

**PROTEIN CONFORMATIONAL DYNAMICS IN A TWO-COMPONENT
SIGNAL TRANSDUCTION SYSTEM**

By

Rahul Banerjee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry – Doctor of Philosophy
Quantitative Biology – Dual Major

2014

ABSTRACT

PROTEIN CONFORMATIONAL DYNAMICS IN A TWO-COMPONENT SIGNAL TRANSDUCTION SYSTEM

By

Rahul Banerjee

Two-component system (TCS) signal transduction is the predominant mechanism in bacteria to adapt to environmental changes. Protein conformational dynamics in various proteins (or individual domains) of this signaling system facilitate signal transduction from the exterior to the interior of the cell. A membrane bound histidine kinase (HK) is autophosphorylated at a conserved histidine residue, upon signal recognition in the extracellular domain. A crucial loop region in the ATP-binding catalytic domain (CA) of HK, namely, the ATP-lid loop, is implicated in the autophosphorylation reaction of the HK. Integral to the TCS pathways is a phosphotransfer reaction between HK and a cognate response regulator (RR) protein. The RR protein undergoes conformational changes due to phosphorylation of a conserved aspartate residue. The phosphorylated form of the RR carries out the response by interacting with its targets, typically some gene promoters for downstream regulation. The autophosphorylation and/or RR phosphatase activities of the HK control the level of RR phosphorylation and hence the output response. The key for understanding of the mechanisms of biological signal transduction is to elucidate the conformational dynamics of its signaling proteins, as the activation of a signaling protein is fundamentally a process of conformational transition from an inactive to an active state. Activation of a response regulator protein has been investigated using RR468 as a model system using equilibrium and non equilibrium simulation techniques. Conformational dynamics in the CA domain (of HK) and RR protein are studied here using various computational tools to characterize intermediate states in the process of conformational transition. A protein is likely to

have intramolecular interactions that are not present in the experimentally observed stable states but should be there to stabilize intermediates in the process of conformational transition. These non native interactions were identified using molecular dynamics simulations. It is hypothesized that any perturbation in these interaction hotspots can severely impair the process of conformational transition. Identification of metastable intermediates and non native interactions in the CA domain and RR enables us to elucidate the underlying mechanism of signal transmission through the TCS.

Copyright by
RAHUL BANERJEE
2014

ACKNOWLEDGEMENTS

- I would like to thank my co-advisors, Dr. Robert Cukier and Dr. Honggao Yan for great support and guidance in research and life. Thank you for guiding me through this great journey of learning and making it an enjoyable experience overall. I am fortunate enough for I was given freedom to explore various research projects, both experimental and computational, and my advisors always acted as pathfinders whenever I was directionless in this vast field. I was often given opportunity to collaborate within the group and elsewhere to learn and share new ideas. I am forever indebted for teaching me to think critically and making me into a scientist.
- I would like to express my sincerest gratitude to my collaborators, Dr. Sheng Yang He and Dr. Thomas D. Sharkey. I would also like to acknowledge Dr. Zachary Burton for giving me the opportunity to work in a very interesting project on RNA polymerase. It is a great learning experience interacting with Dr. John Withers, Dr. Nadesan Sudhakar, Dr. Jian Yao in the He-lab, Aparajita Banerjee in the Sharkey-lab and Kris Opron in the Burton-lab.
- I would like to acknowledge members of my guidance committee including Dr. Leslie Kuhn, Dr. Xuefei Huang, and Dr. Aizhuo Liu for critical insight.
- I am grateful to present and past members of the Yan-group. I would like to thank Dr. Jifeng Wang and Dr. Zhenwei Lu for helping me to adapt in the new lab during my initial days. I would like to thank Dr. Yue Li and Yan Wu for their patience to train me in different molecular biology and protein biochemistry techniques. We were able to make an excellent team effort in some collaborative projects within the lab. Much credit goes to Dr. Aizhuo Liu and Dr. David Achila for this effort. I would like to thank Dr. Aizhuo Liu once again for providing the coordinates of the NMR ensemble of the HK853 catalytic and ATP-binding domain for structural comparison.

- I would like to thank Dr. Kaillathe Padmanabhan at the Macromolecular Computer Facility in the Department of Biochemistry, Michigan State University and Dr. Dirk Colbry at High Performance Computer Center at Michigan State University for all technical support while using the computer facilities.
- I would also like to extend my appreciation to Dr. Michael Feig and present and former members of the Feig-lab for helpful discussions.
- My path in the graduate school would not have been possible without the support and encouragement of my family, my parents, in-laws, my wife and kid. In spite of some great difficulties associated with their age and illness, my Mom and Dad always sounded so encouraging to shape my career path. Special thanks go out to my sister and her family for all good wishes. I would like to thank my in laws for enormous support and encouragement.
- I am forever indebted to Dr. Santanu Maitra in California State University, Fresno, CA for inspiration and encouragement. I am grateful to Dr. Iwan Setiawan, Dr. Anindya Chanda and Dr. Surajit Banerjee as I would have never survived my first year of graduate school without their kind friendship. Thank you all for some great memories.
- My wife, Koyeli epitomizes resilience, tenacity and perseverance against all odds. She is my colleague in the graduate school as well. There is no word to say thank you for supporting my passion for higher education. Together we came here six years back with our toddler son, Raunak who is now eight years old “grown up”. During the most difficult days of graduate life, he had been such a great source to rejuvenate ourselves. He often enjoys watching “dancing of protein molecule” in molecular viewers together with me. Koyeli and Raunak, you both will remain the greatest source of inspiration for me.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xii
CHAPTER 1: Introduction	1
1.1. Two-component system signal transduction	2
1.2. Distribution and importance	2
1.3. Domain architecture	5
1.3.1. Histidine kinase	7
1.3.2. Response regulator	10
1.4. Modular architecture	13
1.5. Discussion of the problem: conformational dynamics of TCS proteins	13
1.6. Summary	16
REFERENCES	18
CHAPTER 2: Conformational Transition of Response Regulator RR468 in a Two-component System Signal Transduction Process	23
2.1. Abstract	24
2.2. Introduction	25
2.3. Method	31
2.3.1. MD simulation of RR468 without phosphoryl group and Mg^{2+}	31
2.3.2. MD simulation of active form RR468 (P~RR468)	34
2.3.3. Targeted Molecular Dynamics simulations	34
2.3.4. Principal Component Analysis (PCA)	35
2.3.5. DEVIATION Method	36
2.3.6. EXTEND Method	37
2.3.7. PMF construction using WHAM	39
2.3.8. Conformational clustering of the transition path using Wordom	39
2.3.9. Change of pairwise interaction along transition path using Wordom	40
2.3.10. Hydrogen Bond Analysis	41
2.4. Results	43
2.4.1. Pairwise interactions among residues in the crystal structures are maintained in the MD trajectories	45
2.4.2. Hydrogen bond comparison of the crystal structures and MD trajectories	47
2.4.3. PCA, DEVIATION and EXTEND show that the $\beta 3$ - $\alpha 3$ loop can partially transit between inactive and active forms in IAC and ACT simulations	48
2.4.4. Targeted molecular dynamics (TMD) can span inactive to active, and vice versa, conformations	58
2.4.5. Potentials of Mean Force (PMF) for transitions of key residues in the TMD trajectory	60
2.4.6. Hydrogen bond analysis of the TMD trajectory identifies residues	

critical to the conformational transition	64
2.4.7. Conformational clustering of the TMD trajectories permits identification of intermediates	67
2.4.8. PSN analysis with the I2A and A2I TMD trajectories	69
2.4.9. Hydrogen bond analysis with the I2A and A2I TMD trajectories	71
2.5. Discussion	74
2.6. Conclusion	82
APPENDIX	83
REFERENCES	100
CHAPTER 3: Conformational Dynamics in a Two-component System: Molecular Dynamics and Markov State Model Analysis of a Response Regulator Protein	103
3.1. Abstract	104
3.2. Introduction	105
3.3. Method	109
3.3.1. Data Accumulation	109
3.3.2. Markov State Modeling	109
3.3.3. Clustering and Microstates	110
3.3.4. Lumping of Microstates into Macrostates	111
3.3.5. Free energy landscapes	114
3.3.6. Insertion of Mg^{2+} for selected macrostates	114
3.4. Results	115
3.4.1. Microstate and macrostate clustering of the trajectory	115
3.4.2. Residue distance based structural characterization of macrostates	115
3.4.3. Transition path theory (TPT) analysis	119
3.5. Discussion	121
3.5.1. Conformational sampling by nonequilibrium simulation followed by multiple equilibrium simulations	121
3.5.2. Macrostates 23 and 16, respectively, most resemble the inactive and active like conformations	121
3.5.3. Critical macrostate intermediates along the transition path can be parameterized by $\Delta Q_{\beta 3 \alpha 3}^M$ and $\Delta Q_{\alpha 4}^M$ values	122
3.5.4. Macrostate 22 is the most stable intermediate where the $\beta 3 \alpha 3$ loop conformation is significantly different from the inactive or active state	123
3.5.5. Critical intermediates preceding macrostate 22	126
3.5.6. Macrostate 20 exhibits enhanced flexibility of the $\beta 3 \alpha 3$ loop	126
3.5.7. Macrostate 16 has a stable $\beta 3 \alpha 3$ loop due to proximal residue interactions	130
3.5.8. Proximal contacts control exposure of the D53 side-chain to bulk water	130
3.5.9. Macrostate 20 is a transition state separating inactive and active-like states	133
3.5.10. Free energy landscapes can suggest a most stable intermediate	134
3.5.11. Active site architecture in the presence of a Mg^{2+} ion	137
3.6. Conclusion	140
APPENDIX	143
REFERENCES	152

CHAPTER 4: Conformational Dynamics of the ATP-lid Loop in the Histidine Kinase HK853 in Two-component System Signal Transduction	156
4.1. Abstract	157
4.2. Introduction	158
4.3. Method	163
4.3.1. Structural models for ATP-Mg ²⁺ bound form of the HK853 CA domain	163
4.3.2. Equilibrium MD simulations	163
4.3.3. Targeted MD simulations	166
4.3.4. Pairwise interaction analysis	166
4.4. Results	167
4.4.1. ATP binding models of the CA domain of HK853	167
4.4.2. The ATP-lid is the most flexible region of the CA domain	172
4.4.3. Orientation angle: the angle K468-I476-S434 provides a reaction coordinate for the ATP-lid orientation	174
4.4.4. ATP-lid dihedral angles show a characteristic difference between the two starting structures	177
4.4.5. Conformational transition connects two distinct states of the ATP-lid loop in the absence of ATP and Mg ²⁺	179
4.5. Discussion	181
4.5.1. The ATP-lid orientation in HK853 models compared with that in the CheA and PhoQ histidine kinases	181
4.5.2. Interaction between the ATP-lid and MgATP in HK853, CheA and PhoQ histidine kinases	184
4.5.3. Role of a salt bridge in the conformational transition	190
4.5.4. Hydrophobic interactions of Val431 and Leu444 and their role in conformational plasticity	190
4.6. Conclusion	194
REFERENCES	196
CHAPTER 5: Conclusions and Perspective	199
5.1. Understanding signal transduction in a Two-component system	200
5.2. Perspective	201
REFERENCES	203

LIST OF TABLES

Table 1.1. Biological functions of some representative Two-component systems with the name of the organism it was reported with.	4
Table 2.1. A summary of molecular dynamics (MD) and targeted molecular dynamics (TMD) simulations along with the acronyms used in this chapter.	33
Table 2.2. Reaction coordinates for loop motions.	50
Table 2.3. PCA mode eigenvalues derived from the inactive and active trajectories.	60
Table 2.4. Reaction coordinates used for calculation of the PMF for local transitions.	62
Table A1 A. Hydrogen bonds identified from the crystal structures of the active form of RR468 and their frequency during ACT simulation.	87
Table A1 B. Hydrogen bonds identified from the crystal structures of the inactive form of RR468 and their frequency during IAC simulation .	88
Table A2. Change in interaction from one TMD cluster to the next cluster during inactive to active (I2A) conformational transition.	89
Table A3. Change in pairwise interactions in two consecutive TMD clusters during active to inactive (A2I) conformational transition.	91
Table A4. Change in hydrogen bond interactions in two consecutive TMD clusters during inactive to active (I2A) conformational transition.	92
Table A5. Change in hydrogen bond interaction from one TMD cluster to the next cluster during active to inactive (A2I) conformational transition.	93
Table A6. Pair wise interactions involving residue number 53-60 ($\beta 3$ - $\alpha 3$), 83-87 ($\beta 4$ - $\alpha 4$), and 104-108 ($\beta 5$ - $\alpha 5$).	94
Table A7. Hydrogen bonds identified from the I2A TMD trajectory that has rare occurrence in the equilibrium trajectories.	97
Table A8. Hydrogen bonds identified from the A2I TMD trajectory that has rare occurrence in the equilibrium trajectories.	98
Table A9. Details of Steered MD simulation to sample transitional phase for D60 rearrangement and T83.	99
Table 3.1. Accessible surface area (ASA) calculated for D53 and the	

distance between the K105 and D9 side-chains.	132
Table 3.2. Structural features of four critical macrostates identified from the MSM and TPT analysis, and subsequent docking and structural refinement.	142
Table B1. Critical $C\alpha$ - $C\alpha$ distances.	144
Table B2. Critical $\beta 3\alpha 3$ loop distances in macrostates.	145
Table B3. Critical $\beta 4\alpha 4$ loop distances in macrostates.	146
Table B4. Transition path and reactive flux between macrostate 23 and 16.	147
Table B5. Characterization of macrostates.	149
Table 4.1. Summary of structural models.	165
Table 4.2. Polar and ionic interactions of ATP with Mg^{2+} and CA domain residues in COM2C2A and COM3DGE simulations .	169
Table 4.3. Hydrophobic interactions of Val431 and Leu444.	192

LIST OF FIGURES

Figure 1.1. Domain architecture of Two-component systems and signal propagation	6
Figure 1.2. Structures of the kinase core in HKs.	9
Figure 1.3. Structures of RR.	11
Figure 2.1. A comparison between the inactive and active conformations of RR468.	28
Figure 2.2. RMSD and RMSF during equilibrium simulations.	44
Figure 2.3. Intra and inter loop interactions of the proximal loop residues.	46
Figure 2.4. Reaction coordinates for loop motions	51
Figure 2.5. DEV analysis.	55
Figure 2.6. EXTEND analysis.	57
Figure 2.7. Critical residue conformation or orientation during TMD simulations.	59
Figure 2.8. PMF analysis.	63
Figure 2.9. Critical transient hydrogen bonds.	66
Figure 2.10. DRMS during TMD simulations.	68
Figure 2.11. Change in pairwise interaction during RR468 activation.	70
Figure 2.12. Critical salt bridge interactions during inactive and active form simulation of RR468.	77
Figure A1. System considered for the DFT calculation.	84
Figure A2. Time series shows parameters that best describe conformation and position of M56, D60 and Th83.	85
Figure A3. PMF calculation.	86
Figure 3.1. Relaxation timescale and population.	113
Figure 3.2. Connectivity map.	118
Figure 3.3. Transition path.	120
Figure 3.4. Conformational change in macrostate 22.	125

Figure 3.5. Conformational change in macrostate 20 and 16.	128
Figure 3.6. Free energy landscape.	136
Figure 3.7. Active site architecture in presence of Mg^{2+} .	139
Figure B1. Characterization of macrostates based on critical side chain orientation.	151
Figure 4.1. Binding models MgATP in the CA domain of HK853.	170
Figure 4.2. RMSDs and RMSFs of the MD simulations of HK853 CA domain.	173
Figure 4.3. Definition of orientation angle .	175
Figure 4.4. Change in the ATP-lid orientation angle in the MD simulations.	176
Figure 4.5. Ramachandran plot analysis of ATP-lid loop residues.	178
Figure 4.6. Change in the ATP-lid orientation angle along the TMD simulations.	180
Figure 4.7. Structural comparison of CA domain of HK853, PhoQ and CheA.	183
Figure 4.8. Active site in HK853, PhoQ and CheA.	186
Figure 4.9. Interaction of the ATP molecule with ATP-lid loop residue.	189
Figure 4.10. Critical salt bridge interaction.	193

CHAPTER 1

Introduction

1.1. Two-component system signal transduction

Two-component system (TCS) signal transduction is the predominant mechanism in bacteria and other simpler forms of life to enable them to adapt to harsh changes in environmental or growth conditions such as temperature, nutrients etc. (Stock et al., 2000). This stress related signaling system is activated during host invasion, drug resistance, motility, phosphate uptake, osmoregulation, and nitrogen fixation, amongst others in bacteria, and some eukaryotes like fungi, protozoa (Blanco et al., 2002). This mode of signal transduction is abundant in plant systems as well. In 1980s, genetic sequencing of various regulatory systems led to finding that these diverse and seemingly unrelated processes were in fact controlled by closely-related pairs of this class of regulatory proteins. The term “Two Component System” was first coined by Nixon et. al. where the signal transduction involves a sensor kinase and a response regulator protein (Nixon et al., 1986).

1.2. Distribution and importance

More than 50,000 TCS proteins identified so far suggests that they allow adaptational responses to a huge variety of environmental stimuli in microorganisms. The Microbial Signal Transduction database (MiST) does the book keeping for the signal transduction proteins in bacterial and archaeal genomes (Ulrich and Zhulin, 2010). The TCS proteins are identified using various domain profiling technique based on sequence and structural motifs that directly or indirectly implicate a particular protein participating in signal transduction (Ulrich and Zhulin, 2010). The genomic surveys show that TCSs are abundant in super kingdoms: Eubacteria, Archaea, Protista, Fungi, and Plantae. However this mode of signal transduction is completely absent in the animal kingdom, Animalia. TCS components are present in 864 out of 899

available completely sequenced bacterial genomes (Wuichet et al., 2010). The TCSs are important for adaptational response in microorganisms to cope with harsh environmental conditions. This mode of signal transduction controls responses like spore formation, energy homeostasis and flagellar movement to survive against the harsh condition caused by nutrient depletion or change in growth condition such as temperature, pH etc. There are TCSs that controls bacterial adaption in the host system by means of biofilm production, quorum sensing, toxin production, antibiotic resistance etc. A short list of Two Component Systems with their origin and function is given in Table 1.1, which helps understand their importance in variety of biological functions.

There has been multitude of effort to characterize TCSs as future target for therapeutic intervention to treat bacterial infection mainly for two reasons (Gotoh et al., 2010b; Worthington et al., 2013). Firstly, gene clusters contributing to the processes such as cell growth and pathogenicity are often controlled by this mode of signal transduction microbial systems (Gotoh et al., 2010b). Secondly, this mode of signal transduction is found in all forms of life except in the animal world (Thomason and Kay, 2000). Preliminary studies are reported where drugs that target a TCS required for growth serve as novel bactericidal agents for multi-drug-resistant bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococcus* (VRE) (Gotoh et al., 2010a; Okada et al., 2010). Inhibitors of TCSs that control virulence factors, such as biofilm production and quorum sensing (QS), are also reported as bacteriostatic agents that could control virulence without killing the pathogenic bacteria (Cegelski et al., 2008; Rasko et al., 2008). TCSs are also present in eukaryotic microorganisms and are involved in pathogenicity. Thus, TCSs in medically important fungal pathogens are also considered as potential drug targets (Gotoh et al., 2010b).

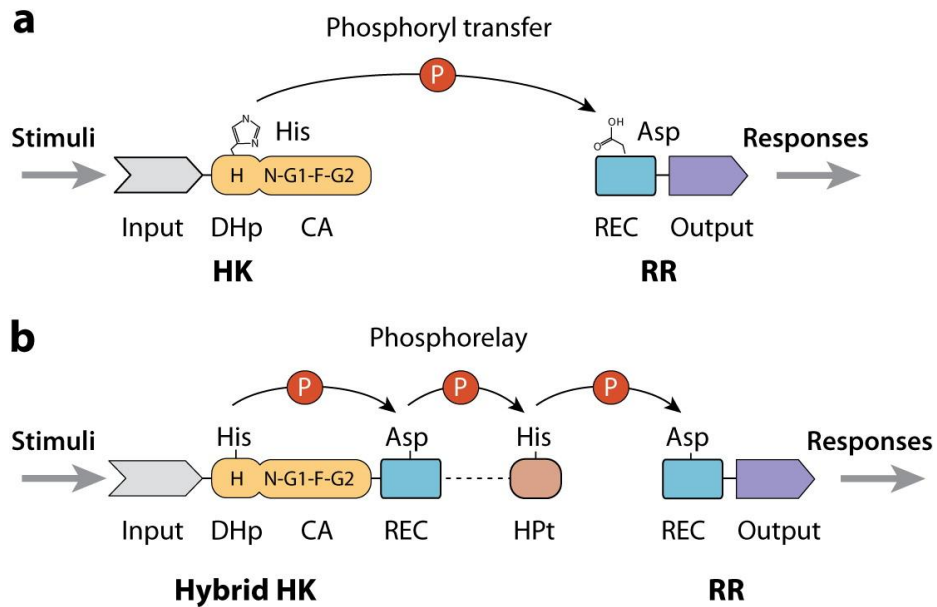
Table 1.1. **Biological functions of some representative Two-component systems with the name of the organism it was reported with.**

Biological Function	Organism	Name [†]	Reference
Nutrient acquisition (K ⁺ uptake)	<i>Escherichia coli</i>	KdpD-KdpE	(Kremling et al., 2004)
Energy metabolism	<i>Staphylococcus epidermidis</i>	LytS-LytR	(Zhu et al., 2010)
Chemotaxis	<i>Escherichia coli</i>	CheA-CheY	(Hess et al., 1991)
pH sensing	<i>Agrobacterium tumefaciens</i>	VirA-VirG	(Gao and Lynn, 2005)
Osmolarity	<i>Escherichia coli</i>	EnvZ-OmpR	(Cai and Inouye, 2002)
Light quality	<i>Caulobacter crescentus</i>	LovK-LovR	(Purcell et al., 2007)
Sporulation	<i>Bacillus subtilis</i>	KinA-KinD	(Dago et al., 2012)
Virulence/Pathogenesis	<i>Coxiella burnetii</i>	PmrA-PmrB	(Zusman et al., 2007)
Toxin production	<i>Clostridium perfringens</i>	VirS-VirR	(Cheung et al., 2010)
Nitrogen fixation	<i>Sinorhizobium meliloti</i>	FixL-FixJ	(Roche et al., 2002)
Quorum-sensing	<i>Staphylococcus aureus</i>	AgrC-AgrA	(Sun et al., 2012)
Nitrogen regulation	<i>Klebsiella pneumoniae</i>	NtrB-NtrC	(Martinez-Argudo et al., 2001)
Antibiotic (Polymyxin) resistance	<i>Escherichia coli</i>	PhoP-PhoQ PmrA-PmrB	(Tatsis et al., 2009)

[†] Histidine kinase-Response regulator

1.3. Domain architecture

Integral to TCS signal transduction is a phosphotransfer reaction between a membrane-bound, histidine kinase (HK) and a cognate response regulator (RR) protein (Figure 1.1). The HK is typically the input component of the pathway with a sensor domain, designed to sense extracellular stimuli. A conserved histidine residue in the intracellular region of HK gets autophosphorylated in response to the external stimuli by the ATP-bound catalytic (CA) domain of HK. The RR typically carries out the output of the system, regulated by the phosphorylation of a conserved aspartate residue in the RR by the HK (Gao and Stock, 2009). Signaling is terminated by dephosphorylation of the phosphorylated RR (P~RR) either by phosphatase activity of the HK or due to transfer of the phosphoryl group to a downstream substrate.



