

# **Molecular Thermodynamic View of Biomolecular Signaling: Allostery and Protein-Protein Interactions**

Thesis Submitted to AcSIR for the Award of  
the Degree of  
**DOCTOR OF PHILOSOPHY**  
In  
Chemical Sciences



By

**Amit Kumawat**

10BC14A26038

Under the guidance of

**Dr. Kumar Vanka**

**Dr. Suman Chakrabarty**

**CSIR-National Chemical Laboratory  
Pune, India**



# सीएसआईआर - राष्ट्रीय रासायनिक प्रयोगशाला

(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)

डॉ. होमी भाभा मार्ग, पुणे - 411 008, भारत



## CSIR - NATIONAL CHEMICAL LABORATORY

(Council of Scientific & Industrial Research)

Dr. Homi Bhabha Road, Pune - 411 008, India

### Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled “**Molecular Thermodynamic View of Biomolecular Signaling: Allosteric and Protein-Protein Interactions**” submitted by **Mr. Amit Kumawat** to Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy in Chemical Sciences**, embodies original research work under our supervision and guidance. We further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material obtained from other sources has been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged. It is also certified that this work done by the student, under our supervision, is plagiarism free.

.....*Kumar Vanka*.....

(Supervisor)

Dr. Kumar Vanka

CSIR-National Chemical Laboratory

Pune, India

.....*Amit*.....

(Student)

Amit Kumawat

CSIR-National Chemical Laboratory

Pune, India

.....*Suman Chakrabarty*.....

(Co-supervisor)

Dr. Suman Chakrabarty

S. N. Bose National Centre for Basic Sciences

Kolkata, India



#### Communication Channels

NCL Level DID : 2590  
NCL Board No. : +91-20-25902000  
EPABX : +91-20-25893300  
: +91-20-25893400

#### FAX

Director's Office : +91-20-25902601  
COA's Office : +91-20-25902660  
SPO's Office : +91-20-25902664

#### WEBSITE

[www.ncl-india.org](http://www.ncl-india.org)



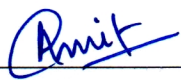


## Declaration of Authorship

I, hereby declare that the thesis entitled “**Molecular Thermodynamic View of Biomolecular Signaling: Allostery and Protein-Protein Interactions**” submitted for the degree of Doctor of Philosophy in Chemical Sciences to the Academy of Scientific & Innovative Research (AcSIR), has been carried out by me at the CSIR-National Chemical Laboratory, Pune under the supervision of Dr. Kumar Vanka (CSIR-NCL) and Dr. Suman Chakrabarty (S. N. Bose NCBS, Kolkata). Research material obtained from other sources has been duly acknowledged in the thesis. The work is original and has not been submitted in part or full by me for any other degree or diploma to any other Institution or University.

Place: Pune

Date: 21-06-2019

  
Amit Kumawat



# Abstract

*“...when you have eliminated the impossible, whatever remains,  
however improbable, must be the truth.”*

– Sir Arthur Conan Doyle, *The Sign of the Four*, Chap. 6, 1890

A biological cell communicates with the environment through numerous pathways initiating from the receptors at the membranes and extending to the interior of the cell, coordinating a large number of functional activities. To understand the underlying molecular mechanisms behind these phenomena are of primary concern towards rational drug designing. This thesis aims to investigate the thermodynamic basis of two major aspects of information propagation in biomolecular signaling, i.e. allostery (intra-protein) and protein-protein interactions (inter-protein), by capturing the complexity of the underlying free energy landscape in the form of physicochemical interactions. We have used conventional molecular dynamics simulations and metadynamics to investigate the conformational changes, dynamics and thermodynamics of these biophysical processes. Two model systems: PDZ domain and Rho GTPases were chosen to understand the molecular details of the allosteric regulation and biomolecular recognition process in signaling pathways, respectively.

PDZ domain proteins are classic examples of dynamic allostery where allostery has been attributed to purely entropic effects. In contrast, in this thesis, we show that the signal induced by perturbation (ligand binding, pH-induced protonation) propagates in the form of internal redistribution and rewiring of electrostatic interactions (enthalpic contributions) leading to a “population shift” in the hydrogen bonded network and salt bridges. In addition, we have explored the underlying mechanism of molecular recognition processes in Rho family of GTPases. We demonstrate a fine energetic balance of molecular interactions that leads to nucleotide dependent conformational selections of the switch-I region of RhoA GTPase and its subsequent role in effector recognition. Moreover, we extended our investigation to understand the phosphorylation-mediated regulation of Rac1 GTPase and RhoGDI interaction. We derive a mechanistic model based on conformational free energy landscape that exhibits rearrangement of hydrogen bonds between Rac1 and GDI upon phosphorylation. Our results are important in the context that understanding of the intricate aspects of signaling in the form of the molecular interactions would be helpful in the prediction of allosteric sites and drug discovery.



# Acknowledgements

“Anyone who ever gave you confidence, you owe them a lot.”

– Truman Capote, *Breakfast at Tiffany's*

The work presented in this thesis would not have been possible without my close association with many people. I take this opportunity to extend my sincere gratitude and appreciation to all those who made this Ph.D. thesis possible.

First and foremost, I would like to extend my sincere gratitude to my research guide, Dr. Suman Chakrabarty. Being one of the first students in his group, I had the privilege to gain experience in my research work directly from him. I appreciate all his contributions of time, ideas and funding to make my Ph.D. experience productive and stimulating. I would never have made it this far in my scientific career if it wasn't for his ever-supportive nature, patience and trust in me. Suman has inspired in me a capability to learn, to teach and to think. I would consider it to be one of my greatest realisations if I could become even a fraction of the scientist that he is. Thank you, Suman.

I would like to thank Dr. Kumar Vanka for his immense support and encouragement in ensuring that my research work continued smoothly without any bureaucratic hassle. I am privileged to have his constant and unlimited support throughout my research period in NCL. I am most grateful to our collaborator Dr. Kiran Kulkarni (CSIR-NCL), Prof. Ruth Nussinov (CCR-NCI, Bethesda) and Dr. Hyunbum Jang (CCR-NCI, Bethesda) for lending me their expertise and intuition to my scientific and technical problems.

I would like to express my sincere gratitude to the members of my Doctorial advisory committee: Dr. Ashok Giri, Dr. Ram Rup Sarkar, and Dr. Sayan Bagchi, for their insightful comments and encouragement which incented me to widen my research from various perspectives. I am extremely grateful to Dr. Debashree Ghosh for her helpful suggestions and comments during my work presentations.

I gratefully acknowledge Department of Biotechnology, for providing me financial support towards research fellowship and travel support for attending an international conference, and Director, CSIR-NCL to allow me to carry out my research in this esteemed laboratory. I would like to thank Director, S.N. Bose National Centre for Basic Sciences for granting permission to visit and work at the institute under Visitor, Associates and Students' Programme and use the available computational facilities.

My time at NCL was made enjoyable in large part due to the many friends and groups that became a part of my life. I would like to thank all my friends and well-wishers at NCL for being a friendly and cheerful group of colleagues. A special thanks to my friends Anand, Bhagyashri, Neharika, Nalini, Rahul, Samik, Sneha and Vrushali for many scientific and non-scientific discussions and for being excellent friends throughout this journey. I would also like to thank past and present SC-Lab, DG-Lab, KK-Lab and KV-Lab members for their help and support.

A special thanks to Ms. Anita Sharma (DBT-CTEP Management Cell) and Shraddha Puntambekar (CSIR-NCL) for their guidance during the travel support application process. I would also like to extend huge, warm thanks to my friends, Mantu, Prasun, Tanmoy, Susmita, and Debayan at SNBNCBS, Kolkata for providing me stimulating and fun-filled environment. I would also like to thank my friends Anil, Ankita, Ajinkya, Bora & Daga, Madhumita, Parul, Radhe and Tanushree for many unforgettable memories which I will cherish forever.

A special thanks to my sister Rima, brother-in-law Nitin and my niece Samiha for their love, affection and moral support. Last, but not least, I would like to express my deepest gratitude to my parents, who supported me in all my pursuits. Their patience and sacrifice will remain as an inspiration throughout my life.

Amit Kumawat  
CSIR-NCL  
June 2019

# List of Publications

"Study the science of art. Study the art of science. Develop your senses  
-especially learn how to see. Realize that everything connects to everything else."

– Leonardo da Vinci

1. Hidden electrostatic basis of dynamic allostery in a PDZ domain. **A. Kumawat** and S. Chakrabarty, *Proc. Nat. Acad. Sci. USA.*, 114, E5825 - E5834 (2017).
2. Nucleotide dependent switching in Rho GTPase: Conformational heterogeneity and competing molecular interactions. **A. Kumawat**, S. Chakrabarty, K. Kulkarni, *Scientific Reports.*, 7, 45829 (2017).
3. Protonation Induced Dynamic Allostery in PDZ Domain: Evidence of Perturbation Independent Universal Response Network. **A. Kumawat** and S. Chakrabarty (*Under submission*).
4. Molecular insights into the regulation of Rac1 by selective phosphorylation of RhoGDI. **A. Kumawat**, H. Jang, S. Chakrabarty and R. Nussinov (*Under submission*).
5. Organoselenium Compounds as Acetylcholinesterase Inhibitors: A Comprehensive Investigation of Mixed Inhibition Mechanism. **Amit Kumawat**, Fasil Ali, Tanveer Ali Dar, Shabnam Raheem, Suman Chakrabarty\* and Masood Ahmad Rizvi\* (*Under submission*).
6. Addressing the role of structural stability and dynamics of  $\alpha$ 1-helix in PDZ2 domain. **A. Kumawat** and S. Chakrabarty (*Under submission*).





# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>List of Publications</b>	<b>v</b>
<b>Contents</b>	<b>vii</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xiii</b>
<b>1. Introduction</b>	<b>1</b>
1.1 Protein-Protein Interactions .....	2
1.2 Activation and Deactivation of Proteins.....	4
1.3 Molecular Basis of Inter-protein Signal Propagation in Rho GTPases .....	5
1.3.1 Thermodynamic View of Effector Recognition Mechanism .....	9
1.3.2 Regulation of Rho GTPase Activities by RhoGDI .....	10
1.4 Molecular Basis of Intra-protein Signal Propagation alias ‘Allostery’ .....	13
1.4.1 Ensemble Nature and Thermodynamic View of Allostery .....	15
1.4.2 Experimental Methods to Probe Allosteric Mechanism .....	18
1.4.3 Computational Approaches to Probe Allosteric Mechanism .....	19
1.4.4 Dynamic Allostery in PDZ Domain: An Unsolved Puzzle.....	23
1.5 “Thermodynamics” and “Molecular Insights” into the Biomolecular Signaling .....	24
1.6 Bibliography .....	26
<b>2. Methodology</b>	<b>45</b>
2.1 Molecular Dynamics Simulations: Background and Theory.....	46
2.1.1 Molecular Dynamics Integrators.....	47
2.2 Force Field.....	50
2.3 Technical Details of Molecular Dynamics Simulations .....	52
2.3.1 Periodic Boundary Conditions .....	52

2.3.2	Minimum Image Convention and Truncation of Interactions.....	53
2.3.3	Long-range Interactions .....	54
2.3.4	Neighbour and Cell Lists .....	55
2.3.5	Thermodynamics Ensembles .....	56
2.4	Steps in Running Molecular Dynamics Simulation.....	59
2.5	Data Analysis Methods.....	60
2.6	Free Energy Calculations.....	65
2.6.1	Potential of Mean Force .....	66
2.6.2	Rare Event Sampling.....	67
2.7	Other Computational Details .....	71
2.8	Bibliography .....	72
<b>3.</b>	<b>Hidden Electrostatic Basis of Dynamic Allostery in a PDZ3 Domain</b>	<b>77</b>
3.1	Introduction.....	77
3.2	Methods .....	79
3.3	Results.....	82
3.4	Discussion.....	103
3.5	Bibliography .....	105
<b>4.</b>	<b>pH-dependent Dynamic Allostery in PDZ3 Domain</b>	<b>111</b>
4.1	Introduction.....	111
4.2	Methods .....	114
4.3	Results.....	115
4.4	Discussion.....	125
4.5	Bibliography .....	126
<b>5.</b>	<b>Nucleotide Dependent Conformational Switching in Rho GTPase</b>	<b>131</b>
5.1	Introduction.....	131
5.2	Methods .....	134
5.3	Results.....	138
5.4	Discussion.....	156
5.5	Bibliography .....	157

<b>6. Molecular Insights into the Regulation of Rac1 by Phosphorylation of GDI</b>	<b>163</b>
6.1 Introduction.....	163
6.2 Methods .....	166
6.3 Results.....	169
6.4 Discussion.....	188
6.5 Bibliography .....	190
<b>7. Conclusions and Future Outlook</b>	<b>195</b>
7.1 Decoding Dynamic Allostery in PDZ3 Domain Using Electrostatic Interactions .....	195
7.2 Molecular Insights into Rho GTPase Activation and GDI-mediated Regulation .....	197
7.3 Bibliography .....	199



# List of Figures

<b>Figure 1.1</b>	Crystal structure of RhoA and classification of Rho subfamily of GTPases .....	6
<b>Figure 1.2</b>	Rho GTPase cycle .....	7
<b>Figure 1.3</b>	Types of lipid modifications in Rho GTPases.....	8
<b>Figure 1.4</b>	Rho GTPase and GDI regulation cycle .....	12
<b>Figure 1.5</b>	Induced fit model (KNF) and conformational selection model (MWC) .....	14
<b>Figure 1.6</b>	Dynamic continuum of allostery phenomena .....	16
<b>Figure 1.7</b>	Thermodynamic view of allostery via a bi-stable switch .....	17
<b>Figure 2.1</b>	Representation of potential energy terms of a typical force field.....	51
<b>Figure 2.2</b>	Schematic representation of periodic boundary condition in a 2D system .....	52
<b>Figure 2.3</b>	Schematic representation of K-means algorithm.....	64
<b>Figure 2.4</b>	Illustration of umbrella sampling technique .....	68
<b>Figure 2.5</b>	Schematic representation of metadynamics technique .....	70
<b>Figure 3.1</b>	Structural analysis between unbound and bound PDZ3 domain .....	82
<b>Figure 3.2</b>	Differential contact map and energetic perturbation of minimised structures. ....	84
<b>Figure 3.3</b>	Differential contact frequency map from the MD simulation trajectories .....	86
<b>Figure 3.4</b>	Residue-wise changes in interaction energy between bound and unbound states....	90
<b>Figure 3.5</b>	Electrostatic perturbation network based on pair-wise interaction energies .....	93
<b>Figure 3.6</b>	Rearrangement and re-wiring of side chain interaction network in PDZ3 domain..	96
<b>Figure 3.7</b>	Population shift in hydrogen bond network upon ligand binding .....	99
<b>Figure 3.8</b>	Residue-wise energetic fluctuations of protein-protein, protein-water interaction	101
<b>Figure 3.9</b>	Pairwise energetic fluctuations of protein-protein and protein-water interaction ..	102
<b>Figure 4.1</b>	Protonation of histidine residues at positions 317 and 372 in PDZ3 domain.....	113
<b>Figure 4.2</b>	Residue-wise fluctuation profile (RMSF) in protonation states .....	115
<b>Figure 4.3</b>	Residue-wise comparison between average C $\alpha$ minimum distance.....	116

<b>Figure 4.4</b>	Residue-wise change in average electrostatic energy in protonation states .....	118
<b>Figure 4.5</b>	Network representation of the perturbation in pairwise interaction energies.....	120
<b>Figure 4.6</b>	Population shift in hydrogen bond network upon protonation. ....	123
<b>Figure 5.1</b>	Structural analysis of RhoA.....	138
<b>Figure 5.2</b>	Conformational similarities of switch I region between GDP/GTP/Freeform.....	140
<b>Figure 5.3</b>	Conformational free energy landscape of GDP/GTP/Freeform/GTP-mutant.....	143
<b>Figure 5.4</b>	Correlation between Phe39 and Tyr34 orientation.....	148
<b>Figure 5.5</b>	Conformational states of Switch I based on the Tyr34 orientation .....	149
<b>Figure 5.6</b>	Frequency of Tyr34 sidechain orientation.....	150
<b>Figure 5.7</b>	Solvent exposure of polar and nonpolar residues of Switch I region.....	151
<b>Figure 5.8</b>	Multiple sequence alignment of human Rho GTPases.....	152
<b>Figure 5.9</b>	Structural comparison of H-Ras and RhoA protein in their effector bound state ..	153
<b>Figure 5.10</b>	Stabilization of the GTP bound state with respect to the GDP bound state .....	154
<b>Figure 5.11</b>	Interaction energy for GDP and GTP bound state of switch-I region. ....	155
<b>Figure 6.1</b>	Structural deviation and fluctuations in Rac1-GDI complex .....	170
<b>Figure 6.2</b>	Superimposition of structures from phosphorylated MD trajectory.....	171
<b>Figure 6.3</b>	Minimum distance analysis between the interacting regions of Rac1 and GDI.....	173
<b>Figure 6.4</b>	Principal component analysis for Rac1 and GDI in different states.....	175
<b>Figure 6.5</b>	Extreme conformations and correlated motions in Rac1-GDI complex .....	176
<b>Figure 6.6</b>	Residue-wise change in the interaction energy of Rac1 and GDI.....	178
<b>Figure 6.7</b>	Movement (destabilisation) of the prenyl group in the hydrophobic cavity .....	179
<b>Figure 6.8</b>	Interactions that control the drawbridge-like motion of the HTH domain.....	181
<b>Figure 6.9</b>	Molecular basis for signal propagation in Rho GTPase-GDI complex.....	182
<b>Figure 6.10</b>	Free energy surfaces corresponding to (a) wild type (b) Phosphorylated state....	185
<b>Figure 6.11</b>	Mechanistic model for the transition states based on free energy landscape .....	186

# List of Tables

<b>Table 1.1</b>	Diversity of interaction between Rho GTPases and GDI. ....	11
<b>Table 3.1</b>	List of residues with significant changes in differential contact frequency map.....	88
<b>Table 3.2</b>	Breakdown of perturbation in pair-wise interactions in PDZ3 domain.....	95
<b>Table 4.1</b>	Change in the pairwise interaction energy between different protonation states .....	121
<b>Table 4.2</b>	Hydrogen bond occupancy (percentage) between different protonation states.....	124
<b>Table 5.1</b>	Selection criteria for residues used in unsupervised clustering .....	136
<b>Table 5.2</b>	List of residues used for k-means clustering.....	137
<b>Table 5.3</b>	Summary of crystal structures overlaid on free energy of RhoA, Cdc42, Rac1.....	144
<b>Table 5.4</b>	Free energy barrier between different minima in GDP/GTP/Freeform states.....	146
<b>Table 6.1</b>	Post-translational phosphorylation sites in Rho GTPase-GDI regulation .....	164





# Chapter 1

## Introduction

*“Ce que nous connaissons est peu de chose; ce que nous ignorons est immense.*

*What we know is not much. What we do not know is immense.”*

– Pierre Simon Laplace

**S**ignal transduction process is a complex chain of organised biochemical reactions generated in response to the external or intracellular stimuli. These external or internal stimuli which initiate or activate the signal transduction pathways are known as primary messengers and can be chemical stimuli (e.g. metal ions, ligands, hormones, neurotransmitters, chemokines, paracrine and autocrine factors) or physical stimuli (e.g. pressure, light, heat)<sup>1-5</sup>. These signals are capable of generating cascades of tightly controlled responses at multiple steps in the pathways. Proteins form an elementary unit of these signal transduction cascades. It is a highly specific protein-protein interaction driven process which regulates events linked for the survival of the cell (e.g. gene expression, enzyme activity and cell division activity). A general notion is that these intermediate steps amplify the signal so that a small signal can generate a large response leading to change in cell function. The binary switching *on* and *off* states of a signaling protein plays a crucial role in the signal propagation. A typical cell membrane hosts a variety of transmembrane proteins such as G protein-coupled receptors (GPCRs), receptor tyrosine kinases (RTK), ion channels acting as receptors for these external stimuli<sup>6-10</sup>. Over the last few decades, the receptor activation mechanism has been extensively studied to understand the selectivity and specificity between the interacting entities. The effect of receptor activation is reflected in terms of protein activation/deactivation through binding of small molecules termed as secondary messengers (cAMP, cGMP, Ca<sup>2+</sup> ions, IP3) or protein directed post-translational modifications such as phosphorylation, ubiquitination or localisation in the cell<sup>11</sup>. This chapter provides an overview of the evidences from previous studies to understand the molecular basis of signaling mechanism. In addition, a greater emphasis has been provided to understand how molecular thermodynamics and dynamics govern the biomolecular signaling processes.

## 1.1 Protein-Protein Interactions

---

### 1.1 Protein-Protein Interactions

Protein-protein interactions play a pivotal role in regulating signaling pathways<sup>11</sup>. In particular, how two proteins recognise each other is one of the most fundamental problems in biology. The entire activity is highly specific, based on sequence-based recognition or structural properties of proteins such that the information is transduced via molecular recognition process to downstream effectors. The complex nature of signaling mechanism could be approximated with the information that a single protein can interact with multiple proteins at the same time and in the same space or at different times or locations. These interactions are dictated by various kinds of non-bonded interactions at the molecular scale, namely electrostatic, hydrogen bonding, van der Waals and hydrophobic interactions. The binding interfaces could differ depending on the interacting partners such as protein-membrane<sup>10</sup>, protein-protein<sup>12</sup>, or protein-nucleic acid<sup>13</sup>. A comparative analysis of the proteins and their complexes reveals that the spatial distribution of charged and non-polar amino acids differs on these interfaces<sup>14</sup>. For example, in DNA-Histone complex, DNA is highly negatively charged, and histone proteins are positively charged largely due to lysine side chains<sup>15</sup>. These differences in the distribution of the amino acids have been exploited as targets for drug designing and prediction of protein-protein interaction sites using machine learning techniques.

Most of the signaling proteins are stable three-dimensional domains that are capable of independent functions. Multiple domains can coexist in a single signaling protein. These domains recognise post-translational modifications such as tyrosine phosphorylation (pTyr), methylation or short amino acid sequence motifs containing a specific pattern of amino acids such as hydrophobic amino acids, proline-rich and evolutionary conserved amino acid at specific positions<sup>16-18</sup>. It is well established that the binding site in proteins is characterised by the presence of hydrophobic residues surrounded by charged residues or vice-versa<sup>19</sup>. Such is the case of ligand binding in PDZ domain. PDZ domains are one of the most potential regulators of signal transduction that bind to the specific recognition sequences at the C-terminal of proteins, often transmembrane receptors (more than 80 GPCRs, ion channels)<sup>20</sup>. The carboxylate-binding site in PDZ domain is characterised by the highly conserved loop (*R/K-XXX-G-Φ-G-Φ* motif, where *X* is any amino acid residue, and *Φ* is any hydrophobic residue) that creates a hydrophobic pocket for ligand recognition<sup>21,22</sup>. Many proteins contain several PDZ domains (e.g. GRIP

contains 7 PDZ domains) which can bind individually to multiple subunits of a channel inducing localisation and clustering of transmembrane receptors<sup>23</sup>. A collective analysis of more than 3000 ligands reveals three major classes of PDZ domains which bind to specific sequence motifs at the C-terminal of ligands containing Serine/Threonine/ $\Phi$ (*hydrophobic amino acid*)<sup>20,24</sup>. Several studies have investigated the crucial role of the pairing of electrostatic interactions (salt-bridges, hydrogen bonds) and hydrophobicity in ligand recognition specificity in PDZ domain<sup>25-27</sup>. However, a mechanism that ensures the selectivity and specificity of protein-protein interactions at the molecular level remains elusive.

In addition, these binding interfaces constitute amino-acids with ionizable side-chains (e.g. Arg, Lys, His, Asp and Glu) that undergoes post-translational charge altering modifications such as phosphorylation, acetylation or pH-induced protonation-deprotonation<sup>15,19,28</sup>. For example, SH2, SH3 or PTB (pTyr-binding) domains present in cytosolic signaling proteins such as phospholipase C $\gamma$ , Ras GTPase activating protein (GAP) and protein-tyrosine kinases recognise phosphotyrosine containing sequence motifs in activated receptor tyrosine kinases (RTK)<sup>29,30</sup>. Interestingly, under physiological conditions, the protein-protein interfaces exhibit different degrees of solvation and spatial distribution of water depending on the structural properties (polarity and geometry) of the interface. Several studies suggest water-mediated interactions as a part of biomolecular recognition process where it can act as a shield that weakens the unfavourable charge-charge interactions upon ligand binding<sup>31</sup> or as a linker between indirect charge-charge interactions<sup>13,32</sup>. Furthermore, it insists that these interactions favour enthalpically but can also enhance ligand binding affinity entropically, in case the water resides in the hydrophobic cavities<sup>33,34</sup>. It was observed that water molecules inside a non-polar cavity would have more freedom and hence, higher entropy as compared to the bulk, where the water molecules participate in the water network. This diversity of possible interactions requires spatiotemporal regulation of protein-protein interactions, and although our knowledge about the diverse interacting partners in signaling process is broad, it is important to have a detailed molecular view of the undergoing processes.

## 1.2 Activation and Deactivation of Proteins

---

### 1.2 Activation and Deactivation of Proteins

Protein regulation is one of the most complex and ubiquitous processes in biology. With the development of biophysical and biochemical techniques, the major focus is to understand the activation/deactivation mechanism and the molecular nature of the inactive and active states of the signaling proteins which are part of signal transduction cascade<sup>35</sup>. Previous studies reveal two major modes of mechanism, namely, covalent modification of amino acid residues at different sites<sup>36</sup> and conformational changes in the proteins<sup>37</sup> as imprints of activated/deactivated state of a protein in signaling pathways. One of the most diverse and important covalent modification is protein phosphorylation. This post-translational modification method imparts a molecular switch behaviour through activation and deactivation of a large number of enzymes and transmembrane receptors by kinases and phosphatases respectively<sup>38</sup>. The signaling cascade involving protein kinase activity aims to propagate and concurrently amplify the signal in terms of promoting or inhibiting the protein-ligand or protein-protein interactions. For example, in case of receptor tyrosine kinases (RTKs) activated signaling pathway, several proteins (Src, PLC  $\gamma$ , Ras-GAP, Raf, PI3K) and secondary messengers (cAMP, cGMP, Ca<sup>2+</sup> and diacylglycerol (DAG)) facilitate the activation of downstream protein kinases (PKC, PKA, CaM)<sup>30</sup>. Experimental studies show PKA mediated serine phosphorylation at the PDZ recognition site of K<sup>+</sup> channels results in the uncoupling of the channel from second PDZ domain of PSD-95 and subsequent inhibition of K<sup>+</sup> conductance<sup>39</sup> whereas PKC mediated serine phosphorylation of RhoGDI stimulates its dissociation from Rho GTPase, thereby resulting in Rho GTPase activation<sup>40,41</sup>.

In addition, the activation and deactivation mechanisms are dominated by allosteric events leading to the change in molecular conformation of the target molecule<sup>42</sup>. These events have been underappreciated and often intertwine with covalent modifications in the cell. Over recent years, there have been evidences of conformational driven signaling pathways<sup>43</sup>. The active and inactive states of signaling proteins exist in a conformational equilibrium with tunable population distribution, where there can be multiple metastable states separated by activation barrier over a free energy landscape<sup>44</sup>. The existence of multiple metastable states highlight that there may not be a single active state and ligand binding can activate diverse downstream signaling pathways. The binding of an external ligand to a transmembrane receptor or cytoplasmic protein within a cell shifts the conformational equilibrium from an inactive to a

functionally active state<sup>45</sup>. Several experimental evidences and computational studies on transmembrane receptors and protein complexes provide detailed insight into the dynamic processes and conformational changes during the activation<sup>9,46,47</sup>. It is evident from these studies that the activation and deactivation mechanisms involve a wide range of phenomenon ranging from conformational population shift<sup>47</sup>, allosteric network<sup>48</sup> or in some case exposure of novel interfaces (T cell receptor (TCR)-CD3 complex)<sup>49</sup>. Collectively, all these pathways involve signal propagation either through protein-protein interactions (inter-protein) or by allosteric mechanisms (intra-protein). In this thesis, we aim to understand the underlying mechanism of signal transduction pathways at these two different levels, i.e. intra-protein and inter-protein using two representative protein systems involved in a large number of signaling pathways, PDZ domain and Rho GTPases respectively.

### 1.3 Molecular Basis of Inter-protein Signal Propagation in Rho GTPases

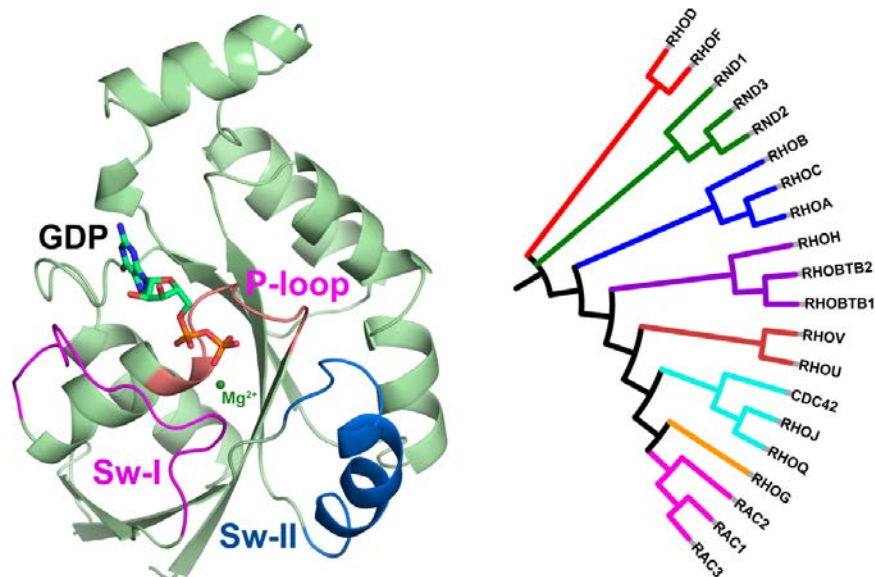
Protein-protein interaction between two signal transduction pathways, i.e. crosstalk, increases the complexity and specificity of actions of different extracellular signals. The molecular interactions that underline the binding specificity between this crosstalk represent the key events in defining their functions. Unlike heterotrimeric GTPases which interact with the activated GPCRs, there exist a large number of other G-proteins that take part in intracellular signaling<sup>50</sup>. Small GTPases of Ras superfamily is one such class of G-proteins which are involved in regulatory processes such as cell division, cell polarity, shape and migration that require crosstalk between multiple signaling pathways<sup>51</sup>. The Ras superfamily is further divided into Ras, Rho, Ran, Rab and Arf GTPases. These GTPases undergo protein-protein interaction mediated complex cycle of regulation and deregulation in the signal propagation pathways<sup>52-54</sup>.

Rho GTPases are part of Ras superfamily of G-proteins that regulate actin cytoskeleton dynamics and thereby coordinate a large number of signal transduction pathways in eukaryotic cells<sup>53-58</sup>. Alteration in Rho GTPases expression and activation contributes to the human cancer development<sup>59-62</sup>, neurological abnormalities<sup>63</sup> and bacterial infections<sup>64-66</sup>. These are small, monomeric GTPases which comprises of a common G-domain containing 6 $\beta$  strands and 5 $\alpha$  helices, and differ from other GTPases with the presence of a distinct insert region between  $\beta$ 5 and  $\alpha$ 4 region<sup>67-69</sup>. Previous studies have shown that the insert region is important for interaction with certain effector proteins for downstream signaling<sup>70</sup>. Rho GTPases share common sequence

### 1.3 Molecular Basis of Inter-protein Signal Propagation in Rho GTPases

---

and structural features of G-domain, which consists of switch I and switch II regions involved in nucleotide exchange mechanism<sup>71-76</sup>. Over recent years, more than 20 members of Rho GTPases have been identified and classified into eight subgroups<sup>77</sup> (Fig. 1.1).



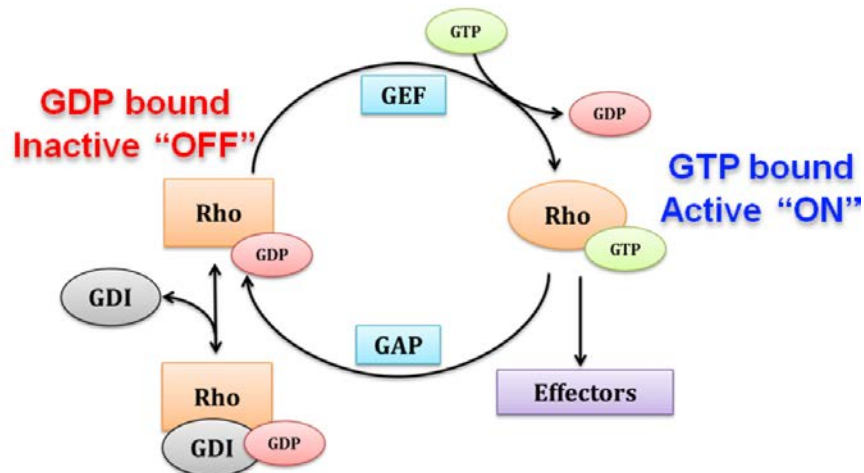
**Figure 1.1** (a) Crystal structure of RhoA in the GDP bound state (PDB ID: 1FTN). Switch I and II regions are highlighted in magenta and blue, respectively. The highly conserved P-loop motif is highlighted in orange. GDP and Mg<sup>2+</sup> ion are in ball-and-stick model in green colour. (b) Classification of Rho subfamily of GTPases according to the sequence similarity. Members of each subgroup are highlighted using the same colour code.

Rho GTPases share around 30% amino acid sequence identity with Ras protein and 45-90% identity within the family<sup>51,78,79</sup>. Majority of the functional information about the Rho GTPases is obtained from three proteins, namely RhoA Rac1 and cdc42. Rho protein has three isoforms RhoA, RhoB and RhoC, which share >85% amino acid sequence identity<sup>80</sup>. Despite the high sequence similarity between these isoforms, distinct modifications at the C-terminal and subsequent different localisation results in different expression level and cellular functions in the cell.

#### Role of regulatory proteins: GEFs, GAPs and GDIs

Like all other G-proteins, these proteins function as molecular switches that regulate cellular functions by using a simple biochemical strategy of switching between an active GTP bound

state and inactive GDP bound state. This catalytic cycle is regulated by three proteins, namely GTPase-activating proteins (GAPs), guanine nucleotide exchange factors (GEFs) and guanine nucleotide dissociation inhibitors (GDIs)<sup>76,81,82</sup> (Fig. 1.2).



**Figure 1.2** Rho GTPase cycle between GTP-bound active and GDP-bound inactive states. In the active state, Rho GTPases interact with downstream effector proteins. The GTPase cycle is regulated by three classes of proteins: (1) Guanine nucleotide exchange factors (GEFs) that catalyze nucleotide exchange and activates GTPase (2) GTPase-activating proteins (GAPs) catalyze GTP hydrolysis, and guanine nucleotide dissociation inhibitors (GDIs) inhibit the dissociation of GDP and subsequent binding of GTP.

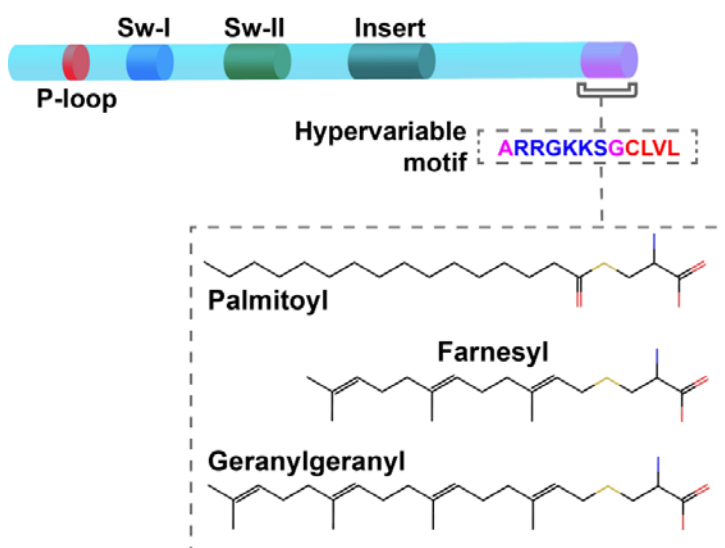
GAPs stimulate the intrinsic GTPase activity of Rho proteins and accelerate the formation of GDP-bound inactive state. Biochemical and structural data suggest the role of evolutionary conserved catalytic arginine residue (R305) of RhoGAP in accelerating GTP hydrolysis by stabilising charges developed during the transition state<sup>83-85</sup>. In the Rho GTPase cycle, GDIs act as a negative regulator by binding to the prenylated form of Rho GTPases in the GDP bound state<sup>82</sup>. Several transmembrane proteins such as GPCRs, RTKs, TK-associated receptors and integrins activate Rho GTPases directly or indirectly through activation of GEFs<sup>10,86,87</sup>. Upon activation, Rho GTPase proteins interact with a wide range of effectors initiating a network of cytoplasmic and nuclear signaling cascade that control processes ranging from cytoskeletal rearrangements<sup>54,88,89</sup> to gene transcription<sup>90,91</sup>. Thus, Rho GTPases act as signaling nodes that can transmit information from upstream receptors activated by diverse external stimuli to specific downstream targets<sup>92</sup>.

### 1.3 Molecular Basis of Inter-protein Signal Propagation in Rho GTPases

---

#### Regulation by post-translational modifications

In addition to the GTP hydrolysis mechanism, Rho GTPases are also regulated by the specific protein-protein complex formation and post-translation modifications<sup>93</sup>. This ensures the spatiotemporal regulation of GTPase in order to interact with various effectors. The post-translational modifications include lipid modification (prenylation, palmitoylation), phosphorylation<sup>94-98</sup> and ubiquitylation<sup>99-102</sup> for regulating GTPase activity and levels inside the cell. Lipid modifications play a pivotal role in subcellular localisation of Rho GTPases to distinct membrane compartments and regulate the downstream signaling pathways (Fig. 1.3).



**Figure 1.3** Types of lipid modifications in Rho GTPases. (a) Illustration of a typical G-domain in Rho GTPases. The P-loop motif (red) and switch regions (I (blue) & II (green)) are highly conserved and involved in binding to the nucleotide (GDP/GTP). The hypervariable region defines the specificity in Rho GTPases. The sequence shown is specific for RhoA. The CAAX motif (red) at the C-terminal undergoes post-translational lipid modifications for membrane attachment. (b) The box shows the chemical structure of palmitoyl, farnesyl and geranylgeranyl group attached to the cysteine residue. These modifications defer in their hydrophobicity, selectivity and position of modification and regulation.

Rho GTPases are characterised by the tetrapeptide motif CAAX (C- cysteine, A-any aliphatic amino acid, X-any amino acid) at carboxyl-terminal. This motif acts as an initiation site for the addition of farnesyl (15-carbon chain) or geranylgeranyl (20-carbon chain) moiety



mediated by FTase (farnesyl transferase) and GGTase (geranylgeranyl transferase) enzymes respectively<sup>103,104</sup>. This is followed by the cleaving of –AAX peptide by endoprotease<sup>105</sup> and finally, carboxymethylation of cysteine residue<sup>106,107</sup>. The addition of lipid tail increases the hydrophobicity of the terminal and facilitates membrane association of the Rho GTPase at different locations<sup>108</sup>. On the other hand, phosphorylation of residues close to the lipid modifications can alter the localisation at the membrane by enhancing its interaction with RhoGDIs. Experimentally, it has been shown that phosphorylation on Ser188 of RhoA inhibits the activity by decreasing its association with effector protein ROCK and increasing its association with RhoGDI<sup>109</sup>.

### 1.3.1 Thermodynamic View of Effector Recognition Mechanism

Rho GTPases are capable of interacting with a large number of effectors regulating diverse cellular functions<sup>108,110</sup>. These effectors bind specifically to the GTP-bound active state of Rho GTPases. Several crystal structures highlight the conformational changes in the switch regions between the GDP-bound inactive state and GTP-bound active state<sup>68,69,111</sup>. Interestingly, effector proteins must recognise these differences/distinct features between the GTP-bound and GDP-bound states for downstream signaling. The lists of Rho binding proteins include upstream regulators (GEFs, GAPs) and downstream effectors such as Ser/Thr kinases (Protein Kinase N, Rho-kinase, PAK), lipid kinases, SH3 domain containing proteins and others. Mutations and structural studies identify an essential role of switch I region for the effector protein recognition and hence termed “effector region”<sup>112</sup>. Sequence analysis reveals diverse and variable amino acid residues in the switch I region, suggesting its association with different effector proteins. For example, TPR domain of NADPH oxidase binds to Rac1 with specific residues Ala27 and Gly30 in the switch I region<sup>113</sup>. Experimentally, TPR domain was shown to bind cdc42 after mutating residues at these key positions 27 and 30 to Alanine and Glycine respectively. Similarly, protein kinase N recognises specific residues in the switch I of RhoA which are different from corresponding residues in cdc42 and thus maintaining specificity for effector activation<sup>114</sup>. Further studies indicate that some effectors bind outside the Switch I region for activation. Using point mutations, it has been shown that ROCK and rhotillin require sequence outside effector region (switch I) for binding to RhoA<sup>115</sup>. Similarly using Rac1/RhoA chimeras, the importance

### 1.3 Molecular Basis of Inter-protein Signal Propagation in Rho GTPases

---

of C-terminal of Rac1 for the activation of PAK and p67<sup>phox</sup> (a subunit of NADPH oxidase) was highlighted<sup>116</sup>.

Over the last two decades, several mutational and structural analyses have revealed multiple mechanisms behind the selectivity and specificity between the effectors and Rho GTPases<sup>117-121</sup>. Comparison between the binding affinities for different effectors binding to the members of Ras superfamily suggests large differences in the energetics<sup>122-126</sup>. Further, energy decomposition of Ras-effector complexes show that the enthalpy and entropy balance each other, contributing favourably towards the binding free energy of the complex formation. For example, Rho proteins exhibit different thermodynamic signatures upon binding to peptide effectors (Cdc42-WASP complex) and large domains (Rac1-TPR domain of p67<sup>phox</sup>). Here, Cdc42-WASP complex has high enthalpic contribution along with high negative entropy, whereas Rac1-TPR domain complex has positive enthalpy compensated by high entropic contributions for favourable binding<sup>113,127</sup>. Previous studies have shown that Ras superfamily proteins and effectors exhibit charge complementarity on the binding surfaces showing the involvement of electrostatic interactions or hydrophobic contributions in protein-protein interactions<sup>31,127</sup>. Interestingly these energetic signatures that govern the effector activation are still elusive, and therefore, it is important to understand the molecular mechanism of Rho GTPase recognition by the effectors thermodynamically.

#### 1.3.2 Regulation of Rho GTPase Activities by RhoGDI

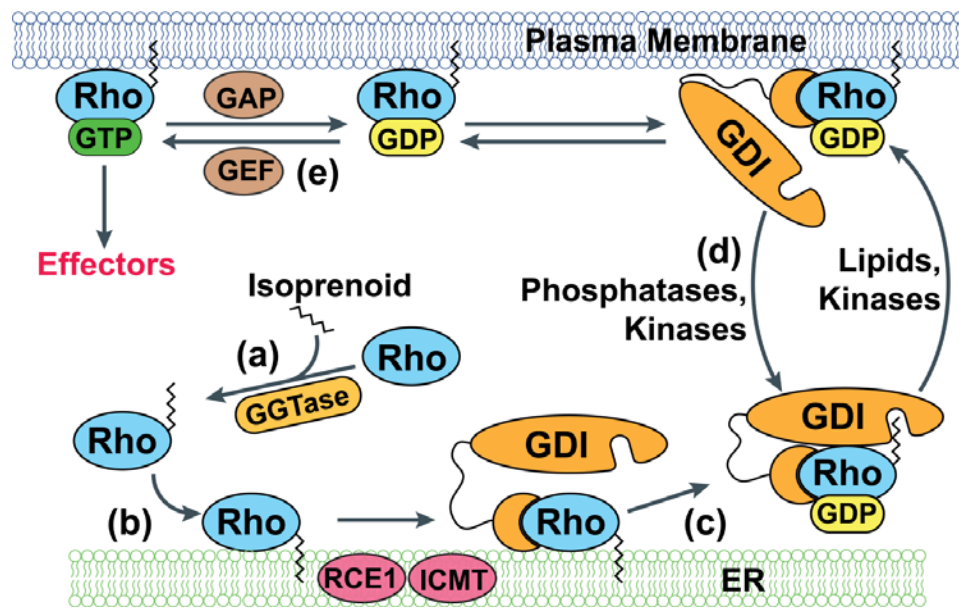
RhoGDIs (Guanine nucleotide dissociation inhibitors) play a critical role in the spatiotemporal regulation of signaling events by controlling the timing and localisation of Rho proteins<sup>110,128,129</sup>. RhoGDIs act in the background like an ‘invisible hand’ as described by Garcia-Mata et al., regulating the level of activated/deactivated GTPases in the cell<sup>82</sup>. Interestingly, only three genes encode for RhoGDIs ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) as compared to the diverse and large number of Rho GTPase family members<sup>130</sup>. RhoGDI $\alpha$  is ubiquitously expressed<sup>131</sup>, whereas RhoGDI $\beta$  is expressed selectively in haematopoietic cells<sup>132,133</sup>. Another one, RhoGDI $\gamma$  is associated with cellular membranes<sup>134,135</sup> (Golgi complexes) and interacts specifically with RhoB and RhoG. RhoGDI $\alpha$  is the best-characterised member and interacts with a number of Rho GTPases (RhoA, Cdc42, Rac1, Rac2). Several experimental evidences suggest multiple interactions between GDIs and

GTPases (Table 1.1) and changes in GDI expression levels have been associated with cancers<sup>136</sup>. The up-regulated or down-regulated expression levels of GDI in different cancer types have been correlated with increased invasion, resistance to chemotherapy or degree of malignancy<sup>137</sup>. These changes in the expression are manifested through their interactions with multiple Rho GTPases, which are further regulated by different mechanisms.

**Table 1.1** Diversity of interaction between Rho GTPases and GDI. Here, the '+/-' symbol denotes the experimentally observed presence or absence of interaction between Rho protein and GDI. The symbol '?' denotes that no experimental proof has been found so far.

Rho GTPases	RhoGDI (GDI-1, $\alpha$ )	D4GDI (GDI-2, $\beta$ )	RhoGDI $\gamma$ (GDI-3)
Cdc42 <sup>138,139</sup>	+	+	-
Rac1 <sup>140,141</sup>	+	+	-
Rac2 <sup>140,141</sup>	+	+	-
Rac3	?	?	-
RhoG <sup>12,134</sup>	+	?	+
Rac1b <sup>142</sup>	-	?	-
RhoA <sup>131</sup>	+	+	-
RhoB <sup>12,134,143</sup>	-	?	+
	+	?	-
RhoC <sup>12</sup>	+	?	-
Rnd3/RhoE <sup>144,145</sup>	+/-	?	?
RhoH/TTF12	Weak	?	?

### 1.3 Molecular Basis of Inter-protein Signal Propagation in Rho GTPases



**Figure 1.4** Newly synthesized Rho protein is geranylgeranylated by GGTase enzyme followed by membrane attachment at the cytoplasmic side of endoplasmic reticulum (ER). (b) After geranylgeranylation, Rho protein is post-translationally modified by RCE1 and ICMT for the removal of –AAX motif at the C-terminal and subsequent carboxyl methyltransferase. (c) RhoGDI binds to the Rho protein and sequesters the prenyl moiety to prevent its degradation in soluble cytosolic form. (d) RhoGDI-GTPase complex is regulated between the cytosol and membrane by post-translational modifications by lipids, phosphatases and kinases. (e) Rho protein attached to the membrane in the absence of GDI is regulated between GTP bound active state and GDI bound inactive state. RhoGDI extracts the inactive form of Rho-GDP complex.

RhoGDI controls the activity of Rho GTPases by three distinct biochemical mechanisms. GDIs were initially found to inhibit the dissociation of GDP from the inactive GTPase by blocking GEF-mediated exchange, thus acting as a negative regulator of Rho proteins<sup>131,138</sup>. Later, it was also shown in Cdc42 that GDIs are also capable of binding to the GTP-bound active state of Rho proteins and block both interactions with the downstream effectors and GAP-catalysed GTP hydrolysis<sup>146</sup>. Structural studies on Rho-GDI complexes suggest that the binding of the N-terminal region of RhoGDI at switch regions of Rho proteins prevents the nucleotide exchange required for GTPase activity. Interestingly, GDI forms a hydrogen bond with Thr35 (cdc42) in the switch-I region which is also required for GEF-catalysed nucleotide exchange.

Similarly, in the case of Rac1, structural studies show that GEF Tiam1 and GDI interact with similar residues in the switch regions to exhibit opposite effects<sup>129</sup>. However, the most crucial function of RhoGDI is the ability to regulate the GTPases between the soluble cytosolic pool and membrane-bound states. RhoGDI plays a unique role by appropriate localisation of Rho GTPases (Fig. 1.4). In the absence of GDI, Rho proteins are attached to the cell membrane through an isoprenyl moiety (geranylgeranyl or farnesyl)<sup>103,147-149</sup>. However, GDI binds to these prenylated forms creating a soluble cytosolic Rho.GDP-RhoGDI complex regulating the cytoplasmic pool of each of the Rho family GTP-binding proteins.

### **Regulation of GTPase-GDI complexes by diverse stimuli/factors**

RhoGDI expression levels in a cell are roughly equal to the total amount of Rho GTPases (RhoA, cdc42, Rac1) combined, suggesting a large amount of GTPase existing as cytosolic GDI complexes. The activation of Rho GTPases mediated by GEF-catalysed nucleotide exchange or GAP-catalyzed GTP hydrolysis requires its release from the RhoGDI complex. Unlike Rab family of GTPases, the existence of conserved GDI dissociation factor (GDF) which would promote dissociation of the RhoGDI-RhoGTPase complex and render the Rho protein active and free<sup>150</sup>, has not been found. Experimental studies highlight different mechanisms for the dissociation of Rho GTPases from the RhoGDI complexes to facilitate the downstream activation process. Early work suggests dissociation or decrease in the affinity of RhoGDI-GTPase complexes by lipids such as phosphatidic acids, phosphoinositides, saturated and unsaturated fatty acids<sup>151,152</sup>. In case of RhoA-GDI complex, phosphoinositides disrupt the complex partially to facilitate GEF-catalyzed nucleotide exchange. Interestingly, experimental studies suggest phosphorylation/acetylation of RhoGDIs by diverse kinases (Pak1, PKA, PKC $\alpha$ ) as key post-translational modifications for the dissociation of the complexes where some kinases act as bonafide GDFs (GDI dissociation factor)<sup>41,130,147,153-156</sup>. However, relatively little is known about the molecular basis of dissociation between specific Rho GTPase and RhoGDIs.

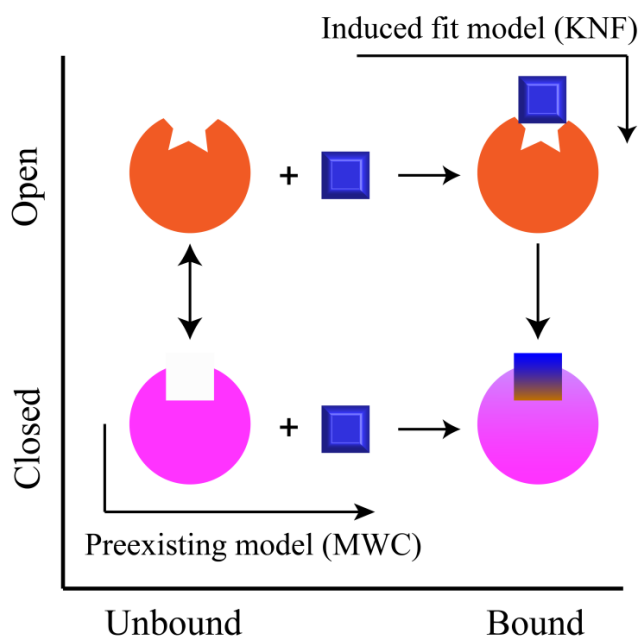
### **1.4 Molecular Basis of Intra-protein Signal Propagation alias ‘Allostery’**

Allosteric regulation of proteins is a process in which the protein activity is modulated through perturbation/stimulus at one site (allosteric site) altering the function/properties of another distant site (orthosteric site)<sup>43,157,158</sup>. The perturbation at the allosteric site can result from a wide range

## 1.4 Molecular Basis of Intra-protein Signal Propagation alias ‘Allostery’

---

of effects such as ligand binding, post-transcriptional modifications, pH and temperature induced changes<sup>159</sup>. Allosteric events involve signal propagation across membranes, cells or large multiple protein complexes which can occur through multiple pathways. Since the discovery of allostery, various models have been proposed to understand and describe the mechanism of allosteric regulation. In the 1960s, two phenomenological models were proposed, namely, the Monod-Wyman-Changeux (MWC) model<sup>160</sup> and the Koshland-Nemethy-Filmer (KNF) model<sup>161</sup>. These models described the allosteric effect in terms of conformational changes of the active site through ligand binding (Fig. 1.5).



**Figure 1.5** Illustrative representation of the difference between the induced fit model (KNF) and conformational selection model (MWC). According to these models, proteins exist in two different predefined conformations (Open and Closed). The ligand binds either to the “Closed” conformation which exists in equilibrium with the Open conformation (MWC), or it binds to the Open conformer and induces conformational transition (KNF).

According to the MWC model, allosteric protein exists in equilibrium between a relaxed (R) conformation and a tensed (T) conformation and can shift between these two states upon ligand binding. It was used to describe allostery in proteins such as GroEL, CheY, and others. On the contrary, KNF model suggests induced-fit mechanism to exhibit the inherent flexibility of the binding site. The model, also known as sequential model, proposes a ligand-induced

conformational change in the protein, which further induces positive or negative cooperativity in the protein (e.g. CAP protein and its ligand c-AMP). For decades, the structure based definition of allostery aimed to elucidate the allosteric mechanism through conformational changes on effector binding. For example, in the case of haemoglobin, with the knowledge of the T and R conformational states, it supported the structure based allostery phenomenon. However, these models do not provide any mechanistic insight into how signal propagation occurs between two distantly situated sites in allosteric proteins. This is because, the biomolecules exist in conformational ensembles rather than explicitly in two states, open and closed as considered by these models. These biomolecules continuously shift between these dynamic ensembles over a free energy landscape, and hence, they must be described statistically and not statically<sup>162</sup>.

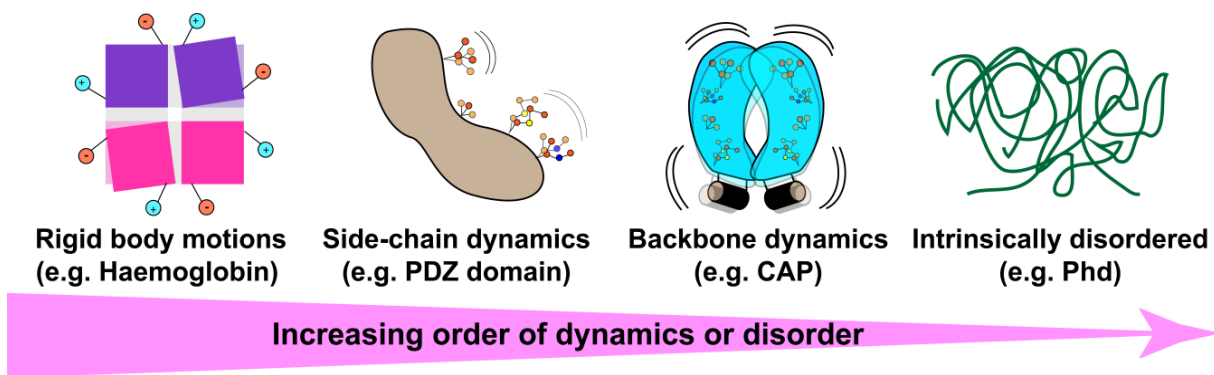
Later in 1984, Cooper and Dryden introduced the term “dynamic allostery” for the allosteric regulation without structural changes breaking the long-established paradigm of structure-based allostery<sup>163</sup>. In an inspiring work, these authors based on the statistical thermodynamics correlate thermal fluctuations and dynamic behaviour with the cooperative binding energies on the order of few kcal/mol, emphasising the role of conformational entropy in allostery. Apparently, the changes in the dynamics linked to the conformational entropy provide potential information about the free energy of protein-ligand association<sup>42,164-166</sup>. This idea established that allostery is a thermodynamically driven process which is governed by both enthalpic and entropic changes<sup>167-172</sup>.

### 1.4.1 Ensemble Nature and Thermodynamic View of Allostery

Solution NMR provides useful insight into the dynamics and characteristic thermodynamics of the protein<sup>173,174</sup>. It highlights the ensemble nature of protein dynamics over a wide range of timescales. These fluctuations range from side chain motions with ps-ns timescales to collective motion of large domains on the  $\mu$ s-ms timescale. The internal motions (thermal fluctuations) are phenomenologically associated with the conformational entropy of the system and provide an indirect measure of the structural ensembles that describes the thermodynamics of the system. NMR analyses have revealed varying degree of changes in dynamics that are associated with the allosteric transitions.

## 1.4 Molecular Basis of Intra-protein Signal Propagation alias 'Allostery'

---

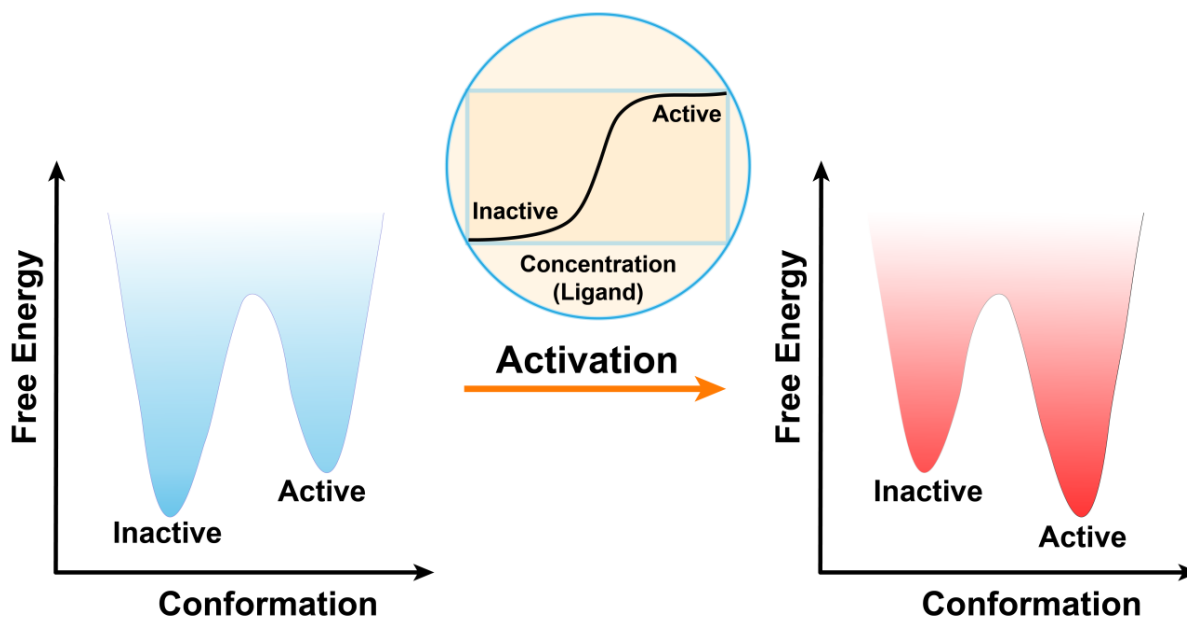


**Figure 1.6** Illustrative representation of dynamic continuum of allostery phenomena that shows the allosteric systems with increasing dynamics or disorder.

For example, in the case of haemoglobin, the allosteric mechanism supports the notion that it can be explained by observable changes in the ensemble average structure. In contrast, CAP dimer and PDZ domain show that the conformational entropy is associated with an allosteric response via backbone and side-chain dynamics, respectively<sup>165,175</sup>. These systems highlight not only the importance of dynamics in allostery but also the limitations of the static view of allosteric mechanisms. Interestingly, intrinsically disordered proteins (IDPs) that lack a proper structure have also been shown to exhibit allostery through disordered to ordered transition. (e.g. Phd/Doc toxin-antitoxin system). This association between different contributions of conformational entropy (measure of dynamics) and degree of structural changes is also described as the “dynamic continuum of allostery” (Fig. 1.6) ranging from ordered rigid body conformational changes (e.g. haemoglobin) to the highest conformational fluctuations in intrinsically disordered proteins<sup>158</sup>.

Protein regulation mechanisms can be best understood in terms of modulation in the free energy landscape where different energy states represent conformational ensembles with a certain distribution. Allosteric regulation has been viewed in terms of population shift or redistribution of conformational states under various conditions such as ligand binding /unbinding, temperature, pH, mutations and other physical or chemical factors<sup>42,176-179</sup>. For an allosteric protein, the conformational states pre-exist for the active, inactive and transition states over a free energy landscape. The population shift stabilises one of these conformational states depending upon the allosteric trigger event (ligand, mutation). Fig. 1.7 represents a simple two-state activation model of allostery which correlates the thermodynamic view of allostery with





**Figure 1.7** Thermodynamic view of allostery via a bi-stable switch. The figure illustrates the free energy landscape with respect to the active and inactive conformational states that are separated by a free energy barrier. Each minima represents an ensemble of conformations, and the depth of the basin determines the relative population in each state. Before activation, the inactive state population dominates the active conformational ensemble. With the increase in ligand concentration, the allostery event is observed as a population shift in favour of the active state.

population shift between active and inactive conformations. Nussinov and co-workers highlight that the extent of the population shift or the free energy difference between conformational states ( $\Delta\Delta G$ ) determines the allosteric efficacy rather than the binding affinity of the ligand to the protein<sup>162</sup>. Recently, a thermodynamic basis for allostery based on the degree of conformational changes ranging from entropy-driven side chain dynamics to the enthalpy-driven large domain motions has been proposed<sup>162</sup>. Allosteric proteins may have multiple pre-existing allosteric pathways and favour one depending on the perturbation events such as ligand binding, mutations, covalent modifications and changes in the cellular physiological conditions<sup>180</sup>. Over last 40 years, experimental techniques and several computational methods have been developed to characterise allostery in proteins; however, a physical understanding of signal propagation between distant sites of an allosteric protein at the molecular level is still lacking.

## 1.4 Molecular Basis of Intra-protein Signal Propagation alias 'Allostery'

---

### 1.4.2 Experimental Methods to Probe Allosteric Mechanism

X-ray crystallography is one of the most widely used experimental techniques to understand major biological processes such as structure-based drug-design, enzyme mechanisms or allostery at the molecular level. Allostery involves conformational changes that can be captured using X-ray crystallography under different conditions such as ligand bound/unbound state, enzyme complexes, and site-directed mutagenesis<sup>181-185</sup>. Majority of the proposed models of allosteric mechanisms (classical MWC and KNF model<sup>160,161</sup>, Cooperon model, TTS model<sup>186</sup>, Szabo and Karplus model<sup>187</sup>, and Lee and Karplus model<sup>188</sup>) are based on the classical example of cooperativity and allostery observed in haemoglobin<sup>189,190</sup>. Most of the results regarding the functional mechanism of haemoglobin are based on the information from the X-ray crystal structures. The allosteric effect in haemoglobin is characterised by the rotation of one  $\alpha\beta$  unit of the dimer by  $15^\circ$  with respect to its partner. Perutz proposed a novel stereochemical mechanism for cooperativity and allostery in haemoglobin in the form of inter- and intra- subunit salt bridges that drives the transition between the low oxygen affinity (*Tensed*, T-state) and high oxygen affinity (*Relax*, R-state) states<sup>191,192</sup>.

Evidently, this classic example shows how charge-charge interactions can govern the equilibrium shift between different conformations. Recently data collections at synchrotron radiation sources have enabled a time-resolved crystallographic analysis to obtain time-dependent structural information of the protein molecule at high resolution as it undergoes a conformational change<sup>193,194</sup>. However, the allosteric effect may not always be defined based on the noticeable conformational changes that can be captured by X-ray crystallography. This is predominantly observed in case of proteins exhibiting allostery with no significant change in the conformation (also called dynamic allostery).

Recent advancements in the NMR methods highlight use of chemical shifts, residual dipolar coupling and relaxation dispersion to probe and characterise equilibrium dynamics in the free and bound states of the allosteric protein<sup>195,196</sup>. Interestingly, allostery associated conformational changes at *ps-ns* time scale can be probed using NMR spin relaxation that captures the fluctuations of bond vectors<sup>197,198</sup>. This method is highly successful in understanding the dynamic allostery using the backbone amide order parameter ( $S^2$ ) in several systems such as cAMP binding in CAP<sup>199</sup>, CRIPT ligand binding in PDZ domain and functional role of  $\alpha 3$  helix

in PDZ3<sup>175</sup>. In addition, biochemical technique such as double-mutant cycle analysis (DMC) has been successfully used to quantitatively determine the energetic connectivity between any two protein residues<sup>200</sup>. This technique can be used to characterise/measure higher order cooperatively between intermolecular interactions. DMC analysis has been extensively used to map energetic networks in PDZ domain in homologous proteins (e.g. PTP-BL PDZ2, SAP97 PDZ2 and PSD-95 PDZ3)<sup>201-204</sup>.

