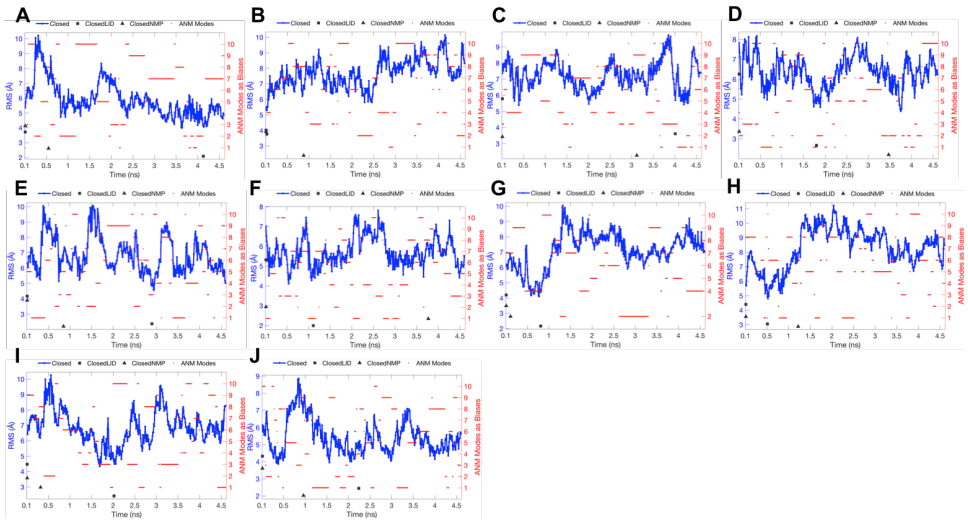


Figure 7.11. RMSD and the Biased CV Versus Simulation Time of Each Replica for apo AdK with Non-Targeted simulation. Replicas 1 to 10 are given in A-J, Respectively.

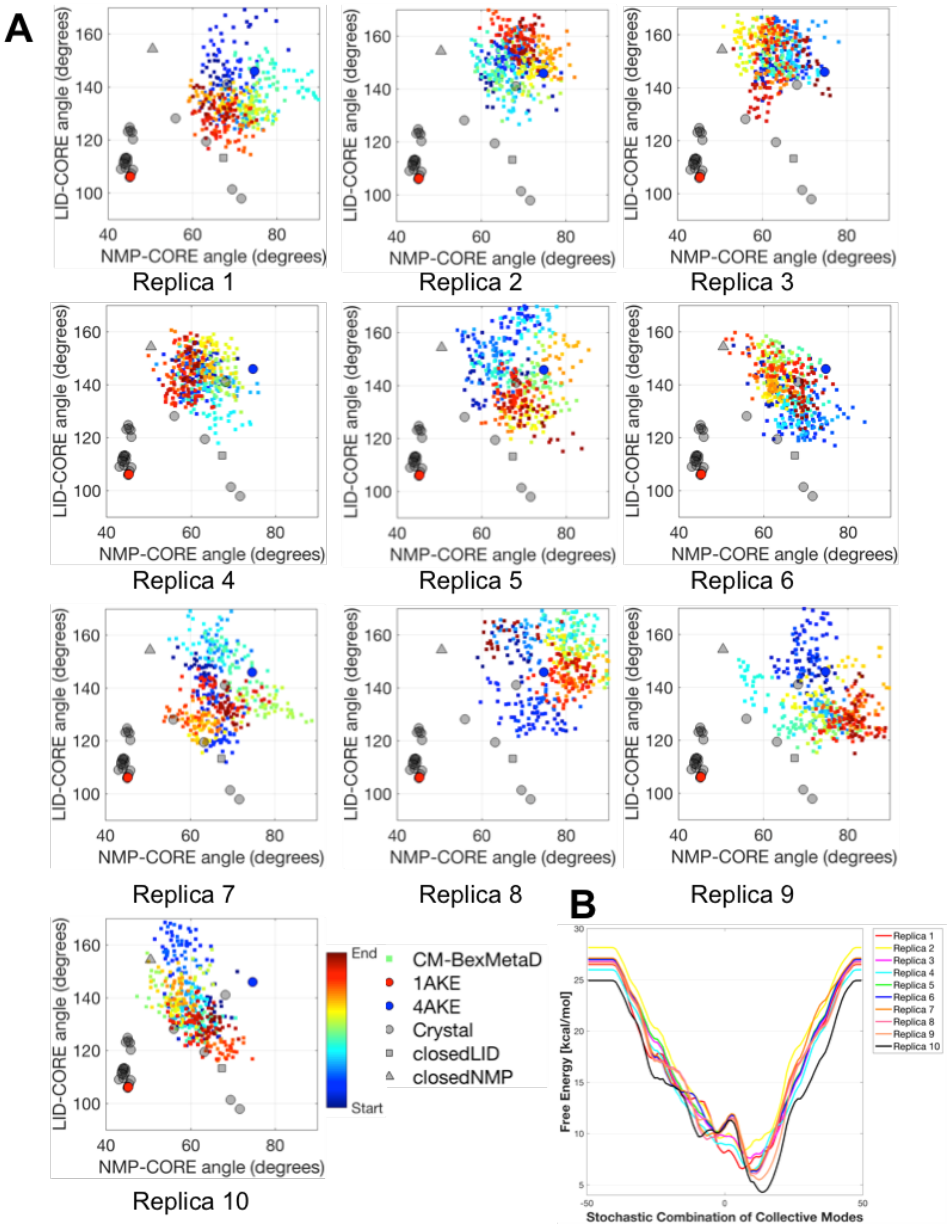


Figure 7.12. a) Conformational Sampling of AdK in terms of LID - CORE and NMP - CORE Angles for Non-Targeted Simulation of apo AdK The Angles are Color-Coded According to the Simulation Time. b) FEL Underlying the Conformational Transition of the Enzyme.

8. ALLOSTERIC CONTROL OF BINDING DYNAMICS via HINGE RESIDUES

Binding is one of the main functions of proteins and the understanding of protein-protein interactions is at the heart of the signaling processes occurs in the cell. The phenomenon allostery might closely be coupled to the binding mechanism. Proteins are potentially allosteric structures and allostery is a key process in regulation of cellular pathways. In allosteric proteins a perturbation with a ligand (small or large molecule) binding, may lead to a change at the substrate binding sites through conformational change and dynamics and end up with reshaping the energy landscape. These allosteric interactions may enhance or restrict a function such as complex formation for the signal transduction. Knowledge of the conformational states in the native state ensemble can provide significant information to the understanding of the fundamental rules governing the allosteric interactions. Understanding the relation between binding and allosteric regulation will give an opportunity to control and even interfere the mechanism that could be very important for the targeted drug design.

To this end, in order to address the importance of the hinge residues in the allosteric control of the binding dynamics, a detailed research is performed. It is illustrated that there is a significant association of allosteric mutations, which lead to high binding affinity changes, with the hinge positions of global modes by the analysis of a structural dataset of kinetics and energetics of mutant protein interactions (SKEMPI). We elaborate on the mechanism of allosteric dynamics further, through exemplary case studies on human growth hormone (hGH) and pyrin domain (PYD), that show how the mutations at hinge regions could be instrumental to allosterically affect the binding site dynamics or dispose alternative binding modes by modifying the ensemble of accessible conformations. This research is published in Biophysical Journal [52].

The inflammasome complexes are activated by rapid formation of ASC (Apoptosis associated Speck-like protein containing CARD) speck which acts as an adaptor that

bridges procaspase-1 to the receptors. The human ASC protein contains two death domains (N-terminal Pyrin (PYD) and C-terminal CARD) and these two domains are connected via flexible 23-residue linker (starting from residue 90 to residue [112]). The PYD domain is composes of residues [1-89] and the CARD domain is between residues [113-195]. The linker who connects the PYD and CARD domains has some local structure, which leads to a back-to-back orientation of those domains. As a result of this orientation, these domains can interact with multiple partners.

The resemblance between the ASC speck and aggresome raises the question whether the ASC speck formation is a result of specific interactions between PYD and CARD which both belong to the death fold superfamily or simply aggregation of hydrophobic patches of ASC proteins. To address this question, we performed structure and dynamics based analyses on the ASC protein using Gaussian Network Model (GNM) of PYD and CARD, and Molecular Dynamics (MD) simulations of the wild type and $in-silico$ mutated PYD, with the mutational analysis on the ASC structure and its separated domains in human cells, we show that the ASC speck formation is an organized structure based on specific homophilic but not heterophilic interactions by PYD and CARD separately. PYD is able to use alternative interaction modes other than type I that might be important in compaction of the ASC speck. We propose a model in which filament formation is the first level of organization in the ASC speck and filaments further compact in a higher organization level. This research is published in Structure [143].

9. MUTATION INDUCED ALLOSTERIC REGULATION OF KINESIN - TUBULIN BINDING AFFINITY INVESTIGATED BY AFM

9.1. Introduction

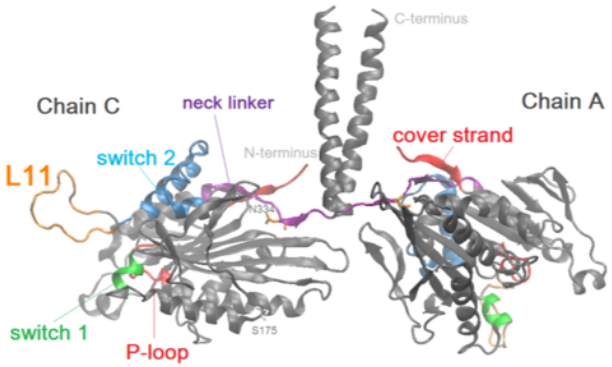
Molecular motors are key components of the cell responsible for all kinds of movements to realize cellular processes in living organisms. Further understanding of the motility and the force generating mechanisms of molecular motors requires the ability to apply forces on them and test their response. This can be performed in great control and precision with single-molecule experiments. Advanced single-molecule spectroscopic and microscopic techniques have made possible the visualization and engineering of motor proteins and led to the development of multiple theoretical approaches describing different aspects of their dynamics [224].

Atomic force microscopy (AFM) has proven to be a suitable approach to study the intermolecular forces and dynamics controlling the protein-protein/protein-ligand interactions, having the unique advantage over other single-molecule techniques: to probe molecular interactions at high resolutions in combination with force sensitivity [225]. AFM requires attaching one of the molecule pair to the AFM tip, and the other to the substrate surface, bringing the tip into contact with the surface so as to form a complex. Pulling the tip until the interaction bond breaks enables the estimation of the unbinding forces and the characterization of the force-driven pathway along the pulling direction until the bond ruptures, revealing the energy landscape of a bond rupture.

As one of the smallest molecular nanomachines, kinesin-1 (conventional kinesin) is a prototype of molecular motors responsible for carrying cargoes along microtubules (MT) and regulating the microtubule dynamics. Human kinesin is a dimeric protein containing a smaller motor/head domain (M1-T323). The functional sites of the motor

domain are microtubule binding sites, the nucleotide-binding pocket (catalytic site), the neck-linker (NL), which connects the motor domain to the neck-helix domain (see Figure 9.1). ATP binding pocket of kinesin is formed by the P-loop, switch I and II regions where a conformational change of these regions is observed with the ATP hydrolysis.

Kinesin molecule moves along MT by adapting its conformation according to the different nucleotide states of the motor domain, such as ATP, ADP.Pi, ADP bound forms. ATP molecules hydrolyzed on kinesin and the energy generated are used for the walking process of kinesin on MT. Kinesin walking on MT is an asymmetric hand-over-hand mode, which is two head domains are alternating in the lead [226-130]. During action, the rear head domain moves around 15 nm, whereas the other head does not translate due to strong binding to microtubule affected by the ATP binding. This argument was corroborated by fluorescence imaging techniques [231].



	Functional Site	Resid*
	Cover strand	M1-I9
Nucleotide Binding Pocket	P loop (N1)	G86-T93
	Switch 1 (N2)	N199-R204
	Switch 2 (N3)	D232-E237
	N4	R14-P17
Microtubule Binding Residues	β5 strand and Loop 8	S150-T170
	Loop 11	S240-E251
	Switch 2 cluster	N256-I287
	Neck Linker	I327-T338
	Catalytic	G235

Based on rat kinesin

Figure 9.1. Kinesin Functional Sites are Shown on Structures.

In the absence of ATP, the kinesin molecule is with ADP bound and the two head domains are bound to the tubulin. In this state, the NL points forward and rearward on the trailing head and leading head, respectively. The ATP binding on the trailing head initiates and completes the docking of NL to the leading head, rotates the trailing head to 160 $\tilde{A}$ forward, starts to perform a diffusional search. Then the head tightly binds to the next tubulin, which makes it new leading head [232]. This process completes within a few-milliseconds. During this marching process, the new trailing

head completes the ATP hydrolysis and becomes ready for a new ATP binding. Kinesin generates forces in the order of some piconewtons through this stepping mechanism. However, loading kinesin bound on MT with a force, it stops its motion and eventually dissociates from MT. The amount of force that dissociates kinesin from MT is called the stall force.

Comparison of NL regions in the available crystal structures of kinesin motor domains reveals the importance of NL in the force generation mechanism. A novel force-generating mechanism has been proposed for the cover-neck-bundle (CNB) formed by the interaction of the cover strand (M1-I9) with N332 of the NL regions. Besides, mutagenesis studies in the NL region resulted in impaired motility while the ATPase and MT binding activities of the motor domain stayed relatively unchanged [233, 234].

In a recent experimental study of full-length human kinesin-1 (KIF5B), mutations were introduced on the phosphorylation site of kinesin-1 (S175D and S175A) to explore the effect of phosphorylation on fast axonal transport (FAT). In both cases ATPase, microtubule binding affinity and processivity were unchanged, but application of force differs. Placement of negative charge at S175, through phosphorylation or mutation, leads to a lower stall force and decreased velocity under a load of 1 pN or greater in optical trap experiments [235]. Also, addition of a negative charge at S175 -at physiologically relevant salt concentrations- favors the compact/autoinhibited conformation of kinesin, since the tail contains a conserved positively charged motif that binds/docks to motor domain near residue 175 [235, 236]. Since residue 175 does not fall within the MT- or ATP-binding domains, the effect of a modification at this position cannot be readily inferred by its location on the motor domain. The biochemical and biophysical effects of altering (phosphorylation, alteration of charge, etc.) residue 175 should be thoroughly investigated [235].

Point mutations on the KIF5A neuronal kinesin motor domain leading to SPG10 disease (which is a form of the hereditary spastic paraplegia, HSP) have been detected including D73N in rat (which is homologues to D72N in human kinesin). Majority of these mutations are located in the P-loop, switch I and II regions [27], [238]. D72N

is in close vicinity of the G76, which forms an H-bond with N332 and therefore forms the major interaction between the neck linker and the motor domain. D72N mutation results in dysfunctional kinesin motors with reduced transport velocity on microtubules resulting in impaired axonal transport and axonal degeneration [237].

Based on the mutagenesis studies given in the literature that led neurodegenerative diseases and listed in the cancer mutation database (COSMIC) [239] on kinesins that overlaps with the hinge sites determined by Aykac et al. D72N, S175A, N332A mutations are considered to study via AFM-DFS experiments. Kinesin's allosteric network underlying its mechanochemical cycle is based on the large conformational change triggered by the hydrolysis of ATP. This network of communication probably involves many more unidentified allosteric sites in addition to the key residues already determined in the literature. Here, we have chosen to work on the wild type and mutant (D72N, S175A, N332A mutations in human) kinesin-1 to explore their allosteric effect in the $\alpha\beta$-tubulin binding, considering previously discussed molecular/functional aspects of these residues to contribute to the understanding of the potential defective functioning mechanism of the protein. In this respect, we have performed AFM single molecule pulling experiments probing the unbinding force ranges of kinesin-$\alpha\beta$-tubulin complexes for studying the protein allostery by mechanical signals.

9.2. Materials and Methods

9.2.1. Preparation of Molecules and AFM Setup

All the biotinylated wild type and mutant human Kinesin (family of KIF5) Heavy Chain motor domain (will be called as Kinesin throughout this chapter) produced in a bacterial expression system, was purchased from Cytoskeleton Inc. USA, as lyophilized powder. The powdered form of the molecules were reconstituted and diluted to $2\mu g$/ml with CMW Buffer 1. The CMW Buffer 1 (containing 100 mM PIPES pH 7.0, 200 mM KCl, 2 mM $MgCl_2$, 1 mM DTT, and 200 μM ATP) was also purchased from Cytoskeleton Inc. USA in liquid form. Tubulin protein (α/β tubulin) was purchased from Cytoskeleton Inc. USA, as white solid, which has been purified from porcine

brain. Tubulin molecules were reconstituted and diluted to 2μg/ml with General Tubulin Buffer (GTB). GTB was also purchased from Cytoskeleton Inc. USA, as white lyophilized powder, which was reconstituted with de-ionized water to give 1X strength buffer (80 mM PIPES, 2 mM $MgCl_2$, 0.5 mM EGTA pH 7.0). Streptavidin molecules were purchased as lyophilized powder from Sigma-Aldrich (St. Louis, MO, USA) and reconstituted in PBS to have final concentration of 100 μg/ml. Bovine serum albumin (BSA) was purchased from Sigma-Aldrich (St. Louis, MO, USA). Mica surface was purchased from Ted Pella Inc. Biotinylated PEG attached Si_3N_4 AFM probes (MLCT. BIO, Novascan, Ames, IA, USA) were used in the experiments. The nominal spring constants of the cantilevers used in the experiment are 0.01 and 0.1 N/m for lower and higher velocities, respectively. A commercial AFM system (Dimension Edge, Bruker Nano, Santa Barbara, CA, USA) was used in the experiments.

9.2.2. Immobilization of Tubulin Molecules and Functionalization of AFM Tips with Kinesin

Prior to AFM data collection, Tubulin molecules were immobilized on a mica surface via physical adsorption by incubating 100 μl of previously prepared Tubulin aqueous solution on mica surface for 20 min. Tubulin molecules have high affinity for silica, therefore they can strongly sticked to the surface. After incubation completed, tubulin solution was pipeted and 100 μl bovine serum albumin (BSA) was dropped onto the surface and incubated for 20 min. Then the mica surface was gently washed with deionized (DI) water to remove loosely adhered molecules. The mica surface with immobilized molecules was kept hydrated via GTB prior to data collection.

Streptavidin was used as a mediator to functionalize biotinylated-PEG attached Si_3N_4 AFM tips with biotinylated Kinesin molecules. Biotinylated cantilevers were first incubated with 20 μl streptavidin (100 μg/ml in PBS) for 15 minutes at room temperature. Afterwards, AFM tips were washed two times with DI water to remove unbound streptavidin molecules. The streptavidin-functionalized tips were dipped into 100 μl of Kinesin solution and incubated for 40 minutes at room temperature. The kinesin functionalized AFM tips were used immediately in force measurements after

rinsing with DI water to remove the loosely bound Kinesin molecules. Figure 9.2-A shows the experimental setup for the kinesin-tubulin system and the multi-step functionalization of AFM tip and surface in this experiment. The position of the applied force on the kinesin molecule together with the predicted functionally important residues and the ligand position are shown in Figure 9.2-B.

Multi-step functionalization of AFM tips may introduce measurement of unbinding forces that do not arise from the molecules of interest. To test the existence of such undesired unbinding events, the pulling experiments were performed using: 1) PEG+biotin attached tip and non-treated mica surface, 2) PEG + biotin + streptavidin attached tip and non-treated mica surface, 3) PEG+biotin+streptavidin attached tip and $\alpha\beta$-tubulin covered mica surface, 4) PEG+biotin attached tip and $\alpha\beta$-tubulin covered mica surface, 5) PEG+biotin + streptavidin + kinesin attached tip and non-treated mica surface. No statistically significant specific adhesion was observed between biotin and mica surface, whereas specific adhesion forces were observed in all other cases. Therefore, BSA was used as the nonreactive protein for reducing the nonspecific adhesive forces and undesired interaction of the multi - functionalization elements. Additionally, in order to minimize random errors such as variation in surface coverage, the data collected were repeated at different locations on the surface, with different cantilevers and also at different dates.

9.2.3. Evidence of Specific Interaction-Control Experiments

To confirm if kinesin functionalized on the cantilever has a specific binding ability to $\alpha\beta$-tubulin immobilized on the mica surface, a control experiments was carried out, in which kinesin functionalized cantilever was saturated with excess $\alpha\beta$-tubulin (incubation with 100 ul tubulin solution for 20 min) prior to force-distance data collection.