AR Gao R, Stock AM. 2009.
Annu. Rev. Microbiol. 63:133–54

Figure 1.1. **Domain architecture of Two-component systems and signal propagation.** (a) In the most basic scheme, the TCS pathway features phosphoryl transfer reaction between the highly conserved kinase core of HK and receiver (REC) domains to couple extracellular stimuli to the output responses. The kinase core of HK consists of two domains, DHp and CA domains. The DHp domain hosts a conserved histidine residue, which becomes phosphorylated in response to the extracellular stimuli sensed through the input, extracellular domain of HK. The catalytic domain (CA) of HK contains sequence motifs, N, G1, F and G2, which are conserved across histidine kinases (HKs). These conserved motifs play crucial role in nucleotide recruitment. (b) A phosphorelay scheme is also utilized by some HKs involving additional REC, and a separate histidine phosphotransfer (HPT) domain is responsible for phosphorylation of RR to exert output action. This scheme involves multiple phosphotransfer events.

1.3.1. Histidine kinase

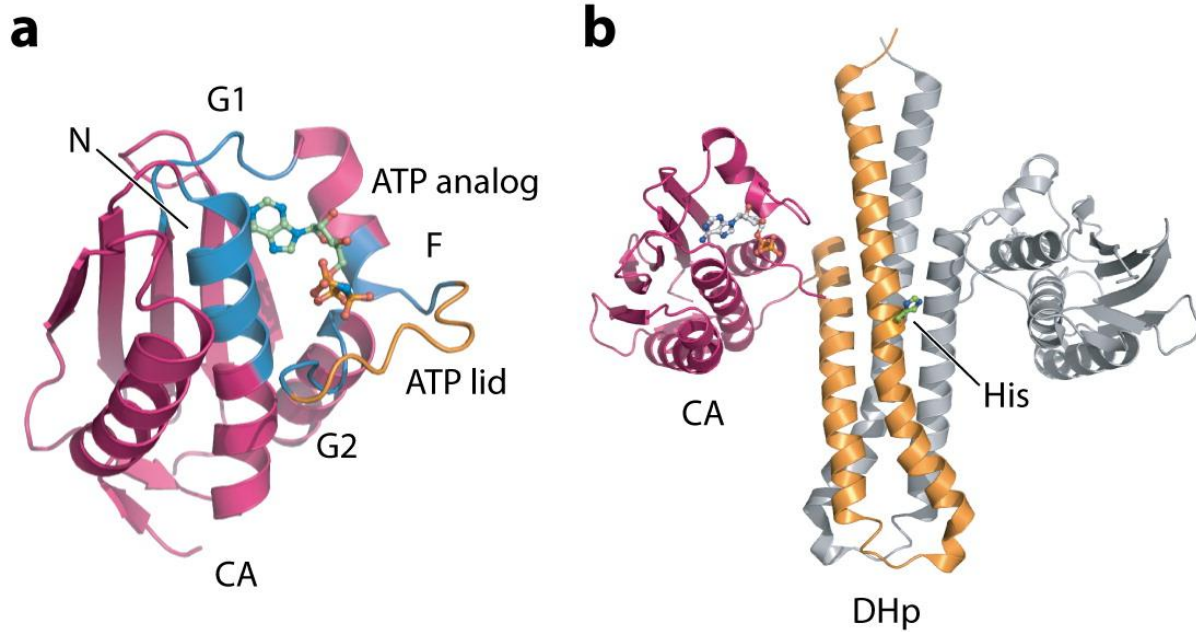
HK is the membrane bound input component of the TCSs, with diverse input domains at the extracellular region linked to a fairly conserved catalytic core in the interior to the cell. This domain architecture allows the coupling of a wide variety of input signals to generate appropriate cellular responses that are controlled by the same chemical mechanism of autophosphorylation and phosphoryl transfer in the intracellular domains (Gao and Stock, 2009). HK plays a dual role in many TCSs, where it also catalyzes the dephosphorylation of its P~RR partner, thus acting as an RR kinase and also P~RR phosphatase (Casino et al., 2010). Apart from the signal recognition at the extracellular sensory domain, autophosphorylation, phosphoryl transfer, and phosphatase reactions involving intracellular domains of HK and RR control the signal transduction in TCSs. The sensory domain is located at the N-terminal of the HK, and they share very little sequence similarity. Despite great advances in the understanding of the structure-function relationship in the intracellular domains, the structures of the sensory domains and the identity of the exact stimuli remain unclear (Mascher et al., 2006; Szurmant et al., 2007).

In a prototypical HK, the cytoplasmic portion consists of two distinct domains: a dimerization and histidine-containing phosphotransfer domain (DHp), and a C-terminal catalytic and ATP binding (CA) domain (Figure 1.2). In HKs belonging to TCSs that are responsible for bacterial chemotaxis, the phosphorylation site is located on a distant histidine phosphotransfer (HPt) domain, and a separate domain is responsible for dimerization (Baker et al., 2006).

HKs function as dimers. Both *cis* and *trans* phosphorylation mechanisms have been observed but ATP bound to one subunit phosphorylates the conserved histidine either in the same subunit or in the other subunit in the dimer, but not both subunits. DHp domain is formed by two antiparallel

α helices, thereby forming a four helix bundle when it dimerizes. The conserved sequence motif in the DHp domain that hosts a highly conserved histidine residue is called an H-box. The H-box is located in the first helix of the DHp domain (Gao and Stock, 2009).

Histidine autophosphorylation is the first step in signal propagation through TCS after signal recognition by the sensor domain of HK on the exterior of the cell. The CA domain binds one ATP molecule via a flexible loop, namely ATP-lid, and a central helix. The γ -phosphate of the ATP is attacked by a conserved His of the DHp domain. A highly conserved ATP binding cavity is defined by a few characteristic sequence motifs, namely the N, G1, F, and G2 boxes. The ATP-lid is a long flexible loop in the CA domain that serves as a cover, as the name suggests, for the ATP-binding pocket (Gao and Stock, 2009).

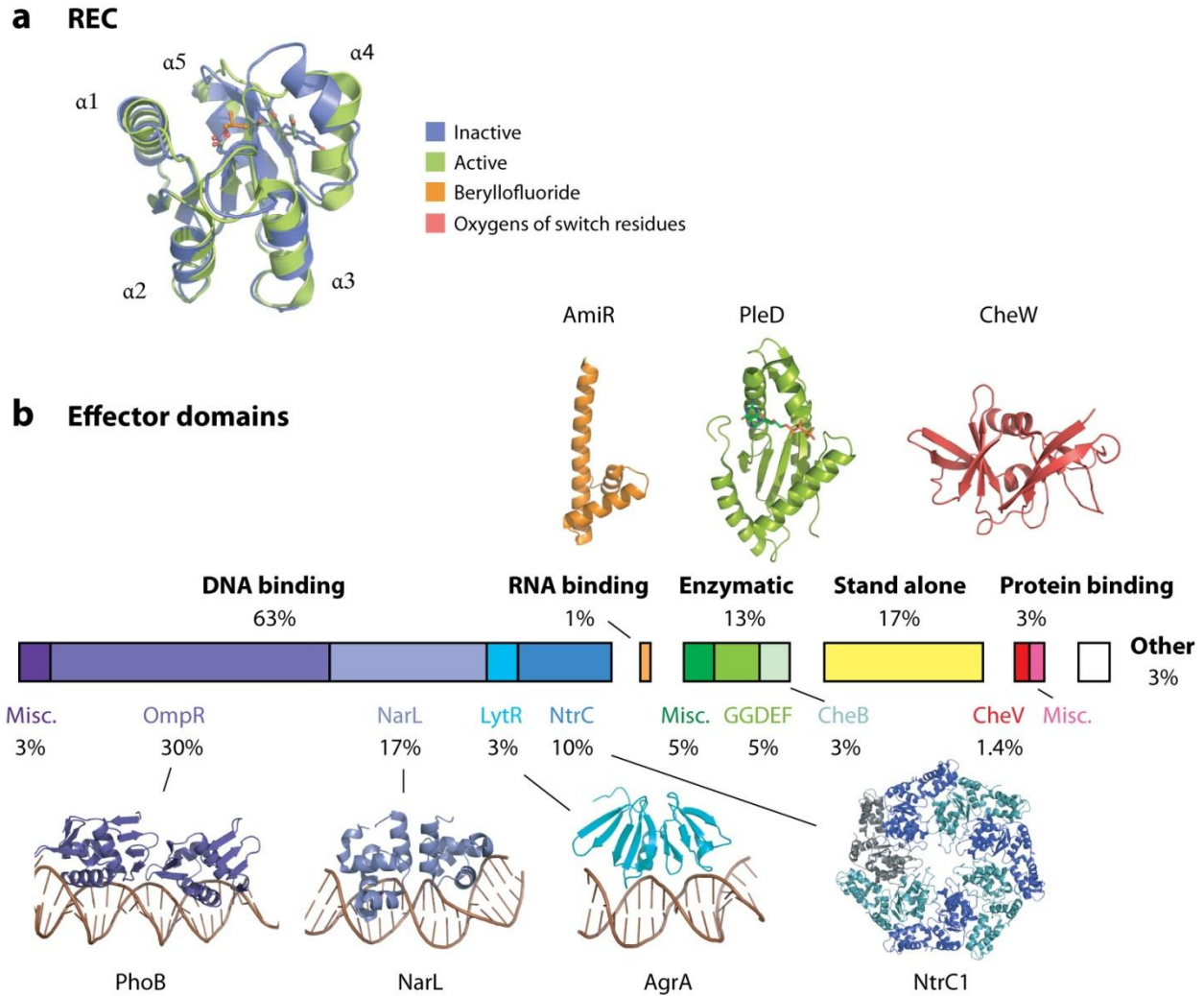


A Gao R, Stock AM. 2009.
R Annu. Rev. Microbiol. 63:133–54

Figure 1.2. **Structures of the kinase core in HKs.** Crystal structures of (a) the CA domain of *Escherichia coli* PhoQ (PDB ID: 1ID0) and (b) the entire kinase core of *Thermotoga maritima* HK853 (PDB ID: 2C2A). (a) Sequence motifs, N, G1, F, G2 play crucial role in nucleotide recruitment are shown in blue in the PhoQ CA domain structure. A flexible loop between conserved motifs, F and G2 is called ATP lid (orange). The ATP-lid loop covers the bound ATP analog, AMPPNP as shown in the crystal structure. (b) HK853 is a dimer, where DHp domains involve in dimerization. Two monomers of HK are shown in orange and grey. The DHp domain in HK853 hosts a conserved histidine (H260 shown) that becomes autophosphorylated by the ATP molecule bound to the CA domain.

1.3.2. Response regulator

Response regulator proteins typically carry out the output of the system, regulated by the phosphoryl transfer activity of the HK (Figure 1.3). A prototypical RR contains two domains, a receiver domain (REC) (Bourret, 2010; Gao and Stock, 2010) that hosts a phosphoacceptor aspartate residue and one effector (EFF) or output domain responsible for downstream action (Galperin, 2010). The effector domain interacts with its targets, typically gene promoters for downstream action. Phosphorylation of the conserved aspartate residue in the REC domain modulates the interactions of the RR that modulates its interaction with its targets. In ~15% of RRs in bacteria and ~50% of RRs in archaea (Jenal and Galperin, 2009), the EFF domain is absent and the effector function is taken over by the REC domain itself.



 Gao R, Stock AM. 2009.
Annu. Rev. Microbiol. 63:133–54

Figure 1.3. Structures of RR. (a) The structural superposition of inactive and active form PhoB receiver (REC) domains illustrates the conformational differences between these two forms. Residues undergo significant conformation change are shown in ball-and-stick mode. Beryllium trifluoride, which mimics the phosphoryl group of the phosphorylated aspartate in the active form, is colored orange. (b) The diverse effector (EFF) domains are grouped together according to their function. Majority (63%) of effector domains exert biological functions through binding to the target gene promoter region.

Figure 1.3 (cont'd)

Representative structures of individual subfamily effector domain folds are shown here to emphasize structural heterogeneity among the effector domain fold.

1.4. Modular architecture

Signal transduction through TCS follows a modular architecture, where the extracellular domain of the HK is very diverse and susceptible to specific extracellular stimuli. In the intracellular region, the signal transduction is carried out or controlled by three reaction mechanisms: autophosphorylation of the HK followed by phosphoryl transfer reaction from HK to RR and finally dephosphorylation of phosphorylated RR. The EFF domain is also very diverse among different TCSs as demanded by specific action required for a specific stimulus (Galperin, 2010). The phosphorylation-dephosphorylation reactions are controlled by reasonably conserved domains that involve the DHp and CA domains of HK and the REC domain of RR. Essentially this is the common mechanism in different TCSs. There are proteins that can augment this basic scheme, and many so called “Two Component” pathways in fact consist of more than two proteins. In such atypical TCSs, phosphoryl transfer takes place multiple times following histidine kinase autophosphorylation. The phosphoryl group is passed to an aspartate residue on a receiver domain, then to a histidine on another sensor domain, and finally to an aspartate on the receiver domain of a response regulator (Goulian, 2010).

1.5. Discussion of the problem: conformational dynamics of TCS proteins

Protein conformational dynamics in various response regulator proteins investigated using molecular dynamics simulations, starting from the experimental structures representing active and inactive forms of the protein (Bourret, 2010). Using experimental structures obtained from either NMR or X-ray crystallography as starting structures, MD simulations were used to understand the conformational dynamics involving different structural elements in the protein (Tatsis et al., 2009). Correlated motions and long range interactions in the chemotaxis protein,

CheY from two different organisms were compared using all-atoms MD simulations (Knaggs et al., 2007). Coarse grained simulation was used to address similar problems in the nitrogen regulatory protein C (NtrC) (Liu et al., 2008). Change in protein flexibility induced by phosphorylation, changes in key interactions at the phosphorylation site and the free energy cost of the phosphorylation event were estimated using MD simulation for a RR protein, Spo0F, in *Bacillus subtilis* responsible for its sporulation (Peters, 2009). Conformational transition is a local unfolding-folding event in a protein, which was obscured in equilibrium simulations due to high energy barriers separating experimentally observed active and inactive states. Conformational transition in the nitrogen fixation protein, FixJ, was investigated using targeted molecular dynamics simulation (TMD), where an additional constraint term was introduced in the potential energy function to force the system towards a given target structure (Roche et al., 2002). This approach was quite popular and followed in other literature examples as well to sample conformational states spanning experimentally observed structures of the response regulator proteins (Lei et al., 2009). A biasing potential was applied in some instances on a specific part of the protein as a way to probe the coupling (not the precise sequence of events) between different structural motifs (Formanek et al., 2006). A different form of non-equilibrium simulation, namely, self-guided Langevin dynamics (SGLD), was used where additional guiding forces was applied (Damjanovic et al., 2009). This external force in the SGLD method is proportional to the momentum of the particle averaged over a predetermined time window. Conformational transition paths were sampled by Ma and Cui using a combination of molecular dynamics (MD) and Monte Carlo (MC) simulations (Ma and Cui, 2007). It samples the phase space using MD and accepts or rejects the paths using MC procedures to obtain the transition path ensemble. The discrete path sampling method constructs a database of local minima on the potential energy

surface (Khalili and Wales, 2008). A transition state is defined geometrically as the point where the Hessian matrix of second derivatives has only one negative eigenvalue, and a minimum is defined as the point where this matrix has no negative eigenvalues. Phenomenological two-state rate constants are calculated from these databases using statistical rate theory for the individual minimum-to-minimum transitions (Khalili and Wales, 2008).

Different simulations apparently do not agree on the key mechanistic details associated with the conformational change upon phosphorylation. The active and inactive conformations differ mainly in the loops and secondary structural elements around the site of phosphorylation, the proximal region of the RR. Conformational dynamics were analyzed mainly in terms of the order of events involving different secondary structural elements (Khalili and Wales, 2008; Lei et al., 2009; Liu et al., 2008).

1.6. Summary

Conformational dynamics in various domains of HK and RR proteins are responsible for signal transmission in this mode of signal transduction. Moreover, a conformational transition associated with RR following the phosphoryl transfer reaction is responsible for generating the output by controlling the downstream regulation in this process. In the following chapters the results from several extensive computer simulations of various TCS domains will be presented. We have used the TCS, HK853-RR468 in *Thermotoga maritima* as a model system toward the understanding of the mechanism of signal transduction. High resolution structures for entire cytoplasmic domains of HK853 (Marina et al., 2005) and HK853 in complex with RR468 (Casino et al., 2009) are available in the Protein Data Bank (Bernstein et al., 1977). RR468 is a single domain response regulator. There are two structures available for the receiver domain of the response regulator protein, RR468; one for the inactive, unphosphorylated form and the other that resembles the active, phosphorylated form of the protein (Casino et al., 2009). Conformational transition between the inactive and the active forms of RR is the key mechanism for downstream action in TCS signaling pathway. Critical intermediate stages and nonnative interactions in RR468 conformational transition are proposed using various equilibrium and non-equilibrium molecular dynamics simulations in Chapters 2. The induced fit (Koshland, 1958) and conformational selection or the population shift models (Ma et al., 1999; Tsai et al., 1999) are two extreme cases of a continuum aimed at understanding the mechanism by which the conformational transition is achieved due to binding of a ligand. Population shift model suggests that various conformational states in a macromolecule like this are sampled in equilibrium irrespective of the fact whether the ligand is present or not. However it is extremely difficult to sample the intermediate states starting from a highly stable state conformation obtained through

structural biology techniques. Conformational snapshots from a non-equilibrium simulation trajectory connecting the inactive and active states of RR468 were used as starting points for multiple equilibrium simulations to sample intermediate states of RR468. A network of metastable states or intermediates in the conformational transition was identified by Markov state analysis of the extensive equilibrium MD data. Transition paths for the conformational transition spanning the inactive and active state of RR468 were obtained using Transition Path Theory (Berezhkovskii et al., 2009; Metzner et al., 2009). These results are presented in Chapter 3.

Conformational dynamics of the ATP-lid loop in HK plays a critical role in the autophosphorylation reaction in the signal transmission process. Molecular dynamics simulation of the catalytic and ATP-binding (CA) domain of HK853 provides insight into the conformational plasticity of the CA domain and the role of ATP-lid loop in nucleotide recruitment, which is in Chapter 4.

REFERENCES

REFERENCES

- Baker, M.D., Wolanin, P.M., and Stock, J.B. (2006). Signal transduction in bacterial chemotaxis. *Bioessays* 28, 9-22.
- Berezhkovskii, A., Hummer, G., and Szabo, A. (2009). Reactive flux and folding pathways in network models of coarse-grained protein dynamics. *J Chem Phys* 130.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). Protein Data Bank - Computer-Based Archival File for Macromolecular Structures. *J Mol Biol* 112, 535-542.
- Blanco, A.G., Sola, M., Gomis-Ruth, F.X., and Coll, M. (2002). Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure* 10, 701-713.
- Bourret, R.B. (2010). Receiver domain structure and function in response regulator proteins. *Curr Opin Microbiol* 13, 142-149.
- Cai, S.J., and Inouye, M. (2002). EnvZ-OmpR interaction and osmoregulation in *Escherichia coli*. *J Biol Chem* 277, 24155-24161.
- Casino, P., Rubio, V., and Marina, A. (2009). Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell* 139, 325-336.
- Casino, P., Rubio, V., and Marina, A. (2010). The mechanism of signal transduction by two-component systems. *Curr Opin Struct Biol* 20, 763-771.
- Cegelski, L., Marshall, G.R., Eldridge, G.R., and Hultgren, S.J. (2008). The biology and future prospects of antivirulence therapies. *Nature Reviews Microbiology* 6, 17-27.
- Cheung, J.K., Keyburn, A.L., Carter, G.P., Lanckriet, A.L., Van Immerseel, F., Moore, R.J., and Rood, J.I. (2010). The VirSR Two-Component Signal Transduction System Regulates NetB Toxin Production in *Clostridium perfringens*. *Infect Immun* 78, 3064-3072.
- Dago, A.E., Schug, A., Procaccini, A., Hoch, J.A., Weigt, M., and Szurmant, H. (2012). Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis. *Proc Natl Acad Sci USA* 109, E1733-E1742.
- Damjanovic, A., Garcia-Moreno, B., and Brooks, B.R. (2009). Self-guided Langevin dynamics study of regulatory interactions in NtrC. *Proteins* 76, 1007-1019.
- Formanek, M.S., Ma, L., and Cui, Q. (2006). Reconciling the "old" and "new" views of protein allostery: A molecular simulation study of chemotaxis Y protein (CheY). *Proteins* 63, 846-867.
- Galperin, M.Y. (2010). Diversity of structure and function of response regulator output domains. *Curr Opin Microbiol* 13, 150-159.