### 1.4.3 Computational Approaches to Probe Allosteric Mechanism

#### Sequence-based Methods

Amino acid sequences in a given protein family are not only constrained by the biochemical functionality but also by the evolutionary history of the proteins. Protein sequence analysis can be a useful tool in detecting allosteric pathways based on evolutionary and energetically coupled sites<sup>205,206</sup>. One of the remarkable methods, statistical coupling analysis (SCA)<sup>207-211</sup> successfully predicted surface sites/residues that are energetically linked to the functional sites in *E. coli* DHFR (dihydrofolate reductase), G-protein<sup>212,213</sup>, GPCRs<sup>214</sup> and PDZ domain. SCA has also been used to engineer a light-sensitive LOV2 domain<sup>215,216</sup>, design an artificial WW domain<sup>217</sup> and predict potential sites for drug design for antigen 85C of *M. tuberculosis*<sup>218</sup> and cathepsin K<sup>219</sup>. However, it has been debated that the energetic coupling based on the statistical analysis (SCA) may not be a true reporter of allostery<sup>201</sup>. Also, sequence-based methods are limited by input sequence selection and required number of sequence for multiple sequence alignment (MSA). Often lowering the stringency of the search parameters to obtain a statistically meaningful number of sequences is susceptible to the increased noise level in MSA. Hence, such methods primarily based on MSA conservation poses a challenging task to determine the specific biological roles for the identified evolutionary significant residue.

#### Topology and feature based predictive methods

The ability to determine the protein structure using X-ray and NMR technique has given rise to the development of structural analysis methods focused on decoding allostery. These methods are based on several intrinsic features such as contact patterns<sup>220</sup>, hydrophobicity, interatomic distances<sup>221</sup>, bond-to-bond propensity<sup>222</sup> and protein-protein interactions for a given protein molecule. The developed algorithms reproduce previously known allostery networks in

## 1.4 Molecular Basis of Intra-protein Signal Propagation alias 'Allostery'

---

DHFR<sup>223</sup>, caspase-1, CheY and H-Ras and CAP (Catabolite activator protein) and further predict key allosteric interactions in a set of 17 other proteins<sup>222</sup>. Thus far, several methods have been developed to identify and characterise allosteric sites in the protein and validated through the high throughput screening by experimental techniques. Allosteric Database (ASD) serves as a comprehensive platform with manually curated data about all the experimentally confirmed allosteric evidence along with allosteric molecules and features and allosteric networks<sup>224</sup>. Based on this incredible data resource of the allosteric mechanisms, feature-based predictive models have been developed<sup>225</sup>, but at the same time, these models suffer from biased application as often the information used for the prediction is limited depending on the available experimental data. However, it is challenging to predict allosteric sites from structure or topological based methods, as in some of the cases, allosteric sites cannot be easily predicted from the apo structure as they may be present in the intermediate states which is difficult to be captured by X-ray/NMR technique. The mechanism of dynamic allostery requires further investigation as the average structure does not change, but the allostery effect is transmitted through changes in dynamics.

### Normal mode analysis based methods

Normal mode analysis is an alternative approach to computationally intensive MD simulation, which assumes that the protein undergoes harmonic fluctuations around energy minimised equilibrium conformation. It has been a powerful method to explore the dynamics of proteins that do not undergo major allosteric transitions and functions close to their native structure. NMA has been coupled with techniques such as coarse grain MD simulations, protein modelling, modification of selective spring constants or incorporation of electrostatic forces to better account for structural dynamics of residues that are likely to participate in the allosteric mechanism<sup>226-230</sup>. Further, an atomistic ensemble NMA approach based on all available crystal structures for a given protein family under study was applied to two allosteric proteins, namely Hemoglobin and Caspase 7, after an improved performance in accurate prediction of residue level dynamic coupling in heterotrimeric  $G\alpha$ <sup>231</sup>. NMA is a computationally inexpensive high-throughput approach that can be automated. Taking advantage of this, a number of NMA method based web servers such as SPACER<sup>232-235</sup>, PARS<sup>236,237</sup>, DynOmics<sup>238</sup> have been used to predict allosteric pathways with its ability to provide global/local modes that imply functional significance. However, it is still doubtful that NMA (ENM model) is sufficiently accurate to

capture the essential features required for predicting allosteric network analysis. The method is focused on examining the harmonic fluctuations of a protein around an energetically minimised conformation, which means that it misses out on large-scale conformational changes involving local unfolding or rigid body movements<sup>158</sup>.

### **Molecular dynamics based methods**

Molecular dynamics (MD) simulation has proven to be an important method that can reproduce/explore local and large-scale conformational changes in the protein. The structural changes associated with the allosteric mechanism have been supported by two major ideas: induced-fit (conformational change) or population-shift (conformational selection). Further, it is defined as the change in the relative population of the preexisting accessible conformational ensembles upon perturbation. In principle, using long molecular dynamics simulations can generate sufficient conformation sampling (ensembles) to understand allostery mechanism<sup>44,239</sup>. Multiple methods such as correlation analysis, mutual information, and ENM based dynamical network are used to identify/evaluate a pathway of coupled residues based on the simulation data<sup>240-244</sup>. However, allostery phenomenon is characterised by motions and conformational changes with a wide range of timescales (ns-ms). Atomistic MD simulation technique is restricted with the timescale limitation to explore the entire energy landscape or significant large-scale conformational changes corresponding to allostery effects. Hence MD simulations are often complemented with the use of enhanced sampling technique, which enables to overcome the free energy barriers and explore the conformational space between the two states.

Various enhanced sampling techniques such as bias-exchange metadynamics (BEMD) simulations, targeted molecular dynamics (TMD), umbrella sampling have been extensively used to explore transition conformations and propose the allosteric activation mechanism in multiple allosteric systems (for example; bovine chymosin upon P8-P4  $\kappa$ -casein or open/close transition in calmodulin) within feasible simulation and computational time<sup>245</sup>. Moreover, classical MD simulation technique for all-atom systems is limited by the system size because the computational cost and the time required are directly proportional to the number of degrees of freedom. As an alternative, Coarse-grained method has enabled the simulation of large systems of realistic size (e.g. Virus envelope, actin filaments) with the reduction in complexity by representing atomic systems in the form of beads with unique parameterisation. The CG models

## 1.4 Molecular Basis of Intra-protein Signal Propagation alias 'Allostery'

---

have been used to understand induced-fit or population shift models of allostery in different protein systems (e.g. glutamate-binding protein, adenylate kinase)<sup>246,247</sup>. This method has been hugely successful in understanding the molecular mechanism and nature of conformational changes that take place during the transmembrane receptor activation (GPCRs, T-cell receptor).

Interestingly, MD simulations play an essential role in understanding the allosteric mechanism in proteins that do not undergo structural change upon perturbation and the information is transmitted to the active site without detectable conformational changes (for example; methionine repressor MetJ, Catabolite activator protein (CAP) and PDZ domain). This phenomenon is also characterised as dynamic allostery. With the development of new methods and analysis algorithm based on molecular dynamics simulations data, multiple structural perturbations sites and propagating pathways have been defined for these systems. A method termed as force distribution analysis which can reveal how a ligand-induced mechanical strain is dissipated through the protein in the form of inter-atomic forces was applied to methionine repressor MetJ and Catabolite activator protein (CAP) to propose well-defined signal propagation pathway<sup>248,249</sup>. Both the studies suggest change in dynamics as the regulatory driving force and the communication pathway largely of entropic nature.

Another typical example of dynamic allostery is PDZ domain where several MD based studies have been performed to identify the allosteric signaling pathway based on structural or energetic fluctuations<sup>24,167,172,175,250-259</sup>. In addition, two non-equilibrium MD methods, namely, PPMD- Pump-probe MD<sup>260</sup> and ATD- anisotropic thermal diffusion<sup>261</sup>, have also been used to study PDZ domain. In the first method, selective residue His76 mimicking the ligand binding effect was excited with a set of oscillating forces and the propagation of the perturbation is probed using Fourier transform of the atomic fluctuations while the second method measures the thermal diffusion of kinetic energy from His76 within PDZ domain. The signaling pathway was elucidated by tracking the most significant thermal diffusion pathway. However, these methods mimic the ligand binding event by perturbing a single residue (His76) which does not correspond to an exact ligand bind effect. Hence the generated dynamic response in the protein may not be the true reporter of the ligand binding event.

### 1.4.4 Dynamic Allostery in PDZ Domain: An Unsolved Puzzle

The existential models for allostery mechanism are based on the structural approach in which the two states of protein must differ significantly so that the conformational ensembles of these states are separated by a large thermodynamic barrier on the energy landscape. However, these models do not explain dynamic allostery where the conformational ensembles exhibit structural changes which are accessible to thermal fluctuations. Despite numerous studies on proteins such as GPCRs, heat shock protein (HSP), MetJ repressor and immune cell receptors, a physical understanding of how communication/signal propagation occurs between the distant sites in allosteric proteins remains elusive. With the identification of allostery in monomeric single-domain, PDZ domain has been a popular model system to study single domain allostery<sup>163,177,262</sup>. PDZ domains are novel modules implicated in the localisation of membrane receptors and ion-channels and are associated with the cellular signal transduction<sup>20,23</sup>. These are ubiquitous protein-protein interacting domains structurally conserved throughout evolution in various species. They can bind to the specific recognition sequences at the c-terminal of proteins<sup>21</sup>, or they can dimerise with other modular protein domains (WW, SH2, SH3, PH, etc.)<sup>20</sup>.

A number of communication networks have been proposed based on key residues identified using both theoretical<sup>256,258,260</sup> and experimental techniques<sup>175,250</sup> to provide insights into allosteric communication in PDZ domains. Such structural perturbations have been studied based on the normal mode analysis, elastic network models and correlation of dynamical changes<sup>251,252,263</sup>. With the evolving paradigm of dynamic allostery, the long range communication between the ligand binding site and the distant regions of in PDZ domain has been described based on entropy changes resulting from dynamics without changing the average position of the structure. Recent studies demonstrate a direct interrelation between the allosteric communications and energy propagation along the pathways<sup>24,264-267</sup>. Interestingly, dynamic allostery can also be explained in terms of population shift between two distinct conformations driven by entropy and/or enthalpy<sup>159,162,180</sup>. A thermodynamic basis of dynamic allostery will provide physical and molecular insight into how the interactions and subsequent dynamics of these conformational ensembles play an important role in signal propagation across a protein.

### 1.5 “Thermodynamics” and “Molecular Insights” into the Biomolecular Signaling

Deciphering the molecular details of the protein-protein interactions has been a crucial and challenging problem in the signal transduction pathways. One of the fundamental aspects of the biological process is the understanding of functional mechanisms in terms of physical interactions, for example, electrostatic complementarity, van der Waals interactions and hydrogen bonds that contribute to the thermodynamic properties of the system. Experimental evidences based on kinetic assays along with the sequence and structural data have been the primary source of information to improve our knowledge and understanding of the functional mechanisms. However, many of these experimental techniques do not provide a direct molecular interaction-based information/picture between the proteins. With the recent development in the computations modelling and simulation techniques, it has been possible to furnish real-time imaging and information of the system at atomistic-level resolution. Unlike experimental methods, molecular dynamics simulations have been useful for *in silico* investigation of biological processes to provide molecular interactions and structural dynamical information.

Biological cells employ a large number of intracellular signaling pathways to regulate their functional activity in response to external/internal stimuli. All these signaling pathways involve the transfer of information either through biomolecular recognition or by allosteric mechanisms. The work in this thesis is directed towards understanding the underlying mechanism of signal transduction pathways at these two different levels of activation and deactivation of signaling proteins. With respect to this, the thesis is organized into seven different chapters.

This chapter (**chapter 1**) provides an overview of the molecular basis of the signal transduction pathways. We review historical developments and evidences that show how conformational ensembles and molecular interactions play a role in biomolecular recognition and allostery, and illustrate using relevant examples, how interactions modulate thermodynamics and structural properties in the signaling process.

**Chapter 2** discusses the basic theory of molecular dynamics simulations, free energy calculation methods and enhanced sampling techniques to explore rare events. In this thesis, we have performed conventional molecular dynamics simulations and metadynamics simulation for



enhanced sampling. This chapter discusses the practical and technical aspects of the methods used to analyse data discussed in the following chapters.

**Chapter 3** investigates the energetic basis of dynamic allostery in PDZ3 domain. Our work provides molecular insights into the internal redistribution and re-wiring of side-chain interactions and accompanied population shift in specific electrostatic interactions that drives the allosteric modulation in PDZ3 domain.

In **Chapter 4**, we examine pH-driven dynamic allostery in PDZ3 domain in terms of change in the electrostatic interaction network identified in our previous work (Chapter 3). Here, we demonstrate that various perturbation events such as ligand binding, or change in cellular conditions (protonation) follow a universal response system in PDZ3 domain in terms of modulation in the dynamics and intra-protein interaction network.

**Chapter 5** investigates the nucleotide dependent conformational signatures in Rho GTPases that facilitates effector recognition. This chapter discusses the energetic balance of molecular interactions between the switch-I region residues with the nucleotide and solvent that leads to preferential stabilisation of some conformational states over other depending on the nucleotide and its subsequent role in effector recognition.

In **Chapter 6**, we extend our investigation to understand the molecular view of the dissociation of Rac1 GTPase and RhoGDI complex. This chapter provides insights into the conformational dynamics and interactions that play an important role in the phosphorylation-mediated regulation of GTPase-GDI complex. Here, we propose a molecular-interaction based mechanistic model for the dissociation of the complex as an effect of phosphorylation.

**Chapter 7** concludes the thesis by highlighting the important findings from the molecular investigation of the protein-protein interactions and allosteric regulation processes. It also discusses the implications of these findings in the field of drug designing.

## 1.6 Bibliography

---

### 1.6 Bibliography

1. Dalous, J., et al., *Reversal of Cell Polarity and Actin-Myosin Cytoskeleton Reorganization under Mechanical and Chemical Stimulation*. *Biophysical Journal*, 2008. 94(3): p. 1063-1074.
2. Artemenko, Y., T.J. Lampert, and P.N. Devreotes, *Moving towards a paradigm: common mechanisms of chemotactic signaling in Dictyostelium and mammalian leukocytes*. *Cellular and Molecular Life Sciences*, 2014. 71(19): p. 3711-3747.
3. Artemenko, Y., et al., *Chemical and mechanical stimuli act on common signal transduction and cytoskeletal networks*. *Proceedings of the National Academy of Sciences*, 2016. 113(47): p. E7500.
4. Stroka, K.M. and K. Konstantopoulos, *Physical Biology in Cancer. 4. Physical cues guide tumor cell adhesion and migration*. *American Journal of Physiology-Cell Physiology*, 2013. 306(2): p. C98-C109.
5. Cortese, B., et al., *Influence of electrotaxis on cell behaviour*. *Integrative Biology*, 2014. 6(9): p. 817-830.
6. Hubbard, S.R. and W.T. Miller, *Receptor tyrosine kinases: mechanisms of activation and signaling*. *Current Opinion in Cell Biology*, 2007. 19(2): p. 117-123.
7. Wieduwilt, M.J. and M.M. Moasser, *The epidermal growth factor receptor family: biology driving targeted therapeutics*. *Cellular and molecular life sciences : CMLS*, 2008. 65(10): p. 1566-1584.
8. Lee, Y., S. Choi, and C. Hyeon, *Communication over the network of binary switches regulates the activation of A2A adenosine receptor*. *PLoS computational biology*, 2015. 11(2): p. e1004044-e1004044.
9. Lefkowitz, R.J. and S.K. Shenoy, *Transduction of Receptor Signals by  $\beta$ -Arrestins*. *Science*, 2005. 308(5721): p. 512.
10. Sah, V.P., et al., *The Role of Rho in G Protein-Coupled Receptor Signal Transduction*. *Annual Review of Pharmacology and Toxicology*, 2000. 40(1): p. 459-489.
11. Lee, M.J. and M.B. Yaffe, *Protein Regulation in Signal Transduction*. *Cold Spring Harbor Perspectives in Biology*, 2016. 8(6).
12. Fauré, J. and M.-C. Dagher, *Interactions between Rho GTPases and Rho GDP dissociation inhibitor (Rho-GDI)*. *Biochimie*, 2001. 83(5): p. 409-414.
13. Schwabe, J.W.R., *The role of water in protein—DNA interactions*. *Current Opinion in Structural Biology*, 1997. 7(1): p. 126-134.
14. Nakamura, H., *Roles of electrostatic interaction in proteins*. *Q Rev Biophys*, 1996. 29(1): p. 1-90.
15. Zhou, H.-X. and X. Pang, *Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation*. *Chemical reviews*, 2018. 118(4): p. 1691-1741.

16. Duan, G. and D. Walther, *The roles of post-translational modifications in the context of protein interaction networks*. PLoS computational biology, 2015. 11(2): p. e1004049-e1004049.
17. Ingham, R.J., et al., *WW domains provide a platform for the assembly of multiprotein networks*. Molecular and cellular biology, 2005. 25(16): p. 7092-7106.
18. Ardito, F., et al., *The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review)*. International journal of molecular medicine, 2017. 40(2): p. 271-280.
19. Zhang, Z., S. Witham, and E. Alexov, *On the role of electrostatics in protein-protein interactions*. Physical biology, 2011. 8(3): p. 035001-035001.
20. Lee, H.J. and J.J. Zheng, *PDZ domains and their binding partners: structure, specificity, and modification*. Cell Commun Signal, 2010. 8: p. 8.
21. Doyle, D.A., et al., *Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ*. Cell, 1996. 85(7): p. 1067-76.
22. Zhang, J., et al., *Crystallographic and nuclear magnetic resonance evaluation of the impact of peptide binding to the second PDZ domain of protein tyrosine phosphatase 1E*. Biochemistry, 2010. 49(43): p. 9280-91.
23. Kim, E. and M. Sheng, *PDZ domain proteins of synapses*. Nat Rev Neurosci, 2004. 5(10): p. 771-81.
24. Ishikura, T., et al., *Energy exchange network of inter-residue interactions within a thermally fluctuating protein molecule: A computational study*. J Comput Chem, 2015. 36(22): p. 1709-18.
25. Harris, B.Z., et al., *Role of Electrostatic Interactions in PDZ Domain Ligand Recognition*. Biochemistry, 2003. 42(10): p. 2797-2805.
26. Ivarsson, Y., *Plasticity of PDZ domains in ligand recognition and signaling*. FEBS Letters, 2012. 586(17): p. 2638-2647.
27. Fuh, G., et al., *Analysis of PDZ Domain-Ligand Interactions Using Carboxyl-terminal Phage Display*. Journal of Biological Chemistry, 2000. 275(28): p. 21486-21491.
28. Hunter, T., *Why nature chose phosphate to modify proteins*. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2012. 367(1602): p. 2513-2516.
29. Uezu, A., et al., *Modified SH2 domain to phototrap and identify phosphotyrosine proteins from subcellular sites within cells*. Proceedings of the National Academy of Sciences, 2012. 109(43): p. E2929.
30. Moran, M.F., et al., *Src homology region 2 domains direct protein-protein interactions in signal transduction*. Proceedings of the National Academy of Sciences of the United States of America, 1990. 87(21): p. 8622-8626.

## 1.6 Bibliography

---

31. Kiel, C., et al., *Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex*. Proceedings of the National Academy of Sciences of the United States of America, 2004. 101(25): p. 9223.
32. Woda, J., et al., *An analysis of the relationship between hydration and protein-DNA interactions*. Biophysical journal, 1998. 75(5): p. 2170-2177.
33. Petrone, P.M. and A.E. Garcia, *MHC–Peptide Binding is Assisted by Bound Water Molecules*. Journal of Molecular Biology, 2004. 338(2): p. 419-435.
34. Sorenson, J.M. and T. Head-Gordon, *The importance of hydration for the kinetics and thermodynamics of protein folding: simplified lattice models*. Folding and Design, 1998. 3(6): p. 523-534.
35. Chilvers, E.R. and T. Sethi, *How receptors work: mechanisms of signal transduction*. Postgraduate medical journal, 1994. 70(829): p. 813-817.
36. Pawson, T. and J.D. Scott, *Signaling Through Scaffold, Anchoring, and Adaptor Proteins*. Science, 1997. 278(5346): p. 2075.
37. Tompa, P., *The principle of conformational signaling*. Chemical Society Reviews, 2016. 45(15): p. 4252-4284.
38. Neel, B.G. and N.K. Tonks, *Protein tyrosine phosphatases in signal transduction*. Current Opinion in Cell Biology, 1997. 9(2): p. 193-204.
39. Cohen, N.A., et al., *Binding of the Inward Rectifier K<sup>+</sup> Channel Kir 2.3 to PSD-95 Is Regulated by Protein Kinase A Phosphorylation*. Neuron, 1996. 17(4): p. 759-767.
40. Mehta, D., A. Rahman, and A.B. Malik, *Protein Kinase C- $\alpha$  Signals Rho-Guanine Nucleotide Dissociation Inhibitor Phosphorylation and Rho Activation and Regulates the Endothelial Cell Barrier Function*. Journal of Biological Chemistry, 2001. 276(25): p. 22614-22620.
41. Dovas, A., et al., *Serine 34 Phosphorylation of Rho Guanine Dissociation Inhibitor (RhoGDI $\alpha$ ) Links Signaling from Conventional Protein Kinase C to RhoGTPase in Cell Adhesion*. Journal of Biological Chemistry, 2010. 285(30): p. 23296-23308.
42. Tsai, C.J., A. del Sol, and R. Nussinov, *Allostery: absence of a change in shape does not imply that allostery is not at play*. J Mol Biol, 2008. 378(1): p. 1-11.
43. Liu, J. and R. Nussinov, *Allostery: An Overview of Its History, Concepts, Methods, and Applications*. PLoS Comput Biol, 2016. 12(6): p. e1004966.
44. Fenwick, R.B., S. Esteban-Martín, and X. Salvatella, *Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles*. European Biophysics Journal, 2011. 40(12): p. 1339-1355.
45. Nussinov, R. and C.-J. Tsai, *The Role of Allostery in the Termination of Second Messenger Signaling*. Biophysical journal, 2015. 109(6): p. 1080-1081.

46. Bissantz, C., *Conformational Changes of G Protein-Coupled Receptors During Their Activation by Agonist Binding*. Journal of Receptors and Signal Transduction, 2003. 23(2-3): p. 123-153.
  47. Milburn, M.V., et al., *Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins*. Science, 1990. 247(4945): p. 939.
  48. Nussinov, R., C.-J. Tsai, and B. Ma, *The Underappreciated Role of Allostery in the Cellular Network*. Annual Review of Biophysics, 2013. 42(1): p. 169-189.
  49. Moogk, D., A. Natarajan, and M. Krogsgaard, *T cell receptor signal transduction: affinity, force and conformational change*. Current Opinion in Chemical Engineering, 2018. 19: p. 43-50.
  50. Syrovatkina, V., et al., *Regulation, Signaling, and Physiological Functions of G-Proteins*. Journal of molecular biology, 2016. 428(19): p. 3850-3868.
  51. Colicelli, J., *Human RAS superfamily proteins and related GTPases*. Science's STKE : signal transduction knowledge environment, 2004. 2004(250): p. RE13-RE13.
  52. Etienne-Manneville, S. and A. Hall, *Rho GTPases in cell biology*. Nature, 2002. 420(6916): p. 629-635.
  53. Hall, A., *Rho family GTPases*. Biochemical Society Transactions, 2012. 40(6): p. 1378.
  54. Sadok, A. and C.J. Marshall, *Rho GTPases: masters of cell migration*. Small GTPases, 2014. 5: p. e29710-e29710.
  55. Nobes, C.D. and A. Hall, *Rho GTPases Control Polarity, Protrusion, and Adhesion during Cell Movement*. The Journal of Cell Biology, 1999. 144(6): p. 1235.
  56. Jaffe, A.B. and A. Hall, *Rho GTPases: biochemistry and biology*. Annu Rev Cell Dev Biol, 2005. 21: p. 247-69.
  57. Wittinghofer, A. and I.R. Vetter, *Structure-function relationships of the G domain, a canonical switch motif*. Annu Rev Biochem, 2011. 80: p. 943-71.
  58. Sit, S.-T. and E. Manser, *Rho GTPases and their role in organizing the actin cytoskeleton*. Journal of cell science, 2011. 124: p. 679-683.
  59. Sahai, E. and C.J. Marshall, *RHO-GTPases and cancer*. Nat Rev Cancer, 2002. 2(2): p. 133-42.
  60. Parri, M. and P. Chiarugi, *Rac and Rho GTPases in cancer cell motility control*. Cell Commun Signal, 2010. 8: p. 23.
  61. Porter, A.P., A. Papaioannou, and A. Malliri, *Deregulation of Rho GTPases in cancer*. Small GTPases, 2016. 7(3): p. 123-138.
  62. Jansen, S., et al., *Paving the Rho in cancer metastasis: Rho GTPases and beyond*. Pharmacology & Therapeutics, 2018. 183: p. 1-21.
  63. Aguilar, B.J., Y. Zhu, and Q. Lu, *Rho GTPases as therapeutic targets in Alzheimer's disease*. Alzheimer's research & therapy, 2017. 9(1): p. 97-97.
-

## 1.6 Bibliography

---

64. Schmidt, G., et al., *Gln 63 of Rho is deamidated by Escherichia coli cytotoxic necrotizing factor-1*. *Nature*, 1997. 387(6634): p. 725-729.
65. Aktories, K., *Bacterial protein toxins that modify host regulatory GTPases*. *Nat Rev Microbiol*, 2011. 9(7): p. 487-98.
66. Lemichez, E. and K. Aktories, *Hijacking of Rho GTPases during bacterial infection*. *Experimental Cell Research*, 2013. 319(15): p. 2329-2336.
67. Abdul Azeez, Kamal R., et al., *The crystal structure of the RhoA-AKAP-Lbc DH-PH domain complex*. *Biochemical Journal*, 2014. 464(2): p. 231-239.
68. Dias, S.M.G. and R.A. Cerione, *X-ray Crystal Structures Reveal Two Activated States for RhoC*. *Biochemistry*, 2007. 46(22): p. 6547-6558.
69. Ihara, K., et al., *Crystal structure of human RhoA in a dominantly active form complexed with a GTP analogue*. *J Biol Chem*, 1998. 273(16): p. 9656-66.
70. Miyazaki, K., S. Komatsu, and M. Ikebe, *Dynamics of RhoA and ROK $\alpha$  translocation in single living cells*. *Cell Biochemistry and Biophysics*, 2006. 45(3): p. 243-254.
71. Mosteller, R.D., J. Han, and D. Broek, *Identification of residues of the H-ras protein critical for functional interaction with guanine nucleotide exchange factors*. *Molecular and Cellular Biology*, 1994. 14(2): p. 1104-1112.
72. Segal, M., et al., *Two Distinct Regions of Ras Participate in Functional Interaction with GDP-GTP Exchangers*. *European Journal of Biochemistry*, 1995. 228(1): p. 96-101.
73. Créchet, J.-B., A. Bernardi, and A. Parmeggiani, *Distal Switch II Region of Ras2p Is Required for Interaction with Guanine Nucleotide Exchange Factor*. *Journal of Biological Chemistry*, 1996. 271(29): p. 17234-17240.
74. Quilliam, L.A., et al., *Involvement of the Switch 2 Domain of Ras in Its Interaction with Guanine Nucleotide Exchange Factors*. *Journal of Biological Chemistry*, 1996. 271(19): p. 11076-11082.
75. Day, G.-J., R.D. Mosteller, and D. Broek, *Distinct Subclasses of Small GTPases Interact with Guanine Nucleotide Exchange Factors in a Similar Manner*. *Molecular and Cellular Biology*, 1998. 18(12): p. 7444-7454.
76. Cherfils, J. and M. Zeghouf, *Regulation of Small GTPases by GEFs, GAPs, and GDIs*. *Physiological Reviews*, 2013. 93(1): p. 269.
77. Eliáš, M. and V. Klimeš, *Rho GTPases: Deciphering the Evolutionary History of a Complex Protein Family*, in *Rho GTPases: Methods and Protocols*, F. Rivero, Editor. 2012, Springer New York: New York, NY. p. 13-34.
78. Wennerberg, K. and C.J. Der, *Rho-family GTPases: it's not only Rac and Rho (and I like it)*. *Journal of Cell Science*, 2004. 117(8): p. 1301.
79. Narumiya, S. and D. Thumkeo, *Rho signaling research: history, current status and future directions*. *FEBS letters*, 2018. 592(11): p. 1763-1776.

80. Wheeler, A.P. and A.J. Ridley, *Why three Rho proteins? RhoA, RhoB, RhoC, and cell motility*. Experimental Cell Research, 2004. 301(1): p. 43-49.
  81. Lowy, D.R. and B.M. Willumsen, *Function and Regulation of Ras*. Annual Review of Biochemistry, 1993. 62(1): p. 851-891.
  82. Garcia-Mata, R., E. Boulter, and K. Burridge, *The 'invisible hand': regulation of RHO GTPases by RHOGDIs*. Nat Rev Mol Cell Biol, 2011. 12(8): p. 493-504.
  83. Hall, B.E., D. Bar-Sagi, and N. Nassar, *The structural basis for the transition from Ras-GTP to Ras-GDP*. Proceedings of the National Academy of Sciences of the United States of America, 2002. 99: p. 12138-12142.
  84. Prakash, P., A. Sayyed-Ahmad, and A.A. Gorfe, *The Role of Conserved Waters in Conformational Transitions of Q61H K-ras*. PLOS Computational Biology, 2012. 8(2): p. e1002394.
  85. Scheffzek, K., M.R. Ahmadian, and A. Wittinghofer, *GTPase-activating proteins: helping hands to complement an active site*. Trends in Biochemical Sciences, 1998. 23(7): p. 257-262.
  86. Snyder, J.T., et al., *Structural basis for the selective activation of Rho GTPases by Dbl exchange factors*. Nature Structural Biology, 2002. 9(6): p. 468-475.
  87. Rossman, K.L., C.J. Der, and J. Sondek, *GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors*. Nature Reviews Molecular Cell Biology, 2005. 6(2): p. 167-180.
  88. Chimini, G. and P. Chavrier, *Function of Rho family proteins in actin dynamics during phagocytosis and engulfment*. Nature Cell Biology, 2000. 2(10): p. E191-E196.
  89. Hall, A., *Rho GTPases and the Actin Cytoskeleton*. Science, 1998. 279(5350): p. 509.
  90. Liu, M., et al., *Rho GTPase regulation by miRNAs and covalent modifications*. Trends in cell biology, 2012. 22(7): p. 365-373.
  91. Lin, Y. and Y. Zheng, *Approaches of targeting Rho GTPases in cancer drug discovery*. Expert opinion on drug discovery, 2015. 10(9): p. 991-1010.
  92. Karnoub, A.E., et al., *Molecular Basis for Rho GTPase Signaling Specificity*. Breast Cancer Research and Treatment, 2004. 84(1): p. 61-71.
  93. Olson, M.F., *Rho GTPases, their post-translational modifications, disease-associated mutations and pharmacological inhibitors*. Small GTPases, 2018. 9(3): p. 203-215.
  94. Schoentaube, J., et al., *Serine-71 phosphorylation of Rac1/Cdc42 diminishes the pathogenic effect of Clostridium difficile toxin A*. Cellular Microbiology, 2009. 11(12): p. 1816-1826.
  95. Kwon, T., et al., *Akt Protein Kinase Inhibits Rac1-GTP Binding through Phosphorylation at Serine 71 of Rac1*. Journal of Biological Chemistry, 2000. 275(1): p. 423-428.
-

## 1.6 Bibliography

---

96. Trost, M., et al., *Posttranslational regulation of self-renewal capacity: insights from proteome and phosphoproteome analyses of stem cell leukemia*. *Blood*, 2012. 120(8): p. e17.
97. Tu, S., et al., *Epidermal Growth Factor-dependent Regulation of Cdc42 Is Mediated by the Src Tyrosine Kinase*. *Journal of Biological Chemistry*, 2003. 278(49): p. 49293-49300.
98. Chang, F., et al., *Tyrosine Phosphorylation of Rac1: A Role in Regulation of Cell Spreading*. *PLOS ONE*, 2011. 6(12): p. e28587.
99. Torrino, S., et al., *The E3 Ubiquitin-Ligase HACE1 Catalyzes the Ubiquitylation of Active Rac1*. *Developmental Cell*, 2011. 21(5): p. 959-965.
100. Dong, S., et al., *F-box protein complex FBXL19 regulates TGF $\beta$ 1-induced E-cadherin down-regulation by mediating Rac3 ubiquitination and degradation*. *Molecular Cancer*, 2014. 13(1): p. 76.
101. Chen, Y., et al., *Cullin Mediates Degradation of RhoA through Evolutionarily Conserved BTB Adaptors to Control Actin Cytoskeleton Structure and Cell Movement*. *Molecular Cell*, 2009. 35(6): p. 841-855.
102. Wei, J., et al., *A new mechanism of RhoA ubiquitination and degradation: Roles of SCFFBXL19 E3 ligase and Erk2*. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2013. 1833(12): p. 2757-2764.
103. Casey, P.J. and M.C. Seabra, *Protein Prenyltransferases*. *Journal of Biological Chemistry*, 1996. 271(10): p. 5289-5292.
104. Winter-Vann, A.M. and P.J. Casey, *Post-prenylation-processing enzymes as new targets in oncogenesis*. *Nature Reviews Cancer*, 2005. 5(5): p. 405-412.
105. Boyartchuk, V.L., M.N. Ashby, and J. Rine, *Modulation of Ras and a-Factor Function by Carboxyl-Terminal Proteolysis*. *Science*, 1997. 275(5307): p. 1796.
106. Cushman, I. and P.J. Casey, *RHO methylation matters: a role for isoprenylcysteine carboxylmethyltransferase in cell migration and adhesion*. *Cell adhesion & migration*, 2011. 5(1): p. 11-15.
107. Dai, Q., et al., *Mammalian Prenylcysteine Carboxyl Methyltransferase Is in the Endoplasmic Reticulum*. *Journal of Biological Chemistry*, 1998. 273(24): p. 15030-15034.
108. ten Klooster, J.P. and P.L. Hordijk, *Targeting and localized signalling by small GTPases*. *Biology of the Cell*, 2007. 99(1): p. 1-12.
109. Ellerbroek, S.M., K. Wennerberg, and K. Burridge, *Serine Phosphorylation Negatively Regulates RhoA in Vivo*. *Journal of Biological Chemistry*, 2003. 278(21): p. 19023-19031.
110. Dovas, A. and J.R. Couchman, *RhoGDI: multiple functions in the regulation of Rho family GTPase activities*. *Biochem J*, 2005. 390(Pt 1): p. 1-9.



111. Muraoka, S., et al., *Crystal structures of the state 1 conformations of the GTP-bound H-Ras protein and its oncogenic G12V and Q61L mutants*. FEBS Letters, 2012. 586(12): p. 1715-1718.
112. Spoerner, M., et al., *Dynamic properties of the Ras switch I region and its importance for binding to effectors*. Proceedings of the National Academy of Sciences of the United States of America, 2001. 98: p. 4944-4949.
113. Lapouge, K., et al., *Structure of the TPR Domain of p67phox in Complex with Rac·GTP*. Molecular Cell, 2000. 6(4): p. 899-907.
114. Maesaki, R., et al., *The Structural Basis of Rho Effector Recognition Revealed by the Crystal Structure of Human RhoA Complexed with the Effector Domain of PKN/PRK1*. Molecular Cell, 1999. 4(5): p. 793-803.
115. Fujisawa, K., et al., *Different Regions of Rho Determine Rho-selective Binding of Different Classes of Rho Target Molecules*. Journal of Biological Chemistry, 1998. 273(30): p. 18943-18949.
116. Diekmann, D., et al., *Rac GTPase interacts with GAPs and target proteins through multiple effector sites*. The EMBO Journal, 1995. 14(21): p. 5297-5305.
117. Westwick, J.K., et al., *Rac regulation of transformation, gene expression, and actin organization by multiple, PAK-independent pathways*. Molecular and cellular biology, 1997. 17(3): p. 1324-1335.
118. Burbelo, P.D., D. Drechsel, and A. Hall, *A Conserved Binding Motif Defines Numerous Candidate Target Proteins for Both Cdc42 and Rac GTPases*. Journal of Biological Chemistry, 1995. 270(49): p. 29071-29074.
119. Morreale, A., et al., *Structure of Cdc42 bound to the GTPase binding domain of PAK*. Nature Structural Biology, 2000. 7(5): p. 384-388.
120. Sahai, E., A.S. Alberts, and R. Treisman, *RhoA effector mutants reveal distinct effector pathways for cytoskeletal reorganization, SRF activation and transformation*. The EMBO Journal, 1998. 17(5): p. 1350-1361.
121. Bishop, A.L. and A. Hall, *Rho GTPases and their effector proteins*. The Biochemical journal, 2000. 348 Pt 2(Pt 2): p. 241-255.
122. Herrmann, C., G.A. Martin, and A. Wittinghofer, *Quantitative Analysis of the Complex between p21 and the Ras-binding Domain of the Human Raf-1 Protein Kinase*. Journal of Biological Chemistry, 1995. 270(7): p. 2901-2905.
123. Gorman, C., et al., *Equilibrium and Kinetic Measurements Reveal Rapidly Reversible Binding of Ras to Raf*. Journal of Biological Chemistry, 1996. 271(12): p. 6713-6719.
124. Boettner, B., C. Herrmann, and L. Van Aelst, *Ras and rap 1 interaction with af-6 effector target*, in *Methods in Enzymology*. 2001, Academic Press. p. 151-168.
125. Wohlgemuth, S., et al., *Recognizing and Defining True Ras Binding Domains I: Biochemical Analysis*. Journal of Molecular Biology, 2005. 348(3): p. 741-758.

## 1.6 Bibliography

---

126. Nakhaeizadeh, H., et al., *The RAS-Effector Interface: Isoform-Specific Differences in the Effector Binding Regions*. PloS one, 2016. 11(12): p. e0167145-e0167145.
127. Rudolph, M.G., et al., *Thermodynamics of Ras/Effector and Cdc42/Effector Interactions Probed by Isothermal Titration Calorimetry*. Journal of Biological Chemistry, 2001. 276(26): p. 23914-23921.
128. DerMardirossian, C. and G.M. Bokoch, *GDI: central regulatory molecules in Rho GTPase activation*. Trends Cell Biol, 2005. 15(7): p. 356-63.
129. Robbe, K., et al., *Dissociation of GDP dissociation inhibitor and membrane translocation are required for efficient activation of Rac by the Dbl homology-pleckstrin homology region of Tiam*. J Biol Chem, 2003. 278(7): p. 4756-62.
130. DerMardirossian, C., A. Schnelzer, and G.M. Bokoch, *Phosphorylation of RhoGDI by Pak1 mediates dissociation of Rac GTPase*. Mol Cell, 2004. 15(1): p. 117-27.
131. Ueda, T., et al., *Purification and characterization from bovine brain cytosol of a novel regulatory protein inhibiting the dissociation of GDP from and the subsequent binding of GTP to rhoB p20, a ras p21-like GTP-binding protein*. Journal of Biological Chemistry, 1990. 265(16): p. 9373-9380.
132. Lelias, J.M., et al., *cDNA cloning of a human mRNA preferentially expressed in hematopoietic cells and with homology to a GDP-dissociation inhibitor for the rho GTP-binding proteins*. Proceedings of the National Academy of Sciences of the United States of America, 1993. 90(4): p. 1479-1483.
133. Scherle, P., T. Behrens, and L.M. Staudt, *Ly-GDI, a GDP-dissociation inhibitor of the RhoA GTP-binding protein, is expressed preferentially in lymphocytes*. Proceedings of the National Academy of Sciences, 1993. 90(16): p. 7568.
134. Zalcman, G., et al., *RhoGDI-3 Is a New GDP Dissociation Inhibitor (GDI): identification of a non-cytosolic GDI protein interacting with the small GTP-binding proteins RhoB AND RhoG*. Journal of Biological Chemistry, 1996. 271(48): p. 30366-30374.
135. Adra, C.N., et al., *RhoGDIgamma: a GDP-dissociation inhibitor for Rho proteins with preferential expression in brain and pancreas*. Proceedings of the National Academy of Sciences of the United States of America, 1997. 94(9): p. 4279-4284.
136. Wu, Y., et al., *Src phosphorylation of RhoGDI2 regulates its metastasis suppressor function*. Proceedings of the National Academy of Sciences of the United States of America, 2009. 106(14): p. 5807-5812.
137. Harding, M.A. and D. Theodorescu, *RhoGDI signaling provides targets for cancer therapy*. European Journal of Cancer, 2010. 46(7): p. 1252-1259.
138. Leonard, D., et al., *The identification and characterization of a GDP-dissociation inhibitor (GDI) for the CDC42Hs protein*. Journal of Biological Chemistry, 1992. 267(32): p. 22860-22868.

139. Regazzi, R., et al., *Characterization of small-molecular-mass guanine-nucleotide-binding regulatory proteins in insulin-secreting cells and PC12 cells*. European Journal of Biochemistry, 1992. 208(3): p. 729-737.
140. Hiraoka, K., et al., *Both stimulatory and inhibitory GDPGTP exchange proteins, smg GDS and rho GDI, are active on multiple small GTP-binding proteins*. Biochemical and Biophysical Research Communications, 1992. 182(2): p. 921-930.
141. Ando, S., et al., *Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase*. Journal of Biological Chemistry, 1992. 267(36): p. 25709-25713.
142. Matos, P., J.G. Collard, and P. Jordan, *Tumor-related Alternatively Spliced Rac1b Is Not Regulated by Rho-GDP Dissociation Inhibitors and Exhibits Selective Downstream Signaling*. Journal of Biological Chemistry, 2003. 278(50): p. 50442-50448.
143. Michaelson, D., et al., *Differential localization of Rho GTPases in live cells: regulation by hypervariable regions and RhoGDI binding*. The Journal of cell biology, 2001. 152(1): p. 111-126.
144. Fiegen, D., et al., *Crystal structure of Rnd3/RhoE: functional implications I*. FEBS Letters, 2002. 525(1-3): p. 100-104.
145. Forget, M.-A., et al., *Phosphorylation states of Cdc42 and RhoA regulate their interactions with Rho GDP dissociation inhibitor and their extraction from biological membranes*. The Biochemical journal, 2002. 361(Pt 2): p. 243-254.
146. Hart, M.J., et al., *A GDP dissociation inhibitor that serves as a GTPase inhibitor for the Ras-like protein CDC42Hs*. Science, 1992. 258(5083): p. 812.
147. Tnimov, Z., et al., *Quantitative analysis of prenylated RhoA interaction with its chaperone, RhoGDI*. J Biol Chem, 2012. 287(32): p. 26549-62.
148. Cox, A.D. and C.J. Der, *Protein prenylation: more than just glue?* Current Opinion in Cell Biology, 1992. 4(6): p. 1008-1016.
149. Wong, K.W., S. Mohammadi, and R.R. Isberg, *Disruption of RhoGDI and RhoA regulation by a Rac1 specificity switch mutant*. J Biol Chem, 2006. 281(52): p. 40379-88.
150. Sivars, U., D. Aivazian, and S.R. Pfeffer, *Yip3 catalyses the dissociation of endosomal Rab-GDI complexes*. Nature, 2003. 425(6960): p. 856-859.
151. Fauré, J., P.V. Vignais, and M.-C. Dagher, *Phosphoinositide-dependent activation of Rho A involves partial opening of the RhoA/Rho-GDI complex*. European Journal of Biochemistry, 1999. 262(3): p. 879-889.
152. Chuang, T.H., B.P. Bohl, and G.M. Bokoch, *Biologically active lipids are regulators of Rac.GDI complexation*. Journal of Biological Chemistry, 1993. 268(35): p. 26206-26211.
153. Qiao, J., et al., *Phosphorylation of GTP dissociation inhibitor by PKA negatively regulates RhoA*. American Journal of Physiology-Cell Physiology, 2008. 295(5): p. C1161-C1168.

## 1.6 Bibliography

---

154. DerMardirossian, C., et al., *Phosphorylation of RhoGDI by Src regulates Rho GTPase binding and cytosol-membrane cycling*. *Molecular biology of the cell*, 2006. 17(11): p. 4760-4768.
155. Kuhlmann, N., et al., *RhoGDIalpha Acetylation at K127 and K141 Affects Binding toward Nonprenylated RhoA*. *Biochemistry*, 2016. 55(2): p. 304-12.
156. Knezevic, N., et al., *GDI-1 phosphorylation switch at serine 96 induces RhoA activation and increased endothelial permeability*. *Mol Cell Biol*, 2007. 27(18): p. 6323-33.
157. Laskowski, R.A., F. Gerick, and J.M. Thornton, *The structural basis of allosteric regulation in proteins*. *FEBS Lett*, 2009. 583(11): p. 1692-8.
158. Motlagh, H.N., et al., *The ensemble nature of allostery*. *Nature*, 2014. 508(7496): p. 331-9.
159. Nussinov, R., C.J. Tsai, and J. Liu, *Principles of allosteric interactions in cell signaling*. *J Am Chem Soc*, 2014. 136(51): p. 17692-701.
160. Monod, J., J. Wyman, and J.P. Changeux, *On the Nature of Allosteric Transitions: A Plausible Model*. *J Mol Biol*, 1965. 12(1): p. 88-118.
161. Koshland, D.E., Jr., G. Nemethy, and D. Filmer, *Comparison of experimental binding data and theoretical models in proteins containing subunits*. *Biochemistry*, 1966. 5(1): p. 365-85.
162. Tsai, C.J. and R. Nussinov, *A unified view of "how allostery works"*. *PLoS Comput Biol*, 2014. 10(2): p. e1003394.
163. Cooper, A. and D.T.F. Dryden, *Allostery without conformational change*. *European Biophysics Journal*, 1984. 11(2): p. 103-109.
164. Gunasekaran, K., B. Ma, and R. Nussinov, *Is allostery an intrinsic property of all dynamic proteins?* *Proteins: Structure, Function, and Bioinformatics*, 2004. 57(3): p. 433-443.
165. Popovych, N., et al., *Dynamically driven protein allostery*. *Nat Struct Mol Biol*, 2006. 13(9): p. 831-838.
166. Bu, Z. and D.J. Callaway, *Proteins move! Protein dynamics and long-range allostery in cell signaling*. *Adv Protein Chem Struct Biol*, 2011. 83: p. 163-221.
167. Basdevant, N., H. Weinstein, and M. Ceruso, *Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study*. *J Am Chem Soc*, 2006. 128(39): p. 12766-77.
168. Frederick, K.K., et al., *Conformational entropy in molecular recognition by proteins*. *Nature*, 2007. 448(7151): p. 325-9.
169. Ferreiro, D.U., et al., *On the role of frustration in the energy landscapes of allosteric proteins*. *Proc Natl Acad Sci U S A*, 2011. 108(9): p. 3499-503.

170. Di Russo, N.V., M.A. Martí, and A.E. Roitberg, *Underlying Thermodynamics of pH-Dependent Allostery*. The Journal of Physical Chemistry B, 2014. 118(45): p. 12818-12826.
171. Ribeiro, A.A. and V. Ortiz, *Energy propagation and network energetic coupling in proteins*. J Phys Chem B, 2015. 119(5): p. 1835-46.
172. Kalescky, R., et al., *Rigid Residue Scan Simulations Systematically Reveal Residue Entropic Roles in Protein Allostery*. PLoS Comput Biol, 2016. 12(4): p. e1004893.
173. Lipchock, J.M. and J.P. Loria, *Monitoring Molecular Interactions by NMR*, in *Protein Structure, Stability, and Interactions*, J.W. Shriver, Editor. 2009, Humana Press: Totowa, NJ. p. 115-134.
174. Palmer, A.G., C.D. Kroenke, and J. Patrick Loria, *[10] - Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules*, in *Methods in Enzymology*, T.L. James, V. Dötsch, and U. Schmitz, Editors. 2001, Academic Press. p. 204-238.
175. Petit, C.M., et al., *Hidden dynamic allostery in a PDZ domain*. Proc Natl Acad Sci U S A, 2009. 106(43): p. 18249-54.
176. Cui, Q. and M. Karplus, *Allostery and cooperativity revisited*. Protein Sci, 2008. 17(8): p. 1295-307.
177. Volkman, B.F., et al., *Two-state allosteric behavior in a single-domain signaling protein*. Science, 2001. 291(5512): p. 2429-33.
178. Weinkam, P., J. Pons, and A. Sali, *Structure-based model of allostery predicts coupling between distant sites*. Proc Natl Acad Sci U S A, 2012. 109(13): p. 4875-80.
179. Ellis, C.R. and J. Shen, *pH-Dependent Population Shift Regulates BACE1 Activity and Inhibition*. Journal of the American Chemical Society, 2015. 137(30): p. 9543-9546.
180. del Sol, A., et al., *The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways*. Structure (London, England : 1993), 2009. 17(8): p. 1042-1050.
181. Thompson, J.R., et al., *Vmax Regulation through Domain and Subunit Changes. The Active Form of Phosphoglycerate Dehydrogenase*. Biochemistry, 2005. 44(15): p. 5763-5773.
182. Bauman, J.D., et al., *Detecting Allosteric Sites of HIV-1 Reverse Transcriptase by X-ray Crystallographic Fragment Screening*. Journal of Medicinal Chemistry, 2013. 56(7): p. 2738-2746.
183. Liu, X., et al., *Mechanism of intracellular allosteric  $\beta(2)AR$  antagonist revealed by X-ray crystal structure*. Nature, 2017. 548(7668): p. 480-484.
184. Brecher, M., et al., *A conformational switch high-throughput screening assay and allosteric inhibition of the flavivirus NS2B-NS3 protease*. PLoS Pathogens, 2017. 13(5): p. e1006411.
185. Wellington, S., et al., *A small-molecule allosteric inhibitor of Mycobacterium tuberculosis tryptophan synthase*. Nature Chemical Biology, 2017. 13: p. 943.

## 1.6 Bibliography

---

186. Henry, E.R., et al., *A tertiary two-state allosteric model for hemoglobin*. Biophysical Chemistry, 2002. 98(1): p. 149-164.
187. Szabo, A. and M. Karplus, *A mathematical model for structure-function relations in hemoglobin*. Journal of Molecular Biology, 1972. 72(1): p. 163-197.
188. Lee, A.W. and M. Karplus, *Structure-specific model of hemoglobin cooperativity*. Proceedings of the National Academy of Sciences of the United States of America, 1983. 80(23): p. 7055-7059.
189. Eaton, W.A., et al., *Evolution of allosteric models for hemoglobin*. IUBMB Life, 2007. 59(8-9): p. 586-599.
190. Yuan, Y., et al., *New look at hemoglobin allostery*. Chemical reviews, 2015. 115(4): p. 1702-1724.
191. Bettati, S., A. Mozzarelli, and M.F. Perutz, *Allosteric mechanism of haemoglobin: rupture of salt-bridges raises the oxygen affinity of the T-structure*. Journal of Molecular Biology, 1998. 281(4): p. 581-585.
192. Perutz, M.F., *Stereochemistry of Cooperative Effects in Haemoglobin: Haem-Haem Interaction and the Problem of Allostery*. Nature, 1970. 228(5273): p. 726-734.
193. Šrajer, V., et al., *Photolysis of the Carbon Monoxide Complex of Myoglobin: Nanosecond Time-Resolved Crystallography*. Science, 1996. 274(5293): p. 1726.
194. Knapp, J.E., et al., *Allosteric action in real time: Time-resolved crystallographic studies of a cooperative dimeric hemoglobin*. Proceedings of the National Academy of Sciences, 2006. 103(20): p. 7649.
195. Grutsch, S., S. Brüschweiler, and M. Tollinger, *NMR Methods to Study Dynamic Allostery*. PLOS Computational Biology, 2016. 12(3): p. e1004620.
196. Palazzesi, F., et al., *The allosteric communication pathways in KIX domain of CBP*. Proceedings of the National Academy of Sciences of the United States of America, 2013. 110(35): p. 14237-14242.
197. Frederick, K.K., et al., *Conformational entropy in molecular recognition by proteins: A first example from the calmodulin system*. Nature, 2007. 448(7151): p. 325-329.
198. Stone, M.J., *NMR Relaxation Studies of the Role of Conformational Entropy in Protein Stability and Ligand Binding*. Accounts of Chemical Research, 2001. 34(5): p. 379-388.
199. Tzeng, S.-R. and C.G. Kalodimos, *Dynamic activation of an allosteric regulatory protein*. Nature, 2009. 462: p. 368.
200. Horovitz, A., *Double-mutant cycles: a powerful tool for analyzing protein structure and function*. Folding and Design, 1996. 1(6): p. R121-R126.
201. Chi, C.N., et al., *Reassessing a sparse energetic network within a single protein domain*. Proc Natl Acad Sci U S A, 2008. 105(12): p. 4679-84.
202. Hultqvist, G., et al., *Energetic Pathway Sampling in a Protein Interaction Domain*. Structure, 2013. 21(7): p. 1193-1202.

203. Gianni, S., et al., *Sequence-specific Long Range Networks in PSD-95/Discs Large/ZO-1 (PDZ) Domains Tune Their Binding Selectivity*. Journal of Biological Chemistry, 2011. 286(31): p. 27167-27175.
204. Gautier, C., et al., *Seeking allosteric networks in PDZ domains*. Protein Engineering, Design and Selection, 2019. 31(10): p. 367-373.
205. Zhang, S.W., et al., *Estimating residue evolutionary conservation by introducing von Neumann entropy and a novel gap-treating approach*. Amino Acids, 2008. 35(2): p. 495-501.
206. Lichtarge, O., H.R. Bourne, and F.E. Cohen, *An Evolutionary Trace Method Defines Binding Surfaces Common to Protein Families*. Journal of Molecular Biology, 1996. 257(2): p. 342-358.
207. Lockless, S.W. and R. Ranganathan, *Evolutionarily conserved pathways of energetic connectivity in protein families*. Science, 1999. 286(5438): p. 295-9.
208. Gasper, P.M., et al., *Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities*. Proc Natl Acad Sci U S A, 2012. 109(52): p. 21216-22.
209. Van Wart, A.T., et al., *Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis*. J Chem Theory Comput, 2014. 10(2): p. 511-517.
210. Du, Q.-S., et al., *Correlation Analysis for Protein Evolutionary Family Based on Amino Acid Position Mutations and Application in PDZ Domain*. PLOS ONE, 2010. 5(10): p. e13207.
211. Dima, R.I. and D. Thirumalai, *Determination of network of residues that regulate allostery in protein families using sequence analysis*. Protein Science : A Publication of the Protein Society, 2006. 15(2): p. 258-268.
212. Suel, G.M., et al., *Evolutionarily conserved networks of residues mediate allosteric communication in proteins*. Nat Struct Biol, 2003. 10(1): p. 59-69.
213. Sayar, K., et al., *Exploring allosteric coupling in the  $\alpha$ -subunit of Heterotrimeric G proteins using evolutionary and ensemble-based approaches*. BMC Structural Biology, 2008. 8: p. 23-23.
214. Nichols, S.E., et al., *Structure-based network analysis of an evolved G protein-coupled receptor homodimer interface*. Protein Science : A Publication of the Protein Society, 2013. 22(6): p. 745-754.
215. Reynolds, Kimberly A., Richard N. McLaughlin, and R. Ranganathan, *Hot Spots for Allosteric Regulation on Protein Surfaces*. Cell, 2011. 147(7): p. 1564-1575.
216. Socolich, M., et al., *Evolutionary information for specifying a protein fold*. Nature, 2005. 437: p. 512.
217. Russ, W.P., et al., *Natural-like function in artificial WW domains*. Nature, 2005. 437: p. 579.

## 1.6 Bibliography

---

218. Baths, V. and U. Roy, *Identification of distant co-evolving residues in antigen 85C from Mycobacterium tuberculosis using statistical coupling analysis of the esterase family proteins*. Journal of Biomedical Research, 2011. 25(3): p. 165-169.
219. Novinec, M., et al., *A novel allosteric mechanism in the cysteine peptidase cathepsin K discovered by computational methods*. Nature Communications, 2014. 5: p. 3287.
220. van den Bedem, H., et al., *Automated identification of functional dynamic contact networks from X-ray crystallography*. Nature Methods, 2013. 10: p. 896.
221. Greener, J.G., I. Filippis, and M.J.E. Sternberg, *Predicting Protein Dynamics and Allostery Using Multi-Protein Atomic Distance Constraints*. Structure(London, England:1993), 2017. 25(3): p. 546-558.
222. Amor, B.R.C., et al., *Prediction of allosteric sites and mediating interactions through bond-to-bond propensities*. Nature Communications, 2016. 7: p. 12477.
223. Boehr, D.D., et al., *A Distal Mutation Perturbs Dynamic Amino Acid Networks in Dihydrofolate Reductase*. Biochemistry, 2013. 52(27): p. 4605-4619.
224. Shen, Q., et al., *ASD v3.0: unraveling allosteric regulation with structural mechanisms and biological networks*. Nucleic Acids Research, 2016. 44(D1): p. D527-D535.
225. Huang, W., et al., *Allosite: a method for predicting allosteric sites*. Bioinformatics, 2013. 29(18): p. 2357-2359.
226. Martin, D.R. and D.V. Matyushov, *Solvated dissipative electro-elastic network model of hydrated proteins*. The Journal of Chemical Physics, 2012. 137(16): p. 165101.
227. Tehver, R., J. Chen, and D. Thirumalai, *Allostery Wiring Diagrams in the Transitions that Drive the GroEL Reaction Cycle*. Journal of Molecular Biology, 2009. 387(2): p. 390-406.
228. Zheng, W., B.R. Brooks, and D. Thirumalai, *Allosteric Transitions in the Chaperonin GroEL are Captured by a Dominant Normal Mode that is Most Robust to Sequence Variations*. Biophysical Journal, 2007. 93(7): p. 2289-2299.
229. Dykeman, E.C. and R. Twarock, *All-atom normal-mode analysis reveals an RNA-induced allostery in a bacteriophage coat protein*. Physical Review E, 2010. 81(3): p. 031908.
230. De Los Rios, P., et al., *Functional Dynamics of PDZ Binding Domains: A Normal-Mode Analysis*. Biophysical Journal, 2005. 89(1): p. 14-21.
231. Yao, X.-Q., L. Skjærven, and B.J. Grant, *Rapid Characterization of Allosteric Networks with Ensemble Normal Mode Analysis*. The journal of physical chemistry. B, 2016. 120(33): p. 8276-8288.
232. Goncarenco, A., et al., *SPACER: server for predicting allosteric communication and effects of regulation*. Nucleic Acids Research, 2013. 41(Web Server issue): p. W266-W272.
233. Mitternacht, S. and I.N. Berezovsky, *Binding Leverage as a Molecular Basis for Allosteric Regulation*. PLOS Computational Biology, 2011. 7(9): p. e1002148.



234. Udi, Y., et al., *Unraveling Hidden Regulatory Sites in Structurally Homologous Metalloproteases*. Journal of Molecular Biology, 2013. 425(13): p. 2330-2346.
235. Nagase, H. and J.F. Woessner, *Matrix Metalloproteinases*. Journal of Biological Chemistry, 1999. 274(31): p. 21491-21494.
236. Panjkovich, A. and X. Daura, *Exploiting protein flexibility to predict the location of allosteric sites*. BMC Bioinformatics, 2012. 13(1): p. 273.
237. Panjkovich, A. and X. Daura, *PARS: a web server for the prediction of Protein Allosteric and Regulatory Sites*. Bioinformatics, 2014. 30(9): p. 1314-1315.
238. Li, H., et al., *DynOmics: dynamics of structural proteome and beyond*. Nucleic Acids Research, 2017. 45(Web Server issue): p. W374-W380.
239. Hertig, S., N.R. Latorraca, and R.O. Dror, *Revealing Atomic-Level Mechanisms of Protein Allostery with Molecular Dynamics Simulations*. PLOS Computational Biology, 2016. 12(6): p. e1004746.
240. Lange, O.F. and H. Grubmuller, *Generalized correlation for biomolecular dynamics*. Proteins, 2006. 62(4): p. 1053-61.
241. Ghosh, A. and S. Vishveshwara, *A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis*. Proc Natl Acad Sci U S A, 2007. 104(40): p. 15711-6.
242. VanWart, A.T., et al., *Exploring Residue Component Contributions to Dynamical Network Models of Allostery*. Journal of Chemical Theory and Computation, 2012. 8(8): p. 2949-2961.
243. Allain, A., et al., *Allosteric pathway identification through network analysis: from molecular dynamics simulations to interactive 2D and 3D graphs*. Faraday Discuss, 2014. 169: p. 303-21.
244. Lin, M.M., *Timing Correlations in Proteins Predict Functional Modules and Dynamic Allostery*. J Am Chem Soc, 2016. 138(15): p. 5036-43.
245. Ovchinnikov, V. and M. Karplus, *Analysis and Elimination of a Bias in Targeted Molecular Dynamics Simulations of Conformational Transitions: Application to Calmodulin*. The Journal of Physical Chemistry B, 2012. 116(29): p. 8584-8603.
246. Okazaki, K.-i. and S. Takada, *Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms*. Proceedings of the National Academy of Sciences, 2008. 105(32): p. 11182.
247. Daily, M.D., G.N. Phillips, Jr., and Q. Cui, *Interconversion of Functional Motions between Mesophilic and Thermophilic Adenylate Kinases*. PLOS Computational Biology, 2011. 7(7): p. e1002103.
248. Stacklies, W., et al., *Mechanical network in titin immunoglobulin from force distribution analysis*. PLoS Comput Biol, 2009. 5(3): p. e1000306.
249. Stacklies, W., C. Seifert, and F. Graeter, *Implementation of force distribution analysis for molecular dynamics simulations*. BMC Bioinformatics, 2011. 12(1): p. 101.
-

## 1.6 Bibliography

---

250. Fuentes, E.J., et al., *Evaluation of energetic and dynamic coupling networks in a PDZ domain protein*. J Mol Biol, 2006. 364(3): p. 337-51.
251. Dhulesia, A., J. Gsponer, and M. Vendruscolo, *Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein*. J Am Chem Soc, 2008. 130(28): p. 8931-9.
252. Gerek, Z.N. and S.B. Ozkan, *Change in Allosteric Network Affects Binding Affinities of PDZ Domains: Analysis through Perturbation Response Scanning*. Plos Computational Biology, 2011. 7(10): p. e1002154.
253. Münz, M., J. Hein, and P.C. Biggin, *The Role of Flexibility and Conformational Selection in the Binding Promiscuity of PDZ Domains*. PLOS Computational Biology, 2012. 8(11): p. e1002749.
254. Buchli, B., et al., *Kinetic response of a photoperturbed allosteric protein*. Proc Natl Acad Sci U S A, 2013. 110(29): p. 11725-30.
255. Morra, G., A. Genoni, and G. Colombo, *Mechanisms of Differential Allosteric Modulation in Homologous Proteins: Insights from the Analysis of Internal Dynamics and Energetics of PDZ Domains*. J Chem Theory Comput, 2014. 10(12): p. 5677-89.
256. Kalescky, R., J. Liu, and P. Tao, *Identifying key residues for protein allostery through rigid residue scan*. J Phys Chem A, 2015. 119(9): p. 1689-700.
257. Mino-Galaz, G.A., *Allosteric communication pathways and thermal rectification in PDZ-2 protein: a computational study*. J Phys Chem B, 2015. 119(20): p. 6179-89.
258. Lu, C., V. Knecht, and G. Stock, *Long-Range Conformational Response of a PDZ Domain to Ligand Binding and Release: A Molecular Dynamics Study*. J Chem Theory Comput, 2016. 12(2): p. 870-8.
259. Kong, Y. and M. Karplus, *Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis*. Proteins, 2009. 74(1): p. 145-54.
260. Sharp, K. and J.J. Skinner, *Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling*. Proteins, 2006. 65(2): p. 347-61.
261. Ota, N. and D.A. Agard, *Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion*. J Mol Biol, 2005. 351(2): p. 345-54.
262. Feher, V.A., et al., *Computational approaches to mapping allosteric pathways*. Curr Opin Struct Biol, 2014. 25: p. 98-103.
263. Raimondi, F., et al., *A Mixed Protein Structure Network and Elastic Network Model Approach to Predict the Structural Communication in Biomolecular Systems: The PDZ2 Domain from Tyrosine Phosphatase 1E As a Case Study*. Journal of Chemical Theory and Computation, 2013. 9(5): p. 2504-2518.
264. Ho, B.K. and D.A. Agard, *Conserved tertiary couplings stabilize elements in the PDZ fold, leading to characteristic patterns of domain conformational flexibility*. Protein Sci, 2010. 19(3): p. 398-411.

265. Xiang, Y., et al., *Simulating the effect of DNA polymerase mutations on transition-state energetics and fidelity: evaluating amino acid group contribution and allosteric coupling for ionized residues in human pol beta*. *Biochemistry*, 2006. 45(23): p. 7036-48.
266. Vijayabaskar, M.S. and S. Vishveshwara, *Interaction energy based protein structure networks*. *Biophys J*, 2010. 99(11): p. 3704-15.
267. Bhattacharyya, M. and S. Vishveshwara, *Probing the Allosteric Mechanism in Pyrrolysyl-tRNA Synthetase Using Energy-Weighted Network Formalism*. *Biochemistry*, 2011. 50(28): p. 6225-6236.



# Chapter 2

## Methodology

*“...everything that living things do can be understood in terms of the jiggling and wiggling of atoms.”*

– Richard Feynman, Lectures on Physics Vol. 1, Chap. 3, 1963

**B**iological macromolecules are inherently dynamic systems in which the internal motions and the resulting conformational changes play an important role in their functional activity. A fundamental understanding of biological processes would require the knowledge and connection between the structure, dynamics, energetics and function of the macromolecule.<sup>1,2</sup> The traditional approach is to understand underlying physiological mechanisms through experimental techniques such as X-ray crystallography, NMR and Raman spectroscopy. These mechanisms involve interactions that are governed at macroscopic (thermodynamic properties) and microscopic level (atomistic). The experimental techniques provide an inadequate physicochemical based understanding of interactions at the molecular level. This information gap is overcome by the atomistic level description of the systems using concepts of theoretical chemistry and statistical mechanics.<sup>3,4</sup> Several computer simulations have been carried out by generating a physical model of the system to study the equilibrium properties of the system. Two main simulation techniques used to obtain structural and thermodynamic properties of the system are molecular dynamics simulations and Monte Carlo simulations.<sup>5</sup> Molecular dynamics simulation is a powerful technique used to study the dynamics and molecular interactions of biological systems such as protein, nucleic acids and cell membranes at the atomic level.<sup>6</sup> It enables the visualisation of events such as protein folding, protein-drug interaction or functional dynamics associated with the biomolecules, which cannot be observed by experimental techniques. In this thesis, we have used molecular dynamics simulation as the primary method to gain insight into the dynamics and energetics of the protein systems.