9.2.4. Atomic Force Microscopy (AFM): The Experiments Performed

We have performed two different sets of AFM pulling experiments to investigate the unbinding force distributions involved in the interactions of Kinesin -$\alpha\beta$ -tubulin

complexes. All force measurements were performed in fluid-phase at room temperature. The cantilevers were calibrated using thermal noise method (Hutter and Bechhoefer, 1993). AFM pulling experiments were performed by the approach of the Kinesin attached cantilevers -manipulated by a piezo- on a $\alpha\beta$- tubulin adsorbed onto the surface (AFM protocol described above). The deflection signal, which gives the position of the AFM cantilever, was recorded during the experiment. To obtain the unbinding forces at various loading rates, force-distance data were recorded at different approach/retract velocities -spanning three order of magnitudes- (0.5, 1, 2.5, 5, 10, 20, 50 μm/s). At each speed, we collected sufficient force curves to obtain statistically significant number of force curves. The trigger force was fixed to 1 nN.

The unbinding force at a specific loading rate was determined by inspecting the force-extension curve. Figure 9.2-C shows a typical force curve (representing just retract) indicating no adhesion/rupture event on the right and two successive unbinding events between a single pair of molecules. The effective loading rates were estimated by fitting a linear line to the force-time curve immediately before the rupture point in each approach/retract. The effective loading rates obtained from the experiment were equally quantized in the logarithmic space to get uniform number of most probable rupture force vs logarithm of loading rates relation. Force distributions are plotted as histograms for each loading rate-range- and the most probable rupture force at each loading rate is determined by inspecting the force histograms. The most probable rupture force of the kinesin-tubulin complex at a given loading rate interval was determined by calculating the median of the force histograms. The most probable rupture force f*, versus the natural logarithm of the loading rate, ln(r_f), was plotted and the Bell parameters (k_{off} and x_β; where k_{off} is the unbinding rate constant and x_β is the distance from the bound state of the activated state), which are characteristic values of the investigated molecular pair, were estimated from the slope, $\frac{k_BT}{x_\beta}$, and the intercept, $\frac{k_bT}{x_\beta k_{off}}$, of the fitting linear curves (See Equation 9.1).

$$f^* = \frac{k_BT}{x_\beta}\left(lnr_f - \ln\frac{k_BT}{x_\beta}k_{off}\right) \tag{9.1}$$

where k_B and T are the Boltzmann constant and the absolute temperature, respectively.

9.3. Results and Discussion

9.3.1. Kinesin-Tubulin interactions via AFM experiments

Force measurements of the wild type and mutant (S175A, N332A, D72N) kinesin-tubulin interaction were carried out using Dynamic Force Spectroscopy (DFS) via AFM as described in Materials and Methods (Figure 9.2-A) to measure the magnitude of unbinding forces and to explore the binding dynamics of the kinesin-tubulin complex.

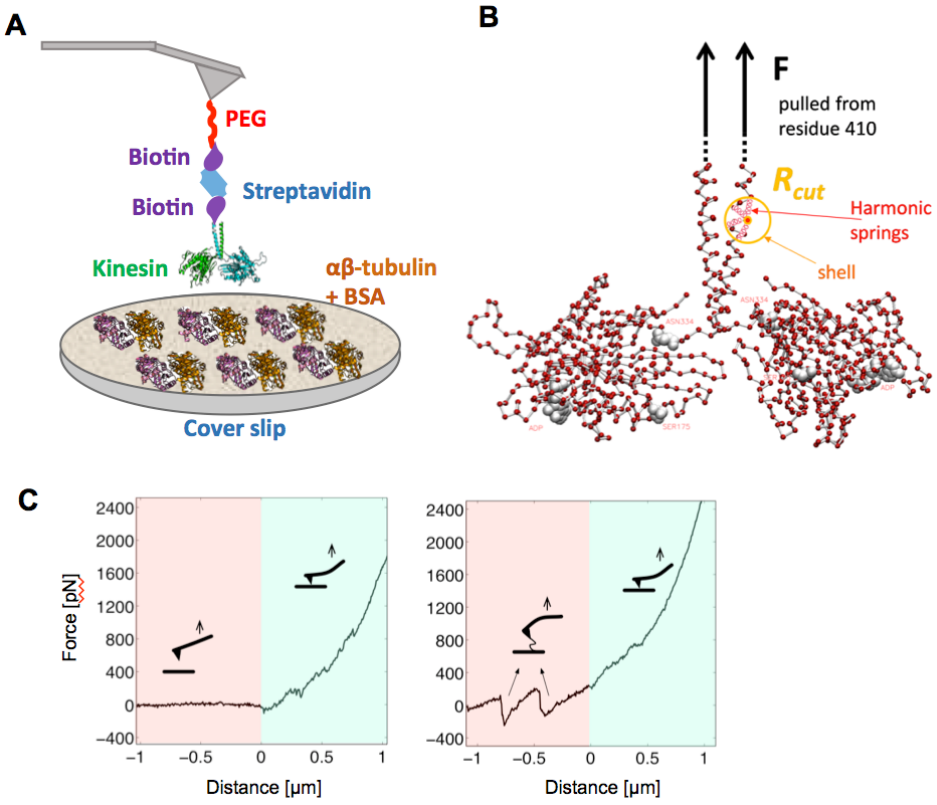


Figure 9.2. Illustration of AFM Pulling Experiments with and Example Force - Distance Curve. a) The experimental Setup. b) The Functionalization of Kinesin Molecule c) A Typical Force Curve Exhibiting no Adhesion/Rupture Events (Left) and an Unbinding Event (Right) with a Force Strength ($F_{unbinding}$=250 pN).

9.3.2. Confirmation of Specific Interactions

The functionalization and immobilization protocol optimization experiments revealed that the intermediate agents used in the experimental setup causes some non-specific force measurements. To reduce the probability of unbinding forces coming from multi-functionalization agents, the wild type kinesin-tubulin experiments were repeated by adding a nonreactive protein (BSA) on the surface to coat the uncovered mica surface. The wild type kinesin-tubulin interaction experiment was repeated via the updated protocol. The specificity of the measured unbinding forces between kinesin and tubulin was checked by saturating the kinesin functionalized AFM cantilevers via tubulin solution before collecting kinesin-tubulin unbinding data as a control experiment. The probability of adhesion curves of the experiments together with the control experiments were calculated and shown in Figure 9.3. Saturating kinesin functionalized AFM tip before measuring the kinesin-tubulin interaction reduces the probability of adhesion to half of the value obtained via unsaturated experiment implying that the specific kinesin-tubulin interaction was causing the observations in unbinding experiments.

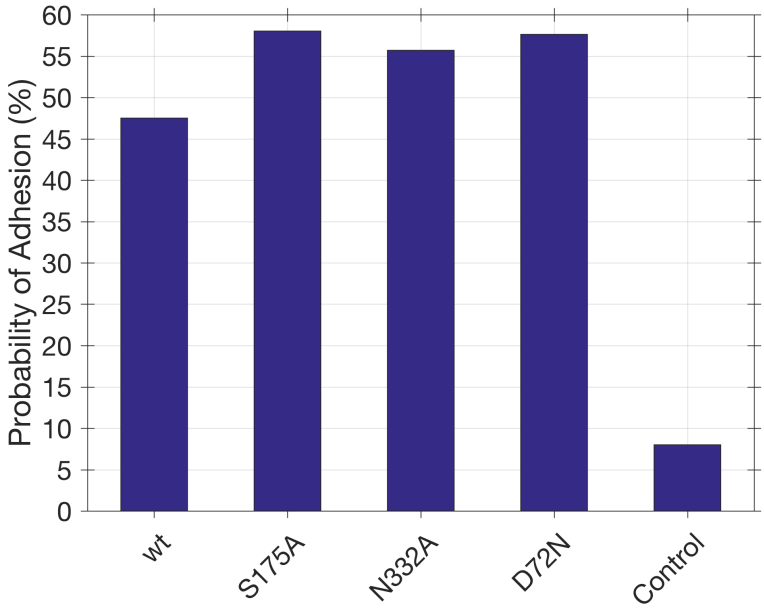


Figure 9.3. Comparison of the Probability of Adhesion of Regular Experiments with the Control Experiment (Functionalized AFM Cantilevers Were Saturated with Excess Tubulin Molecules before the Pulling).

9.3.3. Kinesin Tubulin Dissociation Reaction Follows Alterative

The adhesion forces between kinesin and $\alpha\beta$-tubulin molecules in the wild type and the S175A, N332A, and D72N mutants were extracted from the collected force measurement data. Both the wild type and mutant kinesin-tubulin complexes reflect unimodal distribution at lower loading rates. However, at higher loading rates the dissociation process follows bimodal distribution for all cases as reflected in the force histograms given in Figure 9.4. This bimodality suggests the existence of two distinct states of kinesin-tubulin association where a low-strength state dissociates at lower forces (20-120pN) and a high-strength state dissociates at higher forces (180-300pN). The two different states may imply the existence of two distinct population of the molecular complex in the ensemble of conformations (existence of two different conformational isomers or/and different dynamic behavior) where each of them has its own force distribution. However, since kinesin molecule is in dimeric state, this two state dissociation reaction may overlap with the binding of a single monomer and both monomer at the same time, which gives the low-strength state and high-strength state (at higher loading rates), respectively. Moreover, the difference in the probabilities of these two states could be associated with the relative probabilities of the different populations (either two distinct conformations (ensemble) of a given monomer or two monomers) such that the low-strength state still has greater fraction compared to high-strength state.

9.3.4. S175A and N332A mutations weakens the low-strength state of the kinesin-tubulin complex

The most probable rupture forces of the kinesin-tubulin complex at a given loading rate were determined from the histograms of measured unbinding forces. Figure 9.5 plots the most probable rupture forces against the loading rates in semilog scale for the wild-type and for all three mutant kinesin-tubulin dissociation reaction. Within the window of measured loading rates, the rupture force of all kinesin-tubulin complexes showed a gradual increase with a single linear relation in both low-strength (see Figure 9.5) and high-strength (inset of Figure 9.5) states.

As seen in Figure 9.5, the mutations do not affect the unbinding forces observed at high-strength state for all cases (inset). However S175A and N332A mutations decreases the unbinding forces of low-strength state of kinesin-tubulin complex at comparable loading rates whereas the D72N mutation has the similar unbinding forces with wild-type.

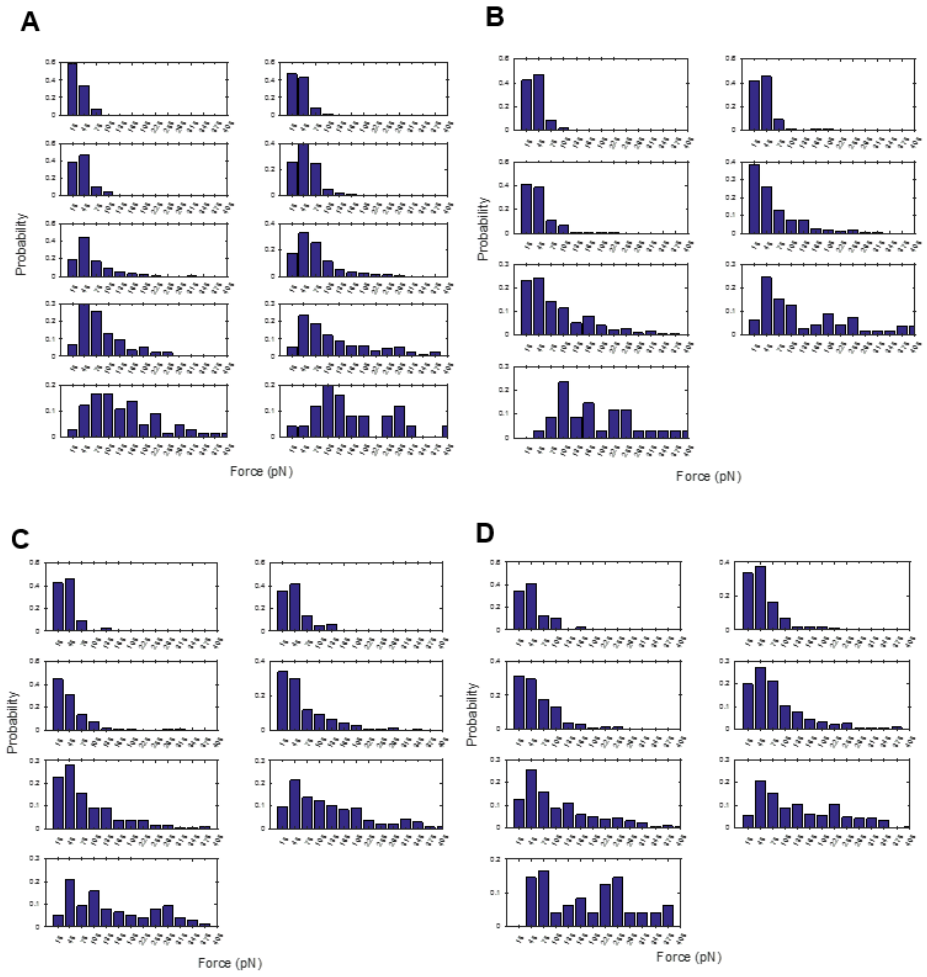


Figure 9.4. Adhesion Force Histograms Obtained at Different Loading Rates. a) Wild Type Kinesin b) S175A Kinesin c) cN332A Kinesin d) cD72N Kinesin. There is Unimodal Distribution at Low Loading Rates and Bimodal Distribution at High Loading Rates in All Cases.

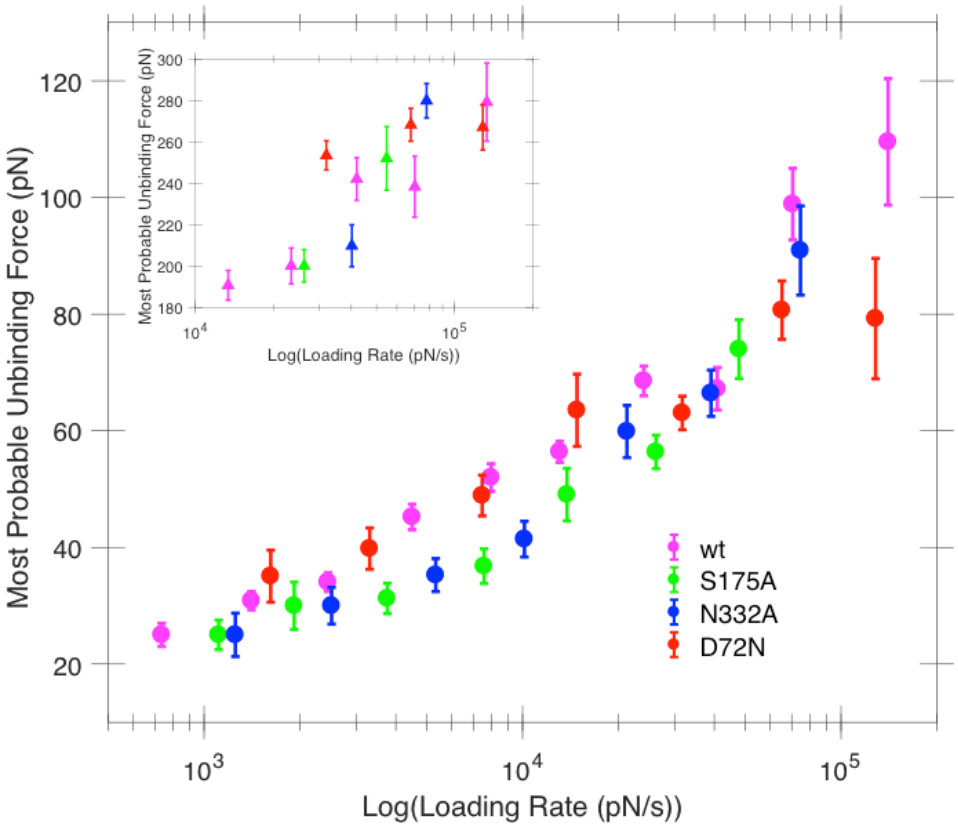


Figure 9.5. The Dynamic Adhesion Force Spectra of Kinesin-Tubulin Complex Depending on the Loading Rates. The Most Probable Rupture Force vs Logarithm of the Loading Rates Determined as Low-Strength State is given as Main Plot, whereas the High-Strength State is given as Inset.

9.3.5. Effect of Mutations on the Binding Energy Landscape of the Complex

The observed rupture forces vs loading rates relation can also be examined in the context of the Bell model (Bell G.I., 1978), which predicts a linear relationship between the rupture force and the natural logarithm of the loading rate. As shown in Figure 9.6, the dynamic spectrum of both low and high-strength states in all kinesin-tubulin binding events revealed a single linear regime within the range of the experimental loading rates. This observation indicates that all kinesin-tubulin complexes have a single minimum in the energy landscape of the disassociation process of the kinesin-tubulin complex in both low-strength (shown in spheres) and high-strength (shown in triangles) states as shown in Figure 9.6. The parameters of the dissociation energy

landscape and kinetics of the complex obtained from the Bell's model is calculated as described in Materials and Methods.

The kinetic parameters calculated using the Bell's model such as the dissociation constant k_{off} and the distance of the activated complex from the bound state x_β are given in Table 9.1 and placed on the energy landscape diagram in Figure 9.6-E and F.

In addition to decreasing the mean unbinding forces, S175A and N332A mutations also decrease the activation energy barrier of the dissociation reaction of kinesin-tubulin complex at both low and high-strength states as shown in Figure 9.6-E and F where the wild-type is shown in magenta, S175A is in green, N332A is in blue, and D72N is in red together with the calculated kinetic parameters and activation energy change with respect to wild type. The effect of N332A mutation is much effective than S175A in terms of lowering the activation energy barrier. This result explains the importance of neck linker in the force generation mechanism and motility of kinesin. Since neither S175 nor N332 is at the binding sites of kinesin - tubulin interaction, the effects of the perturbation in terms of introducing a mutation are allosterically in effect of the kinesin-tubulin complex association.

D72N is a disease causing point mutation, which results in neurodegenerative disorder. According to the DFS results, although this mutation does not affect the mean rupture force, it decreases dissociation constant of both low and high-strength states of kinesin-tubulin complex. The activation energy barrier of high-strength state is drastically increased via D72N mutation, which indicates that dissociation of D72N kinesin-tubulin complex is extremely difficult. These observation are in accordance with the mutational studies in the literature as this mutation results in dysfunctional kinesin motors with reduced transport velocity and reduced binding on microtubules resulting in impaired axonal transport and axonal degeneration (Kawaguchi, 2013).

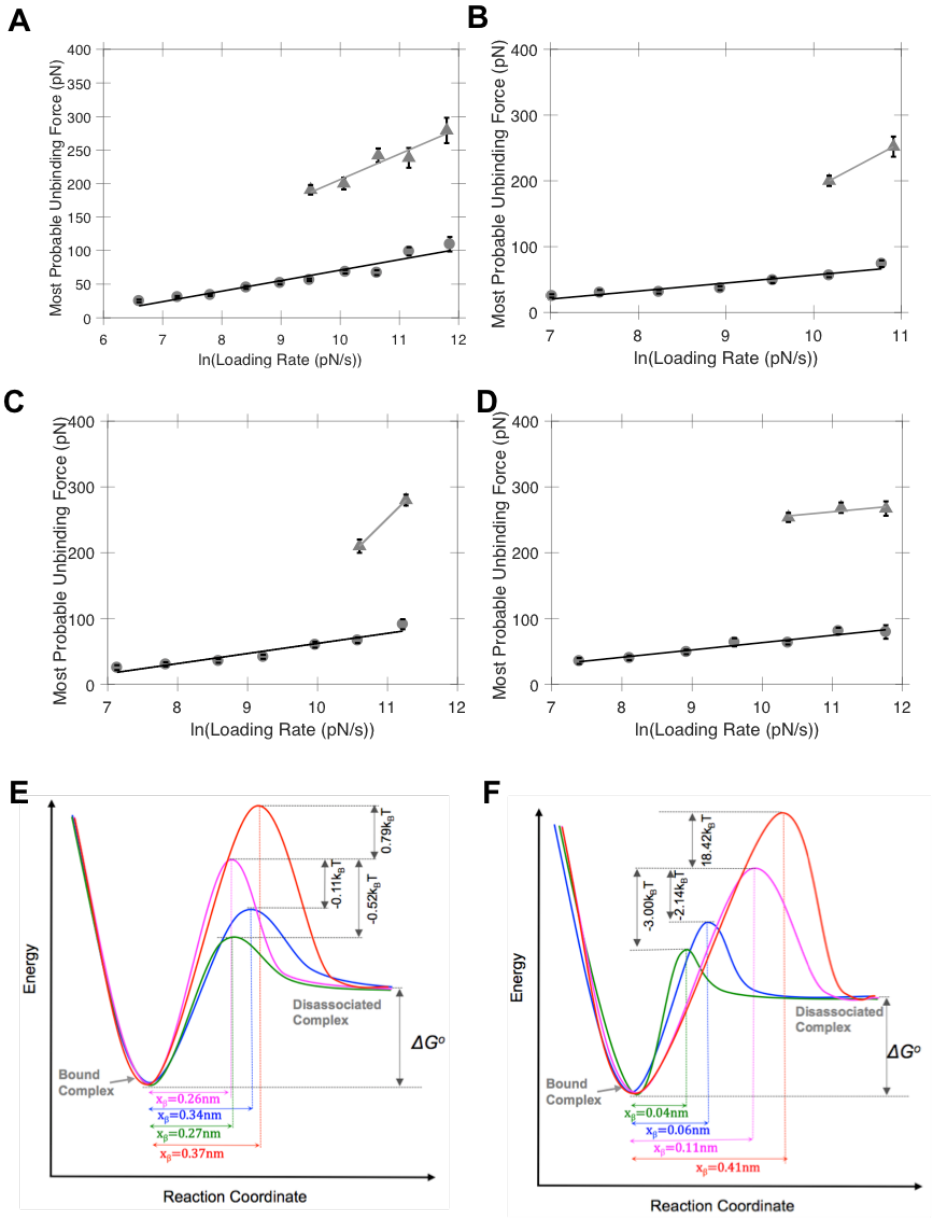


Figure 9.6. DFS Results of Kinesin-Tubulin Complex. a) Wild Type Kinesin b) S175A Kinesin c) N332A Kinesin d) D72N Kinesin. A Smooth Diagram of the Binding Energy Landscape of the Low-Strength State of Kinesin-Tubulin Complex e) and the High-Strength State f).