- Gao, R., and Lynn, D.G. (2005). Environmental pH sensing: Resolving the VirA/VirG two-component system inputs for *Agrobacterium* pathogenesis. *J Bacteriol* 187, 2182-2189.
- Gao, R., and Stock, A.M. (2009). Biological Insights from Structures of Two-Component Proteins. *Annu Rev Microbiol* 63, 133-154.
- Gao, R., and Stock, A.M. (2010). Molecular strategies for phosphorylation-mediated regulation of response regulator activity. *Curr Opin Microbiol* 13, 160-167.
- Gotoh, Y., Doi, A., Furuta, E., Dubrac, S., Ishizaki, Y., Okada, M., Igarashi, M., Misawa, N., Yoshikawa, H., Okajima, T., *et al.* (2010a). Novel antibacterial compounds specifically targeting the essential WalR response regulator. *J Antibiot* 63, 127-134.
- Gotoh, Y., Eguchi, Y., Watanabe, T., Okamoto, S., Doi, A., and Utsumi, R. (2010b). Two-component signal transduction as potential drug targets in pathogenic bacteria. *Curr Opin Microbiol* 13, 232-239.
- Goulian, M. (2010). Two-component signaling circuit structure and properties. *Curr Opin Microbiol* 13, 184-189.
- Hess, J.F., Bourret, R.B., and Simon, M.I. (1991). Phosphorylation Assays for Proteins of the 2-Component Regulatory System Controlling Chemotaxis in *Escherichia-Coli*. *Methods Enzymol* 200, 188-204.
- Jenal, U., and Galperin, M.Y. (2009). Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics. *Curr Opin Microbiol* 12, 152-160.
- Khalili, M., and Wales, D.J. (2008). Pathways for conformational change in nitrogen regulatory protein C from discrete path sampling. *J Phys Chem B* 112, 2456-2465.
- Knaggs, M.H., Salsbury, F.R., Edgell, M.H., and Fetrow, J.S. (2007). Insights into correlated motions and long-range interactions in CheY derived from molecular dynamics simulations. *Biophys J* 92, 2062-2079.
- Koshland, D.E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci USA* 44, 98-104.
- Kremling, A., Heermann, R., Centler, F., Jung, K., and Gilles, E.D. (2004). Analysis of two-component signal transduction by mathematical modeling using the KdpD/KdpE system of *Escherichia coli*. *BioSyst* 78, 23-37.
- Lei, M., Velos, J., Gardino, A., Kivenson, A., Karplus, M., and Kern, D. (2009). Segmented Transition Pathway of the Signaling Protein Nitrogen Regulatory Protein C. *J Mol Biol* 392, 823-836.
- Liu, M.S., Todd, B.D., Yao, S.G., Feng, Z.P., Norton, R.S., and Sadus, R.J. (2008). Coarse-grained dynamics of the receiver domain of NtrC: Fluctuations, correlations and implications for allosteric cooperativity. *Proteins* 73, 218-227.

- Ma, B.Y., Kumar, S., Tsai, C.J., and Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein Eng* 12, 713-720.
- Ma, L., and Cui, Q. (2007). Activation mechanism of a signaling protein at atomic resolution from advanced computations. *J Am Chem Soc* 129, 10261-10268.
- Marina, A., Waldburger, C.D., and Hendrickson, W.A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J* 24, 4247-4259.
- Martinez-Argudo, I., Martin-Nieto, J., Salinas, P., Maldonado, R., Drummond, M., and Contreras, A. (2001). Two-hybrid analysis of domain interactions involving NtrB and NtrC two-component regulators. *Mol Microbiol* 40, 169-178.
- Mascher, T., Helmann, J.D., and Unden, G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev* 70, 910-+.
- Metzner, P., Schutte, C., and Vanden-Eijnden, E. (2009). Transition Path Theory for Markov Jump Processes. *Multiscale Model Sim* 7, 1192-1219.
- Nixon, B.T., Ronson, C.W., and Ausubel, F.M. (1986). 2-Component Regulatory Systems Responsive to Environmental Stimuli Share Strongly Conserved Domains with the Nitrogen Assimilation Regulatory Genes Ntrb and Ntrc. *Proc Natl Acad Sci USA* 83, 7850-7854.
- Okada, A., Igarashi, M., Okajima, T., Kinoshita, N., Umekita, M., Sawa, R., Inoue, K., Watanabe, T., Doi, A., Martin, A., *et al.* (2010). Walkmycin B targets WalK (YycG), a histidine kinase essential for bacterial cell growth. *J Antibiot* 63, 89-94.
- Peters, G.H. (2009). The effect of Asp54 phosphorylation on the energetics and dynamics in the response regulator protein SpoOF studied by molecular dynamics. *Proteins* 75, 648-658.
- Purcell, E.B., Siegal-Gaskins, D., Rawling, D.C., Fiebig, A., and Crosson, S. (2007). A photosensory two-component system regulates bacterial cell attachment. *Proc Natl Acad Sci USA* 104, 18241-18246.
- Rasko, D.A., Moreira, C.G., Li, D.R., Reading, N.C., Ritchie, J.M., Waldor, M.K., Williams, N., Taussig, R., Wei, S.G., Roth, M., *et al.* (2008). Targeting QseC signaling and virulence for antibiotic development. *Science* 321, 1078-1080.
- Roche, P., Mouawad, L., Perahia, D., Samama, J.P., and Kahn, D. (2002). Molecular dynamics of the FixJ receiver domain: movement of the beta 4-alpha 4 loop correlates with the in and out flip of Phe101. *Protein Sci* 11, 2622-2630.
- Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-component signal transduction. *Annu Rev Biochem* 69, 183-215.
- Sun, F., Liang, H.H., Kong, X.Q., Xie, S., Cho, H., Deng, X., Ji, Q.J., Zhang, H.Y., Alvarez, S., Hicks, L.M., *et al.* (2012). Quorum-sensing agr mediates bacterial oxidation response via an

intramolecular disulfide redox switch in the response regulator AgrA. *Proc Natl Acad Sci USA* 109, 9095-9100.

Szurmant, H., White, R.A., and Hoch, J.A. (2007). Sensor complexes regulating two-component signal transduction. *Curr Opin Struct Biol* 17, 706-715.

Tatsis, V.A., Tsoulos, I.G., Krinas, C.S., Alexopoulos, C., and Stavrakoudis, A. (2009). Insights into the structure of the PmrD protein with molecular dynamics simulations. *Int J Biol Macromol* 44, 393-399.

Thomason, P., and Kay, R. (2000). Eukaryotic signal transduction via histidine-aspartate phosphorelay. *J Cell Sci* 113, 3141-3150.

Tsai, C.J., Kumar, S., Ma, B.Y., and Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Sci* 8, 1181-1190.

Ulrich, L.E., and Zhulin, I.B. (2010). The MiST2 database: a comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res* 38, D401-D407.

Worthington, R.J., Blackledge, M.S., and Melander, C. (2013). Small-molecule inhibition of bacterial two-component systems to combat antibiotic resistance and virulence. *Future Med Chem* 5, 1265-1284.

Wuichet, K., Cantwell, B.J., and Zhulin, I.B. (2010). Evolution and phyletic distribution of two-component signal transduction systems. *Curr Opin Microbiol* 13, 219-225.

Zhu, T., Lou, Q.A., Wu, Y., Hu, J.A., Yu, F.Y., and Qu, D. (2010). Impact of the *Staphylococcus epidermidis* LytSR two-component regulatory system on murein hydrolase activity, pyruvate utilization and global transcriptional profile. *BMC Microbiol* 10.

Zusman, T., Aloni, G., Halperin, E., Kotzer, H., Degtyar, E., Feldman, M., and Segal, G. (2007). The response regulator PmrA is a major regulator of the icm/dot type IV secretion system in *Legionella pneumophila* and *Coxiella burnetii*. *Mol Microbiol* 63, 1508-1523.

CHAPTER 2

Conformational Transition of Response Regulator RR468 in a Two-component System Signal Transduction Process

Rahul Banerjee, Honggao Yan and Robert I. Cukier

Adapted from

J. Phys. Chem. B 2014, 118, 4727–4742

2.1. Abstract

Signal transduction can be accomplished via a Two Component System (TCS) consisting of a histidine kinase (HK) and a response regulator (RR). In this work, we simulate the response regulator RR468 from *Thermotoga maritima*, in which phosphorylation and dephosphorylation of a conserved aspartate residue acts as a switch via a large conformational change concentrated in three proximal loops. A detailed view of the conformational transition is obscured by the lack of stability of the intermediate states, which are difficult to detect using common structural biology techniques. Molecular Dynamics (MD) trajectories of the inactive and active conformations were run, and show that the inactive (or active) trajectories do not exhibit sampling of the active (or inactive) conformations on this time scale. Targeted MD (TMD) was used to generate trajectories that span the inactive and active conformations and provide a view of how a localized event like phosphorylation can lead to conformational changes elsewhere in the protein, especially in the three proximal loops. The TMD trajectories are clustered to identify stages along the transition path. Residue interaction networks are identified that point to key residues having to rearrange in the process of transition. These are identified using both hydrogen bond analysis and residue interaction strength measurements. Potentials of mean force are generated for key residue rearrangements to ascertain their free energy barriers. We introduce methods that attempt to extrapolate from one conformation to the other and find that the most fluctuating proximal loop can transit part way from one to the other, suggesting that this conformational information is embedded in the sequence.

2.2. Introduction

Two Component System (TCS) signal transduction is the predominant mechanism for bacteria to sense environmental conditions such as temperature, nutrients, and osmopressure (Bourret and Silversmith, 2010; Stewart, 2010). The critical role of TCS signal transduction in bacteria for their virulence and pathogenesis and the absence of this mode of signal transduction in higher organisms make TCSs an important target for future therapeutic interventions to treat bacterial infections (Casino et al., 2010). Integral to TCS pathways is a phosphotransfer reaction between a membrane bound histidine kinase (HK) and a cognate response regulator (RR) protein. A prototypical RR contains two domains, a receiver domain (REC) that hosts a phosphoacceptor aspartate residue and an effector or output domain responsible for downstream action (Bourret, 2010; Gao and Stock, 2010). In RR468, the system of our interest, the effector domain is absent and the receiver domain itself functions as an effector domain (Casino et al., 2009). RR468 in *Thermotoga maritima* is involved in relaying a phosphoryl group from its cognate pair histidine kinase, HK853, in response to extracellular stimuli (Casino et al., 2009; Gao and Stock, 2009). The activated TCS is characterized by phosphorylated RR468 where the phosphoryl group is covalently bound to a conserved aspartate (D53) residue.

In spite of an extensive literature focused on biochemical and mutational studies of TCS signaling, there was almost no structural basis to understand the mechanism until 2009 when Casino and coworkers published a series of high resolution crystal structures of HK853, RR468 and a HK853-RR468 complex (Casino et al., 2009). The conformations of RR468 for free and BeF_3^- -bound forms highlighted the structural difference between active (phosphorylated) and inactive (dephosphorylated) conformations of this protein. Beryllium trifluoride, covalently attached to a aspartate residue mimics the phosphorylated D53 at the site of phosphorylation in

phosphorylated RR468 (P~RR468) (Casino et al., 2009; Knaggs et al., 2007). Because the conformations of the two end states (active and inactive) are known for RR468, an excellent model system is now available to computationally address some key questions related to the conformational transition in RR468 in the presence or absence of phosphorylation. The underlying hypothesis being that a simulation of the active form, without the phosphoryl group or metal ion, should enable the system to adopt the conformation of the inactive form. We also wanted to understand the energetic contributions associated with the event of phosphorylation that produce a conformational change by lowering energetic barriers associated with various local rearrangements.

The active and inactive conformations of RR468 differ most significantly in the region adjacent to the site of phosphorylation. We will refer to it as the proximal region (Figure 2.1 a). There are five loops towards the proximal side. Three proximal loops surround the site of phosphorylation closely: $\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$, and $\beta 5$ - $\alpha 5$ loops. The difference between active and inactive conformations is most conspicuous in the $\beta 3$ - $\alpha 3$ loop (Figure 2.1 b). This is the longest loop in the proximal end (residues 54 to 59) and is closest to the site of phosphorylation, D53 (Figure 2.1 b, c). In the HK853-RR468 complex the $\beta 3$ - $\alpha 3$ loop of RR468 has contacts with two regions of HK853: the $\alpha 4$ - $\beta 4$ loop and the ATP-lid region of the CA domain (Casino et al., 2009). Due to these contacts, D53 of RR468 is lined up with H260 of HK853 in the complex, favoring the phosphoryl transfer reaction. Residue M56, in the middle of the $\beta 3$ - $\alpha 3$ loop, is known to be crucial for complex formation in this and several other HK-RR cognate pairs (Casino et al., 2009). M56 is buried inside a hydrophobic pocket in RR468 consisting of V8, L52 and V64 in the active conformation whereas it is solvent exposed in the inactive conformation. D60 is the

first residue in the $\alpha 3$ helix immediately following the $\beta 3$ - $\alpha 3$ loop. It interacts with the N-termini of the $\alpha 3$ and $\alpha 2$ helices in the active and inactive conformations, respectively (Figure 2.1).

The RR468 $\beta 4$ - $\alpha 4$ loop interacts with DHp-CA linker residues in the HK853-RR468 complex crystal structure. Highly conserved T83 is the last residue the $\beta 4$ strand. This residue is known to be important for neutralizing or shielding the negative charge on the phosphoryl group and assisting phosphoryl transfer to a water molecule (Knaggs et al., 2007). Hence, in the active conformation, the T83 side chain flips to an inward orientation to interact with the phosphoryl group on P~D53, where it is outward towards the exterior of the protein in the inactive conformation.

The last, crucial RR468 loop $\beta 5$ - $\alpha 5$ clamps the $\alpha 1$ -helix from the DHp domain of HK853 with the $\alpha 1$ helix of RR468 in the HK853-RR468 complex. K105 is the first residue in this loop. In the active conformation, K105 interacts with the highly conserved D9 ($\beta 1$) and the phosphoryl group on D53 (or phosphoryl group mimic BeF_3^-). The sidechain of K105 has two alternative conformations in the crystal structure of the inactive form: one outward and another inward relative to the site of phosphorylation. In the inward conformation, the K105 sidechain fills the void space of the phosphoryl group, hence close to the D53 sidechain, and the salt bridge between K105 and D9 is missing in this conformation.

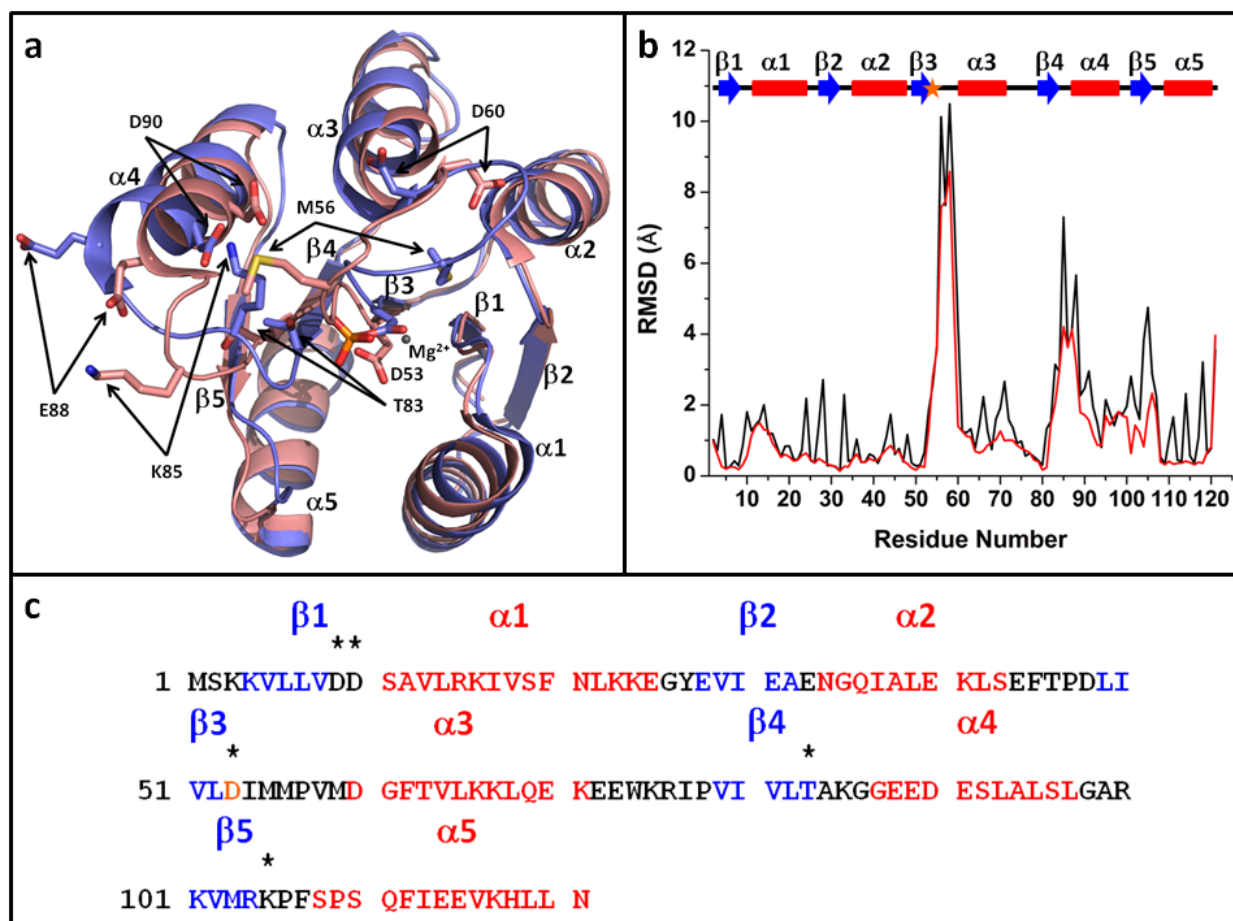


Figure 2.1. **A comparison between the inactive and active conformations of RR468.** Comparison between the inactive (PDB id 3DGF in pink) and active conformations of RR468 (PDB id 3GL9 in blue) along with the residues that differ most significantly in conformation, orientation and interaction in the active and the inactive form of RR468. (b) The RMSD between active and inactive form crystal structures of all heavy atoms (black) and backbone heavy atoms (red), and (c) the sequence of RR468. The α -helices and the β -strands are shown in red and blue respectively. The site of phosphorylation is D53, which is the last residue of β 3 (shown in orange in (b) and (c)). In (c), the residue positions marked with ‘*’ are highly conserved residues in related response regulator proteins.

In this work, structural determinants of the active and inactive states and the transition between them are proposed using, respectively, equilibrium molecular dynamics (MD) and Targeted MD (TMD) simulations. TMD has proved to be a useful tool to identify high-energy intermediate conformational states along a transition path that are difficult to identify using common structural biology techniques (Huang et al., 2009). MD simulations together with TMD simulations were shown to be very useful to investigate the molecular basis for the conformational transition in various response regulator proteins (Formanek et al., 2006; Knaggs et al., 2007; Lei et al., 2009; Ma and Cui, 2007; Roche et al., 2002; Tatsis et al., 2009). In this work, how a local event like the phosphorylation or dephosphorylation of a particular amino acid cascades down to other local conformational rearrangements, producing a global conformational change in RR468 is investigated by analyzing the MD simulations of the inactive and active forms, and the TMD simulations connecting them, using protein structure network (PSN) (Seeber et al., 2007) and hydrogen bond analyses. NMR experiments suggest that the conformational transition in RR468 takes place on μ s to ms time scale (unpublished data). This necessitates the introduction of TMD to span the transition. From the TMD trajectories, there are key residues, M56, D60, and T83, as noted above, which undergo substantial conformational changes. For these residues, we use biased MD simulations, started from various points along the transition path, to estimate their free energy barriers to rearrangement from the potentials of mean force (PMFs) along relevant reaction coordinates.

With mobile loops in the active and inactive forms, it could be that the inactive conformation would sample active conformations, and vice versa, but that overlap would not be accessible within the MD timescale. To see if information about the active conformation is encoded in the inactive sequence, and vice versa, we use principle component analysis (PCA) (Amadei et al.,

1993; García, 1992; Jolliffe, 2004) to find the major modes of motion of the inactive and active MD trajectories. Then, these modes are used to probe whether the mobile loop fluctuations of one form point in a direction to sample the conformations of the other form.

In Section 2.3, we present the MD and TMD simulation details, along with the hydrogen bond and PSN analysis methods, and the construction of the free energy (potentials of mean force) for selected residue conformational transitions. Two methods for using PCA modes to investigate the encoding behavior of inactive and active forms are presented. Section 2.4 gives the results of the hydrogen bond and PSN analyses of the MD and TMD trajectories, and the PCA-based analysis. The PMFs of the critical residues for the transition between protein forms are shown. Our results are discussed in Section 2.5 to indicate which residues are mainly responsible for the conformational transition. Section 2.6 presents our conclusions.

2.3. Methods

2.3.1. MD simulation of RR468 without phosphoryl group and Mg^{2+} . The inactive conformation of RR468 (PDB id 3DGF) was simulated (designated as IAC) using Amber10 (Case et al., 2008) in the presence of explicit water molecules (See Table 2.1 for a list of all simulations with their designators). The ff99SB force field was used for the simulation. Water molecules in the crystal structure were removed, hydrogen atoms were added, and the system was neutralized by addition of an appropriate number of Na^+ ions. The system was solvated with a 12 Å layer of TIP3P waters in a rectangular water box using tleap. The system was minimized first using a combination of steepest descent and conjugated gradient methods for 8,000 steps. A weak positional restraint (2 kcal/mole) was applied on all protein atoms during this minimization step. The whole system was minimized again using a combination of steepest descent and conjugated gradient methods for 20,000 steps without any positional restraint on any atoms. A time step of 2 fs was used for all subsequent heating, equilibration and production runs with the SHAKE option on all bonds containing H-atoms. Langevin dynamics was used for temperature control in the heating, equilibration and production steps. The minimized system was heated from 0 to 300 K in 1 ns. Weak positional restraint (2 kcal/mole) on all protein atoms were applied during the heating cycle. Constant pressure equilibration was done at 300K for 2 ns without any positional restraints. Finally, a trajectory of 100 ns was generated during a production run.

Root mean square deviation (RMSD) from the crystal structure was obtained along the time course of simulation and root mean square fluctuation (RMSF) for backbone heavy atoms for each residue was obtained using the Ambertools 10 ptraj module (Cheatham, 2008). All

individual frames were aligned on the first frame along the C α atoms, excluding the three mobile loop regions β 3- α 3, β 4- α 4, and β 5- α 5.

The active conformation of RR468 was simulated for 100 ns without phosphate (or BeF $^{3-}$) and Mg $^{2+}$ ion to investigate whether or not a conformational transition from active to inactive conformation takes place during this time scale. The explicit solvent simulation was done in the same way as outlined above.

Table 2.1. **A summary of molecular dynamics (MD) and targeted molecular dynamics (TMD) simulations along with the acronyms used in this chapter.** The results and conclusions were derived from one particular trajectory for each MD and TMD simulation unless exclusively mentioned in the text.