## 2.1 Molecular Dynamics Simulations: Background and Theory

---

### 2.1 Molecular Dynamics Simulations: Background and Theory

The earliest known computer simulation study was carried out in 1953 to compute the thermodynamic properties of liquid modelled as hard spheres.<sup>7</sup> This study laid down the foundations of presently known as ‘Metropolis Monte Carlo simulation’. The very first molecular dynamics study was performed by Alder and Wainright for a system of hard spheres to obtain dynamic properties using the solutions of Newton’s equation of motions.<sup>8</sup> This was followed by the remarkable work by Rahman in 1964 on the molecular dynamics simulation of liquid Argon using soft sphere (LJ) potential which proved as an important step for modern simulations.<sup>9</sup> With these background studies, rapid attempts were followed to model large systems such as proteins. The first molecular dynamics simulation for a protein was performed in vacuum in 1977.<sup>10</sup> Since then, there has been phenomenal development in the simulation techniques and computing power resulting in simulation of larger systems ( $> 10^6$  atoms) and long production run (ns- $\mu$ s). Molecular dynamics simulation is a technique in which the positions and velocities of atoms of an  $N$ -particle system is calculated using classical mechanics Newton’s equation of motion,

$$F_i(t) = m_i \frac{d^2 r_i}{dt^2} \quad (2.1)$$

Here,  $F_i$  is the force acting on the  $i^{\text{th}}$  atom at time  $t$ ,  $r_i(t)$  is the position of the atom at time  $t$ ,  $m_i$  is the mass of the atom. The idea is to calculate the force on each atom using potential energy function based on the interactions between the  $N$ -particles. The force on atom  $i$  can be given as,

$$F_i(t) = -\frac{\partial V_{(r_i)}}{\partial r_i} \quad (2.2)$$

where  $V_{(r_i)}$  is the potential of the system. Thus, the potential energy of the system is a function of the atomic coordinates of the  $N$ -particles. This force can be used to calculate the position, velocity and acceleration of each individual atom by solving numerically Newton’s equation of motions using an integration algorithm. Repeating this process after discrete time steps gives the evolution of the individual particle motions of a molecular system as a function of time. Thus,

molecular dynamics simulation technique aims to explore the energy landscape of a system by generating an ensemble of successive conformations.

### 2.1.1 Molecular Dynamics Integrators

Multiple algorithms can be used to integrate Newton's equation of motions. A good molecular dynamics program requires a good integration algorithm. The criteria for a good integration algorithm are:

1. Requires minimum number of cycle to calculate forces per time step, i.e. computationally efficient and occupy less memory.
2. Should conserve energy and momentum.
3. Should be time reversible.
4. Should conserve phase volume.
5. Should have higher accuracy.
6. Should be stable and produce no artefact with larger  $\Delta t$  (time steps).

The algorithm can be derived using simple Taylor series expansion for the positions  $r(t)$  to determine the time evolution of positions, velocities and accelerations of the system in phase space.

$$r(t + \delta t) = r(t) + \frac{dr}{dt} \delta t + \frac{1}{2} \frac{d^2 r}{dt^2} \delta t^2 + \frac{1}{3!} \frac{d^3 r}{dt^3} \delta t^3 + \dots \quad (2.3)$$

$$r(t + \delta t) = r(t) + v(t) \delta t + \frac{1}{2} a(t) \delta t^2 + \dots \quad (2.4)$$

Steps in a typical molecular dynamics program:

1. Initialisation of the system (Assign initial positions and velocities. Positions are obtained from Cartesian coordinates and velocities are assigned based on the Maxwell distribution at a given temperature.)
2. Compute forces on all the atoms of the system using the forcefield potential function form.
3. Integrate Newton's equations of motion. Compute the positions, velocities and accelerations at every  $\Delta t$  time step.

## 2.1 Molecular Dynamics Simulations: Background and Theory

---

4. Repeat steps 2 and 3 to obtain the time evolution of the system for the desired time length.

Here we describe some most commonly used integrators in molecular dynamics which are stable, accurate and computationally efficient.

### 1. Verlet algorithm

This algorithm is one of the simplest and earliest used in molecular dynamics simulations. The verlet algorithm determines the positions of atoms from the Taylor expansion of the coordinate of the particles at time  $(t - \Delta t)$  and  $(t + \Delta t)$ .

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \dots \quad (2.5)$$

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 - \dots \quad (2.6)$$

After adding these two equations, we get:

$$r(t + \Delta t) \approx 2r(t) - r(t - \Delta t) + a(t)\Delta t^2 \quad (2.7)$$

Here  $\Delta t$  is the time step in the simulation. This algorithm is also known as position verlet algorithm, as the velocities are not used to calculate the new positions. However, one can derive velocities using

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} \quad (2.8)$$

The computed velocities are used to calculate the kinetic energy and hence the instantaneous temperature.

Highlights:

1. At time  $t=0$ , the algorithm requires position at  $-\Delta t$ , which is undetermined. The problem can be solved by using a different algorithm for the first time step or by using Taylor expansion about  $r(t)$ .
2. Minimum storage requirement and straightforward.



3. Large errors in velocity approximation (Moderate precision).

## 2. Leap-frog algorithm

Leap-frog algorithm is derived from the Verlet scheme and hence give rise to identical trajectories. It is the most commonly used integrator. In this algorithm, the velocities are evaluated at half-integer time steps and subsequently used to compute the new positions. The algorithm can be written as:

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + a(t)\Delta t \quad (2.9)$$

$$r(t + \Delta t) = r(t) + v\left(t + \frac{\Delta t}{2}\right)\Delta t \quad (2.10)$$

Here, the position at time  $t$  and velocity at previous half time step  $(t - \Delta t/2)$  are used to calculate the velocity at next half time step  $(t + \Delta t/2)$ . From the latter, one can calculate the positions at next integer time step  $(t + \Delta t)$ . Thus, the velocity takes a leap over the position by half time step, and then the position leaps over the velocity to give position at full-time step.

Highlights:

1. Velocities are explicitly calculated. Elimination of addition of small terms to larger terms.  
Improved evaluation of velocities.
2. Computationally expensive than Verlet.

## 3. Velocity-verlet algorithm

This algorithm is a better implementation of the Leap-frog integration method. It calculates the position, velocities and accelerations at the same time.

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \quad (2.11)$$

$$v(t + \Delta t) = v(t) + \frac{1}{2}\Delta t[a(t) + a(t + \Delta t)] \quad (2.12)$$

## 2.2 Force Field

---

Highlights:

1. Explicitly incorporates velocity.
2. System at  $t+\Delta t$  can be directly calculated from quantities at time  $t$ .
3. More precise. Error in positions and velocities is of the order  $O(\Delta t^4)$ .

## 2.2 Force Field

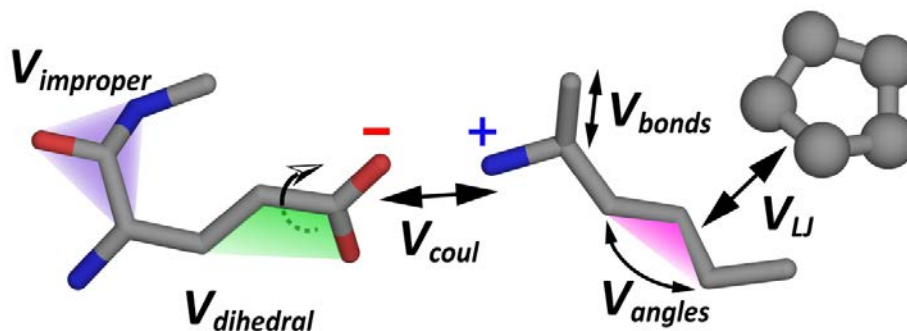
Ideally, the molecular dynamics simulation of a system can be performed using interatomic forces calculated from *ab-initio* quantum mechanical calculations. However, this approach is computationally expensive and limits the use of a system of reasonable size (several hundreds of particles). This limitation is overcome by the use of empirical force field based methods with higher approximations. The use of force field allows simulating a system consisting of several thousands of atoms for time scales of the order of microseconds. A force field defines an empirical set of potential functions and parameters to describe the interactions between atoms and the energy of the system as a function of the coordinates of each atom. The accuracy of the simulations is directly related to the parameters, and potential energy function used to describe the interatomic interactions. A number of force fields have been developed depending upon the energy function formulae and the strategy used for parameterisation. The potential energy function consists of terms representing bonded (covalent) and non-bonded (non-covalent) interactions (Fig. 2.1). A typical potential function in Amber force field can be written as,

$$V = V_{bonds} + V_{angles} + V_{dihedral} + V_{impropers} + V_{LJ} + V_{coul} \quad (2.13)$$

$$\begin{aligned} V = & \sum_{bonds} \frac{1}{2} k_r (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \\ & + \sum_{dihedral} \frac{1}{2} V_n [1 + \cos(n\phi - \gamma)] + \sum_{improper} \frac{K_{imp}}{2} [1 + \cos(2\omega - \pi)] \\ & + \sum_{i=1}^N \sum_{j=i+1}^N 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \end{aligned} \quad (2.14)$$

Here, the first term in the function describes bond length between two particles as harmonic spring with spring constant  $K_r$  and the displacement from the equilibrium bond length ( $r_0$ ) as ( $r$ -

$r_0$ ). The description of bond vibration as the harmonic function implies that the bonds cannot be broken and hence, no chemical process can be studied using classical force field based MD. The bond angles are also modelled by the harmonic function where  $K_\theta$  represents the bending force constant and the displacement from the reference bond angle ( $\theta$ ) as  $(\theta - \theta_0)$ .



**Figure 2.1** Representation of potential energy terms of a typical force field. The carbon atoms are shown in black, nitrogen in blue and oxygen in red.  $V_{bonds}$  is the bond-stretching potential,  $V_{angles}$  angle-bending potential,  $V_{dihedral}$  and  $V_{improper}$  are torsional potentials,  $V_{LJ}$  and  $V_{coul}$  are the non-bonded van der Waals and coulomb's potentials respectively.

In any molecule with more than four atoms, the variation in the energy due to rotation about bonds is included as dihedral torsional potential as described by the third term. Torsional energy is expressed as a cosine function where  $V_n$  represents the height of the potential barrier,  $\varphi$  is the dihedral angle,  $\gamma$  is the phase factor and  $n$  is the multiplicity. In addition, improper torsion potential term is included to ensure planarity of  $sp^2$  hybridised carbon in amino acids carbonyl group or aromatic rings. This term accounts for positive contribution to the potential energy of out-of-plane motions. Here,  $\omega$  is the improper angle related to the deviation from the planarity. The last two terms represent a contribution to the potential energy based on interactions between particles separated by more than three covalent bonds. Here, the fifth term corresponds to the short-range interatomic van der Waals interaction modelled by the 12-6 form of Lennard-Jones potential. In equation 2.14,  $\epsilon$  is the well depth,  $\sigma$  is the width of LJ potential, and the repulsion and attractive forces vary as  $r^{-12}$  and  $r^{-6}$  respectively. The final term in the equation represents the long range electrostatic interactions, which are described by Coulomb's law.

## 2.3 Technical Details of Molecular Dynamics Simulations

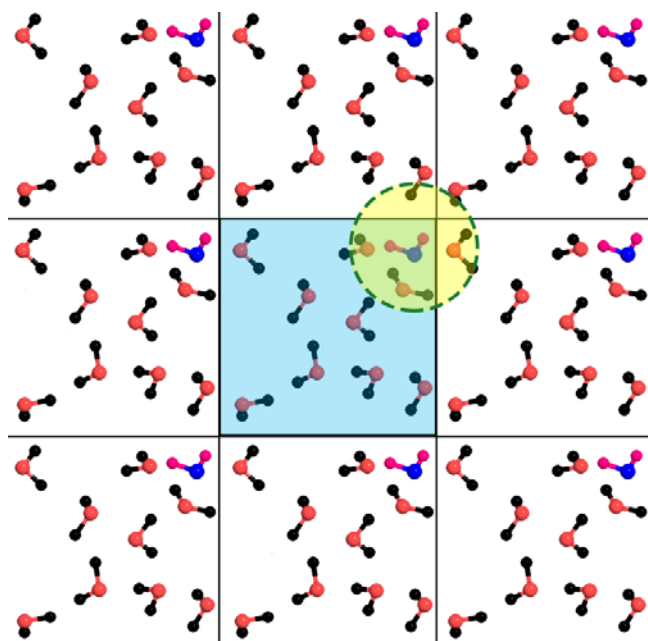
---

The force constants, equilibrium structural parameters and the partial atomic charges are derived from *ab-initio* quantum mechanical calculations and refined by fitting to experimental data from X-ray diffraction, NMR, Raman, infrared and neutron spectroscopy. The potential function form and/or parameters may differ for various available force fields depending on the systems and properties against which it has been characterised. In this thesis, we have used different force fields, namely Charmm27<sup>11</sup>, Charmm36<sup>12</sup> and Amber99sb-Ildn<sup>13</sup> for protein simulations.

## 2.3 Technical Details of Molecular Dynamics Simulations

### 2.3.1 Periodic Boundary Conditions

Molecular dynamics simulations aim to provide the structural and thermodynamic properties of a macroscopic system with a large number of particles. In order to simulate bulk properties of  $N$ -particle system, the surface effects can be overcome by introducing periodic boundary conditions (Fig. 2.2). The periodic boundary conditions mimic the presence of infinite bulk surrounding by replicating an infinite replica of the original simulation box around itself. This removes the artefact of unwanted boundaries and subsequent edge/surface effects.



**Figure 2.2** Schematic representation of periodic boundary condition in a 2D system. A cutoff radius ( $r_c$ ) is shown around the water in blue-magenta colour in the central box. This molecule

can interact with any other atom that lies within the cutoff radius, whether it is in the same box or across PBC.

It is necessary because the fraction of all particles/atoms present at the surface/boundary walls is proportional to  $N^{-1/3}$ . While this fraction is negligible in a macroscopic system with a large number of particles ( $\sim 10^{23}$ ), the surface effects become significant for typical systems used in computer simulations in the order of  $10^3 \sim 10^6$ . In PBC, at any time during the simulation, only  $N$ -particles inside the original cell are considered explicitly. The implications of PBC ensures that if a particle leaves the simulation box from one side, its image from the replica instantaneously enters the box from the opposite direction, thus keeping the number of particles in the simulation box constant. The idea of using periodic boundary condition introduces an artificial periodicity which consists of an original box and all its replicas at the multiples of box length. It is problematic to simulate fluctuations with wavelengths greater than the length of the simulation box. Hence, the box length must be sufficiently large to study properties with long range-correlations.

### 2.3.2 Minimum Image Convention and Truncation of Interactions

With the inclusion of periodic boundary conditions in the system, the number of interacting particles increases enormously. This is because the particles in the original simulation box can interact with all its replica images. This limitation is overcome by using minimum image convention, which ensures that at a time an atom can interact with at most one image of other atoms in the box or nearest images. However, for an  $N$  particle system, the computational time required for pairwise interactions scales with an order of  $N^2$ . Hence, the minimum image convention is used along with the truncation of interactions method to reduce the computational time by restricting the calculation of number of interactions. The method is much effective for calculation of short-range interactions as the interaction potential decays more rapidly than  $r^{-3}$  ( $u(r) \sim 1/r^n$ ;  $n > 3$ ). In this method, a spherical cut-off is defined taking into account all the pairwise interaction in the closest image.

$$U_{trunc}(r) = \begin{cases} U(r), & r \leq r_c \\ 0, & r > r_c \end{cases} \quad (2.15)$$

## 2.3 Technical Details of Molecular Dynamics Simulations

---

Here, the potential  $U(r)$  beyond the cutoff ( $r > r_c$ ) is set zero where  $r_c$  is the cutoff distance. The cutoff must be chosen such that the potential beyond  $r_c$  can be safely neglected. Further, in consistence with the minimum image convention, the cutoff radius cannot be larger than half the length of the simulation box to prevent interaction of a particle with more than two images. The use of simple truncation scheme leads to large fluctuations in energy due to discontinuity in the potential at  $r_c$ . This is overcome by the use of truncated and shifted potential:

$$U_{trunc}(r) = \begin{cases} U(r) - U(r_c), & r \leq r_c \\ 0, & r > r_c \end{cases} \quad (2.16)$$

This method shifts the potential energy surface such that the potential would be zero at the cutoff  $r_c$ . Further, it must be highlighted that the truncation methods must be used only to describe short-range interactions and never be used for long-range interactions such as electrostatic interactions.

### 2.3.3 Long-range Interactions

The calculation of non-bonded interactions, i.e. van der Waals and electrostatic interactions are the most computationally expensive step in molecular dynamics simulation. With the inclusion of periodic boundary condition, a particle interacts with all other particles in its own image and also within all periodic images. For an N-particle system, the Coulomb contribution of the potential energy is given by,

$$V = \frac{1}{2} \sum_{|n|=0}^{\prime} \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon |r_{ij} + nL|} \quad (2.17)$$

where  $L$  is the cubic box length,  $n$  represents all periodic images, prime on the summation represents the sum runs over all the periodic images. Unlike short-range interactions, the truncation method cannot be applied to long range interactions as the potential function decays as  $r^{-1}$  (faster than  $r^{-3}$ ) giving rise to huge errors in force calculations. Different techniques have been developed to handle long-range interactions such as Ewald summation,<sup>14</sup> Particle mesh Ewald summation,<sup>15</sup> and multipole method.

In Ewald summation, the long range interactions are calculated in two parts: the real-space contribution and the reciprocal space contribution. It is assumed that a point charge  $q_i$  is surrounded by a Gaussian charge distribution cloud of opposite sign. Hence, the electrostatic potential energy has three contributions; (1) electrostatic potential due to a point charge  $q_i$  (2) due to screening charge cloud with net charge  $-q_i$  (3) due to the counter neutralising charge cloud with charge  $q_i$ . The electrostatic potential between the point charge and screening charge distribution can be computed by direct summation and is performed in real space. The interaction with the compensating charge cloud is performed in the reciprocal space. Further, the final potential form includes two more terms: (1) to correct for the inclusion of self-interaction between the point charge  $q_i$  and the compensating charge distribution and (2) in case the medium surrounding the charge has infinite dielectric constant. The total electrostatic potential energy is given by,

$$V = V^{(r)} + V^{(k)} + V^{(self)} + V^{(dipolar)} \quad (2.18)$$

$$V = \frac{1}{2} \sum_{i=1}^N \sum_{j=i+1}^N \left\{ \sum_{|n|=0}^{\infty} \frac{q_i q_j}{4\pi\epsilon_0} \frac{\frac{2}{\sqrt{\pi}} \int_r^{\infty} \exp(-t^2) dt (\alpha |r_{ij} + n|)}{|r_{ij} + n|} \right. \\ \left. + \sum_{k \neq 0} \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\epsilon_0} \frac{4\pi^2}{k^2} \exp\left(-\frac{k^2}{4\alpha^2}\right) \cos(k \cdot r_{ij}) - \frac{\alpha}{\sqrt{\pi}} \sum_{k=1}^N \frac{q_k^2}{4\pi\epsilon_0} + \frac{2\pi}{3L^3} \left| \sum_{k=1}^N \frac{q_k}{4\pi\epsilon_0} r_k \right|^2 \right\} \quad (2.19)$$

Here  $V^{(r)}$  is the real space contribution,  $V^{(k)}$  is the reciprocal space contribution,  $V^{(self)}$  is the self energy correction and  $V^{(dipolar)}$  is the dipolar correction. In the case of large systems, the traditional Ewald summation method is inefficient as the calculation in the reciprocal space scales as  $N^2$ . The limitation is overcome by Particle mesh approaches where the charges in the systems are interpolated on a grid which are solved using Fast Fourier Transform algorithm.<sup>15</sup> The FFT based particle mesh method scales as  $M \log N$ .

### 2.3.4 Neighbour and Cell Lists

As described earlier, the truncation method reduces the total number of pair interactions. However, this method requires the calculation of the distance between an atom  $i$  with the remaining  $N-1$  atoms to determine which ones lie within the cutoff radius ( $r_c$ ). Thus at any time

## 2.3 Technical Details of Molecular Dynamics Simulations

---

step, despite truncating potential, one needs to calculate the  $N(N-1)/2$  pairwise distances and the time scales as  $N^2$ . Hence, to speed up the calculation of non-bonded interactions, neighbour and cell list techniques are developed such that the computing time scales as  $N^{3/2}$  or  $N$ . In neighbour list (verlet list) method, another cutoff ( $r_{list}$ ) is introduced at the beginning of the simulation such that  $r_{list} > r_c$  and a list of all the atoms within  $r_{list}$  is maintained. In the subsequent steps, interactions are calculated for only atoms included in the neighbour list, thus reducing a large number of unnecessary calculations and increasing overall performance. The list is updated every 10-20 steps when the atoms are displaced more than  $r_{list} - r_c$ . The cell list method is used when the number of atoms is large, and the length of the simulation box is much greater than the cutoff radius ( $r_c$ ). In this method, the box is divided into several cells with edge equal to or slightly larger than  $r_c$ . At the beginning of the simulation, a list of atoms in the same or neighbouring cell is maintained and remains unchanged unless the domain changes during the simulation. Thus at any time step, the interaction for any atom  $i$  is calculated with all atoms within the volume defined by same and neighbouring cell (27 cells), i.e.  $4/3 \cdot \pi r_c^3$ . The cell list method is fast and efficient as compared to the verlet list as the method scales as  $N$ .

### 2.3.5 Thermodynamics Ensembles

The thermodynamic properties of a system can be derived using a number of ensembles based on the concepts of statistical mechanics. Ideally, molecular dynamics simulation is an evolution of system of  $N$  particles in a fixed volume  $V$  and constant total energy  $E$ . Thus, the time average thermodynamic properties for this system can be derived from the ensemble average in the microcanonical ensemble (NVE). However, most reactions and experimental observations are performed at constant pressure and temperature with the possibility for the system to interact with the surrounding and undergo fluctuations in the energy or volume or number of particles. In molecular dynamics simulations, this is achieved using two thermodynamics ensembles, namely canonical ensemble (N,V,T) and isothermal-isobaric ensemble (N,P,T). In NVT ensemble, the temperature, number of particles and volume is kept fixed whereas, in NPT ensemble, temperature, number of particles and pressure are constant while the volume of the system is allowed to fluctuate. To achieve these ensembles, the system is coupled to thermostat or barostat (external baths) at a constant temperature and pressure, respectively. In this thesis,



NVT ensemble was used during temperature equilibration while NPT ensemble was used during all production runs.

### Thermostat (Constant Temperature Simulations)

The temperature of the system is directly correlated to the kinetic energy by the equipartition theorem of energy,

$$\frac{1}{2}mv^2 = \frac{3}{2}Nk_bT \quad (2.20)$$

Based on this well-known theorem, several ways have been proposed to control the temperature of the system.

#### 1. Velocity rescaling

The most trivial way of controlling the temperature is by modifying the velocities of atoms. The idea is to scale the velocities of the particles in the system such that the average kinetic energy of the systems matches with the system at target temperature. In this method, velocities are scaled by a factor,

$$\lambda = \sqrt{\frac{T_B}{T(t)}} \quad (2.21)$$

where  $T_B$  is the desired temperature, and  $T(t)$  is the instantaneous temperature before scaling.

#### 2. Berendsen thermostat

The Berendsen thermostat is an example of modified velocity rescaling and weak coupling thermostat.<sup>16</sup> It is a more physical way of controlling temperature by weakly coupling of the system to an external heat bath maintained at the desired temperature ( $T_B$ ). It is achieved by modifying Newton's equation of motions by introducing a pseudo friction coefficient  $\gamma(t)$ . The temperature is scaled by factor  $\lambda$  given by,

$$\lambda = \left[ 1 + \frac{\Delta t}{\tau} \left( \frac{T_B}{T} - 1 \right) \right]^{1/2} \quad (2.22)$$

where  $\Delta t$  is the time step, and  $\tau$  is the coupling parameter between the heat bath and system.

## 2.3 Technical Details of Molecular Dynamics Simulations

---

In this thesis, we have used velocity rescaling temperature coupling method<sup>17</sup> that is basically a modified Berendsen thermostat with an additional stochastic term to generate a correct Canonical ensemble given by,

$$dK = (K_0 - K) \frac{dt}{\tau_T} + 2 \sqrt{\frac{KK_0}{N_f}} \frac{dW}{\sqrt{\tau_T}} \quad (2.23)$$

where  $K$  is the kinetic energy,  $N_f$  is the number of degrees of freedom,  $dW$  a Wiener process and  $\tau_T$  is the temperature coupling time constant.

### 3. Anderson thermostat

In this constant temperature method, the system is coupled to a heat bath represented by stochastic collisions of randomly selected particles with the heat bath.<sup>18</sup> At each collision, new velocity is assigned to the particle from a Maxwell-Boltzmann distribution corresponding to the desired bath temperature keeping all other particles unaffected. The collision frequency ( $\nu$ ) determines the strength of the coupling with the heat bath. The distribution of time intervals between two stochastic collisions is given as,

$$P(\nu, t) = \nu e^{-\nu t} \quad (2.24)$$

### 4. Nose-Hoover thermostat

This method is based on extended Lagrangian with an additional new degree of freedom ( $s$ ) and effective mass  $Q$  associated with a fictional heat bath.<sup>19,20</sup> The Lagrangian can be written as

$$L_{N-H} = \sum_{i=1}^N \frac{m_i}{2} s^2 \left( \frac{dr_i}{dt} \right)^2 - U(r^N) + \frac{Q}{2} \left( \frac{ds}{dt} \right)^2 - \frac{L}{\beta} \ln s \quad (2.25)$$

Here, the last two terms represent the potential energy and kinetic energy of the heat bath.

### Barostat (Constant Pressure Simulations)

Similar to the thermostat, in constant pressure simulations, the system can be coupled to pressure bath with desired pressure, which is achieved by scaling the volume of the simulation box. In

other terms, in the NPT ensemble, the shape and size of the cell are rescaled to maintain constant pressure. The rate of change of pressure with respect to time is given by,

$$\frac{dP}{dt} = \frac{1}{\tau_p} (P_B - P) \quad (2.26)$$

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_p} (P_B - P) \quad (2.27)$$

where  $\kappa$  is the isothermal compressibility of the system and is given by:

$$\kappa = -\frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_T \quad (2.28)$$

Thus, at each step, the coordinates, as well as the box lengths, are scaled by  $\lambda$  to maintain the pressure at  $P_0$ . In this thesis, we have used Parrinello-Rahman Pressure Coupling method.<sup>21</sup>

## 2.4 Steps in Running Molecular Dynamics Simulation

The entire molecular dynamic simulation process can be divided into four steps:

### 1. System Setup

To start a simulation, one needs to generate initial conformation of the system from pre-existing models such as structural data or by placing the molecules or particles randomly inside a box which represents the volume of the system. In the case of biological entities such as proteins and nucleic acids, the atomic coordinates of the system can be obtained from the three-dimensional structure solved using x-ray diffraction or NMR technique. Appropriate force-field is selected to assign non-bonded and bonded parameters of the atoms/particles in the system which will be used for the calculation of potential energy at a later stage. To replicate biological conditions, the system (protein/nucleic acids) is solvated by adding water and ions in the simulation box.

### 2. Energy Minimization and Equilibration

The system is then energy minimised using either steepest descent or conjugate gradient method to remove any atom overlapping and steric stress between two extremely close atoms. The initial velocities of the atoms are assigned randomly from a Maxwellian

## 2.5 Data Analysis Methods

---

distribution centred around the desired temperature, and then the resulting velocities are scaled to modify the average kinetic energy to the desired value. The initial state of the system could be far away from the thermodynamic equilibrium conditions in which the simulation has to be performed. Hence, a two-step NVT and NPT equilibration are performed by coupling system to thermostat and barostat respectively, so that the system reaches equilibrium. Sometimes the solute part in the system is position restrained to enable the equilibration of solvent part at the desired conditions.

### 3. Production Run

Once the system is equilibrated, it is made to evolve through the conformational space in the production run. The time steps for the calculation of position and velocities must be sufficient to account for the fastest event in the simulation ( $\sim 2$ fs), and the production run must be sufficiently long to ensure adequate sampling of the phase space.

### 4. Data Analysis

The production run generates data about the position and velocities of the atoms with time, which can be used to calculate various physical properties of the system. Some of the conventional analyses performed for protein systems are root-mean-square deviation, root-mean-square fluctuations and radius of gyration, geometric clustering. In this thesis, we have used two other techniques, namely principal component analysis and k-means clustering to explore the conformational dynamics of the system.

## 2.5 Data Analysis Methods

### 2.5.1 Root Mean Square Deviation

Root mean square deviation (*RMSD*) is defined as the deviation of a set of atoms in a molecule with respect to the similar set of atoms in reference structure  $r^{ref}$  with time and is calculated as,

$$RMSD(t) = \left[ \frac{1}{M} \sum_{i=1}^N m_i |r_i(t) - r_i^{ref}|^2 \right]^{1/2} \quad (2.29)$$

where  $N$  is the total number of atoms,  $m_i$  is the mass of the atom  $i$ ,  $r_i(t)$  is the position of atom  $i$  at time  $t$  and  $M = \sum_{i=1}^N m_i$  is the total mass. The above equation (2.29) is for mass-weighted RMSD calculation after least square fitting the structure at time  $t$  to the reference structure.

### 2.5.2 Root Mean Square Fluctuation

Root mean square fluctuation (*RMSF*) is the deviation of an atom  $i$  with respect to the reference position  $r_i^{ref}$  over the time ( $T$ ). The reference position can be also a time-average position of the atom  $i$ , i.e.,  $r_i^{ref} = \bar{r}_i$ .

$$RMSF_i = \left[ \frac{1}{T} \sum_{t=1}^T |r_i(t) - r_i^{ref}|^2 \right]^{1/2} \quad (2.30)$$

The difference between RMSF and RMSD is that the RMSF is averaged over time, whereas the RMSD is the deviation with time.

### 2.5.3 Radius of Gyration

Radius of gyration defines the distribution of atoms about an axis through the center of mass of the molecule.

$$R_g = \left( \frac{\sum_i |r_i|^2 m_i}{\sum_i m_i} \right)^{1/2} \quad (2.31)$$

where  $m_i$  is the mass and  $r_i$  is the position of atom  $i$ . This technique is used to define the compactness or size of the molecule; for example proteins, polymer chains

### 2.5.4 Hydrogen Bonds

A hydrogen bond is defined between a donor atom (hydrogen attached) and an acceptor atom by following geometric criterion:

- (i) distance between the donor atom and the acceptor atom  $r \leq 0.35 \text{ nm}$
- (ii) angle A-D-H  $\alpha \leq 30^\circ$

In this thesis, we have used Gromacs utility `g_hbond` and VMD Hbonds plugin to identify all hydrogen bonds between any two groups of atoms.

### 2.5.5 Principal Component Analysis

A typical molecular dynamics trajectory contains a large number of variables (atomic coordinates or internal coordinates) that can be correlated or dependent on each other. In molecular dynamics simulation, principal component analysis (PCA) is one of the most commonly used techniques to probe the correlated functionally significant motions in the protein. PCA is a multivariate statistical technique used to reduce the dimension of data obtained from  $3N$  atomic coordinates to minimal selective degrees of freedom while retaining maximum variation in the data set.

Steps in Principal Component Analysis:

1. Calculate the covariance matrix:

Any PCA calculation starts with the construction of a covariance matrix or correlation matrix for a set of atomic coordinates from the molecular dynamics trajectory. Initially, a rotational-fit is performed over the trajectory to remove translational and rotational motions to remove dynamics that do not contribute to the internal motions of the system. The standard procedure is to align the sampled structures to a reference structure which can be an average structure from the trajectory or native structure obtained from NMR/X-ray diffraction technique. Once aligned, one can calculate the covariance matrix. For a system with  $N$  atoms, the covariance matrix can be described as,

$$C_{ij} = \left\langle \left( r_i - \langle r_i \rangle \right) \left( r_j - \langle r_j \rangle \right) \right\rangle \quad (2.32)$$

where  $C_{ij}$  is the covariance between  $i^{\text{th}}$  and  $j^{\text{th}}$  atoms,  $r_1 \dots r_{3N}$  are mass-weighted cartesian coordinates and  $\langle \dots \rangle$  is the average over all conformations sampled. The element  $C_{ij}$  will be large and positive if the atoms  $i$  and  $j$  deviate largely from their respective equilibrium positions and in the same direction and vice-versa. One of the crucial steps is the selection of atoms for the calculation of covariance matrix where depending on the type of motion to be identified such as large-scale global motions or localised motions, the atoms selection varies from all atoms, all non-hydrogen atoms or  $\alpha$ -carbon atoms.

2. Calculate eigenvectors and eigenvalues:

The next step is to diagonalise the covariance matrix to obtain the eigenvector ( $v^i$ ) and eigenvalues ( $\lambda_i$ ) which describe the modes and amplitude of collective motion.

3. Arrange the eigenvectors in decreasing order of eigenvalues:

The eigenvalue decomposition of the covariance matrix gives a set of orthogonal eigenvectors and corresponding eigenvalue which characterises different motions of the system. The eigenvalues ( $\lambda_i$ ) are arranged in the decreasing order of the magnitude i.e.  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \dots \geq 0$ . The largest eigenvalues correspond to the reduced representation of large motions in the protein.

4. Calculate the principal components:

The principal components (PC) can be calculated by the projection of the data  $r = (r_1 \dots r_{3N})^T$  over the eigenvectors ( $v^i$ ).

$$V^i = v^i \cdot r \quad (2.33)$$

All PCs are uncorrelated and geometrically orthogonal. A scatter plot of two PC modes shows how the system explored the conformational space defined by the selected PCs.

Based on the atomic fluctuations, a protein's conformational space has been divided into two subspaces, namely essential subspace and constrained subspace. It has been shown that the first few PC modes represent fluctuations/motions which are relevant to protein function (opening and closing of the loop, hinge bending, domain motions) and constitute the essential subspace. While the remaining PC modes correspond to the irrelevant local fluctuations described as constraint subspace. However, this technique is limited by the conformational space sampled in the molecular dynamics simulations.

### 2.5.6 K-means Clustering

Cluster analysis is another common method used to analyse multivariate data. K-means is an unsupervised clustering algorithm which is used to partition a given data set into a priori defined number of clusters. Let  $X = \{x_i\}$ ,  $i = 1, \dots, n$  be the set of  $n$  data points to be partitioned into  $K$  clusters,  $C = \{c_k\}$ ,  $k = 1, \dots, K$ . The aim is to partition  $n$  data points into  $k$  clusters such that the sum of the squared error over all  $K$  clusters is least. Let  $\mu_k$  be the mean of cluster  $c_k$ . Then,

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \quad (2.34)$$

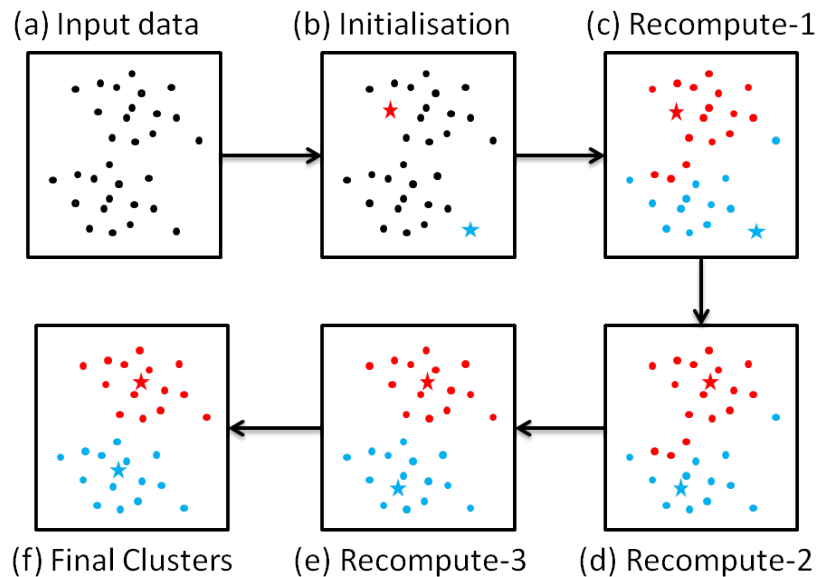
$$J(C) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \quad (2.35)$$

## 2.5 Data Analysis Methods

---

where  $\|x_i - \mu_k\|^2$  is the squared error between the data points in the cluster  $c_k$  and mean  $\mu_k$ . The following steps are involved in K-means algorithm:

1. Initialise  $k$  data points as initial cluster centers,  $c_1 \dots c_k$
2. For each data point, assign it to the closest cluster.
3. Compute new cluster centers by averaging over all the points assigned to it.
4. Repeat steps 2 and 3 until convergence (No change in the cluster centers).
5. Return  $\{c_1 \dots c_k\}$ .



**Figure 2.3** Schematic representation of K-means algorithm: (a) Cluster input data into two clusters ( $k=2$ ). (b) Initialise two data points as cluster centers (c, d, e) Compute the distance between data points and cluster center. Recompute cluster center and reassign data points. Repeat these steps until no change in cluster center is observed (f) Final clusters after convergence.

The k-means clustering algorithm requires three pre-defined parameters: number of clusters ( $k$ ), distance metric and cluster initialisation. K-means clustering can be viewed as a greedy algorithm that converges to local minima with different initialisation leading to different clusters. This can be overcome if the clusters are well separated or by running algorithm multiple times using different initialisation for given  $k$  clusters. Euclidean metric is usually used to calculate the distance between the cluster center and the data points. There is no theoretical



solution for determining the optimal number of clusters a priori for a given data set. A simple approach is to perform k-means clustering for a given data set independently for multiple values of k and choose the partition based on known criteria.

### 2.6 Free Energy Calculations

In nature, all the biochemical and physical processes such as protein-ligand binding, molecular recognition and protein folding are governed by the underlying free energy landscape. The understanding of their functional mechanisms at the molecular level requires the description of such processes in terms of thermodynamics. Statistical mechanics describes the equilibrium ensembles by a probability distribution and connects the microstates of a system to the macroscopic thermodynamic properties such as free energy, temperature, density. Free energy is a thermodynamic quantity that describes the spontaneity of a chemical process and thus determines the equilibrium state of a system.

The free energy of a system is usually described as Helmholtz or Gibb's free energy. Using statistical mechanics, the Helmholtz free energy in a Canonical ensemble can be expressed in terms of the partition function  $Q$ :

$$A = -k_B T \ln Q(N, V, T) \quad (2.36)$$

Similarly, the characteristic state function of isothermal-isobaric ensemble, i.e. Gibbs free energy can be described as,

$$G = -k_B T \ln Z(N, V, T) \quad (2.37)$$

In the case of canonical ensemble, the partition coefficient  $Q(N, V, T)$  for  $N$  particle system is defined as,

$$Q = \frac{1}{h^{3N} N!} \iint e^{-\frac{H(p,r)}{k_B T}} dp dr \quad (2.38)$$

where  $r$  and  $p$  are  $3N$  dimensional vector corresponding to the coordinates and momenta of all particles of the system. Complex processes such as enzyme-catalysed transformations or protein

## 2.6 Free Energy Calculations

---

folding occurs in a high-dimensional phase space that involves a large number of degrees of freedom resulting in rough energy landscapes. Moreover, the calculation of free energy becomes difficult as a realistic biochemical system involves interaction with solvent molecules, ions and other substrates generating local minima and sub-minima on free energy surface populated with conformations. Hence, the free energy calculations are performed along a set of coordinates which defines the progress of a chemical process along the reduced degrees of freedom or characterise the end states of the reaction. These are typically referred to as collective variables or reaction coordinates. The reaction coordinates are usually defined as the functions of Cartesian coordinates of atoms in the system. The choice of the reaction coordinate (collective variable) entirely depends upon the reaction process to be monitored. For example, in a bond-dissociation process, the bond length between the associated atoms can be used to describe the progress of the reaction. The distinct folded and unfolded states of protein can be characterised by using different sets of reaction coordinates such as dihedral angles ( $\phi$ ,  $\psi$ ), radius of gyration ( $R_g$ ) and hydrogen bonds.

### 2.6.1 Potential of Mean Force

Theoretically, one can calculate exact free energy for a system using equation 2.36; however, practically, it is not computationally possible to calculate it for a system which has a high free energy barrier. Hence, once a reaction coordinate is identified that describes the physical progress of a reaction, one can construct a free energy profile based on it. At the same time, one cannot ignore the remaining degrees of freedom that determines how free energy changes along the reaction coordinate. This is achieved by averaging over all other degrees of freedom. This free energy profile is referred to as potential of mean force (PMF). The PMF is given by,

$$A(X) = -k_B T \ln \int dr^N \delta(f(r^N) - X) \exp[-\beta U(r^N)] \quad (2.39)$$

where  $\delta(f(r^N) - X)$  is the dirac delta function for the reaction coordinate  $X$  with respect to position coordinates of all particles ( $r^N$ ). The above equation (2.39) gives the reduced free energy along a reaction coordinate with average effects of all other degrees of freedom. Based on the relationship between partition coefficient and free energy, the PMF along a reaction coordinate can also be defined from the average distribution function,

$$A(X) = -k_B T \ln P(X) + c \quad (2.40)$$

where  $P(X)$  is the probability distribution for any reaction coordinate  $X$  of interest. The method has been successfully used for free energy calculations in processes such as organic reactions in water<sup>22</sup> and biological systems such as nucleic acid base flipping,<sup>23</sup> ion transport through membranes.<sup>24</sup>

### 2.6.2 Rare Event Sampling

In nature, several biochemical and physical events occur at time scales which are inaccessible by conventional molecular dynamics simulations. Such systems exhibit complex free energy landscape with stable states separated by large free energy barrier, e.g. protein folding,<sup>25,26</sup> nucleation of first-order phase transitions.<sup>27</sup> In such processes, the probability distribution  $P(X)$  in the equation 2.40 cannot be computed correctly as the probability that a fluctuation will sample the stable states becomes exponentially small with the increase in the barrier height greater than  $k_B T$ . Thus, during conventional molecular dynamics, the system gets trapped into the local minima and the free energy surface isn't sufficiently sampled. To overcome this problem, a range of enhanced sampling techniques<sup>28</sup> have been developed, such as umbrella sampling,<sup>29</sup> metadynamics,<sup>30,31</sup> and replica exchange molecular dynamics.<sup>32</sup>

#### Umbrella Sampling

The umbrella sampling technique is often used to overcome the sampling problem by introducing a bias potential so that the probability of visiting the stable states separated by large free energy barriers is increased.<sup>29,33</sup> The idea is to calculate the PMF along a chosen reaction coordinate by adding a bias which is usually a harmonic potential of the form,

$$W(X) = k(X - X_0)^2 \quad (2.41)$$

where  $k$  is the harmonic restraint on the CV,  $X_0$  is the equilibrium state in each window. The harmonic restraint or bias function is applied to keep the system close to the equilibrium state. This is added to the original potential  $U_0(r)$  so the total potential becomes,

$$U(r) = U_0(r) + W(X) \quad (2.42)$$

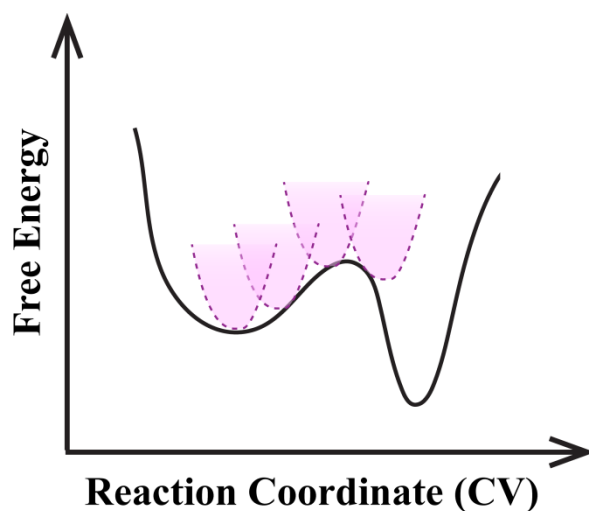
## 2.6 Free Energy Calculations

---

The reaction coordinate is defined such that the equilibrium value of bias potential can be chosen as a set of intermediate points between the two end states. Sampling in each window/umbrella will yield biased probability distribution around each point. Then, the ensemble average of quantity  $F$  for an unbiased trajectory becomes,

$$\langle F \rangle_{U_0} = \frac{\langle F \exp(\beta W) \rangle_U}{\langle \exp(\beta W) \rangle_U} \quad (2.43)$$

where  $U$  indicates, the system is sampled using biased potential. Umbrella sampling technique involves simulations at multiple windows, thus generating multiple PMF along the reaction coordinate.



**Figure 2.4** *Illustration of umbrella Sampling technique. The attractive harmonic potential is added along the free energy landscape in the form of multiple windows. The conformations are sampled using these parabola potentials and are processed using WHAM method to reconstruct the original free energy surface.*

Weighted Histogram Analysis Method (WHAM) is a standard technique used to combine the multiple simulations to generate a complete PMF.<sup>34,35</sup> The accuracy of the umbrella sampling lies in the sufficient overlapping of the windows and the choice of biased potential. Several methods have been introduced to estimate and reduce error due to insufficient sampling.<sup>36-39</sup> This method has been widely used to study large conformational changes in protein such as protein misfolding, ligand binding events and movement of permeants in ions channels.<sup>38,40,41</sup>

## Metadynamics

Metadynamics is another technique used for enhanced sampling and to reconstruct free energy landscape as a function of selective collective variables.<sup>30,42</sup> Unlike umbrella sampling, this method pushes the system out of the local minima by addition of repulsive bias potential to accelerate the sampling of rare events.<sup>43</sup> A history-dependent bias potential is added with time, such that the bias potential at time  $t$  can be written as the sum of repulsive Gaussian deposited along the trajectory and is given by,

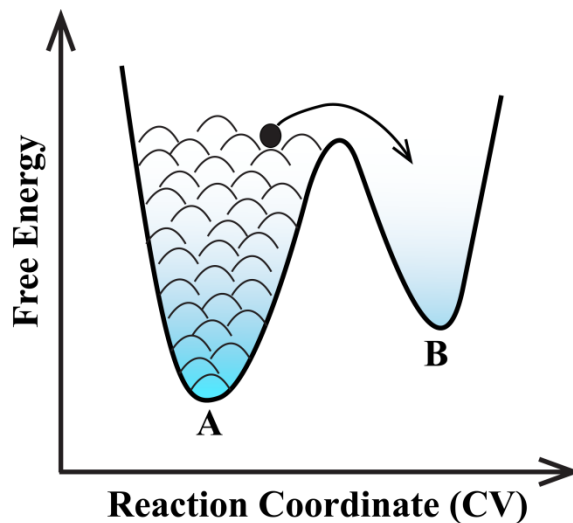
$$V_G(S(x), t) = w \left[ \exp\left(-\frac{(S(x) - S(x(t_0)))^2}{2\sigma^2}\right) + \exp\left(-\frac{(S(x) - S(x(t_1)))^2}{2\sigma^2}\right) + \dots \right] \quad (2.44)$$

$$V_G(S(x), t) = w \sum_{\substack{i=0,1,\dots \\ t_i < t}} \exp\left(-\frac{(S(x) - S(x(t_i)))^2}{2\sigma^2}\right) \quad (2.45)$$

$$V_G(S(x), t) = w \sum_{\substack{i=0,1,\dots \\ t_i < t}} \exp\left(-\frac{(S(x) - s(t_i))^2}{2\sigma^2}\right) \quad (2.46)$$

where  $s(t) = S(x(t))$  is the value of collective variable at time  $t$ ,  $\sigma$  is the Gaussian width and  $w$  is the Gaussian height.

The addition of biased potential pushes the system away from the free energy minima. Thus, the time interval between the addition of successive hills (deposition frequency) must allow the system to relax to the nearest local minimum. The Gaussian width and height, along with the deposition frequency, determine the accuracy and efficiency of the reconstructed free energy profile. Addition of larger hills or too quickly (higher deposition frequency), results in large errors as the system fails to get back to nearest minima whereas addition of smaller hills though produces more accurate FES, increases the computational time. Hence, there should be a balance between the selections of these three parameters.



**Figure 2.5** Schematic representation of metadynamics technique: In this technique, the repulsive bias added in the form of Gaussians, which acts as sand that fills the valleys on the free energy surface. With time, the addition of sufficient Gaussian hills, i.e. accumulation of sand in the minima A forces the system to cross the barriers and explore new minima B.

In metadynamics, it is assumed that after addition of sufficient number of Gaussian hills over a long time, the bias potential provides an estimate of the unbiased energy landscape,

$$V_G(S, t \rightarrow \infty) = -F(S) + c \quad (2.47)$$

where  $c$  is a constant. However, the first version of metadynamics technique is limited by the convergence, often leading to overfilling of the free energy space. This limitation is overcome by another approach referred as well-tempered metadynamics.<sup>44</sup> In well-tempered metadynamics, the bias potential is given by,

$$V(S, t) = k_B \Delta T \ln \left( 1 + \frac{\omega N(S, t)}{k_B \Delta T} \right) \quad (2.48)$$

where  $V(S, t)$  is the bias potential,  $N(S, t)$  is the histogram of the collective variable  $S$ ,  $\Delta T$  has a unit of temperature and controls the degree to which the biased trajectory can explore free energy surface away from minima. In other words, the CV is sampled at enhanced temperature ( $T + \Delta T$ ) and by tuning  $\Delta T$  one can limit the exploration of the FES region to the physically

meaningful regions in the CV space. In this method, the rate of Gaussian deposition decrease over time which is implemented by rescaling the height of the Gaussians as,

$$w = \omega \tau_G e^{-\left(\frac{V(S,t)}{\Delta T}\right)} \quad (2.49)$$

where  $\omega$  is the initial bias deposition rate,  $\tau_G$  is the time interval. The rate of deposition decreases as  $1/t$ . Further, the FES can be estimated as,

$$V_G(S, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + c \quad (2.50)$$

Therefore, in two limiting cases, at  $\Delta T = 0$ , the bias is equal to zero and represents unbiased simulation and at  $\Delta T \rightarrow \infty$ , the deposition rate becomes constant and  $V_G(S, t \rightarrow \infty) = -F(S)$ . Metadynamics technique has been widely used to study reaction pathways and identify structural entities corresponding to metastable states as the system escapes the minima through the lowest energy saddle point.<sup>45-48</sup> It is also useful to reconstruct complete free energy landscape without a prior knowledge about the landscape. In this thesis, we have used well-tempered metadynamics to generate free energy surfaces with desired collective variables.

## 2.7 Other Computational Details

The graphs were plotted using the Xmgrace and Gnuplot plotting program. The structural figures were generated using VMD<sup>49</sup> and PyMOL<sup>50</sup> molecular visualisation program.

## 2.8 Bibliography

---

### 2.8 Bibliography

1. Chipot, C. and A. Pohorille, *Free energy calculations theory and applications in chemistry and biology*. 2007, Berlin: Springer.
2. Bagchi, B. *Water in biological and chemical processes : from structure and dynamics to function*. 2013.
3. Chandler, D., *Introduction to modern statistical mechanics*. 1987, New York: Oxford University Press.
4. Tuckerman, M.E., *Statistical mechanics : theory and molecular simulation*. 2018, Oxford [etc.]: Oxford University Press.
5. Bagchi, B. *Statistical mechanics for chemistry and materials science*. 2019.
6. Frenkel, D. and B. Smit, *Understanding molecular simulation : from algorithms to applications*. 2002, San Diego: Academic Press.
7. Metropolis, N., et al., *Equation of State Calculations by Fast Computing Machines*. The Journal of Chemical Physics, 1953. 21(6): p. 1087-1092.
8. Alder, B.J. and T.E. Wainwright, *Phase Transition for a Hard Sphere System*. The Journal of Chemical Physics, 1957. 27(5): p. 1208-1209.
9. Rahman, A., *Correlations in the Motion of Atoms in Liquid Argon*. Physical Review, 1964. 136(2A): p. A405-A411.
10. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins*. Nature, 1977. 267(5612): p. 585-590.
11. MacKerell, a.D., N. Banavali, and N. Foloppe, *Development and current status of the CHARMM force field for nucleic acids*. Biopolymers, 2001. 56(4): p. 257-65.
12. Huang, J. and A.D. MacKerell Jr, *CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data*. Journal of Computational Chemistry, 2013. 34(25): p. 2135-2145.
13. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field*. Proteins, 2010. 78(8): p. 1950-8.
14. Ewald, P.P., *Die Berechnung optischer und elektrostatischer Gitterpotentiale*. Annalen der Physik, 1921. 369(3): p. 253-287.
15. Essmann, U., et al., *A smooth particle mesh Ewald method*. The Journal of Chemical Physics, 1995. 103(19): p. 8577-8593.
16. Berendsen, H.J.C., et al., *Molecular dynamics with coupling to an external bath*. The Journal of Chemical Physics, 1984. 81(8): p. 3684-3690.
17. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. The Journal of Chemical Physics, 2007. 126(1): p. 014101.
18. Andersen, H.C., *Molecular dynamics simulations at constant pressure and/or temperature*. The Journal of Chemical Physics, 1980. 72(4): p. 2384-2393.



19. Nosé, S., *A molecular dynamics method for simulations in the canonical ensemble*. Molecular Physics, 1984. 52(2): p. 255-268.
20. Hoover, W.G., *Canonical dynamics: Equilibrium phase-space distributions*. Physical Review A, 1985. 31(3): p. 1695-1697.
21. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*. Journal of Applied Physics, 1981. 52(12): p. 7182-7190.
22. Gao, J., *A priori computation of a solvent-enhanced SN2 reaction profile in water: the Menshutkin reaction*. Journal of the American Chemical Society, 1991. 113(20): p. 7796-7797.
23. Huang, N., N.K. Banavali, and A.D. MacKerell, *Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase*. Proceedings of the National Academy of Sciences, 2003. 100(1): p. 68.
24. Allen, T.W., O.S. Andersen, and B. Roux, *Energetics of ion conduction through the gramicidin channel*. Proceedings of the National Academy of Sciences, 2004. 101(1): p. 117.
25. Jha, S. and J. Udgaonkar, *Free energy barriers in protein folding and unfolding reactions*. Vol. 99. 2010. 457-475.
26. Dinner, A.R., et al., *Understanding protein folding via free-energy surfaces from theory and experiment*. Trends in Biochemical Sciences, 2000. 25(7): p. 331-339.
27. Oxtoby, D.W., *Nucleation of First-Order Phase Transitions*. Accounts of Chemical Research, 1998. 31(2): p. 91-97.
28. Maximova, T., et al., *Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics*. PLOS Computational Biology, 2016. 12(4): p. e1004619.
29. Kästner, J., *Umbrella sampling*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2011. 1(6): p. 932-942.
30. Barducci, A., M. Bonomi, and M. Parrinello, *Metadynamics*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2011. 1(5): p. 826-843.
31. Marinelli, F. and José D. Faraldo-Gómez, *Ensemble-Biased Metadynamics: A Molecular Simulation Method to Sample Experimental Distributions*. Biophysical Journal, 2015. 108(12): p. 2779-2782.
32. Abrams, C. and G. Bussi, *Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration*. Entropy, 2014. 16(1).
33. Torrie, G.M. and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*. Journal of Computational Physics, 1977. 23(2): p. 187-199.
34. Souaille, M. and B.t. Roux, *Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations*. Computer Physics Communications, 2001. 135(1): p. 40-57.

## 2.8 Bibliography

---

35. Kumar, S., et al., *THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method*. Journal of Computational Chemistry, 1992. 13(8): p. 1011-1021.
36. Zhu, F. and G. Hummer, *Convergence and error estimation in free energy calculations using the weighted histogram analysis method*. Journal of computational chemistry, 2012. 33(4): p. 453-465.
37. Rajamani, R., K.J. Naidoo, and J. Gao, *Implementation of an adaptive umbrella sampling method for the calculation of multidimensional potential of mean force of chemical reactions in solution*. Journal of Computational Chemistry, 2003. 24(14): p. 1775-1781.
38. Bartels, C., M. Schaefer, and M. Karplus, *Adaptive umbrella sampling of the potential energy: modified updating procedure of the umbrella potential and application to peptide folding*. Theoretical Chemistry Accounts, 1999. 101(1): p. 62-66.
39. Meng, Y. and B. Roux, *Efficient Determination of Free Energy Landscapes in Multiple Dimensions from Biased Umbrella Sampling Simulations Using Linear Regression*. Journal of Chemical Theory and Computation, 2015. 11(8): p. 3523-3529.
40. Mahdavi, S. and S. Kuyucak, *Mechanism of Ion Permeation in Mammalian Voltage-Gated Sodium Channels*. PLOS ONE, 2015. 10(8): p. e0133000.
41. Akhshi, P. and G. Wu, *Umbrella sampling molecular dynamics simulations reveal concerted ion movement through G-quadruplex DNA channels*. Physical Chemistry Chemical Physics, 2017. 19(18): p. 11017-11025.
42. Laio, A. and M. Parrinello, *Escaping free-energy minima*. Proceedings of the National Academy of Sciences, 2002. 99(20): p. 12562.
43. Valsson, O., P. Tiwary, and M. Parrinello, *Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint*. Annual Review of Physical Chemistry, 2016. 67(1): p. 159-184.
44. Barducci, A., G. Bussi, and M. Parrinello, *Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method*. Physical Review Letters, 2008. 100(2): p. 020603.
45. Kumar, P.P., A.G. Kalinichev, and R.J. Kirkpatrick, *Dissociation of carbonic acid: Gas phase energetics and mechanism from ab initio metadynamics simulations*. The Journal of Chemical Physics, 2007. 126(20): p. 204315.
46. Petersen, L., et al., *Mechanism of Cellulose Hydrolysis by Inverting GH8 Endoglucanases: A QM/MM Metadynamics Study*. The Journal of Physical Chemistry B, 2009. 113(20): p. 7331-7339.
47. Alfonso-Prieto, M., et al., *The Molecular Mechanism of the Catalase Reaction*. Journal of the American Chemical Society, 2009. 131(33): p. 11751-11761.
48. Fiorin, G., et al., *Using Metadynamics to Understand the Mechanism of Calmodulin/Target Recognition at Atomic Detail*. Biophysical Journal, 2006. 91(8): p. 2768-2777.

49. Humphrey, W., Dalke, A. and Schulten, K., *VMD - Visual Molecular Dynamics*. J. Molec. Graphics. 14: p. 33-38.
50. *The PyMOL Molecular Graphics System*. Schrödinger, LLC.



# Chapter 3

## Hidden Electrostatic Basis of Dynamic Allostery in a PDZ3 Domain

### 3.1 Introduction

Allosteric regulation of proteins plays a key role in physiological cell functions, biochemical and signal transduction pathways and drug discovery<sup>1-3</sup>. It has remained a challenge to understand how the thermodynamic perturbation caused by ligand binding at one site would propagate and modulate the structure and dynamics of distal regions of proteins. The prevailing models of structure-based allostery<sup>4,5</sup> do not apply to the more recent examples of allostery without conformational change such as PDZ domain<sup>6</sup>, CAP dimer<sup>7</sup> and met repressor<sup>8</sup>. These examples have triggered the concept of “dynamic allostery”, where the side-chain dynamics is modulated on ligand binding, and the origin has often been attributed to changes in the conformational entropy<sup>9,10</sup>. The modern view of allostery invokes a thermodynamic picture, where a “population shift” among pre-existing conformational states occurs upon binding the allosteric effector<sup>11,12,13</sup>. It has also been suggested in the context of allostery without conformational change that “not observed does not imply that it is not there”<sup>10</sup> since crystallographic techniques may not resolve the relatively minor population shifts. An interesting idea has emerged that all proteins might be allosteric in nature!<sup>14</sup>

PDZ domain has been a classic model system to study single domain allostery without major structural changes<sup>9,15,16</sup>. PDZ domains are evolutionary conserved protein-protein interaction modules associated with the cellular signaling and are implicated in localization of membrane receptors and ion-channels<sup>17,18</sup>. They can dimerise with other modular protein domains (e.g. WW, SH2, SH3, PH etc.) or can bind specific recognition sequences at the C-terminus of proteins in a hydrophobic groove between  $\beta 2$ - $\alpha 2$  regions<sup>17,19</sup>. At present, more than 200 structures of PDZ domains and its complexes are available in PDB database which provides detailed information about the specificity and selectivity in ligand binding. PDZ domains exhibit

### 3.1 Introduction

---

a highly conserved carboxylate-binding loop which forms a hydrophobic pocket for protein-protein and protein-ligand interactions<sup>17,19,20</sup>. The loop is characterised by a conserved motif R/K-XXX-G- $\Phi$ -G- $\Phi$ , (where X is any amino acid and  $\Phi$  is any hydrophobic residue) which recognises specific C-terminal motifs of the binding partner. PDZ domains are divided into three main classes<sup>17</sup>: (1) class I domains recognise the motif S/T-X- $\Phi$ ; (2) class II domains recognise the motif  $\Phi$ -X- $\Phi$ ; and (3) class III domains recognise the motif D/E-X- $\Phi$ . The variability in the motif allows the interaction of PDZ domains with multiple proteins in the signaling pathway. Allosterity has been one of the profound regulatory mechanisms of PDZ protein-protein interactions. The signature of the allosteric effects observed in PDZ domains has been purely in terms of the dynamics of the side chains. Solution NMR studies<sup>20,21</sup> as well as molecular dynamics (MD) simulations<sup>22,23</sup> have confirmed that binding the effector ligand leads to substantial modulation of side chain dynamics in the PDZ domain. In particular, Lee and coworkers have unearthed a “hidden dynamic allosterity” in the PDZ3 domain, where deletion of a non-canonical distal  $\alpha$ 3 helix domain reduces the ligand binding affinity by 21 times<sup>6</sup>. Their work has highlighted the role of differential side chain motions towards the allosteric response and attributed the origin of dynamic allosterity to purely entropic effects since the enthalpic contribution towards change in binding affinity upon  $\alpha$ 3 helix deletion was minimal. Interestingly, they have also hinted towards a possibility that internal structural adjustments could lead to cancellations in individual changes of enthalpy, which will be demonstrated to be the case in our work.

Prior simulation studies have attempted to understand the allosteric communication pathways in terms of correlations in structural or energetic fluctuations in PDZ domain<sup>24,25</sup>. Karplus and coworkers have revealed two continuous correlation pathways in a PDZ2 domain and highlighted the existence of such pathways even in the absence of the ligand<sup>24</sup>. A number of theoretical<sup>23,26,27</sup> and experimental studies<sup>6,28</sup> propose multiple allosteric pathways for PDZ domains based on evolutionary information<sup>29,30</sup>, local structural changes<sup>22,31,32</sup>, heat diffusion pathways<sup>33,34</sup> and energy connectivity networks<sup>35,36</sup>. Most of the existing approaches look for correlated motions or energy fluctuations to characterise the allosteric effects. While such correlations may demonstrate the effect of allosterity, they do not explain the specific origin of such coupling. It has also been debated that the statistical analysis of evolutionarily coupled residues may not be the true reporter of functional coupling since the evolutionary information

does not include the molecular details of the interactions<sup>31,37</sup>. Thus, a molecular thermodynamic approach that utilises the perturbations in the non-bonded interactions upon ligand binding would provide a more direct view of the functional energetic coupling between the protein residues<sup>38-40</sup>.

In this chapter, we argue that protein dynamics is governed by the underlying energy landscape. Our objective is to understand the perturbation in the internal energy landscape of the allosteric protein due to ligand binding and how that manifests into functional changes in distal sites. To achieve this goal, we performed atomistic MD simulations both in the ligand bound and unbound states of a PDZ3 domain protein and compared the changes in non-bonded interactions energies for each residue as well as contributions from individual pair-wise interactions. We track the large energetic perturbations upon ligand binding as signature of allosteric effects. We show that despite the subtle structural changes, the protein-protein and protein-water electrostatic interactions undergo dramatic re-distribution upon ligand binding. We construct a residue-pair-wise energetic perturbation network where the binding site and distal allosteric regions are connected by the non-canonical  $\alpha 3$  helix, which has been suggested to play a significant role in the dynamic allostery in PDZ3 domain<sup>6</sup>. Finally, we elucidate the molecular basis of this perturbation network to be a “population shift” between the pair-wise hydrogen bonded network involving the protein side-chains to be the primary reason behind the observed thermodynamic and dynamical effects. Based on our observations, we suggest a “hidden energetic allostery” driven by the population shift in specific electrostatic interactions and internal re-distribution of non-bonded interactions to be the driving force behind the “dynamic allostery” in PDZ3 domain.

## 3.2 Methods

### 3.2.1 System Setup

The peptide bound and unbound structures for MD simulations were obtained from the crystal structure of PDZ3 domain bound with peptide CRIPT (PDB ID: 1BE9)<sup>19</sup>. The unbound structure was obtained by removing the peptide CRIPT (KQTSV) and equilibrating for 100ns. The C-terminus of the protein and N-termini of both protein and ligand were capped by N-methyl amide and acetyl groups, respectively.