Table 9.1. Bell Model Parameters of the Dissociation Reaction and the Difference in the Activation Energy of the Kinesin-Tubulin Complex.

	Kinesin Type	x_β (nm)	k_{off} (s-1)	$\Delta\Delta$E (k_BT)*
Low-Strength	Wild type	2.62E-01	1.52E+01	
	S175A	3.37E-01	1.71E+01	-0.11
	N332A	2.67E-01	2.55E+01	-0.52
	D72N	3.68E-01	6.89E+00	0.79
High-strength state	Wild type	1.08E-01	2.58E+00	
	S175A	5.79E-02	2.20E+01	-2.14
	N332A	3.89E-02	5.21E+01	-3
	D72N	4.10E-01	2.58E-08	18.42
*$\Delta\Delta$E is relative to wild type kinesin-tubulin binding energy				

9.3.6. Dissociation Rate of D72N kinesin has Higher Force Sensitivity

The separation between the transition state and bound state (the width of the barrier in the energy landscape, x_β), which dictates the force-resistance of the complex, is the largest for D72N compared to both wild type and other mutations in both low and high-strength states and therefore the effect of pulling force on the dissociation constant is largest (see Figure 9.6-E, F and Table 9.1). Wider energy barrier indicates the higher number of entropically favorable microstates within the well. S175A and N332A also increase the width of the energy barrier compared to the wild type kinesin at the low-strength state, whereas decreases the barrier width at the high-strength state. These two results indicate that the S175A and N332A mutant kinesin-tubulin dissociation constants are more force sensitive than the wild-type but less force sensitive compared to D72N at the low-strength state. On the other hand, the effect of external force on the dissociation constant is smaller in S175A and N332A mutant kinesin-tubulin unbinding at high-strength state compared to both wild-type and D72N.

10. ALLOSTERIC REGULATION OF RAC1-PAK1 BINDING AFFINITY BY MUTANT RESIDUES THROUGH MOLECULAR SIMULATIONS AFM

10.1. Introduction

Rac1 (Ras-related C3 botulinum toxin substrate 1) is a small member of Ras superfamily and Rho family of GTPases, associated with the plasma membrane, acting as a binary switch between the GTP-bound active state (ON) and the GDP-bound inactive state (OFF) [240]. Rac1 activation is regulated by guanine nucleotide exchange factors (GEFs), which activate Rac1 to transmit an incoming signal through catalysis of GDP release and GTP binding via stabilizing the nucleotide-free state during the GDP-GTP exchange, and GTPase activating proteins (GAPs) that inactivate Rac1 by accelerating GTP cleavage [241]. Rac1 is an important GTPase, found in all eukaryotic organisms, regulating cell responses such as cell adhesion, cytoskeleton rearrangement, lamellipodia and membrane ruffle formation, nicotinamide adenine dinucleotide phosphate (NADPH) oxidase activation (for antibacterial defense) and induction of gene expression, through binding of numerous effector proteins in its active state [242, 243]. PAK1 (p21-activated kinase 1) belongs to Pak family and it was first identified as an interaction partner of Rac1 [244], being one of the most important downstream effectors of Rac1. Paks are serine/threonine kinases interacting with multiple partners to regulate essential cellular processes, such as cytoskeletal dynamics and actin depolymerization, cell polarity and motility, cell growth signaling and transformation, and cell death and survival signaling [245].

There are four main functional sites in Rac1 protein: Insert region, p-loop, and Switch regions (Switch I and Switch II) (Figure 10.1). Insert region (residues 124-135) is specific to Rho subfamily and regulates mitogenesis and apoptosis [246]. p-loop (res. 10-17) functions in binding to the phosphate groups of the nucleotide bound to Rac1. Switch I region (res. 26-45), also called the "effector region", is respon-

sible for the interaction of Rac1 with downstream effectors (e.g. PAK1 and MLK3). Whereas, Switch II region (res. 59-74) is important for the binding of GEFs or proteins that activate Rac1 in its GTP-bound state [246]. During GAP binding, some of the predominant contacts on Rho family GTPases were detected as Val36 and Phe37 on Switch I and Asp63, Tyr64 and Leu67 on Switch II [247]. Hence, switch regions are crucial for the activity of Rac1 protein, except interactions with the membrane [246]. GTP/GDP controls the positioning of Switch I and II [240] by binding to the pocket between p-loop, Switch I (residues 28, 32, 34, 35), Switch II (residue 60), Helix 6 (H6, residues 116, 118 and 119) and residues 158-160. The conformation of switch loops in the GDP-bound form changes dramatically upon GTP binding; as hydrogen bonds are formed between the γ-phosphate and residues Thr35 and Gly60 [247]. The affinity of the protein for nucleotide binding is enhanced via the coordination of the main chain carbonyl of Thr35 both with the Mg^{2+} ion and with the γ-phosphate [247]. New interactions introduced with GTP binding thus changes the dynamics and structure of the switch loops, which are unstructured or highly flexible otherwise, so that Switch I gets reoriented and can interact with effector proteins [247].

From the known complex structures of Rac1-like proteins and PAK1: Rac3-PAK1 (CRIB domain) complex structure with PDB ID: 2qme; and Cdc42-PAK1 (GTPase binding domain) complex structure with PDB ID: 1e0a, it is known that PAK1 interacts with Rac mostly through Switch I region and C-terminal residues (residues 165-175) (Figure 10.1) where structurally important regions on Rac3 and Cdc42 are shown and mutant residues analyzed in this study (T17, Q61, and Y72) are labeled. Rac1 binds to the CRIB (Cdc42- and Rac-interactive binding, res. 75-90) domain of its downstream effector PAK1, which is necessary but not sufficient for high-affinity binding, whereas PBD (p21-binding domain - also called GTPase binding domain, res. 67-113) in PAK1 is responsible for the overall high-affnity binding [248-250]. The basal kinase activity of Paks is controlled by the Inhibitory switch (IS) domain (res. 87-136), which has common residues with PBD [250]. In its OFF state, PAK1 forms an asymmetric homodimer through the interaction of N-terminal regulatory domain (auto-regulatory segment, res. 70-149, including PBD/CRIB and IS domain) and C-terminal kinase/catalytic domain (res. 249-545) [249, 250]. Activation of PAK1 takes

place through binding of GTP-bound Rac1 to the CRIB domain in PAK1 causing the inactive dimer to dissociate via unfolding of IS domain. Thus, the kinase domain is unblocked for downstream signaling and the activation loop is released for auto-phosphorylation [250].

Rac1 is shown to have a role in cancer cell migration and invasion [251-253] and regulating tumor metastasis and progression [253-255]. Hence, problems in the regulation of Rac1 might play crucial roles in various steps of cancer such as initiation, progression, invasion, and metastasis [256, 257]. The Q61L mutation on Switch II region of Rac1 keeps the protein always in the GTP-bound active state preventing the GAP-stimulated GTPase activity (i.e. GTP hydrolysis is prevented resulting in no GDP production) so that the mutation is accountable for a constitutively active protein and cells with Q61L mutant Rac1 show increased multinucleation [241]. On the other hand, T17N is a dominant-negative mutation in Rac1 that prevents GTP binding and reduces GDP binding, hence the protein is in either nucleotide free or inactive state [243] causing it to competitively bind to GEFs with higher affinity than wild-type and to block GEF binding and activation of wild-type Rac1 protein [257]. Both dominant-negative and constituvely active mutants have important roles in the relation of Rho GTPase family (of which Rac1 is a member) with oncogenesis [257]. Moreover, COSMIC (Catalogue of Somatic Mutations in Cancer) database [239] lists Y72C mutation on Rac1 as being related to liver carcinoma. However, no publication describing the functional effect of this oncogenic mutation is available. It is thus a good candidate for understanding the functional effect through molecular simulations and single molecule experiments, which is one of main aims of this study.

Being the downstream effectors for several signaling pathways (e.g. ERK, AKT and WNT), Paks are key regulators of cell signaling networks in cancer for their ability to amplify and propagate upstream oncogenic signals, which makes them potential therapeutic targets [259, 260]. Overexpression or mutational activation of Pak isoforms (mostly PAK1) by upstream elements (e.g. Rac1) have oncogenic signaling effects in cell proliferation, survival, invasion and metastasis [259]. Although there have been drugs targeting several oncogenic protein kinases successfully, new targets are

needed due to the drug resistance developed. Hence, Rac1-PAK1 interaction is a good target for studying the effects of allosteric Rac1 mutations on the interaction with PAK1 and developing inhibitors for the cancer treatment. This study aims to enlighten the intrinsic changes in the dynamics caused by mutations on Rac1 structure through molecular simulations as well as verifying the allosteric effects of constitutively active Q61L, oncogenic Y72C, and dominant-negative T17N mutations on Rac1-PAK1 binding through single molecule Atomic Force Microscopy (AFM) experiments.

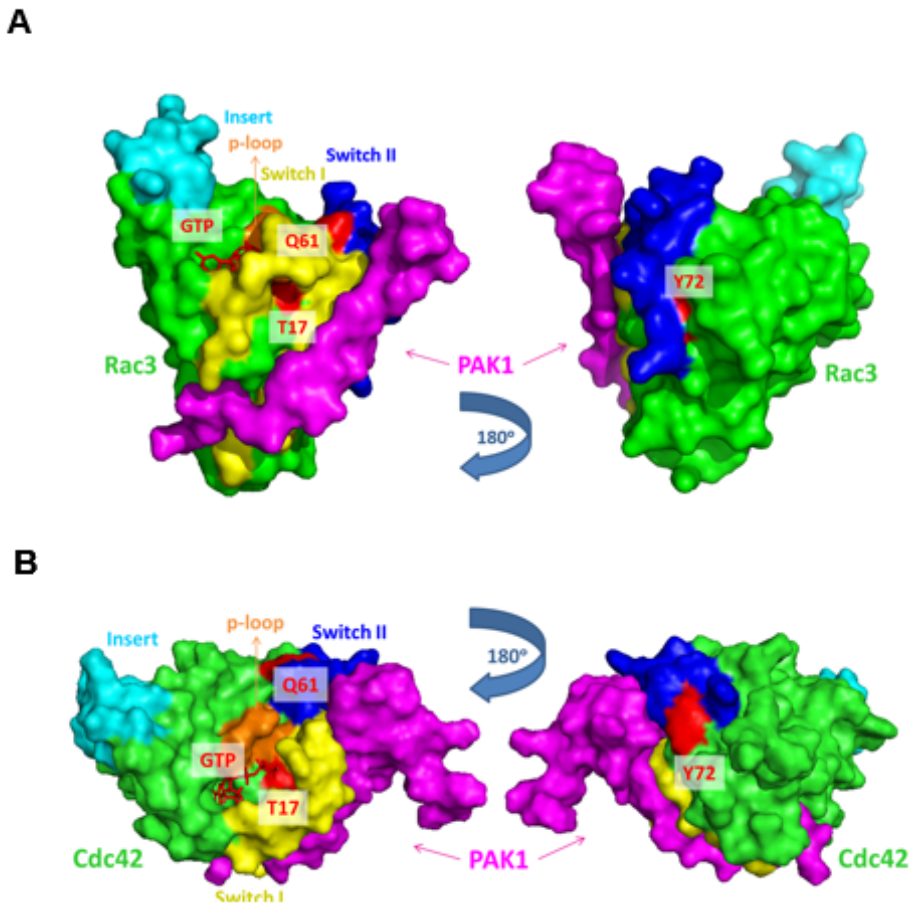


Figure 10.1. Available Structures of Rac1-Like Proteins Bound to PAK1. a) Rac3 - PAK1 (CRIB Domain) Complex Structure (PDB ID: 2qme). b) Cdc42 - PAK1 (GTPase Binding Domain) Complex Structure (PDB ID: 1e0a).

10.2. Materials and Methods

10.2.1. Gaussian Network Model (GNM)

The GNM [9, 139] is an elastic network model in which a protein is represented as a network of amino acids, where α-carbon atoms are nodes, and edges are the springs (with a uniform force constant γ) combining residues within a cutoff distance (r_{cut}). Details about the theoretical derivation are given in Materials and Methods section of this thesis. In the GNM calculations, rcut was taken as 10 $\tilde{A}$ and the calculations were done via Python 2.7 [260]. During the hinge site analysis, nearby residues with α-carbon distances to hinge residues within 6 $\tilde{A}$ in space were also considered.

10.2.1.1. Molecular-Dynamics (MD) Simulations. Starting structures for molecular - dynamics (MD) simulations were the GTP-bound wild-type Rac1 (PDB ID: 3th5), Q61L mutant Rac1 (PDB ID: 4gzl), Y72C mutant Rac1 and nucleotide free T17N mutant Rac1 (PDB ID: 3b13). To obtain GTP-bound structures, starting coordinates were modified by changing the β-γ bridging nitrogen atom of GNP (a non-hydrolyzable analog of GTP) with oxygen. The in silico mutation of the wild-type Rac1 (PDB ID: 3TH5), Y72C, was created using VMD 1.9.1 [261]. The Q61L mutant Rac1 structure (PDB ID: 4gzl) lacks residues 31 and 48 and they were modeled using the wild-type structure (PDB ID: 3th5). A total of 0.8 μs-long MD simulations, including two parallel runs for each case, was performed for the wild-type and mutant Rac1 structures. The details of simulations are given in Table 10.1.

MD simulations were carried out with the all-atom CHARMM27 force field [193] of NAMD 2.7 [192] using a 2 fs integration time step, with periodic boundary conditions. The temperature was maintained at 310 K with a Langevin damping coefficient of 1 ps^{-1} [195]. The pressure was kept at 101.3 kPa by means of NosÃ©-Hoover Langevin piston pressure control. The SHAKE algorithm was used to restrain the length of bonds involving hydrogen atoms for a time step of 2 fs [262]. The initial crystal structures were immersed in a TIP3P-type water box [196] with at least 10 $\tilde{A}$ of

padding between the solute and the edge of the box. The system was neutralized with Cl- and Na+ ions. A nonbonded cutoff of 12 $\tilde{A}$ was used for all Lennard-Jones interactions with a switching function starting at 10 $\tilde{A}$, and the long-range electrostatics was treated according to the particle-mesh Ewald method [197]. The nonbonded pair list distance was 14 $\tilde{A}$. All systems were energetically minimized by the conjugated gradient method for steric crush and crystal contact removal. The trajectories generated by MD simulations were saved every 10 ps for structural and dynamic properties. The analyses were performed using VMD 1.9.1 [261] and all additional calculations were done using MATLAB version R2015a (The MathWorks, Natick, MA). MD trajectories are then analyzed, including calculations of the Root Mean Square Deviation (RMSD) and Mean Square Fluctuation (MSF) profiles and principal component analysis.

Table 10.1. The Details of the Molecular-Dynamics (MD) Simulations.

Parallel Simulation	Structure (PDB ID)	Total Simulation Length (ns)	Equilibration Time (ns)
2-Jan	Rac1 wt structure (3th5)	100 - 100	4-5
2-Jan	Rac1 T17N Mutant (3b13)	100 - 100	50 - 40
2-Jan	Rac1 Q61L Mutant (4gzl)	100 - 100	(10-85) - 25
2-Jan	Rac1 Y72C Mutant (in silico)	100 - 100	5-10

10.2.1.2. Atomic Force Microscopy (AFM). AFM single molecule pulling experiments were performed to explore the allosteric effects of hinge-site-mutations on the Rac1-PAK1 binding.

10.2.1.3. Preparation of Molecules and AFM Setup. The His-tagged wild-type (wt), GST-tagged T17N, His-tagged Q61L and His-tagged Y72C mutant human Rac1 and GST-tagged human PAK1 (p21 binding domain-PBD) molecules, produced in a bacterial expression system, were purchased from Creative BioMart USA, as lyophilized powder. The powdered form of the molecules were reconstituted and diluted to 5, 1, 1, and 3 μg/ml, respectively, with regular Tris buffer. The Tris Buffer solution (containing 50 mM Tris pH 7.5, 0.5 mM $MgCl_2$ and 50 mM NaCl) was prepared in

liquid form. GTP molecules were purchased as lyophilized powder from Sigma-Aldrich (St. Louis, MO, USA). In the beginning of each experiment, wt and mutant Rac1 molecules were loaded with freshly reconstituted GTP (in deionized (DI) water to have a final concentration of 2 mM) by preparing the following solution: 0.1-1 mg/ml Rac1, 15mM EDTA and 2mM GTP. After 30 minute incubation at room temperature, 60 mM MgCl2 (final) was added to lock Rac1 and GTP molecules.

Polyethylene glycol (PEG) and Glutathione (GSH) attached Si3N4 (silicon nitride) AFM probes, as well as NTA (Nitrilotriacetic acid) and GSH covered mica surfaces were purchased from Novascan Technologies, Inc., USA. The nominal spring constants of the cantilevers used in the experiments are 0.01 and 0.1 N/m. The experiments were performed using a commercial AFM system (Dimension Edge, Bruker Nano, Santa Barbara, CA, USA).

10.2.1.4. Tip and Surface Coating. In all of the pulling experiments, PAK1 was functionalized on AFM tip and wt/mutant Rac1 molecule was immobilized on the functionalized mica surface via physical adsorption. GST is adsorbed to GSH and His is adsorbed to NTA. Thus, GSH and NTA mica surfaces respectively were used to functionalize GST-tagged T17N and His-tagged wild-type (wt), Q61L and Y72C mutant human Rac1 (see Figure 10.2-A).

Wild-type and mutant Rac1 molecules (100 μl from each with 20 μg/ml in Tris) were immobilized on the mica surface through a 30-minute incubation at room temperature. 50 μl of EDTA and GTP (solution with deionized (DI) water) was also added for GTP loading of Rac1 molecules, so that GTP-loaded Rac1 molecules got attached to the mica surface. After the incubation, the mica surface was gently rinsed with DI water to remove unbound/loosely bound molecules. The mica surface with immobilized molecules was kept hydrated via Tris prior to data collection.

Functionalized cantilevers (GSH and PEG attached) were dipped into the 100 μl GST-tagged PAK1 (20 μg/ml in Tris buffer) and incubated for 30 minutes at room

temperature, then gently rinsed two times with DI water to remove unbound/loosely bound PAK1 molecules. PAK1 functionalized AFM tips were then used immediately in force measurements. Figure 10.2-A shows the multi-step functionalization of AFM tip and surface in this experiment.

Multi-step functionalization of AFM tips may introduce measurement of unbinding forces that do not arise from the molecules of interest. To test the existence of such undesired unbinding events, the pulling experiments were performed using: 1) PEG + Gluthation attached tip and NTA covered mica surface, 2) PEG + Gluthation+PAK1 attached tip and NTA covered mica surface, 3) PEG + Gluthation attached tip and NTA+Rac1 covered mica surface. No statistically significant specific adhesion was observed between multi-step functionalization elements (see Figure 10.2-B). Additionally, in order to minimize random errors such as variation in surface coverage, the data was collected at different locations on the surface.

10.2.1.5. Experimental Procedure. Three different sets of AFM pulling experiments were performed to investigate the unbinding force distributions involved in the interactions of Rac1-PAK1 complexes. Figure 10.2-A shows the experimental setup for the Rac1-PAK1 systems. All force measurements were performed in the fluid-phase at room temperature. Three different experiments were conducted to ensure the specificity of measured forces coming from the molecules of interest. Exp.1 is the pulling experiment of PEG+Gluthation attached AFM tip and NTA covered mica surface, Exp.2 is the pulling experiment of PEG+Gluthation+PAK1 attached tip and NTA covered mica surface, Exp.3 is the pulling experiment of PEG+Gluthation attached tip and NTA+Rac1 covered mica surface. An example of force-time data collected during the experiment is also given in Figure 10.2-C. 100 pN unbinding force was measured during presented approach/retract cycle. Extension of the flexible PEG linker is observed as parabolic delay before the unbinding takes place.