Acronym	Starting conformation	Target conformation	Method	Length (ns)	Number of simulations
IAC	Inactive conformation	-	MD	100	3
P~RR468	Active conformation with P~D53 and Mg ²⁺	-	MD	100	1
ACT	Active conformation without PO ₃ ²⁻ and Mg ²⁺	-	MD	100	3
I2A	Inactive conformation	Active conformation	TMD	20	3
A2I	Active conformation without PO ₃ ²⁻ and Mg ²⁺	Inactive conformation	TMD	20	3

2.3.2. MD simulation of active form RR468 (P~RR468). The crystal structure of RR468 (3GL9) in its active conformation has a phosphoryl group mimic, BeF_3^- , covalently attached to the aspartate (D53) residue. An Mg^{2+} ion is present in the active site. We replaced BeF_3^- with phosphoryl group (PO_3^{2-}). An *ab initio* quantum chemical calculation was performed on a phosphorylated aspartate residue using Density Functional Theory, with the B3LYP functional and 6-31G* basis set in Gaussian 2003 (Frisch et al., 2003). A system comprised of phosphorylated aspartate capped with Acetyl (Ace) and N-methyl (Nme) groups at the N and C-terminals, respectively, was used for this calculation (Figure A1, Appendix A). The Mg^{2+} ion, three water molecules, and an acetate ion were considered as a part of the system, resulting in a net charge -1 for this calculation. One of these water molecules mimics the backbone carbonyl group of M55, and the acetate ion mimics D10, which interacts with the Mg^{2+} ion, as in the crystal structure. The Gaussian input structure was built based on all heavy atom positions in the crystal structure (3GL9). MD charges on each atom in the phosphorylated aspartate residue were obtained by processing the Gaussian output using Antechamber (Wang, 2008). The other force field parameters for phosphorylated aspartate were obtained using the parmchk program in the Amber10 suite. Explicit solvent simulations were done using the same protocol as described above for the simulation of the inactive conformation.

2.3.3. Targeted Molecular Dynamics simulations. To simulate a transition pathway between the inactive to active (I2A) structures, targeted Molecular Dynamics (TMD) was performed, starting from the equilibrated structure in the simulation of the inactive conformation. The target reference structure was the crystal structure conformation of the active form of the protein (PDB id 3GL9). In this method, a harmonic restraint in the mass-weighted RMSD between the initial and target structures is introduced that can, over a suitably slow transformation, provide a

reasonable set of intermediate conformations between the initial and target structures. All C α atoms were best fit on the target reference structure and the RMSD was calculated for all heavy atoms in the protein. A force constant of 8 kcal/mol was used for the TMD. A similar transition pathway along the active to inactive (A2I) conformation was simulated starting from the equilibrated structure in the simulation of the active conformation. The target structure was the crystal structure (PDB id 3DGF) conformation of the inactive form of the protein in this case. The simulation length was 20 ns for each TMD trajectories. Two additional sets of 20 ns independent transition pathways were simulated for I2A and A2I to reach a consensus among the transition path. Slightly different conformations, after equilibration in the IAC and ACT simulations, were used as starting structures in these TMD simulations.

2.3.4. Principal Component Analysis (PCA). Methods based on principal component analysis (PCA) were used to investigate conformational transitions of the $\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$, $\beta 5$ - $\alpha 5$ loops in RR486 in IAC and ACT simulations. PCA diagonalizes the covariance matrix,

$$\mathbf{C} = \frac{1}{T} \int_0^T \delta \mathbf{R}(t) \delta \mathbf{R}^T(t) dt = \langle \delta \mathbf{R}(t) \delta \mathbf{R}^T(t) \rangle \quad (1)$$

of the atom fluctuations from their trajectory average values, $\delta \mathbf{R}(t) = \mathbf{R}(t) - \langle \mathbf{R}(t) \rangle$.

Here, $\mathbf{R}(t) = (x_1(t), y_1(t), \dots, z_N(t))^T$ denotes configurations along the trajectory and $\langle \dots \rangle$ denotes a time average over trajectory snapshots. The trajectory fluctuations are expressed as

$$\delta \mathbf{R}(t) = \sum_{\nu=1}^{3N} [\delta \mathbf{R}(t) \cdot \mathbf{m}_{\nu}^T] \mathbf{m}_{\nu} \equiv \sum_{\nu=1}^{3N} \delta p_{\nu}(t) \mathbf{m}_{\nu} \quad (2)$$

with the $\mathbf{m}_{\nu}$ ($\nu=1, 2, \dots, 3N$) the (orthonormal) eigenvectors and λ_{ν}^2 the corresponding eigenvalues of the covariance matrix $\mathbf{C}$. The principal component $\delta p_{\nu}(t)$ is the projection of the

trajectory onto the v 'th eigenvector. In the rotated Cartesian coordinate basis defined by the $\mathbf{m}_v$, the largest eigenvalue λ_1^2 captures the largest fraction of the mean square fluctuation (MSF), the second largest the next largest fraction of the MSF, etc. In applying PCA, before diagonalization of the covariance matrix, the overall translational and rotational motion of the protein has to be removed. We do so in the present context by superimposing all trajectory snapshots onto the C α atoms of the crystal structure, except for those in $\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$, and $\beta 5$ - $\alpha 5$ loops. That helps to isolate the fluctuations of the least stable parts of RR486.

PCA carried out on protein MD trajectories often show that their fluctuations correspond to many small amplitude, one-basin oscillatory modes along with a small number of large fluctuation and, typically, non-harmonic modes. If this separation held for the inactive protein and the amplitudes of all the modes were artificially increased, then it would be a remote possibility that active conformations would result or *vice versa*. Note that for PCA (along with Normal Mode Analysis) only directions of motion are found, their amplitudes are not determined. On the other hand, if it is true that active conformations are encoded in the inactive fluctuations, but they are not seen in the MD trajectory because the trajectory is not run for a sufficiently long time, then it may be that some set of the large “productive” modes from the inactive PCA will point toward active conformations. If these properties are manifest in the trajectory then a small subspace contained in the covariance matrix $\mathbf{C}$ characterizes the important inactive motions and that subspace is the relevant one for active-directed motion. Based on these considerations, we present two methods that are useful for investigating if the fluctuations around one structural basin can point to the other structure’s conformational basin.

2.3.5. DEVIATION Method. To explore the possibility that the IAC trajectory that is started from the inactive form crystal structure (here denoted as initial) encodes information about the

active conformation (here denoted as final), we use a previously developed method (Cukier, 2009) that projects the final structure onto the PCA modes of the initial structure. Thus, express the difference $\Delta\mathbf{X}$ between final and initial conformations in the initial state MD trajectory PCA basis:

$$\Delta\mathbf{X} = \mathbf{R}^{final} - \mathbf{R}^{initial} = \sum_v^{3N} (\Delta\mathbf{X} \bullet \mathbf{m}_v^T) \mathbf{m}_v \equiv \sum_v^{3N} a_v \mathbf{m}_v \quad (3)$$

In this fashion, the active conformation is expressed in the inactive PCA basis. If the inactive fluctuation directions that are expressed by the $\mathbf{m}_v$ basis do encode the active direction then the a_v expansion coefficients in equation 3.1 should track the λ_v coefficients of the initial trajectory

PCA. Equivalently, if this encoding hypothesis is true, then the function $\Delta\mathbf{X} - \Delta\mathbf{X}^{M_v} \equiv \sum_{v>M_v}^{3N} a_v \mathbf{m}_v$

should also decay rapidly with v (again, on the same scale as the λ_v). A convenient (normalized) measure to use is

$$\begin{aligned} \text{DEV} &= (\Delta\mathbf{X} - \Delta\mathbf{X}^{M_v})^T \bullet (\Delta\mathbf{X} - \Delta\mathbf{X}^{M_v}) / (\Delta\mathbf{X})^T \bullet (\Delta\mathbf{X}) \\ &= (\Delta\mathbf{X} - \Delta\mathbf{X}^{M_v})^T \bullet (\Delta\mathbf{X} - \Delta\mathbf{X}^{M_v}) / \text{MSD} \end{aligned} \quad (4)$$

2.3.6. EXTEND Method. Another approach to see if e.g. the inactive trajectory fluctuations point in the active form direction can be formulated. Consider the set of all inactive trajectory snapshots written in the basis of M_v PCA modes. Again, a small set of modes are used to eliminate the small unproductive fluctuations from the trajectory. For each snapshot introduce a multiplier, α ($-\infty$ to $+\infty$) and construct the following mean square metric:

$$\begin{aligned} f^{M_v}(\alpha(t)) &= (\mathbf{R}^{in} + \alpha \delta \mathbf{R}^{M_v}(t) - \mathbf{R}^{ac}) \bullet (\mathbf{R}^{in} + \alpha \delta \mathbf{R}^{M_v}(t) - \mathbf{R}^{ac}) \\ &= (\alpha \delta \mathbf{R}^{M_v}(t) - \Delta\mathbf{R}) \bullet (\alpha \delta \mathbf{R}^{M_v}(t) - \Delta\mathbf{R}) \end{aligned} \quad (5)$$

Scan α on some grid to measure the overlap with the active conformation, for each snapshot, and find the snapshot that minimizes $f^{M_v}(\alpha(t))$. Then, the snapshot that can be extended to be as close as possible to the active conformation has been generated. The EXTEND method is more objective than the DEV method because the active form is only used to monitor the extent to which the inactive form extension overlaps with the active form. There is certainly no reason for closer approach to the active conformation with the addition of more modes in $\delta\mathbf{R}^{M_v}(t)$ as in the DEV method. To actually accomplish this minimization, the derivative $\partial f^{M_v}(\alpha|t)/\partial\alpha = 0$ leads to

$$f^{M_v}(\alpha_{\min}(t)) = \Delta\mathbf{R}^2 \sin^2 \theta(t) \quad (6)$$

with

$$\cos(\theta(t)) = \frac{\Delta\mathbf{R} \bullet \delta\mathbf{R}^{M_v}(t)}{\|\Delta\mathbf{R}\| \|\delta\mathbf{R}^{M_v}(t)\|} \quad (7)$$

the cosine of the angle between the indicated vectors and

$$\alpha_{\min}(t) = \cos(\theta(t)) \frac{\|\Delta\mathbf{R}\|}{\|\delta\mathbf{R}^{M_v}(t)\|} \quad (8)$$

It is straightforward to show that the second derivative, $\partial^2 f^{M_v}(\alpha|t)/\partial\alpha^2 \Big|_{\alpha_{\min}} > 0$ thus a minimum results. Then, searching all the times (snapshots) for the $t = t_{\min}$ for which $f^{M_v}(\alpha_{\min}(t))$ is minimized produces the smallest deviation from the active form. This value of α is used to generate the extended protein conformation from the inactive trajectory. It is guaranteed to be as close as possible in the mean square sense to the active conformation based on a one-parameter measure.

In a similar manner, the PCA analysis was carried out using the 100 ns MD trajectory of the active conformation (ACT) of the APO-protein and projected on the crystal structure of the inactive conformation.

2.3.7. PMF construction using WHAM. When using a window method to obtain a potential of mean force (PMF) for some reaction coordinate q , the trajectory data for the restrained (biased) windows need to be combined. The Weighted Histogram Analysis Method (WHAM) (Kumar et al., 1992; Souaille and Roux, 2001) combines such data from the N different bias window potentials by expressing the true, unbiased estimated probability density $\rho^{(u)}(q)$ along q as a linear combination of the window biased probability densities

$$\rho_w^{(b)}(q) \rho^{(u)}(q) = \sum_{w=1}^N c_w \rho_w^{(b)}(q) \quad (9)$$

The coefficients, c_w in this linear combination, found by minimizing the statistical error of the density estimation along the reaction coordinate, lead to the iterative determination of the window free energies and are used to construct

$$\text{PMF}(q) = -k_B T \ln \rho^{(u)}(q) \quad (10)$$

using a program developed by Grossfield (Grossfield, 2003). A few snapshots along the transition path in the I2A TMD trajectory were used directly as endpoints for biased window simulations for local transitions in M56, D60, and T83.

2.3.8. Conformational clustering of the transition path using Wordom. Changes in interactions among different residues in RR468 along the TMD trajectory were studied using the Protein Structure Network (PSN) analysis in Wordom (Seeber et al., 2007; Seeber et al., 2010). The TMD trajectory was segregated into 4 conformational clusters based on Distance RMS

(DRMS) of all C_α atoms in the $\beta 3-\alpha 3$, $\beta 4-\alpha 4$, and $\beta 5-\alpha 5$ loops with respect to that in the active conformation of RR468. The DRMS between two structures (a and b) is defined as the square root of the averaged squared distance deviations (d_{ij}) between pairs of atoms i and j , i.e.

$$DRMS = \sqrt{\frac{1}{N} \sum_{ij}^n (d_{ij}^a - d_{ij}^b)^2} \quad (11)$$

The DRMS for n atoms, thus, takes into account the fluctuations of $N = n(n-1)/2$ distance pairs. Clustering was done using the Leader algorithm in Wordom. A DRMS cutoff 1.65 Å was used for clustering.

2.3.9. Change of pairwise interaction along transition path using Wordom. A PSN analysis was done on each individual conformational cluster obtained from the I2A TMD trajectory (Seeber et al., 2010). The PSN output was compared for two subsequent clusters to determine interactions formed/broken during transition. The interaction strength (I_{ij}) between residue pair i and j , is defined as

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} 100 \quad (12)$$

Here I_{ij} is the interaction percentage of residue i and j , n_{ij} is the number of side-chain atom pairs within a given distance cutoff (4.5Å as a default), and N_i and N_j are, respectively, the normalization factors for residues i and j , that take into account the differences in size of the different residues as larger residues are likely to make more contact pairs, n_{ij} (Seeber et al., 2010). This interaction strength provides a quantitative estimate of the overall interaction between two residues based on their proximity. In our experience it is a better estimate of the hydrophobic interaction because n_{ij} will be higher in the case of hydrophobic interactions compared to polar interactions, the latter being strongly directional and involving fewer atoms.

Pair wise interactions with interaction percentage 2.0% (I_{min}) in two consecutive clusters were compared. If a given pair of residues interacts with strength ($\geq 2.0\%$) in a significant number of snapshots (frequency $\geq 50\%$) in one cluster has significantly lower frequency (frequency $\leq 20\%$) in the next cluster it was considered that the interaction has broken for the transition between one cluster and the next cluster. Similarly, residue pairs with interaction frequency less than 20% in one cluster but more than 50% in the next cluster were considered to represent the formation of an interaction.

2.3.10. Hydrogen Bond Analysis. Minimized conformations of the inactive and active crystal structures of RR468 were compared for their hydrogen bonding patterns using the “ptraj” module in AmberTools 10. Also hydrogen bond analyses for the four independent trajectories, IAC, ACT, and I2A and A2I were done using ptraj. In the ptraj output for hydrogen bond analysis, all hydrogen bonds with percentage occurrences more than 1% were listed. Distance cutoff 3.5 Å (between two heavy atoms involved in the hydrogen bond) and angle cutoff 120° (heavy atom-hydrogen-heavy atom) was used to detect the hydrogen bonds. Multiple ptraj outputs were analyzed together using a script to extract meaningful information. All the hydrogen bonds in the crystal structures protein can be classified in 3 categories: hydrogen bonds that are present only in the inactive conformation, only in the active conformation and hydrogen bonds that are there in both conformations. Hydrogen bonding patterns in two consecutive clusters along the I2A transition path (as obtained using Wordom) were analyzed to obtain changes in hydrogen bonding patterns along the transition path. The hydrogen bond analysis was done for each TMD cluster and the formation and dissociation of characteristic hydrogen bonds (of the inactive and active conformations) were compared in two consecutive clusters. The hydrogen bond analysis in each TMD cluster was obtained using ptraj and the change in

hydrogen bonding pattern in two consecutive TMD cluster was obtained. The same distance and angular criterion for hydrogen bonding was used as before. A percentage occurrence of less than 15% was considered as absent and over 40% was considered as present.

2.4. Results

MD trajectories of Apo-RR468 were generated starting from the inactive conformation (PDB id 3DGF) and active conformation (PDB id 3GL9) without BeF_3^- attached to D53 and without a Mg^{2+} ion. A phosphorylated RR468 complex was obtained by modifying BeF_3^- to PO_3^{2-} . All three systems, the inactive, apo-RR468 denoted as IAC, the active without BeF_3^- and Mg^{2+} denoted as ACT, and active with BeF_3^- and Mg^{2+} denoted as P~RR468, (see Table 2.1 for a summary of the simulations) were simulated for 100 ns in the presence of explicit water molecules. Since root-mean-square deviation (RMSD) versus time traces reached a plateau during the 5 ns equilibration cycle (not shown) prior to the 100 ns production simulation, the whole 100 ns segments of the three simulation trajectories were considered equilibrated and used for further analysis (Figure 2.2 a). Backbone root-mean-square fluctuation (RMSF) during the production runs show the inactive conformation of RR468 is more fluctuating than that of the active conformation with or without the phosphoryl group and the Mg^{2+} ion. Apart from N and C-terminal residues, fluctuations are greater in the loop regions, especially in three proximal loops where the inactive and active conformations differ significantly (Figure 2.2 b).

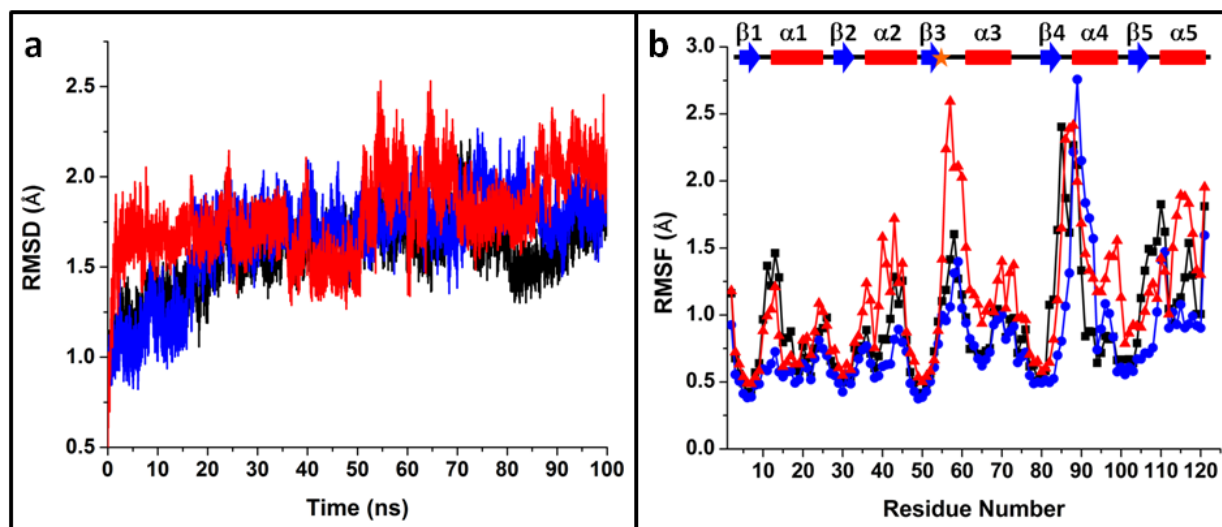


Figure 2.2. **RMSD and RMSF during equilibrium simulations.** Backbone RMSD versus time (a) and backbone RMSF (b) plots for the 100 ns simulations starting from the inactive conformation (red), active conformation with PO_3 and Mg^{2+} (black), and active conformation without PO_3^{2-} and Mg^{2+} (blue). It is evident from the RMSF plot that the secondary structural elements in RR468 remain stable during all MD simulations.

2.4.1. Pairwise interactions among residues in the crystal structures are maintained in the MD trajectories. The protein structure networks in the crystal structures of RR468 and in the MD trajectories ACT and IAC were studied using Wordom (Seeber et al., 2010). Protein Structure Network (PSN) analysis on a given segment of trajectory provides a pairwise interaction map among all residue pairs. For a given pair of residues it provides in how many snapshots the interaction strength, I_{ij} , as defined in the Methods section, is greater than a cutoff, I_{min} . Intra and inter loop interactions among residues in the three proximal mobile loops are summarized in Figure 2.3 (a, b). All the common interactions among the $\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$, and $\beta 5$ - $\alpha 5$ loops, as observed in the crystal structures of RR468, are stable during MD simulation (Figure 2.3 c, d). Interaction between D53 and T83 in the active conformation and the ACT trajectory is due to orientation of the T83 sidechain towards the site of phosphorylation (D53) in the active form. In the active conformation, M56 is buried in a hydrophobic pocket. In this conformation, M56 is in contact with D53. This interaction is stable in the ACT simulation, but in the IAC simulation M56 makes weak contact with other loop residues. All the interactions involving the $\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$, and $\beta 5$ - $\alpha 5$ loops along with one residue each following and preceding these loops are summarized in Table A6, Appendix A. For the sake of simplicity, interactions not there in the crystal structures but formed during simulations were ignored in this analysis.

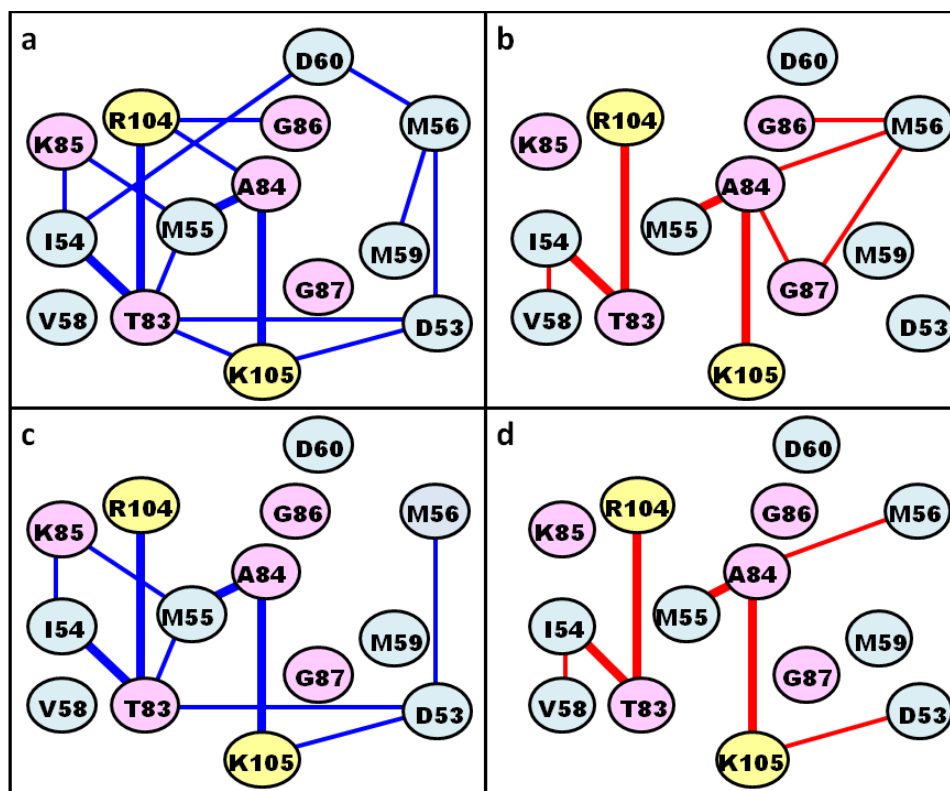


Figure 2.3. **Intra and inter loop interactions of the proximal loop residues.** One residue adjacent to the loops in each N and C-terminus of these three loops were also considered as part of the loop for their contacts in the crystal structure of the (a) active form, (b) inactive form and corresponding simulation trajectories for (c) ACT, and (d) IAC simulations. All interactions in the active and inactive form crystal structures are shown with blue and red connectors, respectively. Intra and inter loop interactions in the ACT and IAC simulations are shown with same color code in (c) and (d), respectively. Interactions shown in (c) and (d) have interaction strength more than 5.0% in more than 30% of the total number of snapshots during MD trajectories. Common interactions that are present in both crystal structures or during simulations are shown with bold connectors in all the interaction maps. Residues from $\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$, and $\beta 5$ - $\alpha 5$ loops are shown as cyan, pink, and yellow circles, respectively.