## 3.2 Methods

---

### 3.2.2 Simulation Parameters

All MD simulations were performed using GROMACS 5.0.7 software<sup>41</sup> with Amber99SB-ILDN force field<sup>42</sup> and TIP4P-Ew water model<sup>43</sup>. The resultant solvated boxes contained around 8800 water molecules for all the systems. The protonation states for the titrable residues were determined using MCCE method<sup>44</sup> as follows: all Asp, Glu, Arg and Lys residues were charged, and His residues were neutral. Although the MCCE protocol identified a minor population of the protonated charged species for the two histidine residues, we have decided to use the neutral species with major population for both bound and unbound states. The systems were neutralised by adding appropriate number of Na<sup>+</sup> ions. The structures were energy minimised followed by two-step equilibration, namely NVT equilibration followed by NPT equilibration. Temperature was controlled through velocity rescaling<sup>45</sup> at 300K with a time constant of 0.1 ps and pressure was controlled using Parrinello-Rahman barostat<sup>46</sup> at 1bar. The particle mesh Ewald algorithm<sup>47</sup> was applied to calculate long-range electrostatic interactions. The cutoff for short-range electrostatics and van der Waals' interaction was 1.0 nm. Four independent MD simulations of 500ns each were performed (total 2.0 $\mu$ s) for the PDZ3 domain in the bound and unbound states with LINCS constraints for all bonds<sup>48</sup> and frames were recorded at every 2ps.

### 3.2.3 Analysis

#### Differential contact map

Two residues were defined to be in contact if the distance between any two atoms of these residues was less than 4.5 $\text{\AA}$ <sup>49</sup>. Since a particular contact may form and break during the course of a dynamic trajectory, we defined a “contact frequency map” by  $f_{ij} = n_{ij}/N$ , where  $n_{ij}$  is the number of frames where the residues  $i$  and  $j$  were in contact and  $N$  is the total number of frames. Thus,  $f_{ij} = 1$  for a contact that is present in all the frames. The “differential contact map” ( $C_{ij}$ ) is the difference between the contact frequency map ( $f_{ij}$ ) obtained from simulations of the bound state and the unbound state:  $C_{ij} = f_i^{bound} - f_i^{unbound}$ , where the values of  $C_{ij}$  would lie between -1 and +1, where  $C_{ij} = -1$  and  $C_{ij} = +1$  would indicate a contact between residue pairs  $i$  and  $j$  to be present exclusively in unbound and bound states exclusively, respectively.



#### Perturbation in non-bonded interaction energies

The average non-bonded interaction energy was computed for each residue ( $E_i$ ) as well as all residue pairs ( $E_{ij}$ ) and compared between the ligand bound and unbound states in the following manner. The change in average non-bonded energy of  $i$ -th residue is given by:  $\Delta E_i = \langle E_i \rangle_{bound} - \langle E_i \rangle_{unbound}$ , where the  $\langle \rangle$  notation indicates an ensemble average over the trajectory for that particular state (bound/unbound). Note that this difference in average energy can be further broken down in terms of the contributions from ligand, protein and water in the following manner:

$$\Delta E_i = \Delta E_i^{protein} + \Delta E_i^{ligand} + \Delta E_i^{water} \quad (3.1)$$

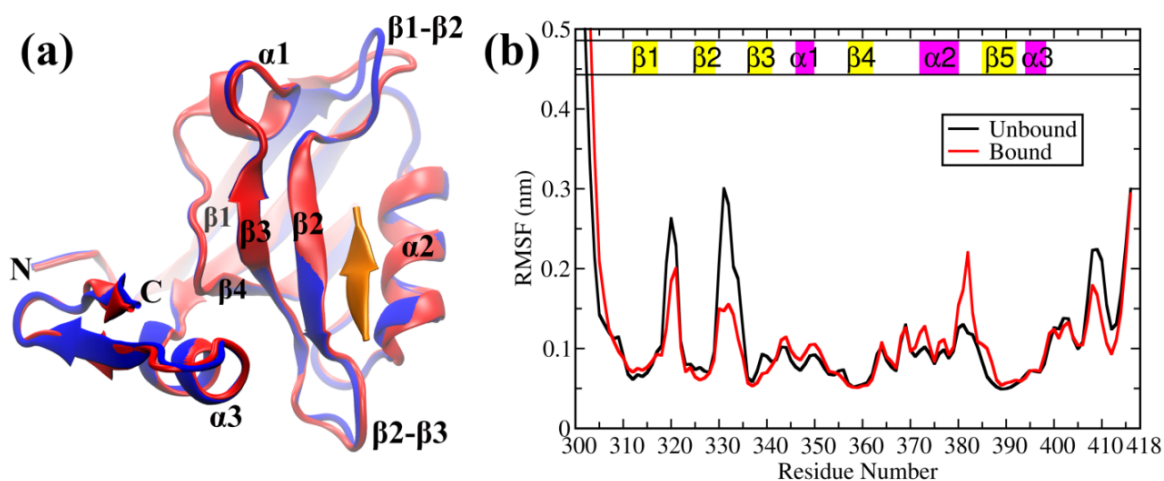
$$\Delta E_i = E_i^{ligand,b} + \left[ E_i^{protein,b} - E_i^{protein,u} \right] + \left[ E_i^{water,b} - E_i^{water,u} \right] \quad (3.2)$$

where the terms  $\Delta E_i^{protein}$ ,  $\Delta E_i^{ligand}$  and  $\Delta E_i^{water}$  denote the change in the average non-bonded interaction energy between the bound (b) and unbound (u) state due to the interactions between the  $i$ -th residue and ligand/protein/water, respectively. Here the interaction energies were calculated for all atoms of protein and ligand, whereas for computing the interaction energy with water molecules, a large cut-off of 2nm has been used. We have separately computed the contributions due to Lennard Jones (LJ) and electrostatic (Coulomb) non-bonded interactions to  $\Delta E_i$ , but the LJ terms were generally found to be numerically much smaller than the respective electrostatic terms, so we have primarily focused on the electrostatic interactions while dissecting the perturbation in pair-wise interactions  $\Delta E_{ij}$ . This helped us to discover the significant role of specific electrostatic interactions towards observed allosteric response as discussed later.

## 3.3 Results

## Side-chain rearrangement leads to dramatic change in energetics

The basic premise of the “dynamic allostery” phenomenon has been the lack of structural changes between the ligand bound and unbound states as observed in the respective crystal structures. To be precise, the presence or absence of significant structural changes is evaluated based only on the backbone structure, and side-chains are not invoked in this description in general. As a first step towards characterising the structural features and differences (if any) between the available crystal structures for the PDZ3 domain with and without the ligand (PDB ID: 1BE9 and 1BFE, respectively), we performed energy minimisation of the crystal structures while restraining the position of the backbone atoms. Only the side chain atoms were allowed to move in order to investigate the significance of side-chain rearrangement on the interaction pattern. In order to maintain a consistent comparison, we used an identical length of the two protein structures (residues 306 to 415). Fig. 3.1a shows a superposition of the structures (bound and unbound) after energy minimisation, which preserves the well-known characteristics of the crystal structures that there is almost no structural difference (in backbone) between the two states except minor rearrangement in the  $\beta 1$ - $\beta 2$  and  $\beta 2$ - $\beta 3$  loop regions.



**Figure 3.1** (a) Superposition of the structures after energy minimisation of the respective crystal structures in the unbound and bound states (PDB IDs: 1BFE and 1BE9, respectively). Bound and unbound states are coloured in red and blue, respectively. The peptide ligand is highlighted in orange. (b) Residue-based RMSF profile for PDZ3 domain in the unbound state (black) and bound state (red). There is overall decrease in fluctuations in the ligand bound state as

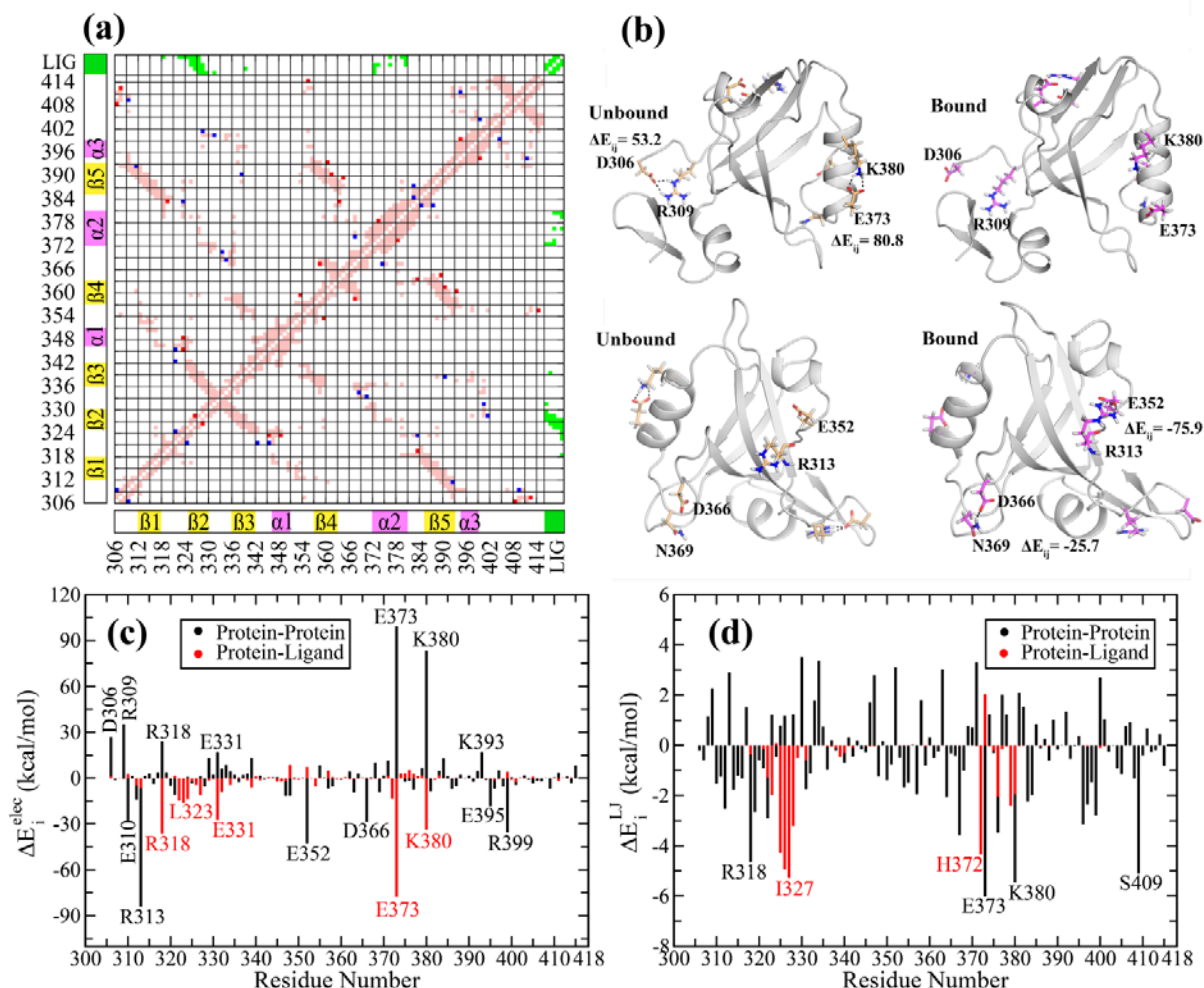
*compared to the unbound state exhibiting flexibility in the domain in the absence of ligand, and subsequently, ligand binding activity quenches fluctuations in the loop regions prominently.*

Subsequently, we turn our attention to the possible rearrangement of the side-chains. Looking for differences in the residue pair-wise contact map has been a popular choice for identifying the significant interaction pathways. After all, the non-bonded interactions are intimately coupled to the presence of contacts and their rearrangements. As a matter of fact, in the context of allosteric coupling it has been suggested that the residues which are in spatial proximity forming contacts are more likely to be coupled as compared to the distal residues<sup>37</sup>, and models have been proposed to construct a biophysical framework based on contacts acting as structural support for the propagation of information<sup>50-52</sup>. We took a similar approach of identifying the possible differences in the side-chain contacts between the two end states by constructing the differential contact map in Fig. 3.2a. The contacts present exclusively in the ligand bound and unbound states are marked in red and blue, respectively. This visual representation clearly indicates that majority of the pair-wise contacts remain the same (light pink regions), whereas there are a few inter-residue contacts which completely disappear or appear on ligand binding. Evidently, the side chain contact map reveals relatively minor structural rearrangement in the contact pattern.

Finally, we turn our attention to the arguably most fundamental parameter: non-bonded interaction energy. Ligand binding is likely to cause significant local energetic perturbation in the binding site of the protein, which should propagate through the intra-protein interaction network to the allosteric site. Thus we have investigated the changes in the residue-wise non-bonded interaction energy ( $\Delta E_i$ ) with the rest of the protein (and ligand). Figs. 3.2c and 3.2d present the electrostatic ( $\Delta E_i^{elec} = E_i^{elec,bound} - E_i^{elec,unbound}$ ) and van der Waals' ( $\Delta E_i^{LJ} = E_i^{LJ,bound} - E_i^{LJ,unbound}$ ) components of the change in total interaction energy of the  $i$ -th residue upon ligand binding. Remarkably, we find that there is a wide range of variation in the change of electrostatic energy ( $\Delta E_i^{elec}$ ) (up to  $\pm 90$  kcal/mol) for both protein-only and ligand-only interactions. On the other hand, the changes in van der Waals' interaction energy ( $\Delta E_i^{LJ}$ ) show relatively modest range of variation up to  $\pm 6$  kcal/mol. Evidently, the non-bonded interactions, particularly the electrostatic interaction energy, provide a highly sensitive probe towards identifying the subtle structural changes. Thus, even though the backbone and side chain

### 3.3 Results

structural parameters (e.g. RMSD, cut-off based contact map etc.) show a minor change, the interaction energies show a substantial change between the two states and in our subsequent analysis we shall further demonstrate that the electrostatic interaction energy indeed captures the nature of allosteric coupling in the PDZ3 domain. We further emphasize that our analysis proves that indeed there is structural change at the side chain level (even in the crystal structure) that can be significant in terms of energetics (Fig. 3.2b).



**Figure 3.2** (a) Differential contact map between the bound and unbound states (energy minimised structures). Contacts unique in bound (including the contacts with the ligand) and unbound states are shown in red and blue, respectively. The common contacts are shown in light pink. The secondary structural elements have been highlighted along both axes as visual guides. (b) Energetic perturbation based on minimised crystal structures with identical backbone structure. The figure highlights four such pairs of residues with large  $\Delta E_{ij}$  as a consequence of

### 3. Hidden Electrostatic Basis of Dynamic Allostery

---

*explicit hydrogen bond formation between the unbound and bound states. (c) Electrostatic (Coulomb) and (d) van der Waals' (Lennard Jones) components of the residue-wise change in total interaction energy between the bound and unbound states ( $\Delta E_i = E_i^{\text{bound}} - E_i^{\text{unbound}}$ ). A few residues with significant change between the two states have been marked on each figure. The  $\Delta E_i$  due to interaction with ligand ( $\Delta E_i^{\text{ligand}}$ ) and other protein residues ( $\Delta E_i^{\text{protein}}$ ) have been marked in red and black, respectively.*

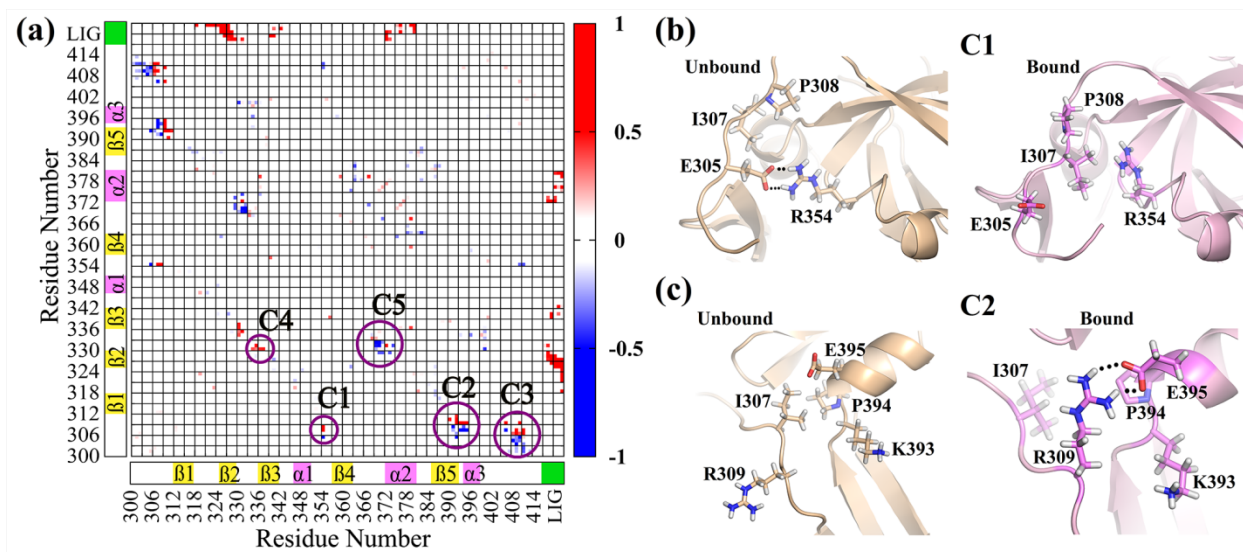
The results presented in Fig. 3.2 confirm that the inherent energy landscapes of the bound and unbound states can be drastically different even when the backbone structures are almost identical. The subtle rearrangement and rewiring of the side chain interactions can lead to dramatically different energetic coupling (Fig. 3.2b). But the crystal structures may not be a true representative of the respective structural ensemble in solution due to crystal packing and crystallisation conditions. Moreover, a crystallographic view of single static structure often does not capture the biologically significant ensemble of conformations and dynamic fluctuations. Thus, in order to investigate the signatures of “dynamic allostery” in PDZ3 domain, we shall base our subsequent analyses on the MD simulation trajectories to capture the “population” of the significant interactions (between the side chains) that might lead to allosteric modulation.

The local dynamical variations in the PDZ3 domain has been captured in the residue-wise root mean square fluctuations (RMSF) with respect to the average structure for both the ligand bound and unbound states (Fig. 3.1b). Overall the fluctuations decrease in the presence of the ligand in the  $\beta 1$ - $\beta 2$  and  $\beta 2$ - $\beta 3$  loop regions and towards the C-terminal region. In contrast, the fluctuation increases in the presence of ligand around residues 381-385 ( $\alpha 2$ - $\beta 5$  loop). This emphasizes greater flexibility and plasticity in the unbound state as compared to the bound state as already discussed in the literature<sup>25</sup>. Although it has been well established that the overall dynamics of the PDZ3 domain is modulated upon ligand binding (hence the term “dynamic allostery”), but the reason behind it is not completely clear. We shall argue that dynamics is governed by the underlying energy landscape. Thus, understanding the perturbation or modulation in the intra-protein interaction network is likely to provide the answer.

### 3.3 Results

#### Contact maps do not capture the specific nature of contacts (ionic/polar/non-polar)

We have shown in Fig. 3.2a that the side-chain contact maps undergo only a minor rearrangement. Of course, proteins in solution are flexible entities, and thus contacts may break/form with certain population. In order to understand the differences in the contact pattern in the dynamic system, we have computed the differential contact frequency map between the trajectories for bound and unbound states (Fig. 3.3a). Five major regions (C1 to C5) are marked on Fig.3.3a corresponding to the significant changes in the intra-protein contact map (involving both backbone and side chain atoms). Red and blue regions signify exclusive contacts present in bound and unbound states, respectively.



**Figure 3.3** (a) Differential contact frequency map between the bound and unbound states as obtained from the MD simulation trajectories. The regions with high differential contact frequency ( $|C_{ij}| > 0.5$ ) are highlighted as clusters (C1 to C5). Positive values (red regions) and negative values (blue regions) indicate exclusive contacts present in the bound and unbound states, respectively. The values of differential contact frequency of these residue pairs have been shown in Table 3.1. (b) The representative snapshots showing changes in contact pattern between unbound and bound states for cluster C1 (ionic: R354-E305 to non-polar: R354-I307) and (c) C2 (non-polar: E395-I307 to ionic: E395-R309).

We notice that the N-terminus region (residues 305-309) is involved in contact(s) with the  $\alpha$ 1- $\beta$ 4 (C1),  $\beta$ 5- $\alpha$ 3 (C2) and C-terminus (C3) loop/coil regions. For all these clusters, there is

an upward movement (N- to C- terminus) on going from unbound to bound state. This minor change in contact pattern seems insignificant since the contact map is agnostic to the specific nature of the contacts (ionic/polar/non-polar). Let us focus on two specific cases for the clusters C1 (Fig. 3.3b) and C2 (Fig. 3.3c) to understand the nature of interactions between these contacts. For cluster C1: there exists a salt bridge (ionic interaction) between the oppositely charged species E305 (negative)-R354 (positive) in the unbound state, whereas in the bound state this converts to non-polar contacts between hydrophobic I307/P308 and charged R354 (positive). Similarly, for the cluster C2: there is non-polar contact between I307 (hydrophobic) and E395 (negative) in the unbound state, and this converts to a salt bridge between R309 (positive) and E395 (negative) in the bound state.

Note that in both the cases, the contact region shifts by only 2 residues in the N terminal region, whereas the specific nature of the interaction dramatically changes between non-polar to ionic and vice versa. Thus, we argue that looking at purely structural parameters (distance/position based) is not enough to understand the allosteric modulation unless the chemical identity (charge distribution) is being considered. Thus, in subsequent sections, we shall exclusively focus on the energetics as the yardstick to identify the interaction network that connects the binding site to the allosteric site. Here the non-bonded interaction energy captures both the structural parameters (through distance dependence) and chemical identities (through partial charges for electrostatic and Lennard-Jones parameters for van der Waals' interactions) of the molecular systems.

### 3.3 Results

**Table 3.1** List of residues with net contact frequency ( $|C_{ij}| \geq 0.5$ ), positive value would indicate a contact between residue pairs  $i$  and  $j$  to be present exclusively in bound state and negative value states contact is predominantly formed in unbound state.

ResID	Residues list with net contact frequency $ C_{ij}  \geq 0.5$
G303	S409 (-0.6)
E304	S408 (-0.5)
E305	R354 (-0.8), Y392 (-0.8), G410 (-0.7)
D306	S409 (0.7), R411 (0.6)
I307	P394 (-0.9), K393 (-0.9), V406 (-0.8), S409 (0.7), E395 (-0.6), R411 (0.9), R354 (0.9)
P308	R354 (0.6), Q391 (-0.9)
R309	Y392 (0.9), K393 (0.8), P394 (0.9), E395 (0.7), V406 (0.8), G410 (0.8)
E310	Y392 (0.9)
E331	G335 (0.7), A370 (-0.5)
D332	N369 (-0.6)
G335	E331 (0.7)
R354	E305 (-0.8), I307 (0.9), P308 (0.6)
N369	D332 (-0.6)
A370	E331 (-0.5)
Q391	P308 (-0.9)
Y392	E305 (-0.8), R309 (0.9), E310 (0.9)
K393	I307 (-0.9), R309 (0.8)
P394	I307 (-0.9), R309 (0.9)
E395	I307 (-0.6), R309 (0.7)
V406	I307 (-0.8), R309 (0.8)
S408	E304 (-0.5)
S409	G303 (-0.6), D306 (0.7) I307 (0.7)
G410	E305 (-0.7), R309 (0.8)
R411	D306 (0.6), I307 (0.9)
K5	I328 (0.8), G329 (0.7), H372 (0.8)
Q6	N326 (1.0), I327 (0.9), I328 (0.9), S339 (0.8), F340 (0.7), H372 (0.7)
T7	F325 (0.7), N326 (1.0), I327 (1.0), I328 (0.9), H372 (1.0), A376 (1.0)
S8	F325 (1.0), N326 (1.0), I327 (1.0), L342 (0.7), K380 (0.6)
V9	R318 (0.8), T321 (0.7), G322 (1.0), L323 (1.0), G324 (1.0), F325 (1.0), N326 (1.0), I327 (1.0), A376 (0.9), L379 (1.0), K380 (0.7)



#### Electrostatic energy is the key determinant of allosteric modulation in PDZ3 domain

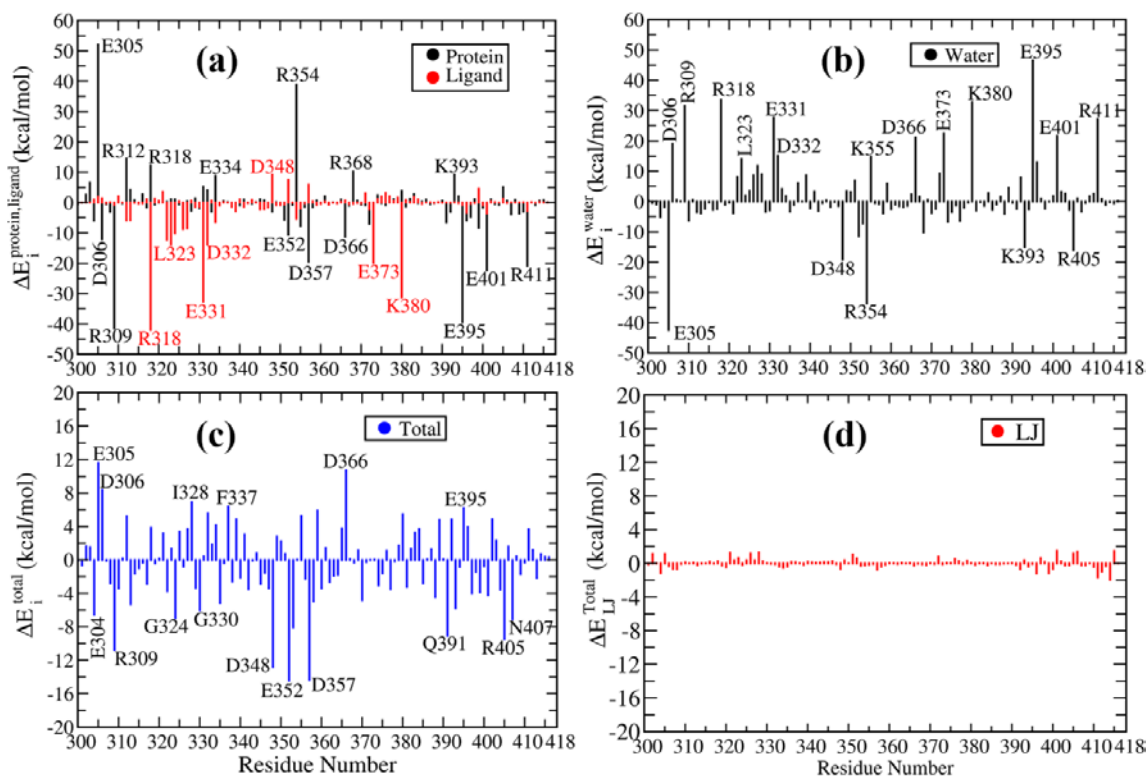
So far, we have established that non-bonded interactions provide the most sensitive yardstick to capture the structural rearrangement as compared to any purely position/distance-based parameters. Now we are going to demonstrate the perturbation in the intra-protein interaction network caused by the ligand and how this perturbation might propagate through pair-wise interactions between protein residues. For this purpose, we have evaluated the average interaction energy of each residue with the various components of its environment, namely protein, ligand and water. The changes in each of these energy terms were calculated between the ligand bound and unbound states. All of these results are summarized in Fig. 3.4a: protein-only and ligand-only, Fig. 3.4b: water-only and Fig. 3.4c: total (protein+ligand+water). Here we focus on only the electrostatic component of the respective  $\Delta E_i = \langle E_i \rangle_{bound} - \langle E_i \rangle_{unbound}$  terms in Fig. 3.4. Corresponding van der Waals' (Lennard Jones) terms are much smaller in magnitude ( $< 3$  kcal mol) as shown in Fig. 3.4d.

Our results unequivocally prove that the electrostatic interaction energy provides a sensitive yardstick towards capturing the allosteric effects in this system. While the van der Waals' interactions are crucial towards attaining the functional structure of the protein, the allosteric modulation due to ligand binding seems to be strongly associated with the electrostatic interactions. Our findings provide further support to the already established view regarding the significant role of electrostatic interactions in biomolecular functions including allostery<sup>53-58</sup>. Evidently, a negative  $\Delta E_i$  signifies an interaction more favorable in the bound state as compared to the unbound state and vice versa. In Fig. 3.3a we can easily recognise the residues interacting strongly and favorably with the ligand (red lines), e.g. R318, E331, E373 and K380 have contributions more negative than  $-30$  kcal/mol. On the other hand, there are residues (black lines) that do not interact with the ligand at all or spatially far away from the ligand, but still exhibit very large magnitude of  $\Delta E_i$ , e.g. E305, R309, R354, E395, E401 etc. These residues clearly demonstrate the effect of "allosteric" modulation in terms of their energetics. Thus, interactions with the ligand are leading to internal structural rearrangements (or population shift) in the protein in such a way that the intra-protein interaction network is significantly perturbed.

Another interesting observation to make here is the contribution from the ligand and protein-only interactions towards  $\Delta E_i$  are often in the reverse direction, e.g. R312, R318, E331,

### 3.3 Results

E334, E352 etc. The favorable interaction with the ligand ( $\Delta E_i^{ligand} < 0$ ) forces the side-chains of these residues to attain certain orientations that lead to unfavorable interactions (or break previously favorable interactions) with the other residues of the protein ( $\Delta E_i^{protein} > 0$ ). Such locally unfavorable protein-protein interaction is likely to initiate further downstream rearrangement to release the energetic stress, much like a domino effect! We must also note that there are positive  $\Delta E_i^{protein}$  values for a few residues, e.g. E305, R312, R318, E334, R354, R368, K393 etc. This signifies that there exist certain favorable interactions for these residues in the unbound state, which was broken in the process of ligand binding and subsequent allosteric modulation. We shall further dissect the molecular basis of these long range perturbations in the subsequent sections.



**Figure 3.4** Residue-wise changes in average electrostatic energy between bound and unbound states ( $\Delta E_i^{elec} = \langle E_i^{elec} \rangle_{bound} - \langle E_i^{elec} \rangle_{unbound}$ ) as obtained from MD simulation trajectories: (a) The contribution due to ligand and rest of the protein have been shown separately as red and black lines, respectively. (b) Contribution due to water only, where average interaction energy between the  $i$ -th residue and water molecules with 2nm cut-off radius has been used to compute

$\Delta E_i^{water}$ . (c) The  $\Delta E_i$  due to the full environment comprising of ligand, protein and water molecules within 2nm, i.e.  $\Delta E_i = \Delta E_i^{protein} + \Delta E_i^{ligand} + \Delta E_i^{water}$ . (d) Van der Waals' (Lennard Jones) component of the residue-wise change in total interaction energy between the bound and unbound states  $\Delta E_{LJ}^{Total} = \langle E_{i,LJ}^{Total} \rangle_{bound} - \langle E_{i,LJ}^{Total} \rangle_{unbound}$ . The change in interaction energy is less than 3 kcal/mol which is relatively very low as compared to the difference in the electrostatic interaction energy between the bound and unbound state. This highlights the major contribution of electrostatic effect in the allosteric phenomenon upon ligand binding.

Prior studies on allostery have primarily focused on the structure and dynamics of the protein itself; whereas a few studies highlight the effect of solvation forces and water mediated interactions towards allosteric modulation<sup>59,60</sup>. Fig. 3.4b demonstrates the modulation in the residue-wise solvation energy (to be precise, the average electrostatic interaction energy with water) upon ligand binding. Here the electrostatic interaction energy for each residue was calculated with respect to the water molecules within 2nm cutoff. As expected, the solvent exerts a dielectric screening effect that goes in the reverse direction of the electrostatic contributions due to protein and ligand only. The residue-wise changes in the total interaction energy including the solvation energy ( $\Delta E_i = \Delta E_i^{protein} + \Delta E_i^{ligand} + \Delta E_i^{water}$ ) is shown in Fig. 3.4c. Evidently, the contributions due to the charged residues still remain the most significant. In addition, Fig. 3.4c highlights the relatively large  $\Delta E_i$  values and oscillatory patterns for the residues in and around the binding site, e.g. G324, I328, G330, F337 etc. Most of these residues have favorable interaction with the ligand, but an unfavorable desolvation penalty associated with ligand binding. The local solvation environment may have substantially altered for certain residues, e.g. for residues D306, D348, E352, D366, R405 we observe that  $|\Delta E_i^{water}| > |\Delta E_i^{protein}|$ . This implies that the solvation energy may overcompensate the changes due to protein-protein interactions in these cases due to redistribution in the protein-protein versus protein-water interaction pattern. Our future work would dissect the more specific roles of water mediated interactions towards allosteric regulation.

An interesting feature of the total  $\Delta E_i$  reported in Fig. 3.4c is that there is an oscillatory pattern of residues with large positive  $\Delta E_i$  and large negative  $\Delta E_i$  values. This indicates that the

### 3.3 Results

---

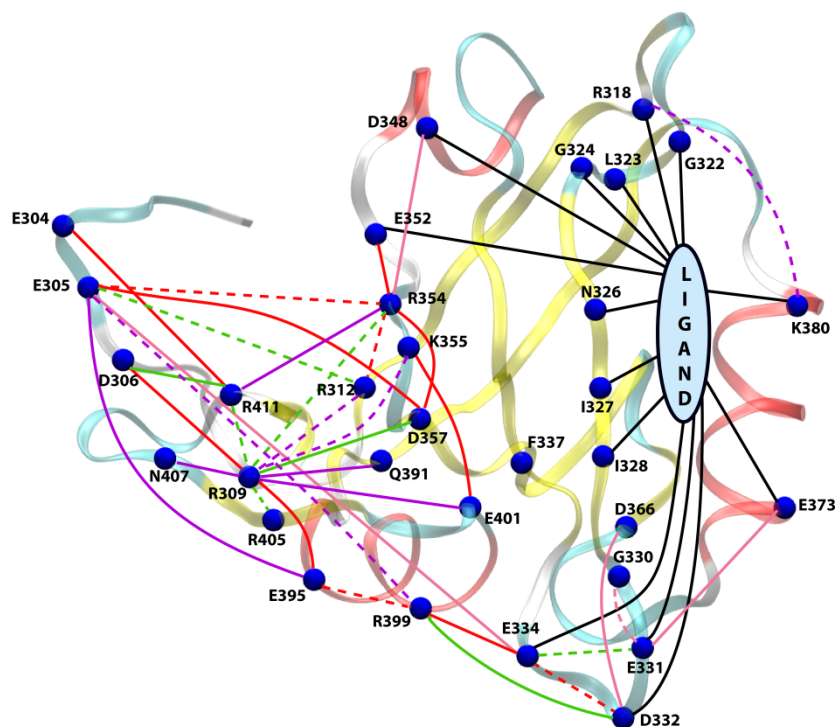
ligand binding leads to a massive internal rearrangement or redistribution of the non-bonded interactions. There is a clear separation of the residues in terms of favorable and unfavorable interactions. Due to these cancellation effects, the total non-bonded interaction energy of the protein is perturbed to a much lower extent as compared to the local perturbation or energy redistribution at the residue level. Thus, it is not that enthalpy does not play a role in the “dynamic allostery” observed in PDZ3 domain; rather it plays a significant role in terms of the local rearrangement and rewiring of the specific interactions.

#### **Energetic perturbation network connects the binding site and allosteric site**

Now that we have identified the protein residues that undergo a magnificent energetic perturbation between the bound and unbound states, we attempt to dissect the residue-pair-wise contributions towards this change so that we can build a connectivity network of the energetic perturbation. In order to achieve this, we have dissected the  $\Delta E_i^{protein}$  (shown in Fig. 3.4a) into all residue-pair-wise contribution terms:  $\Delta E_{ij}$ , where  $\Delta E_{ij} = \langle E_{ij} \rangle_{bound} - \langle E_{ij} \rangle_{unbound}$ . Note that  $\Delta E_{ij}$  involves the short range electrostatic component of the interaction only. The numerical values of  $\Delta E_{ij}$  for residues with  $|\Delta E_i| > 6$  kcal/mol have been shown in Table 3.2. The  $\Delta E_{ij}$  values are summarized and visualized in Fig. 3.5, which now leads to an energetic perturbation network that dramatically connects the ligand binding site to the allosteric side (distal regions) of PDZ3 domain. The caption of Fig. 3.5 provides a detailed description of the visualization scheme used to build the network.

A detailed analysis of the network presented in Fig. 3.5 gives us a multitude of significant insights into the nature of the communication pathways between the binding site and allosteric site: (i) Direct perturbation by ligand: The solid black lines indicates that the ligand perturbs most of the binding site residues (in  $\beta 1$ - $\beta 2$  loop,  $\beta 2$  sheet and  $\alpha 2$  helix (K380, E373)), but also induces changes in energetics of the  $\beta 2$ - $\beta 3$  loop and distant residues (D348, E352) through long range electrostatic interactions. Our analysis successfully captures the effect on residues already known for their involvement in ligand binding e.g. R318, G322, N326, I327, I328, D332, K380<sup>27,29,31,33,35</sup> (ii) Perturbation in intra-protein interactions: Interestingly, we can observe an extensive energetic redistribution at the distal side (possibly, the allosteric site) comprised of the N- and C-termini,  $\alpha 3$ -helix and  $\alpha 1$ - $\beta 4$  unstructured regions. The solid lines indicate the

interactions that have become more favourable in the ligand bound state as compared to unbound state, and the dashed lines indicate more unfavourable ones.



**Figure 3.5** A comprehensive network view of the perturbation in pair-wise electrostatic interaction energies ( $\Delta E_{ij} = \langle E_{ij} \rangle_{\text{bound}} - \langle E_{ij} \rangle_{\text{unbound}}$ ). Visualization scheme: (i) The blue spheres indicate residues with  $|\Delta E_i^{\text{total}}|$  or  $|\Delta E_i^{\text{ligand}}| > 6$  kcal/mol. A few residues with large  $|\Delta E_{ij}|$ , but  $|\Delta E_i^{\text{total}}| < 6$  kcal/mol have been highlighted as blue spheres as well, e.g. R312, R354, K355, R399, E401 and R411. (ii) Connections with negative and positive  $\Delta E_{ij}$  values are indicated with solid and dashed lines, respectively, i.e. a solid (or dashed) line indicates a contact more (or less) favourable in bound (or unbound) state. (iii) The connections are coloured on the basis of magnitude of  $|\Delta E_{ij}| > 10$  kcal/mol (red),  $> 6$  kcal/mol (green),  $> 4$  kcal/mol (purple) and  $> 3$  kcal/mol (pink). The black lines represent connections between peptide ligand and residues with  $|\Delta E_{ij}| > 6$  kcal/mol. The connections based on  $\Delta E_{ij} > 3$  kcal/mol were considered only for residues which are directly perturbed upon ligand binding. The  $\Delta E_{ij}$  values for all significant pairs have been reported in Table 3.2.

### 3.3 Results

---

The presence of comparable number of solid and dashed lines in the allosteric site indicates that an extensive rearrangement and rewiring of interactions have taken place in this region. In particular, a few pairs connected by red lines ( $|\Delta E_{ij}| > 10$  kcal/mol) indicate that D332-E334, E395-R399 and E305-R354 interactions are more favourable in the unbound state, whereas E334-R399, E395-R309, R309-D306, E305-D357, E304-R411, E352-R354, R354-D357 and K355-E401 become more favorable in the bound state. As we can speculate, many of these pairwise interactions between charged residues would involve strong electrostatic interaction (salt bridges) or hydrogen bonded interactions as we shall show in the next section.

Interestingly, the binding site and allosteric site are connected through the  $\alpha 3$  helix mediated by a salt bridge interaction between E334 and R399. Thus, our energy perturbation network re-establishes the experimental results of Lee and coworkers regarding the significant role of  $\alpha 3$  helix towards “hidden dynamic allostery” in PDZ3 domain<sup>6</sup> and subsequent simulation studies showing the hydrogen bonded interactions between the  $\alpha 3$  helix and  $\beta 2$ - $\beta 3$  loop region that controls this connection<sup>61</sup>. As we have already seen before, both  $\Delta E_i^{protein}$  and  $\Delta E_{ij}$  values show large variation with opposite signs that leads to cancellations resulting in small change in the total energy/enthalpy of the protein. Similarly, a few residues have a large change in  $\Delta E_{ij}$  with  $|\Delta E_i| < 6$  kcal/mol, e.g. R312, R354, K355, R399, E401 and R411 (Table 3.2). This observation again highlights the phenomena of massive internal rearrangement without changing the total interaction energy due to cancellation between pair-wise interactions in reverse directions. The residues with large number of connections as shown in Fig. 3.5 are likely to play the role of intermediate hubs in the network of energetic propagation of allosteric modulation.

### 3. Hidden Electrostatic Basis of Dynamic Allostery

**Table 3.2** Breakdown of perturbation in pair-wise interactions, where  $|\Delta E_{ij}| > 2$  kcal/mol. Residues with  $|\Delta E_{ij}| > 10$  kcal/mol are highlighted in red.

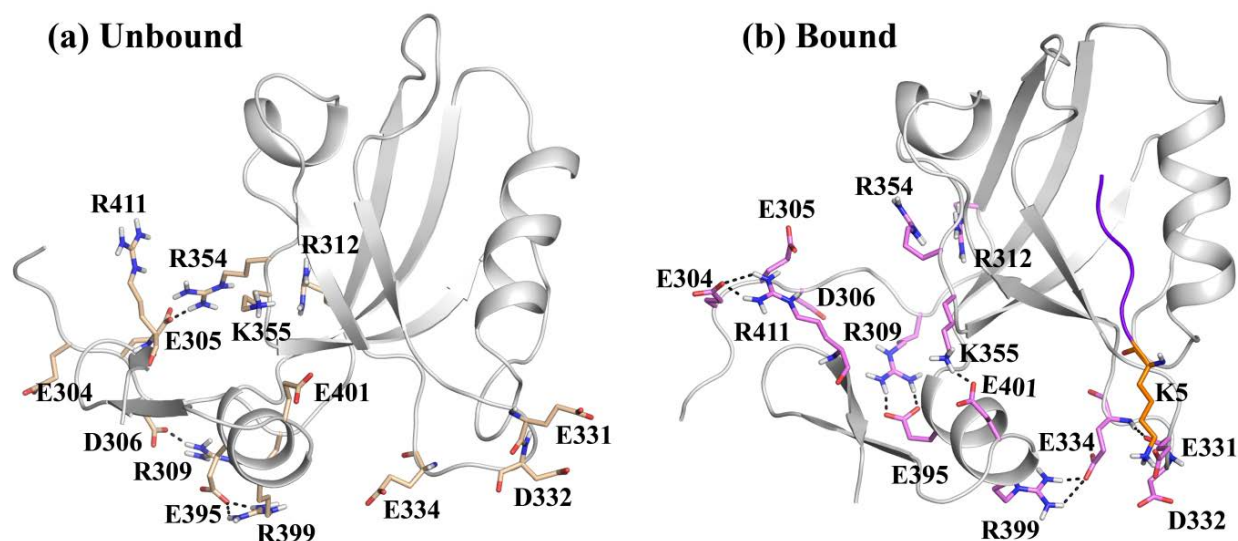
ResID	Residue-pair-wise contribution : $\Delta E_{ij}$ (kcal/mol)
E304	<b>R411 (-10.8)</b>
E305	<b>R354 (57.8), D357 (-11.4)</b> , R312 (8.8), K393 (6.4), E396 (-5.0), E395 (-4.9), R399 (4.7), E334 (-3.4), K355 (3.2), R405 (3.1), E401 (-3.1), E352 (-2.6), E304 (2.3), R368 (2.1), N407 (-2.1)
D306	<b>R309 (-12.1)</b> , R411 (-9.5), E395 (3.7), R354 (2.8)
R309	<b>E395 (-41.3), D306 (-12.1)</b> , R411 (9.9), D357 (-7.3), R405 (6.8), R354 (6.1), K355 (5.3), Q391 (-5.1), E401 (-4.9), R312 (4.9), E310 (4.7), N407 (-4.3), P394 (3.6), K393 (3.6), D348 (-2.9), E396 (-2.7), I307 (-2.6), E304 (-2.4), E352 (-2.4), R399 (2.3)
R312	<b>R354 (13.1)</b> , E305 (8.8), E310 (-8.4), P308 (-5.2), R309 (4.9)
R318	K380 (4.9), N381 (-4.5), T321 (3.9), G383 (3.8)
E331	E334 (8.3), E373 (-3.6), G330 (3.2), E401 (2.9), R399 (-2.9), E396 (2.6), K355 (-2.4), H372 (-2.2), A370 (2.2), Q374 (2.2), G333 (-2.0)
D332	<b>E334 (15.6)</b> , R399 (-8.4), E396 (4.2), D366 (-3.5), R368 (2.9), G333 (-2.7)
E334	<b>D332 (15.6), R399 (-13.3)</b> , E331 (8.3), K393 (4.7), E305 (-3.4)
E352	<b>R354 (-11.1)</b>
R354	<b>E305 (57.8), R312 (13.1), D357 (-11.3), E352 (-11.1)</b> , R309 (6.1), E310 (-6.0), G351 (-5.0), P308 (-4.3), R411 (-4.0), D348 (-3.9), D306 (2.8), R313 (2.1), R405 (-2.1)
K355	<b>E401 (-13.2)</b> , R309 (5.3), E305 (3.2), Y397 (2.8), E304 (-2.4), E331 (-2.4)
D357	<b>E305 (-11.4), R354 (-11.3)</b> , R309 (-7.3), E310 (4.7), P308 (4.3), E395 (3.0), R411 (-2.7)
D366	D332 (-3.5)
R368	E334 (3.0), D332 (2.9), K393 (2.3), E305 (2.1)
H372	E331 (-2.2)
E373	E331 (-3.6)
K380	R318 (4.9)
Q391	R309 (-5.1), K393 (-3.5)
K393	E305 (6.4), E395 (-4.7), E334 (4.7), R309 (3.6), Q391 (-3.5), R368 (2.3), P308 (2.0)
E395	<b>R309 (-41.3), R399 (11.0)</b> , E305 (-4.9), K393 (-4.7), E396 (-4.5), D306 (3.7), E310 (3.3), D357 (3.0), P394 (-2.8), R411 (-2.6), R312 (-2.5), I307 (2.3)
R399	<b>E334 (-13.3), E395 (11.0)</b> , D332 (-8.4), E305 (4.7), E331 (-2.8), E405 (-2.0)
E401	<b>K355 (-13.2)</b> , Y397 (-6.8), R309 (-4.9), E305 (-3.1), E331 (2.9)
R405	R309 (6.8), E305 (3.1)

### 3.3 Results

N407	R309 (-4.3)
R411	E304 (-10.8), R309 (9.9), D306 (-9.5), E310 (-4.1), R354 (-4.0)

#### Population shift of hydrogen bonded network leads to allosteric modulation.

So far, we have unravelled that ligand binding leads to significant perturbation in the intra-protein electrostatic interaction pattern. Let us now investigate the nature of these specific interactions that rearranges and rewires upon ligand binding. Towards this goal, we have shortlisted the residue pairs with  $|\Delta E_{ij}| > 8$  kcal/mol and found that most of these pairs are capable of forming hydrogen bonds (H-bonds) either through side chains or backbones. Fig. 3.6 shows representative snapshots from the unbound and bound trajectories that highlight the possible differences in the H-bonding pattern between these selected residue pairs. Interestingly, these H-bonds are not exclusive in nature, i.e. a certain pair that forms H-bond in unbound state may be present in the bound state as well, but with a different population.



**Figure 3.6** Rearrangement and re-wiring of side chain interaction network between unbound and bound states. These pairs have been chosen based on large  $|\Delta E_{ij}|$  (see Fig. 6 for values). Note that these specific interactions can be transient and not all hydrogen bonds are present at all frames due to inherent dynamical fluctuations (Fig. 6 for population distribution of these interactions). We have shown selected frames that highlight the nature of the hydrogen bonded network leading to the electrostatic coupling between the binding site and allosteric site. Polar



*contacts (e.g. hydrogen bonds or salt bridges) formed between the side chains are shown using dashed lines.*

This aspect of “population shift” of the pair-wise interactions has been further elucidated in Fig. 3.7. Here we show the population distribution of the minimum approach distance between a set of representative residue pairs with large  $|\Delta E_{ij}|$  values. The minimum distance has been computed between all possible pairs of atoms (including hydrogens) between the residues. Interestingly, for most of the cases, we observe that there are at least two peaks where one corresponds to strong interaction at a short distance (around 0.2nm) and broken/weaker interactions at larger distances for both the unbound and bound states. The strong peaks observed around 0.2nm signify the existence of specific polar interactions (e.g. hydrogen bond or salt bridge) between these pairs. For example, Fig. 3.7a shows the interaction between E331-E334, where the peak around 0.2nm signify the H-bonded interaction between the backbone atoms of these residues (Fig. 3.6b for a representative structure). The population of this state is significantly higher in the bound state as compared to unbound state, but it is still present in the unbound state. Similar observations can be made for other pairs as well.

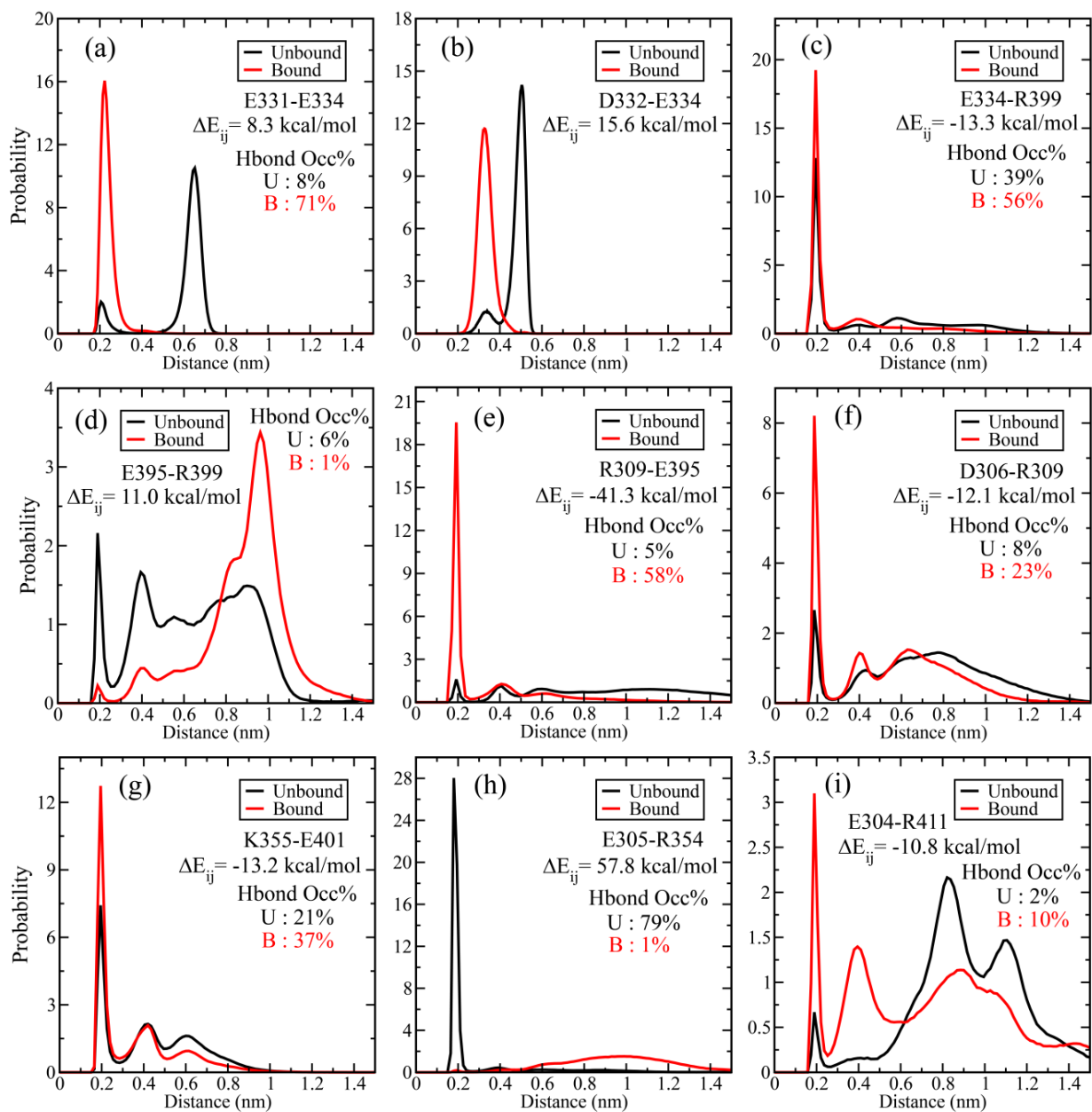
Interestingly, there are examples where this “population shift” in pair-wise interaction is very subtle, e.g. E334-R399 (Fig. 3.7c) and K355-E401 (Fig. 3.7g). Here the peak for the H-bonded species gets stronger (increased population) without changing the overall shape of the distribution function. Yet, these pairs exhibit large  $\Delta E_{ij}$  values (around -13 kcal/mol). There are only two examples, where the distribution function changes in an almost exclusive non-overlapping manner and very large values for  $\Delta E_{ij}$ , e.g. E305-R354 (Fig. 3.7h,  $\Delta E_{ij} = 57.8$  kcal/mol) and R309-E395 (Fig. 3.7e,  $\Delta E_{ij} = -41.3$  kcal/mol). As indicated by the signs of the  $\Delta E_{ij}$  values and distance distribution, the E305-R354 salt bridge interaction exists almost exclusively in the unbound state. On the other hand, the R309-E395 interaction becomes highly favourable instead in the ligand bound state. A structural implication of this observation is that in the unbound state the disordered N terminus region interacts strongly with the  $\alpha 1$ - $\beta 4$  coil region (involving R354), whereas on ligand binding the interactions shift to  $\alpha 3$  helix region (R309-E395) implicated as mediator in dynamic allostery, and also the C terminus coil region (e.g. E304-R411). Fig. 3.6 provides a structural view of these specific H-bonded interactions that undergoes subtle to dramatic population shift upon ligand binding.

### 3.3 Results

---

The above analysis also provides a clear mechanistic picture into the initiation and propagation of the energetic perturbation through the population shift in the H-bonded network much like trapeze artists changing their partners. As shown in Fig. 3.6, the positively charged lysine (K5) of ligand interacts strongly with the E331 and E332 residues in the  $\beta$ 2- $\beta$ 3 loop, which in turn makes the backbone H-bonding between E331-E334 stronger (Fig. 3.7a). This controls the orientation of E334 side-chain to increase the population of the H-bond (salt bridge) formed with R399 in the  $\alpha$ 3 helix (Fig. 3.7c). The role of the interaction between  $\beta$ 2- $\beta$ 3 loop and  $\alpha$ 3 helix has been already shown in an earlier study<sup>61</sup>. Interestingly, there exists a significant population (6%) for the intra-helical *i-i+4* salt bridge formation between the side-chains of E395-R399 in the unbound state (Fig. 3.7d). This interaction gets weaker (1%) in the ligand bound state as E395 starts interacting with R309, and R399 interacting with E334 (Figs. 3.6, 3.7e and 3.7c). Thus, the combination of Figs. 3.6 and 3.7 paint a detailed mechanistic picture of the rearrangement of H-bonded interactions induced by ligand and resultant changes in pair-wise electrostatic interaction energies. Also, we have proven the molecular basis of the strong influence of the extra-domain  $\alpha$ 3 helix on the ligand binding and the conformational preference of sidechain rotamers. These dynamic interactions and energetic coupling between the side chain orientations/interactions provide the key towards the dynamic allosteric modulation in PDZ3 domain.

### 3. Hidden Electrostatic Basis of Dynamic Allostery



**Figure 3.7** Probability distribution of pairwise minimum distance between residue pairs with  $|\Delta E_{ij}| > 10$  kcal/mol for unbound (black line) and bound (red line) states. The  $\Delta E_{ij}$  values have been indicated for all pairs. The strong peak around 0.2nm would signify presence of polar contact (e.g., hydrogen bond or salt bridge). The hydrogen bond occupancy (percentage) based on standard geometric criteria have been marked for all the pairs (except D332-E334).

We must note that for most of the residue pairs dissected in Fig. 3.7, there is a significant tightening of the polar contacts (H-bonds) on ligand binding. Thus the underlying free energy

### 3.3 Results

---

landscape of these pair-wise interactions become narrower, thus providing a direct connection with the entropic view of “dynamic allostery”, where the tightening of the H-bonding interactions leads to reduced conformational entropy for these residues.

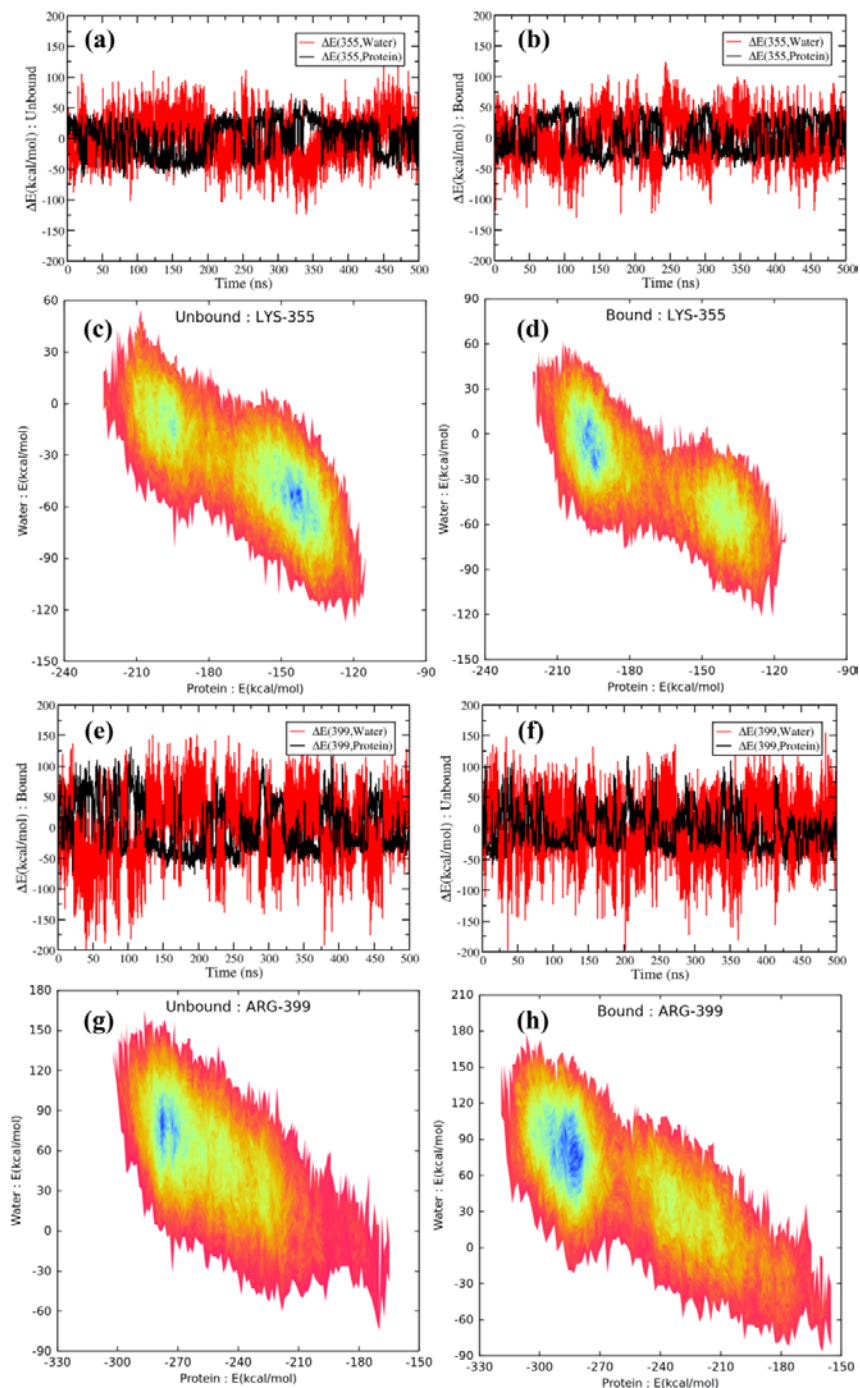
#### Role of solvation energy in allosteric regulation

Prior studies on protein-water interactions highlight the role of bulk and hydration water in the conformational fluctuations of proteins and there exists a strong coupling between solvent and protein dynamics<sup>62,63</sup>. It is debated that water slaves/controls the protein motions or *vice-versa*.<sup>64-68</sup>. So far, we have defined the allostery in terms of perturbation in the change in the interaction energy in protein. In our previous analysis, we have dissected the change in total interaction energy for each residue into three components depending upon the interacting partner: protein, ligand and water. We have observed large values for the residue-wise average electrostatic interaction energy with water i.e.  $E_{i,water}$  which is in the opposite direction to the  $E_{i,protein}$ . We have attributed this neutralizing effect to the dielectric screening. However, it will be interesting to observe if there is any anticorrelation between the energetic contributions of protein and water. Recently, Bagchi and coworkers showed that the self energy (protein-protein interaction) and cross-interaction energy fluctuations (protein-water interaction) are anticorrelated and exhibit power law decay<sup>69</sup>. The study was performed on five different protein-water systems where the whole system was divided into two ensembles: protein and water. Hence, the anticorrelated energy fluctuations exhibit a global picture between water and whole protein. Our aim is to explore if any such anticorrelations exist for the residue-wise interaction energy of the protein and water.

Figure 3.8 shows the residue-wise protein and water interaction energy fluctuations for selective residues which show large change in  $\Delta E_{i,protein}$  and are involved in hydrogen bond rearrangement network. Further, a comparison was done between the bound and unbound states. As shown in Fig. 3.8a, b we observe dramatic anticorrelation between  $E_{i,protein}$  and  $E_{i,water}$  for residues E355 and R399. Both these residues are involved in hydrogen bonding network with difference in hydrogen occupancy > 50%, accompanied by rearrangement in side chain orientation between unbound and bound states. In addition, Fig. 3.8c, d highlight the signatures

### 3. Hidden Electrostatic Basis of Dynamic Allostery

of two states in the energy distributions of these residues. Since, the small scale motions such as side chain fluctuations are coupled to the water molecules in the hydration layer.

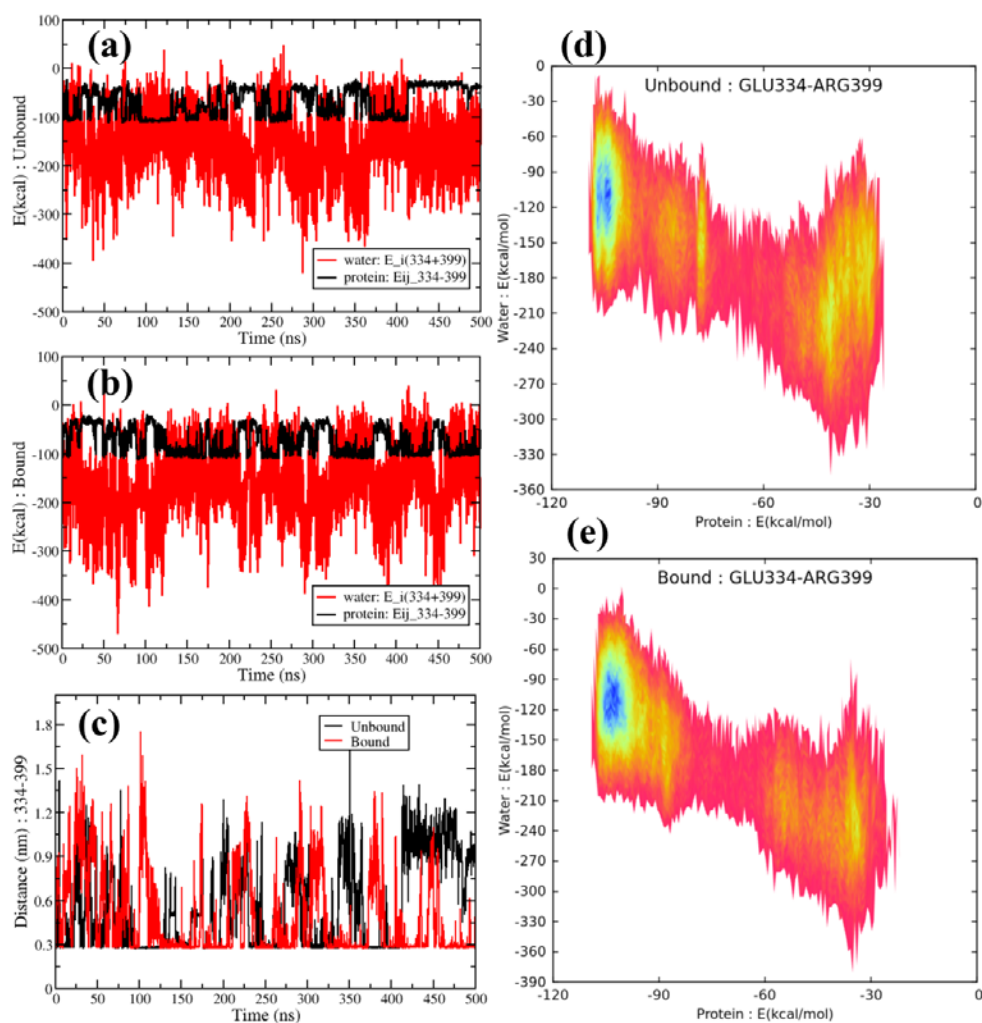


**Figure 3.8** (a, b, e, f) Energetic fluctuations of protein-protein interaction and protein-water interaction with time for residues K355 and R399 for unbound and bound states.  $\Delta E_i(t) = E_i(t) - \langle E_i \rangle$ . Anticorrelation between  $\Delta E_{i,protein}$  and  $\Delta E_{i,water}$  fluctuations can be observed for

### 3.3 Results

residues K355 and R399. (c, d, g, h) 2D probability distribution for energy contributions from protein ( $E_{i,protein}$ ) and water ( $E_{i,water}$ ) for residues K355 and R399. The negative slope of the distribution indicates anticorrelation.