Cantilevers were calibrated using thermal noise method [263]. AFM pulling experiments, the multi-functionalized AFM tip coated with p21 binding domain of PAK1

was brought in contact -manipulated by a piezo- with Rac1 adsorbed surface (immobilization and functionalization details are described above). The deflection signal, which gives the position of the AFM cantilever, was recorded during the experiment. An example approach/retract cycle is given in Figure 10.2-C. To obtain the unbinding forces at various loading rates, force-distance data were recorded at different approach/retract velocities -spanning three order of magnitudes- (0.1, 0.5, 1, 2.5, 5, 10, 20, 50, 100 μm/s). At each velocity, we collected statistically significant number of force-distance data for further analysis. The trigger force was fixed to 2 nN.

The unbinding force at a specific loading rate was determined by inspecting the force-extension curve. The effective loading rates were estimated by fitting a linear line to the force-time curve immediately before the rupture point in each approach/retract. The effective loading rates obtained from the experiment were equally quantized in the logarithmic space to get uniform number of most probable rupture force vs logarithm of loading rates relation [172, 173]. Force distributions are plotted as histograms for each loading rate-range- and the most probable rupture force at each loading rate is determined by inspecting the force histograms. The most probable rupture force of the Rac1-PAK1 complex at a given loading rate interval was determined by calculating the median of the force histograms. The most probable rupture force f*, versus the natural logarithm of the loading rate, $\ln(r_f\ )$, was plotted and the Bell parameters (k_{off} and x_β; where k_{off} is the unbinding rate constant and x_o is the distance from the bound state of the activated state), which are characteristic values of the investigated molecular pair, were estimated from the slope, $\frac{k_BT}{x_\beta}$, and the intercept, $\frac{k_BT}{x_\beta}k_{off}$, of the fitting linear curves (See Equation 10.1).

$$f^* = \frac{k_BT}{x_\beta}\left(lnr_f - \ln\frac{k_BT}{x_\beta}k_{off}\right) \tag{10.1}$$

where k_B and T are the Boltzmann constant and the absolute temperature, respectively.

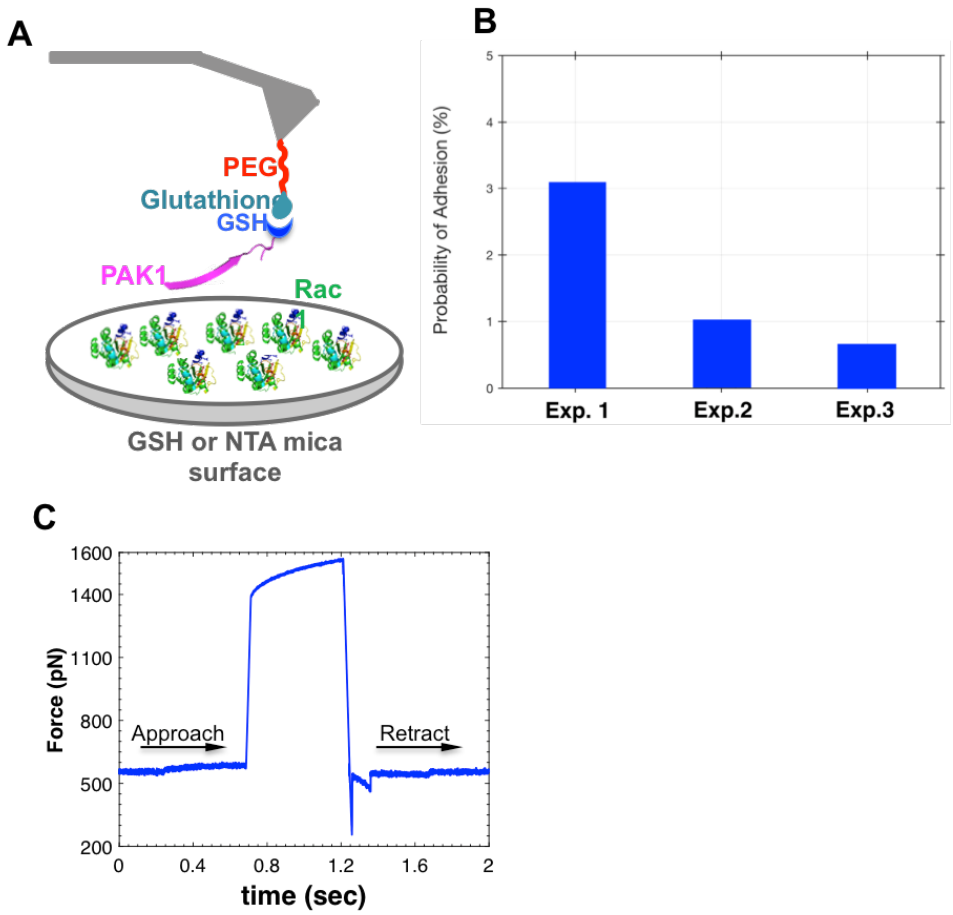


Figure 10.2. The Design of AFM Pulling Experiments. a) Schematic Representation of Experimental Setup. b) The Probability of Adhesion Obtained from Control Experiments of Multi-Step Functionalization. c) An Example of Force-Time Data Collected During the Experiment.

10.3. Results and Discussion

10.3.1. The Relation of Functional Mutations and Hinges on Rac1

Hinge positions are mechanistically key sites of the structure that have significance in mediating cooperative dynamics in the structure. The first step in the computational part was to explore the positions of functionally important mutations in Rac1, such as constitutively active Q61L and oncogenic Y72C mutations on Switch II region and T17N dominant-negative mutation on p-loop with respect to global hinge regions. Using GNM, hinge regions on the wild-type structure of Rac1 (PDB ID: 3th5) for the

first three slowest modes of motion were identified and all these three residues (T17, Q61, and Y72), were found to correspond to hinge or nearby residues (Table 10.2). The mutant structures (PDB IDs: 4gzl, 3b13 and *insilico* Y72C mutation on 3th5) were also investigated to see the changes in the hinge regions upon mutations. When compared to the wild-type structure, residue position 61 on all mutant structures is no more a hinge/nearby residue, whereas positions 17 and 72 correspond to hinge/nearby residues in all cases (Table 10.2). For further dynamic analyses, note that residues 39 and 40 on Switch I are predicted as hinges in the second slowest mode and hence they mainly divide Switch I into two dynamic regions. Similarly, residues 61-65, being predicted as a hinge region in the third slowest mode, divide Switch II into two dynamic regions. Therefore, mainly two opposite behaviors are observed during the cross-correlation analysis of the switch regions (see the next Section).

Table 10.2. Hinge and Nearby Residues Corresponding to Functionally Important Mutations in the First Three Slowest Modes in the Wild-Type (PDB ID: 3th5), Q61L Mutant (PDB ID: 4gzl), T17N Mutant (PDB ID: 3b13) and Y72C Mutant (in silico) Rac1.

	Mode 1 (Nearby)	Mode 2 (Nearby)	Mode 3 (Nearby)
wt Rac1	T17, Q61 (G12)	Y72 (P73)	Q61
Q61L Rac1	T17 (K16)	Y72 (T75)	-
T17N Rac1	N17 (K16)	N17	Y72 (P69)
Y72C Rac1	T17 (K16)	-	C72

10.3.1.1. Molecular Simulation Results. A total of 0.8 μs-long MD simulation, including two parallel runs for each case, is performed for the GTP-bound wild-type Rac1 (PDB ID: 3th5), Q61L mutant Rac1 (PDB ID: 4gzl), Y72C mutant Rac1 and nucleotide free T17N mutant Rac1 (PDB ID: 3b13) to observe the changes in the dynamic behavior of functionally important regions upon mutation (10.3). The focus of this study is on possible allosteric effects of constitutively active Q61L and oncogenic Y72C mutations (Switch II, activation loop) and dominant-negative T17N mutation (p-loop, GTP-binding) on Rac1 Switch I (downstream effector) that is responsible for

the PAK1 binding. MD simulation results are based on the time frames after the equilibration (see Table 10.1), determined using the Root Mean Square Deviation (RMSD) in each run (Figure 10.3). RMSD values of the structures during MD simulations are comparably low with means below 2 $\tilde{A}$, whereas RMSD means of the two parallel runs for the T17N mutant are slightly higher, which can be explained with the lack of a nucleotide bound in the structure. All of the structures seem to equilibrate around 10-20 ns, with the exception of T17N mutant, whose structures equilibrate around 40-50 ns (Table 10.1).

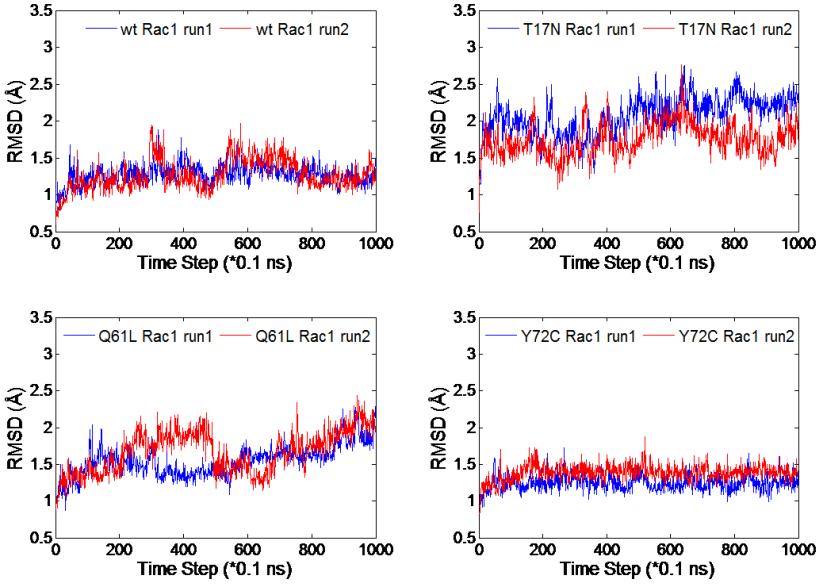


Figure 10.3. RMSD Profiles during Parallel MD Simulations for the Wild-Type and Mutant Rac1 Structures.

10.3.1.2. Dynamic Analyses on Wild-Type and Mutant Rac1 Proteins. MSF profiles and the changes in the correlation of fluctuations of dominant-negative T17N (PDB ID: 3b13), constituvely active Q61L (PDB ID: 4gzl) and oncogenic Y72C mutant (*insilico* mutation) Rac1 proteins with respect to wild-type (wt) Rac1 (PDB ID: 3th5) are analyzed in this part. The average cross correlations of two parallel runs were calculated for each case and the differences of mutants from the wild type were analyzed to detect the changes in the dynamic behavior with the mutations. The main focus is the changes in the MSF values and correlations of important functional regions on Rac1 (p-loop,

Switch I, Switch II, insert domains and C-terminus; shown in Figure 10.1-A and Figure 10.5-A) upon mutation.

10.3.1.3. The Dynamics of T17N Mutation Reveals the Effect of Nucleotide Binding. Nucleotide-free dominant-negative mutant Rac1 (T17N) has increased fluctuations specifically at the nucleotide binding region containing p-loop (residues 10-17) propagating across Switch I (residues 26-45) compared to the wild-type and other Rac1 mutants (Figure 10.4), and this behavior can be attributed to the lack of a bound nucleotide to stabilize those regions [240]. T17N mutation in p-loop also displays a significant decrease in the correlation between the fluctuations of p-loop and a part of the Switch I region (residues 26-38) (Figure 10.5-B). If we combine this observation with the significant increase in the mobility of p-loop and Switch I, we can infer that the lack of a bound-nucleotide in the structure disrupts the coupling of p-loop and Switch I, which has an important role in nucleotide binding, hence regulation of the activation of Rac1. This mutation also allosterically decreases the correlation between Switch I and C-terminus which is important in the PAK1 binding (Figure 10.5-C). The correlations between residues of Switch I and Switch II regions are also allosterically affected as such that the correlations between residues 59-60 and 26-38 decrease; while the correlations between residues 68-72 and 26-38 increase (Figure 10.5-C). Note that residues 39 and 40 on Switch I and residues 61-65 on Switch II are predicted as hinge regions and they mainly divide the loops into two dynamic regions. Another observation is that the highly flexible p-loop starts to positively correlate with insert domain upon T17N mutation (Figure 10.5-B). T17N mutant is known to bind to GEFs competitively with higher affinity than the wild-type [257] and gained correlation of the p-loop-insert domain may have a role in stronger GEF binding.

10.3.1.4. The Mobility of Switch I Significantly Increases Upon Q61L Mutation. The mobility of Switch I (especially between res. 31-35 which are GTP-binding residues close to the interface for Rac1-PAK1 interaction) is increased allosterically upon the Q61L mutation (Figure 10.4). This constitutively active mutation on Switch II is known to prevent GAP-stimulated hydrolysis of GTP and hence the protein stays

in the GTP-bound active form [241]. For the hydrolysis, it is important to fix the position of Thr35, which enhances the affinity of the protein for nucleotide binding via coordinating with the Mg^{2+} ion and γ-phosphate, and Val36 and Phe37 (on Switch I) are among the predominant contacts of GAP proteins on Rac1 [247]. Therefore, the stabilization of Switch I loop is necessary for the GAP binding and this is supported by the MSF results such that the allosteric increase in the mobility of Switch I loop upon Q61L mutation (Figure 10.4) may account for a shift in the Thr35 position disrupting the metal ion coordination and not being able to perform the hydrolysis reaction so that the protein stays in the active form (247). As such, protein staying in the active form in the cell for longer periods of time may lead to stimulated downstream interactions, such as the Rac1 PAK1 interaction, and the mobile Switch I may span the conformation space with an increased probability of capturing a favorable conformation for PAK1 binding. Although Q61L mutation is allosteric for Switch I and p-loop, the structure shows similar behavior to T17N in terms of the decrease in the correlations between Switch I-p-loop and Switch I-C-terminus with respect to the wild-type (Figure 10.5-D). This might be due to the increased mobility of Switch I upon both T17N and Q61L mutations (Figure 10.4). Although T17N mutant is nucleotide-free and Q61L mutant is GTP-bound, the alteration in the correlations of the two switch domains is similar and two opposite behaviors are observed due to existence of hinge regions (Figure 10.5-D and E). Oncogenic Y72C mutation, on the other hand, does not have such a significant effect on the residue fluctuations in Switch I and on the correlations of p-loop and C-terminus with Switch I (Figure 10.5-G) but rather Switch regions are more positively correlated than the T17N and Q61L mutants and they act as a whole (Figure 10.5-G and I).

10.3.1.5. Switch II Significantly Rigidifies in Y72C and Q61L Mutants. Oncogenic Y 72 C and constituvely active Q 61 L mutations have a significant rigidifying effect on Switch II (activation region) with respect to the wild-type, whereas T17N also slightly decreases the mobility in Switch II (Figure 10.4). This rigidity may lead to a more stable and already entropically penalized binding site for the activators of Rac1 (such as GEFs) and hence may give rise to over activated Rac1 mutants. An increase in

the correlation of Switch II region (especially residue 70) with GTP-binding residues 158-160 can be observed upon Q61L mutation, which is directly in the affected area (Figure 10.5-F). Considering that Switch II loop becomes stabilized upon Q61L mutation, it makes sense that constitutively active mutant has higher correlations between the activation domain (Switch II) and GTP binding residues, which are responsible for activation through GEF and GTP binding, respectively. Additionally, the most significantly increasing correlations upon Q61L mutation seem to be between GTP binding residues-Helix 7 of insert domain and between GTP binding residues-Switch II (Figure 10.5-E and F). Helix 6 (H6) embodies GTP-binding residues 116, 118 and 119 and their correlations with Helix 7 (H7), which is part of the insert loop, is one example where Q61L allosterically relates GTP binding and insert domain by altering dynamics with respect to the wild-type (Figure 10.5-E). Y72C on Switch II (activation loop), significantly increases the correlation of Switch II with the p-loop, which is responsible for nucleotide binding (Figure 10.5-H and I). Combining this observation with the MSF results showing the rigidifying effect of Y72C mutation on Switch II, we can infer that a more stable and already entropically penalized binding site for the activators of Rac1 (such as GEFs) coordinating with the nucleotide binding site may give rise to over activated Rac1 mutants leading to stimulated downstream interactions, such as Rac1 - PAK1.

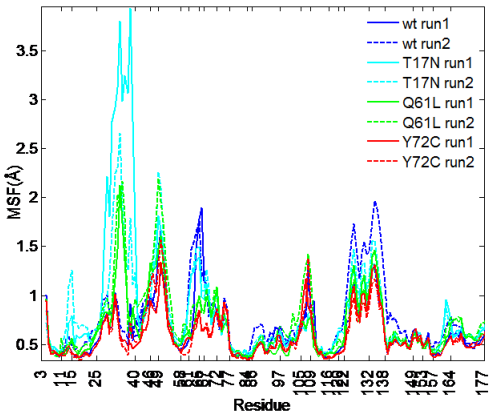


Figure 10.4. MSF Profiles during Parallel MD Simulations for the Wild-Type and Mutant Rac1 Structures. Functional Regions on Rac1: p-loop (res. 10-17, GTP-binding), Switch I (res. 26-45, Downstream Effector), Switch II (Res. 59-74, Activation Loop) and Insert (res. 124-135).

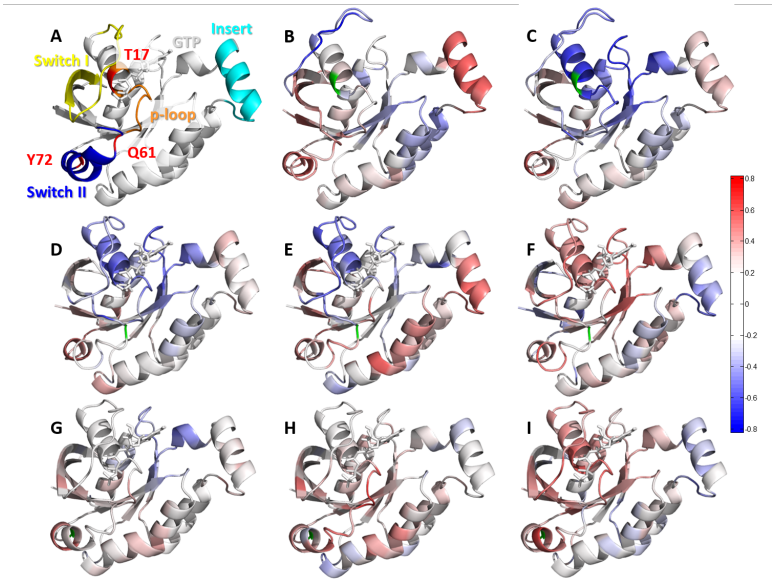


Figure 10.5. a) Functional Regions and Mutations on Wild-Type Rac1. Average PCA Cross Correlations at P-Loop and Switch I for T17N (b-c). P-Loop, Switch I, and Switch II (d-f) for Q61L and (g-i) for GTP-Bound Y72C Mutant Rac1 (*in silico*).

10.3.1.6. Rac1-PAK1 Association can be characterized via AFM. AFM dynamic force spectroscopy (DFS) is a powerful single-molecule approach to measure the binding behavior of two interacting molecules [265-268]. Here, we have conducted AFM pulling experiments to measure the magnitude of binding forces as well as to determine the force distribution/localization and the binding dynamics of Rac1-PAK1, and the effects of Rac1 mutations on these interactions. In AFM pulling experiments, GST-tagged PAK1 GTPase binding domain is functionalized on the AFM cantilever and His-tagged and GST-tagged Rac1 proteins are immobilized respectively on NTA and GSH covered mica surfaces (See Figure 10.2-A).

The adhesion forces between PAK1 and Rac1 were extracted from the collected deflection-distance data. The probabilities of adhesion (defined as the ratio of the number of force curves showing specific unbinding events to the total number of force curves) are 19.1, 6.4, 17.6, and 18.7% for the wild-type, T17N, Q61L, and Y72C Rac1, respectively, as illustrated in Figure 10.6. The negative mutant (T17N) shows almost three-fold decrease in the probability of adhesion confirming the reliability of experimental results as a control experiment. This indicates that we are measuring

specifically the unbinding of Rac1-PAK1 in the experiments.