2.4.2. Hydrogen bond comparison of the crystal structures and MD trajectories. We first consider hydrogen bonding interactions in the crystal structures and simulation trajectories, ACT and IAC, starting from the respective crystal structure conformations in order to obtain endpoint patterns that will be connected later in the TMD simulations. All hydrogen bonds in RR468 can be classified in three categories: hydrogen bonds that are present only in the inactive conformation, only in the active conformation, and hydrogen bonds that are there in both. Obviously not all of these hydrogen bonds are stable in presence of explicit water molecules during the simulation and there are new hydrogen bonds formed during the equilibration process. The results of the hydrogen bond analysis based on the minimized crystal structures of the both forms of the RR468 are summarized in Table A1, Appendix A.

There are 14 hydrogen bonds that are identified in the inactive conformation but not there in the active conformation crystal structures. But only three of these hydrogen bonds are stable (percent occurrence more than 40%) during the IAC trajectory. These three hydrogen bonds are also somewhat stable in the ACT simulation as well. It can be concluded that these characteristic hydrogen bonds present in the inactive but not in the active conformation as identified from the crystal structure are not very stable during MD simulation.

There are 26 hydrogen bonds that are there in the active conformation but not in the inactive conformation. Two out of these 26 hydrogen bonds have comparable stability in the IAC trajectory as well (backbone-backbone hydrogen bond, T83-K105 and sidechain-sidechain hydrogen bond, R15-E31). Eight of these hydrogen bonds are stable only in the ACT trajectory but not in the IAC trajectory. The salt bridge between D9 and K105 brings the $\beta 1$ - $\alpha 1$ and $\beta 5$ - $\alpha 5$ loops closer. The N-termini of all 5 α -helices are proximal to the site of phosphorylation. The N-termini of the $\alpha 1$, $\alpha 2$, and $\alpha 3$ helices are more stable in the ACT simulation. This is evident from

the stability of the backbone hydrogen bonds for A12-K16 ($\alpha 1$), I37-K41 ($\alpha 2$), G61-K65 and T63-K67 ($\alpha 3$) in the ACT simulation to that compared to the IAC simulation.

There are 35 hydrogen bonds that are there in both the active and inactive conformations. Most of these common hydrogen bonds involve backbone interactions only and have comparable stability in both these trajectories.

2.4.3. PCA, DEVIATION and EXTEND show that the $\beta 3$ - $\alpha 3$ loop can partially transit between inactive and active forms in IAC and ACT simulations. During simulations starting from the inactive (IAC) and active (ACT) states, the trajectories fluctuate around their respective starting conformations in the sense that the flexible regions, the three crucial loops ($\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$, and $\beta 5$ - $\alpha 5$), do not, on the 100 ns MD time scale, sample each other's conformations. It could be that, with mobile loops in the active and inactive forms, an overlap of loop conformations would occur over times beyond MD practicality. To monitor the relative motions of the $\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$, and $\beta 5$ - $\alpha 5$ loops, representative distances between residues were chosen in such a way that the difference in their distances were maximized between the active and inactive conformations in the crystal structures (Table 2.2). The M56-N34, V58-G87 and V58-P106 distances maintain their characteristic values during the simulations starting from the active conformation with or without phosphoryl group and from the inactive conformation (Figure 2.4). The possibility of any rearrangement spanning the other's conformation in the $\beta 3$ - $\alpha 3$ loop was ruled out using this analysis. Since the difference in distance among any such residue pairs from the $\beta 4$ - $\alpha 4$ and $\beta 5$ - $\alpha 5$ loops are small in the active and inactive conformation, the possibility of any conformational transition among backbone atoms in three mobile proximal loops were studied based exclusively on the position and relative orientation of $\beta 3$ - $\alpha 3$ with respect to the

core of the protein and with respect to the other two mobile proximal loops: $\beta 4$ - $\alpha 4$ and $\beta 5$ - $\alpha 5$. Note that the stable (non-loop) residues are very closely aligned in the two structures (Figure 2.1 a b). Using these criteria, we do find that there is no loop conformational overlap in the 100 ns IAC and ACT MD trajectories.

Table 2.2. **Reaction coordinates for loop motions.** The relative orientation of the $\beta 3\text{-}\alpha 3$ loop with respect to the $\beta 4\text{-}\alpha 4$ and $\beta 5\text{-}\alpha 5$ loops in RR468 was monitored by tracking the distances between the $C\alpha$ atoms of the following residues: M56 and V58 from the $\beta 3\text{-}\alpha 3$ loop, G87 from the $\beta 4\text{-}\alpha 4$ and P106 from the $\beta 5\text{-}\alpha 5$ loop. N34 marks a very stable region in the protein, as clear from the RMSF plot in Figure 2.2. Residues in the $\beta 3\text{-}\alpha 3$, $\beta 4\text{-}\alpha 4$, and $\beta 5\text{-}\alpha 5$ loops are shown in bold.

Residue	Distance between $C\alpha$ atoms (Å)		
	Inactive	Active	Difference
M56-N34	13.1	5.8	7.3
V58-G87	9.1	23.4	14.3
V58-P106	17.8	23.0	5.2

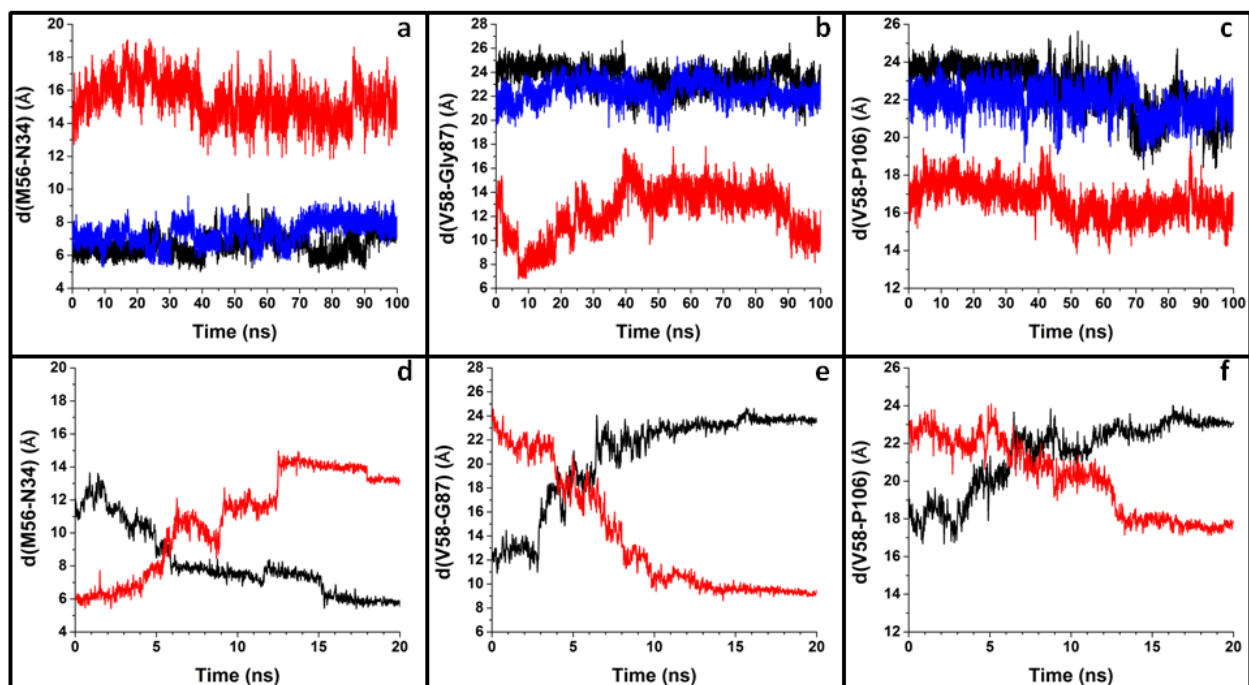


Figure 2.4. **Reaction coordinates for loop motions.** Top panel shows distances between C α atoms of (a) M56 and Asn34, (b) V58 and G87, (c) V58 and P106 residues in the IAC, ACT and P~RR468 trajectories (red, black and blue respectively). Bottom panel shows the same distances during TMD pathways I2A and A2I (in black and red respectively).

To see if this lack of connection is due to the physically short time scale that is practical with MD we use the DEVIATION and EXTEND methods (Sections 2.3.5 and 2.3.6) that are based on a principal component analysis of the trajectory data. PCA decomposes the atom fluctuations of a trajectory into modes that successively incorporate decreasing amounts of the total mean square fluctuation over the trajectory duration. If there is a separation in the sizes of the mode eigenvalues λ_v^2 with a small set of large eigenvalues and a large set of small eigenvalues, then those first few modes of large amplitude may account for most of the “productive” motion while the many, small amplitude modes may correspond to “noise”. PCA was carried out using the backbone heavy atoms after best fitting the trajectory snapshots onto the crystal structure on all CA atoms of the residues, excluding the loops, as summarized in Table 2.3. The first 10 and 20 eigenvalues out of 1440 eigenvalues obtained from the inactive trajectory account for 69.4% and 80.0% of the total fluctuation in the system, respectively. The fluctuations are relatively less in the active simulation, where the first 10 and 20 principle components obtained from active trajectory account for 60.3% and 72.0% of the total fluctuation in the system, respectively. Thus, in both simulations, a small number of modes do capture the bulk of the loop fluctuations.

Table 2.3. **PCA mode eigenvalues derived from the inactive and active trajectories.**

Trajectory	Mode	λ_v (Å)	$\lambda_v^2(\text{Å}^2)$	Cumulative fraction
IAC	1	0.48079	0.23116	0.23116
	10	0.14396	0.02073	0.69399
	20	0.08588	0.00738	0.79990
	30	0.06290	0.00396	0.85290
	47	0.04528	0.00205	0.90030
ACT	1	0.45916	0.21083	0.21083
	10	0.13094	0.01715	0.60257
	20	0.09096	0.00827	0.72003
	30	0.07115	0.00506	0.78383
	74	0.03758	0.00141	0.90085

As discussed in the Methods Section, if it were true that e.g. active conformations are encoded in the inactive trajectory fluctuations, but they are not seen in the MD trajectory because the trajectory is not run for a sufficiently long time, then it may be that some set of the large “productive” modes from the inactive trajectory PCA will point toward active conformations, and *vice versa*. The DEV method is one way to address this issue. It projects a final state structure onto the PCA modes of a trajectory based on an initial state structure. If the initial state trajectory does “point” in the direction of the final state structure, then DEV defined in Equation (2.4) should track the decay of the λ_v^2 eigenvalues from the initial state PCA. The scaling of DEV with λ_v^2 is shown in Figure 2.5. As noted above, there is good separation between a small number of the large PCA eigenvalues from the many small eigenvalues. The DEV measure modestly tracks the decay of the λ_v^2 indicating some degree of (loops) movement of the inactive towards the active form. Also displayed in Figure 2.5 is the backbone conformation (gray) of RR468 reconstructed from the first 50 PCA modes of the inactive trajectory, the equilibrated inactive conformation (red) and active conformation crystal structure (violet). Focusing on loop $\beta 3$ - $\alpha 3$, the predicted position is intermediate between the inactive and the active conformation, which is consistent with the modest agreement between DEV and the eigenvalue decays. Also displayed are snapshots at 2 (yellow), 4 (green) and 6 (blue) ns of the 20 ns TMD trajectory that was used to span the inactive to active conformations. The loop movement in DEV is between the 4 and 6 ns of the TMD trajectory, indicating that a reasonable direction of loop movement toward the active form is encoded in the inactive trajectory.

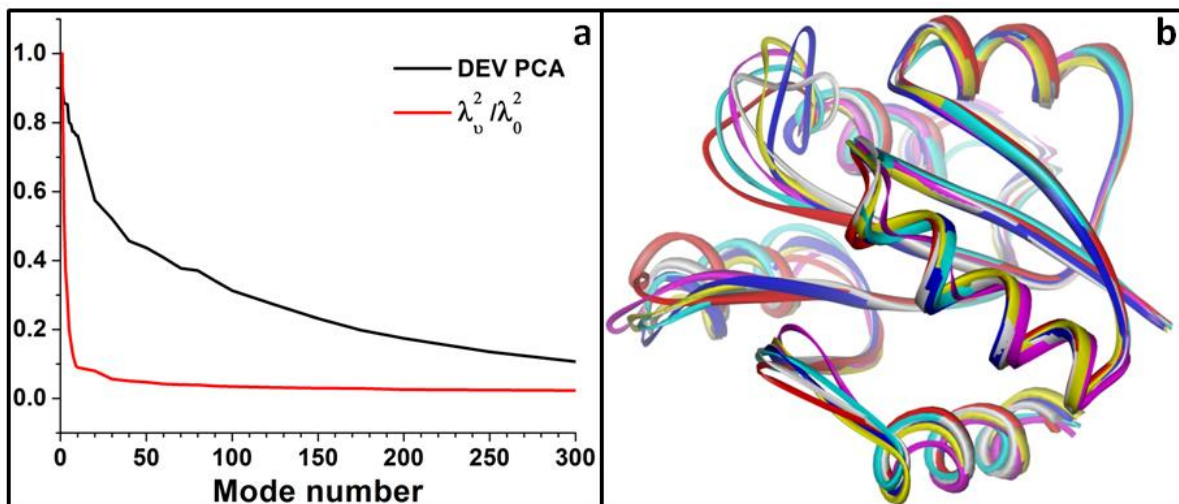


Figure 2.5. **DEV analysis.** (a) DEV that monitors the approach toward the inactive conformation as a function of the number of modes used in the expansion of the difference of equilibrated inactive conformation and active conformation crystal structures in the PCA basis, along with the decay of the inactive fluctuation eigenvalues, λ_v^2 . The decay of DEV with the number of modes is slower than that of the eigenvalues, showing that the inactive fluctuations modestly encode active conformations. (b) Backbone conformation (gray) of RR468 reconstructed from first 50 PCA modes in the inactive trajectory compared with the equilibrated inactive conformation (red) and active conformation crystal structure (blue). Conformational snapshots at 2, 4 and 6 ns of the TMD trajectory are shown in cyan, magenta, and yellow respectively. The reconstructed $\beta 3$ - $\alpha 3$ loop structure (top-left loop) shows that an intermediate position between inactive and active conformations can be achieved.

In the DEV method, as more modes are included, the reconstructed coordinates will eventually match the target conformation (active conformation in this case). Thus, a more objective method is desired. The EXTEND method only uses the active conformation to measure how well the inactive trajectory succeeds in matching its configuration when minimizing the RMS distance between them. Atomic coordinates of the backbone atoms reconstructed from the first few PCA-modes in the inactive trajectory were compared with the equilibrated conformation of the inactive form and the crystal structure of the active form. Similarly, atomic coordinates of the backbone atoms reconstructed from the first few PCA-modes in the active trajectory were compared with the equilibrated conformation of the active form and the crystal structure of the inactive form. In either case, the large amplitude fluctuations in a trajectory started with one end state indeed points in the direction that samples the other substate as well, as shown in Figure 2.6. Again, as in the DEV method, the $\beta 3$ - $\alpha 3$ loop distinctly points in a direction that, for the active trajectory, spans the inactive-to-active conformations, and *vice versa*.

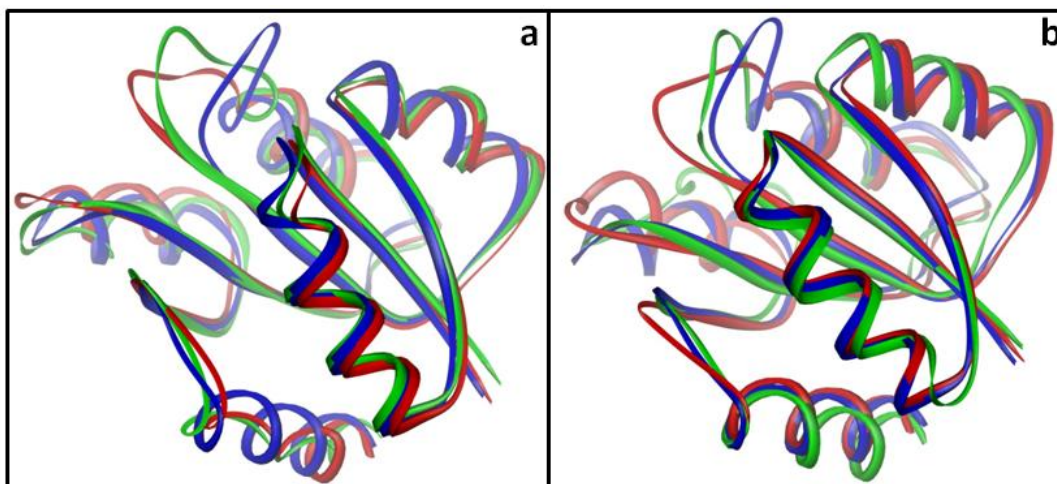


Figure 2.6. **EXTEND analysis.** (a) Equilibrated structure of the inactive (red) conformation and crystal structure of the active conformation (blue) of RR468 were aligned with the atomic coordinates reconstructed from first 10 PCA modes derived from the inactive trajectory, (b) Equilibrated structure of the active (blue) conformation and crystal structure of the inactive conformation (red) of RR468 were aligned with the atom coordinates reconstructed from the first 20 PCA modes derived from the active trajectory. Loop $\beta 3$ - $\alpha 3$ in the inactive trajectory points toward the active conformation and *vice versa*.

2.4.4. Targeted molecular dynamics (TMD) can span inactive to active, and vice versa, conformations. Starting from the inactive or active conformation of RR468, a conformational transition to the other form was not observed within the 100 ns equilibrium simulation (IAC or ACT). Thus, we applied TMD to find a possible transition path from the inactive to active conformation, or the reverse direction.

The convergence of beginning to final conformations in TMD was confirmed by monitoring three representative distances noted in Table 2.2, and from the positions and interactions of three residues M56, D60, and T83, that undergo significant rearrangement during transition (Figures 2.3 and 2.7). We repeated the TMD simulations two more times for I2A or A2I with slightly different starting points (from the equilibrated structures during MD simulations) to reach a consensus. The active and inactive conformations from the crystal structure were considered as target structures for the I2A and A2I trajectories, respectively. In all I2A trajectories, the $\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$, and $\beta 5$ - $\alpha 5$ loop conformations finally converge to that of the active form between 10-15 ns (Figure 2.3 d, e, and f respectively). Rearrangements in the M56, D60 and T83 side chains occur somewhere near to 15 ns in all three I2A transition paths (Figure 2.7 a, b, and c respectively). In the A2I trajectories (Figure 2.7 d, e, and f respectively) transitions are more-or-less a reflection of the I2A path. During conformational transitions in either direction, the M56 rearrangement follows multiple stable intermediate states but the D60 and T83 rearrangements are more drastic in nature (Figure 2.7).

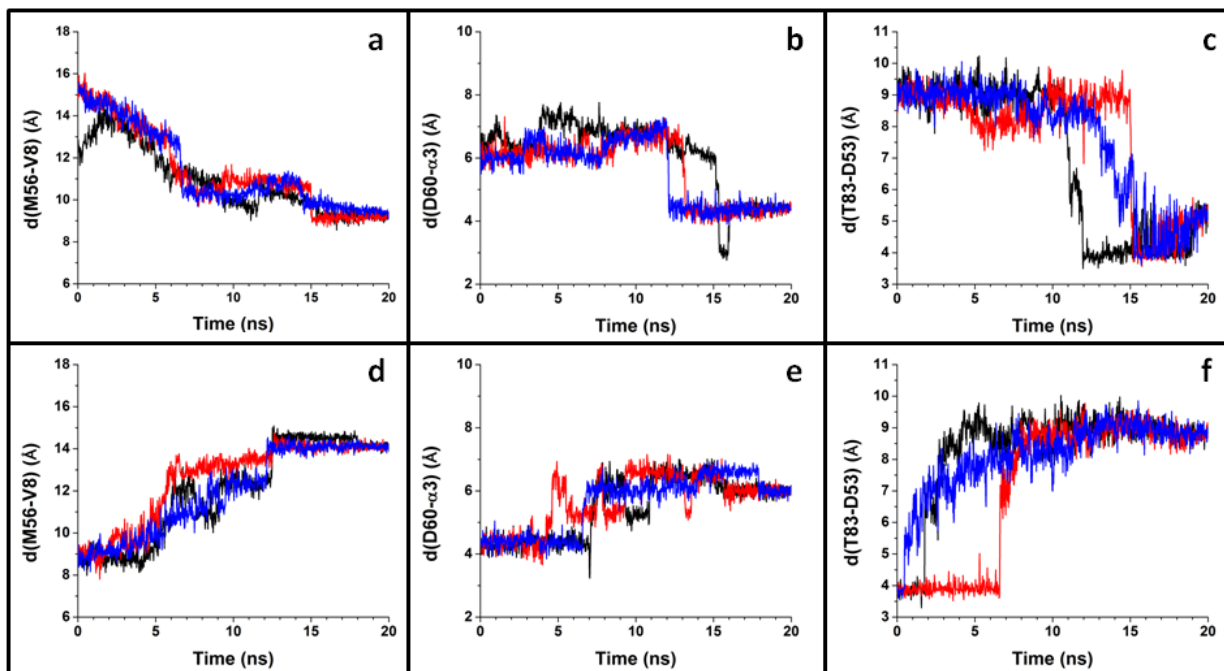


Figure 2.7. **Critical residue conformation or orientation during TMD simulations.** Time series of parameters that best describe the conformation and position of M56, D60 and Th83 in the I2A (a, b, c) and A2I TMD trajectories (d, e, f): Distance between C α atoms of M56 and V8 in a hydrophobic core (a, b), Average distance between C γ atom of D60 and backbone N atoms of three N-terminal residues of the α 3 helix (G61, F62, T63) (c, d) and Distance between the O γ atom of T83 and the C γ atom of D53 (e, f). Results from 3 independent TMD trajectories in either directions (I2A or A2I) are marked in 3 colors.