We have also looked into the pairwise interaction energy i.e.  $E_{ij,protein}$  and  $E_{ij,water}$  for residues involved in network (Fig 3.9). For primary analysis, we selected residue pair E334-R399. In the previous chapter, we have shown that, this pair undergoes intermittent breaking and formation of hydrogen bond in bound and unbound states. A comparison between the energetic fluctuations and H-bond dynamics shows a significant anticorrelation.



**Figure 3.9** (a, b) Pairwise interaction energy fluctuation for residue pair E334-R399 with time shows strong anticorrelation. Here, the interaction energy with water is calculated as  $E_{E334-R399,water} = E_{E334,water} + E_{R399,water}$  (c) Minimum distance plot for residue pair E334-

*R339 between bound and unbound states. The fluctuations show intermittent formation and breaking of hydrogen bond with time. On comparison with energy fluctuations, one could see correlation between H-bond dynamics and interaction energy. (d, e) 2D probability distribution for energy contributions from protein and water. The negative slope of the distribution indicates anticorrelation.*

Based on this preliminary analysis, we speculate that this intermittent H-bond formation and breaking could be responsible for the power law decay of interaction energy as observed by Bagchi and co-workers. We observe signatures of allostery in the solvation energy contribution (interaction energy with water) of the total interaction energy, and it is possible that the water dynamics can provide a probe into the dynamics of intermittent intra-protein interactions.

#### 3.4 Discussion

The basic premise of this work is to demonstrate that the “dynamic allostery” phenomena in the PDZ3 domain protein originates from the modulation of underlying energy landscape dictated by ligand binding. Although the prevailing view considers dynamic allostery to be purely entropy driven, we argue and demonstrate for the first time that there exists significant energetic redistribution in terms of the specific electrostatic interactions between the protein residues. First, we demonstrate that the structure/position-based parameters like RMSD, contact map etc. are not able to capture the extent of modulation in specific interactions that undergo upon ligand binding. It is crucial to understand the nature of the rearranging contacts, namely hydrogen bonding, salt bridge, non-polar etc. The non-bonded interaction energy provides the most fundamental and robust yardstick for capturing the subtle changes in the side-chain orientation.

We have shown that the electrostatic interaction energy becomes the key determinant in distinguishing the structural ensemble between the ligand bound and unbound states, and elucidating the allosteric modulation. There exist extensive competing interactions and cancellation effects due to interactions between the protein residues, protein-protein, protein-ligand and protein-water interactions. Such cancellations lead to relatively minor changes in the total enthalpy of the protein, whereas there exists substantial rearrangement or redistribution of the interactions at a local level. We have identified the allosteric network by decomposing the average pair-wise interaction energies for all PDZ3 domain residues and their perturbation upon

### 3.4 Discussion

---

ligand binding. This network clearly connects the ligand binding site with the distal allosteric site through the  $\alpha 3$  helix domain, which has been implicated to play a crucial role in mediating the dynamic allostery in this system.

Our detailed analysis has identified an extended network of hydrogen bonded pairs that control the interaction network. The population distribution of these pair-wise interactions indicate that a “population shift” mechanism prevails, where the pre-existing conformational distribution gets modulated on ligand binding with the tightening of most of the H-bonding interactions upon ligand binding (related to the previous reports of reduction in conformational entropy). Our study, for the first time, identifies the role of specific electrostatic interactions and their population shift towards the allosteric modulation in PDZ3 domain. Moreover, we elucidate the molecular basis of the allosteric modulation in the N- and C-terminal regions through major rearrangement in the interaction pattern, which would provide valuable mechanistic insights into the role of PDZ3 domains in cell signaling. Since the PDZ domains are usually chained together in sequence, it is interesting that the allosteric effects propagate to the termini regions possibly leading to a global response in terms of the spatial arrangement of these domains upon binding with the effector ligands/proteins.

We must note that the substantial rearrangement of electrostatic interactions upon ligand binding and associated structural population shift might alter the pKa values of the titrable residues. That would modulate the population of the protonation states of these residues with further changes in the electrostatic interaction network. Similar effects have been reported earlier for other enzymatic systems<sup>70</sup>, and our future work would explore the intricacies of such altered population of protonation states with the possible consequence of pH-dependent allostery.

On a futuristic note, our approach of energetic perturbation map would be useful in the identification of putative target sites for allosteric drugs. There is a rapidly growing interest in allosteric drugs in contrast to the competitive orthosteric inhibitors due to their selectivity and ability to both enhance and inhibit the activity in a controllable manner. Prior studies have suggested that allosteric drugs work through “anchor” and “driver” atoms, where the anchor atom attaches itself to the binding pocket without causing any conformational change and the driver atom exerts a “push” or “pull” action to modulate the conformational ensemble towards the active or inactive states<sup>71</sup>. A rigorous analysis of the energetic balance of such interactions



(electrostatic, van der Waals' and water mediated) along with their coupling with the inherent energy flow network of the enzyme as elucidated in the current work would be essential for rational design of allosteric drugs.

#### 3.5 Bibliography

1. Nussinov, R. and C.J. Tsai, *Allostery in disease and in drug discovery*. Cell, 2013. 153(2): p. 293-305.
2. Motlagh, H.N., et al., *The ensemble nature of allostery*. Nature, 2014. 508(7496): p. 331-9.
3. Szilagyi, A., R. Nussinov, and P. Csermely, *Allo-network drugs: extension of the allosteric drug concept to protein-protein interaction and signaling networks*. Curr Top Med Chem, 2013. 13(1): p. 64-77.
4. Koshland, D.E., Jr., G. Nemethy, and D. Filmer, *Comparison of experimental binding data and theoretical models in proteins containing subunits*. Biochemistry, 1966. 5(1): p. 365-85.
5. Monod, J., J. Wyman, and J.P. Changeux, *On the Nature of Allosteric Transitions: A Plausible Model*. J Mol Biol, 1965. 12(1): p. 88-118.
6. Petit, C.M., et al., *Hidden dynamic allostery in a PDZ domain*. Proc Natl Acad Sci U S A, 2009. 106(43): p. 18249-54.
7. Popovych, N., et al., *Dynamically driven protein allostery*. Nat Struct Mol Biol, 2006. 13(9): p. 831-838.
8. Rafferty, J.B., et al., *Three-dimensional crystal structures of Escherichia coli met repressor with and without corepressor*. Nature, 1989. 341(6244): p. 705-710.
9. Cooper, A. and D.T.F. Dryden, *Allostery without conformational change*. European Biophysics Journal, 1984. 11(2): p. 103-109.
10. Nussinov, R. and C.J. Tsai, *Allostery without a conformational change? Revisiting the paradigm*. Curr Opin Struct Biol, 2015. 30: p. 17-24.
11. Tsai, C.J. and R. Nussinov, *A unified view of "how allostery works"*. PLoS Comput Biol, 2014. 10(2): p. e1003394.
12. Nussinov, R., C.J. Tsai, and J. Liu, *Principles of allosteric interactions in cell signaling*. J Am Chem Soc, 2014. 136(51): p. 17692-701.
13. Weber, G., *Ligand binding and internal equilibria in proteins*. Biochemistry, 1972. 11(5): p. 864-878.
14. Gunasekaran, K., B. Ma, and R. Nussinov, *Is allostery an intrinsic property of all dynamic proteins?* Proteins: Structure, Function, and Bioinformatics, 2004. 57(3): p. 433-443.

### 3.5 Bibliography

---

15. Feher, V.A., et al., *Computational approaches to mapping allosteric pathways*. *Curr Opin Struct Biol*, 2014. 25: p. 98-103.
16. Volkman, B.F., et al., *Two-state allosteric behavior in a single-domain signaling protein*. *Science*, 2001. 291(5512): p. 2429-33.
17. Lee, H.J. and J.J. Zheng, *PDZ domains and their binding partners: structure, specificity, and modification*. *Cell Commun Signal*, 2010. 8: p. 8.
18. Kim, E. and M. Sheng, *PDZ domain proteins of synapses*. *Nat Rev Neurosci*, 2004. 5(10): p. 771-81.
19. Doyle, D.A., et al., *Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ*. *Cell*, 1996. 85(7): p. 1067-76.
20. Zhang, J., et al., *Crystallographic and nuclear magnetic resonance evaluation of the impact of peptide binding to the second PDZ domain of protein tyrosine phosphatase 1E*. *Biochemistry*, 2010. 49(43): p. 9280-91.
21. Fuentes, E.J., C.J. Der, and A.L. Lee, *Ligand-dependent dynamics and intramolecular signaling in a PDZ domain*. *J Mol Biol*, 2004. 335(4): p. 1105-15.
22. Dhulesia, A., J. Gsponer, and M. Vendruscolo, *Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein*. *J Am Chem Soc*, 2008. 130(28): p. 8931-9.
23. Lu, C., V. Knecht, and G. Stock, *Long-Range Conformational Response of a PDZ Domain to Ligand Binding and Release: A Molecular Dynamics Study*. *J Chem Theory Comput*, 2016. 12(2): p. 870-8.
24. Kong, Y. and M. Karplus, *Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis*. *Proteins*, 2009. 74(1): p. 145-54.
25. Morra, G., A. Genoni, and G. Colombo, *Mechanisms of Differential Allosteric Modulation in Homologous Proteins: Insights from the Analysis of Internal Dynamics and Energetics of PDZ Domains*. *J Chem Theory Comput*, 2014. 10(12): p. 5677-89.
26. Kalescky, R., J. Liu, and P. Tao, *Identifying key residues for protein allostery through rigid residue scan*. *J Phys Chem A*, 2015. 119(9): p. 1689-700.
27. Sharp, K. and J.J. Skinner, *Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling*. *Proteins*, 2006. 65(2): p. 347-61.
28. Fuentes, E.J., et al., *Evaluation of energetic and dynamic coupling networks in a PDZ domain protein*. *J Mol Biol*, 2006. 364(3): p. 337-51.
29. Lockless, S.W. and R. Ranganathan, *Evolutionarily conserved pathways of energetic connectivity in protein families*. *Science*, 1999. 286(5438): p. 295-9.
30. Dima, R.I. and D. Thirumalai, *Determination of network of residues that regulate allostery in protein families using sequence analysis*. *Protein Sci*, 2006. 15(2): p. 258-68.

31. Gerek, Z.N. and S.B. Ozkan, *Change in Allosteric Network Affects Binding Affinities of PDZ Domains: Analysis through Perturbation Response Scanning*. Plos Computational Biology, 2011. 7(10): p. e1002154.
32. Raimondi, F., et al., *A Mixed Protein Structure Network and Elastic Network Model Approach to Predict the Structural Communication in Biomolecular Systems: The PDZ2 Domain from Tyrosine Phosphatase 1E As a Case Study*. Journal of Chemical Theory and Computation, 2013. 9(5): p. 2504-2518.
33. Ota, N. and D.A. Agard, *Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion*. J Mol Biol, 2005. 351(2): p. 345-54.
34. Mino-Galaz, G.A., *Allosteric communication pathways and thermal rectification in PDZ-2 protein: a computational study*. J Phys Chem B, 2015. 119(20): p. 6179-89.
35. Ho, B.K. and D.A. Agard, *Conserved tertiary couplings stabilize elements in the PDZ fold, leading to characteristic patterns of domain conformational flexibility*. Protein Sci, 2010. 19(3): p. 398-411.
36. Ishikura, T., et al., *Energy exchange network of inter-residue interactions within a thermally fluctuating protein molecule: A computational study*. J Comput Chem, 2015. 36(22): p. 1709-18.
37. Chi, C.N., et al., *Reassessing a sparse energetic network within a single protein domain*. Proc Natl Acad Sci U S A, 2008. 105(12): p. 4679-84.
38. Xiang, Y., et al., *Simulating the effect of DNA polymerase mutations on transition-state energetics and fidelity: evaluating amino acid group contribution and allosteric coupling for ionized residues in human pol beta*. Biochemistry, 2006. 45(23): p. 7036-48.
39. Vijayabaskar, M.S. and S. Vishveshwara, *Interaction energy based protein structure networks*. Biophys J, 2010. 99(11): p. 3704-15.
40. Bhattacharyya, M. and S. Vishveshwara, *Probing the Allosteric Mechanism in Pyrrolysyl-tRNA Synthetase Using Energy-Weighted Network Formalism*. Biochemistry, 2011. 50(28): p. 6225-6236.
41. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. 1-2: p. 19-25.
42. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field*. Proteins, 2010. 78(8): p. 1950-8.
43. Horn, H.W., et al., *Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew*. J Chem Phys, 2004. 120(20): p. 9665-78.
44. Georgescu, R.E., E.G. Alexov, and M.R. Gunner, *Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins*. Biophysical Journal, 2002. 83(4): p. 1731-1748.
45. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. J Chem Phys, 2007. 126(1): p. 014101.

### 3.5 Bibliography

---

46. Parrinello, M. and A. Rahman, *Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method*. Journal of Applied Physics, 1981. 52(12): p. 7182-7190.
47. Essmann, U., et al., *A Smooth Particle Mesh Ewald Method*. Journal of Chemical Physics, 1995. 103(19): p. 8577-8593.
48. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations*. Journal of Computational Chemistry, 1997. 18(12): p. 1463-1472.
49. Yuan, C., H. Chen, and D. Kihara, *Effective inter-residue contact definitions for accurate protein fold recognition*. BMC Bioinformatics, 2012. 13(1): p. 292.
50. Di Paola, L. and A. Giuliani, *Protein contact network topology: a natural language for allostery*. Curr Opin Struct Biol, 2015. 31: p. 43-8.
51. Johnson, Q.R., et al., *Mapping allostery through computational glycine scanning and correlation analysis of residue-residue contacts*. Biochemistry, 2015. 54(7): p. 1534-41.
52. van den Bedem, H., et al., *Automated identification of functional dynamic contact networks from X-ray crystallography*. Nat Methods, 2013. 10(9): p. 896-902.
53. Warshel, A., *Electrostatic Basis of Structure-Function Correlation in Proteins*. Accounts of Chemical Research, 1981. 14(9): p. 284-290.
54. Warshel, A. and J. Aqvist, *Electrostatic energy and macromolecular function*. Annu Rev Biophys Chem, 1991. 20: p. 267-98.
55. Nakamura, H., *Roles of electrostatic interaction in proteins*. Q Rev Biophys, 1996. 29(1): p. 1-90.
56. Sharp, K.A. and B. Honig, *Electrostatic interactions in macromolecules: theory and applications*. Annu Rev Biophys Chem, 1990. 19: p. 301-32.
57. Kumar, S. and R. Nussinov, *Close-Range Electrostatic Interactions in Proteins*. ChemBioChem, 2002. 3(7): p. 604-617.
58. Xiang, Y., et al., *Simulating the Effect of DNA Polymerase Mutations on Transition-State Energetics and Fidelity: Evaluating Amino Acid Group Contribution and Allosteric Coupling for Ionized Residues in Human Pol  $\beta$* . Biochemistry, 2006. 45(23): p. 7036-7048.
59. Prakash, P., A. Sayyed-Ahmad, and A.A. Gorfe, *The Role of Conserved Waters in Conformational Transitions of Q61H K-ras*. PLOS Computational Biology, 2012. 8(2): p. e1002394.
60. Buchli, B., et al., *Kinetic response of a photoperbated allosteric protein*. Proc Natl Acad Sci U S A, 2013. 110(29): p. 11725-30.
61. Mostarda, S., D. Gfeller, and F. Rao, *Beyond the binding site: the role of the beta(2)-beta(3) loop and extra-domain structures in PDZ domains*. PLoS Comput Biol, 2012. 8(3): p. e1002429.

62. Mukherjee, S., S. Mondal, and B. Bagchi, *Distinguishing dynamical features of water inside protein hydration layer: Distribution reveals what is hidden behind the average*. The Journal of Chemical Physics, 2017. 147(2): p. 024901.
63. Mondal, S., S. Mukherjee, and B. Bagchi, *Origin of diverse time scales in the protein hydration layer solvation dynamics: A simulation study*. The Journal of Chemical Physics, 2017. 147(15): p. 154901.
64. Frauenfelder, H., P.W. Fenimore, and B.H. McMahon, *Hydration, slaving and protein function*. Biophysical Chemistry, 2002. 98(1): p. 35-48.
65. Caliskan, G., et al., *Protein and solvent dynamics: How strongly are they coupled?* The Journal of Chemical Physics, 2004. 121(4): p. 1978-1983.
66. Fenimore, P.W., et al., *Slaving: Solvent fluctuations dominate protein dynamics and functions*. Proceedings of the National Academy of Sciences, 2002. 99(25): p. 16047.
67. Beece, D., et al., *Solvent viscosity and protein dynamics*. Biochemistry, 1980. 19(23): p. 5147-5157.
68. Bellissent-Funel, M.-C., et al., *Water Determines the Structure and Dynamics of Proteins*. Chemical Reviews, 2016. 116(13): p. 7673-7697.
69. Mukherjee, S., S. Mondal, and B. Bagchi, *Mechanism of Solvent Control of Protein Dynamics*. Physical Review Letters, 2019. 122(5): p. 058101.
70. Alexov, E., *Calculating proton uptake/release and binding free energy taking into account ionization and conformation changes induced by protein-inhibitor association: Application to plasmepsin, cathepsin D and endothiapepsin-pepstatin complexes*. Proteins: Structure, Function, and Bioinformatics, 2004. 56(3): p. 572-584.
71. Nussinov, R. and C.-J. Tsai, *Unraveling structural mechanisms of allosteric drug action*. Trends in Pharmacological Sciences, 2014. 35(5): p. 256-264.



# Chapter 4

## pH-dependent Dynamic Allostery in PDZ3 Domain

### 4.1 Introduction

Allosteric regulation of proteins is a complex phenomenon which may involve pH and concentration driven dynamic conformations, where often electrostatics plays a dominant role.<sup>1-7</sup> pH varies significantly in different diseased states affecting the ligand binding affinity and/or functional activity of various allosteric enzymes. Proteins contain ionizable amino acid residues that are either exposed to the solvent or buried, partially or completely in a folded protein. The physiochemical environment of the residue strongly influences the  $pK_a$  values of the ionizable groups. The  $pK_a$  values are largely affected by charge-charge interactions and charge-dipole interaction, which includes salt-bridges and hydrogen bonds.<sup>8</sup> In addition, the ligand binding process is associated with change in the electrostatic interactions that are responsible for  $pK_a$  shifts of the charged groups/amino acids in protein.<sup>9-12</sup> These events induce ionisation changes through proton uptake/release.<sup>13,14</sup> The resulting charge alteration of the ionizable residues as a response to change in microscopic  $pK_a$  is known to trigger protein conformational changes.<sup>10,15,16</sup> Previous studies highlight the coupling between pH-dependent protonation and associated structural rearrangements such that, in equilibrium, one of the conformational states will differ in  $pK_a$  values for ionizable residues.<sup>17,18</sup> pH-dependent conformational changes and coupled ionisation-conformational equilibrium have been studied using molecular dynamics simulations techniques for Nitrophorin 4 and Staphylococcal nuclease respectively.<sup>18,19</sup>

Majority of the pH-regulated biochemical processes occur within pH range of 5~8. Several studies show that at this intermediate pH range, histidine ( $pK_a=6.5$  in water) plays an important role in regulating two different energetic states of protein in equilibrium.<sup>20-23</sup> Conformational changes coupled to the change in the physiological pH and subsequent effects on the protein functions can be associated with pH-dependent allosteric regulation of proteins.<sup>24</sup> The

## 4.1 Introduction

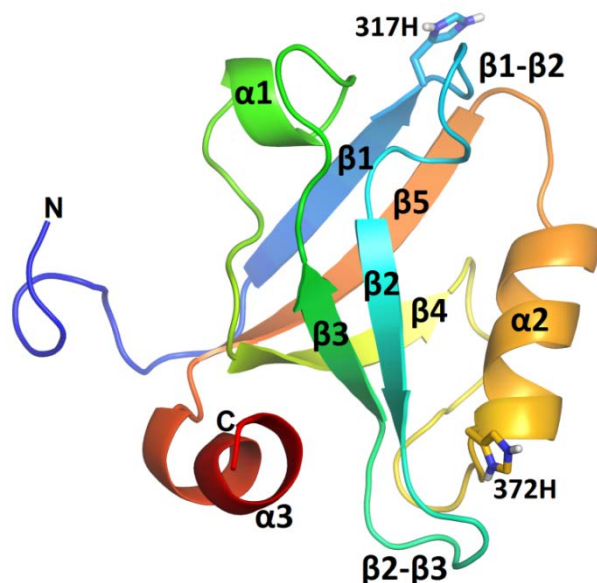
---

Bohr effect in haemoglobin is the classic example of pH-dependent allosteric regulation.<sup>25</sup> The increase in affinity for oxygen binding is coupled to the deprotonation of His146 residue with increasing pH (reduced acidity). This results in the disruption of charge-charge interaction (salt-bridge) between His146 and Asp94, shifting the equilibrium towards R-state.<sup>26,27</sup> Similarly, several proteins (also known as pH sensors) are functionally regulated by a change in the physiological pH through ionisation of pH sensing residues.<sup>28</sup> Computer simulations and experimental studies suggest allosteric regulation of pH-sensitive systems such as ion transport proteins, channels, pumps and enzymes through protonation of histidine residues.<sup>29-33</sup> This posttranslational modification by protons can regulate specificity (e.g. histidine protonation mediated interaction between RhoGEF and phosphoinositides), allosteric effects (protonation induced structural changes) and cooperativity (multiple protonation induced  $pK_a$  shifts and electrostatic coupling).

More recently, allostery has been viewed in terms of “population shift” of conformational sub-states without considerable change in the backbone upon ligand binding.<sup>34,35</sup> Previous studies based on pH dependent allosteric transitions, exhibit shift in the relative population of the states.<sup>36</sup> Narayan and Naganathan show a unique role of pH in population redistribution of transcriptional co-repressor Cnu.<sup>37</sup> Spectroscopic and calorimetric studies reveal that the protonation of His45 switches the conformation of Cnu protein to an alternate native ensemble which is characterised by the disordered helix. The protonation/deprotonation of ionizable residues can cause unfavourable electrostatic frustrations that can drive conformational changes in the equilibrium populations. These changes have a marked effect on thermodynamics and underlying energy landscape of the system.<sup>24</sup>

In the previous chapter, we have highlighted the role of specific electrostatic interactions and their population shift towards the allosteric modulation in PDZ3 domain upon ligand binding.<sup>38</sup> Nussinov and co-workers highlighted that proteins may have multiple pre-existing allosteric pathways and favor one depending on the perturbation events such as ligand binding, mutations, covalent modifications and changes in the cellular physiological conditions.<sup>39</sup> Previous observation suggests pH dependence of the interaction between PDZ domains and their ligand (peptide sequences) through a conserved histidine residue at position 372.<sup>40</sup> PDZ3 domain has two histidine residues at positions 317 and 372 situated at the  $\beta$ 1- $\beta$ 2 loop and  $\alpha$ 2 helix respectively which interacts directly/indirectly with the ligand (Fig. 4.1).





**Figure 4.1** Representative image of PDZ3 domain (PDB: 3I4W). Histidine residues at positions 317 and 372 are highlighted in the protonated state. Peptide ligand binds in the cavity between  $\beta 2$  and  $\alpha 2$  where 372H forms a hydrogen bond with threonine at position -2 of Class-I type peptide ligands.

As shown in Fig. 4.1, both the histidine residues are solvent exposed. PDZ domain is involved in membrane localisation of multi-protein signaling complexes which can occur in different cellular conditions. Change in the cellular environment or ligand binding event can alter the ionisation state of His-317 (317H) or/and His-372 (372H), further modulating the charge distribution and affecting the overall electrostatics of the binding site residues. It is also possible that the protonation in the ligand binding site may introduce unfavourable/favourable electrostatic interactions that will alter the ligand specificity and/or binding affinity. In this chapter, we have investigated the effect of protonation states of histidine (at position 317 or 372) (another kind of perturbation) on the electrostatic interaction network of the protein and associated population shift (if any) in a PDZ3 domain in the unbound state. Moreover, it will be interesting to see if there are signatures of pH-dependent dynamic allostery in PDZ3 domain. Here, we demonstrate for the first time the existence of pH-driven dynamic allostery in PDZ3 domain in terms of modulation in distal conformational fluctuations and changes in the electrostatic interaction network already identified in our previous work. Further, we hypothesise that various perturbation events such as ligand binding, or change in cellular conditions

## 4.2 Methods

---

(protonation) may trigger the same universal response system in PDZ3 domain in terms of modulation in the dynamics and intra-protein interaction network.

## 4.2 Methods

### 4.2.1 System Setup

MD simulations were performed using the crystal structure of PDZ3 domain in the ligand unbound state (PDB ID: 3I4W).<sup>41</sup> The crystal structure is without cloning artefact as it is assumed that its presence may impede certain dynamic features and thus modulate allosteric response network. The termini of the protein were capped to avoid any charge bias. Succinimide residue at the position 332 in the  $\beta$ 2- $\beta$ 3 loop was modified to aspartate residue as per PDZ3 domain sequence. Three different systems were generated based on the protonation state of histidine residue at positions 317 and 372; (1) neutral- both the histidine at 317 and 372 are uncharged, (2) protonation at 317-His only and (3) protonation at 372-His only.

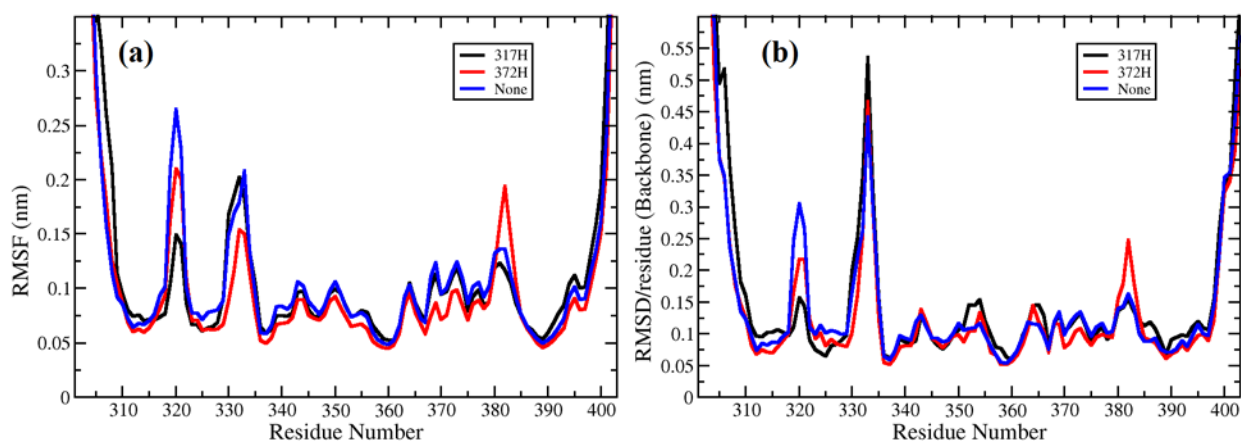
### 4.2.2 Simulation Parameters

We have performed MD simulations of PDZ3 domain in the ligand unbound state (PDB: 3I4W). Simulations were performed in three different states: (1) neutral- both the histidine at 317 and 372 were uncharged, (2) protonation at 317H only and (3) protonation at 372H only. All simulations were performed using Gromacs 5.0.7<sup>42</sup> with AmberSB99-ildn force field<sup>43</sup> and the TIP4P-Ew water model.<sup>44</sup> The systems were neutralised by adding counter ions followed by energy minimisation using steepest descent method. The temperature was maintained at 300K using velocity rescale (modified Berendsen) thermostat<sup>45</sup> and pressure was kept constant at 1 atm using Parinello-Rahman barostat.<sup>46</sup> The simulations were performed using periodic boundary conditions and particle mesh Ewald method<sup>47</sup> was used to treat long range electrostatic interactions. The cut-off distance for short-range electrostatic and van der Waals interactions was set to 1.0 nm. The bonds were constrained with LINCS.<sup>48</sup> Four independent simulations were run for 2 $\mu$ s each with frames saved at every 2ps. The initial 500ns of each trajectory was considered as a relaxation period (equilibration). The individual trajectories post 500ns were concatenated yielding total simulation run of 6 $\mu$ s.

### 4.3 Results

#### Thermodynamic fluctuations as signatures of dynamic allostery

To elucidate the effect of protonation on the dynamics of PDZ domain, we examined the root-mean-square fluctuations (RMSF) about the mean conformation with (317H and 372H protonation) and without protonation states. Previous studies show that PDZ domain exhibits greater flexibility and plasticity in the loop regions ( $\beta$ 1- $\beta$ 2 and  $\beta$ 2- $\beta$ 3) in the unbound state (neutral) which is suppressed upon ligand binding introducing rigidity in the domain.<sup>49-54</sup> We have shown the comparison of fluctuation profile and average structural deviation of each residue for the different protonation states in Fig. 4.2. Surprisingly, the conformational flexibility of the PDZ domain changes upon protonation at selective sites. The  $\beta$ 1- $\beta$ 2 loop shows the least structural deviation and reduced fluctuations upon protonation of 317H.

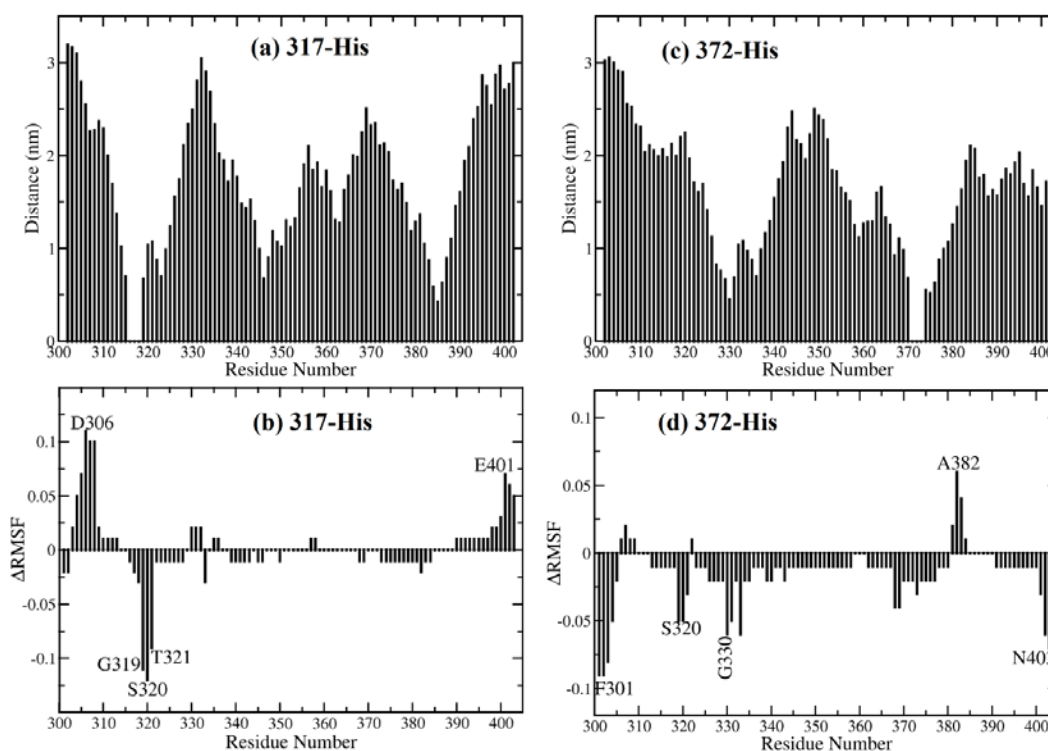


**Figure 4.2** (a) Residue-wise fluctuation profile (RMSF) for PDZ3 domain in different protonation states. It shows the fluctuations for each residue from the average structure. An enhanced flexibility is observed in the  $\beta$ 1- $\beta$ 2 and  $\beta$ 2- $\beta$ 3 loop regions in the neutral state (absence of protonation on histidine). (b) The plot shows the backbone structural deviation (RMSD) for each residue from the initial structure.

This also establishes that protonation selectively enhances the rigidity of the  $\beta$ 1- $\beta$ 2 loop situated next to the protonation site. In contrast, upon protonation of 372H, overall the fluctuations decrease in the loop regions ( $\beta$ 1- $\beta$ 2 and  $\beta$ 2- $\beta$ 3) with the exceptional increase of flexibility towards the C-terminal of  $\alpha$ 2 helix ( $\alpha$ 2- $\beta$ 5 loop). The results tend to be similar to the

### 4.3 Results

fluctuation profile obtained for ligand bound state in chapter 3 (Fig. 3.1b) emphasising decrease in flexibility and plasticity in the  $\beta$ 1- $\beta$ 2 and  $\beta$ 2- $\beta$ 3 loops and ligand binding site of PDZ domain upon ligand binding. Such modulation of conformational flexibility in the binding pocket may alter the ligand binding activity.<sup>55</sup> Furthermore, we have shown in Fig. 4.3, the comparison of average  $C\alpha$  distance of protein residues from the protonation sites (317H and 372H) and net change in fluctuation ( $\Delta$ RMSF) from the neutral state. It is worth noting that the positive charge introduced by the protonation has a pronounced effect on the flexibility of the loop regions ( $\beta$ 1- $\beta$ 2 and  $\alpha$ 2- $\beta$ 5) situated at a distance. The modulation of sidechain dynamics of distal residues upon protonation represents the signatures of pH-dependent dynamic allostery.



**Figure 4.3** (a, c) Comparison between average  $C\alpha$  minimum distance for each residue from the protonation site (317H and 372H), and (b, d) difference in the average fluctuations ( $\Delta$ RMSF) upon protonation from the neutral state.

#### Effect of protonation on the energetics of PDZ domain

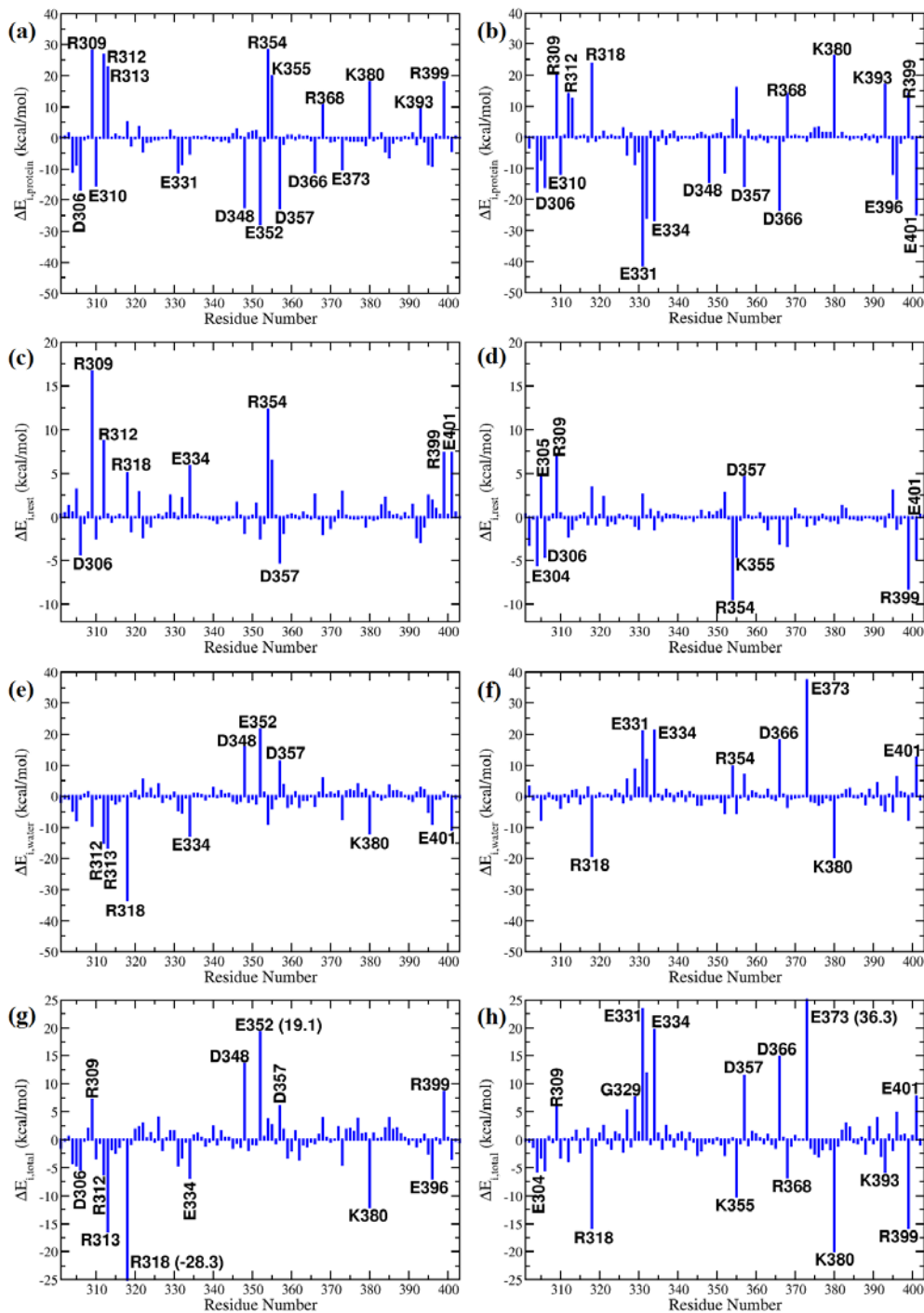
In chapter 3, we have highlighted that the electrostatic interaction energy provides a sensitive yardstick to capture the allosteric effects in PDZ domain. It is expected that the changes in the

#### 4. pH-dependent Dynamic Allostery in PDZ3 Domain

---

charge composition of the protein will directly affect the coulomb (electrostatic) interactions. To understand how does a perturbation other than the ligand binding will affect the underlying energy landscape, we have calculated the change in the average interaction energy of each residue between the protonated state and neutral state. The energy contribution has been evaluated for protein and water separately as  $\Delta E_{i,protein}$  and  $\Delta E_{i,water}$  respectively. Fig 4.4a and 4.4b demonstrate the modulation in the residue-wise interaction energy upon protonation at 317H and 372H. We observe large values of  $\Delta E_{i,protein}$  ( $> \pm 15$  kcal/mol) for residues that are located far from the protonation site, for example, D306, R309, R312, E331, etc.

## 4.3 Results



**Figure 4.4** Residue-wise change in average electrostatic energy between the protonated and neutral state ( $\Delta E_i = \langle E_{i,protonated} \rangle - \langle E_{i,neutral} \rangle$ ) upon protonation at 317H (a, c, e, g) and 372H (b, d, f, h). The change in interaction energy due to protein with and without contribution from the protonated histidine residue are shown separately in figures (a, b) and (c, d) as  $\Delta E_{i,protein}$  and  $\Delta E_{i,rest}$ . Contribution due to water only ( $\Delta E_{i,water}$ ) are shown in figures e and f.

#### 4. pH-dependent Dynamic Allostery in PDZ3 Domain

---

Figs. g and h show the change in total interaction energy, i.e.  $E_{i,total} = (\Delta E_{i,protein} - \Delta E_{i,His}) + \Delta E_{i,water}$ .

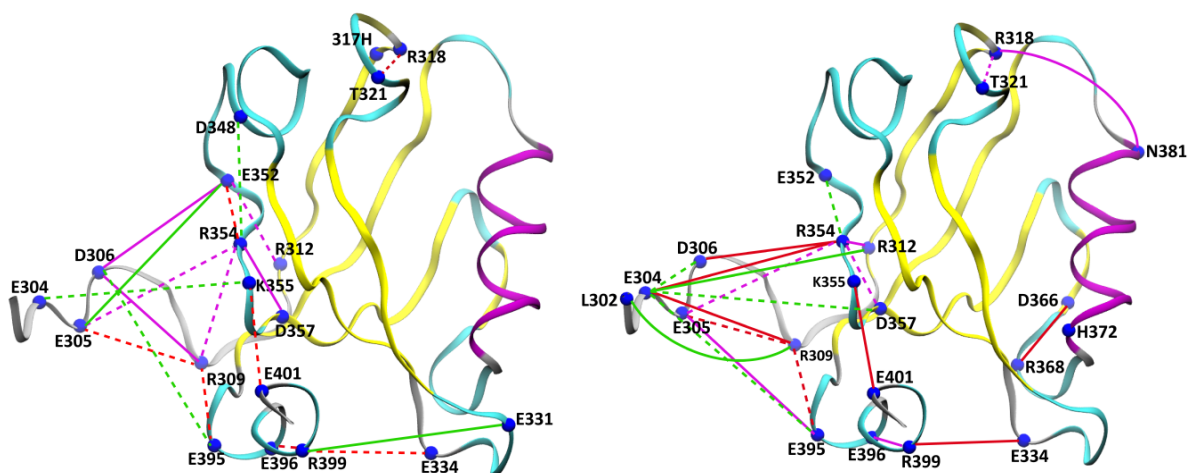
It is obvious that the introduction of a charge on histidine will contribute significantly to the interaction energy due to direct charge-charge interaction. Evidently, we observe that the contribution from the protonated histidine residue,  $\Delta E_{i,His}$  towards  $\Delta E_{i,protein}$  is as large as +/- 10kcal/mol. Hence in order to identify allosteric response generated due to other protein-protein interactions, we calculate  $\Delta E_{i,rest} = \Delta E_{i,protein} - \Delta E_{i,His}$  where  $\Delta E_{i,rest}$  represents the effect of protonation on the  $i^{th}$  residue with respect to the rest of the protein. In Fig. 4.4c and 4.4d, a negative  $\Delta E_{i,rest}$  shows interaction of residue  $i^{th}$  with other residues as favorable in the protonated state and vice versa for the positive  $\Delta E_{i,rest}$ . Interestingly we observe more favorable protein-protein interactions ( $\Delta E_{i,rest} < 0$ ) for residues, for example, E334, R354, D366, R399, E401, etc upon protonation at position 372H as compared to 317H ( $\Delta E_{i,rest} > 0$ ) where they interact strongly in the neutral state. These results signify that protonation at different sites may perturb the intra-protein interaction pattern even far from the protonation site.

Another interesting observation is that these (energetically perturbed) residues form an integral part of the interaction network that undergoes extensive rewiring and rearrangements in terms of allosteric modulation upon ligand binding. This clearly demonstrates the effect of protonation on the intra-protein interaction network and allosteric modulation in terms of electrostatic interaction energy. Figs. 4.4e and 4.4f demonstrate the effect of protonation on the protein-water interactions in terms of change in average electrostatic interaction energy with water (within 2 nm cutoff). We observe that water exerts a dielectric screening effect. The protonation (charge addition) may alter the local solvation environment around residues, for example, E331, D332, E334, D348, D366 and E401 where we observe either favourable ( $\Delta E_{i,water} < 0$ ) or unfavourable ( $\Delta E_{i,water} > 0$ ) interaction with water. We observe large desolvation penalties ( $\Delta E_{i,water}$ ) as compared to the  $\Delta E_{i,rest}$  contribution due to the rest of the protein. Figs. 3g and 3h demonstrate the changes in total interaction energy ( $\Delta E_{i,total}$ ) that indicates that protonation generates both favorable ( $-\Delta E_{i,total}$ ) and unfavorable ( $+\Delta E_{i,total}$ ) energetic perturbations, and there exists local rearrangement and energy redistribution similar to our earlier observation related to ligand binding.

### 4.3 Results

#### pH-driven energetic perturbation network between protonation site and allosteric site

To create a network from the identified energetically perturbed residues, we dissect the electrostatic component of the protein-protein interaction  $\Delta E_{i,Protein}$  into residue pairwise contributions ( $\Delta E_{ij}$ ), where  $\Delta E_{ij} = \langle E_{ij} \rangle_{Protonated} - \langle E_{ij} \rangle_{Neutral}$ . As mentioned previously, any discrete charge addition (histidine protonation) will contribute to the protein-protein interaction energy ( $> 10$  kcal/mol) significantly. Hence, the connections are based on the energetic perturbation between residues other than the protonated histidine and the contribution of  $\Delta E_{i,His}$  is not considered for constructing network. Residues with  $|\Delta E_{i,(Pro-His)}| \geq 5$  kcal/mol and  $|\Delta E_{i,Pro}| \geq 10$  kcal/mol were selected for residue pairwise contribution. In Fig. 4.5, the selective residues are highlighted as blue spheres. The solid lines represent interactions that are more favorable in the protonated state as compared with neutral state and vice versa for the dashed lines.



**Figure 4.5** Network representation of the perturbation in pairwise interaction energies: The network is mapped onto the three-dimensional protein structure (3I4W). The blue spheres represent residues with  $|\Delta E_{i,(Pro-His)}| \geq 5$  kcal/mol and  $|\Delta E_{i,Pro}| \geq 10$  kcal/mol. Connections with negative and positive  $\Delta E_{ij}$  values are indicated with solid and dashed lines, respectively. The connections are colored on the basis of  $\Delta E_{ij}$  values where Red:  $|\Delta E_{i,Pro}| > 4$  kcal/mol, Purple:  $|\Delta E_{i,Pro}| > 2.50$  kcal/mol and Green:  $|\Delta E_{i,Pro}| > 2.0$  kcal/mol. ( Table 4.1 contains  $\Delta E_{ij}$  values for all significant pairs).



#### 4. pH-dependent Dynamic Allostery in PDZ3 Domain

**Table 4.1** Change in the interaction energy per residue for all residues with  $|\Delta E_{i,(Pro-His)}| \geq 5$  kcal/mol and  $|\Delta E_{i,Pro}| \geq 10$  kcal/mol (a) 317H and (b) 372H

(a)	ResID- $\Delta E_{i,pro} / \Delta E_{(i,pro-i,His)}$	$\Delta E_{ij}$ (kcal/mol)
1	E304(-10.8/0.5)	K355(2.5), H317(-11.2)
2	D306(-16.5/-4.3)	R309(-2.6), E352(-2.8), E395(2.4), H317(-12.3)
3	R309(28.0/16.6)	E305(7.3), D306(-2.6), R354(2.6), E395(6.6), H317(11.5)
4	R312(26.7/8.6)	E352(3.0), H317(18.1)
5	E331(-11.1/-0.2)	R399(-2.3), H317(-10.9)
6	E334(-5.0/5.8)	R399(6.0), H317(-10.8)
7	D348(-22.3/-1.8)	R354(2.1), H317(-20.4)
8	E352(-27.6/-2.4)	E305(-2.2), D306(-2.8), R312(3.0), R354(4.4), H317(-25.3)
9	R354(28.1/12.3)	E305(2.7), R309(2.6), D348(2.1), E352(4.4), D357(-2.7), H317(15.8)
10	K355(19.7/6.4)	E304(2.5), E401(4.6), H317(13.3)
11	D357(-22.4/-5.2)	R354(-2.7), H317(-17.1)
12	R399(18.0/7.3)	E331(-2.3), E334(6.0), E396(5.0), H317(10.7)
13	E401(-4.1/7.3)	K355(4.6), H317(-11.4)
(b)	ResID- $\Delta E_{i,pro} / \Delta E_{(i,pro-i,His)}$	$\Delta E_{ij}$ (kcal/mol)
1	E304(-17.3/-5.5)	D306(2.2), R309(-4.6), R312(-2.2), R354(-7.7), D357(2.3), E395(2.1), H372(-11.8)
2	D306(-16.0/-4.5)	E304(2.2), R354(-7.5), H372(-11.5)
3	R309(20.2/6.9)	L302(-2.4), E304(-4.6), E305(6.4), E395(5.6), H372(13.3)
4	R312(13.9/-2.2)	E304(-2.2), R354(-3.0), H372(16.1)
5	R318(23.6/3.3)	T321(3.3), N381(-2.6), H372(20.3)
6	E334(-26.6/-1.4)	R399(-5.0), H372(-25.2)
7	E352(-11.3/2.7)	R354(2.5), H372(-14.0)
8	R354(5.4/-9.4)	E304(-7.7), E305(2.8), D306(-7.5), R312(-3.0), E352(2.5), D357(3.1), H372(14.7)
9	K355(15.7/-4.5)	E401(-5.5), H372(20.3)
10	D357(-15.8/4.6)	E304(2.3), R354(3.1), H372(-20.3)
11	D366(-23.2/-3.0)	R368(-4.4), H372(-20.2)
12	R368(13.9/-3.3)	D366(-4.4), H372(17.2)
13	E395(-11.7/3.0)	E304(2.1), E305(-3.2), R309(5.6), H372(-14.7)
14	E396(-19.8/-1.3)	R399(-3.0), H372(-18.5)
15	R399(13.9/-8.2)	E334(-5.0), E396(-3.0), H372(22.1)
16	E401(-24.8/-4.8)	K355(-5.5), H372(-20.0)

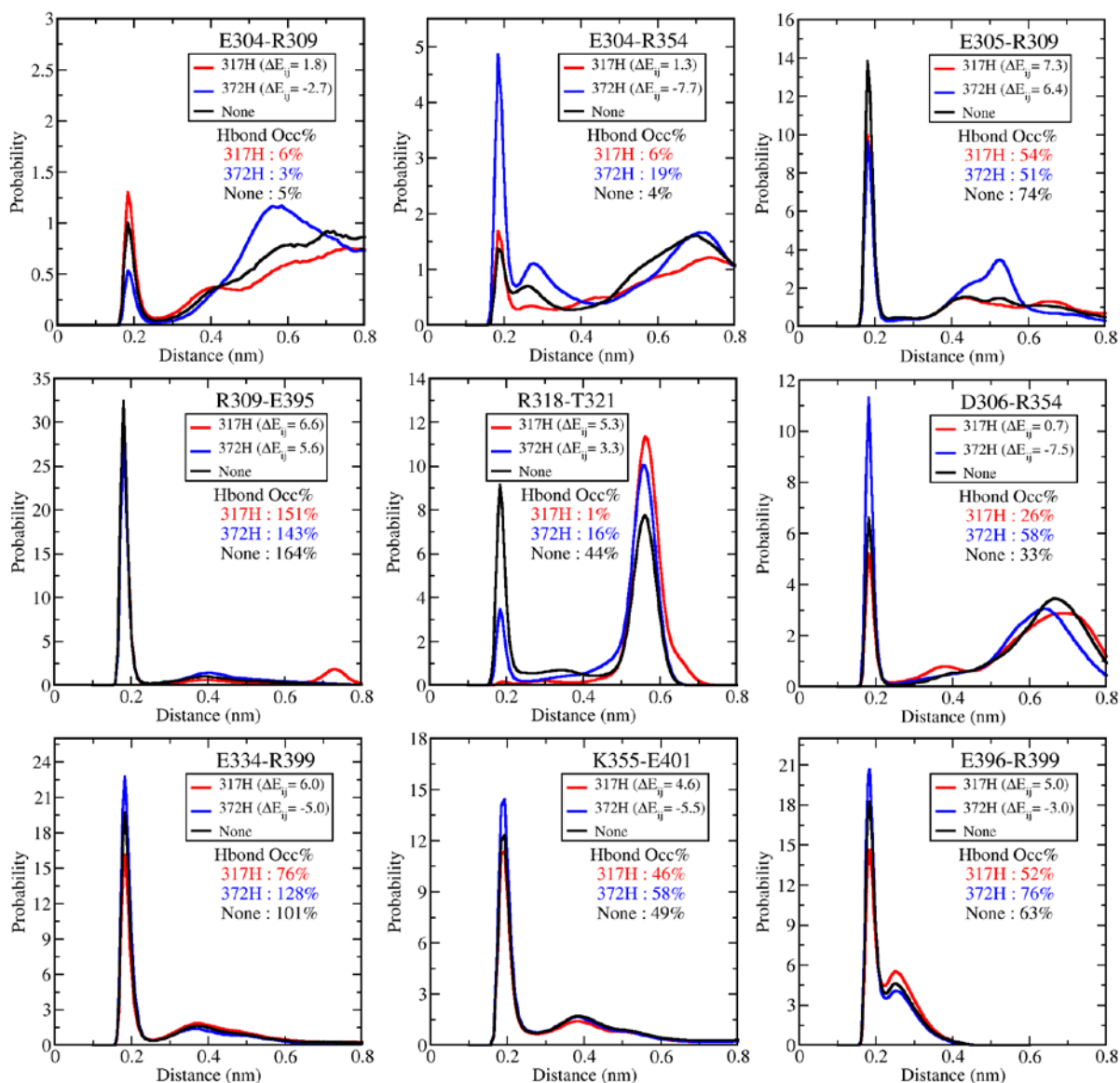
### 4.3 Results

---

A comparison between the two networks gives us significant insights into the change in the interaction pattern upon histidine protonation at position 317 or 372. In Fig. 4.5, the network changes extensively at the distal region (towards N-terminal) from the ligand binding site with a differential pattern of pairwise energetic connections between 317H and 372H protonation states. Interestingly, we observe a significant reduction in  $\Delta E_{ij}$  values (more negative) for certain interaction pairs upon 372H protonation as compared to 317H. (Table S4) In particular, a few pairs such as E334-R399, K355-E401, D306-R354, E304-R354, E304-R309 connected by solid red lines exhibit favorable interaction upon 372H protonation as compared to 317H where the interaction is considerably more favorable in the neutral state (dashed lines). There are examples where the change in pairwise interaction energy ( $\Delta E_{ij}$ ) is positive (favourable in neutral state) in both the networks, for example, E305-R309, R309-E395, R318-T321, E352-R354 and E305-R354. This observation highlights that the introduction of a charge on histidine leads to weaker interactions among certain residue pairs irrespective of the protonation site.

Interestingly few connections, for example, E305-R309, R309-E395, E334-R399, K355-E401, and E305-R354 also appear in energetic perturbation network identified as a consequence of ligand binding in chapter 3. This indicates that the PDZ domain has an inherent and intrinsic network of nonbonded interactions (H-bonded/salt bridges) that undergo energetic redistribution (change in the local protein-protein interaction energy) upon perturbation. Few of these interactions, for example, E334-R399, K355-E401 are also present in the ligand bound state with significantly higher population. Evidently, a negative  $\Delta E_{ij}$  signifies that distal region of the protein exhibits more favorable interactions (stronger connections) and extensive rearrangements in the intra-protein interaction network upon protonation at 372H (more solid lines) as compared to 317H (more dashed lines).

#### 4. pH-dependent Dynamic Allostery in PDZ3 Domain



**Figure 4.6** Probability distribution of pairwise minimum distance between residues with  $|\Delta E_{i,Pro}| > 4$  kcal/mol. The minimum distance has been calculated for all pairs of atoms (including hydrogen) between the residues. The  $\Delta E_{ij}$  values for different protonation states are indicated in the plot. The hydrogen bond occupancy (percentage) based on standard geometric criteria has been reported in Table 4.2.

#### Hydrogen bond population redistribution as a means of dynamic allostery

In PDZ3 domain, we identified an interaction network based on hydrogen-bonded pairs that undergo extensive rearrangement and rewiring on ligand binding in PDZ3 domain. Furthermore, we have shown a population shift mechanism for these pairwise hydrogen bonded interactions to

### 4.3 Results

---

demonstrate the allosteric response in PDZ3 domain. Here, we highlight a similar aspect of population shift of the pairwise interactions upon protonation in Fig.4.6. For most of the cases, we observe that the overall shape of the population distribution does not change with at least two peaks present where one at shorter distance signifies strong/specific polar interactions (mostly hydrogen bond) and another at a larger distance for weaker interactions. However, we observe that a particular interaction (strong/weak) may be present in different protonation states simultaneously, but the population distribution may vary between these states. For example, residue pairs E304-R309, E304-R354, D306-R354, E334-R399, K355-E401 and E396-R399 with negative  $\Delta E_{ij} (\geq 5 \text{ kcal/mol})$  exhibit significantly higher population in the 372H protonation state. On the other hand, these interactions become less favorable (decreased population as compared to the neutral state) in 317H protonation with positive  $\Delta E_{ij}$ . There are examples (E305-R309 and R309-E395) where the distribution decreases upon protonation (at either site) from the neutral state. In Fig. 4.6, the interaction between R318-T321 in  $\beta 1$ - $\beta 2$  loop shows equal population for two probable conformations at 0.2nm (backbone H-bonded) and 0.6nm in neutral state. Upon 317H protonation, the population exclusively shifts to a single conformational ensemble towards weaker interaction between the residue pair.

**Table 4.2** *Hydrogen bond occupancy (percentage) for residues with  $|\Delta E_{i,pro}| > 4 \text{ kcal/mol}$  based on standard geometric criteria ( $\theta \leq 30^\circ, r < 3.5\text{\AA}$ ). The table shows residue pairs that are connected by red lines in the network in Fig. 3. The H-bond occupancy was calculated for residue pairs in neutral and protonated states (317H and 372H).*

Sr. No.	Residue Pair	317H	372H	Neutral
1	E304-R309	6	3	5
2	E304-R354	6	19	4
3	E305-R309	54	51	74
4	R309-E395	151	143	164
5	R318-T321	1	16	44
6	D306-R354	26	58	33
7	E334-R399	76	128	101
8	K355-E401	46	58	49
9	E396-R399	52	76	63

In the previous chapter, we have highlighted the strong interaction between the N-terminus region and the  $\alpha 2$ - $\beta 5$  coil region (involving R354) in the ligand unbound state.

However, in the present work, we observe a further decrease in the average  $E_{ij}$  values of the pairwise interactions from the neutral state (representative for the ligand unbound state) exhibiting stronger interactions upon protonation at 372H. The above observation provides insight onto the effect of protonation on the energy landscape of the protein. It is possible that the underlying free energy landscape becomes deeper and narrower leading to an increase in the barrier between the neutral and protonated state.

### 4.4 Discussion

In summary, our work revealed that protonation of histidine residues in PDZ3 domain as a consequence of physiological pH changes in the cellular environment modulates the conformational dynamics and the energetics distribution in the domain. In this chapter, we highlight that there exist multiple interaction networks in PDZ3 domain, and the allosteric modulation can be observed in terms of energetic propagation through any of these networks depending upon the perturbation factor. First, we demonstrate that the fluctuations/dynamics in the loop region ( $\beta 1$ - $\beta 2$  and  $\beta 2$ - $\beta 3$ ) and  $\alpha 1$ - $\beta 5$  coiled region are suppressed upon protonation at position 317 or 372. The enhanced rigidity in the domain may affect the known two-step binding process in PDZ domain via a sequential (induced fit) mechanism.<sup>56</sup> These decreased fluctuations (increased rigidity) in the domain upon protonation at 372H may result in decreased binding affinity as observed experimentally. Further, it is important to understand how these conformational changes in the loops can play a significant role in peptide binding specificity and allosteric transitions. The increased rigidity in PDZ3 domain also provides signatures of pH-induced dynamic allostery.

Furthermore, we have identified network based on residue pairwise interaction energy that highlights extensive rearrangements of interactions at the distal region of PDZ domain. Our work demonstrates how introduction of positive charge on a histidine residue results in subtle changes in the population distribution of pairwise interactions in the network. We have shown that protonation at different sites have variable effects on the energetics of the protein. Together with dynamics, these changes may alter the ligand binding affinity and allosteric signatures in PDZ3 domain. In this chapter, we propose that there exists an inherent network of interactions in PDZ domain that undergo population redistribution upon perturbation by different factors such as ligand binding, pH-induced or any other. This network may respond differently depending

## 4.5 Bibliography

---

upon the perturbation factor. Also, we must note that these changes modulate the underlying energy landscape and thus affect the functional aspects of the protein.

## 4.5 Bibliography

1. Zhang, Z., S. Witham, and E. Alexov, *On the role of electrostatics on protein-protein interactions*. *Physical biology*, 2011. 8(3): p. 035001-035001.
2. Shindler, J.S. and K.F. Tipton, *A model for the allosteric regulation of pH-sensitive enzymes*. *The Biochemical journal*, 1977. 167(2): p. 479-482.
3. Demaurex, N., *pH Homeostasis of Cellular Organelles*. *Physiology*, 2002. 17(1): p. 1-5.
4. Doi, T., et al., *Extracellular K<sup>+</sup> and Intracellular pH Allosterically Regulate Renal Kir1.1 Channels*. *Journal of Biological Chemistry*, 1996. 271(29): p. 17261-17266.
5. Webb, B.A., et al., *Dysregulated pH: a perfect storm for cancer progression*. *Nature Reviews Cancer*, 2011. 11: p. 671.
6. P Barros, E., et al., *Electrostatic Interactions as Mediators in the Allosteric Activation of Protein Kinase A R1a*. *Biochemistry*, 2017. 56(10): p. 1536-1545.
7. Serber, Z. and J.E. Ferrell, *Tuning Bulk Electrostatics to Regulate Protein Function*. *Cell*, 2007. 128(3): p. 441-444.
8. Pace, C.N., G.R. Grimsley, and J.M. Scholtz, *Protein Ionizable Groups: pK Values and Their Contribution to Protein Stability and Solubility*. *Journal of Biological Chemistry*, 2009. 284(20): p. 13285-13289.
9. Onufriev, A.V. and E. Alexov, *Protonation and pK changes in protein-ligand binding*. *Quarterly reviews of biophysics*, 2013. 46(2): p. 181-209.
10. Czodrowski, P., C.A. Sotriffer, and G. Klebe, *Protonation Changes upon Ligand Binding to Trypsin and Thrombin: Structural Interpretation Based on pKa Calculations and ITC Experiments*. *Journal of Molecular Biology*, 2007. 367(5): p. 1347-1356.
11. Petukh, M., S. Stefl, and E. Alexov, *The role of protonation states in ligand-receptor recognition and binding*. *Current pharmaceutical design*, 2013. 19(23): p. 4182-4190.
12. Misra, V.K. and B. Honig, *On the magnitude of the electrostatic contribution to ligand-DNA interactions*. *Proceedings of the National Academy of Sciences*, 1995. 92(10): p. 4691.
13. Alexov, E., *Calculating proton uptake/release and binding free energy taking into account ionization and conformation changes induced by protein-inhibitor association: Application to plasmepsin, cathepsin D and endothiapepsin-pepstatin complexes*. *Proteins: Structure, Function, and Bioinformatics*, 2004. 56(3): p. 572-584.
14. Mason, A.C. and J.H. Jensen, *Protein-protein binding is often associated with changes in protonation state*. *Proteins: Structure, Function, and Bioinformatics*, 2008. 71(1): p. 81-91.

15. Dai, Z., et al., *pH-induced conformational change of IscU at low pH correlates with protonation/deprotonation of two conserved histidine residues*. *Biochemistry*, 2014. 53(32): p. 5290-5297.
16. Sun, J., et al., *Study of the mechanism of protonated histidine-induced conformational changes in the Zika virus dimeric envelope protein using accelerated molecular dynamic simulations*. *Journal of Molecular Graphics and Modelling*, 2017. 74: p. 203-214.
17. Swails, J.M., et al., *pH-dependent mechanism of nitric oxide release in nitrophorins 2 and 4*. *The journal of physical chemistry. B*, 2009. 113(4): p. 1192-1201.
18. Di Russo, N.V., et al., *pH-Dependent conformational changes in proteins and their effect on experimental pK(a)s: the case of Nitrophorin 4*. *PLoS computational biology*, 2012. 8(11): p. e1002761-e1002761.
19. Liu, J., et al., *A Coupled Ionization-Conformational Equilibrium Is Required To Understand the Properties of Ionizable Residues in the Hydrophobic Interior of Staphylococcal Nuclease*. *Journal of the American Chemical Society*, 2018. 140(5): p. 1639-1648.
20. Perier, A., et al., *Concerted Protonation of Key Histidines Triggers Membrane Interaction of the Diphtheria Toxin T Domain*. *Journal of Biological Chemistry*, 2007. 282(33): p. 24239-24245.
21. Murtaugh, M.L., et al., *A combinatorial histidine scanning library approach to engineer highly pH-dependent protein switches*. *Protein science : a publication of the Protein Society*, 2011. 20(9): p. 1619-1631.
22. Lee, S.A., et al., *Targeting of the FYVE domain to endosomal membranes is regulated by a histidine switch*. *Proceedings of the National Academy of Sciences of the United States of America*, 2005. 102(37): p. 13052-13057.
23. Hanakam, F., et al., *Binding of Hisactophilin I and II to Lipid Membranes Is Controlled by a pH-Dependent Myristoyl-Histidine Switch*. *Biochemistry*, 1996. 35(34): p. 11036-11044.
24. Di Russo, N.V., M.A. Martí, and A.E. Roitberg, *Underlying Thermodynamics of pH-Dependent Allostery*. *The Journal of Physical Chemistry B*, 2014. 118(45): p. 12818-12826.
25. Perutz, M.F., *Stereochemistry of Cooperative Effects in Haemoglobin: Haem-Haem Interaction and the Problem of Allostery*. *Nature*, 1970. 228(5273): p. 726-734.
26. Bettati, S., A. Mozzarelli, and M.F. Perutz, *Allosteric mechanism of haemoglobin: rupture of salt-bridges raises the oxygen affinity of the T-structure*. *Journal of Molecular Biology*, 1998. 281(4): p. 581-585.
27. Cui, Q. and M. Karplus, *Allostery and cooperativity revisited*. *Protein Sci*, 2008. 17(8): p. 1295-307.