10.3.1.7. Wild-type Rac1-PAK1 Complex Dissociate Alternative Pathways. The interaction of the wild-type Rac1 loaded with GTP with the p21 binding domain of PAK1 was first explored. The force distributions (with normalized probabilities) at different loading rates (LR) obtained during the unbinding process of PAK1 from the wild-type Rac1 is given in Figure 10.7-A. The dissociation of the complex reflects unimodal distribution at lower loading rates but a bimodal distribution at higher loading rates (a low-strength state which is observed at all loading rates, dissociates at lower forces; and a high-strength state, which emerges as loading rate increases, dissociates at higher forces). This bimodality indicates the existence of two distinct populations of the molecular complex in the ensemble of conformation (existence of two different conformational isomers or/and different dynamic behavior) where each of them has its own force distribution. Moreover, the difference in the probabilities of these two mean forces could be associated with relative probabilities of different populations as such that the low-strength state still has greater fraction compared to high-strength state (at all loading rate intervals that give bimodal force distribution). Presence of two states (low and high-strength states) in the unbinding reaction of the wild-type Rac1-PAK1 complex supports the observation of two distinct behaviors in the parallel MD simulations of wild-type Rac1.

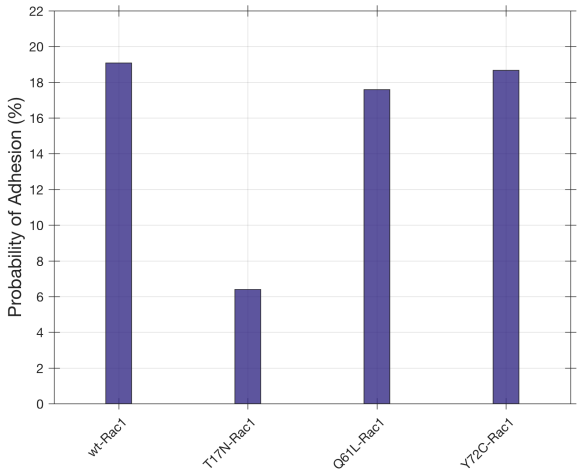


Figure 10.6. The Probability of Adhesion of Wild-Type, T17N, Q61L and Y72C Rac1-PAK1 Observed in the Experiments.

10.3.1.8. T17N Rac1 forms weak association with PAK1. Next, interaction between the GTP loaded negative mutant (T17N) Rac1 and PAK1 was explored. T17N mutation significantly reduces the probability of Rac1-PAK1 association as well as measured rupture forces (see Figure 10.7-B). The distribution of rupture forces along the dissociation of T17NRac1-PAK1 complex were unimodal in all measured loading rates indicating that the complex exhibits a single weak bound state. This loss of binding capability could be associated with the observed high mobility in the p-loop and switch I, where switch I is the PAK1 binding region with the T17N mutation as observed in MD simulations.

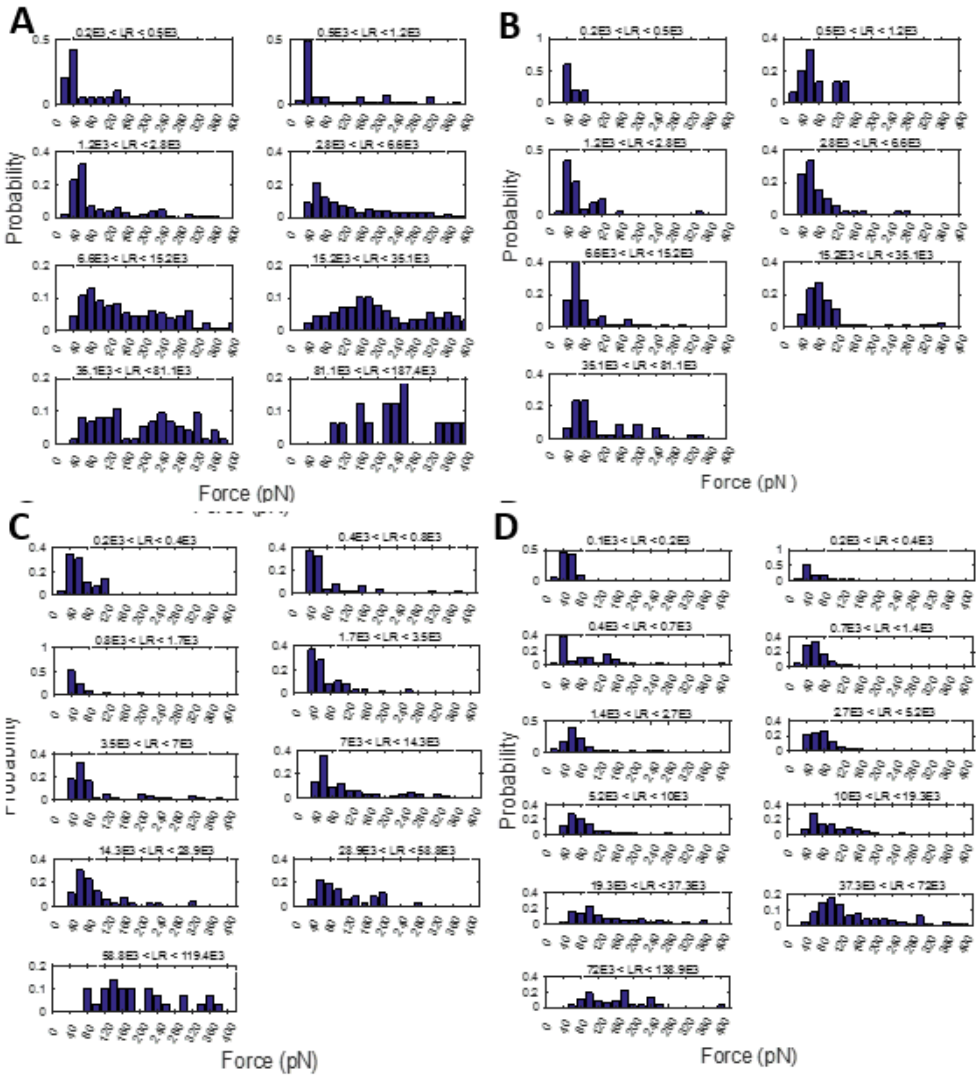


Figure 10.7. Adhesion Force Histograms at Different Loading Rates GTP Loaded Wild-Type T17N, Q61L and Y72C Rac1-PAK1 Complex are given in A-D, Respectively.

10.3.1.9. Mutations Disclose a New Binding State in Rac1. The Q61L mutation leads to a constitutively active state in Rac1, although the Q61R mutation is shown to be oncogenic. On the other hand, Q61L in Ras is an oncogenic mutation. Y72C is known as another oncogenic mutation in Rac1. Both of these mutations are allosteric sites and were seen to be associated with the global hinge sites of Rac1 We examined the effect of these mutations on Rac1-PAK1 binding with respect to the wild type behavior using DFS.

Both Q61L and Y72C mutations give rise to unimodal distribution of rupture forces in all loading rates (see Figure 10.7Figure 10.7-C and D). The extracted most probable rupture forces for the mutants at loading rates higher than 104 pN/s are even lower than the ones observed at the low-strength dissociation state of the wild-type Rac1-PAK1 (see Figure 10.8). This result indicates that both constitutively active and oncogenic mutations induce a shift to a new population in the ensemble of conformations. The differentiation in the correlation network of residue fluctuations, specifically between Switch I, Switch II and p-loop, by the MD simulations of the wild-type and mutant Rac1 is consistent with the emergence of a new state via mutations.

10.3.1.10. Effect of Mutations on the Binding Free Energy Landscape of the Complex. The most probable rupture force against the natural logarithm of loading rates for the low-strength state of the wild-type (shown with straight line in Figure 10.9-A), constitutively active Q61L (shown with straight line in Figure 10.9-B), and oncogenic Y72C (shown with straight line in Figure 10.9-C) Rac1-PAK1 leads to two distinct linear relations, whereas high-strength state of the wild-type Rac1-PAK1 has a single linear relation (shown with dashed line in Figure 10.9-A) within the window of measured loading rates. The rupture force of the low-strength state of the wild-type, Q61L, and Y72C Rac1-PAK1 complexes showed an initial gradual increase, followed by a more rapid increase with increasing loading rates. This result indicates that the low-strength state of the wild-type, Q61L and Y72C Rac1-PAK1 complexes involves at least two transition states along the dissociation reaction pathway and thus one relatively stable intermediate complex structure. Here, since the reaction coordinate is

the distance between molecular pairs, the distance between the chains of the complex is higher at the intermediate state compared to the bound state. These mutations, Q61L and Y72C, are not at the binding interface of the molecules, which means they allosterically control the binding affinity of the complex. Their allosteric affect could be the disruption or addition non-bonded / hydrogen bonds that defines a differentially different bound state in between bound and unbound forms. In other words a redistribution of contacts along changes in the dynamics.

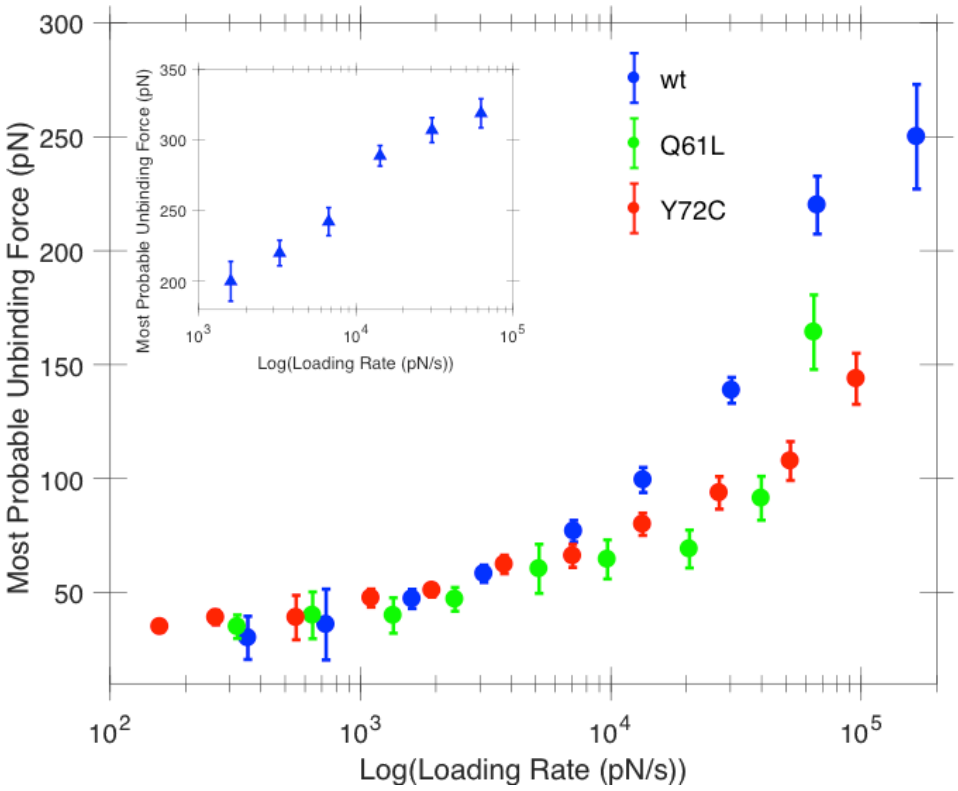


Figure 10.8. Observed Most Probable Rupture Forces Along the Measured Loading Rates.

The observed rupture force vs loading rate relation can also be examined in the context of Bell and Evans model [173] which predicts a linear relationship between the rupture force and the natural logarithm of the loading rates. The two distinct linear relation in the most probable rupture forces against the natural logarithm of the loading rates corresponds two activation energy barrier; an inner barrier and an outer barrier corresponding to larger and smaller rupture forces, respectively [172], [173].

The Bell's parameters of the transitions of the wild-type, Q61L and Y72C Rac1-PAK1 complexes were determined from the linear relations and tabulated in Table 10.3 where blue: wild-type Rac1-PAK1 complex green: Q61L Rac1-PAK1 complex red: Y72C Rac1-PAK1 complex). The disassociation rate (k_{off}) is a function of the activation energy (ΔE) according to the transition state theory. The difference between activation energies of the two systems, system 1 and 2, ($\Delta\Delta E = \Delta E_2 - \Delta E_2$) can be calculated from the ratio of the disassociation rates (i.e. k_2/k_1=exp $-\Delta\Delta E/k_B T$). The calculated difference energies of the Q61L and Y72C Rac1-PAK1 pair relative to the wild-type Rac1-PAK1 are also tabulated in Table 10.3 and labeled on free energy landscape in Figure 10.9-D. Our analysis reveal that the enhanced duration of activity of Q61L Rac1 may be attributed to a larger activation energy barrier $1.5k_B$ T at the rate-limiting transition. The Y72C mutation further increases the barrier height of the rate-limiting transition with $2.4k_B$ T.

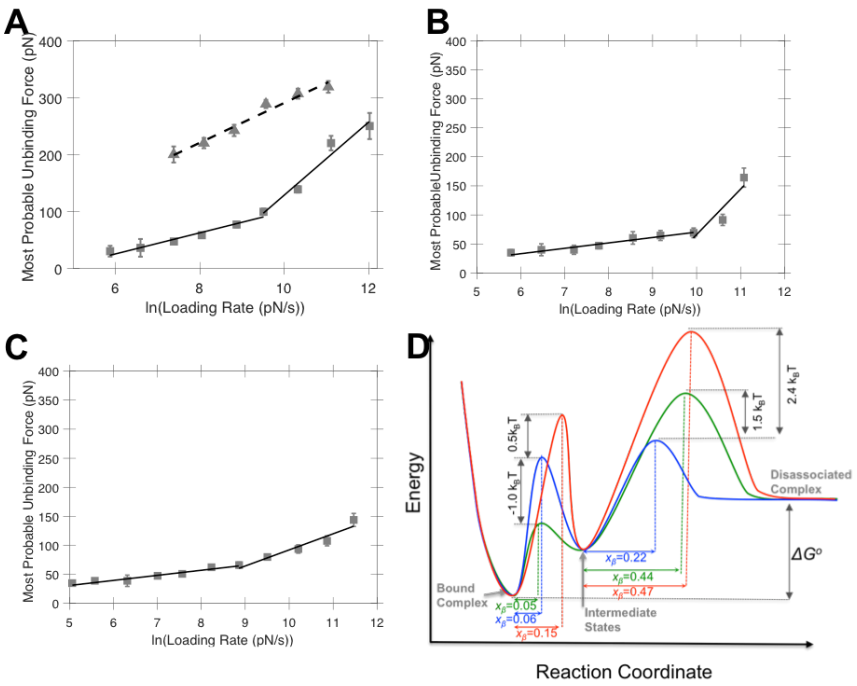


Figure 10.9. The Dynamic Rupture Force Spectra and the Corresponding Binding Energy Landscape of Rac1-PAK1 Complexes. Most Probable Rupture Forces of a) Wild-Type b) Q61L c) Y72C Rac1-PAK1 Complexes. d) A Smooth Diagram of the Binding Energy Landscape.

Table 10.3. Bell Model Parameters and Difference in the Activation Energy of Rac1-PAK1 Complexes.

Ligand-Receptor	Loading rate range. (x10^3 pN/s)	x_β (nm)	$k_{off}(s^{-1})$	$\Delta\Delta E$(kBT)*
Wild type Rac1-PAK1	0.2-190	0.11	8.45	
Q61L Rac1-PAK1	0.2-30	0.44	1.22	1.93
	30-120	0.05	120.98	-2.66
Y72C Rac1-PAK1	0.1-10	0.47	0.49	2.84
	10-140	0.15	28.54	-1.22
* $\Delta\Delta E$ is relative to wild type Rac1-PAK1 binding energy				

10.3.1.11. Longer Duration in Bound State makes Mutation Oncogenic. Although the dissociation of the wild-type Rac1-PAK1 complex requires higher forces to separate the complex even at the low-strength state than both constitutively active and oncogenic mutations (see Figure 10.8), DFS of both Q61L and Y72C mutations resulted in lower dissociation constant for the rate-limiting transition. The dissociation of the Y72C Rac1-PAK1 complex is even slower than the Q61L Rac1-PAK1 complex. Although both barriers of Y72C are higher than both barriers of Q61L, the energy difference in the outer activation energy barrier, which is the rate-limiting step for both mutants, is higher than the inner barrier. This difference makes the probability of the intermediate state in the ensemble of conformations of Y72C higher than the probability of the intermediate state in the ensemble of conformations of Q61L.

10.3.1.12. Dissociation Rate of Mutants Higher Force Sensitivity. Another parameter that characterizes the dissociation reaction of the Rac1-PAK1 complex is x_β, which is the separation between the transition state and the closer stable state (the width of the barrier in the energy landscape). The calculated x_β values for both wild-type and mutants of Rac1-PAK1 dissociation are listed in Table 10.3 and projected onto the free energy landscape on Figure 10.9-D. The wild-type Rac1-PAK1 complex has smaller barrier width for the rate limiting transition in the low-strength state compared to

the barrier width of the rate-limiting transition of both mutants of the Rac1-PAK1 complex, meaning that the wild type has narrower activation energy barrier and therefore the effect of external pulling force on the force dependent dissociation constant is smaller. Both Q61L and Y72C mutants have wider barrier width corresponding to higher force sensitivity of the complex to an external force. Moreover, at lower loading rates corresponding to outer energy barrier, the dissociation kinetics of both mutants are different in terms of barrier height but nearly identical in terms of the barrier width. However, the width of the inner barrier in the Q61L Rac1-PAK1 dissociation reaction is smaller than the Y72C mutant, where the height on the inner barrier is lower in case of the Q61L mutation. These differences in the dissociation kinetics mean a lower but steeper inner barrier in the case of the Q61L mutation.

11. CONCLUSION

Interaction between two biomolecules such as proteins, DNA, RNA, drug molecules is the key phenomena in many biological processes. They recognize and bind each other in the cell environment to conduct their function. In a standard point of view this recognition and binding are determined via their structures. However, in natural mechanism binding of two or more than two molecules requires some conformational and dynamical changes.. It is thus essential to disclose the underlying dynamics and dynamic determinants of these complex systems and their functions to understand how these molecules act to function.

The motivation of this thesis has been develop a novel integrated computational and experimental approach to understand the allosteric control of large conformational transitions and molecular recognition in proteins. In the first part of this thesis, a novel hybrid methodology is called CM-BexMetaD was developed for the sampling of rare events in the conformational transitions. In the second part, mutation induced allosteric control of binding mechanism of proteins were studied by a sequence of studies involving computational and experimental means.

11.1. Enhancing Conformational Sampling via Collective Modes of Motion

Timescale of computer simulations significantly increased with the recent development in both dedicated machines and distributed computing protocols (parallel computing). However, most of the biologically relevant conformational transitions are still not accessible by conventional all-atom simulations methods. The dynamic behavior of biomolecules is characterized by a large number of molecular interactions that makes the underlying FEL extremely rough and complex. However, in order to understand the conformational transitions of biomolecules, one has to characterize the FEL of the dynamic nature of molecules. Enhanced sampling methods are promising solutions developed to overcome the limitations of conventional all-atom simulation protocols. Metadynamics is one of the enhanced sampling methods that are able to

provide feasible solutions for overcoming high energy barriers separating the free energy basins of locally stable conformational states.

Here, a hybrid methodology, which combines ANM with the metadynamics simulations, is developed to sample the large conformational transition of molecules and to deduce the underlying FEL, called CM-BexMetaD. The methodology consists application of ANM on an energetically minimized initial conformation, and extraction of global modes. Then, the dynamics of protein is biased with the selected number of slowest modes of motion (as collective variables) with metadynamics simulation. Three different protocols were applied to select the slowest modes that will be used as CVs. First, only slowest ten modes were selected but the directions of the eigenvectors were corrected according to the difference vector of the initial and target states. Second, only the slowest modes overlapping the distance vector of the initial and target states were used to bias the simulation. Third, only slowest five modes were selected but the simulation was biased in both directions of the eigenvectors corresponding the slowest modes. The third protocol requires only the initial structure of the protein, which is in advantage if the target state is not known. Here, bias-exchange well-tempered metadynamics approach is employed to be able to use relatively higher number of CVs and to provide better sampling of conformational space with the stochastic combination of the collective modes. The success of the method could be amplified if one has any experimental data such as FRET distances, topographic images obtained via AFM, NMR chemical shift data, or SAXS data.

The developed hybrid methodology is applied to sample the open to closed conformational transition of a well-studied protein AdK in apo state. Although fully-closed state could not be sampled within the simulation time windows, some of the transient conformation and semi-closed states were sampled even at a very short simulation time scale. The reason for the failure in sampling the fully-closed conformation might be the requirement of the substrate to overcome the high energy barrier enclosing the fully-closed conformation or insufficient number of CVs to enhance the transition. The proposed protocol also allowed estimating the FEL underlying the conformational transition of AdK and characterizing the energetically favorable sequence of transition of

the enzyme, where closure of LID domain precedes closure of NMP domain.