2.4.5. Potentials of Mean Force (PMF) for transitions of key residues in the TMD

trajectory. As shown in Figure 2.7, residues M56, D60 and T83 undergo significant rearrangements during the TMD trajectories, which were not observed during the MD simulations (Figure A2, Appendix A). These transitions were monitored by various atom distances. However, to construct potentials of mean force (PMF), more meaningful reaction coordinates that best describe these transitions needed to be introduced (Table 2.4). For M56 we use the distance between the C α atoms in M56 and V8. For this gradual M56 transition, endpoints for the PMF are 10.5 Å and 9.25 Å. To span the reaction coordinate, four windows were needed to obtain good window overlap (Figure A3, Appendix A). For D60 and T80, with their abrupt nature of rearrangement in the TMD trajectory, we had to perform brief steered molecular dynamics (SMD) spanning the reaction coordinates, from a characteristic value in the inactive conformation to that in the active conformation (details in Table A9, Appendix A). Starting conformations for the biased window simulations were obtained from these SMD trajectories. Because of the mainly side chain based rearrangements of these residues, we found that the T83 transition was best described by the indicated side chain atoms and the D60 transition best described by the indicated dihedral angle in Table 2.4. To obtain good window overlap with these reaction coordinates, eight and ten windows were needed for the construction of the PMF for the D60 and T83 rearrangements, respectively (Figure A3, Appendix A).

The PMF plots are shown in Figure 2.8. From the offset of the PMF value in the inactive conformation versus that in the active conformation, it is evident that the active conformation is a relatively high energy conformation. The energy barrier for the M56 rearrangement is relatively low compared to that for D60 and T83. Conformational rearrangement in M56 follows several

intermediate conformations whereas for D60 and T83 the rearrangements are very abrupt in nature (Figure 2.8 and Figure A3, Appendix A).

Table 2.4. **Reaction coordinates used for calculation of the PMF for local transitions.**

Local changes	Reaction coordinate	Acronym
M56 transition from exposed (inactive) to buried (active) state	Distance between $C\alpha$ atoms in M56 and V8	d(M56-V8)
Shift in D60 interaction with $\alpha 2$ in inactive to that with $\alpha 3$ in active conformation	Torsion angle $C\gamma$ - $C\beta$ - $C\alpha$ -C of D60	tD60($C\gamma$ - $C\beta$ - $C\alpha$ -C)
T83 flip from outward to inward	Distance between T83 $O\gamma$ and D53 $C\gamma$	d(T83-D53)

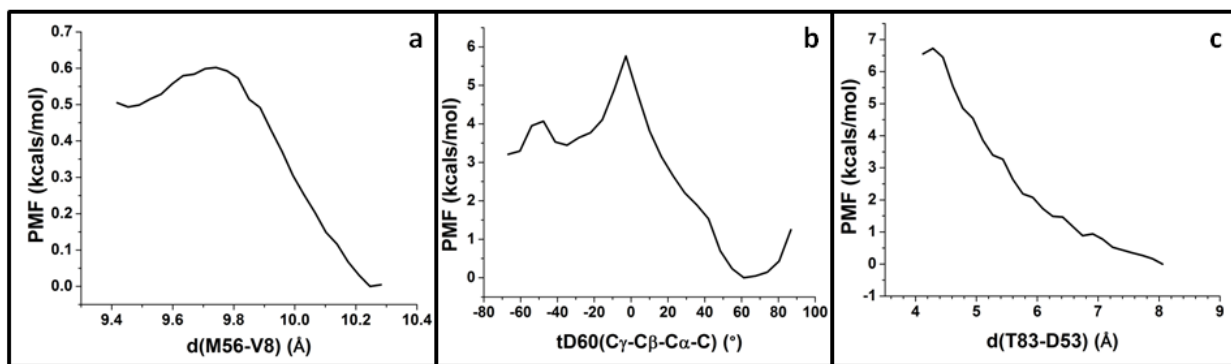


Figure 2.8. **PMF analysis.** Potentials of mean force along reaction coordinates that best describe local conformational rearrangements in (a) M56, (b) D60, and (c) T83 for the inactive to active transition. In all three plots, the reaction coordinates proceed from left to right during the inactive to active transition.

2.4.6. Hydrogen bond analysis of the TMD trajectory identifies residues critical to the conformational transition. In the TMD trajectories, several hydrogen bonds, which we will refer to as non-native, transiently exist during the transition path, and are absent in either of the end state simulations. The role of a particular non-native hydrogen bond was confirmed using the multiple TMD trajectories. To identify critical non-native hydrogen bonds that may assist conformational transitions in RR468, we selected hydrogen bonds with lower occurrences in both the ACT and IAC simulations (less than 20%) but relatively higher occurrence in the TMD trajectories (e.g. I2A). We have chosen a small cutoff for occurrences (10%) in TMD simulations for this purpose since conformational sampling is limited in the biased simulation. In the I2A TMD trajectory, 26 such hydrogen bonds were found that are present in at least 15 consecutive snapshots (lifetime 300 ps or more). Out of 26 hydrogen bonds identified 19 involve side chain interactions (Table A7, Appendix A).

Two notable non-native hydrogen bonds involve residues D60 and R104. Hydrogen bonds between the D60 side-chain and main chain N of G35 can be crucial for the D60 rearrangement. This hydrogen bond either never appears or has very rare appearances during the IAC and ACT simulations (Table A7, Appendix A). The occurrence of the hydrogen bond between D60 sidechain and G35 backbone N-atom is 54.0% in the I2A trajectory, but this hydrogen bond never forms during the IAC or ACT simulations (Figure 2.9 a). The same hydrogen bond is present in 29.3% and 24.4% of the total number of snapshots in the other two I2A trajectories. D60 forms similar non-native hydrogen bonds with adjacent residues (F62 and T63) prior to its local conformational rearrangement in the other I2A trajectories as well. Significant interactions of D60 with T63 were also observed in A2I TMD trajectory (occurrence 23.10 %) (Table A8, Appendix A). Transient hydrogen bonds formed by R104 with the $\beta 4$ - $\alpha 4$ loop residues and their

adjacent residues are likely to couple the $\beta 5$ - $\alpha 5$ and $\beta 4$ - $\alpha 4$ loop movements during the transition. This feature was observed in two out of three independent I2A trajectories. A salt bridge between R104 and E91 found in the A2I TMD trajectory is also crucial for the conformational transition in the reverse direction (occurrence 48.60 % in A2I) (Table A7, Appendix A). R104 remains solvated most of the time during the ACT and IAC equilibrium simulations.

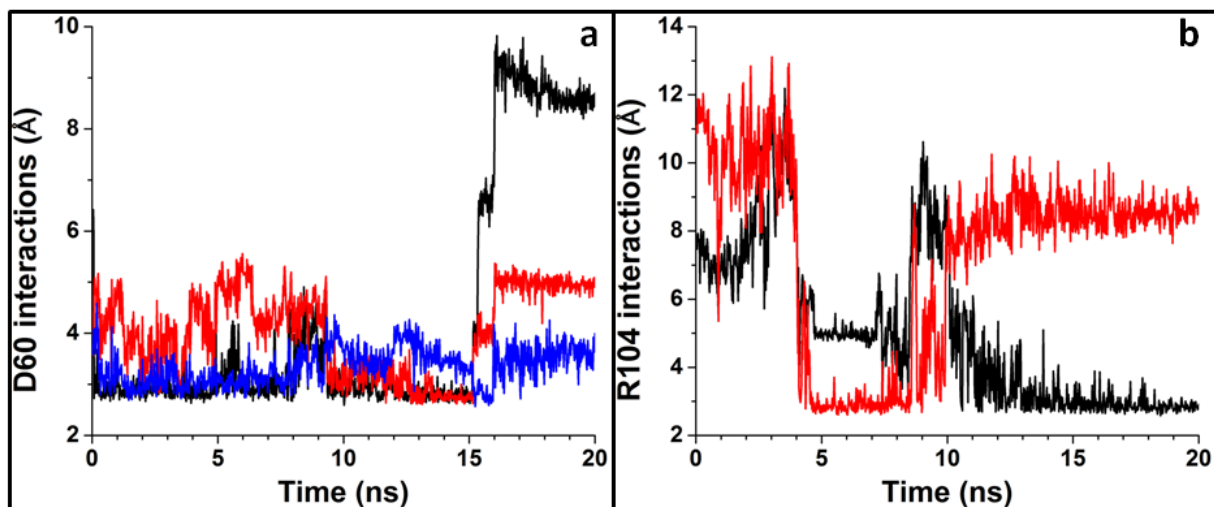


Figure 2.9. **Critical transient hydrogen bonds.** (a) Transient hydrogen bond formed by D60 O δ with backbone N of G35 (black), G61 (red), and D60 (blue) during the inactive to active TMD trajectory. (b) Hydrogen bonding of R104 with several residues in the β 4- α 4 loop is a significant feature of the I2A TMD trajectory.

2.4.7. Conformational clustering of the TMD trajectories permits identification of

intermediates. The above hydrogen bond analysis presents a global view over the entire TMD trajectories. A similar analysis based on conformational clustering along the TMD trajectory was carried out to identify changes in pairwise interactions among different residues. Conformational clustering of the TMD trajectories was based on the DRMS (Distance RMS) of all the C α atoms in the β 3- α 3, β 4- α 4, and β 5- α 5 loop regions to identify crucial intermediate states during the conformational transition (Figure 2.10). The 20 ns I2A TMD trajectory was segmented into 4 distinct clusters (0-2.88 ns, 2.90-4.02 ns, 4.04-9.22 ns, and 9.24-20.00 ns) based on the DRMS with respect to the active conformation. The β 3- α 3, β 4- α 4, and β 5- α 5 loop conformations during the first 2.88 ns of the TMD trajectory remain close to those of the inactive conformation; while these loop conformations converge to those of the active conformation during the entire last half of the TMD trajectory (cluster IV).

The TMD trajectory A2I was segmented into four distinct clusters: I (0-6.58 ns), II (6.60-8.12 ns), III (8.14-12.50 ns), and IV (12.52-20.00 ns). Conformations of the β 3- α 3, β 4- α 4, and β 5- α 5 loops remain close to that of the active form during first 6-7 ns of the A2I TMD trajectory.

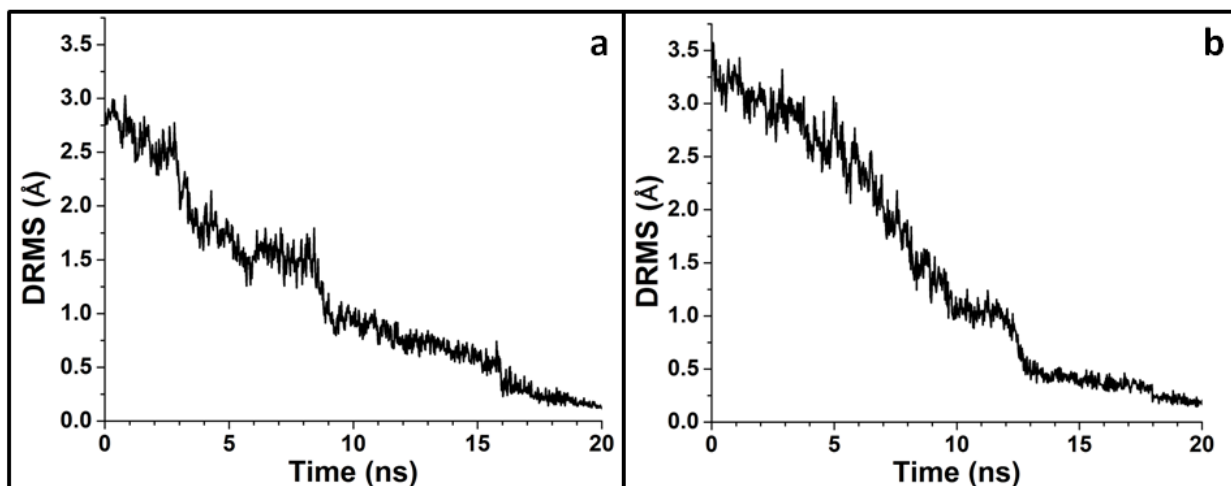


Figure 2.10. **DRMS during TMD simulations.** (a) DRMS of the all C α atoms in the β 3- α 3, β 4- α 4, and β 5- α 5 loops with respect to the active conformation of RR468 in the I2A TMD trajectory. The partitioning into four clusters is I (0-2.88 ns), II (2.90-4.02 ns), III (4.04-9.22 ns), and IV (9.24-20.00 ns). (b) DRMS of the all C α atoms in the β 3- α 3, β 4- α 4, and β 5- α 5 loops with respect to the inactive conformation of RR468 in the A2I TMD trajectory. The partitioning into four clusters is I (0-6.58 ns), II (6.60-8.12 ns), III (8.14-12.50 ns), and IV (12.52-20.00 ns).

2.4.8. PSN analysis with the I2A and A2I TMD trajectories. As noted above, PSN analysis on a given segment of trajectory provides a pairwise interaction map among all residue pairs. Residue pairs having significant changes in the frequency of interaction at a given interaction strength in two consecutive clusters are likely to point to the interactions formed or broken during this stage of the transition. Changes in interaction patterns from one to the next cluster during the inactive to active or reverse direction, of the conformational change are summarized in the Table A2 and A3, Appendix A, respectively. Some notable changes in the pair wise interactions along with the hydrogen bond analysis among the TMD clusters are given below.

For the inactive to active (I2A) transition, interactions between the $\beta 5$ - $\alpha 5$ loop and the other two mobile loops become weak during the initial stage of the I2A TMD simulation. This is evident from the breaking of interactions of R104 and K105 with $\beta 3$ - $\alpha 3$ and $\beta 4$ - $\alpha 4$ residues M55 and T83, respectively (Figure 2.11). Among other significant changes involving three proximal mobile loop residues there is the loss of interaction between V58 and F62, signifying weakening of the N-terminus of the $\alpha 3$ helix. Movement of the N-terminus of the $\alpha 4$ helix is implicated in the loss of interaction between T83 and K91. The N-terminus of the $\alpha 4$ helix is not stable during any simulation. A stable salt bridge between D9 and K105 is a feature in the ACT simulation. This salt bridge forms in the early stage of the I2A simulation. Changes in the interaction pattern involving proximal mobile loop residues are summarized in Table A2, Appendix A.

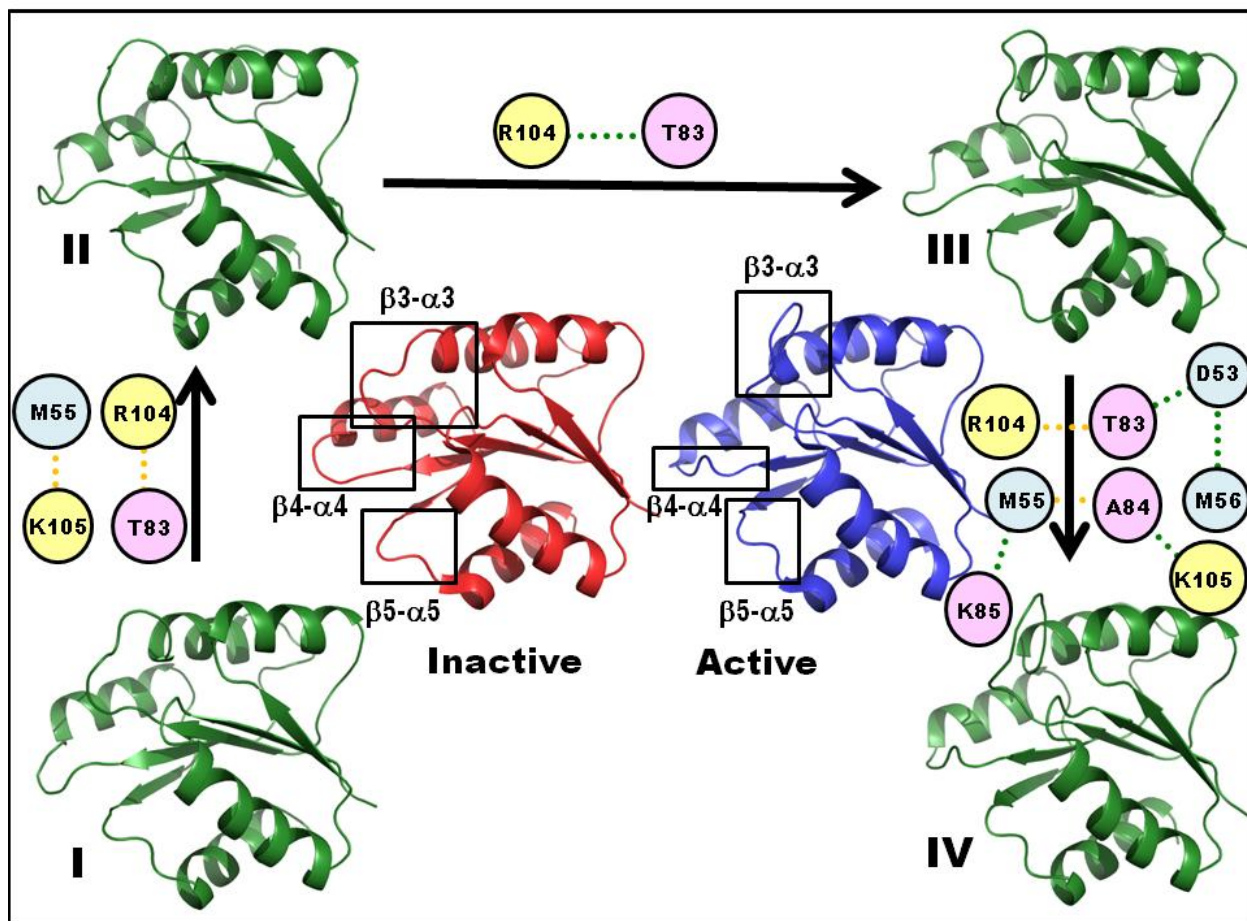


Figure 2.11. **Change in pairwise interaction during RR468 activation.** A representative structure from each cluster during the I2A TMD trajectory is shown in green cartoon. Inter and intra-loop pair wise interactions among $\beta 3-\alpha 3$, $\beta 4-\alpha 4$, and $\beta 5-\alpha 5$ loops that are broken and formed during each transition are shown with orange and green broken lines respectively. Conformations of the three proximal mobile loops are shown for the inactive (red) and active (blue) forms of RR468.

The interaction between T83 and R104 forms and breaks in two intermediate segments of the I2A simulation (Figure 2.11). The side-chain of M56 resides in the void space of the phosphoryl group and becomes close to the D10 residue. D10 is coordinated with the Mg^{2+} ion in the phosphorylated complex. D60 and P57 form a contact with N34 prior to side-chain rearrangement of D60. The T83 sidechain is tethered in an outward orientation and interacts with V102 and R104 (Figure 2.11). This interaction finally breaks during the last stage of the TMD simulation and the T83 side chain flips towards D53, forming a hydrogen bond with the D53 side-chain. Hydrophobic collapse was observed in the side-chains of residues in the proximal end that involves L14, V18, L22, L39, M59, and F107 during the last phase of the conformational transition. The frequency of interaction involving these residues for the interaction strength cutoff, 2% increases for the TMD cluster IV compared to that in cluster III (Table A2, Appendix A). M56 forms contacts with D53 in the buried conformation, as observed in the active form.

For the active to inactive (A2I) transition, many hydrophobic interactions involving V8, A32, I37, L52, V79, and M56 are lost during the early stage of the TMD trajectory. Interactions of M56 with V8 and D53 are broken and that with M59 is formed as a result of the expulsion of M56 sidechain from the deep hydrophobic pocket consisting of V8, L52 and V64 (Table A3, Appendix A). Towards the end of the A2I TMD simulation, the hydrophobic core of the protein is regenerated through expulsion of solvent molecules. Formation of hydrophobic contacts involving A32, I37, I54, V58, F62 in the later stages of this simulation leaves the distortion among two end states to a minimum, except for the three crucial loop regions.

2.4.9. Hydrogen bond analysis with the I2A and A2I TMD trajectories. The definition of interaction in the PSN analysis is based on total number of atom pairs in contact between two residues and thus can be ambiguous about the exact nature of the interaction. Therefore, as

complementary information to PSN, the hydrogen bonding pattern was also constructed that compares each of the consecutive TMD clusters. The hydrogen bond analysis performed on each of these TMD clusters was carried out as discussed in the Methods Section, 2.3.10. The criterion for formation and dissociation of different hydrogen bonds along two consecutive segments along TMD trajectories are as follows. If a hydrogen bond is present significantly (occurrence > 40.0%) in one segment is found to be rare (occurrence < 15.0%) in the next segment, it was considered as a critical hydrogen bond that has to dissociate for the formation of the second segment from the first. Similarly, a hydrogen bond that is rare (<15.0%) in one segment but significant (>40.0%) in the next segment, was considered a critical hydrogen bond that has to form for the formation of the second segment from the first. A summary of all hydrogen bonds that are broken or formed during conformational transition in I2A or A2I trajectory is given in Tables A4 and A5, Appendix A, respectively. Some notable changes in the hydrogen bonding interaction are as follows.

In the inactive to active transition direction, at the beginning of the transition the T83 side chain is stabilized in an outward conformation due to its hydrogen bonding interaction with M103. This hydrogen bond is observed in the IAC simulation as well. This interaction breaks towards the end of the I2A simulation, and T83 flips inward to form a hydrogen bond with D53, as in the active conformation. Formation and dissociation of several hydrogen bonds among the R104 sidechain and different residues of the β 4- α 4 loop is a notable feature of the I2A simulation. Loss of backbone interaction between D60 and V64 at the early stage of the simulation also points towards the weakening of the N-terminus of the α 3 helix.

In the active to inactive transition direction, breaking of the N-terminal of the α 3 backbone interaction was also observed. A backbone hydrogen bond between G61 and L65 breaks during

the early stage of this transition path. R104 also plays a crucial role in this process as it forms a hydrogen bond with T83 in the intermediate stages, but this hydrogen bond is absent in either end state conformations.