## 4.5 Bibliography

---

28. Schönichen, A., et al., *Considering Protonation as a Posttranslational Modification Regulating Protein Structure and Function*. Annual Review of Biophysics, 2013. 42(1): p. 289-314.
29. Baird, F.E., et al., *Evidence for allosteric regulation of pH-sensitive System A (SNAT2) and System N (SNAT5) amino acid transporter activity involving a conserved histidine residue*. Biochemical Journal, 2006. 397(2): p. 369.
30. Olkhova, E., et al., *Multiconformation continuum electrostatics analysis of the NhaA Na<sup>+</sup>/H<sup>+</sup> antiporter of Escherichia coli with functional implications*. Proceedings of the National Academy of Sciences of the United States of America, 2006. 103(8): p. 2629-2634.
31. Hunte, C., et al., *Structure of a Na<sup>+</sup>/H<sup>+</sup> antiporter and insights into mechanism of action and regulation by pH*. Nature, 2005. 435: p. 1197.
32. Srivastava, J., et al., *Structural model and functional significance of pH-dependent talin-actin binding for focal adhesion remodeling*. Proceedings of the National Academy of Sciences, 2008. 105(38): p. 14436.
33. Fledderman, E.L., et al., *Myristate Exposure in the Human Immunodeficiency Virus Type 1 Matrix Protein Is Modulated by pH*. Biochemistry, 2010. 49(44): p. 9551-9562.
34. Nussinov, R. and C.J. Tsai, *Allostery without a conformational change? Revisiting the paradigm*. Curr Opin Struct Biol, 2015. 30: p. 17-24.
35. Tsai, C.J. and R. Nussinov, *A unified view of "how allostery works"*. PLoS Comput Biol, 2014. 10(2): p. e1003394.
36. Ellis, C.R. and J. Shen, *pH-Dependent Population Shift Regulates BACE1 Activity and Inhibition*. Journal of the American Chemical Society, 2015. 137(30): p. 9543-9546.
37. Narayan, A. and A.N. Naganathan, *Switching Protein Conformational Substates by Protonation and Mutation*. The Journal of Physical Chemistry B, 2018. 122(49): p. 11039-11047.
38. Kumawat, A. and S. Chakrabarty, *Hidden electrostatic basis of dynamic allostery in a PDZ domain*. Proceedings of the National Academy of Sciences, 2017. 114(29): p. E5825-E5834.
39. del Sol, A., et al., *The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways*. Structure (London, England : 1993), 2009. 17(8): p. 1042-1050.
40. Chi, C.N., et al., *Two Conserved Residues Govern the Salt and pH Dependencies of the Binding Reaction of a PDZ Domain*. Journal of Biological Chemistry, 2006. 281(48): p. 36811-36818.
41. Cámara-Artigas, A., et al., *Novel conformational aspects of the third PDZ domain of the neuronal post-synaptic density-95 protein revealed from two 1.4Å X-ray structures*. Journal of Structural Biology, 2010. 170(3): p. 565-569.
42. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. 1-2: p. 19-25.



43. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field*. Proteins, 2010. 78(8): p. 1950-1958.
44. Horn, H.W., et al., *Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew*. The Journal of Chemical Physics, 2004. 120(20): p. 9665-9678.
45. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. The Journal of Chemical Physics, 2007. 126(1): p. 014101.
46. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*. Journal of Applied Physics, 1981. 52(12): p. 7182-7190.
47. Essmann, U., et al., *A smooth particle mesh Ewald method*. The Journal of Chemical Physics, 1995. 103(19): p. 8577-8593.
48. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations*. Journal of Computational Chemistry, 1997. 18(12): p. 1463-1472.
49. Mostarda, S., D. Gfeller, and F. Rao, *Beyond the binding site: the role of the beta(2)-beta(3) loop and extra-domain structures in PDZ domains*. PLoS Comput Biol, 2012. 8(3): p. e1002429.
50. Petit, C.M., et al., *Hidden dynamic allostery in a PDZ domain*. Proc Natl Acad Sci U S A, 2009. 106(43): p. 18249-54.
51. Ota, N. and D.A. Agard, *Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion*. J Mol Biol, 2005. 351(2): p. 345-54.
52. Buchli, B., et al., *Kinetic response of a photoperurbed allosteric protein*. Proc Natl Acad Sci U S A, 2013. 110(29): p. 11725-30.
53. Fuentes, E.J., C.J. Der, and A.L. Lee, *Ligand-dependent dynamics and intramolecular signaling in a PDZ domain*. J Mol Biol, 2004. 335(4): p. 1105-15.
54. Kong, Y. and M. Karplus, *Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis*. Proteins, 2009. 74(1): p. 145-54.
55. Münz, M., J. Hein, and P.C. Biggin, *The Role of Flexibility and Conformational Selection in the Binding Promiscuity of PDZ Domains*. PLOS Computational Biology, 2012. 8(11): p. e1002749.
56. Yang, C., et al., *A two-step binding mechanism for the self-binding peptide recognition of target domains*. Molecular BioSystems, 2016. 12(4): p. 1201-1213.



# Chapter 5

## Nucleotide Dependent Conformational Switching in Rho GTPase

### 5.1 Introduction

Biological cells respond to growth factors and hormones through activation of signaling pathways involved in the regulation of cell adhesion activity and other aspects of cell behaviour, including gene expression. Growth factor-induced alterations frequently include changes in cell movement and shape, which play an important role in processes such as embryonic development and wound healing<sup>1-3</sup>. The actin cytoskeleton predominantly governs these aspects of cell activity. Signaling to the actin cytoskeleton through GPCRs involves a large number of signaling cascades that include Rho family of GTPases (RhoA, cdc42 and Rac1), their regulators (GAPs, GEFs, GDIs) and their downstream effector proteins (Rho-kinase, PAK)<sup>4-6</sup>. These proteins regulate a variety of cellular processes such as cell cycle progression, cell migration and gene transcription through molecular recognition based protein-protein interactions.

Rho GTPases are small (~20 kDa), monomeric, Guanine nucleotide binding proteins that belong to the RAS superfamily. These proteins are well known bio-molecular switches whose activity is regulated by binding to the guanine nucleotide<sup>7-10</sup>. The highly conserved switching mechanism involves shuttling of Rho GTPases between their GTP bound (ON) active state and GDP bound (OFF) inactive state<sup>11</sup>. The transition of Rho GTPases between their ON and OFF states is regulated by GEFs (guanine nucleotide exchange factors) and GAPs (GTPases activating proteins)<sup>12</sup>. GEFs catalyse the activation of Rho GTPases by exchanging their bound GDP with GTP, whereas GAPs inactivate them by promoting their intrinsic GTP hydrolysing capacity. More than 80 GAPs have been identified based on human genome analysis. In GTP-bound active state, these proteins can interact with more than 100 effectors and promote subsequent downstream signaling<sup>13,14</sup>. GEFs belong to the Dbl family and contains Dbl homology (DH or RhoGEF) domain. Another class of structurally distinct GEFs belong to

## 5.1 Introduction

---

DOCK family of proteins. Around 70 members of Dbl RhoGEFs and 11 members of DOCK RhoGEFs have been identified<sup>15,16</sup>. The combined number of RhoGEFs and RhoGAPs outnumbers the known 20 Rho GTPase members. This suggests tight regulation of regulatory proteins spatially and temporally, to balance the availability of the active and inactive forms of Rho proteins within a cell. Guanine nucleotide dissociation inhibitors (GDIs) provide additional regulation by sequestering the GDP bound form to maintain an adequate pool of inactive Rho GTPases in the cytosol<sup>17</sup>. Aberrations in the functional Rho GTPases, mainly in their switching action, have been implicated in a number of diseases like cancers, developmental disorders and bacterial infections, where they occur in constitutively active state<sup>18-22</sup>.

Several structural studies on small GTPases have shown that two divergent regions in these proteins, called switch I (SWI, residues 28 to 40 in RhoA) and switch II (SWII, residues 61 to 81) regions play critical role in exerting biological functions of these proteins<sup>23-28</sup>. These two regions not only constitute the nucleotide binding pocket but also engage with their regulators (GEFs and GAPs) and effectors (like kinases)<sup>22,29</sup>. These loops also perch the Mg<sup>2+</sup> binding site, essential for functioning of the protein. In the ON state, the  $\gamma$ -phosphate of the bound GTP forms two hydrogen bonds between the side-chain oxygen of SWI Thr37 and the main-chain oxygen of SWII Gly62 according to RhoA numbering (Thr35 and Gly60 according to H-Ras numbering, respectively). The change in structure upon the loss of the  $\gamma$ -phosphate, on GTP hydrolysis, is termed as the loaded-spring mechanism<sup>11</sup>. This change defines the activation and inactivation mechanism of the GTPases.

Crystal structures of both GTP and GDP bound states of Rho GTPase exhibit multiple SWI conformations, suggesting the plausibility of an ensemble of “micro” ON and OFF states. Interestingly, NMR studies on H-Ras, a member from the RAS superfamily, have revealed that SWI region of the GTP bound state has multiple interconvertible conformations, whereas no such conformational fluctuations are seen in the GDP bound state<sup>30</sup>. Based on these experimental studies, the GTP bound state for Ras GTPase has been classified into two states, known as state 1 (inactive) and state 2 (active)<sup>31,32</sup>. Modest alterations in the coordination between Thr35 and Mg<sup>2+</sup> are shown to alter these states of the Ras GTPase. Furthermore, state 2, which is most prevalently seen in the crystal structures of the wild type Ras, represents the high affinity state for the effectors.

On the other hand, state 1 represents a different (GDP-like) state of the protein with a substantially reduced affinity for effectors, often seen in the crystal structures of mutant of Ras (T35S)<sup>33</sup>. Conformational fluctuations of these two states for Ras GTPases have been extensively studied using normal mode analysis and molecular dynamic simulations<sup>34-42</sup>. The minimum energy pathway between the two states has been analysed with the conjugate peak refinement method<sup>43,44</sup>. However, there is very limited information on the conformational states of Rho GTPases. Experimental studies on Cdc42, a Rho GTPase, have shown the protein to exist only in the state 2 conformation<sup>45</sup>. It is interesting to note that, despite sharing significant structure and sequence homology with H-Ras and having the conserved threonine in the SWI region, Cdc42 predominantly exists in the state 2 conformation.

In this chapter, we explore the conformational states/space of SWI region of Rho GTPases and demonstrate the molecular nature of interactions that govern the nucleotide dependent conformational selections of the Switch I loop region. To address this problem, we have carried out extensive molecular dynamics (MD) simulations on RhoA, a *bonafide* member of Rho Family of GTPases. Availability of several crystal structures of RhoA in different forms made this protein an amiable model system to investigate the conformational states. To further understand the conformational features, free energy calculations were performed. These studies indicate plausibility of existence of state 1 like conformation of GTP bound Rho GTPases. A comparative analysis of the structural signatures of SWI conformations, corresponding to the free energy minima, provides a qualitative explanation for the sparse occurrence of state 1 conformation in Rho GTPases. Finally, analysis of conformation of SWI residues in the GTP bound state reveals substantial solvent exposure of the hydrophobic residues, where the unfavourable solvation energy gets overcompensated by the favourable electrostatic interaction with the nucleotide (GTP). Perhaps, this unique conformational state with hydrophobic exposure has an important role in effector recognition.

## 5.2 Methods

---

### 5.2 Methods

#### 5.2.1 System Setup

The crystal structures of RhoA for GDP bound and GTP bound states were selected as input structures for MD simulation (PDB: 1FTN and 1A2B, respectively). The wild type structure for GTP bound form was obtained by back mutating the V14G in the crystal structure (PDB: 1A2B) using Coot program<sup>46</sup>. In case of GTP bound form, the GTP analogue was converted to GTP. The unbound form of the RhoA GTPase was simulated too. The nucleotide freeform for RhoA was obtained from the complex with GEF of Dbl family (PDB: 1LB1). All the systems preserved the Mg<sup>2+</sup> ion together with the coordinating water molecules.

#### 5.2.2 Simulation Parameters

MD simulations were performed using GROMACS 4.6.5 package<sup>47</sup>. The Charmm27 force field with CMAP corrections<sup>48</sup> was used for the protein, and the parameters for ligands namely GDP and GTP were obtained from SwissParam<sup>49</sup>, a web service that provides topologies and parameters for small organic molecules, compatible with the Charmm all atoms force field, for use with the GROMACS software. All the structures were inserted into cubic box with explicit solvent described by the TIP3P water model<sup>50</sup> and simulated with periodic boundary conditions. The box sizes were set to ensure a distance of at least 1 nm between the protein and the box boundaries. This results in a 7.78nm wide box with ~14700 water molecules for the nucleotide freeform, 7.72 nm wide box with ~14600 water molecules for the GDP bound form and 7.8 nm wide box with ~14700 water molecules for the GTP bound form. The systems were found to be negatively charged, thus in order to neutralise the simulation systems, 7 Na<sup>+</sup> ions in the freeform system, 8 Na<sup>+</sup> ions in the GDP bound system and 9 Na<sup>+</sup> ions in GTP bound system were added. The solvated proteins were energy minimised using the steepest descent algorithm. All the systems were equilibrated for 300ns in NPT ensemble using a modified Berendsen thermostat<sup>51</sup> at 300K and Parrinello-Rahman barostat<sup>52</sup> at 1 bar. Long-range electrostatic interactions were calculated with the particle mesh Ewald (PME) summation method<sup>53</sup> with a grid spacing of 0.16 nm and fourth-order cubic interpolation. For short-range electrostatics and van der Waals, a cut-off distance of 1 nm was used. All covalent bonds were constrained using the LINCS algorithm<sup>54</sup>. The distance between the Mg<sup>2+</sup> ion and the hydroxyl oxygen in Thr17

of RhoA was restrained to preserve the interactions between  $Mg^{2+}$  ion and its coordinating atoms. The integration time step was set to 2fs. Two different types of simulations were performed, atomistic MD simulation and well-tempered metadynamics simulation. The systems were equilibrated for 150ns. The atomistic MD simulations were further performed in two replicas for 300ns with each frame being saved at every 2ps. Only one replica (300ns) was used for the analysis. The metadynamics simulations were carried out for 1 $\mu$ s.

### 5.2.3 Analysis

#### Distance based RMSD parameter

Distance based RMSD (DRMSD) measures the structural deviation with respect to a reference structure based on the distances between all the pairs of atoms. This approach alleviates the need of aligning to a reference structure as required in the usual position based RMSD. The DRMSD ( $X_t$ ) between any conformation at time  $t$  and the reference structure ( $X_{ref}$ ) can be measured as:

$$d(X^t, X^{ref}) = \frac{1}{N(N-1)} \sum_{i \neq j} \left[ d(x_i^t, x_j^t) - d(x_i^{ref}, x_j^{ref}) \right]^2 \quad (5.1)$$

where  $N$  is the number of atoms and  $d(x_i, x_j)$  represents the distance between atoms  $i$  and  $j$ . We used the set of all  $C_\alpha$  atoms of SWI loop for DRMSD calculation. Two reference structures (crystal structures of GDP and GTP-bound forms) were used to characterise the structural similarity/deviation with respect to the active and inactive states.

#### Unsupervised clustering

We have performed unsupervised clustering using the Weka software to cluster the conformations from the trajectories<sup>55</sup>. The loop and coiled regions in the protein are best described by the dihedral angles of the amino-acid residues. The K-means clustering method was carried out using the feature set consisting of the  $\phi$ ,  $\psi$  and  $\chi$  dihedral angles of a subset of the evolutionary conserved residues of the SWI loop that undergo large changes in the dihedral angles between the active and inactive states. The switch region contains ~13 residues. Table 5.1 shows the list of all residues in the switch-I region and associated dihedral values in the GDP-bound inactive and GTP-bound active state. The residues for machine learning were selected on

## 5.2 Methods

---

the basis of the difference in the torsional angles from the GDP and GTP bound equilibrated structure ( $\Delta X = X_{\text{GDP}} - X_{\text{GTP}}$ ).

**Table 5.1** Selection criteria for residues used in unsupervised clustering. The angles included for the analysis include  $\phi$ ,  $\psi$ ,  $\chi_1$ ,  $\chi_2$ . All those residues with major change in the orientation of the backbone/side chain (highlighted in red) between the active and inactive state were selected for *k*-means clustering.

Switch I	$\phi_{\text{GDP}}$	$\phi_{\text{GTP}}$	$\Delta\phi$	$\psi_{\text{GDP}}$	$\psi_{\text{GTP}}$	$\Delta\psi$	$\Delta\chi_1$	$\Delta\chi_2$
Asp28	65.5	67.4	-1.9	38.7	40.0	-1.3		
Gln29	-161.0	-128.3	-32.6	153.4	141.9	11.5		
<b>Phe30</b>	-78.7	-108.7	30.1	136.4	125.0	11.4	2.6	<b>-165.3</b>
Pro31	-102.6	-53.2	-49.4	107.7	148.6	-40.9		
Glu32	-66.0	-74.0	8.0	-30.5	-36.4	5.9		
Val33	-138.9	-152.5	13.6	157.0	161.7	-4.8		
<b>Tyr34</b>	68.3	-74.0	<b>142.3</b>	41.1	124.0	-82.9	<b>-227.2</b>	48.2
Val35	-70.7	-110.7	40.0	119.9	128.9	-9.0		
<b>Pro36</b>	-64.6	-53.5	-11.1	162.6	122.9	39.6	<b>78.2</b>	3.2
<b>Thr37</b>	-58.4	-73.9	15.5	-47.9	128.2	<b>-176.2</b>	-0.8	
<b>Val38</b>	-147.1	-86.9	-60.1	170.5	-54.8	<b>225.3</b>	<b>105.4</b>	
<b>Phe39</b>	-127.3	-144.6	17.3	134.4	150.3	-16.0	<b>-206.7</b>	-11.2
Glu40	-156.9	-139.3	-17.6	155.6	146.5	9.0		

Thus, the total number of features (dihedral angles) used to describe the switch-I conformation were 22 (Table 5.2). The total number of output cluster was limited to four.



**Table 5.2** List of residues used for k-means clustering. Selection of residues is based on the difference in the dihedral angles in the equilibrated GDP bound and GTP bound form as highlighted in Table 5.1.

Sr. No	Switch I Residues	Dihedral Angles
1	PHE30	$\varphi, \psi, \chi_1, \chi_2$
2	TYR34	$\varphi, \psi, \chi_1, \chi_2$
3	PRO36	$\varphi, \psi, \chi_1, \chi_2$
4	THR37	$\varphi, \psi, \chi_1$
5	VAL38	$\varphi, \psi, \chi_1$
6	PHE39	$\varphi, \psi, \chi_1, \chi_2$

### Metadynamics simulations

We have also carried out well-tempered Metadynamics<sup>56</sup> to accelerate the conformational sampling of the Switch I region and to obtain the free energy landscape corresponding to the different nucleotide-bound states. We have used the Plumed code (ver. 2.0.2)<sup>57</sup> to incorporate the Metadynamics functionalities in Gromacs. We have used a “dihedral similarity” ( $S$ ) parameter based on the backbone dihedral angles ( $\varphi/\psi$ ) as the collective variable to distinguish the various conformational states of the Switch I region as defined below.

$$S^{ref} = \frac{1}{2} \sum_i \left[ 1 + \cos(\varphi_i - \varphi_i^{ref}) \right] \quad (5.2)$$

where the sum runs over the  $\varphi$  and  $\psi$  dihedral angles for all the residues of the Switch I region. The reference values ( $\varphi_i^{ref}$ ) have been taken from the respective angles in the crystal structure. The width of the Gaussian ( $\sigma$ ) has been set to 2.3 degrees for both the CVs with the initial height of Gaussians  $W=0.1$  kcal/mol added every 2ps.

### Interaction energy calculation

The interaction energy for different components for the system, i.e. protein, water, Switch I region and  $Mg^{2+}$  ion was calculated using gromacs inbuilt tool `g_energy`. The total interaction energy is calculated as the sum of the electrostatic columbic interaction energy and Van der

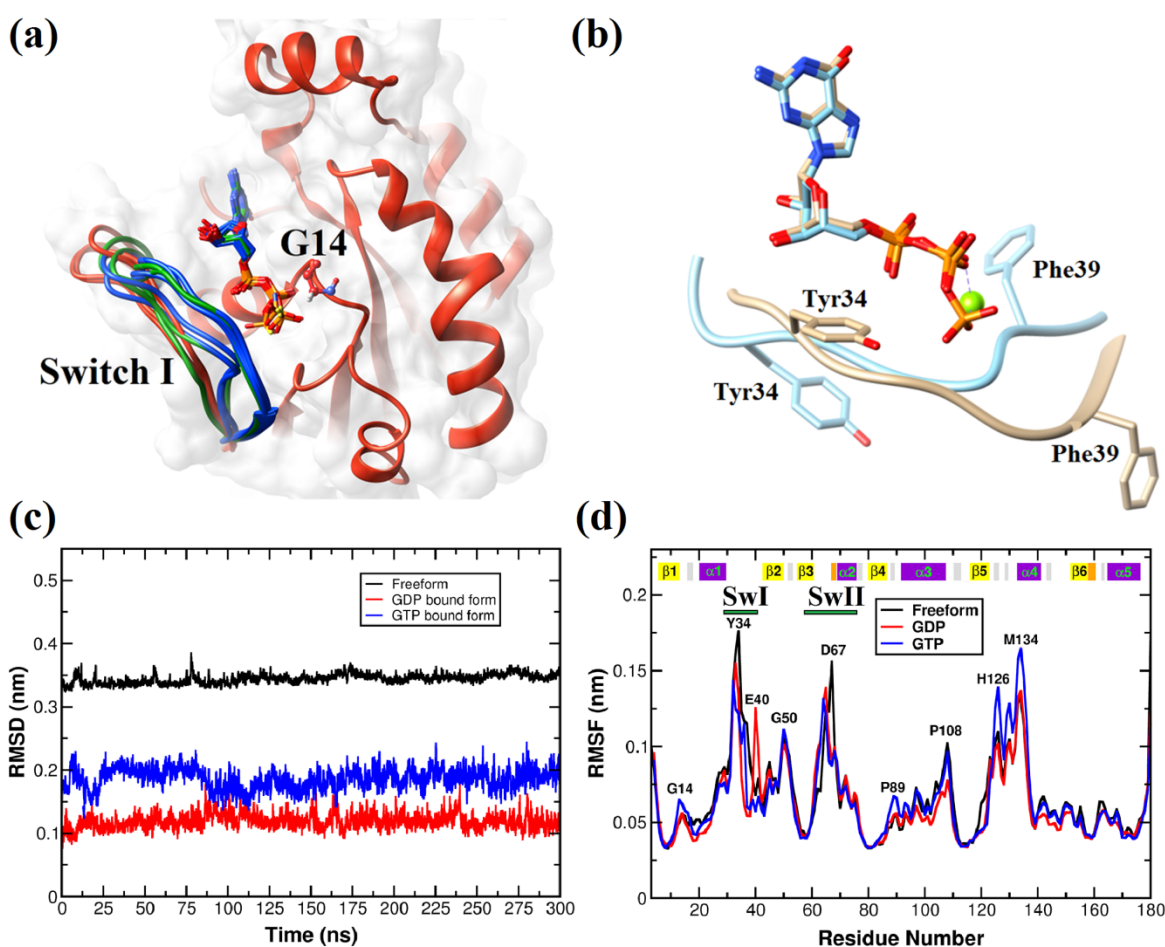
### 5.3 Results

Waals interaction energy terms. The cutoff radius was set to 1nm. All the atoms within the cutoff were included for the energy calculation.

### 5.3 Results

#### MD simulations of GDP, GTP and nucleotide free state of RhoA

To address the conformational heterogeneity in the Rho GTPases, we have superimposed the available crystal structures of RhoA in the nucleotide free form, the GDP bound and the GTP bound form (Fig. 5.1a).



**Figure 5.1** (a) Superimposition of the Switch I loop using the available crystal structures of RhoA. The conformational preference of the loop region has been indicated by the following colours: nucleotide free (red), GTP bound (blue), and GDP bound (green). (b) open and closed conformation in GDP and GTP bound form exhibited by change in the orientation of Tyr34 and

*Phe39 (c) Root mean square deviation (RMSD) plot shows that the systems are well equilibrated and there are no major secondary structural changes in the GTPase. (b) In root mean square fluctuations (RMSF) plot, the protein shows enhanced fluctuations in the SwI and SwII region in the freeform state. In the GTP bound form, increased fluctuations are observed in the insert helix towards the C-terminal.*

Our analysis shows that the high resolution X-ray crystal structures of Rho GTPases in GDP bound and GTP bound form show overall close structural similarity. However, switch I and switch II regions seem to exhibit the largest deviation among crystal structures with particularly large variation between the side chain orientations of Tyr34 and Phe39 (Fig. 5.1b). Although structurally diverse, the SWI and SWII regions exhibit consensus sequence which might play an important role in effector binding. To further explore the nucleotide dependent conformational signatures in Rho GTPases and possible existence of any intermediate state, we performed 1 $\mu$ s of atomistic MD simulations on RhoA in its nucleotide free, GDP bound, GTP bound and the GTP bound G14V mutated form.

In the following sections, we have reported extensive structural and thermodynamic characterisation of this conformational ensemble. We have calculated the RMSD of the backbone atoms with respect to the crystal structures to ensure convergence and stability of the protein (Fig. 5.1c). Although the timescales of MD simulation limit to track the complete transition of the protein between the active and inactive states, shifting of Tyr34 in open and closed conformations, which Rho GTPases exhibit in their GDP and GTP bound state, were observed. This aspect will be further discussed in greater detail in subsequent sections.

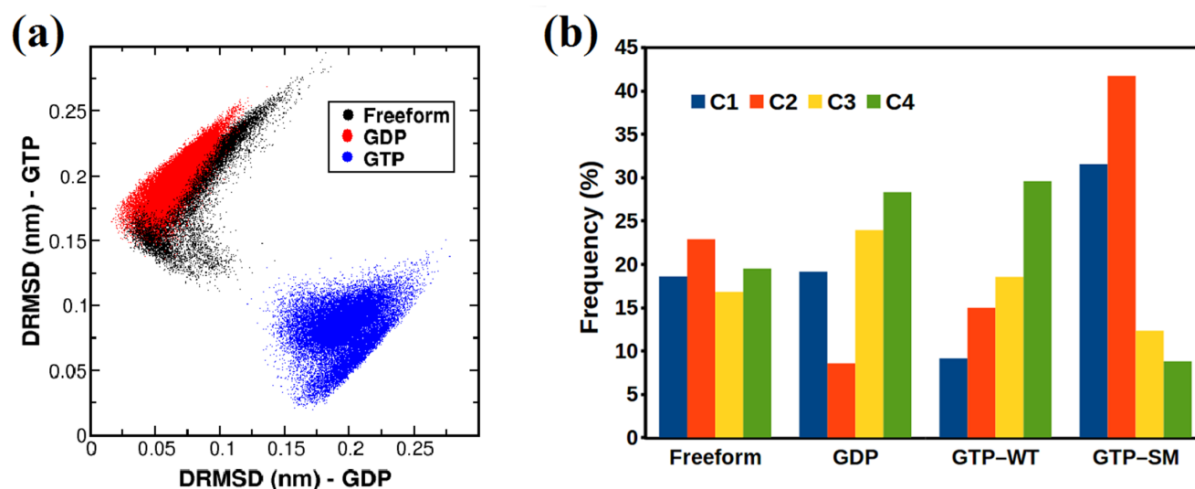
A long-held view is that the flexibility of the switch regions plays an important part in the effector binding in its active state. The RMSF profile showing residue-wise fluctuations exhibits larger fluctuations in the SWI and SWII regions as compared to the other regions of the protein (Fig. 5.1d). However, both the loop regions show significantly lower fluctuations upon nucleotide binding as compared to the nucleotide free state, which signifies the role of these regions in nucleotide binding or effector recognition. Prior simulation studies have shown that mutations in the p-loop or switch regions can significantly alter the degree of flexibility of these regions<sup>58-61</sup>. A point mutation (P29V) locks the RAC1 in the GTP bound activated state, accompanied by the enhanced flexibility in the Switch I region and rigidity in the switch II

### 5.3 Results

region<sup>58</sup>. However, this work would retain the focus on the conformational heterogeneity in the SWI loop, since our objective is to investigate the presence of conformational substates like “state 1” and “state 2” as identified earlier in the H-Ras proteins.

#### Identification of conformational states of SWI region in Rho GTPases

To assess the conformational spread of SWI region from the GTP bound and GDP bound states, distance based RMSD (DRMSD) has been used as a metric. Due to the inherent degeneracy of RMSD-like quantities, a single reference structure is not enough to characterise the structures when the deviation between the reference structure and structures under query are large. Thus, we have used two different references structures (crystal structures for GDP-bound inactive state and GTP-bound active state) to achieve better resolution in classifying the MD trajectories.



**Figure 5.2** (a) Distance based RMSD values (DRMSD) for the trajectories: free form (black), GDP bound (red) and GTP bound (blue). The coordinates are obtained as the DRMSD values of each structure calculated with respect to both the GTP bound structure (Y axis) and the GDP bound structure (X axis) in order to demonstrate the structural distribution and similarities. (b) Population Analysis using K-means Clustering.

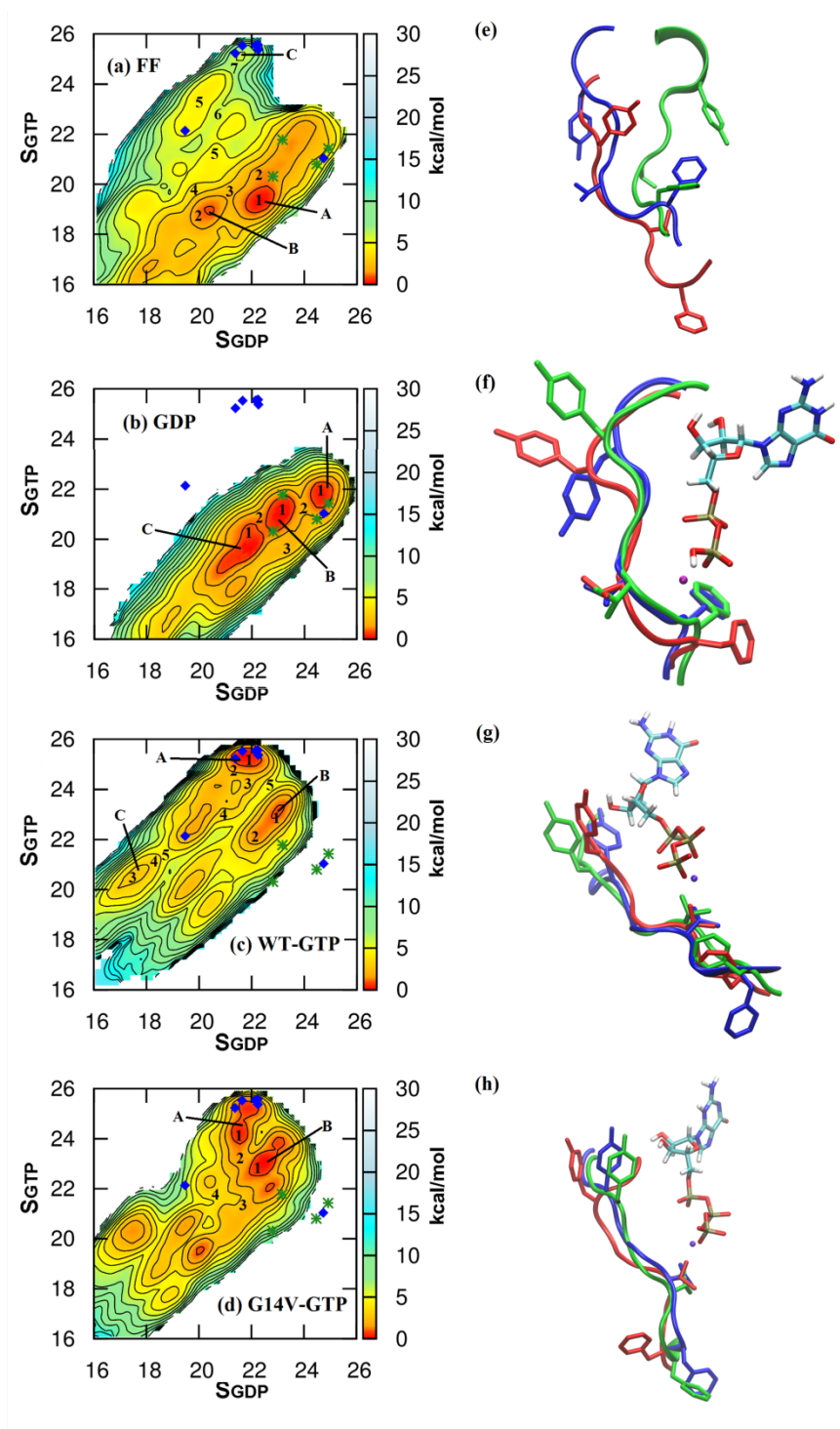
Figure 5.2a demonstrates the spread of conformational states sampled in the MD simulations of the respective systems. Lower X values (DRMSD w.r.t GDP-bound inactive state) and Y values (DRMSD w.r.t GTP-bound active state) indicate higher similarity to GDP and GTP bound structures, respectively. The distribution clearly indicates that the SWI region in the

nucleotide freeform and GDP bound inactive state has very similar conformations, whereas the GTP bound active state has very unique structures corresponding to an isolated domain in this conformational landscape. Interestingly, there is a subtle tendency in the freeform to approach the GTP bound states, which will be further demonstrated through free energy calculations in a later section. Although DRMSD could quantify the similarity/dissimilarity with reference structures, it provides a coarsened global picture of the structural differences. In order to achieve a more microscopic conformational clustering, such as state 1 and state 2 of the 'ON' form (GTP bound form), of the observed MD trajectories an unsupervised machine learning approach was used. We have employed the k-means clustering method<sup>62,63</sup> which has been used earlier to group similar structures of the conformational space in different systems<sup>64-66</sup> (See Methods section for details). A coarse representation of different clusters (Fig. 5.2b) reveals that the GTP bound conformational states of the SWI loop region form two distinct clusters (C3 and C4), whereas the conformational states of nucleotide freeform and GDP bound states show considerable similarity, as observed in case of DRMSD analysis.

To further investigate the effect of mutation on the conformational states of SWI, the procedure was repeated for the simulation results of single mutant (G14V) RhoA-GTP, which is observed in myriad type of cancers<sup>67,68</sup>. Equivalent mutants in H-Ras have shown SWI region of this protein to predominantly adopt the state 1 (inactive GDP bound) like conformation<sup>69</sup>. Hence the objective here was to test whether RhoA has similar tendencies. The results obtained from k-means clustering with the feature set from single mutant RhoA-GTP showed clustering pattern of forming two distinct clusters, as observed for the Wild Type RhoA-GTP structures. However, the occurrence of one of the state (C2) seems more dominant. This analysis clearly demonstrates that GTP bound form of RhoA is capable of adopting two preferential states that might be analogous to the state 1 and state 2 of the GTP bound H-Ras. On the contrary, nucleotide free form and GDP bound form of RhoA do not exhibit preferential conformational states. It is interesting to note that the demarcation of the preferences for the states is more illustrious for the mutant, similar to what has been observed for H-Ras. In case of H-Ras, this mutation drives the GTP bound protein towards the inactive, state 1 conformation. However, for RhoA, this mutation impedes the protein from binding to nucleotides and effectors<sup>70</sup>.

### Conformational free energy landscape of Switch I region

In the case of H-Ras, several MD simulations have been performed to delineate state 1 and state 2 conformations of the protein. In most of these studies, the distance between Thr35 and the  $\gamma$ -phosphate of the bound GTP, coordination between Thr35 and  $Mg^{2+}$  ion and the H-bond between Gly60 and  $\gamma$ -phosphate are considered as an important structural indicator of the states<sup>36,71,72</sup>. To our knowledge, there are no crystal structures of Rho GTPases attributed to these particular conformational states of the SWI loop in the GTP bound form. Hence to identify the structural signatures of state 1 and state 2 like conformations in Rho GTPases, there was a need to explore a large conformational space of SWI region and sample the conformations based on their free energy to identify the energetically stable and metastable states in terms of known active and inactive states. Similar to the unsupervised clustering, the free energy surface incorporates the information about the possible conformation ensemble describing the energetically stable and metastable states. Although unbiased MD simulations provide very useful molecular insights regarding the dynamics of biomolecules, such simulations may have limited capabilities in sampling a complex free energy landscape.



**Figure 5.3** Free energy surfaces corresponding to (a) Freeform, (b) GDP bound form, (c) Wild type (WT) GTP bound form, and (d) GTP bound form with G14V mutation on the left panel. The position of the few representative available crystal structures have been marked on the free energy surface with the following colors: GDP (green star) and GTP (blue diamond). On the

### 5.3 Results

---

right panel (e – h), Structural superimpositions of switch I region conformations for different minima labelled as A(Blue), B(Green), C(Red).

In particular, large-scale conformational changes in biomolecules are often associated with high free energy barriers (greater than a few  $k_B T$ ) making them inaccessible (or poorly sampled) to normal MD simulations. Hence the well-tempered metadynamics method has been used to accelerate the conformational sampling of the SWI region in each of the systems under different consideration, namely (i) the GDP bound form, (ii) the GTP bound form, (iii) the nucleotide freeform, and (iv) the GTP bound form of the G14V mutation, which is known to exist in a constitutively active state. The free energy surfaces computed using the re-weighting approach<sup>73</sup> have been shown in Fig. 5.3.

**Table 5.3** Summary of Crystal structures overlaid on free energy of RhoA, Cdc42, Rac1 of (a) GDP bound form and (b) GTP bound form.

S.No.	PDB	Title: GDP bound form
11 <sup>#</sup>	2G0N	The crystal structure of the human Rac3 in complex with GDP and chloride
13 <sup>#</sup>	2W2T	Rac2 (G12V) in complex with GDP
14 <sup>#</sup>	1A4R	G12V mutant of human placental Cdc42 GTPase in the GDP form
15 <sup>#</sup>	1AN0	Cdc42Hs-GDP complex
S.No.	PDB	Title : GTP bound form
19 <sup>#</sup>	1KMQ	Crystal structure of a constitutively activated RhoA mutant (q63l)
21 <sup>#</sup>	2GCP	Crystal structure of the human RhoC-GSP complex
23 <sup>#</sup>	3TVD	Crystal structure of mouse RhoA-GTP complex
30 <sup>#</sup>	3SBD	Crystal structure of Rac1 P29S mutant
31 <sup>#</sup>	3SUA	Crystal structure of the intracellular domain of Plexin-B1 in complex with Rac1
32 <sup>#</sup>	3TH5	Crystal structure of wild-type Rac1
33 <sup>#</sup>	4GZM	Crystal structure of Rac1 F28L mutant



To identify the ‘active’ and the ‘inactive’ conformational states of the SWI loop “dihedral similarity” ( $S$ ) collective variable was chosen. The two collective variables  $S_{GDP}$  (X axis) and  $S_{GTP}$  (Y axis) are the  $S$  parameters calculated with respect to the GDP bound and GTP bound reference structures, respectively, indicate the dihedral similarity of SWI region. According to the definition of the  $S$  parameter, larger  $S$  values would indicate higher similarity with the reference structure, e.g. structures with higher  $S_{GDP}$  values will have greater degree of similarity to the reference GDP bound structure. The two-dimensional space shown in Fig. 5.3 ( $S_{GDP}, S_{GTP}$ ) is the reduced representation of the full conformational space of the backbone ( $2^N$  for the  $\phi/\psi$  angles of  $N$  residues). In order to validate this representation, the ( $S_{GDP}, S_{GTP}$ ) values for the SWI loop of total 36 crystal structures of RhoA, Rac1 and cdc42 GTPases in their GDP bound and GTP bound form (Table 5.3) were mapped on each of these free energy surfaces. In Fig. 5.3, only crystal structures existing in *apo* nucleotide-bound state have been superimposed. Interestingly the ( $S_{GDP}, S_{GTP}$ ) collective variables corresponding to the majority of GTP and GDP bound crystal structures form clusters, which nicely overlap with the global minima in the FES of the corresponding system. This clearly indicates that the conformational states discovered by the metadynamics sampling closely resemble the available structural data. Also, it is clear from the representation that the ( $S_{GDP}, S_{GTP}$ ) collective variables have the ability to distinguish between the active and inactive conformational states.

A comparative analysis of these maps suggests that in the absence of any nucleotide (Fig. 5.3a), a large number of conformational states are thermally accessible as indicated by the presence of multiple stable states in the dihedral conformational space. This unveils the highly flexible nature of the loop in the absence of nucleotide (GDP/GTP). On the other hand, the GDP bound structure is characterised with much more structured free energy surface with lower number of thermally accessible conformational states compared to the freeform. The FES for the GDP bound form (Fig. 5.3b) has three minima separated by rather small ( $\sim 2$  kcal/mol) barrier (Table 5.4) indicating a high degree of flexibility between these states. However, the GTP bound conformations are never visited in a GDP bound state, which would have required to cross much higher free energy barrier ( $>15$  kcal/mol). However, in the case of GTP bound structure, the number of deep minima (stable states) is even lower (Fig. 5.3c). This signifies the presence of a few distinct conformational states in both nucleotide bound systems unlike the freeform. But the position of the minima in the GTP bound state is significantly different as compared to the GDP

### 5.3 Results

bound system, indicating the unique conformational signatures of the GTP bound ON state. Interestingly, for the GTP bound form other metastable intermediate states, e.g. the state marked “B” with slightly higher free energy (~0.7 kcal/mol) than the state “A”, were observed. However, the barrier of the interconversion (A to B) is 5 kcal/mol.

**Table 5.4** *Relative free energy of the conformational states and the barrier between different Minima on Free Energy Surface in nucleotide unbound and bound states.*

S.No.	System	Transition States	Free Energy (kcal)
1	Nucleotide freeform	$\Delta G$ (A)	0
		$\Delta G$ (C)	4.6
		$\Delta G^\ddagger$ (A→C)	7
2	GDP bound state	$\Delta G$ (A)	0
		$\Delta G$ (B)	0.5
		$\Delta G^\ddagger$ (A→B)	2
3	(a) Wild Type	$\Delta G$ (A)	0
		$\Delta G$ (B)	0.7
		$\Delta G^\ddagger$ (A→B)	5
	(b) Mutant G14V	$\Delta G$ (A)	0
		$\Delta G$ (B)	-0.4
		$\Delta G^\ddagger$ (A→B)	2

In a previous study by Gorfe et al.<sup>35</sup>, it has been shown that the Ras mutants occur in the intermediate region between the GTP bound active and GDP bound inactive states, and hence they have hypothesised the existence of a lower free energy barrier in the oncogenic variants. In the case of RhoA, the oncogenic mutation G14V is known to be in a constitutively active state leading to an uncontrolled growth<sup>68,74</sup>. To investigate the effect this mutation on the conformational preference of the SWI loop of RhoA, FES for the G14V mutation in the GTP bound state was calculated (Fig. 5.3d). Although, the qualitative nature of the FES remains similar to the wild-type GTP bound form (Fig. 5.3c), but in the G14V system the identity of the

global minimum shifts to an intermediate state (marked “B”). The  $\Delta G$  for transition from A to B is +0.7 kcal/mol and -0.4 kcal/mol for the WT and G14V, respectively. Furthermore, the barrier of transition between the GTP bound active and GDP bound inactive states is lowered by 3 kcal/mol, thus signifying enhanced flexibility of the loop. It has been shown earlier that the G14V mutant becomes constitutively active since the chemical step of deactivation (GTP hydrolysis) is hindered in this mutant<sup>75</sup>. Our results for the G14V mutation corroborates the decreased free energy barrier with the increased flexibility in the switch I region as observed in the earlier study<sup>58</sup>. In addition, the current results indicate that the loop region undergoes a population shift of the conformational states in addition to increased flexibility, which might promote binding to any arbitrary effector proteins through induced fit mechanism. Thus, the G14V mutant might be able to activate a large variety of downstream effector with lower degree of selectivity/specificity.

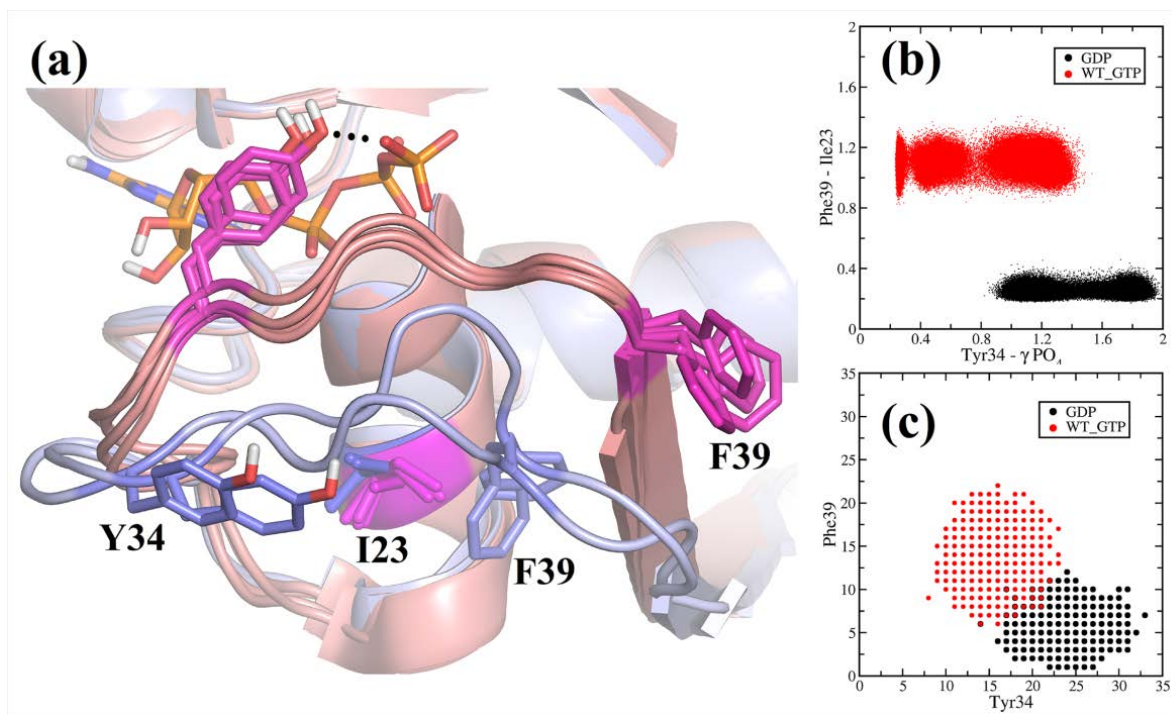
### **Rho GTPases exhibit signatures of state 1 and state 2 in their GTP bound form**

For H-Ras, the state 1 and state 2 conformations of the GTP bound protein was defined in terms of the distance between Thr35 and the  $\gamma$ -phosphate of the bound GTP. However, dynamics of other neighbouring residues, such as Tyr32 in H-ras (Tyr34 in RhoA numbering), were also shown to contribute in stabilising and defining these states<sup>76</sup>. For Rho GTPases there is no structural data attributing their state 1 and state 2 conformations. Hence all available GTP and GDP bound structures of the GTPases belonging to this family were compared to define the “active” and “inactive” states of the protein. The only difference expected in the GTP bound forms, compared to the GDP bound forms, is stabilisation of the SWI region by the Thr37 and  $\gamma$ -phosphate mediated interactions. However, for Rho GTPases, in addition to this interaction there is a correlated motion of Tyr34 and Phe39. In the GDP bound form the side chain of Tyr34 points outward; whereas it flips inside to form hydrogen bond with the  $\gamma$ -phosphate in case of the GTP bound state (Fig. 5.4a). Similarly, the hydrophobic Phe39 side chain is buried in the GDP form, whereas it flips out in the active state to become solvent exposed (Fig. 5.4a)<sup>77</sup>.

We have identified the nucleotide dependent correlation between the orientation of Tyr34 and Phe39 based on the distance between the nearest interacting partner with respective sidechain in the unbound and bound state (Fig. 5.4b). This was further substantiated with the analysis of average number of water molecules around the Tyr34 and Phe39 sidechains (Fig.

### 5.3 Results

5.4c). It is observed that Tyr34 is more solvent exposed in GDP bound form as compared to the GTP bound form, whereas Phe39 is more solvent exposed in the GTP bound form as compared to the GDP bound form. The 2D plot corroborates our claim that these two residues show distinct sidechain orientation in different nucleotide-bound state, and there exists a correlation among their orientation. Since these particular residues, Tyr34 and Phe39 and their correlated motion is uniquely found in Rho GTPases (Fig. 5.5), conformations akin to these were used as a structural signature to define the active and the inactive state of the GTPase.

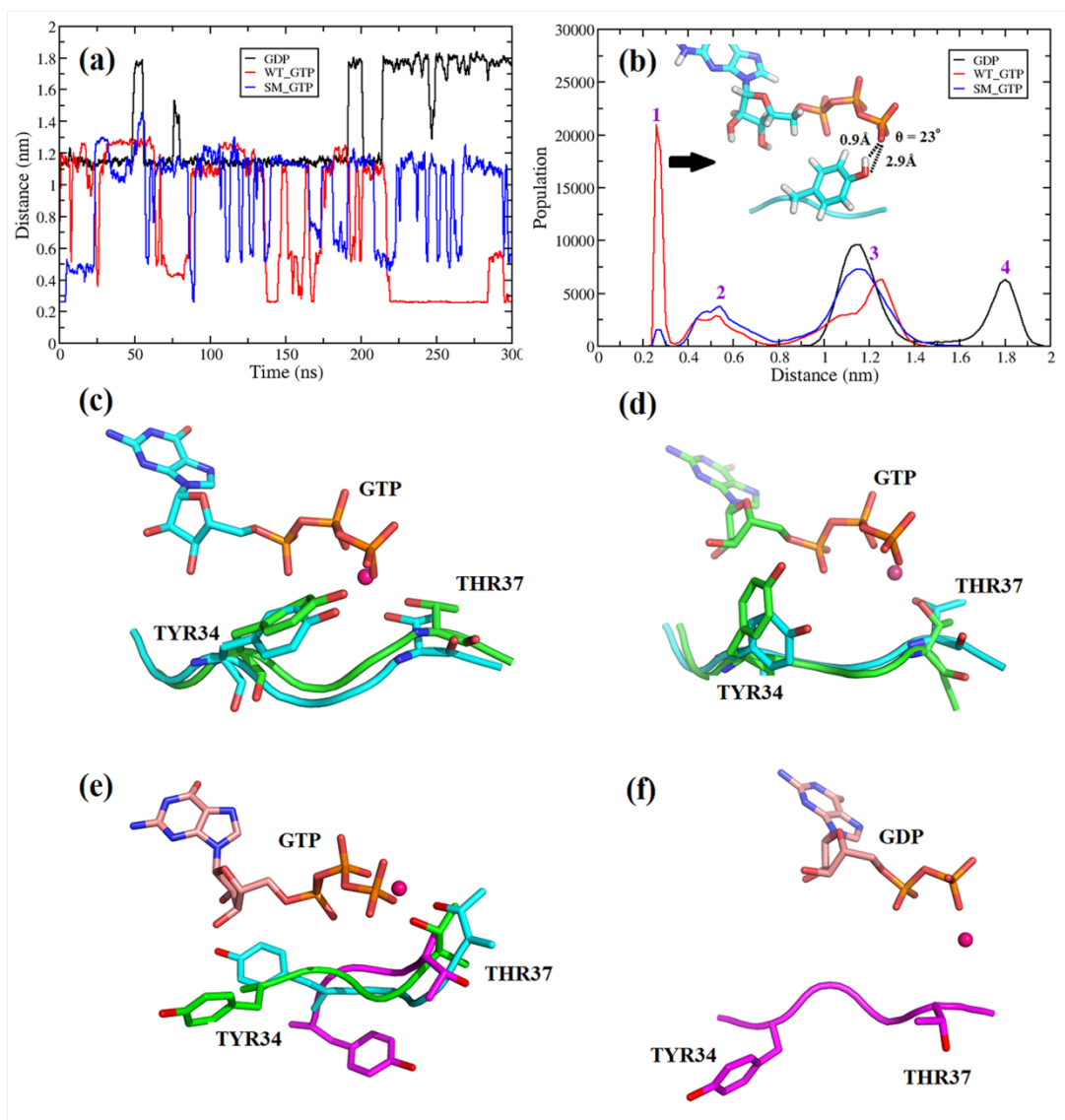


**Figure 5.4** (a) Orientation of Phe39 and Tyr34 in GDP bound crystal structures (Blue) and GTP bound crystal structures (Pink). Tyr34 sidechain forms hydrogen bond with the  $\gamma$ -phosphate oxygen atom whereas in case of GDP bound form, Phe39 sidechain interacts with Ile23 sidechain. (b) Minimum distance between the nearest possible interaction. (c) Solvent exposure of Tyr34 and Phe39 sidechain. Average number of water molecules was calculated around the Tyr34 and Phe39 residues.

To validate this assignment, conformations of SWI regions of the trajectories corresponding to the three minima in the FES of the nucleotide freeform, GDP bound form, GTP bound form and GTP bound form of the G14V mutation were carefully examined (Fig. 5.3e–h).

## 5. Conformational Heterogeneity in Rho GTPases

The minima corresponding to the nucleotide freeform do not show any correlated conformations of Tyr34 and Phe39, indicating the absence of well-defined states of the GTPase in this form. However, the minima corresponding to the GDP bound form have similar conformations of Phe39 and a modest change in the open conformation of the Tyr34.

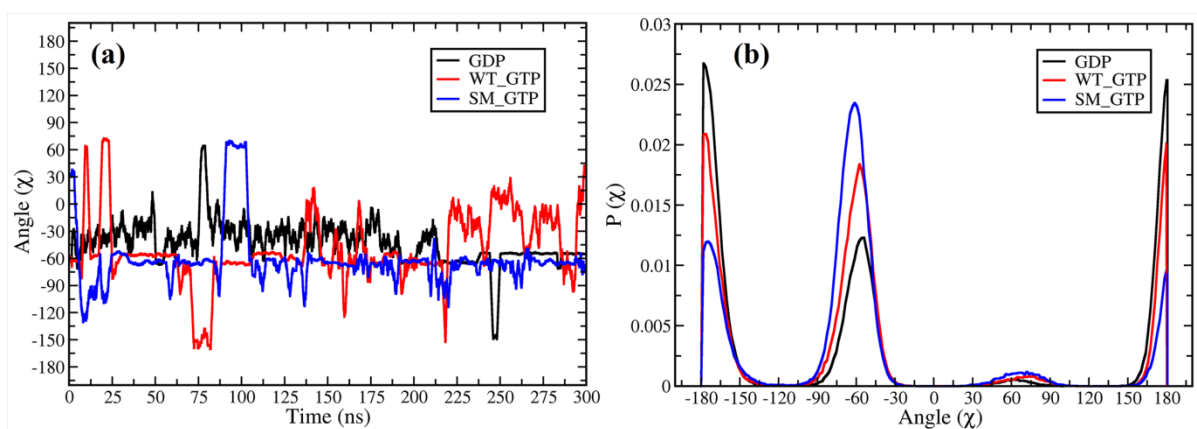


**Figure 5.5** Structural states in different systems based on the Tyr34 orientation. (a) Time evolution and (b) probability distribution of minimum distance between the oxygen atom of hydroxyl group in Tyr34 and oxygen atoms of terminal phosphate group of GDP and GTP molecule in GDP bound form, wild type GTP bound form (WT\_GTP), single mutation G14V-GTP bound form (SM\_GTP). The peak labelled as 1 represents the existence of hydrogen bond

### 5.3 Results

between Tyr34 and the terminal phosphate. The existence of multiple states is evident from the probability distribution plot which also exhibits a distinct nucleotide dependent shift of population. (c - f) Superimposition images of structures from each of the labelled four states coloured as WT-GTP (cyan), G14V-GTP (green), and GDP (magenta).

Perhaps, they all correspond to the same GDP bound state with marginal deviations. Interestingly, for the GTP bound form based on the conformations of the SWI region, the minima can be grouped into two distinct classes: one with Phe39 as in the active state (blue) and the other class related to the energy minima “B” & “C”, belonging to an intermediate state. In these two minima, the Try34 has the active state conformation. Perhaps, these two classes represent state 2 (corresponding to “A”) and state 1 (corresponding to “B” or “C”) of the GTP bound Rho GTPases, equivalent to those observed for the Ras family of GTPases.



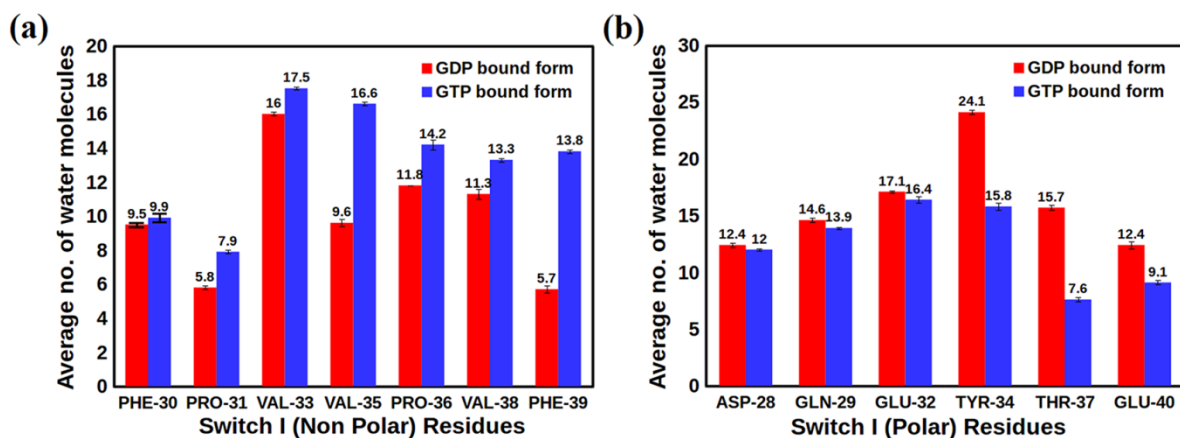
**Figure 5.6** Frequency of Tyr34 sidechain orientation. (a) Time evolution and probability distribution of  $\chi$  dihedral angle of Tyr34.

In case of GTP bound-G14V mutant, which is known to drive the protein towards state 1 (inactive), all the minima have Phe39 in an intermediate conformation similar to the one designated as the WT-GTP state 1. Thus, this particular observation further corroborates the existence of state 1 and state 2 like conformations in Rho GTPases. In Fig. 5.5, we demonstrate the probability distribution of the minimum distance between Tyr34 and GTP/GDP, highlighting the multiple possible intermediate orientations of Tyr34 depending on the nucleotide binding state. The probability distribution of the  $\chi$  dihedral angle of Tyr34 (Fig. 5.6) also clearly demonstrates the dramatic population shift in a nucleotide dependent manner as well as the

G14V point mutation.

### Competing molecular interactions leading to conformational selectivity and correlated motion

To further understand the molecular interactions that stabilise the respective active and inactive conformational states of the Rho GTPases, the role of solvation was investigated. The average number of water molecules within 4Å distance from each residue was calculated. (Fig. 5.7).

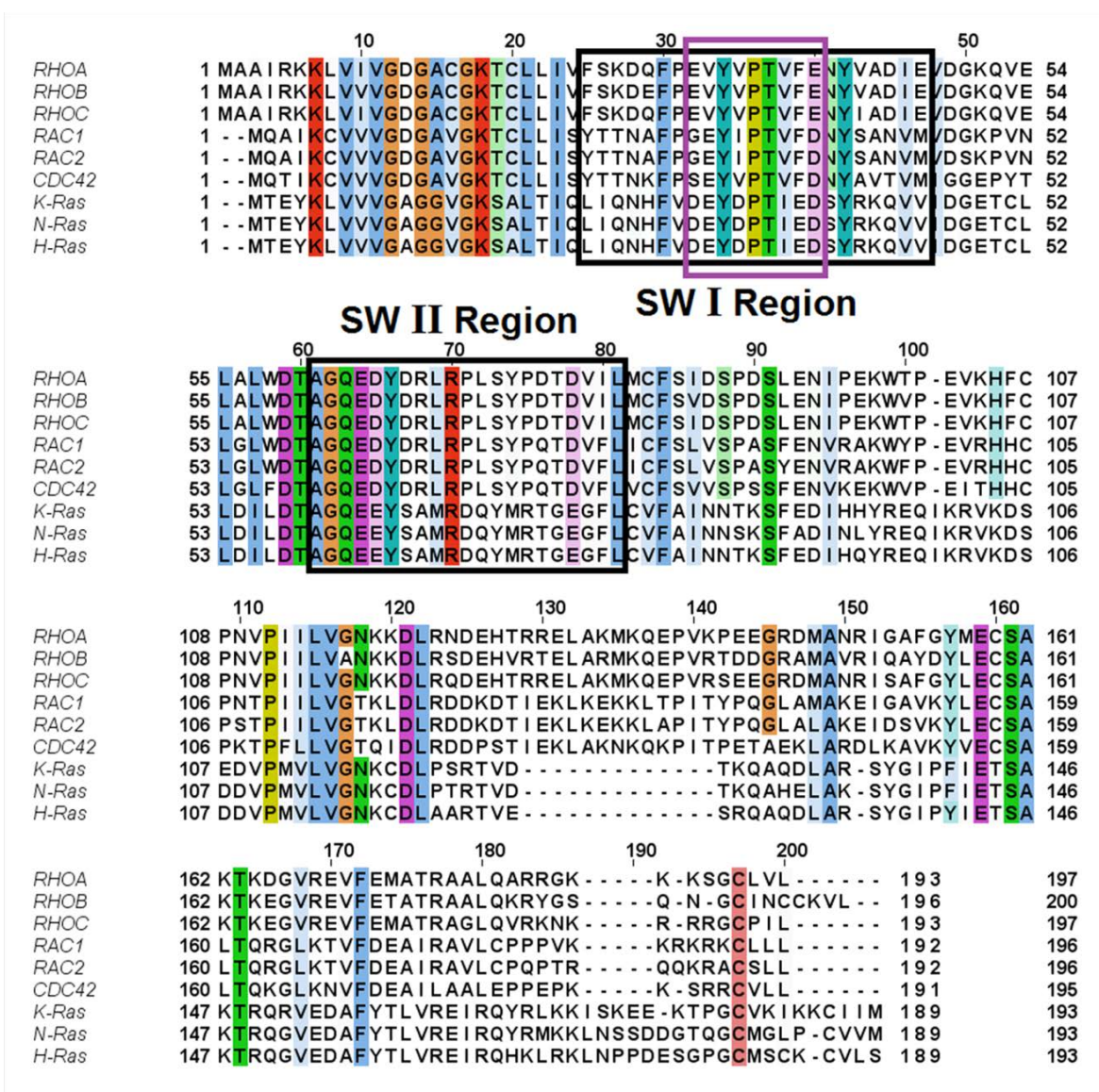


**Figure 5.7** Comparison of the average number of water molecules around (a) non-polar residues and (b) polar residues. The red and blue bars indicate the GDP bound and GTP bound states, respectively.

It was observed that all the hydrophobic (non-polar) residues have consistently higher solvent exposure in the GTP bound state, whereas the hydrophilic (polar/charged) groups have higher exposure in the inactive, GDP bound state, suggesting that the SWI region becomes more hydrophobic in the GTP bound active state which could further facilitate effector recognition, as seen in the case of RhoA-AKAP-Lbc (Rho GEF)<sup>78</sup> and Ras p21-Raf interactions<sup>79</sup>. In order to understand the possible biological significance of the hydrophobic residues in the SWI region, we have presented a sequence alignment comparison between the SWI region in H-Ras and RhoA (Fig. 5.8).

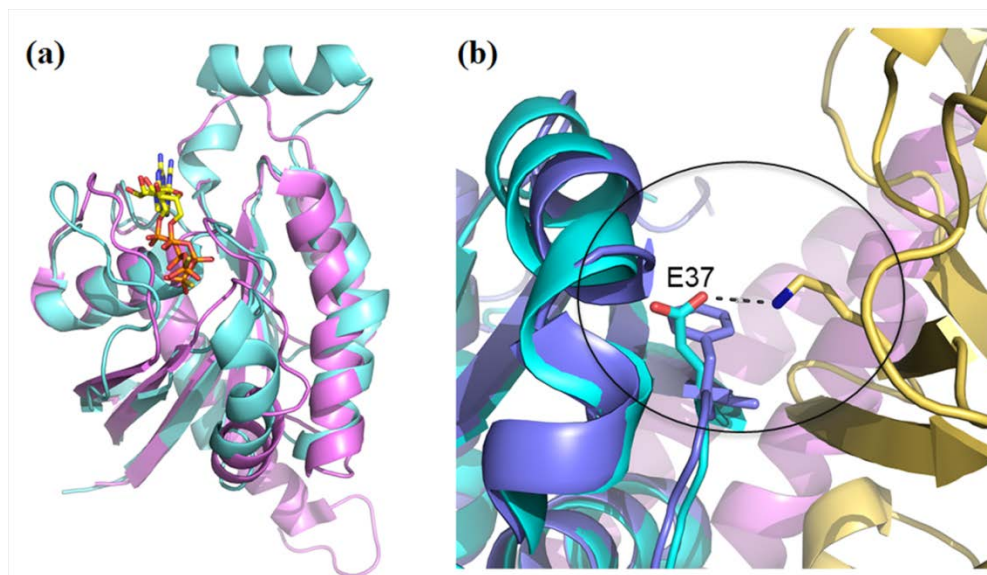


### 5.3 Results



**Figure 5.8** Multiple Sequence Alignment of human Rho GTPases using Clustal-Ω. Residues with greater than 65% identity/conservation are highlighted using Clustal X colouring scheme. The accession numbers are P06749 (RhoA), P01121 (RhoB), P08134 (RhoC), P15154 (Rac1), P15153 (Rac2), P25763 (Cdc42), P01116 (K-Ras), P01111 (N-Ras), and P01112 (H-Ras). The switch regions in Rho and Ras are highlighted in black and violet colour respectively.





**Figure 5.9** Structural comparison of H-Ras and RhoA protein in their effector bound state. (PDB: 1HE8 (H-Ras) and 1CXZ (RhoA)). (a) The figure highlights the difference in SWI loop region between H-Ras (magenta) and RhoA (cyan). (b) Molecular picture of effector recognition between K-Ras (cyan) and RhoA (blue). Superimposition of effector bound states in K-Ras and RhoA highlights the importance of substitution of Glu37 in K-Ras (RhoA numbering) interaction with Ras effector (yellow) eventually increasing the rigidity of the SWI region.