Hinge sites in the CORE domain of the enzyme allosterically regulate the domain closure mechanism in AdK. Y171 is one of the hinge positions in the CORE domain of the enzyme where both apo and holo structures of the Y171W mutant enzyme was determined by x-ray crystallography. The apo form of Y171W mutant AdK was also simulated using developed methodology to understand the effect of this mutation on the structural dynamics and conformational transition of the enzyme. The results of the mutant simulation indicated that, a perturbation on a hinge position via mutation allosterically affects the domain closure dynamics and energy of the enzyme.

11.2. Allosteric Control of Biomolecular Processes via Hinge Residues

Although there are many studies indicating the allosteric control of biomolecular process, there are limited number studies combining both computational experimental results to elucidate the mechanism of allosteric control. A strong association between the hinge positions of global modes and allosteric mutations that lead to high binding affinity changes was indicated via a large scale statistical analysis over the structural dataset of kinetics and energetics of mutant protein interactions (SKEMPI). The mechanism of allosteric dynamics was demonstrated on ASC protein, human growth hormone (hGH) and pyrin domain (PYD) through MD simulations. The results indicate that mutation on a hinge residue discloses alternative binding modes of the proteins.

Allosteric control of binding have been also elucidated on the kinesin- aÃ-tubilin complex. In this study, we have presented the AFM measurements of the unbinding forces of the wild type, S175A, N332A, and D72N kinesin-tubulin complexes. The observed force values in all types kinesin-tubulin AFM-DFS experiment are in accordance with the myosin-actin unbinding and kinesin-microtubule walking experiments available in the literature [268].

The dissociation rates at zero force and the barrier positions of the dissociation process of the kinesin-tubulin complex were determined by fitting the Bell's model

to the force spectra. According the DFS results of the kinesin-tubulin dissociation reaction, all types of kinesin-tubulin dissociation have two distinct states, so-called low and high strength states at higher loading rates, whereas only low-strength states of the molecules are observed at lower loading rates. S175A and N332A mutations decrease the unbinding forces of the low-strength state complex, whereas have no effect on the unbinding forces of the high-strength state. Despite, D72N kinesin-tubulin complex has similar rupture forces with wild type in both low-strength and high-strength states. Moreover, S175A and N332A decrease the activation energy barrier of the dissociation reaction in both observed states, whereas D72N increases the barriers (the increase is drastic in high-strength state).

Since all these residues are associated with the hinge sites of kinesin, these results indicate the allosteric control of kinesin-tubulin interaction through hinge residues and emphasize the importance of the global motion in binding behavior of kinesin and tubulin molecules.

Allosteric effects of hinge point mutations were also illustrated with Rac1-PAK1 interaction. Three different mutants (T17N, Q61L, and Y72C) and the wild-type Rac1 are used to understand mutation induced functional alteration of Rac1's internal dynamics, which is investigated through MD simulations, and the effects of those mutations on Rac1-PAK1 binding behavior, which is investigated via AFM-DFS experiments. Q61L is on Switch II region that leads to a constitutively active protein, whereas T17N mutation at p-loop (GTP/GDP binding site) leads to either a nucleotide free or inactive protein. Y72 is one of the hinge residues predicted via GNM analysis. Moreover, Y72C mutation is listed as oncogenic mutations in COSMIC database. Therefore, studying Y72C mutation in addition to T17N and Q61L provided better understanding of the effect of oncogenic mutations on the internal dynamics and binding behavior of Rac1.

MSF and cross correlation analysis results agree that there is an allosterically regulated interplay between important functional regions on Rac1. All of these analyzed mutations can substantially alter the intrinsic dynamics of Rac1 with global effects

even in the absence of the downstream binding partner PAK1. Hence, MD simulations reveal the alterations in the intrinsic local and global dynamics of Rac1. Moreover, this allosteric affect of these mutations on the Rac1-PAK1 binding also observed via AFM-DFS measurements as alteration of both binding strength and free energy landscape. It can be concluded that, mutations at a hinge position and/or nucleotide binding site allosterically propagated to the binding interface of Rac1and PAK1 which affects the downstream activation of the pathways.

12. FUTURE WORK

CM-BexMetaD methodology is designed to use with many user-defined parameters. Therefore one proposition for the better sampling of conformational space could be using higher number of replicas. Since the number of replicas is associated with the number of collective variables, this will lead to using higher indexed collective modes, which may provide better sampling of rare events because some of the energy barriers could be associated with the higher frequency intramolecular rearrangements.

The height of the Gaussian hills is directly proportional to the added bias potential, therefore lower hills can be used for more accurate sampling at the price of longer simulation time. Besides, lower hill addition frequency may also provide accurate sampling again at the price of longer simulation time.

Using experimental data as additional CV to the global modes might be a promising improvement to the sampling of the conformational space of molecules. Using distance and/or orientation information obtained from AFM topographic images is one the experimental data that could easily implemented to the algorithm.

REFERENCES

1. Csermely, P., R. Palotai, and R. Nussinov, 2010, "Induced fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events", *Trends Biochem*, Vol. 35, pp. 539-546.

2. McCammon, J.A., B.R. Gelin, and M. Karplus. 1977, "Dynamics of Folded Proteins", *Nature*, Vol. 267, pp. 585-90.

3. Bernardi, R.C., M.C.R. Melo, and K. Schulten. 2015, "Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems, Biochim, Biophys", *Acta - General Subjects*, Vol. 1850, pp. 872-877.

4. Pfaendtner, J., and M. Bonomi, 2015, "Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics", Journal Chemical Theory Computation: 151009124113008.

5. Mitsutake, A., Y. Mori, and Y. Okamoto, 2013, "Enhanced Sampling Algorithms", *Methods Molecular Biology*, Vol. 924, pp. 153-195.

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

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

8. Hinsen, K., A. Thomas, and M.J. Field, 1999, "Analysis of Domain Motions in Large Proteins", *Proteins Structure Functions Genetics*, Vol. 34, pp. 369-382.

9. Haliloglu, T., I. Bahar, and B. Erman, 1997, "Gaussian Dynamics of Folded Pro-

teins", *Physical Review Letters*, Vol. 79, pp. 3090-3093.

10. Kurkcuoglu-Soner, Z. 2015, "Elastic Network Model Based Approaches for Conformer", Generation and Docking Applications.

11. Kessel, A., and N. Ben-Tal, 2011, "Introduction to Proteins Structure, Function and Motion", London: CRC Press, Taylor and Francis Group.

12. Weigelt, J., L.D. McBroom-Cerajewski, M. Schapira, Y. Zhao, and C.H. Arrowmsmith, 2008, "Structural Genomics and Drug Discovery", *All in the Family Current Opininion in Chemical Biology*, Vol. 12, pp. 32-39.

13. Lupas, A.N., and M. Gruber, 2005, "The Structure of Alpha-Helical Coiled Coils", *Advanture Protein Chemical*, Vol. 70, pp. 37-78.

14. Branden, C., and J. Tooze, 1999, "Introduction to Protein Structure", 2nd edition New York, Garland Publishing.

15. Woolfson, D.N, 2005, "The Design of Coiled-Coil Structures and Assemblies", *Advanced Protein Chemical*, Vol. 70, pp. 79-112.

16. Fischer, E. 1894, "Einfluss der Configuration auf die Wirkung der Enzyme", *Berichte der Deutschen Chemical Gesellschaft*, Vol. 27, pp. 2985-2993.

17. Koshland, D.E. 1958, "Application of a Theory of Enzyme Specificity to Protein Synthesis", *Proceedings of the National Acadmy of Sciences*, Vol. 44, pp. 98-104.

18. Koshland, D.E. 1963, "The Role of Flexibility in Enzyme Action", *Cold Spring Harbor Symposia Quantitative Biology*, Vol. 28, pp. 473-482.

19. Monod, J., J. Wyman, and J.P. Changeux, 1965, "Nature of Allosteric Transitions: A Plausible Model", *Journal Molecular Biology*, Vol. 12, pp. 88-118.

20. Tsai, C.J., C.J. Tsai, S. Kumar, S. Kumar, B. Ma, B. Ma, R. Nussinov and R.

Nussinov, 1999, "Folding Funnels", *Binding Funnels and Protein Function Protein Science*, Vol. 8, pp. 1181-1190.

21. Bucher, D., B.J. Grant, and J.A. McCammon, 2011, "Induced fit or Conformational Selection the Role of the Semi-Closed State in the Maltose Binding Protein", *Biochemistry*, Vol. 50, pp. 10530-10539.

22. Gerstein, M., and N. Echols, 2004, "Exploring the Range of Protein Flexibility", *From a Structural Proteomics Perspective Current Opinion Chemical Biolology*, Vol. 8, pp. 14-19.

23. Henzler-Wildman, K., and D. Kern, 2007, "Dynamic Personalities of Proteins", *Nature*, Vol. 450, pp. 964-972.

24. Anson, M.L., and A.E. Mirsky, 1925, "On Some General Properties of Proteins", *The Journal General Physiology*, Vol. 9, pp. 169-179.

25. Mirsky, A.E., L. Paulin, 1936, "On the Ttructure of Native", *Denatured and Coagulated Proteins Proceeding National Academy Sciences*, Vol. 22, pp. 439-447.

26. Anfinsen, C.B., 1973, "Principles that Govern the Folding of Protein Chains", *Science*, Vol. 181, pp. 223-230.

27. Levinthal, C., 1968, "Are There Pathways for Protein Folding", *Journal Chimie Physiqe Physico-Chimie Biology*, Vol. 65, pp. 44-45.

28. Thompson, K.S., C.R. Vinson and E. Freire, 1993, "Thermodynamic Characterization of the Structural Stability of the Coiled-Coil Region of the bZIP Transcription Factor GCN4", *Biochemistry*, Vol. 32, pp. 5491-5496.

29. Bryngelson, J.D., J.N. Onuchic, N.D. Socci, and P.G. Wolynes, 1995, "Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis", *Proteins Structural Function Genetics*, Vol. 21, pp. 167-195.

30. Dobson, C.M., A. Sali, and M. Karplus, 1998, "Protein Folding: A Perspective from theory and Experiment", *Angew, Chemie - International Edition*, Vol. 37, pp. 868-893.

31. Dill, K.A., 1985, "Theory for the Folding and Stability of Globular Proteins", *Biochemistry*, Vol. 24, pp. 1501-1509.

32. Brooks III, C.L., J.N. Onuchic and D.J. Wales, 2001, "Taking a Walk on a Landscape", *Science*, Vol. 293, pp. 612-613.

33. Pauling, L., 1940, "A Theory of the Structure and Process of Formation of Antibodies", *Journal of the American Chemical Society*, Vol. 62, pp. 2643-2657.

34. Cui, Q., and M. Karplus, 2008, "Allostery and Cooperativity Revisited", *Protein Scienty*, Vol. 17, pp. 1295-1307.

35. Kitao, A., S. Hayward and N. Go, 1998, "Energy Landscape of a Native Protein: Jumping-Among-Minima Model", *Proteins Structural Function Genetics*, Vol. 33, pp. 496-517.

36. Elber, R., and M. Karplus, 1987, "Multiple Conformational States of Proteins: A Molecular Dynamics Analysis of Myoglobin", *Science*, Vol. 235, pp. 318-321.

37. Austin, R.H., K.W. Beeson, L. Eisenstein, H. Frauenfelder, and I.C. Gunsalus, 1975, "Dynamics of Ligand Binding to Myoglobin", *Biochemistry*, Vol. 14, pp. 5355-5373.

38. Motlagh, H.N., J.O. Wrabl, J. Li, and V.J. Hilser, 2014, "The Ensemble Nature of Allostery", *Nature*, Vol. 508, pp. 331-339.

39. Tsai, C.J., and R. Nussinov, 2014, "A Unified View of "How Allostery Works", *PLoS Compution Biology*, Vol. 10, pp. 101-106.

40. McLeish, T.C.B., M.J. Cann, and T.L. Rodgers, 2015, "Dynamic Transmission of

Protein Allostery without Structural Change: Spatial Pathways or Global Modes", *Biophysical Journal*, Vol. 109, pp. 1-11.

41. Koshland, D.E., G. NÃ©methy, and D. Filmer, 1966, "Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits", *Biochemistry*, Vol. 5, pp. 365-385.

42. Frauenfelder, H., S. Sligar, and P. Wolynes, 1991, "The Energy Landscapes and Motions of Proteins, Science, Vol. 254, pp. 1598-1603.

43. Gunasekaran, K., B. Ma, and R. Nussinov, 2004, "Is Allostery an Intrinsic Property of all Dynamic Proteins Proteins Structual Function", *Genetion*, Vol. 57, pp. 433-443.

44. Kumar, S., B. Ma, C.J. Tsai, N. Sinha, and R. Nussinov, 2000, "Folding and Binding Cascades: Dynamic Landscapes and Population Shiftsion Protein", *Science*, Vol. 9, pp. 10-19.

45. Popovych, N., S. Sun, R.H. Ebright, and C.G. Kalodimos, 2006, "Dynamically Driven Protein Allostery Nature Structural Molecular", *Biology*, Vol. 13, pp. 831-838.

46. Daily, M.D., and J.J. Gray, 2007, "Local Motions in a Benchmark of Allosteric Proteins", *Proteins Structure Function Genetic*, Vol. 67, pp. 385-399.

47. Cooper, A., and D.T.F. Dryden, 1984, "Allostery without Conformational Change - A Plausible Model", *European Biophysics Journal*, Vol. 11, pp. 103-109.

48. Nussinov, R., and C.-J. Tsai, 2015, "Allostery without a Conformational Change Revisiting the paradigm Currently Opinion in Structural", *Biology*, Vol. 30, pp. 17-24.

49. Linder, T., B.L. de Groot, and A. Stary-Weinzinger, 2013, "Probing the Energy Landscape of Activation Gating of the Bacterial Potassium Channel KcsA", *PLoS*

Compution Biology, Vol. 9, pp. 50-51.

50. Korkmaz, E.N., R. Nussinov, and T. Haliloglu, 2012, "Conformational Control of the Binding of the Transactivation Domain of the MLL Protein and c-Myb to the KIX Domain of CREB", *PLoS Compution Biology*, Vol. 8, pp. 1-9.

51. Aykaç Fas, B., Y. Tutar, T. Haliloglu, and T. Haliloglu, 2013, "Dynamic Fluctuations Provide the Basis of a Conformational Switch Mechanism in Apo Cyclic AMP Receptor Protein", *PLoS Compution Biology*, Vol. 9, pp. e1003141.

52. Sumbul, F., S.E. Acuner-Ozbabacan, and T. Haliloglu, 2015, "Allosteric Dynamic Control of Binding", *Biophysical Journal*, Vol. 109, pp. 1190-1201.

53. Autieri, E., 2011, "Development of Free Energy Calculation Methods for the Study of Monosaccharides Conformation in Computer Simulations".

54. Zhao, G., J.R. Perilla, E.L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A.M. Gronenborn, K. Schulten, C. Aiken, and P. Zhang, 2013, "Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics", *Nature*, Vol. 497, pp. 643-646.

55. Lindorff-Larsen, K., S. Piana, R.O. Dror, and D.E. Shaw, 2011, "How Fast-Folding Proteins Foldition", *Science*, Vol. 334, pp. 517-520.

56. Freddolino, P.L., C.B. Harrison, Y. Liu, and K. Schulten, 2010, "Challenges in Protein-Folding Simulationsion National", *Physical*, Vol. 6, pp. 751-758.

57. Abrams, C., and G. Bussi, 2014, "Enhanced Sampling in Molecular Dynamics Using Metadynamics", *Replica-Exchange and Temperature-Acceleration, Entropy*, Vol. 16, pp. 163-199.

58. Shaw, D.E., J.C. Chao, M.P. Eastwood, J. Gagliardo, J.P. Grossman, C.R. Ho, D.J. Lerardi, I. KolossvÃ¡ry, J.L. Klepeis, T. Layman, C. McLeavey, M.M. Deneroff, M. a. Moraes, R. Mueller, E.C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles,

S.C. Wang, R.O. Dror, J.S. Kuskin, R.H. Larson, J.K. Salmon, C. Young, B. Batson, and K.J. Bowers, 2008, "Anton, A Special-Purpose Machine for Molecular Dynamics Simulation", *Association for Computing Machinery*, Vol. 51, pp. 91-92.

59. Beberg, A.L., D.L. Ensign, G. Jayachandran, S. Khaliq, and V.S. Pande, 2009, Folding@home: Lessons from Eight Years of Volunteer Distributed Computing, 2009, "Institute of Electrical and Electronics Engineers", *International Parallel and Distributed Processing Symposium*, Vol. 1, pp. 1-8.

60. Stone, J.E., J.C. Phillips, P.L. Freddolino, D.J. Hardy, L.G. Trobuco, and K. Schulten, 2007, "Accelerating Molecular Modeling Applications with Graphics Processors", *Journal Compution Chemical*, Vol. 28, pp. 2618-2640.

61. Rader, A.J., 2010, "Coarse-Grained Models: Getting More with Less Current Opinion", *Pharmacol*, Vol. 10, pp. 753-9.

62. Marrink, S.J., H.J. Risselada, S. Yefimov, D.P. Tieleman, and A.H. De Vries, 2007, "The Martini Force Field: Coarse Grained Model for Biomolecular Simulations The Martini Force Field: Coarse Grained Model for Biomolecular Simulations", *Journal Physical Chemical Biology*, Vol. 111, pp. 7812-7824.

63. Hamelberg, D., J. Mongan, and J.A. McCammon, 2004, "Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules", *Jouarnal Chem Physical*, Vol. 120, pp. 11919-11929.

64. Isralewitz, B., M. Gao, and K. Schulten, 2001, "Steered Molecular Dynamics and Mechanical Functions of Proteinsion Current Opinion Structural", *Biology*, Vol. 11, pp. 224-230.

65. Faradjian, A.K., and R. Elber, 2004, "Computing Time Scales from Reaction Coordinates by Milestoning", *The Journal Chemical Physics*, Vol. 120, pp. 10880-10889.

66. Bolhuis, P.G., D. Chandler, C. Dellago, and P.L. Geissler, 2002, "Transition Path

Sampling: Throwing Ropes Over Rough Mountain Passes", *In The Dark Annual Review Physical Chemistry*, Vol. 53, pp. 291-318.

67. Patey, G.N., and J.P. Valleau, 1975, "A Monte Carlo Method for Obtaining the Interionic Potential of Mean Force in Ionic Solution", *Journal Chemistry Physical*, Vol. 63, pp. 2334.

68. Sugita, Y., and Y. Okamoto, 1999, "Replica-Exchange Molecular Dynamics Method for Protein Folding", *Chemistry Physical Letters*, Vol. 314, pp. 141-151.

69. Laio, A., and M. Parrinello, 2002, "Escaping Free-Energy Minimal", *Proceeding of the Nastional Acadmy of Siences*, Vol. 99, pp. 12562-12566.

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

71. Wigner, E., 1938, "The Tansition State Method, Tansactions Faraday", *Societny*, Vol. 1, pp. 29-41.

72. Fiorin, G., M.L. Klein, and J. HÃ©nin, 2013, "Using Collective Variables to Drive Molecular Dynamics Simulations Molecular", *Physics*, Vol. 111, pp. 3345-3362.

73. Korkut, A., and W.A. Hendrickson, 2009, "A Force Field for Virtual Atom Molecular Mechanics of Proteins Proceedings", *National Academy of Sciences*, Vol. 106, pp. 15667-72.

74. De Jong, D.H., G. Singh, W.F.D. Bennett, C. Arnarez, T.A. Wassenaar, L. V. Schfer, X. Periole, D.P. Tieleman, and S.J. Marrink, 2013, "Improved Parameters for the Martini Coarse-Grained Protein force Field", *Journal of Chemical Theaory and Pomputation*, Vol. 9, pp. 687-697.

75. Monticelli, L., S.K. Kandasamy, X. Periole, R.G. Larson, D.P. Tieleman, and S.-J.

Marrink, 2008, "The Martini Coarse-Grained Force Field: Extension to Proteins", *Journal of Chemical Theaory and Pomputation*, Vol. 4, pp. 819-834.

76. Shih, A.Y., A. Arkhipov, P.L. Freddolino, and K. Schulten, 2006, "Coarse Grained Protein-Lipid Model with Application to Lipoprotein Particles", *Journal of Chemical Theaory and Pomputation*, Vol. 110, pp. 3674-3684.

77. Marrink, S.J., H.J. Risselada, S. Yefimov, D.P. Tieleman, and A.H. de Vries, 2007, "The Martini Force Field: Coarse Grained Model for Biomolecular Simulations", *Journal of Chemical Theaory and Pomputation*, Vol. 111, pp. 7812-24.

78. Uusitalo, J.J., H.I. Inglfsson, P. Akhshi, D.P. Tieleman, and S.J. Marrink, 2015, "Martini Coarse-Grained Force Field: Extension to DNA", *Journal of Chemical Theory and Computation*, Vol. 11, pp. 3932-3945.