2.5. Discussion

The crystal structure of the inactive and active conformations differ most in the three proximal loops, $\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$ and $\beta 5$ - $\alpha 5$, (Figure 2.1 b). The overall backbone RMSD for the IAC, P~RR468 and ACT trajectories show similar deviations from their respective starting structures, while the respective residue RMSFs show that the three proximal loops closely surrounding the site of phosphorylation exhibit greater fluctuations compared to the rest of the protein (Figure 2.2 b). Backbone atoms in the inactive trajectory fluctuates more, especially around the $\beta 3$ - $\alpha 3$ loop, whereas the active conformation, with or without the phosphoryl group and Mg^{2+} , is relatively stable during the 100 ns simulations. There are four distal loops on the other side of the protein away from the site of phosphorylation: loops $\alpha 1$ - $\beta 2$ (residues 26-27), $\alpha 2$ - $\beta 3$ (44-48), $\alpha 3$ - $\beta 4$ (72-78), and $\alpha 4$ - $\beta 5$ (98-100) (Figure 2.1 a). All these distal loops are very stable in all the trajectories as compared to the proximal loops (Figure 2.2 b). The greater overall fluctuations in the IAC trajectory relative to the other trajectories may indicate a larger conformational space for the inactive conformation. The active conformation without the phosphoryl group on D53 and Mg^{2+} is also trapped in a similar conformation as in the P~RR468 trajectory. Charge-charge repulsions among a set of ionized residues spanning E88, E89, D90, and E91 at the N-terminal of the $\alpha 4$ helix are a somewhat unstable region in the protein (Figure 2.2 b). This region is involved in dimerization of the receiver domain in another response regulator protein and hence may have other physiological significance (Gao and Stock, 2010).

There are a greater number of intra and inter loop interactions of the proximal loops $\beta 3$ - $\alpha 3$, $\beta 4$ - $\alpha 4$, and $\beta 5$ - $\alpha 5$ in the active conformation compared to the inactive conformation crystal structure (Figure 2.4 a, b) and they are more stable during the respective simulation trajectories (Figure 2.4 c, d). Interactions that are present in the crystal structures of both forms are stable in

the dynamic scenario as well. The greater number of contacts of M55, M56, M59, and F107 with other loops and the core of the protein in the ACT simulation (Table A6, Appendix A) indicate more interactions of the proximal loops in the active conformation. A strong hydrogen bond between the conserved T83 and phosphoryl group on D53 (or the D53 sidechain) brings $\beta 3$ and $\beta 4$ closer, which facilitates more hydrophobic contacts between the $\beta 3$ - $\alpha 3$ and $\beta 4$ - $\alpha 4$ loops. The hydrogen bond between T83 and D53 was not observed at all during the IAC simulation.

In the presence of the phosphoryl group, the sidechain of K105 is positioned deep within a pocket making contacts with D9 and the phosphoryl group on D53. The adjacent D10 residue interacts with the Mg^{2+} ion; thus, its conformation is directly affected by phosphorylation or dephosphorylation of the D53 residue. In the absence of the phosphoryl group, K105 can still maintain the salt bridge with D9. A salt bridge between K105 and D9 is one reason for trapping the active conformation in a similar conformation in the absence of the phosphoryl group. In the crystal structure of the inactive form, K105 has two alternative conformations. In one conformation, K105 can form a salt bridge with D53, but this salt bridge is not very stable, as the K105 sidechain can sometimes move towards D9, and D53 becomes exposed to the bulk solvent (Figure 2.12 a, b). The dephosphorylation reaction of the phosphorylated aspartate is water mediated, whereby a water molecule is believed to act as the nucleophile to attack the phosphorus atom (Casino et al., 2009). The salt bridge between the highly conserved D9 from $\beta 1$ - $\alpha 1$ and K105 from $\beta 5$ - $\alpha 5$ (Figure 2.12 b) and several stable hydrophobic contacts involving proximal loop residues and the core of the protein limits the exposure of the site of phosphorylation to the bulk solvent, making the phosphoryl group on D53 very stable (half-life, $t_{1/2}$ is ~ 19 hr. at $4^\circ C$ in P~RR468) (Casino et al., 2009). However P~RR468 is quickly dephosphorylated in the presence of HK, as several residues in the proximal loop regions interact

with the HK domain as shown in the HK853-RR468 complex crystal structure. The K105 position is also inferred to be crucial for complex formation with HK853 (Casino et al., 2009). It seems that the higher mobility of the proximal loops may actually prevent the process of complex formation.

The active conformation of the $\beta 3$ - $\alpha 3$ loop is stabilized by hydrophobic interactions of the M56 side-chain with residues V8, L52 and V64 in a deep hydrophobic pocket (Table A6, Appendix A). M56 is completely solvent exposed during the inactive simulation. Flexibility within the $\beta 3$ - $\alpha 3$ loop is completely arrested when M56 is docked into the deep hydrophobic pocket in the active conformation. This finding is consistent with the fact that complex formation with HK853 is affected when M56 is mutated, as the $\beta 3$ - $\alpha 3$ loop conformation in the inactive form of RR468 is not compatible to interact with HK853, as in the HK853-RR468 crystal structure (3DGE) (Casino et al., 2009). However, the side-chain rearrangements in M56 do not have a significant barrier (~ 0.6 kcal/mol for exposed to buried). The change in the M56-V8 C α distance during the inactive to active transition is from 13-15 Å in the inactive conformation to ~ 9 Å in the active conformation (Figure 2.7). This side-chain rearrangement follows several intermediate states in either direction (Figure 2.7 a, and 2.8 a). For the initial conformational transition from 13-15 to 10.2 Å the side-chain is just sampling solvated states and should have a flat PMF in this range. Therefore, we focused the PMF simulation for side-chain orientation from an exposed (10.2 Å) to buried (9.4 Å) distance that corresponds to the dramatic change in the conformation.

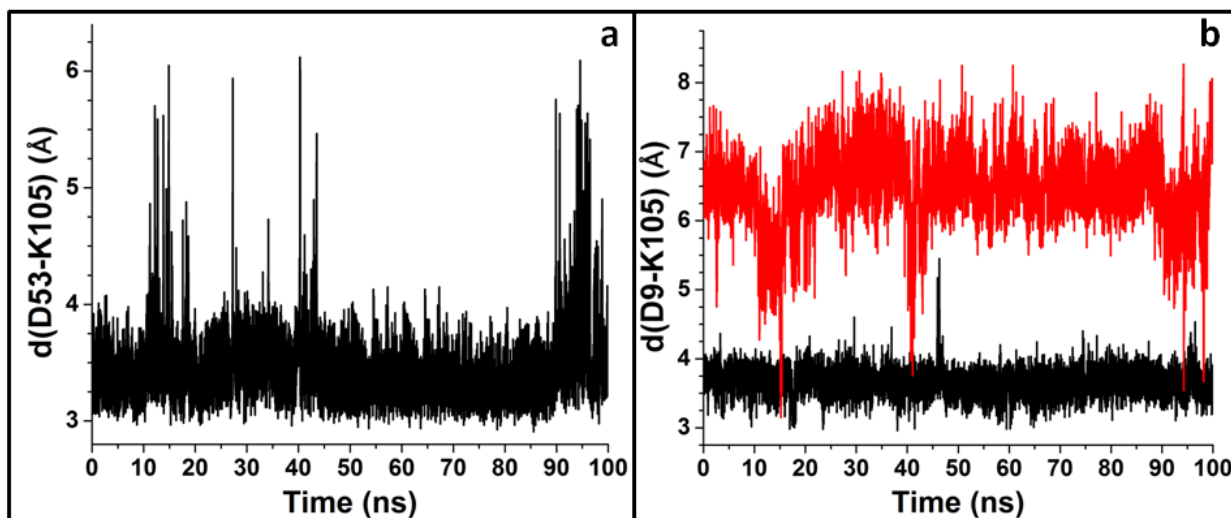


Figure 2.12. **Critical salt bridge interactions during inactive and active form simulation of RR468.** (a) Distance between K105 N ξ and D53 C γ in the IAC simulation. (b) The salt bridge between D9 and K105 in the ACT (black) is very stable whereas this interaction is not present except transiently during the IAC simulation (red).

A specific conformation of the $\beta 3$ - $\alpha 3$ and $\beta 4$ - $\alpha 4$ loops is necessary for the HK853-RR468 interaction (Casino et al., 2009). This conformation in $\beta 3$ - $\alpha 3$ is accompanied by D60 interacting with the N-terminus of the $\alpha 3$ helix in the active conformation while, in the inactive conformation, D60 interacts weakly with the N-terminus of the $\alpha 2$ helix. In the I2A TMD trajectory, the D60 sidechain interaction with G61 (Figure 2.9 a) coincides with formation of cluster IV in the I2A TMD trajectory (Figure 2.10). The energetic barrier of the side-chain rearrangement in D60 is relatively high (~6 kcal/mol). The interaction between RR468 and HK853 may help lower this barrier. Also some transient hydrogen bonds of the D60 side-chain with the backbone N of G35, G61 and backbone N of itself, at the same time brings the N-terminus of the $\alpha 2$ and $\alpha 3$ helices closer in the I2A TMD trajectory, to facilitate the transition of the D60 sidechain (Figure 2.9).

Our PCA based analysis techniques, DEV and EXTEND support the idea that in spite of the lack of transition from inactive to active state and vice versa in the equilibrium trajectories, the primary sequence and the conformation of one state do have information for a transition direction toward the other state. Quasi Harmonic Analysis on the receiver domain of nitrogen regulatory protein C done by Lei et al. is similar to the DEV method in our analysis. They found that e.g. the inactive form modestly points toward the active conformation. Backbone conformations constructed from our PCA based methods point in the right direction of the conformational change, especially for the $\beta 3$ - $\alpha 3$ loop (Figures 2.5 and 2.6). However, the precise conformation that enables side-chain rearrangement and complete conformational transition was not achieved in the equilibrium simulation, suggesting that the state of phosphorylation is also important for the conversion. . The direction of conformational change in the $\beta 4$ - $\alpha 4$ and $\beta 5$ - $\alpha 5$ loops do not clearly approach the other form due to two reasons. First, the difference between the

end states is quite small for these loop regions. Second, PCA modes may be affected by the anomalous fluctuation in the $\alpha 4$, $\beta 5$, $\alpha 5$ regions in the C-terminal region of the protein (Figure 2.2 b).

The TMD simulations were performed sufficiently slowly such that the restraint energy never becomes too large, to minimize deviations from the equilibrium-like scenario. Of course, in finite length trajectories various paths will still be sampled. The relative orientations of the three mobile proximal loops become similar to those of the active conformation somewhere between 10-15 ns in the I2A TMD simulation, and they do not change much thereafter (Figure 2.3 d, e, f). From the DRMS plot it also is evident that the backbone conformations in these loops remain the same during the last half of the I2A trajectory (Figure 2.10). The entire last half of the I2A TMD trajectory fits into the 4th TMD cluster. In all the I2A trajectories, side chain rearrangements in M56, D60 and T83 take place after all three loops adopt a specific active-like conformation i.e. around 15 ns (Figure 2.7 a, b, c). In the A2I trajectories, side chain rearrangements in those three residues take place at an early stage (before 7 ns) of the simulation (Figure 2.7 d, e, f) where the conformation of three proximal loops are still close to that in the active form (Figure 2.3 d, e, f). Side-chain rearrangements are very local conformational changes within the protein that depend on backbone conformations in the system. Apparently, a specific active-like conformation of the three proximal loops is a prerequisite for the sidechain rearrangement in M56, D60, and T83.

Combining hydrogen bond and PSN analyses along the TMD trajectory provides complementary information about the time evolution of pairwise contacts between residues and the physical nature of their interactions. Formation of a contact between R104 and T83 at the third segment of the I2A TMD trajectory (Figure 2.11) is due to hydrogen bond formation between the R104 side-chain and the A84 backbone (Table A4 Appendix A). Breaking of the T83-M103 hydrogen bond

towards the end of I2A trajectory (Table A4, Appendix A) is replaced by one between T83 and D53. Rearrangement of T83 from an outward, relative to the site of phosphorylation, inactive conformation to an inward, active conformation orientation (Figure 2.1) is energetically expensive i.e. free energy barrier ~ 7 kcal/mol (Figure 2.8). This is partly because the T83 side-chain can be stabilized in the outward conformation due to an interaction with the M103 backbone.

Interactions between N34 and V58, D53 and M56, M59 and T63 form towards the end of the I2A conformational transition. These interactions have to break at the early stage of the A2I simulation. The formation of contacts between Q36 and V58, L39 and M59 at the end of I2A and breaking of contacts between Q36 and M59 at the early stage of the A2I simulation also involve the same region in the protein (Tables A2 and A3 in Appendix A). Changes in conformation at the early stage of the A2I trajectory proceed in an approximately reversible manner as compared to the I2A trajectory.

Our hydrogen bond and the Wordom PSN analyses rule out the possibility of any global rearrangement within the protein over the course of the conformational transition. Hydrogen bonds that have more than 50% occurrence in both ACT and IAC trajectories have a very strong presence in the I2A TMD trajectory as well. Thus, the stable regions in the IAC and ACT regions are mostly maintained over the TMD trajectory. The PSN analysis suggests that the conformational rearrangements occur within the proximal loops and their adjacent residues during the transition. Conformational rearrangements in the distal loops and their adjacent residues are minimal during the I2A or A2I TMD trajectories (Table A2 and A3, Appendix A). Also, the distal loops were identified as one of the most stable regions in the protein from the RMSF calculation. The role of R104 seems to be crucial for the conformational transition in the

protein as the interactions of R104 form and break at different stages of the transition (Figure 2.9 b and 2.10). The interaction between the T83 and R104 side-chains might also facilitate this local rearrangement of the T83 residue (Figure 2.11).

2.6. Conclusion

The conformational transition in RR468 is a much slower process (μ s-ms order) than can be captured in 100 ns time scale MD simulations. By complementing the inactive and active form based trajectories with TMD trajectories that span these end states, we have identified structural determinants that may be responsible for the conformational transition. The active conformation without the phosphoryl group is locked close to the conformation of the active form with the phosphoryl group, mainly due to interactions of the M56 and K105 residues. The TMD simulations connecting the active and inactive state conformations of RR468 provide a possible path for this conformational transition and thereby indicate crucial interactions involving residues D60 and R104 that make the conformational transition possible. The relative population of active and inactive states at a given temperature should be affected if these positions are mutated. This hypothesis could be tested using mutation experiments. We demonstrated here how a local event like the phosphorylation or dephosphorylation of a particular amino acid can act as a conformational switch for other local conformational rearrangements leading to a global conformational change in RR468. Biased MD simulations from various starting points along the transition path enabled us to estimate the free energy barrier for several key local conformational rearrangements. From the IAC and ACT trajectories, using PCA-based methods, we found that there is some information embedded in one trajectory about the direction of the β 3- α 3 loop conformational change required to access the other conformation, but other interactions may be required to complete the transition – a situation in between induced fit and conformational selection.

APPENDIX

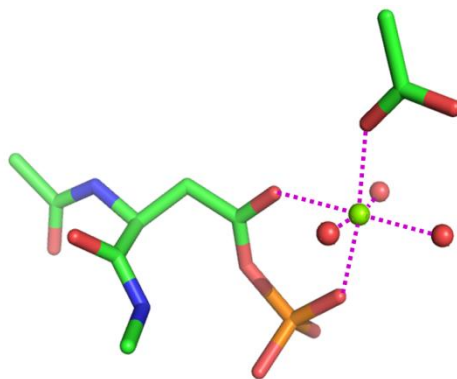


Figure A1. **System considered for the DFT calculation.** The N and C-terminal ends of phosphorylated D53 were capped with Ace and Nme groups, respectively. The acetate ion represents D10 of RR468 that interacts with the Mg^{2+} ion. One water molecule represents the M55 backbone carbonyl O. Two more water molecules were included in the calculation to satisfy the Mg^{2+} coordination sphere.

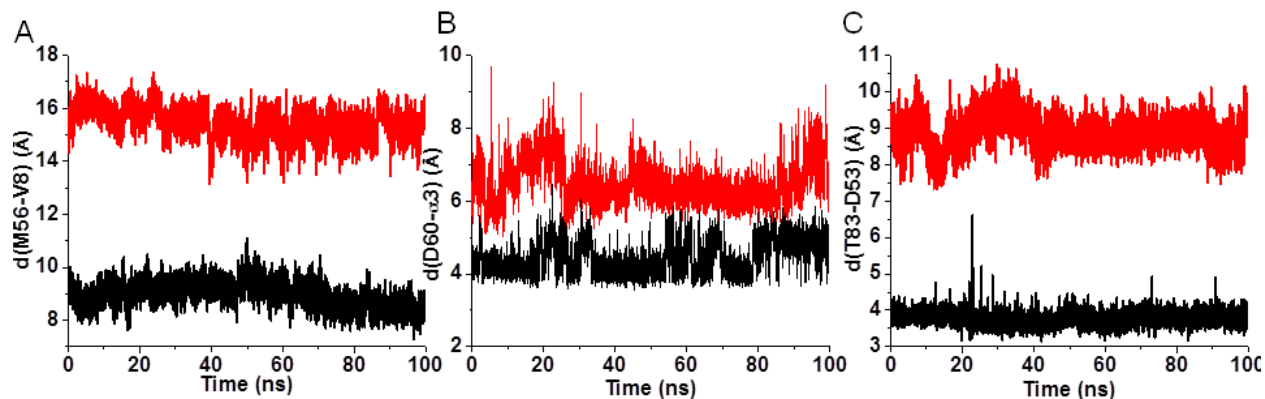


Figure A2. **Time series shows parameters that best describe conformation and position of M56, D60 and Th83.** (A) Distance between C α atoms of M56 and V8 in a hydrophobic core, (B) Average distance between C γ atom of D60 and backbone N atoms of three N-terminal residues of α 3 helix (G61, F62, T63) and (C) Distance between O γ atom of T83 and C γ atom of D53 during IAC (red) and ACT (black).

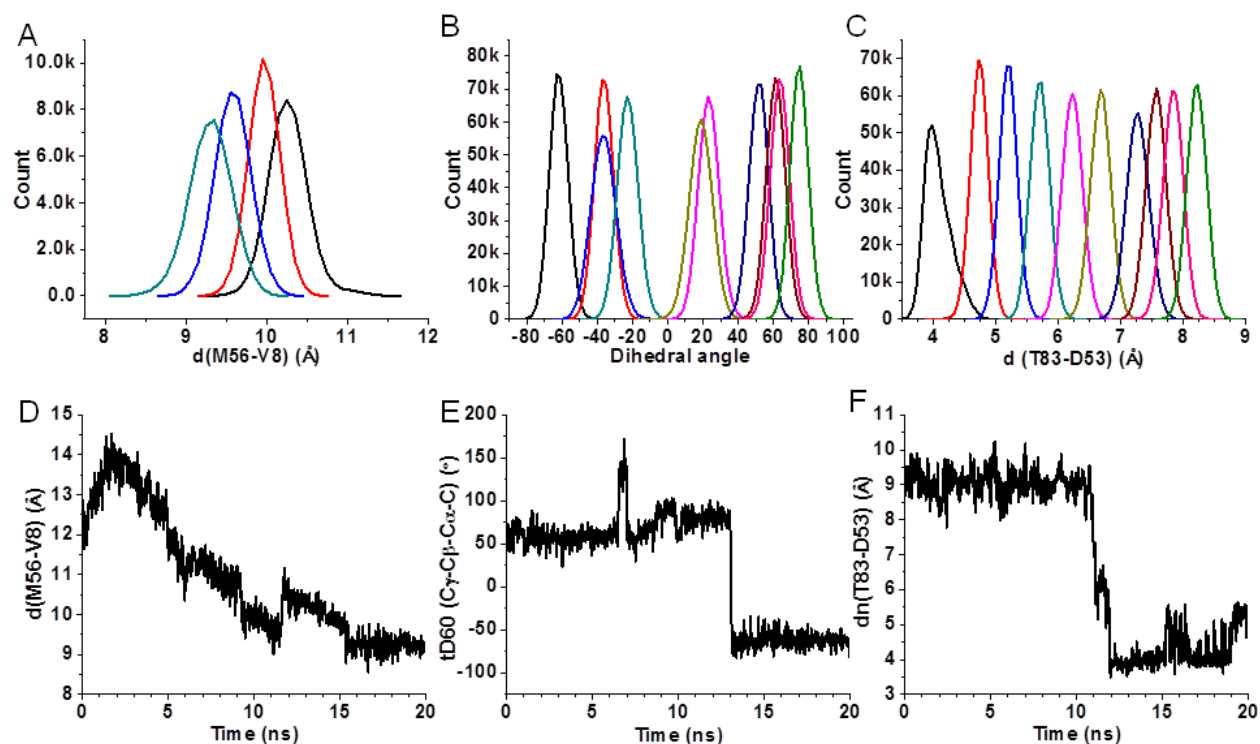


Figure A3. **PMF calculation.** Change in the reaction coordinates along I2A TMD trajectory for (A) M56, (B) D60, and (C) T83 conformational rearrangement. Statistics of the biased simulations that were used for PMF calculation using WHAM for (D) M56, (E) D60, and (F) T83 conformational rearrangement.