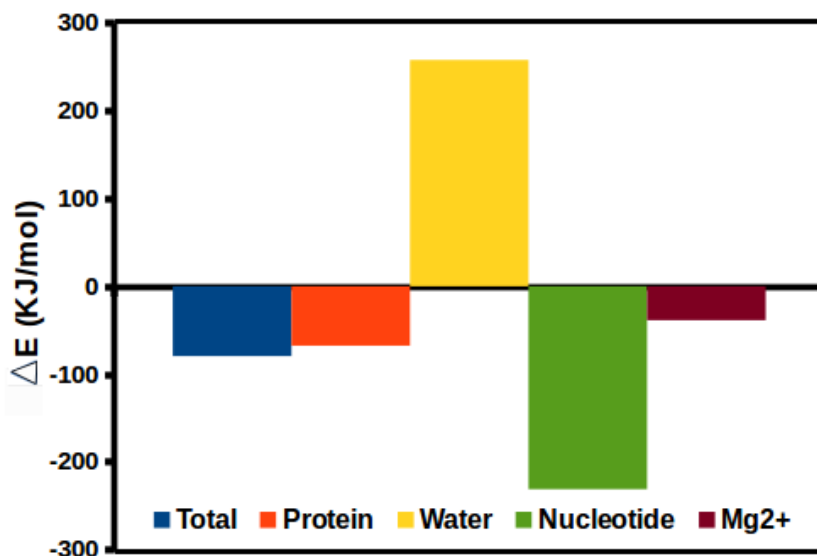
As compared with H-Ras, the SWI region in RhoA exhibits substitution by multiple non-polar residues in place of the charged residues in H-Ras at positions 33, 35 and 39 in accordance with RhoA numbering. These substitutions should result in significantly different solvation behaviour of the solvent exposed region between Ras and Rho GTPases. Prior experimental studies highlight the flipping out of hydrophobic residues Val35, Val38, and Phe39 towards the solvent region in the GTP bound form<sup>74</sup>. The importance of hydrophobic contacts of switch I and II regions in protein-protein interaction has been illustrated by Dvorsky *et.al.*<sup>80</sup> It is interesting to note that the Glu37 (H-Ras Numbering) in H-Ras is substituted by Phe39 (RhoA Numbering) in RhoA, which is also found to be in significant correlation with Tyr34 as demonstrated earlier. Furthermore, comparison of structures of Ras (PDB: 1HE8) and Rho (PDB: 1CXZ) bound to their effectors show significant differences in terms of length of the SWI loop and replacement of Phe39 with Glu37. In Ras GTPases, the length of the SWI loop is shorter by almost 12 residues, compared to its Rho homolog. Also, Glu37 (in H-Ras) provides additional rigidity to

### 5.3 Results

---

the SWI region through either intermolecular or through intra-molecular interactions (Fig. 5.9). Thus, these stabilising factors might make the Ras amiable to study the intermediate conformations SWI region using ensemble-based experimental techniques. However, for Rho GTPases, these additional stabilising factors are unfavourable for capturing the state 1 like conformations. Therefore, biochemical and structural studies of mutants that drive the Rho GTPases towards state 1 confirmation could further augment the current observations.

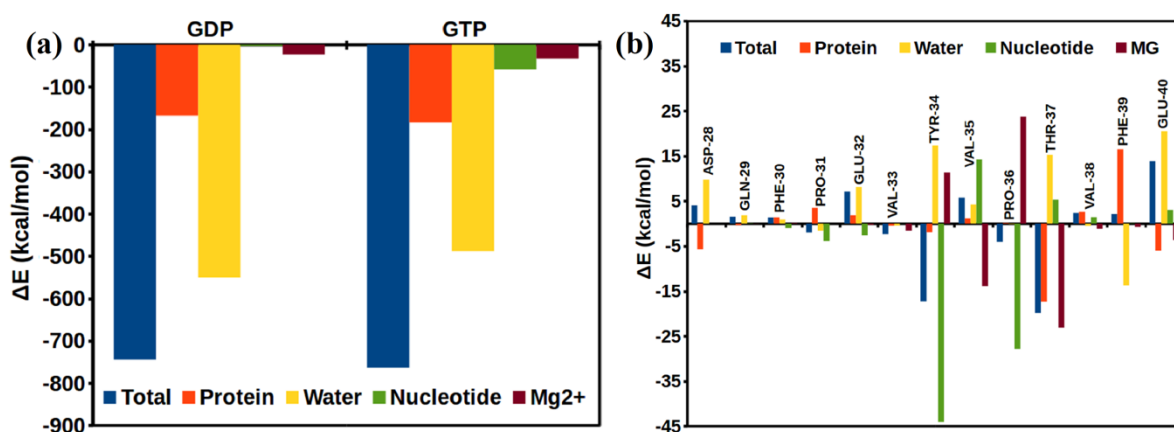
This observation poses two new questions: how does a conformational state with higher solvent exposure of hydrophobic residues can be stabilised in the GTP bound state and why state 1 like conformations are unfavourable to be detected with ensemble-based experimental techniques, such as NMR? Although the stabilisation might come from the hydrogen bonding interaction of a few residues (Tyr34 and Thr37) with the  $\gamma$ -phosphate of GTP, it is important to obtain a more quantitative picture. In order to understand the relative role of stabilisation of the SWI conformations due to specific interactions, the net total energy of the loop between GDP and GTP bound form was computed and the difference of each contribution is shown in Fig. 5.10. We have also decomposed the difference in total energy into individual contributions of the protein, solvent, nucleotide and the  $Mg^{2+}$  ion.



**Figure 5.10** The stabilization energy of the GTP bound state with respect to the GDP bound conformational state, i.e.  $\Delta E = E_{GTP} - E_{GDP}$ , In addition to the total stabilization energy (blue),

we have also shown the various components due to protein (orange), water (yellow), nucleotide (green) and the  $Mg^{2+}$  ion (brown).

These calculations demonstrate the fine energetic interplay that leads to preferential stabilisation of some conformational states over other depending on the nucleotide. In particular, the interaction with nucleotide and solvent seems to be the most important factor that affects the equilibrium (Fig. 5.11a). In the GDP bound inactive state all the hydrophobic groups are buried and the hydrophilic groups are exposed. Thus this conformational state is highly favoured by the solvation energy (interaction with water). In case of the GTP bound active state the interaction with water becomes highly unfavourable (exposed hydrophobic and buried hydrophilic groups), but even higher favourable interaction with the nucleotide (GTP) counterbalances the solvation energy term to make the total energy lower for this conformation. This fine balance (energetic see-saw) is quite remarkably maintained across all the residues as well<sup>81</sup>. The residue-wise distribution of the interaction energies has been shown in Fig. 5.11b.



**Figure 5.11** (a) Interaction energy distribution for GDP and GTP bound state of switch I region. It can be observed that the loss in the interaction energy in the GTP bound form due to the increase in solvent exposure of the switch I region is compensated by the energy contribution due to the nucleotide. (b) Residue-wise distribution of net interaction energy of switch I region between the two states,  $\Delta E = E_{GTP} - E_{GDP}$ , dissected into different components, i.e. total energy (blue), protein (orange), water (yellow), nucleotide (green) and the  $Mg^{2+}$  ion (brown). The energetic balance between different GTP components can be observed among SWI loop residues.

## 5.4 Discussion

---

The differential stability between the state 1 and state 2 conformations in the GTP bound state may also be explained based on the above energetic considerations. Based on the representative structures from free energy landscape, one can speculate that the state 1 conformation exhibits lack of stabilisation due to the reduced favourable interaction with the nucleotide. Furthermore, the energetic balance between the penalty of exposing the hydrophobic residues, and the favourable hydrogen bonded interactions of GTP with Tyr34 and Thr37 would dictate the thermodynamic stability and population of these conformational substrates.

## 5.4 Discussion

Rho GTPase signalling pathways are based on the formation of distinct protein-protein complexes involving a large number of regulators and downstream effectors. A comparative study of the intermolecular interaction sites between Rho GTPase and the binding proteins based on available structures highlight that most proteins interact with the residues in the switch regions of Rho GTPase, thus, making switch regions as an important feature of protein-protein interactions irrespective of the conformational states or the functional output of the complex. In this chapter, we have explored the ability of the switch I region to adopt a range of interfaces (plasticity) and the molecular nature of interactions (specificity) involved in molecular recognition of binding partners. We have presented one of the very few fully atomistic MD simulations for the Rho GTPases. Using an unsupervised machine learning analysis of the conformational ensemble obtained from the MD trajectory as well as free energy calculation, we have demonstrated that the SWI loop of the GTPases may exist in various metastable states. In the GTP bound form, SWI loop has state 1 and state 2 like conformations. For the G14V like mutant, the GTPases shifts more towards the state 1 like conformation. The intermediate conformational sub-states have been characterised with respect to the unique side-chain orientation, particularly with respect to the residues Tyr34 and Phe39, both of which undergo large movement between the active and inactive conformational states.

We have also demonstrated that we can characterise the GTP bound active conformational state by highly exposed hydrophobic groups and less exposed hydrophilic groups, whereas the reverse happens in the GDP bound inactive state. The stabilization of the exposed hydrophobic groups in the active state occurs because of the highly favourable interactions with the GTP. Thus, the fine balance between the interaction between nucleotide

and solvent leads to the shift in the conformational equilibrium depending on the nucleotide binding state. This observation that binding to GTP significantly pushes the conformational ensemble to exposing a hydrophobic patch around the SWI loop in the active state strongly suggests that such hydrophobic interactions might play a dominant role in effector recognition and binding through this region. The current study provides a framework for looking at GTP bound inactive state of Rho GTPases. However, further experimental studies involving measurement of affinities of the SWI mutants (especially Y34S, T37S/T37A and F39E) for the effectors is required to characterise this state. Perhaps, crystal structures of these mutants might throw light on the conformational signatures of these states.

### 5.5 Bibliography

1. Ganguly, S., R. Saxena, and A. Chattopadhyay, *Reorganization of the actin cytoskeleton upon G-protein coupled receptor signaling*. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2011. 1808(7): p. 1921-1929.
2. Vázquez-Victorio, G., et al., *Chapter 9 - GPCRs and actin-cytoskeleton dynamics*, in *Methods in Cell Biology*, A. K. Shukla, Editor. 2016, Academic Press. p. 165-188.
3. Yan, J. and T. Jin, *Signaling network from GPCR to the actin cytoskeleton during chemotaxis*. *Bioarchitecture*, 2012. 2(1): p. 15-18.
4. de Curtis, I. and J. Meldolesi, *Cell surface dynamics – how Rho GTPases orchestrate the interplay between the plasma membrane and the cortical cytoskeleton*. *Journal of Cell Science*, 2012. 125(19): p. 4435.
5. Hanna, S. and M. El-Sibai, *Signaling networks of Rho GTPases in cell motility*. *Cellular Signalling*, 2013. 25(10): p. 1955-1961.
6. Machesky, L.M. and A. Hall, *Rho: a connection between membrane receptor signalling and the cytoskeleton*. *Trends in Cell Biology*, 1996. 6(8): p. 304-310.
7. Wittinghofer, A. and I.R. Vetter, *Structure-function relationships of the G domain, a canonical switch motif*. *Annu Rev Biochem*, 2011. 80: p. 943-71.
8. Jaffe, A.B. and A. Hall, *Rho GTPases: biochemistry and biology*. *Annu Rev Cell Dev Biol*, 2005. 21: p. 247-69.
9. Sit, S.-T. and E. Manser, *Rho GTPases and their role in organizing the actin cytoskeleton*. *Journal of cell science*, 2011. 124: p. 679-683.
10. Nobes, C.D. and A. Hall, *Rho GTPases Control Polarity, Protrusion, and Adhesion during Cell Movement*. *The Journal of Cell Biology*, 1999. 144(6): p. 1235.
11. Vetter, I.R. and a. Wittinghofer, *The guanine nucleotide-binding switch in three dimensions*. *Science (New York, N.Y.)*, 2001. 294(November): p. 1299-1304.

## 5.5 Bibliography

---

12. Lowy, D.R. and B.M. Willumsen, *Function and Regulation of Ras*. Annual Review of Biochemistry, 1993. 62(1): p. 851-891.
13. Moon, S.Y. and Y. Zheng, *Rho GTPase-activating proteins in cell regulation*. Trends in Cell Biology, 2003. 13(1): p. 13-22.
14. Amin, E., et al., *Deciphering the Molecular and Functional Basis of RHOGAP Family Proteins: A systematic approach toward selective inactivation of rho family proteins*. Journal of Biological Chemistry, 2016. 291(39): p. 20353-20371.
15. Laurin, M. and J.-F. Côté, *Insights into the biological functions of Dock family guanine nucleotide exchange factors*. Genes & development, 2014. 28(6): p. 533-547.
16. Rossman, K.L., C.J. Der, and J. Sondek, *GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors*. Nature Reviews Molecular Cell Biology, 2005. 6(2): p. 167-180.
17. Garcia-Mata, R., E. Boulter, and K. Burrridge, *The 'invisible hand': regulation of RHO GTPases by RHOGDIs*. Nat Rev Mol Cell Biol, 2011. 12(8): p. 493-504.
18. Schmidt, G., et al., *Gln 63 of Rho is deamidated by Escherichia coli cytotoxic necrotizing factor-1*. Nature, 1997. 387(6634): p. 725-729.
19. Aktories, K., *Bacterial protein toxins that modify host regulatory GTPases*. Nat Rev Microbiol, 2011. 9(7): p. 487-98.
20. Sahai, E. and C.J. Marshall, *RHO-GTPases and cancer*. Nat Rev Cancer, 2002. 2(2): p. 133-42.
21. Lemichez, E. and K. Aktories, *Hijacking of Rho GTPases during bacterial infection*. Experimental Cell Research, 2013. 319(15): p. 2329-2336.
22. Parri, M. and P. Chiarugi, *Rac and Rho GTPases in cancer cell motility control*. Cell Commun Signal, 2010. 8: p. 23.
23. Mosteller, R.D., J. Han, and D. Broek, *Identification of residues of the H-ras protein critical for functional interaction with guanine nucleotide exchange factors*. Molecular and Cellular Biology, 1994. 14(2): p. 1104-1112.
24. Créchet, J.-B., A. Bernardi, and A. Parmeggiani, *Distal Switch II Region of Ras2p Is Required for Interaction with Guanine Nucleotide Exchange Factor*. Journal of Biological Chemistry, 1996. 271(29): p. 17234-17240.
25. Quilliam, L.A., et al., *Involvement of the Switch 2 Domain of Ras in Its Interaction with Guanine Nucleotide Exchange Factors*. Journal of Biological Chemistry, 1996. 271(19): p. 11076-11082.
26. Cherfils, J. and M. Zeghouf, *Regulation of Small GTPases by GEFs, GAPs, and GDIs*. Physiological Reviews, 2013. 93(1): p. 269.
27. Day, G.-J., R.D. Mosteller, and D. Broek, *Distinct Subclasses of Small GTPases Interact with Guanine Nucleotide Exchange Factors in a Similar Manner*. Molecular and Cellular Biology, 1998. 18(12): p. 7444-7454.

28. Segal, M., et al., *Two Distinct Regions of Ras Participate in Functional Interaction with GDP-GTP Exchangers*. *European Journal of Biochemistry*, 1995. 228(1): p. 96-101.
29. Miyazaki, K., S. Komatsu, and M. Ikebe, *Dynamics of RhoA and ROK $\alpha$  translocation in single living cells*. *Cell Biochemistry and Biophysics*, 2006. 45(3): p. 243-254.
30. Ito, Y., et al., *Regional Polyesterism in the GTP-Bound Form of the Human c-Ha-Ras Protein*. *Biochemistry*, 1997. 36(30): p. 9109-9119.
31. Geyer, M., et al., *Conformational Transitions in p21ras and in Its Complexes with the Effector Protein Raf-RBD and the GTPase Activating Protein GAP*. *Biochemistry*, 1996. 35(32): p. 10308-10320.
32. Matsumoto, S., et al., *Molecular Mechanism for Conformational Dynamics of Ras-GTP Elucidated from In-Situ Structural Transition in Crystal*. *Scientific Reports*, 2016. 6: p. 25931.
33. Spoerner, M., et al., *Dynamic properties of the Ras switch I region and its importance for binding to effectors*. *Proceedings of the National Academy of Sciences of the United States of America*, 2001. 98: p. 4944-4949.
34. Baussand, J. and J. Kleinjung, *Specific Conformational States of Ras GTPase upon Effector Binding*. *J Chem Theory Comput*, 2013. 9(1): p. 738-749.
35. Gorfe, A.A., B.J. Grant, and J.A. McCammon, *Mapping the nucleotide and isoform-dependent structural and dynamical features of Ras proteins*. *Structure*, 2008. 16(6): p. 885-96.
36. Kobayashi, C. and S. Saito, *Relation between the Conformational Heterogeneity and Reaction Cycle of Ras: Molecular Simulation of Ras*. *Biophysical Journal*, 2010. 99(11): p. 3726-3734.
37. Prakash, P. and A.A. Gorfe, *Lessons from computer simulations of Ras proteins in solution and in membrane*. *Biochim Biophys Acta*, 2013. 1830(11): p. 5211-8.
38. Prakash, P. and A.A. Gorfe, *Overview of simulation studies on the enzymatic activity and conformational dynamics of the GTPase Ras*. *Molecular simulation*, 2014. 40(10-11): p. 839-847.
39. Grant, B.J., J.A. McCammon, and A.A. Gorfe, *Conformational Selection in G-Proteins: Lessons from Ras and Rho*. *Biophysical Journal*, 2010. 99(11): p. L87-L89.
40. Grant, B.J., A.a. Gorfe, and J.A. McCammon, *Ras conformational switching: Simulating nucleotide- dependent conformational transitions with accelerated molecular dynamics*. *PLoS Computational Biology*, 2009. 5(3): p. 1-10.
41. Kumar, A., et al., *Molecular dynamic simulation reveals damaging impact of RAC1 F28L mutation in the switch I region*. *PLoS One*, 2013. 8(10): p. e77453.
42. Diaz, J.F., B. Wroblowski, and Y. Engelborghs, *Molecular dynamics simulation of the solution structures of Ha-ras-p21 GDP and GTP complexes: flexibility, possible hinges, and levers of the conformational transition*. *Biochemistry*, 1995. 34(37): p. 12038-12047.

## 5.5 Bibliography

---

43. Noé, F., et al., *Automated computation of low-energy pathways for complex rearrangements in proteins: Application to the conformational switch of Ras p21*. *Proteins: Structure, Function, and Bioinformatics*, 2005. 59(3): p. 534-544.
44. Noé, F., et al., *Transition Networks for the Comprehensive Characterization of Complex Conformational Change in Proteins*. *Journal of Chemical Theory and Computation*, 2006. 2(3): p. 840-857.
45. Phillips, M.J., et al., *Effector Proteins Exert an Important Influence on the Signaling-active State of the Small GTPase Cdc42*. *Journal of Biological Chemistry*, 2008. 283(20): p. 14153-14164.
46. Emsley, P., et al., *Features and development of Coot*. *Acta Crystallographica Section D*, 2010. 66(4): p. 486-501.
47. Hess, B., et al., *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*. *Journal of Chemical Theory and Computation*, 2008. 4(3): p. 435-447.
48. MacKerell, a.D., N. Banavali, and N. Foloppe, *Development and current status of the CHARMM force field for nucleic acids*. *Biopolymers*, 2001. 56(4): p. 257-65.
49. Zoete, V., et al., *SwissParam: A fast force field generation tool for small organic molecules*. *Journal of Computational Chemistry*, 2011. 32(11): p. 2359-2368.
50. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. *The Journal of Chemical Physics*, 1983. 79(2): p. 926-935.
51. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. *The Journal of Chemical Physics*, 2007. 126(1): p. 014101.
52. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*. *Journal of Applied Physics*, 1981. 52(12): p. 7182-7190.
53. Essmann, U., et al., *A smooth particle mesh Ewald method*. *The Journal of Chemical Physics*, 1995. 103(19): p. 8577-8593.
54. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations*. *Journal of Computational Chemistry*, 1997. 18(12): p. 1463-1472.
55. Hall, M., et al., *The WEKA data mining software: an update*. *SIGKDD Explor. Newsl.*, 2009. 11(1): p. 10-18.
56. Barducci, A., G. Bussi, and M. Parrinello, *Well-tempered metadynamics: A smoothly converging and tunable free-energy method*. *Physical Review Letters*, 2008. 100(2): p. 1-4.
57. Tribello, G.A., et al., *PLUMED 2: New feathers for an old bird*. *Computer Physics Communications*, 2014. 185(2): p. 604-613.
58. Rajendran, V., C. Gopalakrishnan, and R. Purohit, *Impact of point mutation P29S in RAC1 on tumorigenesis*. *Tumor Biology*, 2016. 37(11): p. 15293-15304.



59. Lu, S., et al., *The Structural Basis of Oncogenic Mutations G12, G13 and Q61 in Small GTPase K-Ras4B*. Scientific Reports, 2016. 6: p. 21949.
60. Hall, B.E., D. Bar-Sagi, and N. Nassar, *The structural basis for the transition from Ras-GTP to Ras-GDP*. Proceedings of the National Academy of Sciences of the United States of America, 2002. 99: p. 12138-12142.
61. Kuppens, S., J.F. Díaz, and Y. Engelborghs, *Characterization of the hinges of the effector loop in the reaction pathway of the activation of ras-proteins. Kinetics of binding of beryllium trifluoride to V29G and I36G mutants of Ha-ras-p21*. Protein Science : A Publication of the Protein Society, 1999. 8(9): p. 1860-1866.
62. MacQueen, J. *Some methods for classification and analysis of multivariate observations*. in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. 1967. Berkeley, Calif.: University of California Press.
63. Hartigan, J.A. and M.A. Wong, *Algorithm AS 136: A K-Means Clustering Algorithm*. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979. 28(1): p. 100-108.
64. Münz, M., J. Hein, and P.C. Biggin, *The Role of Flexibility and Conformational Selection in the Binding Promiscuity of PDZ Domains*. PLOS Computational Biology, 2012. 8(11): p. e1002749.
65. Sim, A.Y.L. and M. Levitt, *Clustering to identify RNA conformations constrained by secondary structure*. Proceedings of the National Academy of Sciences, 2011. 108(9): p. 3590-3595.
66. Yang, S., N.K. Banavali, and B. Roux, *Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories*. Proceedings of the National Academy of Sciences, 2009. 106(10): p. 3776-3781.
67. Alan, J.K. and E.A. Lundquist, *Mutationally activated Rho GTPases in cancer*. Small GTPases, 2013. 4(3): p. 159-163.
68. Zhao, X., et al., *Overexpression of RhoA Induces Preneoplastic Transformation of Primary Mammary Epithelial Cells*. Cancer Research, 2009. 69(2): p. 483-491.
69. Muraoka, S., et al., *Crystal structures of the state 1 conformations of the GTP-bound H-Ras protein and its oncogenic G12V and Q61L mutants*. FEBS Letters, 2012. 586(12): p. 1715-1718.
70. Sahai, E., A.S. Alberts, and R. Treisman, *RhoA effector mutants reveal distinct effector pathways for cytoskeletal reorganization, SRF activation and transformation*. The EMBO Journal, 1998. 17(5): p. 1350-1361.
71. Shima, F., et al., *Structural Basis for Conformational Dynamics of GTP-bound Ras Protein*. Journal of Biological Chemistry, 2010. 285(29): p. 22696-22705.
72. Sharma, N., U. Sonavane, and R. Joshi, *Probing the wild-type HRas activation mechanism using steered molecular dynamics, understanding the energy barrier and role of water in the activation*. European Biophysics Journal, 2014. 43: p. 81-95.

## 5.5 Bibliography

---

73. Tiwary, P. and M. Parrinello, *A Time-Independent Free Energy Estimator for Metadynamics*. The Journal of Physical Chemistry B, 2015. 119(3): p. 736-742.
74. Ihara, K., et al., *Crystal structure of human RhoA in a dominantly active form complexed with a GTP analogue*. J Biol Chem, 1998. 273(16): p. 9656-66.
75. Khrenova, M.G., et al., *Modeling the Role of G12V and G13V Ras Mutations in the Ras-GAP-Catalyzed Hydrolysis Reaction of Guanosine Triphosphate*. Biochemistry, 2014. 53(45): p. 7093-7099.
76. Buhrman, G., G. Wink, and C. Mattos, *Transformation Efficiency of RasQ61 Mutants Linked to Structural Features of the Switch Regions in the Presence of Raf*. Structure, 2007. 15(12): p. 1618-1629.
77. Dias, S.M.G. and R.A. Cerione, *X-ray Crystal Structures Reveal Two Activated States for RhoC*. Biochemistry, 2007. 46(22): p. 6547-6558.
78. Abdul Azeez, Kamal R., et al., *The crystal structure of the RhoA-AKAP-Lbc DH-PH domain complex*. Biochemical Journal, 2014. 464(2): p. 231-239.
79. Ma, J. and M. Karplus, *Molecular switch in signal transduction: Reaction paths of the conformational changes in ras p21*. Proceedings of the National Academy of Sciences, 1997. 94(22): p. 11905-11910.
80. Dvorsky, R. and M.R. Ahmadian, *Always look on the bright site of Rho: structural implications for a conserved intermolecular interface*. EMBO Rep, 2004. 5(12): p. 1130-6.
81. Miyakawa, T., et al., *Solvent site-dipole fields around guanine nucleotides in the Hras-GTP complex and in the Hras-GDP complex*. AIP Conference Proceedings, 2014. 1599(1): p. 322-325.

# Chapter 6

## Molecular Insights into the Regulation of Rac1 by Phosphorylation of GDI

### 6.1 Introduction

The Guanine Dissociation Inhibitor (RhoGDI) plays a pivotal role in Rho GTPase regulation and acts as a negative regulator by blocking the activation of GTPases from the inactive GDP bound state<sup>1-4</sup>. As discussed in the introduction chapter of this thesis, post-translational modification at the C-terminal of Rho proteins and subsequent membrane attachment plays an important role in the activation of Rho GTPases signaling pathways<sup>5-9</sup>. In the absence of GDI, Rho proteins are membrane localised where the geranylgeranyl moiety attached at the carboxy-terminal cysteine residue is inserted into the lipid bilayer<sup>10, 11</sup> and anchors the Rho proteins to the cellular membrane<sup>12-18</sup>. However, GDI binds to these prenylated forms creating a soluble cytosolic Rho.GDP-RhoGDI complex regulating the cytoplasmic pool of each of the Rho family GTP-binding proteins<sup>19-22</sup>. X-ray crystallographic structures reveal two important sites of interaction between GDI and Rho proteins<sup>23-25</sup>; (a) the N-terminal region of the GDI also called as “regulatory arm” folds into helix-loop-helix and binds to the Switch I and II of the Rho proteins inhibiting the GDP dissociation and GTP hydrolysis, (b) the C-terminal region of GDI adopts an immunoglobulin-like fold with a hydrophobic binding pocket which interacts with the geranylgeranyl moiety of the Rho proteins leading to the complete sequestration of the prenyl moiety from the solvent.

The binding of GDIs to the Rho GTPases at effector recognition sites leave the complex biologically inert with the reduction in the GEF-catalysed nucleotide exchange reaction with Dbl family GEFs<sup>26</sup>, and hence, the activation of Rho proteins by GEFs (Guanine nucleotide exchange factors) requires the release of RhoGDI. Biochemical studies based on *in vitro* kinase assays suggest phosphorylation as a mechanism for the dissociation of classical GTPases (RhoA, Cdc42, Rac1) from GDIs. In addition, post-translational lysine acetylation was also shown to

## 6.1 Introduction

regulate RhoA activity<sup>27</sup>. The phosphorylation and lysine acetylation sites are primarily concentrated in the geranylgeranyl binding pocket (immunoglobulin domain) and regulatory arm of GDI which interacts with switch regions<sup>5</sup>. Recent studies show more than eight sites of lysine acetylation, which also involves the N-terminal domain. Out of this, acetylation at K127 and K141 in RhoGDI were found to interfere with the binding of GDI to nonprenylated RhoA<sup>28</sup>. However, post-translational phosphorylation turns out to be a major regulator of RhoGDI's binding to Rho GTPases<sup>29-38</sup> (Table 6.1).

**Table 6.1** shows all post-translational phosphorylation sites that regulate RhoGDI-GTPase interaction. The residue numbering is based on human RhoGDI-1 (GDI- $\alpha$ ) sequence. The symbol '?' denotes the residue position is not known.

PTM Site	Kinase	RhoGDI	Effect of phosphorylation on the GTPase
Ser34	PKC $\alpha$	GDI- $\alpha$	Promotes dissociation of RhoA
Ser96	PKC $\alpha$	GDI- $\alpha$	Promotes dissociation of RhoA and RhoG
Ser101	PAK1	GDI- $\alpha$	Promotes dissociation of Rac1 (and to a lesser extent Cdc42)
Ser174	PAK1	GDI- $\alpha$	Promotes dissociation of Rac1 (and to a lesser extent Cdc42)
Ser174	PKA	GDI- $\alpha$	Inhibits interaction with RhoA
Ser148	ND	GDI- $\alpha$	Not determined (ND)
Thr?	PKC $\zeta$	GDI- $\alpha$	Promotes dissociation of RhoA, Rac1, Cdc42
Tyr27	SRC	GDI- $\alpha$ , GDI- $\beta$	Promotes dissociation of RhoA, Rac1, Cdc42
Tyr156	SRC	GDI- $\alpha$ , GDI- $\beta$	Promotes dissociation of RhoA, Rac1, Cdc42
Tyr156	FER	GDI- $\alpha$	Promotes dissociation of Rac1

The residues that are phosphorylated in GDIs controls the release of specific GTPase from the complex. For example, phosphorylation of RhoGDI-1 on Ser101 and Ser174 by p21-activated kinase 1 (PAK1) promotes the release of Rac1 (but not RhoA)<sup>29</sup>, whereas phosphorylation of RHOGDI-1 on Ser34 by protein kinase C $\alpha$  (PKC $\alpha$ ) selectively releases RhoA (but not Rac1 or cdc42)<sup>31</sup>. It is speculated that Ser34 interferes the interaction of HTH (GDI) with Arg68 (RhoA) in the switch region triggering dissociation. However, Arg68 (RhoA) or Arg66 (Rac1/Cdc42) is conserved in most of the Rho GTPases, and hence the mechanism behind such specificity is still elusive<sup>39</sup>. Since the number of GDI is limited to three ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) that

interacts with multiple Rho GTPases<sup>3</sup>, it has been suggested that “a unique phosphorylation code may exist” which control the release of specific Rho GTPases based on the stimuli/response. However, relatively little is known about the molecular basis of dissociation between specific Rho GTPase and RhoGDIs.

Interestingly, recent NMR studies show that the N-terminal region of GDI with residues 9-20 and 36-58 are disordered in the absence of Rho proteins<sup>40-43</sup>. Crystal structures show that the later region of N-terminal domain (residues 36-58) adopts an ordered helix-turn-helix (HTH) structure through interactions with switch regions of Rho proteins in all complexes where hydrogen bonds stabilise the HTH-Switch I region interaction between conserved residues Thr35, Val36 of Rac1 and Asp45, Ser47 of GDI respectively<sup>44</sup>. However, the extreme N-terminal residues (9-20) were shown to exist in equilibrium between two conformations (random-coil or helix) depending on the GTPase and GDI member in the complex<sup>45</sup>. It forms a small helix at position 10-15 in the recent crystal structures of RhoA-GDI (PDB: 4F38)<sup>10</sup> and Cdc42-GDI (PDB: 1DOA)<sup>24</sup> complexes while exits as an extended loop in Rac1-GDI complex<sup>25</sup> (PDB: 1HH4). Further functional studies using GDI with truncated N-terminal region elucidate the differential role of N-terminal domain in the inhibition of GDP dissociation and GTP hydrolysis independently.

Determining the key interactions is crucial in understanding the molecular mechanism of the Rho protein and GDI dissociation in the cellular environment. How does phosphorylation of specific residues of Rho GDI affect a unique (individual) member of Rho GTPase (either RhoA / Rac1 / cdc42) remains an open question and demands study of the molecular interactions at the atomistic level. In this chapter, we provide insights into the conformational dynamics and interactions that play an important role in defining the “unique phosphorylation code”. We have performed conventional molecular dynamics simulations and metadynamics simulations for enhanced conformational sampling of the wild-type and phosphorylated state of Rac1-GDI complex and propose a molecular-interaction based mechanistic model for the dissociation of the complex as an effect of phosphorylation.

### 6.2 Methods

#### 6.2.1 System Setup

The classical Rho GTPases namely RhoA, Rac1, cdc42 are geranylgeranylated as compared to the other majority of the Rho GTPases which are farnesylated. The crystal structure for the Rac1 and GDI in the complex<sup>25</sup> was selected with geranylgeranyl at the C-terminal of Rac1 (PDB ID:1HH4). The missing residues in GDI at position 301-308, 359-365 and 502-504 were modelled using Modeller program<sup>46</sup>. The number of amino acid residues were 189 in Rac1 and 204 in GDI. The simulations were performed with GDP and Mg<sup>2+</sup> ion present in the crystal structure. The system was equilibrated for 500ns after modelling the missing residues in the loops. Two different systems were generated from the equilibrated structure: (1) Wild type-without phosphorylation (2) Serine phosphorylation at positions 101, 174 in GDI. The phosphorylated structure was further equilibrated for 500ns. The GDI residues were renumbered as 301-504 where S401 and S474 represent the phosphorylated serine residues.

#### 6.2.2 Simulation Parameters

Molecular dynamics simulations for the wild-type and phosphorylated Rac1-GDI complex were performed using Gromacs software (version 5.0.7)<sup>47</sup>. The Charmm36 force field<sup>48</sup> with cmap corrections was used for the protein and modified cysteine-geranylgeranyl residue. The parameters for GDP molecule were obtained using CGenFF program<sup>4</sup>, which performs automated assignment of parameters and charges by analogy and compatible with Charmm force field. All the structures were solvated using TIP3P water model<sup>50</sup> and simulated with periodic boundary conditions. The systems were found to be negatively charged and were neutralised by adding Na<sup>+</sup> ions as counter ions. The structures were energy minimised using steepest descent algorithm. This was followed by NVT equilibration using modified Berendsen thermostat<sup>51</sup> and NPT equilibration using Parinello-Rahman barostat<sup>52</sup>. For the production run, the temperature was controlled through velocity rescaling at 300K with a time constant of 0.1 ps and pressure was kept constant at 1bar. The cutoff for short-range interactions was 1.0 nm, and the long-range electrostatic interactions were calculated using Particle-Mesh Ewald (PME) method<sup>53</sup>. The bonds were constrained using the LINCS algorithm<sup>54</sup>. We have performed two different types of simulations; conventional MD simulation and well-tempered metadynamics simulation<sup>55</sup>. The

conventional MD simulations were performed for 2 $\mu$ s for both the systems with frames saved at every 2ps. The metadynamics simulations were performed until convergence was achieved.

### 6.2.3 Analysis

#### Principal Component Analysis (PCA)

Principal component analysis is a method used to reduce the dimensionality of data obtained from simulations and identify dominant modes of motions in a system. This is done by diagonalisation of the covariance matrix obtained from the Cartesian coordinates of the superimposed conformations from the trajectory. In the present study, PCA was carried out on the backbone atoms of the individual trajectories of wild type and phosphorylated systems. We have compared the configurational space between wild-type trajectory and phosphorylated serine trajectory by calculating the first five principal components (PC) which account for the motions that constitute the essential subspace. This analysis will help us to identify the coupled low-frequency motions as well as intermediate conformational states between the two states and the effect of phosphorylation in terms of dynamics of the system.

#### Energetic perturbation due to phosphorylation

In chapter 3, we demonstrate that electrostatic interaction provides a highly sensitive yardstick to probe the allosteric modulation in contrast to the traditionally used structure-based parameters. To explore the effect of phosphorylation, i.e. the effect of addition of negative charge in the protein and subsequent energetic perturbation in distal parts of the protein (allostery), we have computed the average electrostatic interaction for each residue ( $E_i$ ) and compared between the wild type and phosphorylated states of the complex. The change in average electrostatic interaction energy of i-th residue is given by:

$$\Delta E_i = E_{i,SP} - E_{i,WT} \quad (6.1)$$

where  $\Delta E_i$  is the change in the average interaction energy between the wild type (WT) and phosphorylated (SP) states due to the interactions between the i-th residue and protein, the notation indicates an ensemble average over the trajectory for that particular state (wild type /

## 6.2 Methods

---

phosphorylated). Similarly, the change in average interaction energy for all residue pairs ( $E_{ij}$ ) is given by  $\Delta E_{ij} = E_{ij,SP} - E_{ij,WT}$ . This analysis will help us to identify how the local energetic perturbation due to phosphorylation propagates to distal parts of the protein through rearrangement of the intra-protein interaction network.

### **Perturbation in intra-protein hydrogen bond occupancy**

One of the major factor that influences the intra-protein interaction pattern is hydrogen bonding between the residues which involve both backbone and sidechain atoms. We have shown earlier that the rearrangement and re-wiring of such hydrogen bonded network plays a crucial role in allosteric signal propagation. The hydrogen bond occupancy (%) is defined as the percentage frequency of the hydrogen bond formation between two different residue pairs over all the frames in the trajectories given by,  $Hb_{ij} = nh_{ij} / N$ , where  $Hb_{ij}$  is the hydrogen bond (H-bond) occupancy (%) between  $ij$ -th residue pair,  $nh_{ij}$  is the number of H-bonds formed between  $ij$ -th residue pair over all the  $N$  frames. It was calculated using HBonds Plugin, Version 1.2 in VMD software. The donor-acceptor cutoff distance used was  $3.5\text{\AA}$ , and the Acceptor-Donor-Hydrogen angle must be less than  $30^\circ$ . The frequency of H-bond formation between a residue pair can be greater than 100% because there may be more than one H-bond between these residues and each H-bond is counted as separately. We have calculated the net H-bond occupancy (%) for all residue pairs between wild type and phosphorylated trajectories, given by,

$$\Delta Hb_{ij} = Hb_{ij}^{SP} - Hb_{ij}^{WT} \quad (6.2)$$

where  $Hb_{ij}$  is the change in H-bond occupancy (%) between  $i$ -th and  $j$ -th residue pair. This analysis will help us to explore how the hydrogen bonded network might be rearranging as a response to the phosphorylation.



### Metadynamics simulations

We have performed well-tempered metadynamics simulations to enhance conformational sampling and explore the free energy surface corresponding to the interaction of the polybasic region (PBR) of Rac1 with the cavity residues of GDI and the N-terminal residues of GDI. All metadynamics simulations<sup>55</sup> were performed using Gromacs code (version 5.0.7)<sup>47</sup> and Plumed plugin (version 2.3.5)<sup>56</sup>. We have used coordination number ( $CN$ ) as the collective variable to bias specific interactions between the polybasic region of Rac1 and the N-terminal/cavity of GDI that governs the conformational states in the wild type and phosphorylated Rac1-GDI complex. The collective variable, coordination number is defined as,

$$CN = \sum_{i=1}^{N_{Rac1}} \sum_{j=1}^{N_{GDI}} \frac{1 - \left(\frac{r_{ij}}{2.5}\right)^6}{1 - \left(\frac{r_{ij}}{2.5}\right)^{12}} \quad (6.3)$$

Here,  $i$  and  $j$  represent indices of a set of selected atoms of the polybasic region of Rac1 ( $N_{Rac1}$ ) and the N-terminal/cavity residues of GDI ( $N_{GDI}$ ) respectively,  $r_{ij}$  is the distance between the  $i$ -th and  $j$ -th atom. The sum is extended over the selected set of  $i$  and  $j$  atoms. The cutoff distance was chosen as 2.5Å in order to include the distance criteria of the H-bond. Thus, the coordination number provides an estimate of the number of hydrogen bonds that are formed between selective regions of Rac1 and GDI. The value of  $CN$  approaches zero when there is no interaction between the defined entities. The hill height ( $\omega$ ) was set to 0.5 kJ/mol, and the width of the Gaussian potential is 2.0 added every 1ps with a bias factor parameter of 5 at 300K.

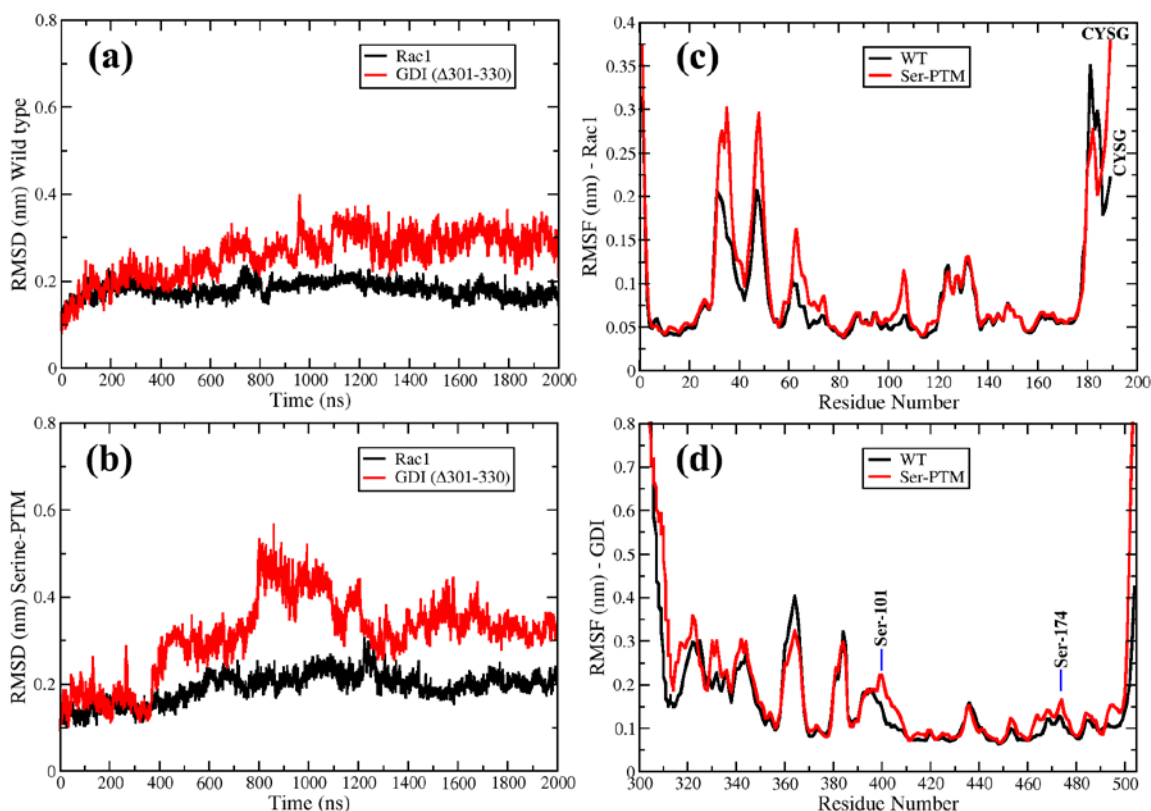
## 6.3 Results

### Structural perturbation in Rac1-GDI complex on phosphorylation

To examine how serine phosphorylation would affect the structure of the Rac1-GDI complex, we have calculated RMSD for Rac1 and GDI separately, of the backbone atoms with respect to the initial equilibrated structure. The structural deviation of Rac1 and GDI was compared between the wild type and the phosphorylated state (Figs. 6.1a-b). Previous NMR studies highlight the flexible nature of GDI N-terminal region<sup>40, 42</sup>; hence, RMSD for GDI was calculated without the

### 6.3 Results

first 30 residues from the N-terminal region. In the case of wild type (Fig. 6.1a), Rac1 and GDI were found to be stable with no large deviation from the initial backbone conformation. However, the RMSD plot in case of phosphorylated serine (Fig. 6.1b, red line) shows drift of the GDI conformation from the initial structure. We observed an increase in the RMSD value at around 400ns to  $\sim 0.3$ nm, which changes drastically at around 800ns to  $\sim 0.5$ nm and returns to  $\sim 0.3$ nm after a brief period of 200ns. However, unlike GDI, we do not observe such deviation for Rac1 in the phosphorylated state.



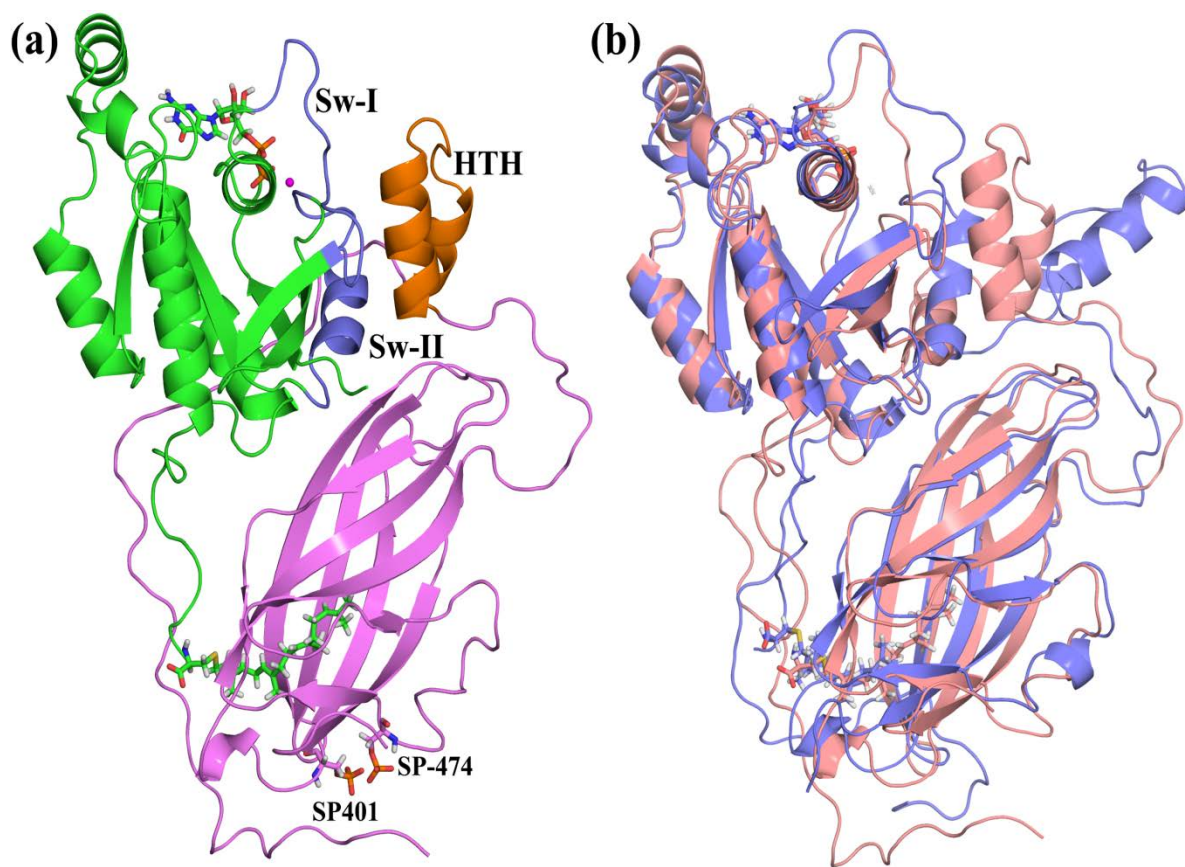
**Figure 6.1** (a and b) RMSD plots of wild-type and phosphorylated complex. The RMSD for Rac1 (black) and GDI without N-terminal residues (blue) was calculated for each system. The structure of Rac1-GDI complex was found stable in wild type. In the case of serine phosphorylation, RMSD value increases drastically for the GDI. (c, d) RMSF plots for the Rac1 and GDI in the wild type and phosphorylated complex. The switch regions in Rac1 show large fluctuations in case of phosphorylated serine.

In addition, we have also computed the fluctuations for these proteins with respect to the average structure to identify any large motions upon phosphorylation. Interestingly, we observe

an increase in fluctuations in switch regions (residues 28-38, 60-72) and geranylgeranylated cysteine residue at the C-terminal of Rac1 upon phosphorylation (Fig. 6.1c). Ideally, the geranylgeranyl moiety at Rac1 C-terminal is buried in the GDI hydrophobic cavity, and an enhanced fluctuation suggests a possible change in dynamics due to phosphorylation. The changes in fluctuations in GDI are at the N-terminal region residues (301-340) and the localised region around the site of phosphorylation.

### Conformational motions in interacting regions of Rac1 and GDI

Crystal structures of Rho GTPase-GDI complex exhibit two major sites of inter-protein interactions that are distantly located from each other (Fig. 6.2a).



**Figure 6.2** (a) Representative image of Rac1-GDI complex with serine phosphorylation at positions 401 and 474 (GDI numbering starts from 301 to 504). Rac1 is shown in green and GDI in magenta. The image highlights the interaction between switch regions (blue) and the helix-turn-helix region. The geranylgeranyl group is shown inside the cavity (sticks). The

### 6.3 Results

---

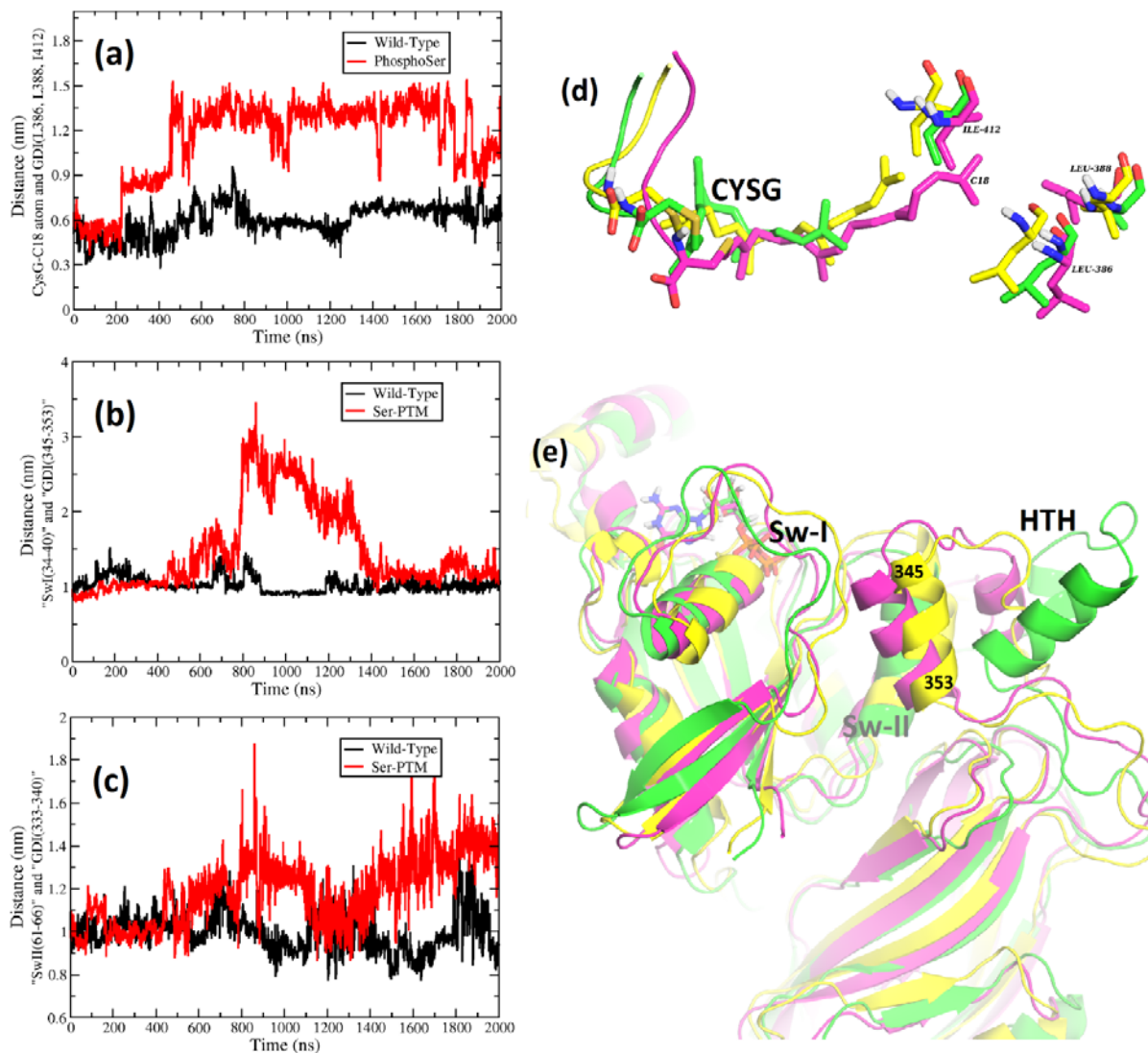
*phosphorylation sites are also shown as SP401 and SP474. The large distance between the Sw-HTH interaction and the geranylgeranyl-cavity interaction can be observed in the figure. (b) Superimposition of structures from time 100ns (pink) and 850ns (blue) show large structural deviation in the switch regions, HTH, and the loop regions.*

For the complex to dissociate, the interactions must break between the switch regions of Rac1 and HTH region of GDI along with the removal of prenyl moiety of Rac1 from the hydrophobic cavity of GDI. Experimentally the dissociation rate for RhoA-GDI complex bound to GDP was observed to be  $10^{-4} \text{ s}^{-1}$  with a half-life of  $\sim 60 \text{ mins}^{10}$ , which highlights that the proteins have a large binding affinity. Hence within MD simulation, it will be computationally inaccessible to observe dissociation between the two entities due to the large free energy barrier between the free states and the complex. The serine phosphorylation sites are located in the immunoglobulin domain of GDI which binds the geranylgeranyl moiety of Rac1. Biochemical studies (kinase and GTPase activation assays) suggest phosphorylation of both S101 and S174 is necessary for GDI dissociation<sup>29</sup>. Further crystal structures show that these phosphorylated residues are within 6-8Å of each other<sup>29</sup>. Hence based on these evidences, it is speculated that the electrostatic repulsion between these negatively charged residues would result in the destabilisation of the interactions between the prenyl moiety and GDI cavity residues.

In the previous section, we have observed a large deviation in the RMSD value of RhoGDI upon phosphorylation, which suggests structural deviation from the initial structure. To identify these structural changes, we superimpose the structures from different time frames along the simulations that show deviation in RMSD values (Fig. 6.2b). Surprisingly, we observe destabilisation of interactions between the HTH region of GDI and Switch regions of Rac1, situated far from the site of phosphorylation. In addition, we observe deviation in the C-terminal polybasic region of the Rac1, which can account for the localised effect of phosphorylation.

To assess the stability of the interactions and to explore any time-dependent connection between the destabilisation of prenyl moiety in the cavity and the HTH-switch regions, we have calculated the minimum distance between these different interacting regions of Rac1 and GDI. The stability of the geranylgeranyl moiety in the hydrophobic cavity was calculated based on the distance between the terminal of the prenyl group (geranylgeranyl) and the selective non-polar residues of GDI (L386, L388, I412) which interact with the terminal carbon atom (C18) of the

prenyl sidechain. An increase in the distance between these regions would indicate the destabilisation of interactions. We observe destabilisation in the position of geranylgeranyl in the binding pocket with time (Fig. 6.3a).



**Figure 6.3** (a, b and c) Minimum distance analysis between the interacting regions of Rac1 and HTH fold of GDI to observe the deviation from the initial structure with time. The distance was calculated between the center of mass of the two groups. (a) The minimum distance between the terminal carbon of the prenyl group and interacting residues of the hydrophobic cavity was calculated to identify destabilisation of prenyl moiety in the cavity. (b, c) Minimum distance calculation between the switch regions and HTH show destabilisation of interactions with increasing distance. (d, e) Superimposed images of the Rac1-GDI complex in the phosphorylated

### 6.3 Results

---

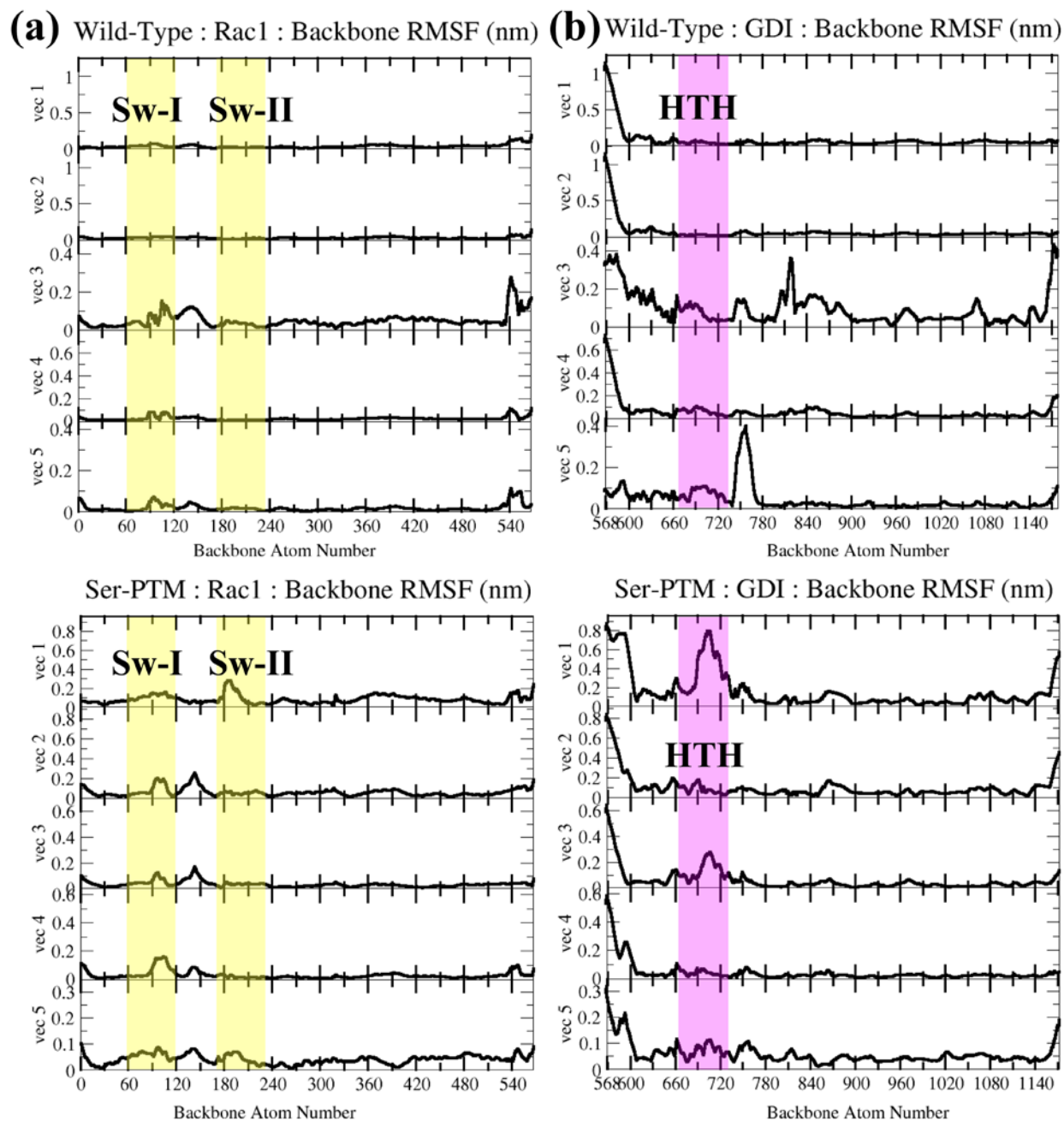
*state at different times of the trajectory. Structures are in magenta at 0ns, green at 1000ns and yellow at 1900ns. The structures show destabilisation of prenyl moiety in the cavity and the HTH interaction with the switch regions.*

Further, we show that the HTH fold of GDI interacts stably with the switch regions in the wild type. However, upon phosphorylation, the distance between the switch regions and HTH increases with time, which indicates the destabilisation of the complex (Figs. 6.3b-c). We observe a drastic change in the distance between the switch I (residues 34-40) and GDI (residues 345-353) at around 800ns. Figs. 6.3d and 6.3e show the superimposed image of the interacting regions in the phosphorylated complex at different times from the trajectory. Interestingly, we find back and forth motion in the HTH region, which suggests intermittent breaking and forming of interactions with the switch regions.

#### **Correlated motions in the Rac1-GDI complex**

We have performed principal component analysis (PCA) using the backbone atoms for the individual trajectories to identify the principal components/correlated motions through the elimination of small amplitude fluctuations and capture the conformational changes in the complex upon phosphorylation in terms of few dominant modes. The results from the PCA are presented in Figs. 6.4-6.5.





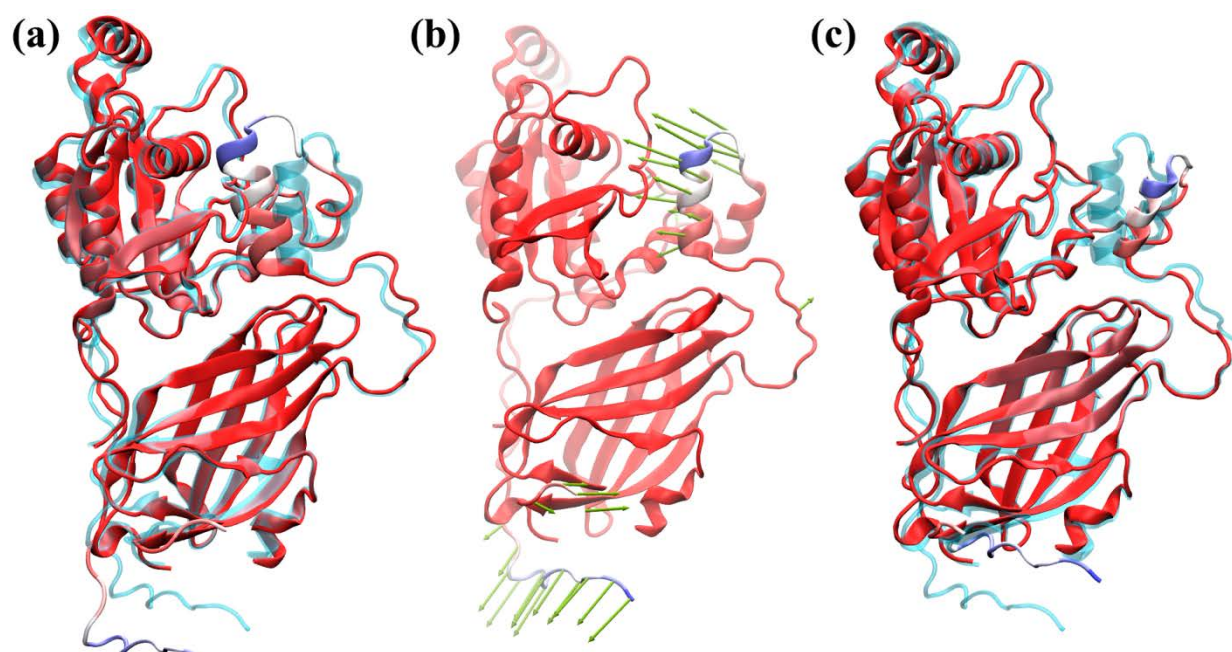
**Figure 6.4** (a, b) Principal component analysis: RMSF of the first five eigenvectors for both the systems (wildtype and phosphorylated) for Rac1 and GDI, respectively. The switch regions and HTH are highlighted in yellow and magenta, respectively.

Figs. 6.4a-b display the backbone RMSF of the first five eigenvectors for both the systems (wildtype and phosphorylated) for Rac1 and GDI respectively. In the principal component analysis, the first PC corresponds to the largest motion observed in the system. In

### 6.3 Results

---

both proteins (Rac1 and GDI), comparison of the first eigenvector shows conformational changes upon phosphorylation and regions showing large shifts comprise residues 60-69 (switch II) of Rac1 and 334-351 (helix turn helix (HTH)) of GDI. Inspection of the motions along the eigenvector 1 reveals that the phosphorylation increases the motion in the regulatory arm (HTH) of GDI which binds to the Sw-I and Sw-II regions of Rac1. In the later modes (PC2-PC5), we also observe conformation fluctuations in the Sw-I region of Rac1 upon phosphorylation. In the case of GDI, the lower modes show increased motions in the loop that connects the HTH region and the C-terminal domain in wild type and phosphorylated systems.