79. Lo, C.A., A.J. Rzepiela, A.H. De Vries, L. Dijkhuizen, P.H. Hu, S.J. Marrink, C.A. Lopez, A.H. de Vries, P.H. Hunenberger, C.A. LoÂ´pez, P.H. HuÂ¨nenberger, A.J. Rzepiela, A.H. de Vries, L. Dijkhuizen, P.H. Hunenberger, S.J. Marrink, C.A. Lo, A.J. Rzepiela, A.H. De Vries, L. Dijkhuizen, P.H. Hu, and S.J. Marrink, 2009, "Martini Coarse-Grained Force Field: Extension to Carbohydrates", *The Journal of Chemical Theory and Computation*, Vol. 5, pp. 3195-3210.

80. Han, K., X. Zhang, and M. Yang, 2014, "Protein Conformational Dynamics", 1st ed. New York, Springer.

81. Li, C., Y. Zhu, Y. Wang, and G. Chen, 2014, "Molecular Dynamics Simulation on the Conformational Transition of the Mad2 Protein from the open to the closed stateion International", *Journal of Molecular Sciences*, Vol. 15, pp. 5553-5569.

82. Schlitter, J., M. Engels, P. Krüger, E. Jacoby, and A. Wollmer, 1993, "Targeted Molecular Dynamics Simulation of Conformational Change-Application to the T.R Transition in Insulin", *Molecular Simulation*, Vol. 10, pp. 291-308.

83. Schlick, T., 2009, "Molecular Dynamics-Based Approaches for Enhanced Sampling of Long-Time, Large-Scale Conformational Changes in Biomolecules", *F1000 Biology Reports*, Vol. 1, pp. 51.

84. Paquet, E., and H.L. Viktor, 2015, "Molecular Dynamics, Monte Carlo Simulations and Langevin Dynamics: A Computational Review", *BioMed Research International*.

85. Spiwok, V., Z. Sucur, and P. Hosek, 2015, "Enhanced Sampling Techniques in Biomolecular Simulations", *Biotechnology Advances*, Vol. 33, pp. 1130-1140.

86. Sugita, Y., and Y. Okamoto, 1999, "Replica-Exchange Molecular Dynamics Method for Protein Folding", *Chemical Physics Letters*, Vol. 314, pp. 141-151.

87. Meli, M., and G. Colombo, 2013, "A Hamiltonian Replica Exchange Molecular Dynamics Method for the Study of Folding, Based on the Analysis of the Stabilization Determinants of Proteins", *International Journal of Molecular Sciences*, Vol. 14, pp. 12157-12169.

88. Nymeyer, H., 2008, "How Efficient is Replica Exchange Molecular Dynamics An Analytic Approach", *The Journal of Chemical Theory and Computation*, Vol. 4, pp. 626-636.

89. Roe, D.R., C. Bergonzo, and T.E. Cheatham, 2014, "Evaluation of Enhanced Sampling Provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods", *The Journal of Chemical Theory and Computation*, Vol. 118, pp. 3543-3552.

90. Torrie, G.M., and J.P. Valleau, 1977, "Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling", *The Journal of Chemical Theory and Computation*, Vol. 23, pp. 187-199.

91. Maximova, T. *et al.*, 2016, "Principles and Overview of Sampling Methods for

Modeling Macromolecular Structure and Dynamics", *PLOS Computers in Biology*, Vol. 12, pp. e1004619.

92. Kastner, J., 2011, "Umbrella Sampling, WIREs Comput", *Molecular Simulation*, Vol. 1, pp. 932-942.

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

94. Mills, M., and I. Andricioaei, 2008, "An Experimentally Guided Umbrella Sampling Protocol for Biomolecules", *Journal of Chemical Physics*, Vol. 1, pp. 129.

95. Formoso, E., V. Limongelli, and M. Parrinello, 2015, "Energetics and Structural Characterization of the Large-Scale Functional Motion of Adenylate Kinase", *Scientific Reports*, Vol. 5, pp. 8425.

96. Darve, E., and A. Pohorille, 2001, "Calculating Free Energies Using Average Force", *Journal of Chemical Physics*, Vol. 115, pp. 9169-9183.

97. Bussi, G., F.L. Gervasio, A. Laio, and M. Parrinello, 2006, "Free-Energy Landscape for Ã Hairpin Folding from Combined Parallel Tempering and Metadynamics", *Journal of the American Chemical Society*, Vol. 128, pp. 13435-13441.

98. Gervasio, F.L., A. Laio, and M. Parrinello, 2005, "Flexible Docking in Solution Using Metadynamics", *Journal of the American Chemical Society*, Vol. 127, pp. 2600-2607.

99. Spiwok, V., P. Lipovova, and B. Kralova, 2007, "Metadynamics in Essential Coordinates: Free Energy Simulation of Conformational Changes", *The Journal of Physical Chemistry*, Vol. 111, pp. 3073-6.

100. Barducci, A., G. Bussi, and M. Parrinello, 2008, "Well-Tempered Metadynam-

ics: A Smoothly Converging and Tunable Free-Energy Method", *Physical Review Letters*, Vol. 100, pp. 20603.

101. Bonomi, M., and M. Parrinello, 2010, "Enhanced Sampling in the Well-Tempered Ensemble", *Physical Review Letters*, Vol.1, pp, 104.

102. Sutto, L., S. Marsili, and F.L. Gervasio, 2012, "New Advances in Metadynamics", *Wiley Interdiscip Reviews: Computational Molecular Science*, Vol. 2, pp. 771-779.

103. Piana, S., and A. Laio, 2007, "A Bias-Exchange Approach to Protein Folding", *The Journal of Physical Chemistry*, Vol. 111, pp. 4553-9.

104. Bahar, I., T.T.R. Lezon, L.L.-W. Yang, and E. Eyal, 2010, "Global Dynamics of Proteins: Bridging Between Structure and Function", *Annual Review of Biophysics*, Vol. 9, pp. 23-42.

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

106. Hayward, S., and B.L. De Groot, 2008, "Normal Modes and Essential Dynamics", *Methods Molecular Biology*, Vol. 443, pp. 89-106.

107. Knapp, E.W., S.F. Fischer, and F. Parak, 1983, "The Influence of Protein Dynamics on MoÂ¨ssbauer Spectra", *Journal of Chemical Physics*, Vol. 78, pp. 4701.

108. Brooks, B., D. Janezic, M. Karplus, and D. Janezic. 1995. "Harmonic Analysis of Large Systems, I. Methodology", *Journal of Computer*, Vol. 16, pp. 1522-1542.

109. Cui, Q., and I. Bahar, 2005, "Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems".

110. Tama, F., F.X. Gadea, O. Marques, and Y.H. Sanejouand, 2000, "Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules, Pro-

teins Structure", *Function, and Bioinformatics*, Vol. 41, pp. 1-7.

111. Levy, R.M., D. Perahia, and M. Karplus, 1982, "Molecular Dynamics of an Alpha-Helical Polypeptide: Temperature Dependence And Deviation from Harmonic Behavior", *The Proceedings of the National Academy*, Vol. 79, pp. 1346-1350.

112. Li, G., and Q. Cui, 2002, "A Coarse-Grained Normal Mode Approach for Macromolecules: an Efficient Implementation and Application to Ca(2+)-ATPase", *Biophysical Journal*, Vol. 83, pp. 2457-74.

113. Durand, P., G. Trinquier, and Y. Sanejouand, 1994, "A New Approach for determining Low-Frequency Normal Modes, Biopolymers, Vol. 1, pp. 759-771.

114. Doruker, P., A.R. Atilgan, and I. Bahar, 2000, Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to Alpha-Amylase Inhibitor", *Proteins*, Vol. 40, pp. 512-24.

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

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

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

118. Yang, L., G. Song, and R.L. Jernigan, 2007, "How Well Can We Understand Large-Scale Protein Motions Using Normal Modes of Elastic Network Models", *Biophysical Journal*, Vol. 93, pp. 920-9.

119. Wako, H., and S. Endo, 2011, "Ligand-Induced Conformational Change of a Protein Reproduced by a Linear Combination of Displacement Vectors Obtained from

Normal Mode Analysis", B*iophysical Chemistry*, Vol. 159, pp. 257-266.

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

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

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

123. Batista, P.R., C.H. Robert, J.-D. MarÃ©chal, M. Ben Hamida-RebaÃ¯, P.G. Pascutti, P.M. Bisch, and D. Perahia, 2010, "Consensus Modes, A Robust Description of Protein Collective Motions from Multiple-Minima Normal Mode Analysis– Application to the HIV-1 Protease", *Physical Chemistry Chemical Physics*, Vol. 12, pp. 2850-2859.

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

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

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

127. Das, A., M. Gur, M.H. Cheng, S. Jo, I. Bahar, and B.B. Roux, 2014, "Exploring

the Conformational Transitions of Biomolecular Systems Using a Simple Two-State Anisotropic Network Model", *PLoS Computational Biology*, Vol. 10, pp. e1003521.

128. Baron, R., 2013, "Fast Sampling of A-to-B Protein Global Conformational Transitions: From Galileo Galilei to Monte Carlo Anisotropic Network Modeling", *Biophysj*, Vol. 105, pp. 1545-1546.

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

130. Zacharias, M., 2008, "Combining Elastic Network Analysis and Molecular Dynamics Simulations by Hamiltonian Replica Exchange", *Journal of Chemical Theory and Computation*, Vol. 4, pp. 477-487.

131. Ostermeir, K., and M. Zacharias, 2014, "Hamiltonian Replica Exchange Combined with Elastic Network Analysis to Enhance Global Domain Motions in Atomistic Molecular Dynamics Simulations, Proteins Structure", *Function, and Bioinformatics*, Vol. 82, pp. 3410-3419.

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

133. Lopez-Blanco, J.R., J.I. Garzon, and P. Chacon, 2011, "iMod: Multipurpose Normal Mode Analysis in Internal Coordinates", *Bioinformatics*, Vol. 27, pp. 2843-2850.

134. Kuriyan, J., 2004, "Allostery and Coupled Sequence Variation in Nuclear Hormone Receptors", *Cell*, Vol. 116, pp. 354-356.

135. Hines, C.S., K. Ray, J.J. Schmidt, F. Xiong, R.W. Feenstra, M. Pras-raves, J.P. De Moes, J.H.M. Lange, M. Melikishvili, M.G. Fried, P. Mortenson, M. Charlton, Y. Patel, S.M. Courtney, C.G. Kruse, and D.W. Rodgers, 2014, "Allosteric Inhibition of the Neuropeptidase Neurolysin", *Journal of Biological Chemistry*, Vol. 289. pp. 35605-35619.

136. Gerek, Z.N., and S.B. Ozkan, 2011, "Change in Allosteric Network Affects Binding Affinities of PDZ Domains: Analysis Through Perturbation Response Scanning", *PLoS Computational Biology 7*.

137. Nevin Gerek, Z., S. Kumar, and S. Banu Ozkan, 2013, "Structural Dynamics Flexibility Informs Function and Evolution at a Proteome Scale", *Evolutionary Applications*, Vol. 6, pp. 423-433.

138. Ozer, N., A. Ozen, C.A. Schiffer, and T. Haliloglu, 2015, "Drug-Resistant HIV-1 Protease Regains Functional Dynamics Through Cleavage Site Coevolution", *Evolutionary Applications*, Vol. 8, pp. 185-198.

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

140. Ertekin, A., R. Nussinov, and T. Haliloglu, 2006, "Association of Putative Concave Protein-Binding Sites with the Fluctuation Behavior of Residues", *Protein Society*, Vol. 15, pp. 2265-2277.

141. Haliloglu, T., E. Seyrek, and B. Erman, 2008, "Prediction of Binding Sites in receptor-Ligand Complexes with the Gaussian Network Model", *Physical Review Letters*, Vol. 100, pp. 1-4.

142. Ozbek, P., S. Soner, B. Erman, and T. Haliloglu, 2010, "DNABINDPROT: Fluctuation-Based Predictor of DNA-Binding Residues within a Network of Interacting Residues", *Nucleic Acids Resistance*, Vol. 38, pp. 417-423.

143. Sahillioglu, A.C., F. Sumbul, N. Ozoren, and T. Haliloglu, 2014, "Article Structural and Dynamics Aspects of ASC Speck Assembly", *Structural Design*, Vol. 22, pp. 1722-1734.

144. Luque, I., and E. Freire, 2000, "Structural Stability of Binding Sites: Consequences for Binding Affinity and Allosteric Effects Proteins", *Supplementary*, Vol. 4, pp. 63-71.

145. Swapna, L.S., R.M. Bhaskara, J. Sharma, and N. Srinivasan, 2012, "Roles of Residues in the Interface of Transient Protein-Protein Complexes Before Complexation", *Science Reports*, Vol. 2, pp. 1-9.

146. Zhang, X., F. Rico, A.J. Xu, and V.T. Moy, 2009, "Atomic Force Microscopy of Protein-Protein Interactions in: Handbook of Single-Molecule Biophysics", *Springer*, Vol. 1, pp. 555-570.

147. Wennmalm, S., and S.M. Simon, 2007, "Studying Individual Events in Biology", *Annual Review of Biochemistry*, Vol. 76, pp. 419-446.

148. Greenleaf, W.J., M.T. Woodside, and S.M. Block, 2007, "High-Resolution, Single-Molecule Measurements of Biomolecular Motion", *Annual Review of Biophysics Biomolecular Structure and Dynamics*, Vol. 36, pp. 171-90.

149. Binnig, G., C. Quate, and C. Gerber, 1986, "Atomic Force Microscope", *Physical Review Letters*, Vol. 56, pp. 930-933.

150. Barkai, E., Y. Jung, and R. Silbey, 2004, "Theory of Single-Molecule Spectroscopy: Beyond the Ensemble Average", *Annual Review of Physical Chemistry*, Vol. 55, pp. 457-507.

151. Bustamante, C., J.C. Macosko, and G.J. Wuite, 2000, "Grabbing the Cat by the Tail: Manipulating Molecules One by One", *Nature Reviews Molecular Cell Biology*, Vol. 1, pp. 130-136.

152. Eghiaian, F., F. Rico, A. Colom, I. Casuso, and S. Scheuring, 2014, "High-Speed Atomic Force Microscopy: Imaging and Force Spectroscopy", *FEBS Letters*, Vol. 588, pp. 3631-8.

153. Rico, F., L. Gonzalez, I. Casuso, M. Puig-Vidal, and S. Scheuring, 2013, "High-Speed Force Spectroscopy Unfolds Titin at the Velocity of Molecular Dynamics Simulations", *Science*, Vol. 342. pp. 741-3.

154. Rico, F., L. Picas, A. Colom, N. Buzhynskyy, and S. Scheuring, 2013, "The Mechanics of Membrane Proteins is a Signature of Biological Function", *Soft Matter*, Vol. 9, pp. 7866.

155. Rico, F., A. Rigato, L. Picas, and S. Scheuring, 2013, "Mechanics of Proteins with a Focus on Atomic Force Microscopy", *Journal of Nanobiotechnology*, Vol. 11, pp. 1-3.

156. Sevim, S., S. Tolunay, and H. Torun, 2014, "Micromachined Sample Stages to Reduce Thermal Drift in Atomic Force Microscopy", *Microsystem Technologies*, Vol. 21, pp. 1559-1566.

157. Hutter, J.L., and J. Bechhoefer, 1993, "Calibration of Atomic-Force Microscope Tips", *Review of Scientific Instruments*, Vol. 64, pp. 1868-1873.

158. Florin, E.L., V.T. Moy, and H.E. Gaub, 1994, "Adhesion Forces Between Individual Ligand-Receptor Pairs", *Science*, Vol. 264, pp. 415-417.

159. Alcaraz, J., L. Buscemi, M. Puig-De-Morales, J. Colchero, A. Baro, and D. Navajas, 2000, "Correction of Microrheological Measurements of Soft Samples with Atomic Force Microscopy for the Hydrodynamic Drag on the Cantilever".

160. Janovjak, H., J. Struckmeier, and D.J. Müller, 2005, "Hydrodynamic Effects in fast AFM Single-Molecule Force Measurements", *European Biophysics Journal*, Vol. 34, pp. 91-96.

161. Lee, G.U., D. a Kidwell, and R.J. Colton, 1994, "Sensing Discrete Streptavidin-Biotin Interactions with Atomic Force Microscopy", *Langmuir*, Vol. 2, pp. 354-357.

162. Yu, J., Y. Jiang, X. Ma, Y. Lin, and X. Fang, 2007, "Energy Landscape of Aptamer/Protein Complexes Studied by Single-Molecule Force Spectroscopy", *Chem Asian*, Vol. 2, pp. 284-9.

163. Neuert, G., C. Albrecht, E. Pamir, and H.E. Gaub, 2006, "Dynamic Force Spectroscopy of the Digoxigenin-Antibody Complex", *FEBS Letters*, Vol. 580, pp. 505-509.

164. Wojcikiewicz, E.P., M.H. Abdulreda, X. Zhang, and V.T. Moy, 2006, "Force Spectroscopy of LFA-1 and its Ligands, ICAM-1 and ICAM-2", *Biomacromolecules*, Vol. 7, pp. 3188-3195.

165. Maki, T., S. Kidoaki, K. Usui, H. Suzuki, M. Ito, F. Ito, Y. Hayashizaki, and T. Matsuda, 2007, "Dynamic Force Spectroscopy of the Specific Interaction Between the PDZ Domain and its Recognition Peptides", *Langmuir*, Vol. 23, pp. 2668-73.

166. Krasnoslobodtsev, A. V., L.S. Shlyakhtenko, and Y.L. Lyubchenko, 2007, "Probing Interactions within the Synaptic DNA-SfiI Complex by AFM Force Spectroscopy", *Journal Molecular Biology*, Vol. 365, pp. 1407-1416.

167. Merkel, R., P. Nassoy, a Leung, K. Ritchie, and E. Evans, 1999, "Energy Landscapes of Receptor-Ligand Bonds Explored with Dynamic Force Spectroscopy", *Nature*, Vol. 397, pp. 50-53.

168. Nevo, R., C. Stroh, F. Kienberger, D. Kaftan, V. Brumfeld, M. Elbaum, Z. Reich, and P. Hinterdorfer, 2003, "A molecular Switch Between Alternative Conformational States in the Complex of Ran and Importin Beta1", *Nature Structural and Molecular Biology*, Vol. 10, pp. 553-7.

169. Yersin, A., T. Osada, and A. Ikai, 2008, "Exploring Transferrin-Receptor Inter-

actions at the Single-Molecule Level", *Biophysical Journal*, Vol. 94, pp. 230-40.

170. Zhang, X., S.E. Craig, H. Kirby, M.J. Humphries, and V.T. Moy, 1997, "Molecular Basis for the Dynamic Strength of the Integrin a 4 b 1 /VCAM-1 Interaction".

171. Hinterdorfer, P., and Y.F. DufrÃªne, 2006, "Detection and Localization of Single Molecular Recognition Events Using Atomic Force Microscopy", *Nature Methods*, Vol. 3, pp. 347-355.

172. Evans, E., and K. Ritchie, 1997, Dynamic Strength of Molecular Adhesion Bonds, Biophysical Journal, Vol 72, pp. 1541-1555.

173. Bell, G., 1978, "Models for the Specific Adhesion of Cells To Cells", *Science*, Vol. 200, pp. 618-627.

174. Dudko, O.K., G. Hummer, and A. Szabo, 2006, "Intrinsic Rates and Activation Free Energies From Single-Molecule Pulling Experiments", *Physical Review Letters*, Vol. 96, pp. 101-102.

175. Friddle, R.W., 2008, "Unified Model of Dynamic Forced Barrier Crossing in Single Molecules", *Physical Review Letters*, Vol. 100.

176. Husson, J., and F. Pincet, 2008, "Analyzing Single-Bond Experiments: Influence of the Shape of the Energy Landscape and Universal Law Between the width, Depth, and Force Spectrum of the Bond, Physical Review. E, Statistical", *Nonlinear and soft Matter Physics*, Vol. 77, pp. 110-112.

177. Evans, E.A., and D.A. Calderwood, 2007, "Forces and Bond Dynamics in Cell Adhesion", *Science*, Vol. 316, pp. 1148-1154.

178. Marshall, B.T., M. Long, J.W. Piper, T. Yago, R.P. McEver, and C. Zhu, 2003, "Direct Observation of Catch Bonds Involving Cell-Adhesion Molecules", *Nature*, Vol. 423, pp. 190-3.

179. Zhu, C., J. Lou, and R.P. McEver, 2005, "Catch Bonds: Physical Models, Structural Bases", *Biological Function and Rheological Relevance, Biorheology*, Vol. 42, pp. 443-62.

180. Nevo, R., V. Brumfeld, M. Elbaum, P. Hinterdorfer, and Z. Reich, 2004, "Direct Discrimination Between Models of Protein Activation by Single-Molecule Force Measurements", *Biophysical Journal*, Vol. 87, pp. 2630-4.

181. Ritco-Vonsovici, M., A. Ababou, and M. Horton, 2007, "Molecular Plasticity of Beta-Catenin: New Insights from Single-Molecule Measurements and MD Simulation", *Protein Society*, Vol. 16, pp. 1984-1998.

182. Giannetti, A.M., P.M. Snow, O. Zak, and P.J. Björkman, 2003, "Mechanism for Multiple Ligand Recognition by the Human Transferrin Receptor", *PLoS Biol*, Vol. 1, pp 1-3.

183. Bahar, I., a Wallqvist, D.G. Covell, and R.L. Jernigan, 1998, "Correlation Between Native-State Hydrogen Exchange and Cooperative Residue Fluctuations from a Simple Model", *Biochemistry*, Vol. 37, pp. 1067-75.

184. Haliloglu, T., and I. Bahar, 1999, "Structure-Based Analysis of Protein Dynamics: Comparison of Theoretical Results for Hen Lysozyme With X-Ray Diffraction and NMR Relaxation Data", *Proteins*, Vol. 37, pp. 654-67.

185. Leach, A.R., 2001, "Molecular Modelling, Principles and Applications", *Prentice Hall.*

186. Laio, A., A. Rodriguez-Fortea, F.L. Gervasio, M. Ceccarelli, and M. Parrinello, 2005, "Assessing the Accuracy of Metadynamics", *The Journal of Physical Chemistry*, Vol. 109, pp. 6714-6721.

187. Barducci, A., M. Bonomi, and M. Parrinello, 2011, "Metadynamics, Wiley Interdiscip. Reviews: Computational Molecular", *Science*, Vol. 1, pp. 826-843.

188. Spiwok, V., Z. Sucur, and P. Hosek, 2014, "Enhanced sampling techniques in biomolecular simulations", *Biotechnol Advanced*, Vol. 1977, pp. 1-11.

189. Spiwok, V., B. KrÃ¡lovÃ¡, and I. Tvaroaka, 2008, "Continuous Metadynamics in Essential Coordinates as a Tool for free Energy Modelling of Conformational Changes", *Journal of Molecular Modeling*, Vol. 14, pp. 995-1002.

190. Gur, M., E. Zomot, M.H. Cheng, and I. Bahar, 2015, "Energy Landscape of LeuT from Molecular Simulations", *Journal of Chemical Physics*, Vol. 143, pp. 243134.

191. Deriu, M.A., G. Grasso, J.A. Tuszynski, D. Gallo, U. Morbiducci, and A. Danani, 2016, "Josephin Domain Structural Conformations Explored by Metadynamics in Essential Coordinates", *PLOS Computational Biology*, Vol. 12, pp. e1004699.

192. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. KalÃ©, and K. Schulten, 2005, "Scalable Molecular Dynamics with NAMD", *Journal Comput Chemical*, Vol. 26, pp. 1781-1802.

193. Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, 1983, "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations", *Journal of Computational Chemistry*, Vol. 4, pp. 187-217.

194. Huang, J., and A.D. Mackerell, 2013, "CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data", *Journal of Computational Chemistry*, Vol. 34, pp. 2135-2145.

195. Levy, R.M., M. Karplus, and J. Andrew McCammon, 1979, "Diffusive Langevin Dynamics of Model Alkanes", *Chemical Physics Letters*, Vol. 65, pp. 4-11.

196. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, 1983, "Comparison of Simple Potential Functions for Simulating Liquid Water", *Journal of Chemical Physics*, Vol. 79, pp. 926.

197. Darden, T., D. York, and L. Pedersen, 1993, "Particle mesh Ewald: An N log(N) method for Ewald sums in Large Systems", *Journal of Chemical Physics*, Vol. 98, pp. 10089.

198. Mackerell, a D., Jr, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, Iii, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J.W. rkiewicz-Kuczera, D. Yin, and M. Karplus, 1998, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins", *Journal of Physical Chemistry Biophysical*, Vol. 102, pp. 3586-3616.

199. Esteban-MartÃn, S., R.B. Fenwick, J. Aden, B. Cossins, C.W. Bertoncini, V. Guallar, M. Wolf-Watz, and X. Salvatella, 2014, "Correlated Inter-Domain Motions in Adenylate Kinase", *PLoS Computational Biology*, Vol. 10, pp. e1003721.

200. Corrons, J.-L.V., E. Garcia, J.J. Tusell, K.I. Varughese, C. West, and E. Beutler, 2003, "Red Cell Adenylate Kinase Deficiency: Molecular Study of 3 New Mutations (118G¿A, 190G¿A, and GAC deletion) Associated with Hereditary Nonspherocytic Hemolytic Anemia", *Blood*, Vol. 102, pp. 353-6.

201. Hardie, D.G., and S. a Hawley, 2001, "Activated Protein Kinase: the Energy Charge Hypothesis Revisited", *Bioessays*, Vol. 23, pp. 1112-1119.

202. Wolf-Watz, M., V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E.Z. Eisenmesser, and D. Kern, 2004, "Linkage Between Dynamics and Catalysis in a Thermophilic-Mesophilic Enzyme Pair", *Nature Structural Molecular Biology*, Vol. 11, pp. 945-949.

203. Shapiro, Y.E., E. Kahana, and E. Meirovitch, 2009, "Domain Mobility in Proteins from NMR/SRLS", *The Journal of Physical Chemistry*, Vol. 113, pp. 12050-12060.

204. Whitford, P.C., S. Gosavi, and J.N. Onuchic, 2008, "Conformational Transitions

in Adenylate Kinase: Allosteric Communication Reduces Misligation", *Journal of Biological Chemistry*, Vol. 283, pp. 2042-2048.

205. ÃdÃ©n, J., and M. Wolf-Watz, 2007, "NMR Identification of Transient Complexes Critical to Adenylate Kinase Catalysis", *Journal of the American Chemical Society*, Vol. 129, pp. 14003-14012.

206. Bilderback, T., T. Fulmer, W.W. Mantulin, and M. Glaser, 1996, "Substrate Binding Causes Movement in the ATP Binding Domain of Escherichia Coli Adenylate Kinase", *Biochemistry*, Vol. 35, pp. 6100-6106.

207. Sinev, M.A., E. V. Sineva, V. Ittah, and E. Haas, 1996, "Domain Closure in Adenylate Kinase", *Biochemistry*, Vol. 35, pp. 6425-6437.

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

209. Ping, J., P. Hao, Y.X. Li, and J.F. Wang, 2013, "Molecular Dynamics Studies on the Conformational Transitions of Adenylate Kinase: A Computational Evidence for the Conformational Selection Mechanism", *BioMed Research International*.

210. Brokaw, J.B., and J.-W. Chu, 2010, "On the Roles of Substrate Binding and Hinge Unfolding in Conformational Changes of Adenylate Kinase", *Biophysj*, Vol. 99, pp. 3420-3429.

211. Pontiggia, F., A. Zen, and C. Micheletti, 2008, "Small- and Large-Scale Conformational Changes of Adenylate Kinase: A Molecular Dynamics Study of the Subdomain Motion and Mechanics", *Biophysical Journal*, Vol. 95, pp. 5901-12.

212. Lou, H., and R.I. Cukier, 2006, "Molecular Dynamics of Apo-Adenylate Kinase: A Principal Component Analysis", *The Journal of Physical Chemistry*, Vol. 110, pp. 12796-12808.

213. Arora, K., and C.L. Brooks, 2007, "Large-Scale Allosteric Conformational Transitions of Adenylate Kinase Appear to Involve a Population-Shift Mechanism", *The Proceedings of the National Academy*, Vol. 104, pp. 18496-18501.

214. Potoyan, D.A., P.I. Zhuravlev, and G.A. Papoian, 1987, "Computing Free Energy of a Large-Scale Allosteric Transition in Adenylate Kinase Using All Atom Explicit Solvent Simulations".

215. Matsunaga, Y., H. Fujisaki, T. Terada, T. Furuta, K. Moritsugu, and A. Kidera, 2012, "Minimum Free Energy Path of Ligand-Induced Transition in Adenylate Kinase", *PLoS Computational Biology*, Vol. 8, pp. 11-13.

216. Zeller, F., and M. Zacharias, 2015, "Substrate Binding Specifically Modulates Domain Arrangements in Adenylate Kinase", *Biophysical Journal*, Vol. 109, pp. 1978-1985.

217. Rgen Ade, J., A. Verma, A. Schug, and M. Wolf-Watz, 1998, "Modulation of a Pre-existing Conformational Equilibrium Tunes Adenylate Kinase Activity".

218. Kovermann, M., J. Aden, C. Grundström, A. Elisabeth Sauer-Eriksson, U.H. Sauer, and M. Wolf-Watz, 2015, "Structural Basis for Catalytically Restrictive Dynamics of a High-Energy Enzyme State", *Nature Communications*, Vol. 6, pp. 7644.

219. Müller, C.W., G.J. Schlauderer, J. Reinstein, and G.E. Schulz, 1996, "Adenylate Kinase Motions During Catalysis: An Energetic Counterweight Balancing Substrate Binding", *Structure*, Vol. 4, pp. 147-156.

220. Daily, M.D., H. Yu, G.N. Phillips, and Q. Cui, 2013, "Allosteric Activation Transitions in Enzymes and Biomolecular Motors: Insights from Atomistic and Coarse-Grained Simulations", *Topics in Current Chemistry*, Vol. 337, pp. 139-164.

221. Beckstein, O., E.J. Denning, J.R. Perilla, and T.B. Woolf, 2009, "Zipping and

Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open - Closed Transitions", *Journal Molecular Biology*, Vol. 394, pp. 160-176.

222. Kurkcuoglu, Z., and P. Doruker, 2016, "Ligand Docking to Intermediate and Close- To-Bound Conformers Generated by an Elastic Network Model Based Algorithm for Highly" *Flexible Proteins.*

223. Matsunaga, Y., Y. Komuro, C. Kobayashi, J. Jung, T. Mori, and Y. Sugita, 2016, "Dimensionality of Collective Variables for Describing Conformational Changes of a Multi-Domain Protein", *The Journal of Physical Chemistry Letters*, Vol.1, pp. 1446-1451.

224. Kolomeisky, A.B., 2013, "Motor Proteins and Molecular Motors: How to Operate Machines at the Nanoscale", *Journal of Physics: Condensed Matter*, Vol. 25, pp. 463101.

225. Lee, C.K., Y.M. Wang, L.S. Huang, and S. Lin, 2007, "Atomic Force Microscopy: Determination of Unbinding Force, off Rate and Energy Barrier for Protein-Ligand Interaction", *Micron*, Vol. 38, pp. 446-461.

226. Hwang, W., M.J. Lang, and M. Karplus, 2008, "Force Generation in Kinesin Hinges on Cover-Neck Bundle Formation", *Structure*, Vol. 16, pp. 62-71.

227. Hua, W., J. Chung, and J. Gelles, 2002, "Distinguishing Inchworm and Hand-Over-Hand Processive Kinesin Movement by Neck Rotation Measurements", *Science*, Vol. 295, pp. 844-8.

228. Asbury, C.L., A.N. Fehr, and S.M. Block, 2003, "Kinesin Moves by an Asymmetric Hand-Over-Hand Mechanism", *Science*, Vol. 302, pp. 2130-4.

229. Yildiz, A., M. Tomishige, R.D. Vale, and P.R. Selvin, 2004, "Kinesin Walks Hand-Over-Hand", *Science*, Vol. 303, pp. 676-8.

230. Kaseda, K., H. Higuchi, and K. Hirose, 2003, "Alternate Fast and Slow Stepping

of a Heterodimeric Kinesin Molecule", *Nature Cell Biology*, Vol. 5, pp. 1079-1082.

231. Korten, T., B. Nitzsche, C. Gell, F. Ruhnow, C. Leduc, and S. Diez, 2011, "Fluorescence Imaging of Single Kinesin Motors on Immobilized Microtubules", *Methods Molecular Biology*, Vol. 783, pp. 121-137.

232. Vale, R.D., and R. a Milligan, 2000, "The way Things Move: Looking Under the Hood of Molecular Motor Proteins", *Science*, Vol. 288, pp. 88-95.

233. Case, R.B., S. Rice, C.L. Hart, B. Ly, and R.D. Vale, 2000, "Role of the Kinesin Neck Linker and Catalytic Core in Microtubule-Based Motility", *Current Biology*, Vol. 10, pp. 157-160.

234. Rice, S., A.W. Lin, D. Safer, C.L. Hart, N. Naber, B.O. Carragher, S.M. Cain, E. Pechatnikova, E.M. Wilson-Kubalek, M. Whittaker, E. Pate, R. Cooke, E.W. Taylor, R.A. Milligan, and R.D. Vale, 1999, "A Structural Change in the Kinesin Motor Protein that Drives Motility", *Nature*, Vol. 402, pp. 778-84.

235. DeBerg, H.A., B.H. Blehm, J. Sheung, A.R. Thompson, C.S. Bookwalter, S.F. Torabi, T.A. Schroer, C.L. Berger, Y. Lu, K.M. Trybus, and P.R. Selvin, 2013, "Motor Domain Phosphorylation Modulates Kinesin-1 Transport", *Journal of Biological Chemistry*, Vol. 288, pp. 32612-32621.

236. Millecamps, S., and J.-P. Julien, 2013, "Axonal Transport Deficits and Neurodegenerative Diseases", *Nature Reviews Neuroscience*, Vol. 14, pp. 161-176.

237. Kawaguchi, K., 2013, "Role of Kinesin-1 in the Pathogenesis of SPG10, a Rare form of Hereditary Spastic Paraplegia", *Neuroscientist*, Vol. 19, pp. 336-44.

238. Musumeci, O., M.T. Bassi, A. Mazzeo, M. Grandis, C. Crimella, A. Martinuzzi, and A. Toscano, 2011, "A Novel Mutation in KIF5A Gene Causing Hereditary Spastic Paraplegia with Axonal Neuropathy", *Neurol Science*, Vol. 32, pp. 665-8.

239. Forbes, S.A., G. Bhamra, S. Bamford, E. Dawson, C. Kok, J. Clements, A. Men-

zies, J.W. Teague, P.A. Futreal, and M.R. Stratton, 2008, "The Catalogue of Somatic Mutations in Cancer", *Current Protocols in Human Genetics Chapter*, Vol. 10, pp. 10-11.

240. Vetter, I.R., and A. Wittinghofer, 2001, "The Guanine Nucleotide-Binding Switch in three Dimensions", *Science*, Vol. 294, pp. 1299-1304.

241. Davis, M.J., B. Hak, E.C. Holman, R. Halaban, J. Schlessinger, T.J. Boggon, and B.H. Ha, 2013, "RAC1P29S is a Spontaneously Activating Cancer-Associated GTPase", *The Proceedings of the National Academy*, Vol. 110, pp. 912-7.

242. Jaffe, A.B., and A. Hall, 2005, "Rho GTPases: Biochemistry and Biology", *Annual Review of Cell and Developmental Biology*, Vol. 21, pp. 247-69.

243. Ridley, A.J., H.F. Paterson, C.L. Johnston, D. Diekmann, and A. Hall, 1992, "The Small GTP-Binding Protein Rac Regulates Growth Factor-Induced Membrane Ruffling", *Cell*, Vol. 70, pp. 401-410.

244. Manser, E., T. Leung, H. Salihuddin, Z. Zhao, and L. Lim, 1994, "A Brain Serine/Threonine Protein Kinase Activated by Cdc42 and Rac1", *Nature*, Vol. 367, pp. 40-46.

245. Bokoch, G.M, 2003, "Biology of the p21-Activated Kinases", *Annual Review of Biochemistry*, Vol. 72, pp. 743-781.

246. Kumar, A., V. Rajendran, R. Sethumadhavan, and R. Purohit, 2013, "Molecular Dynamic Simulation Reveals Damaging Impact of RAC1 F28L Mutation in the Switch I Region", *PLoS One*, Vol. 1, pp. 8.

247. Owen, D., and H.R. Mott, 2005, "Structural Analysis of Rho Protein Complexes, In: RHO Family GTPases", *Berlin/Heidelberg: Springer-Verlag*, Vol. 1, pp. 31-72.

248. Sells, M.A., and J. Chernoff, 1997, "Emerging from the Pak: the p21-Activated Protein Kinase Family", *Trends Cell Biol*, Vol. 7, pp. 162-167.

249. Knaus, U.G., and G.M. Bokoch, 1998, "The p21Rac/Cdc42-activated kinases, International Journal of Biochemistry", *Cell Biol*, Vol. 30, pp. 857-862.

250. Lei, M., W. Lu, W. Meng, M.-C. Parrini, M.J. Eck, B.J. Mayer, and S.C. Harrison, 2000, "Structure of PAK1 in an Autoinhibited Conformation Reveals a Multistage Activation Switch", *Cell*, Vol. 102, pp. 387-397.

251. Price, L.S., and J.G. Collard, 2001, "Regulation of the Cytoskeleton by Rho-Family Gtpases: Implications for Tumour Cell Invasion, Semin", *Cancer Biol*, Vol. 11, pp. 167-173.

252. Schmitz, A.A., E.E. Govek, B. Bottner, and L. Van Aelst, 2000, "Rho GTPases: Signaling, Migration, and Invasion", *Experimental Cell Research*, Vol. 261, pp. 1-12.

253. Chan, A.Y., S.J. Coniglio, Y. Chuang, D. Michaelson, U.G. Knaus, M.R. Philips, and M. Symons, 2005, "Roles of the Rac1 and Rac3 GTPases in Human Tumor Cell Invasion", *Oncogene*, Vol. 24, pp. 7821-9.

254. Espina, C., M.V. CÃ©spedes, M.A. GarcÃa-Cabezas, M.T. GÃ³*mezdelPulgar*, *A.Boluda*, *L.G.Oroz*, *S.A.Benitah*, *P.Cejas*, *M.Nistal*, *R.Mangues*, *and*

66.*Niggli*, *V.*, *D.Schlicht*, *andS.Affentranger*, 2009, "*SpecificRolesofRac*1*andRac*2*inMotil*

255. Vega, F.M., and A.J. Ridley, 2008, "Rho GTPases in Cancer Cell Biology", *FEBS Letters*, Vol. 582, pp. 2093-2101.

256. Ellenbroek, S.I.J., and J.G. Collard, 2007, "Rho GTPases: Functions and Association with Cancer", *Clinical and Experimental Metastasis*, Vol. 24, pp. 657-672.

257. Karnoub, A.E., and C.J. Der, 2013, "Rho Family GTPases and Cellular Transformation".

258. Kumar, R., A.E. Gururaj, and C.J. Barnes, 2006, "p21-Activated Kinases in Cancer", *Nature Reviews Cancer*, Vol. 6, pp. 459-71.

259. Radu, M., G. Semenova, R. Kosoff, and J. Chernoff, 2014, "PAK Signalling During the Development and Progression of Cancer", *Nature Reviews Cancer*, Vol. 14, pp. 13-25.

260. Oliphant, T.E., 2007, "Python for Scientific Computing. Comput", *Science and Engineering*, Vol. 9, pp. 10-20.

261. Humphrey, W., A. Dalke, and K. Schulten, 1996, "Visual Molecular Dynamics", *Journal of Molecular Graphics*, Vol. 14, pp. 33-38.

262. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, 1977, "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes", *Journal of Computational Physics*, Vol. 23, pp. 327-341.

263. Hutter, J.L., and J. Bechhoefer, 1993, "Calibration of Atomic-Force Microscope Tips", *Review of Scientific Instruments*, Vol. 64, pp. 1868-1873.

264. Busscher, H.J., W. Norde, and H.C. van der Mei, 2008, "Specific Molecular Recognition and Nonspecific Contributions to Bacterial Interaction Forces", *Applied and Environmental Microbiology*, Vol. 74, pp. 2559-2564.

265. Henny, F., C. Van Der Mei, M. Rustema-Abbing, J. De Vries, and H.J. Busscher, 2008, "Bond Strengthening in Oral Bacterial Adhesion to Salivary Conditioning Films", *Applied and Environmental Microbiology*, Vol. 74, pp. 5511-5515.

266. Wessel, S.W., Y. Chen, A. Maitra, E.R. van den Heuvel, A.M. Slomp, H.J. Busscher, and H.C. van der Mei, 2014, "Adhesion Forces and Composition of Planktonic and Adhering Oral Microbiomes", *Journal of Dental Research*, Vol. 93, pp. 84-8.

267. Hwang, G., G. Marsh, L. Gao, R. Waugh, and H. Koo, 2015, "Binding Force Dynamics of Streptococcus Mutans-Glucosyltransferase B To Candida Albicans", *Journal of Dental Research*, Vol. 94, pp. 1310-7.

268. Mikhailenko, S. V, Y. Oguchi, and S. Ishiwata, 2010, "Insights into the Mecha-

nisms of Myosin and Kinesin Molecular Motors from the Single-Molecule Unbinding Force Measurements", *Journal of the Royal Society*, Vol. 7, pp. S295-S306.