**Figure 6.5** *The extreme conformations sampled during the simulation from the average structure along the eigenvector1 in case of phosphorylated trajectory. (a, c)The structures on the left and right show the extreme conformations (red) sampled with respect to the average structure (cyan.) (b) The center figure shows the average structure with the conformational projections along the eigenvector1 shown in green. The HTH region and N-terminal region of GDI show correlation with the loop connecting the N-terminal domain and C-terminal domain of GDI.*

We used eigenvector 1 to filter the MD trajectories and isolated two extreme structures associated with this component to interpret and describe the correlated motions in the system. Upon phosphorylation, a large directional motion is visible in the Sw-II of Rac1 and HTH region of GDI (Fig. 6.5). In addition, the motions are correlated in the HTH region and N-terminal



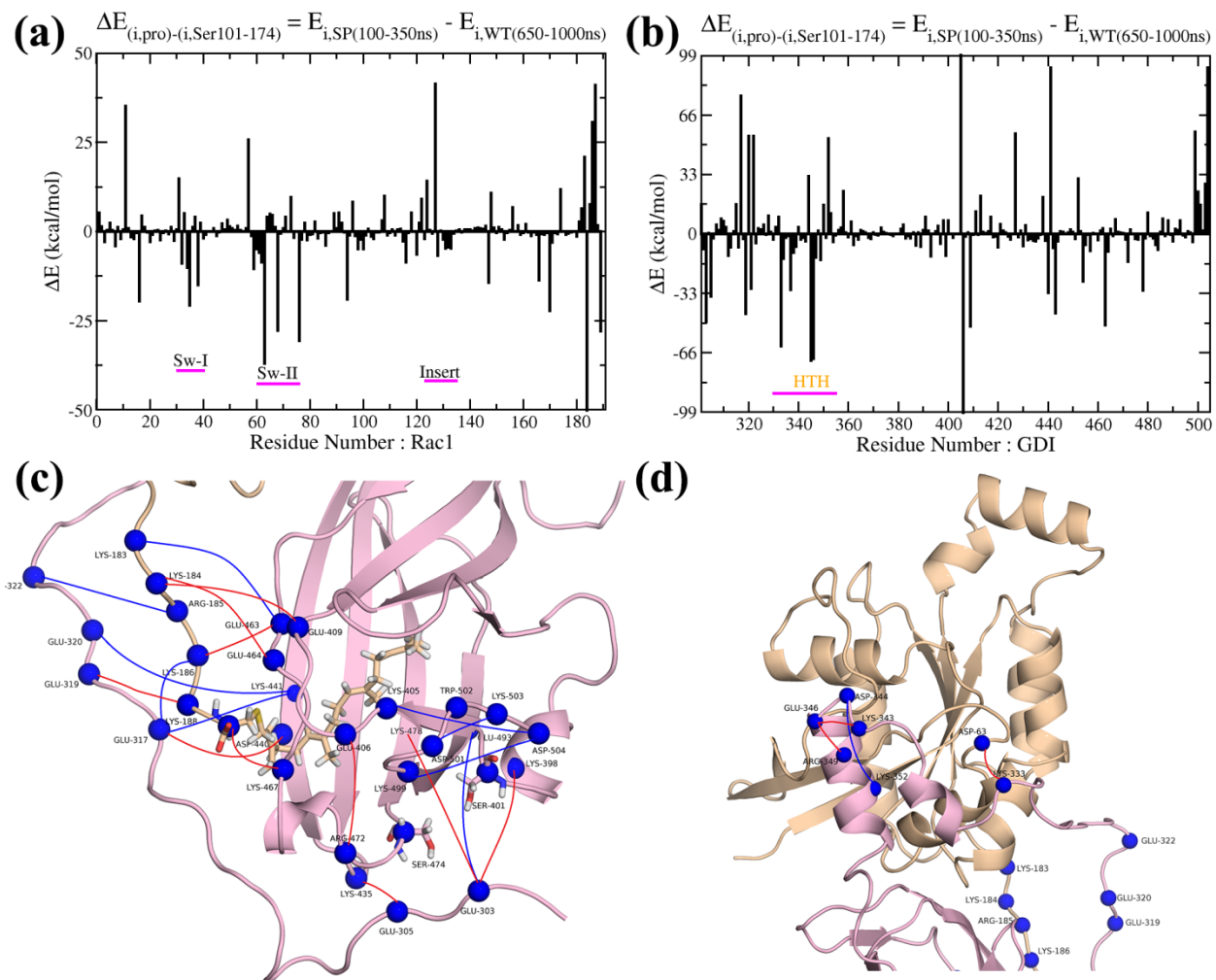
region of GDI with a loop that connects the immunoglobulin domain (C-terminal) with the HTH regions. However, these observations do not provide molecular details and physical understanding of how phosphorylation would promote the dissociation of Rac1-GDI complex. HTH interaction is found to play an important role in the stabilisation of the Rac1-GDI complex and prevent nucleotide exchange in Rho GTPases. It is observed that the distance between the site of phosphorylation (S401, S474) and the HTH region in GDI is greater than 2nm. Thus a long range effect may exist between the site of the phosphorylation and the breaking of favourable interactions between the regulatory arm (HTH) of GDI and switch regions of Rac1. In the previous chapter (Chapter-4), we have demonstrated the effect of protonation in the PDZ3 domain in the form of rearrangement of electrostatic interactions and its significant role in allosteric communication. Hence, it will be interesting to look at the energetics of the Rac1-GDI complex in further analysis to capture any such long range communications due to phosphorylation.

### **Phosphorylation-induced perturbation in the electrostatic interactions and H-bond network**

To probe the effect of negative charge addition on the energetics of the system, we have calculated the interaction energies for Rac1 and GDI residues in the wild type and phosphorylated state. The electrostatic interaction energy is conformation dependent quantity. Any large change in the structure would result in abrupt changes in the interaction energy. Hence, we have divided the phosphorylated trajectory into three different segments based on the RMSD value for GDI system; SPP1 (100-350ns), SPP2 (450-800ns) and SPP3 (800-1000ns). In Figs. 6.6a-b, we can identify residues that show a significant change in interaction energy upon phosphorylation. We identify these residues are from switch II (D57, D63, R68, D76 with  $\Delta E_i \geq \pm 25$  kcal/mol) and polybasic regions in Rac1 (K184, K186, R187, CYSG189). Interestingly, the geranylated modified cysteine residue shows a favourable change in interaction energy ( $\Delta E_i \geq -27.9$  kcal/mol) upon phosphorylation. Similarly, in GDI (Fig. 6.6c), the N-terminal residues comprising of HTH region and residues that are part of the cavity show a large change in interaction energy  $\Delta E_i \geq \pm 25$  kcal/mol. The residue-wise change in interaction energy ( $\Delta E_i$ ) was further dissected into change in pair-wise interaction energy ( $\Delta E_{ij}$ ) between Rac1 and GDI

### 6.3 Results

to probe the origin of such large energetic difference. Figs. 6.6c-d shows all the interactions between the PBR and the N-terminal/cavity region with  $\Delta E_{ij} \geq \pm 30$  kcal/mol. Our analysis captures the effect of phosphorylation (addition of negative charge) on the residues with the large change in inter-residue interaction energy.

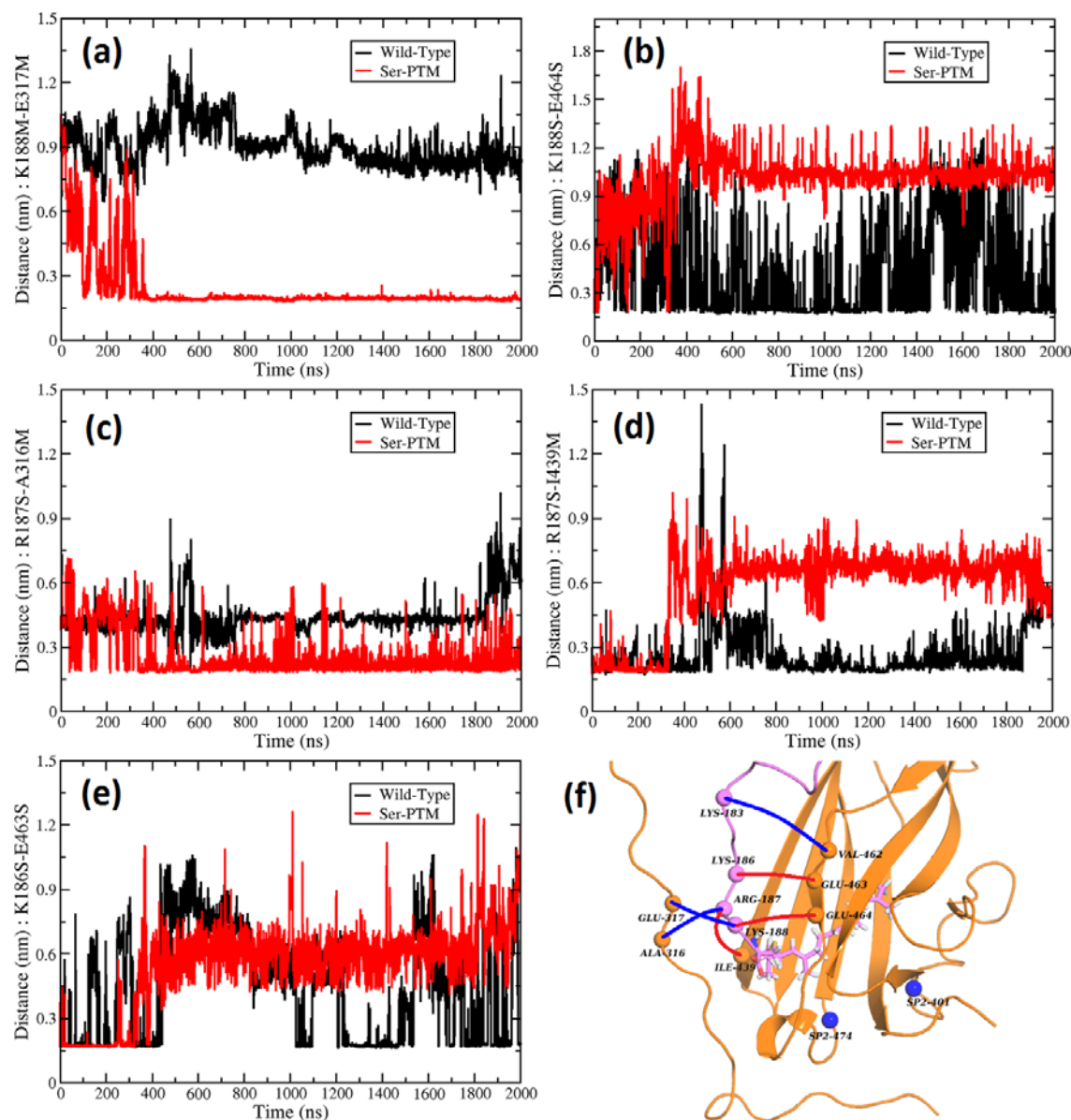


**Figure 6.6** (a, b) Residue-wise change in the interaction energy ( $\Delta E_i$ ) of Rac1 and GDI between wild type and phosphorylated system. (c, d) Representative image that shows the pairwise interaction energy ( $\Delta E_{ij}$ ) for all residues with  $\Delta E_i \geq \pm 25$  kcal/mol (blue spheres). Pairwise interactions with  $\Delta E_{ij} \geq 30$  kcal/mol are shown in blue and  $\Delta E_{ij} \leq -30$  kcal/mol in red.

We observe the localised effect of phosphorylation in terms of change in the charge-charge interactions between Rac1 C-terminal region and GDI. This is characterised by the rearrangement of

## 6. GDI-mediated Regulation of Rho GTPases

Rac1-GDI interacting residues which highlights the loosening of the binding of prenyl moiety in the hydrophobic cavity. It is trivial that to release/pull out the prenyl moiety from the cavity, some interactions must be broken/formed. In Figs. 6.7(a-e), minimum distance plots between selective charge-charge residues pairs show that the polybasic region of C-terminal of Rac1 is being held/(pulled out) by the N-terminal residues of GDI while the interactions with the GDI cavity are broken.



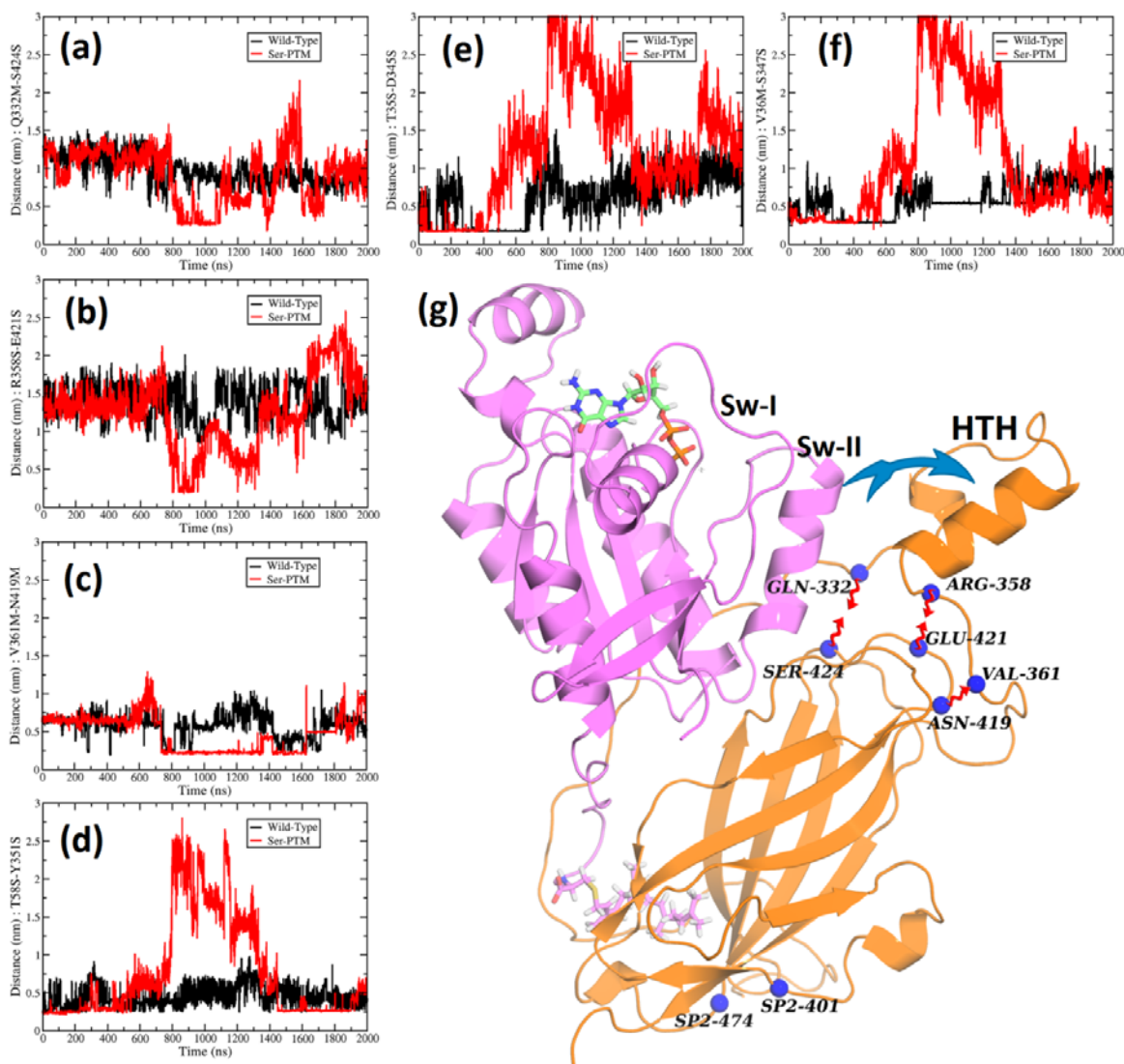
**Figure 6.7** Salt bridges/Hydrogen bonds related to the movement (destabilisation) of the prenyl group in the hydrophobic cavity. (a-e) Minimum distance between selective residue pairs to

### 6.3 Results

---

*identify formation and breaking of interaction with time. (f) Representative image to show the rearrangement of interactions.*

Fig. 6.7f shows these change of interactions between the charged residues in the form of H-bond formation as blue lines and breaking as red lines upon phosphorylation. The time evolution of these distances can be directly correlated with the pulling out of the prenyl group (Fig. 6.3a) around 200-400ns. Thus we have identified the specific interactions responsible for anchoring (in WT) and detachment (upon phosphorylation) of the prenyl group. Further, we have identified a few specific salt bridges/hydrogen bond interactions (e.g. V361-N419, R358-E421, Q332-S424 in Figs. 6.8d-f), which correlate with the outward movement of the HTH domain (Fig. 6.3b) around the 800-1200ns time-scale. It can be hypothesised that some hydrogen bonds (/salt bridges) pull the HTH domain like a drawbridge (Fig. 6.8g). To summarise, we have identified the specific interactions responsible for the structural rearrangements (i) pulling out of prenyl moiety around 200-400 ns, followed by (ii) HTH movement around 800-1200ns.



**Figure 6.8** *Specific interactions (salt bridge/H-bond) that control the drawbridge-like motion of the HTH domain: (a, b) Minimum distance between selective residue pairs show the difference in hydrogen bond formation occupancy in the wild type and phosphorylated state. (g) Representative image that shows the draw-bridge-like motions as a consequence of the tightening of specific interactions.*

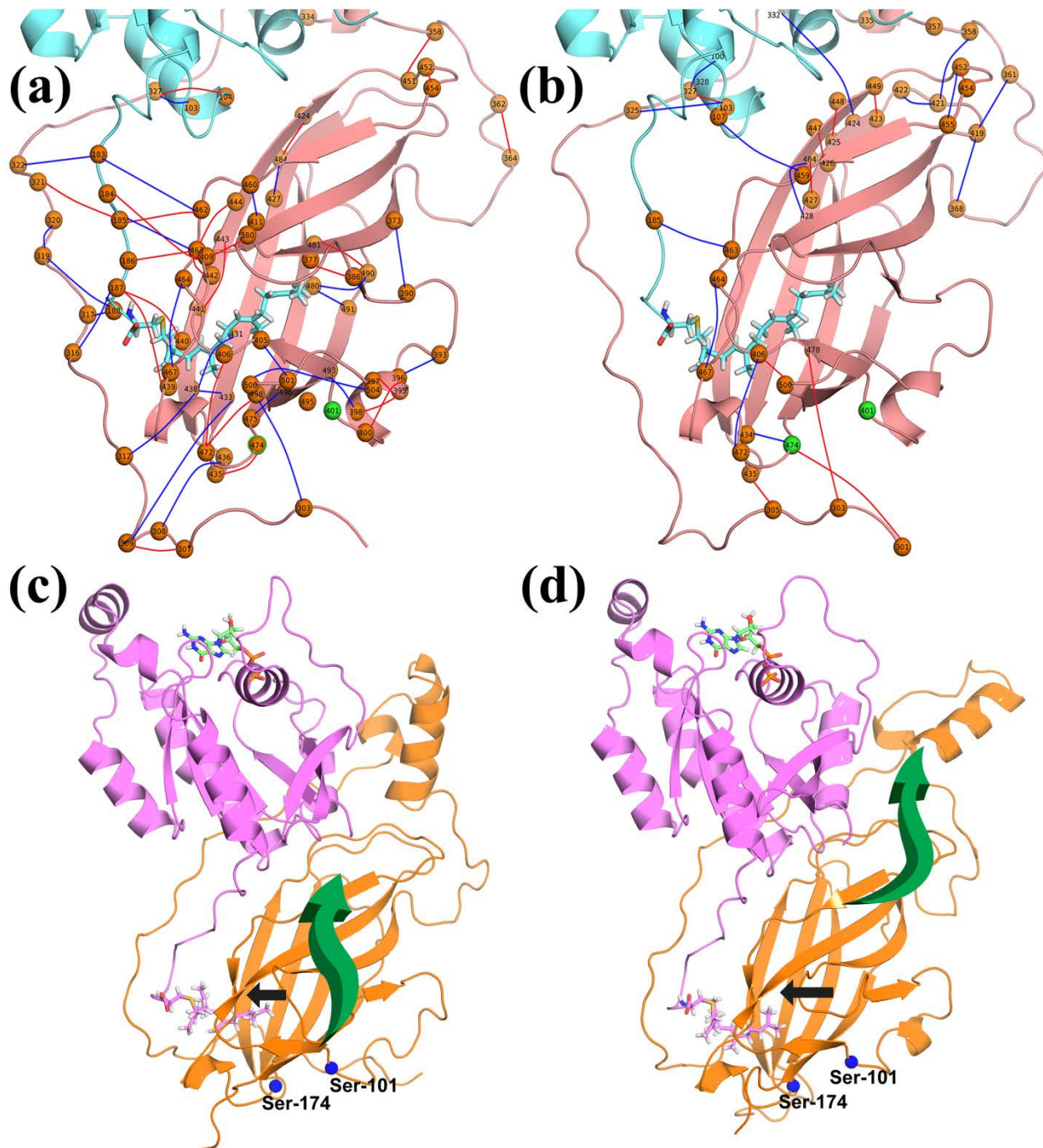
### Mechanistic model for the signal propagation in RhoGDI-GTPase complex

So far, we have established that phosphorylation results in the change in the electrostatic interaction energy of residues of Rac1 and GDI at key interaction sites. We show the destabilisation of the interactions between the prenyl moiety (Rac1) and the hydrophobic cavity (GDI) and, the switch regions of Rac1 and helix-turn-helix (HTH) region of GDI upon phosphorylation. From our



### 6.3 Results

structural and energetic analysis, our results suggest that long range effect exists between the site of the phosphorylation and the breaking of favourable interactions between the switch regions of Rac1 and HTH regions of GDI.



**Figure 6.9** Mechanical model to exhibit the multiple steps involved in the signal propagation

We have used hydrogen bond occupancy percentage to quantify the effect of phosphorylation on the system in terms of hydrogen bonding network. The net H-bond occupancy (%) is calculated as the difference of the H-bond occupancy between different parts of the trajectory as mentioned in our earlier analysis (e.g. SPP2(450-800ns)–SPP1(100-350ns), SPP3(800-1000ns)–SPP2(450-800ns)). In Figs. 6.9a-b, H-bond with occupancy > 30% are shown as blue lines (positive values) and red lines (negative values) indicating formation and breaking of H-bonds respectively.

The effect of phosphorylation on the hydrogen bonding can be elucidated based on the formation of hydrogen bond between the residues of the polybasic region and N-terminal region (R187-316, K188-317, K188-319). This change is accompanied with the breaking of interaction with the cavity residues (R187-440, R187-439, 184-409, R185-462, K186-463). We also observed an increase in the number of H-bond formation (blue lines) within GDI upon phosphorylation. Interestingly, a comparison between Figs. 6.9a-b shows an upward shift in the H-bond pattern exhibiting the time evolution of perturbation in the form of H-bond network. Moreover, in Fig. 6.8, we identified the hydrogen bonds within GDI residues that control the drawbridge-like motion. This induced rigidity in the GDI may result in the dissociation of the complex. In Figs. 6.9c-d, we suggest a two-step mechanical model for the propagation of the signal from the phosphorylation site to the HTH region. In the first step, the effect of phosphorylation results in the localised rearrangements of H-bonds and destabilization of the prenyl moiety. In the second step, the perturbation propagates to the distal region switch region and HTH interaction.

### **Mechanistic model based on conformational free energy landscape**

From our MD simulation studies, it is established that the effect of phosphorylation is mainly dominated as differential interaction of polybasic region of Rac1 between the cavity at the C-terminal domain and N-terminal residues of GDI in wild type and the phosphorylated respectively. According to the hydrogen bonded network and the minimum distance distribution, it is evident that the interactions of PBR (Rac1) between these two regions break and form intermittently, and involve an energy barrier associated with the transition process. With the limitation in standard MD simulations, we have used metadynamics as an enhanced sampling technique to overcome the free energy barrier during the transition process. Considering that the PBR of Rac1 forms greater number of H-bonds with the cavity (GDI) as compared to the N-terminal region (GDI) in the wild type whereas vice-versa occurs upon serine phosphorylation.

### 6.3 Results

---

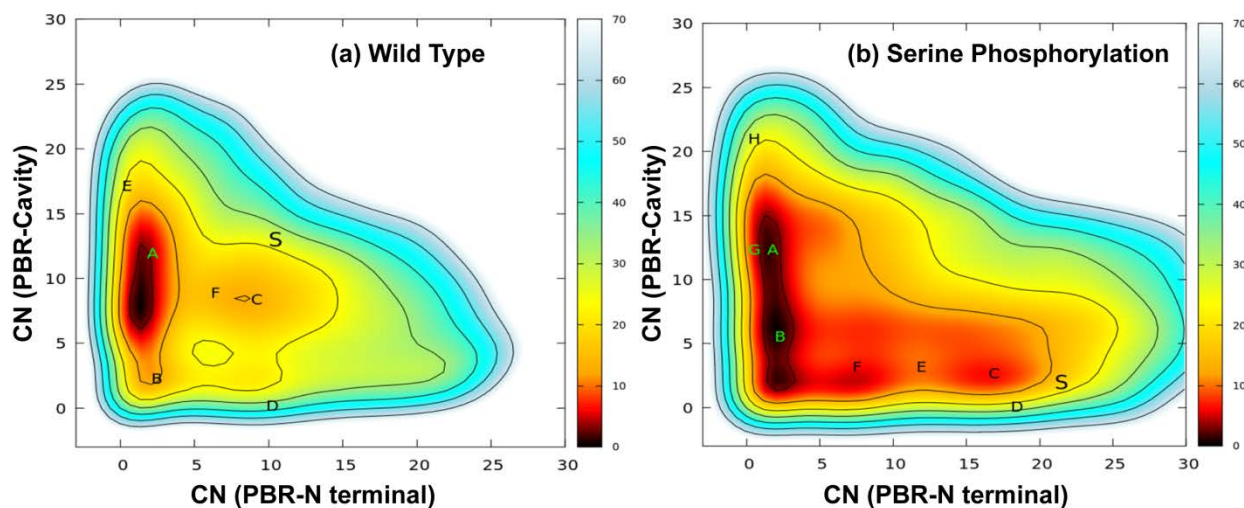
We have used coordination number (*CN*) based on the distance criteria of hydrogen bond as the collective variable to explore the conformational space of the PBR region and identify energetically stable and metastable states in the wild type and phosphorylated systems.

We employ two CVs (1) *CN*(PBR-Cavity) and (2) *CN*(PBR-N terminal) to identify the conformations of PBR with two different regions of GDI in the wild type and phosphorylated systems. The polybasic region in Rac1 consists of residues from position 183-189, which includes the modified geranylgeranylated cysteine. Based on hydrogen bond analysis from unbiased MD simulation, we have identified a list of residues from the N-terminal (GDI) and cavity (GDI) that undergo H-bonds formation and breaking with the PBR (Rac1). The N-terminal region includes residues from position 315-321, and the residues from cavity region are E409, D440, E463, E464 and K467. The CVs are calculated for only atoms that are involved in hydrogen bonding, i.e. donor, acceptor and hydrogen atom for the above-mentioned residues. According to the definition, greater value for *CN* would indicate a higher number of possible hydrogen bonds between the PBR(Rac1) and the cavity/N-terminal region. Further, one must note that the coordination number does not denote the exact total number of H-bonds formed in the interacting regions. This is because the angle criteria is not included in the collective variable definition and each residue has donor, acceptor and hydrogen atom (involved in H-bond) from the main chain and the sidechain (if any).

Fig. 6.10 shows the two-dimensional representation of the conformational free energy surface of the PBR (Rac1) with respect to the interactions with N-terminal/cavity region of GDI. A comparative analysis of the maps between the wild type and phosphorylated state suggests the enhanced interaction of the PBR with the N-terminal regions upon phosphorylation. In the wild type Rac1-GDI complex, the polybasic region strongly interacts with cavity residues of GDI as indicated by the presence of single deep minima. On the other hand, the free energy surface for the phosphorylated system is characterised by the presence of multiple metastable states. This reveals that upon phosphorylation, the interaction of the polybasic region with the cavity weakens, followed by the enhanced interactions with N-terminal region. The phosphorylated metadynamics run starts with an equilibrated structure with higher interaction of PBR with N-terminal (GDI) denoted by label “S” in the FES. The FES in the phosphorylated state shows



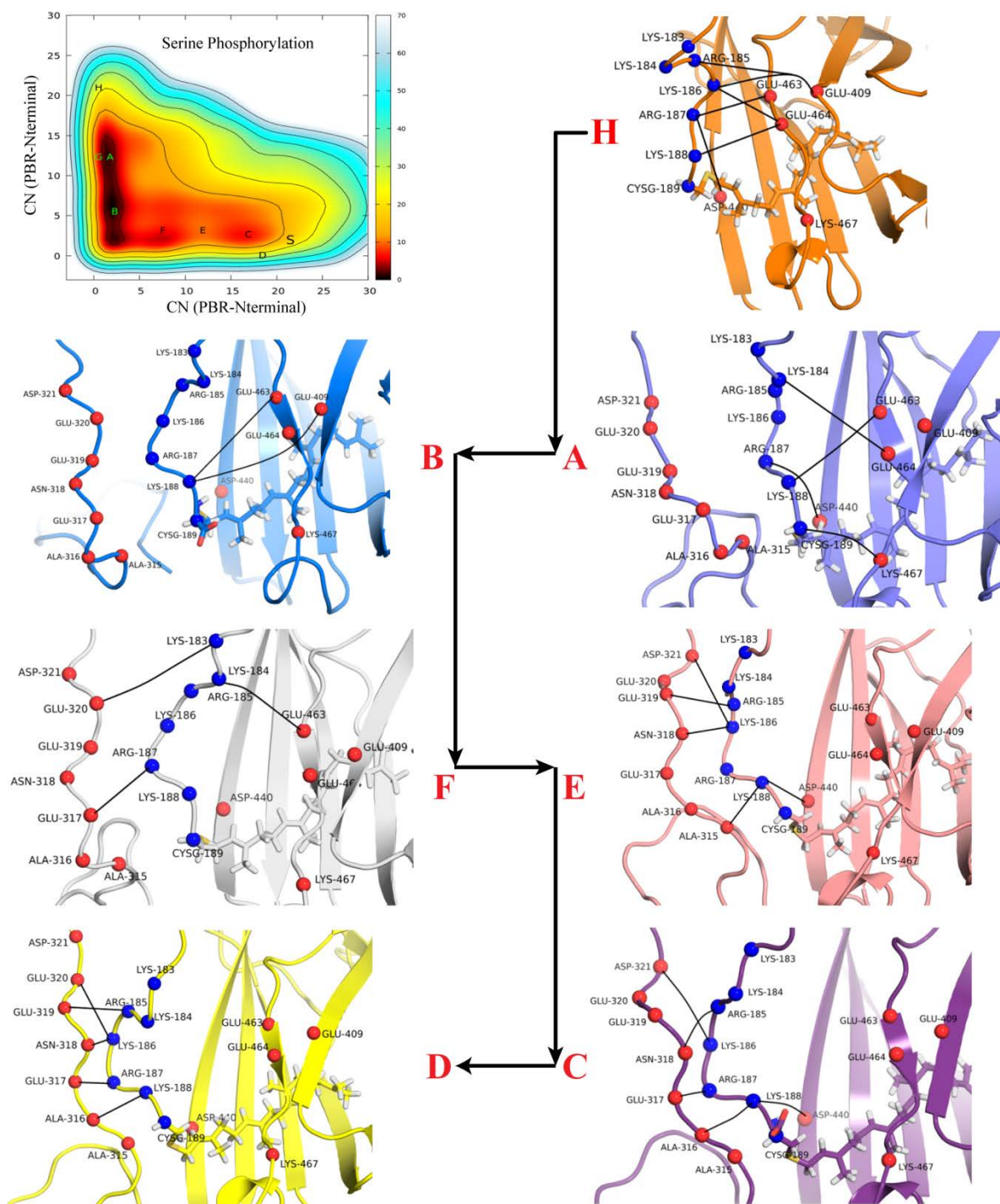
broad deep minima for the interaction between Rac1 and cavity residues with multiple conformational states.



**Figure 6.10** Free energy surfaces corresponding to (a) wild type (b) Phosphorylated state. The higher value for Y-axis and lower value for X-axis indicate higher interactions between PBR-cavity and lower interaction with N-terminal, and vice-versa. The metastable states are marked on individual plots. The position of the starting structure for each system is labelled as “S” on the free energy surface.

Among these states, a region of minima overlaps with the position of minima in the wild type FES (labelled as “A” in both FES). Thus, the phosphorylated state is capable of visiting the wild type PBR interactions and associated conformations. The FES also shows minima where coordination number for both the CVs approaches close to the value of 1 or 2. This signifies that there is a possibility where the polybasic region can break interactions from the cavity as well as from the N-terminal region. Interestingly, the FES shows interactions between PBR and the N-terminal as shallow minima of transitional states that are thermally accessible. This unveils the effect of phosphorylation in the form of an increase in the coordination number (interaction) between PBR (Rac1) and N-terminal (GDI) which is absent in the wild type Rac1-GDI complex. This metastable state labelled as “C” in the phosphorylated FES has slightly higher free energy than minima “A” and “B” and separated by small free energy barrier ( $\sim 2.5$  kcal/mol).

## 6.3 Results



**Figure 6.11** Mechanistic model for the transition of the interaction of PBR (Rac1) from the cavity to N-terminal (GDI) upon phosphorylation. The free energy surface is labelled with different metastable states that correspond to the transition states of interactions of the polybasic region from the cavity (GDI) to N-terminal residues (GDI). The representative structure from each minima is shown in the figure. The transition path is shown as (

$H \rightarrow A \rightarrow B \rightarrow F \rightarrow E \rightarrow C \rightarrow D$ ). The blue spheres represent residues of PBR whereas red spheres are residues from cavity and N-terminal. The H-bond interactions in each representative structure are marked with black lines.

The polybasic region in Rho GTPases is highly flexible and plays an important role in the specificity and selectivity in the formation of the Rho GTPase and GDI complex. Crystallographic studies and kinetic assays show that the modifications (mutation or post-translational modifications) in the polybasic region are closely related to the Rho GTPase and GDI association or dissociation. From our MD studies, we demonstrate the important role of interactions involving PBR (Rac1), N-terminal (GDI) and cavity (GDI) regions in the wild type and phosphorylated state. Based on our metadynamics simulations, we generate a mechanistic model for the transition of interactions of the polybasic region from the cavity to the N-terminal residues of GDI upon phosphorylation (Fig. 6.11). In fig 6.10b, multiple metastable states are labelled as *A*, *B*, *C* and *F*. The labelled “*A*” and “*G*” corresponds to the same minima. The label “*H*” and “*D*” represents conformational states with the interaction of PBR exclusively with cavity and N-terminal, respectively.

Fig. 6.11 explains the mechanism in terms of hydrogen bonds forming and breaking during the conformational transition. The conformations represent the various labelled free energy state in FES. The global minima marked *A/G* represents state which overlaps with the wild type FES (Fig. 6.10a) and shows interactions with the cavity residues. Interestingly, the carboxy-terminal of geranylgeranylated cysteine residue (CYSG189) forms a hydrogen bond with K467 of GDI cavity. This hydrogen bond can act as an anchor to firmly hold the prenyl moiety inside the cavity. This is supported with the interactions of other PBR residues with the cavity region, as shown in the figure. This is followed by the reduction in the number of interactions (conformation labelled as “*B*”) accompanied by the conformation change in the polybasic region. The conformational states marked as *F*, *E*, *C* shows the transition with the breaking of H-bonds with cavity region and forming with the N-terminal residues. The transition from states “*F*” to “*C*” is shown through the conformation labelled “*E*” on FES which is defined with multiple transient interactions (for example; K188-D440, R185-E319, K186-N318, K186-D321, K188-A315). The state “*C*” involves the rearrangement of H-bonds between the PBR and N-terminal residues (R185-N318, K186-D321, R187-E317, K188-A316). Combining the

## 6.4 Discussion

---

conformational reorganisation of the polybasic regions and associated interactions from the metadynamics study, we propose that as a consequence of serine phosphorylation in Rac1, the increased interaction of PBR with the N-terminal in the form of H-bonds generates a pulling effect that forces the outward movement of geranylgeranyl group from the cavity. In addition, the prenyl moiety is tightly bound into the cavity with the favourable interactions with hydrophobic residues. Evidently, we find a range of interactions (electrostatic (hydrogen bonds, salt bridges) and van der Waals) that stabilize the C-terminal residues of Rac1 with the immunoglobulin-like C-terminal domain of GDI and we speculate that the complete removal of this moiety and dissociation of the polybasic region from the cavity would require crossing a high free energy barrier described by multiple reaction coordinates.

## 6.4 Discussion

The regulation mechanism of the interaction between Rho GTPases and GDIs remains still elusive and is considered as a complex process. *In vitro* and *in vivo* studies suggest p21-activated kinase (PAK1) mediated phosphorylation of GDI leads to the selective release of Rac1 for its further downstream activation. We performed molecular dynamics simulation studies of the wild type and phosphorylated state of Rac1-GDI complex to understand the molecular basis of dissociation of the complex upon GDI phosphorylation. Our analyses based on structural differences and motions exhibit destabilisation between key interacting regions of the complex upon phosphorylation. This is evident in the form of intermittent breaking and forming of interactions between these regions in the wild type and phosphorylated state. Using PCA technique, we show large scale correlated motions in the GDI, HTH regulatory arm and the C-terminal domain with the interconnecting loop between the N-terminal domain and C-terminal domain of GDI. Further, we show destabilisation of geranylgeranyl group in the cavity and the HTH-switch I/II interactions. However, it is unclear how the phosphorylation site and the interacting regions of Rac1-GDI complex are connected as the structural and correlated motions do not provide insights into the molecular details of this process.

It is expected that the addition of negative charge upon phosphorylation would affect the overall electrostatics of the system and may induce structural changes in the Rac1-GDI complex. In chapter 3 and 4 of this thesis, we highlight the importance of electrostatic interaction energy as the key determinant in capturing the structural modulation in the system. In accordance with

this hypothesis, we observed extensive rearrangements of interactions between the polybasic region of Rac1 and the N-terminal region or hydrophobic cavity of GDI as a consequence of the change in pairwise electrostatic interaction energy. The resultant energetic perturbation among residues destabilises the geranylgeranyl group in the cavity, thereby promoting dissociation of the complex. Additionally, we show a long range effect of this charge perturbation in terms of change in pairwise electrostatic interaction energy of residues between the switch I/II region and HTH (N-terminal domain) of GDI.

Our analysis on energetic perturbation reveals an increase in interactions between the N-terminal of GDI and polybasic region of Rac1 upon phosphorylation. We speculate that strong interactions with the N-terminal would generate an outward force that would eventually pull the prenyl moiety out of the cavity. Interestingly, we identify specific interactions (V361-N419, R358-E421, Q332-S424) that are coupled to the drawbridge-like motion of the helix-turn-helix region. Moreover, we show this as population shift in the hydrogen bond distribution of these specific interactions. Based on a detailed hydrogen bond occupancy analysis, we propose a mechanistic model for the signal propagation from the site of phosphorylation to the helix-turn-helix (regulatory arm) region of RhoGDI in the form of rearrangements of hydrogen bonds and charge-charge interactions.

Here we also performed metadynamics simulations to capture a series of intermediate states between the wild state and the phosphorylated state of Rac1-GDI complex. We have demonstrated that the polybasic region interacts preferably with the cavity residues with minimum interaction with the N-terminal region in wild type. However, upon phosphorylation, the polybasic region acquires enhanced flexibility generating multiple metastable states with increased interaction with the N-terminal region. During this shifting of interaction from cavity residues to N-terminal, we have captured a series of intermediate residues that correspond to the metastable states according to the free energy landscape. It is speculated that the inherent flexibility of the N-terminal domain plays an important role in overcoming the energetic barriers associated with the complex formation and these structural differences may contribute to the specificity and selectivity between Rho proteins and GDI interactions.

## 6.5 Bibliography

---

### 6.5 Bibliography

1. Hall, A., *Rho family GTPases*. Biochemical Society Transactions, 2012. 40(6): p. 1378.
2. Etienne-Manneville, S. and A. Hall, *Rho GTPases in cell biology*. Nature, 2002. 420(6916): p. 629-635.
3. DerMardirossian, C. and G.M. Bokoch, *GDI: central regulatory molecules in Rho GTPase activation*. Trends Cell Biol, 2005. 15(7): p. 356-63.
4. Falkenberg, C.V. and L.M. Loew, *Computational analysis of Rho GTPase cycling*. PLoS Comput Biol, 2013. 9(1): p. e1002831.
5. Dovas, A. and J.R. Couchman, *RhoGDI: multiple functions in the regulation of Rho family GTPase activities*. Biochem J, 2005. 390(Pt 1): p. 1-9.
6. Garcia-Mata, R., E. Boulter, and K. Burridge, *The 'invisible hand': regulation of RHO GTPases by RHOGDIs*. Nat Rev Mol Cell Biol, 2011. 12(8): p. 493-504.
7. Wittinghofer, A. and I.R. Vetter, *Structure-function relationships of the G domain, a canonical switch motif*. Annu Rev Biochem, 2011. 80: p. 943-71.
8. Jaffe, A.B. and A. Hall, *Rho GTPases: biochemistry and biology*. Annu Rev Cell Dev Biol, 2005. 21: p. 247-69.
9. Sit, S.-T. and E. Manser, *Rho GTPases and their role in organizing the actin cytoskeleton*. Journal of cell science, 2011. 124: p. 679-683.
10. Tnimov, Z., et al., *Quantitative analysis of prenylated RhoA interaction with its chaperone, RhoGDI*. J Biol Chem, 2012. 287(32): p. 26549-62.
11. Muratcioglu, S., et al., *PDEdelta Binding to Ras Isoforms Provides a Route to Proper Membrane Localization*. J Phys Chem B, 2017. 121(24): p. 5917-5927.
12. Cox, A.D. and C.J. Der, *Protein prenylation: more than just glue?* Current Opinion in Cell Biology, 1992. 4(6): p. 1008-1016.
13. Cushman, I. and P.J. Casey, *RHO methylation matters: a role for isoprenylcysteine carboxylmethyltransferase in cell migration and adhesion*. Cell adhesion & migration, 2011. 5(1): p. 11-15.
14. Casey, P.J. and M.C. Seabra, *Protein Prenyltransferases*. Journal of Biological Chemistry, 1996. 271(10): p. 5289-5292.
15. Winter-Vann, A.M. and P.J. Casey, *Post-prenylation-processing enzymes as new targets in oncogenesis*. Nature Reviews Cancer, 2005. 5(5): p. 405-412.
16. Olson, M.F., *Rho GTPases, their post-translational modifications, disease-associated mutations and pharmacological inhibitors*. Small GTPases, 2018. 9(3): p. 203-215.
17. Michaelson, D., et al., *Postprenylation CAAX processing is required for proper localization of Ras but not Rho GTPases*. Mol Biol Cell, 2005. 16(4): p. 1606-16.

18. Wong, K.W., S. Mohammadi, and R.R. Isberg, *Disruption of RhoGDI and RhoA regulation by a Rac1 specificity switch mutant*. J Biol Chem, 2006. 281(52): p. 40379-88.
19. Boulter, E., et al., *Regulation of Rho GTPase crosstalk, degradation and activity by RhoGDII*. Nature cell biology, 2010. 12(5): p. 477-483.
20. Ren, X.D., W.B. Kiosses, and M.A. Schwartz, *Regulation of the small GTP-binding protein Rho by cell adhesion and the cytoskeleton*. The EMBO journal, 1999. 18(3): p. 578-585.
21. Newcombe, A.R., et al., *The interaction between rac1 and its guanine nucleotide dissociation inhibitor (GDI), monitored by a single fluorescent coumarin attached to GDI*. Biochemistry, 1999. 38(21): p. 6879-86.
22. Robbe, K., et al., *Dissociation of GDP dissociation inhibitor and membrane translocation are required for efficient activation of Rac by the Dbl homology-pleckstrin homology region of Tiam*. J Biol Chem, 2003. 278(7): p. 4756-62.
23. Scheffzek, K., et al., *The Rac–RhoGDI complex and the structural basis for the regulation of Rho proteins by RhoGDI*. Nature Structural Biology, 2000. 7(2): p. 122-126.
24. Hoffman, G.R., N. Nassar, and R.A. Cerione, *Structure of the Rho Family GTP-Binding Protein Cdc42 in Complex with the Multifunctional Regulator RhoGDI*. Cell, 2000. 100(3): p. 345-356.
25. Grizot, S., et al., *Crystal Structure of the Rac1–RhoGDI Complex Involved in NADPH Oxidase Activation*. Biochemistry, 2001. 40(34): p. 10007-10013.
26. Li, R. and Y. Zheng, *Residues of the Rho Family GTPases Rho and Cdc42 That Specify Sensitivity to Dbl-like Guanine Nucleotide Exchange Factors*. Journal of Biological Chemistry, 1997. 272(8): p. 4671-4679.
27. Kuhlmann, N., et al., *Structural and Mechanistic Insights into the Regulation of the Fundamental Rho Regulator RhoGDIalpha by Lysine Acetylation*. J Biol Chem, 2016. 291(11): p. 5484-99.
28. Kuhlmann, N., et al., *RhoGDIalpha Acetylation at K127 and K141 Affects Binding toward Nonprenylated RhoA*. Biochemistry, 2016. 55(2): p. 304-12.
29. DerMardirossian, C., A. Schnelzer, and G.M. Bokoch, *Phosphorylation of RhoGDI by Pak1 mediates dissociation of Rac GTPase*. Mol Cell, 2004. 15(1): p. 117-27.
30. Knezevic, N., et al., *GDI-1 phosphorylation switch at serine 96 induces RhoA activation and increased endothelial permeability*. Mol Cell Biol, 2007. 27(18): p. 6323-33.
31. Dovas, A., et al., *Serine 34 Phosphorylation of Rho Guanine Dissociation Inhibitor (RhoGDIα) Links Signaling from Conventional Protein Kinase C to RhoGTPase in Cell Adhesion*. Journal of Biological Chemistry, 2010. 285(30): p. 23296-23308.

## 6.5 Bibliography

---

32. Mehta, D., A. Rahman, and A.B. Malik, *Protein Kinase C- $\alpha$  Signals Rho-Guanine Nucleotide Dissociation Inhibitor Phosphorylation and Rho Activation and Regulates the Endothelial Cell Barrier Function*. *Journal of Biological Chemistry*, 2001. 276(25): p. 22614-22620.
33. Guerrero, I.C., N.H. Keep, and J. Godovac-Zimmermann, *Proteomics Study Reveals Cross-Talk between Rho Guanidine Nucleotide Dissociation Inhibitor 1 Post-Translational Modifications in Epidermal Growth Factor Stimulated Fibroblasts*. *Journal of Proteome Research*, 2007. 6(7): p. 2623-2630.
34. DerMardirossian, C., et al., *Phosphorylation of RhoGDI by Src regulates Rho GTPase binding and cytosol-membrane cycling*. *Molecular biology of the cell*, 2006. 17(11): p. 4760-4768.
35. Wu, Y., et al., *Src phosphorylation of RhoGDI2 regulates its metastasis suppressor function*. *Proceedings of the National Academy of Sciences of the United States of America*, 2009. 106(14): p. 5807-5812.
36. Fei, F., et al., *The Fer tyrosine kinase regulates interactions of Rho GDP-Dissociation Inhibitor  $\alpha$  with the small GTPase Rac*. *BMC biochemistry*, 2010. 11: p. 48-48.
37. Kuribayashi, K., et al., *Essential role of protein kinase C zeta in transducing a motility signal induced by superoxide and a chemotactic peptide, fMLP*. *The Journal of cell biology*, 2007. 176(7): p. 1049-1060.
38. Qiao, J., et al., *Phosphorylation of GTP dissociation inhibitor by PKA negatively regulates RhoA*. *American Journal of Physiology-Cell Physiology*, 2008. 295(5): p. C1161-C1168.
39. Scheffzek, K., M.R. Ahmadian, and A. Wittinghofer, *GTPase-activating proteins: helping hands to complement an active site*. *Trends in Biochemical Sciences*, 1998. 23(7): p. 257-262.
40. Golovanov, A.P., et al., *Structure-activity relationships in flexible protein domains: regulation of rho GTPases by RhoGDI and D4 GDII*. *Journal of Molecular Biology*, 2001. 305(1): p. 121-135.
41. Lian, L.-Y., et al., *Mapping the binding site for the GTP-binding protein Rac-1 on its inhibitor RhoGDI-1*. *Structure*, 2000. 8(1): p. 47-56.
42. Gosser, Y.Q., et al., *C-terminal binding domain of Rho GDP-dissociation inhibitor directs N-terminal inhibitory peptide to GTPases*. *Nature*, 1997. 387(6635): p. 814-819.
43. Ueyama, T., et al., *Negative Charges in the Flexible N-Terminal Domain of Rho GDP-Dissociation Inhibitors (RhoGDIs) Regulate the Targeting of the RhoGDI-Rac1 Complex to Membranes*. *The Journal of Immunology*, 2013. 191(5): p. 2560.



44. Keep, N.H., et al., *A modulator of rho family G proteins, rhoGDI, binds these G proteins via an immunoglobulin-like domain and a flexible N-terminal arm*. Structure (London, England : 1993), 1997. 5(5): p. 623-633.
45. Golovanov, A.P., et al., *Structural consequences of site-directed mutagenesis in flexible protein domains*. European Journal of Biochemistry, 2001. 268(8): p. 2253-2260.
46. Fiser, A., R.K. Do, and A. Sali, *Modeling of loops in protein structures*. Protein science : a publication of the Protein Society, 2000. 9(9): p. 1753-1773.
47. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. 1-2: p. 19-25.
48. Huang, J. and A.D. MacKerell Jr, *CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data*. Journal of Computational Chemistry, 2013. 34(25): p. 2135-2145.
49. Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields*. Journal of Computational Chemistry, 2010. 31(4): p. 671-690.
50. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. The Journal of Chemical Physics, 1983. 79(2): p. 926-935.
51. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. J Chem Phys, 2007. 126(1): p. 014101.
52. Parrinello, M. and A. Rahman, *Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method*. Journal of Applied Physics, 1981. 52(12): p. 7182-7190.
53. Essmann, U., et al., *A Smooth Particle Mesh Ewald Method*. Journal of Chemical Physics, 1995. 103(19): p. 8577-8593.
54. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations*. Journal of Computational Chemistry, 1997. 18(12): p. 1463-1472.
55. Barducci, A., G. Bussi, and M. Parrinello, *Well-tempered metadynamics: A smoothly converging and tunable free-energy method*. Physical Review Letters, 2008. 100(2): p. 1-4.
56. Tribello, G.A., et al., *PLUMED 2: New feathers for an old bird*. Computer Physics Communications, 2014. 185(2): p. 604-613.



# Chapter 7

## Conclusions and Future Outlook

“The woods are lovely, dark and deep,  
But I have promises to keep,  
And miles to go before I sleep,  
And miles to go before I sleep.”

-Robert Frost, *Stopping by Woods on a Snowy Evening*, 1923

Cellular signaling involves a myriad of complex interactions between large number of proteins, lipids, small chemical compounds, ions and water (the universal solvent)<sup>1-4</sup>. Transmembrane receptors (GPCRs, RTKs, ion channels) and their intracellular partners (regulators and effectors) form a standard model of signaling that involves extensive protein-protein interactions between a diverse class of proteins<sup>5,6</sup>. However, how different types of interactions control these cellular processes is yet to be well addressed. The work carried out in this thesis provides a molecular understanding of the interactions that drive allosteric regulation and molecular recognition in signaling proteins.

### 7.1 Decoding Dynamic Allostery in PDZ3 Domain Using Electrostatic Interactions

Allostery is a unique property that allows signal propagation between two distantly located sites in a protein. Understanding allosteric mechanism at the molecular level has been an inspiring quest from decades. It has been more fascinating with Cooper and Dryden adding another dimension of “dynamic allostery” (allostery without macromolecular conformational change) to the existing paradigm of “structure-based allostery”<sup>7</sup>. So far, most of the computational methods used to describe allosteric mechanisms are structure or feature-based approaches that take into account well-defined structures/information from the experimental techniques and attempts to describe the allosteric networks/pathway between the distinct protein regions/sites<sup>8-10</sup>. In the dynamics based methods, variants of MD simulations can be used to enhance conformation sampling and pathway prediction accuracy<sup>11-13</sup>. Though these computational methods are

## 7.1 Decoding Dynamic Allostery in PDZ3 Domain Using Electrostatic Interactions

---

advanced and are modified appropriately for different systems, they fail to provide a physicochemical picture of what drives allostery.

PDZ domains are well known, paradigmatic single domain proteins that show dynamic allostery, where distal side-chain dynamics is modulated on ligand binding. Unlike large conformational changes that are driven by enthalpy, the origin of dynamic allostery in PDZ domain has been attributed to entropic effects<sup>14</sup>. In Chapter 3, we demonstrate that even in dynamic allostery, enthalpy plays an important role. We unearth the energetic basis of the observed dynamic allostery in a PDZ3 domain protein using molecular dynamics simulations. We demonstrate that electrostatic interaction provides a highly sensitive yardstick to probe the allosteric modulation in contrast to the traditionally used structure-based parameters. The ligand creates a local energetic perturbation at the recognition site ( $\beta 2 - \alpha 2$ ) that propagates in the form of domino-like changes in inter-residue interaction pattern towards the N-terminal and  $\alpha 1 - \beta 4$  region (distant region of protein). Albeit without large conformational change, there are significant changes in the nature of specific interactions (nonpolar/ polar) between inter-residue contacts and accompanied side-chain reorientations that drive the major redistribution of energy. Interestingly, this internal redistribution and rewiring of side-chain interactions led to large cancellations resulting in small change in the overall enthalpy of the protein, thus making it difficult to detect experimentally. In contrast to the prevailing focus on the entropic or dynamic effects, we show that the internal redistribution and population shift in specific electrostatic interactions drive the allosteric modulation in the PDZ3 domain protein.

Over the last 60 years, the definition of allostery has emerged from exclusive ligand binding event to more like a quantitative description of structural and dynamical changes in the biological macromolecules as a consequence of range of perturbation events such as mutations, covalent modifications and changes in the cellular physiological conditions<sup>12</sup>. In chapter 4, we extend our investigation to identify signatures of pH-dependent dynamic allostery in PDZ3 domain. We observe that protonation of histidine residues in PDZ3 domain as a consequence of physiological pH changes in the cellular environment modulate the conformational dynamics and the energy distribution in the domain. Further, we demonstrate how introduction of a charge on a histidine residue results in subtle changes in the population distribution of pairwise interactions in the network. Based on our findings, it appears that there exist multiple interactions based

universal network in PDZ3 domain. These interactions could be perturbed to different extents depending upon the perturbation factor (e.g. ligand binding, change in pH) and the allosteric modulation can be observed in terms of redistribution of internal electrostatic interaction energy through shifting of hydrogen bonds between residues.

So far, we have shown that in PDZ3 domain the signal propagation from the ligand binding site to the distant N-terminal region of the domain occurs through the  $\alpha$ 3-helix. Surprisingly, previous studies highlight similar dynamic allostery phenomenon in another PDZ domain i.e. second PDZ (PDZ2) from human tyrosine phosphates (hPTP1E) which lacks the presence of  $\alpha$ 3-helix. Thus, it will be interesting to uncover the molecular basis of dynamic allostery in PDZ2 and understand how signal propagation occurs from the ligand binding site to the distal regions of protein in the absence of  $\alpha$ 3-helix. Moreover, understanding allosteric mechanism in proteins is important for drug discovery and protein engineering. The knowledge of allostery provides an alternative means of regulating protein function by engineering a novel allosteric site. An improved accuracy in the prediction of allosteric sites may provide additional opportunities to design allosteric site targeted drugs for the modulation of pathogenic allosteric signaling.

### 7.2 Molecular Insights into Rho GTPase Activation and GDI-mediated Regulation

One of the most important families of intracellular proteins regulating a myriad of cellular processes is the Rho family GTPases that belong to the Ras superfamily of small GTPases<sup>15</sup>. Rho GTPases exist as conformational switches in alternate *on* (GTP-bound) and *off* (GDP-bound) states that are known to regulate the duration and intensity of the intracellular signal propagation. Once activated and membrane-localised, Rho GTPases can interact with more than 70 downstream effectors<sup>16,17</sup>. Previous experimental evidences highlight the stable association between switch regions of Rho proteins and the effectors<sup>18,19</sup>.

Structural studies on H-Ras, a member of Ras superfamily suggests multiple interconvertible conformations of switch I region in the nucleotide-bound form<sup>20-23</sup>. However, unlike Ras family of GTPases, for the Rho GTPases, there is no clear evidence for the existence of “sub-states” such as state 1 & state 2 in the GTP bound form. In chapter 5, we have explored the nucleotide dependent conformational space of the Switch I loop using atomistic molecular

## 7.2 Molecular Insights into Rho GTPase Activation and GDI-mediated Regulation

---

dynamics and metadynamics simulations on RhoA. These studies demonstrate that both the nucleotide-free state and the GDP bound “OFF” state have very similar conformations, whereas the GTP bound “ON” state has unique conformations with signatures of two intermediate states. We show that the conformational free energy landscape for these systems suggest the presence of multiple intermediate states. Interestingly, the energetic penalty of exposing the non-polar residues in the GTP bound form is counterbalanced by the favourable hydrogen bonded interactions between the  $\gamma$ -phosphate group of GTP with the highly conserved Tyr34 and Thr37 residues. These competing molecular interactions lead to a tuneable energy landscape of the Switch I conformation, which can undergo significant changes based on the local environment, including changes upon binding to effectors.

The remarkable ability of Rho GTPases to interact with a large number of effectors put them into the central role in coordinating distinct signaling pathways that regulate principally actin cytoskeletal rearrangements<sup>24-26</sup> and other various cellular processes<sup>27</sup>. Aberrations in these pathways lead to the development of diseases<sup>28-30</sup> (cancers, autoimmune disorders, etc.) which necessitate the stringent regulation of Rho GTPases at various levels. RhoGDI acts as down-regulator in the GTPase cycling with its ability to prevent GTPase activation and membrane localization<sup>31,32</sup>. These proteins are known to extract Rho GTPases from the membrane in GDP bound inactive state and solubilise them in the cytosol. Thus, further activation of GTPase would require the dissociation Rho GTPase-GDI complex.

Biochemical studies suggest kinase-mediated selective phosphorylation of RhoGDI to promote the release of GTPase for activation through GEFs<sup>33,34</sup>. In chapter 6, we investigate the molecular basis of dissociation of the complex upon GDI phosphorylation through a comparative analysis between the wild-type and phosphorylated state of Rac1-GDI<sup>33</sup>. We demonstrate destabilisation of key interacting regions of Rac1-GDI complex as an effect of phosphorylation of serine residues at positions 101 and 174 in GDI. In the previous work on PDZ3 domain, we demonstrated that the electrostatic interaction energy acts as a key determinant in probing even minor conformational changes as compared to the structure-based parameters. The energetic perturbation analysis in GTPase-GDI complex reveals change in electrostatic interaction energies between Rac1 and GDI residues as a consequence of phosphorylation. Interestingly, phosphorylation induces an increase in the interactions between the N-terminal region of GDI

with the polybasic region, suggesting a mechanistic approach for the extraction of prenyl moiety out of the cavity. The conformational free energy surface captures the transition of interactions of the polybasic regions from the GDI cavity to the GDI N-terminal region. Further, based on a detailed hydrogen bond occupancy analysis, we propose a mechanistic model for the signal propagation from the site of phosphorylation to the helix-turn-helix (regulatory arm) region of RhoGDI in the form of rearrangements of hydrogen bonds and charge-charge interactions.

Rho GTPase signaling pathways are widely associated with cell cycle progression and cell migration activity<sup>27,35</sup>. Mutations are less common in Rho GTPases as compared to the Ras where mutations occur in more than 30% cancers<sup>36,37</sup>. Rho GTPases are hyperactivated by mutations and interact with a unique set of effectors that contribute to tumorigenesis. Interestingly, no clinically effective drug targeting Rho GTPases for cancer treatment is available till date. Rho GTPase signaling is generally targeted by using nucleotide analogues, covalent modifications of Rho proteins (bacterial toxins) or by regulation of interactions with effectors and regulators (GAPs, GEFs, GDIs)<sup>38,39</sup>. Based on our findings, the exposed hydrophobic residues of switch-I region play an important role in the effector recognition in Rho GTPases. Further, it has been observed that targeting the regulators provide better selectivity and success as compared to targeting Rho GTPases directly. This could be useful for identification of novel drug pocket and exploit information from the protein-protein interface for optimization of drug molecules.

### 7.3 Bibliography

1. Narumiya, S. and D. Thumkeo, *Rho signaling research: history, current status and future directions*. FEBS letters, 2018. 592(11): p. 1763-1776.
2. Pawson, T. and J.D. Scott, *Signaling Through Scaffold, Anchoring, and Adaptor Proteins*. Science, 1997. 278(5346): p. 2075.
3. Tompa, P., *The principle of conformational signaling*. Chemical Society Reviews, 2016. 45(15): p. 4252-4284.
4. Syrovatkina, V., et al., *Regulation, Signaling, and Physiological Functions of G-Proteins*. Journal of molecular biology, 2016. 428(19): p. 3850-3868.
5. Hunter, T., *Why nature chose phosphate to modify proteins*. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2012. 367(1602): p. 2513-2516.

### 7.3 Bibliography

---

6. Lee, M.J. and M.B. Yaffe, *Protein Regulation in Signal Transduction*. Cold Spring Harbor Perspectives in Biology, 2016. 8(6).
7. Cooper, A. and D.T.F. Dryden, *Allostery without conformational change*. European Biophysics Journal, 1984. 11(2): p. 103-109.
8. Di Paola, L. and A. Giuliani, *Protein contact network topology: a natural language for allostery*. Curr Opin Struct Biol, 2015. 31: p. 43-8.
9. Grutsch, S., S. Brüsweiler, and M. Tollinger, *NMR Methods to Study Dynamic Allostery*. PLOS Computational Biology, 2016. 12(3): p. e1004620.
10. Liu, J. and R. Nussinov, *Allostery: An Overview of Its History, Concepts, Methods, and Applications*. PLoS Comput Biol, 2016. 12(6): p. e1004966.
11. Di Russo, N.V., M.A. Martí, and A.E. Roitberg, *Underlying Thermodynamics of pH-Dependent Allostery*. The Journal of Physical Chemistry B, 2014. 118(45): p. 12818-12826.
12. Motlagh, H.N., et al., *The ensemble nature of allostery*. Nature, 2014. 508(7496): p. 331-9.
13. Hertig, S., N.R. Latorraca, and R.O. Dror, *Revealing Atomic-Level Mechanisms of Protein Allostery with Molecular Dynamics Simulations*. PLOS Computational Biology, 2016. 12(6): p. e1004746.
14. Petit, C.M., et al., *Hidden dynamic allostery in a PDZ domain*. Proc Natl Acad Sci U S A, 2009. 106(43): p. 18249-54.
15. Hall, A., *Rho family GTPases*. Biochemical Society Transactions, 2012. 40(6): p. 1378.
16. Bishop, A.L. and A. Hall, *Rho GTPases and their effector proteins*. The Biochemical journal, 2000. 348 Pt 2(Pt 2): p. 241-255.
17. Nobes, C.D. and A. Hall, *Rho GTPases Control Polarity, Protrusion, and Adhesion during Cell Movement*. The Journal of Cell Biology, 1999. 144(6): p. 1235.
18. Morreale, A., et al., *Structure of Cdc42 bound to the GTPase binding domain of PAK*. Nature Structural Biology, 2000. 7(5): p. 384-388.
19. Lapouge, K., et al., *Structure of the TPR Domain of p67phox in Complex with Rac·GTP*. Molecular Cell, 2000. 6(4): p. 899-907.
20. Muraoka, S., et al., *Crystal structures of the state 1 conformations of the GTP-bound H-Ras protein and its oncogenic G12V and Q61L mutants*. FEBS Letters, 2012. 586(12): p. 1715-1718.
21. Phillips, M.J., et al., *Effector Proteins Exert an Important Influence on the Signaling-active State of the Small GTPase Cdc42*. Journal of Biological Chemistry, 2008. 283(20): p. 14153-14164.
22. Baussand, J. and J. Kleinjung, *Specific Conformational States of Ras GTPase upon Effector Binding*. J Chem Theory Comput, 2013. 9(1): p. 738-749.



23. Dias, S.M.G. and R.A. Cerione, *X-ray Crystal Structures Reveal Two Activated States for RhoC*. *Biochemistry*, 2007. 46(22): p. 6547-6558.
24. Hall, A., *Rho GTPases and the Actin Cytoskeleton*. *Science*, 1998. 279(5350): p. 509.
25. Sit, S.-T. and E. Manser, *Rho GTPases and their role in organizing the actin cytoskeleton*. *Journal of cell science*, 2011. 124: p. 679-683.
26. Sahai, E., A.S. Alberts, and R. Treisman, *RhoA effector mutants reveal distinct effector pathways for cytoskeletal reorganization, SRF activation and transformation*. *The EMBO Journal*, 1998. 17(5): p. 1350-1361.
27. Sadok, A. and C.J. Marshall, *Rho GTPases: masters of cell migration*. *Small GTPases*, 2014. 5: p. e29710-e29710.
28. Aguilar, B.J., Y. Zhu, and Q. Lu, *Rho GTPases as therapeutic targets in Alzheimer's disease*. *Alzheimer's research & therapy*, 2017. 9(1): p. 97-97.
29. Olson, M.F., *Rho GTPases, their post-translational modifications, disease-associated mutations and pharmacological inhibitors*. *Small GTPases*, 2018. 9(3): p. 203-215.
30. Sahai, E. and C.J. Marshall, *RHO-GTPases and cancer*. *Nat Rev Cancer*, 2002. 2(2): p. 133-42.
31. DerMardirossian, C. and G.M. Bokoch, *GDI: central regulatory molecules in Rho GTPase activation*. *Trends Cell Biol*, 2005. 15(7): p. 356-63.
32. Garcia-Mata, R., E. Boulter, and K. Burridge, *The 'invisible hand': regulation of RHO GTPases by RHOGDIs*. *Nat Rev Mol Cell Biol*, 2011. 12(8): p. 493-504.
33. DerMardirossian, C., A. Schnelzer, and G.M. Bokoch, *Phosphorylation of RhoGDI by Pak1 mediates dissociation of Rac GTPase*. *Mol Cell*, 2004. 15(1): p. 117-27.
34. Dovas, A. and J.R. Couchman, *RhoGDI: multiple functions in the regulation of Rho family GTPase activities*. *Biochem J*, 2005. 390(Pt 1): p. 1-9.
35. Parri, M. and P. Chiarugi, *Rac and Rho GTPases in cancer cell motility control*. *Cell Commun Signal*, 2010. 8: p. 23.
36. Porter, A.P., A. Papaioannou, and A. Malliri, *Deregulation of Rho GTPases in cancer*. *Small GTPases*, 2016. 7(3): p. 123-138.
37. Jansen, S., et al., *Paving the Rho in cancer metastasis: Rho GTPases and beyond*. *Pharmacology & Therapeutics*, 2018. 183: p. 1-21.
38. Lin, Y. and Y. Zheng, *Approaches of targeting Rho GTPases in cancer drug discovery*. *Expert opinion on drug discovery*, 2015. 10(9): p. 991-1010.
39. Harding, M.A. and D. Theodorescu, *RhoGDI signaling provides targets for cancer therapy*. *European Journal of Cancer*, 2010. 46(7): p. 1252-1259.