Table A1 A. **Hydrogen bonds identified from the crystal structures of the active form of RR468 and their frequency during ACT simulation.**

Acceptor		Donor		Distance		% Occurrence
Residue	Atom	Residue	Atom	Crystal	Simulation	
117	O	121	N	3.030	2.983	10.41
117	O	120	N	3.265	3.076	13.04
116	O	119	N	3.228	3.034	38.31
111	O	115	N	3.164	3.044	16.50
95	O	98	N	3.018	3.076	32.51
94	O	97	N	3.23	3.074	51.60
92	O	96	N	2.829	3.015	32.04
90	OD2	85	NZ	2.808	2.871	3.30
90	OD1	85	NZ	2.829	2.833	6.67
84	O	104	NH2	2.820	3.084	12.98
83	OG1	85	N	3.033	3.112	72.55
79	O	100	N	2.816	2.878	99.42
72	OE2	76	NH1	3.022	2.846	5.30
65	O	69	N	3.185	3.069	21.47
60	OD2	63	N	3.041	2.943	27.61
60	OD2	62	N	2.995	3.062	22.72
59	O	35	N	3.089	-----	-----
57	O	59	N	2.771	2.823	83.64
54	O	61	N	2.882	3.052	15.60
53	OD1	54	N	3.041	2.760	20.86
48	OD2	4	N	2.821	2.915	35.88
42	O	74	NE1	3.080	3.052	15.35
39	O	43	N	3.031	3.018	48.81
13	O	17	N	3.224	3.082	33.65
11	OG	14	N	3.127	3.142	15.29
9	OD1/OD2	105	NZ	2.723	2.760	98.40
2	OG	4	NZ	2.888	3.028	0.53

Table A1 B. **Hydrogen bonds identified from the crystal structures of the inactive form of RR468 and their frequency during IAC simulation.**

Acceptor		Donor		Distance		% Occurrence
Residue	Atom	Residue	Atom	Crystal	Simulation	
117	O	121	ND2	2.998	2.946	5.78
108	OG	111	N	3.181	3.092	21.88
96	O	98	N	3.227	3.106	3.76
93	O	96	N	2.639	3.086	35.99
87	O	91	N	2.841	3.053	36.1
84	O	87	N	3.224	-----	----
82	O	54	N	3.117	2.986	82.98
79	O	101	N	3.055	3.039	67.07
73	OE1	71	NZ	2.669	2.817	4.48
72	OE	76	NH	3.015	2.895	29.62
66	O	70	N	3.238	3.031	6.19
60	OD1	35	N	3.292	2.934	2.67
58	O	61	N	3.049	3.056	16.85
53	OD1	55	N	2.957	3.017	0.86
48	OD1	4	N	2.861	2.908	37.69
41	O	45	N	3.220	2.982	62.42
31	OE2	19	OG	3.150	2.723	47.12
28	OE1	4	NZ	2.713	2.849	5.4
22	O	27	N	3.194	3.033	82.39

Table A2. **Change in interaction from one TMD cluster to the next cluster during inactive to active (I2A) conformational transition.** Interactions that are broken and formed during a transition are given in left and right panels respectively. Percentage interaction cutoff 2% was used in PSN analysis in Wordom.

Interactions broken				Interactions formed			
Residue 1	Residue 2	Frequency (%)		Residue 1	Residue 2	Frequency (%)	
		TMD I	TMD II			TMD I	TMD II
M55	K105	65.97	15.79	D9	K105	12.5	91.23
V58	F62	77.08	17.54	S11	L14	15.28	54.39
T83	E91	52.78	14.04	S19	K23	5.56	100
T83	V102	50.69	15.79	Q36	M59	1.39	63.16
T83	R104	76.39	8.77	K41	E44	14.58	56.14
				K41	F45	16.67	70.18

Residue 1	Residue 2	Frequency (%)		Residue 1	Residue 2	Frequency (%)	
		TMD II	TMD III			TMD II	TMD III
L7	E31	52.63	13.85	D9	L14	15.79	53.85
D10	E33	54.39	5.77	D10	M56	0	52.31
S19	E31	98.25	18.08	D10	P57	1.75	51.15
K41	E44	56.14	14.62	E25	K117	5.26	58.85
				Y27	K117	5.26	51.15
				A32	I37	1.75	54.62
				N34	P57	0	65
				T83	V102	15.79	76.15
				T83	R104	8.77	83.85
				K91	L95	14.04	59.62

Table A2 (cont'd)

Residue 1	Residue 2	Frequency (%)		Residue 1	Residue 2	Frequency (%)	
		TMD III	TMD IV			TMD III	TMD IV
K4	F45	55.77	15.96	L14	V18	18.85	89.05
N34	D60	90.38	11.87	L14	F107	15	50.28
M55	A84	50.38	12.62	V18	L22	16.15	62.71
T83	V102	76.15	10.58	S19	E31	18.08	73.47
T83	R104	83.85	11.13	N34	V58	0	83.67
				Q36	E40	3.46	56.22
				Q36	V58	0	55.66
				L39	M59	2.69	56.59
				D53	M56	0	54.36
				D53	T83	0	72.54
				M55	K85	7.69	96.47
				M59	T63	10.38	72.17
				M59	K67	0	78.11
				T63	K67	0	57.14
				V64	L68	5.38	74.21
				A84	K105	3.08	54.92
				K85	D90	9.23	99.81

Table A3. **Change in pairwise interactions in two consecutive TMD clusters during active to inactive (A2I) conformational transition.** Interactions that are broken and formed during a transition are given in left and right panels respectively. Percentage interaction cutoff 2% was used in PSN analysis in Wordom.

Interactions broken				Interactions formed			
Residue 1	Residue 2	Frequency (%)		Residue 1	Residue 2	Frequency (%)	
		TMD I	TMD II			TMD I	TMD II
V8	M56	65.96	0	I37	E40	12.46	57.14
D9	K105	62.31	0	M56	M59	9.12	92.21
D10	P57	70.82	0				
A32	I37	74.77	3.9				
N34	V58	56.84	0				
Q36	M59	66.26	0				
S43	W74	56.53	19.48				
L52	V79	61.70	18.18				
D53	M56	50.46	18.18				
M59	T63	65.65	0				
E72	K75	62.61	12.99				
E114	H118	58.05	0				

Residue 1	Residue 2	Frequency (%)		Residue 1	Residue 2	Frequency (%)	
		TMD II	TMD III			TMD II	TMD III
N21	I113	55.84	0	N21	K24	18.18	50.68
I37	E40	57.14	15.98	A32	I37	3.90	54.34
E88	E91	51.95	11.42	E72	K75	12.99	51.60
				K101	H118	5.19	59.82
				I113	V116	11.69	65.30

Residue 1	Residue 2	Frequency (%)		Residue 1	Residue 2	Frequency (%)	
		TMD III	TMD IV			TMD III	TMD IV
L14	K105	70.32	0.27	S11	L14	10.50	55.47
M56	M59	80.37	0	V18	L22	13.24	54.40
M56	F62	79.00	0	Q36	E40	15.07	67.47
K67	E70	62.56	12	I54	V58	0	50.13
Q69	K75	58.90	16.53	V58	F62	14.16	91.47
				V58	D90	0.91	73.60

Table A4. **Change in hydrogen bond interactions in two consecutive TMD clusters during inactive to active (I2A) conformational transition.** Cutoffs in percent occurrence of 20 and 50% were used to determine the absence and presence of a particular hydrogen bond.

Acceptor		Donor		% Occurrence	
Residue	Atom	Residue	Atom	First cluster	Second cluster
Broken during TMD I to II					
87	O	91	N	84.72	0
53	OD1	55	N	62.5	12.28
84	O	87	N	57.64	0
116	O	119	N	56.94	15.79
8	O	53	N	56.94	1.75
87	O	90	N	56.25	5.26
86	O	88	N	56.25	0
Formed during TMD I to II					
103	O	83	OG1	18.06	84.21
72	OE2	76	NH2	18.75	57.89
Broken during TMD II to III					
60	O	64	N	75.44	0
53	OD2	55	N	75.44	6.92
9	OD2	11	N	56.14	0
Formed during TMD II to III					
53	OD1	55	N	12.28	72.69
87	O	104	NH2	0	61.15
61	O	64	N	15.79	60.77
114	O	118	N	3.51	57.31
117	O	121	N	7.02	56.15
116	O	119	N	15.79	53.08
Broken during TMD III to IV					
103	O	83	OG1	75.38	7.98
87	O	104	NH2	61.15	0.19
66	O	69	N	53.85	17.25
Formed during TMD III to IV					
15	O	19	OG	9.62	69.76
65	O	69	N	19.23	64.38
53	OD1	54	N	0	64.19
84	O	104	NH2	0.77	59.55
40	O	43	OG	7.69	58.07
53	OD1	83	OG1	0	56.4
11	OG	14	N	2.31	54.36

Table A5. **Change in hydrogen bond interaction from one TMD cluster to the next cluster during active to inactive (A2I) conformational transition.** Cutoffs in percent occurrence of 20 and 50% were used to determine absence and presence of a particular hydrogen bond.

Acceptor		Donor		% Occurrence	
Residue	Atom	Residue	Atom	First cluster	Next cluster
Broken during TMD I to II					
9	OD2	11	N	91.19	9.09
31	OE2	15	NE	74.47	0
Formed during TMD I to II					
31	OE2	19	OG	1.52	88.31
31	OE1	15	NE	13.07	84.42
9	OD1	11	N	6.08	76.62
83	OG1	104	NH1	6.08	74.03
14	O	17	N	17.02	71.43
72	OE1	76	NH2	16.72	66.23
56	O	58	N	0	59.74
72	OE2	76	NE	15.81	55.84
21	O	25	N	14.29	51.95
Broken during TMD II to III					
48	OD2	4	N	77.92	0
83	OG1	104	NH1	74.03	14.61
60	OD1	63	OG1	51.95	7.76
Formed during TMD II to III					
48	OD1	4	N	9.09	95.43
87	O	91	N	2.6	85.39
87	O	90	N	9.09	50.68
Broken during TMD III to IV					
91	OE1	104	NH1	95.43	18.67
91	OE2	104	NH2	91.32	19.47
56	O	58	N	57.99	0
72	OE2	76	NE	56.62	0
26	O	2	OG	54.79	19.47
Formed during TMD III to IV					
117	O	121	ND2	10.5	95.73
55	O	58	N	0	89.6
60	OD1	35	N	0	58.67
72	OE2	76	NH1	10.96	56
53	OD1	55	N	11.42	53.87
58	O	61	N	0	52

Table A6. **Pair wise interactions involving residue number 53-60 ($\beta 3-\alpha 3$), 83-87 ($\beta 4-\alpha 4$), and 104-108 ($\beta 5-\alpha 5$).** Intra and inter loop contacts are highlighted in gray.

Residue 1	Residue 2	Interaction strength @crystal structures		Frequency @ trajectory with $I_{\min} > 5.0\%$	
		Active	Inactive	ACT	IAC
8	53	1.46	0	0	0.12
8	56	25.87	0	91.78	0
9	53	33.27	29.27	96.31	100
9	105	24.89	0	99.99	2.23
10	53	3.99	0	0.02	0.29
10	56	29.11	0	63.42	0
10	57	38.64	0	99.5	0
10	59	0	5.54	0	0
11	53	0	2.94	0	0
11	105	4.59	0	6.99	0.02
14	53	0	17.64	0.02	1.24
14	105	69.09	1.41	99.91	31.99
14	107	6.09	26.79	96.47	90.67
17	107	16.33	10.05	98.78	0.67
18	107	20.97	13.11	99.99	36.46
21	107	8.46	0	83.18	0.64
34	56	8.41	0	22.25	0
34	57	11.4	0	35.88	0
34	58	45.81	0	72.66	0
34	59	2.8	0	40.97	0
34	60	0	5.39	0	8.3
35	56	10.48	0	81.08	0
35	59	15.72	0	67.66	0
35	60	0	15.09	0	4.65
36	58	61.6	0	37.9	0
36	59	14.95	0	95.75	0
36	60	0	2.61	0	3.82
39	59	22.62	0	67.68	0
52	56	18.38	0	78.38	0
53	56	29.11	0	73.61	0
53	82	0	32.57	0.12	99.31
53	83	2.89	0	99.96	0
53	105	23.5	0	100	99.9
54	58	0	21.51	0	95.69
54	60	4.2	0	6.96	0.98
54	61	14.11	10.58	76.54	95.34
54	62	38.93	5.02	87.9	41.45
54	65	4.28	2.85	89.94	63.69
54	81	52.23	52.23	99.97	99.94

Table A6 (cont'd)

Residue 1	Residue 2	Interaction strength @crystal structures		Frequency @ trajectory with $I_{\min} > 5.0\%$	
		Active	Inactive	Active	Inactive
54	83	44.08	33.44	99.8	95.91
54	85	15.99	0	99.51	0
54	90	2.8	19.59	3.35	0.22
54	91	0	1.37	0.04	0.08
54	94	8.12	22.75	96.69	63.42
54	102	0	1.54	0.09	0.24
55	83	4.52	0	90.58	0
55	84	14.48	54.71	31.65	55.36
55	85	54.73	0	99.83	0.9
56	59	1.44	0	24.18	0.12
56	60	4.16	0	1.38	0
56	61	13.98	0	94.26	0.02
56	64	13.69	0	71.11	0
56	84	0	1.61	0	83.00
56	86	0	8.73	0	29.59
56	87	0	13.98	0	6.65
56	90	0	6.93	0	28.24
58	61	0	9.2	0	13.15
58	62	0	30.15	0	58.99
58	90	0	23.37	0	32.39
59	63	51.19	0	81.67	31.16
59	64	22.82	0	99.29	0
59	67	41.77	0	68.86	0
60	63	11.56	14.45	98.39	13.08
60	64	0	5.84	0	2.73
82	105	57.81	0	99.98	99.94
82	107	35.32	6.09	97.57	98.91
83	90	1.45	0	0	4.87
83	91	0	40.92	1.68	68.59
83	102	20.62	41.25	96.02	97.71
83	104	6.47	9.06	97.29	97.81
83	105	1.5	0	14.54	0
84	87	0	1.95	0	0
84	104	2.77	0	12.36	0.04
84	105	27.29	4.82	99.98	82.25
85	88	0	25.65	0	2.26
85	90	55.3	0	13.5	0
86	90	1.68	0	20.16	52.4
86	104	30.02	0	4.67	10.34
87	90	1.68	11.74	46.24	43.46

Table A6 (cont'd)

Residue 1	Residue 2	Interaction strength @crystal structures		Frequency @ trajectory with $I_{\min} > 5.0\%$	
		Active	Inactive	Active	Inactive
103	107	0	2.49	0.34	88.45
104	111	0	3.5	10.6	0.05
107	112	16.08	16.08	98.17	80.07
108	111	43.32	23.1	89.85	95.86

Table A7. **Hydrogen bonds identified from the I2A TMD trajectory that has rare occurrence in the equilibrium trajectories.** All these hydrogen bonds at least 1 side chain atom and has more than 15 consecutive occurrences (300 ps) in the TMD trajectory.

Acceptor		Donor		% Occurrence		
Residue	Atom	Residue	Atom	I2A	ACT	IAC
60	OD2	35	N	54.00	0.00	0.00
72	OE2	76	NH1	43.00	0.00	5.33
72	OE1	76	NH1	40.10	0.00	4.66
72	OE1	76	NH2	38.40	1.52	4.63
72	OE2	76	NH2	20.60	2.96	3.20
40	O	43	OG	32.70	4.28	5.45
53	OD1	54	N	30.20	20.54	0.00
84	O	104	NH2	27.10	6.94	0.00
84	O	104	NH1	10.70	0.00	0.00
9	OD2	105	NZ	78.40	41.7	0.00
103	O	83	OG1	25.50	0.00	27.74
31	O	41	NZ	38.80	16.8	15.66
60	OD2	63	OG1	16.20	28.85	3.08
60	OD2	63	N	14.80	18.65	0.00
87	O	104	NH2	13.20	0.00	23.64
25	OE1	117	NZ	13.20	8.54	2.86
90	OD1	85	NZ	12.30	2.09	0.00
31	OE1	19	OG	10.50	13.89	17.16
60	OD1	61	N	10.10	1.24	0.00

Table A8. **Hydrogen bonds identified from the A2I TMD trajectory that has rare occurrence in the equilibrium trajectories.** All these hydrogen bonds at least 1 sidechain atom and has more than 15 consecutive occurrences (300 ps) in the TMD trajectory.

Acceptor		Donor		% Occurrence		
Residue	Atom	Residue	Atom	I2A	ACT	IAC
91	OE1	104	NH1	48.60	6.33	7.63
88	OE2	86	N	36.90	0.00	0.00
117	O	121	ND2	34.70	2.68	3.91
103	O	83	OG1	32.70	0.00	27.74
60	OD1	63	OG1	23.10	27.84	0.00
60	OD2	63	OG1	22.50	28.85	3.08
72	OE2	76	NH2	18.50	2.96	3.20
33	OE2	15	NH1	18.00	22.87	8.17
25	OE1	117	NZ	16.80	8.54	2.86
72	OE1	76	NH1	14.00	0.00	4.66
33	OE1	15	NH1	10.30	10.22	16.59

Table A9. Details of Steered MD simulation to sample transitional phase for D60 rearrangement and T83.

Conformational rearrangement	D60	T83
Reaction coordinate	Torsion angle $C\gamma-C\beta-C\alpha-C$ of D60	Distance between T83 $O\gamma$ and D53 $C\gamma$
Initial value	77.4	8.15
Initial snapshot	12.92 ns of I2A	15.0 ns of I2A
Final value	-63.8	3.79
Final snapshot	13.12 ns of I2A	15.18 ns of I2A
Force constant	20 kcal/mole	20 kcal/mole
Length	5 ns	3 ns

REFERENCES

REFERENCES

- Amadei, A., Linssen, A.B.M., and Berendsen, H.J.C. (1993). Essential Dynamics of Proteins. *Proteins: Struct, Func, Gen* 17, 412-425.
- Bourret, R.B. (2010). Receiver domain structure and function in response regulator proteins. *Curr Opin Microbiol* 13, 142-149.
- Bourret, R.B., and Silversmith, R.E. (2010). Two-component signal transduction. *Curr Opin Microbiol* 13, 113–115.
- Case, D.A., Darden, T.A., III, T.E.C., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Crowley, M., Walker, R.C., Zhang, W., *et al.* (2008). AMBER 10 (University of California, San Francisco.).
- Casino, P., Rubio, V., and Marina, A. (2009). Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell* 39, 325-336.
- Casino, P., Rubio, V., and Marina, A. (2010). The mechanism of signal transduction by two-component systems. *Curr Opin Struct Biol* 20, 763.
- Cheatham, T.E.I. (2008). AmberTools Users' Manual.
- Cukier, R.I. (2009). Apo Adenylate Kinase Encodes Its Holo Form: A Principal Component and Varimax Analysis. *J Phys Chem B* 113, 1662-1672.
- Formanek, M.S., Ma, L., and Cui, Q. (2006). Reconciling the "old" and "new" views of protein allostery: A molecular simulation study of chemotaxis Y protein (CheY). *Proteins: Struct Funct Bioinform* 63, 846-867.
- Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Rob, M.A., Cheeseman, J.R., Jr., J.A.M., Vreven, T., Kudin, K.N., Burant, J.C., *et al.* (2003). Gaussian 03 (Gaussian, Inc., Wallingford, CT).
- Gao, R., and Stock, A.M. (2009). Biological Insights from Structures of Two-Component Proteins. *Annu Rev Microbiol* 63, 133-154.
- Gao, R., and Stock, A.M. (2010). Molecular strategies for phosphorylation-mediated regulation of response regulator activity. *Curr Opin Microbiol* 13, 160–167.
- García, A.E. (1992). Large-Amplitude Nonlinear Motions in Proteins. *Phys Rev Lett* 68, 2696-2699.
- Grossfield, A. (2003). WHAM: the weighted histogram analysis method, version 1.4.

Huang, H., Ozkirimli, E., and Post, C.B. (2009). Comparison of Three Perturbation Molecular Dynamics Methods for Modeling Conformational Transitions. *J Chem Theory Comput* 5, 1304–1314.

Jolliffe, I.T. (2004). *Principal Component Analysis*, second edn (New York: Springer Science).

Knaggs, M.H., Salsbury, J.F.R., Edgell, M.H., and Fetrow, J.S. (2007). Insights into Correlated Motions and Long-Range Interactions in CheY Derived from Molecular Dynamics Simulations. *Biophys J* 92, 2062–2079.

Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A., and Rosenberg, J.M. (1992). The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J Comput Chem* 13, 1011-1021.

Lei, M., Velos, J., Gardino, A., Kivenson, A., Karplus, M., and Kern, D. (2009). Segmented Transition Pathway of the Signaling Protein Nitrogen Regulatory Protein C. *J Mol Biol* 392, 823–836.

Ma, L., and Cui, Q. (2007). Activation Mechanism of a Signaling Protein at Atomic Resolution from Advanced Computations. *J Am Chem Soc* 129, 10261-10268.

Roche, P., Mouawad, L., Perahia, D., Samama, J.-P., and Kahn, D. (2002). Molecular dynamics of the FixJ receiver domain: movement of the b4–a4 loop correlates with the in and out flip of Phe101. *Protein Sci* 11, 2622–2630.

Seeber, M., Cecchini, M., Rao, F., Settanni, G., and Caflisch, A. (2007). Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* 23, 2625–2627.

Seeber, M., Felling, A., Raimondi, F., Muff, S., Friedman, R., Rao, F., Caflisch, A., and Fanelli, F. (2010). Wordom: A User-Friendly Program for the Analysis of Molecular Structures, Trajectories, and Free Energy Surfaces. *J Comput Chem* 32, 1183-1194.

Souaille, M., and Roux, B. (2001). Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput Phys Commun* 135, 40–57.

Stewart, R.C. (2010). Protein histidine kinases: assembly of active sites and their regulation in signaling pathways. *Curr Opin Microbiol* 13, 133–141.

Tatsis, V.A., Tsoulos, I.G., Krinasa, C.S., Alexopoulos, C., and Stavrakoudis, A. (2009). Insights into the structure of the PmrD protein with molecular dynamics simulations. *Int J Biol Macromol* 44, 393–399.

Wang, J. (2008). *AmberTools Users' Manual*.

CHAPTER 3

Conformational Dynamics in a Two-component System: Molecular Dynamics and Markov State Model Analysis of a Response Regulator Protein

3.1. Abstract

Signal transduction in a Two Component System (TCS) is mediated by phosphorylation of a response regulator protein (RR) by a membrane bound histidine kinase (HK). The RR is responsible for carrying out the output of the signaling pathway, as regulated by the phosphoryl transfer activity of the HK to a conserved aspartate residue in the RR. Phosphorylation triggers a conformational change in RR to elicit the response. In response regulator protein RR468 the residues proximal to the site of phosphorylation undergo large conformational changes in their unphosphorylated and phosphorylated forms, as evident from crystal structures. We characterize the conformational transition associated with the unphosphorylated form of RR468 that bridges the inactive and active forms of RR468 using multiple molecular dynamics simulation trajectories analyzed with a Markov state model and transition path theory. Conformational snapshots from a nonequilibrium simulation trajectory connecting inactive and active states of RR468 were used as starting points for multiple equilibrium simulations to sample intermediate states. Discretization of the resulting conformations was done in two steps: first into a large set of microstates clustered geometrically and then reduced to a small set of macrostates based on their kinetic relationship. This strategy, along with identification of a set of distance coordinates to characterize the various macrostates, enabled us to define the nature of the conformational transition that the protein undergoes before acting as a substrate for the phosphoryl transfer reaction. Dominant pathways are identified that show that the transition proceeds through a few critical metastable intermediate states in RR468, where the catalytic center gradually evolves to facilitate the phosphoryl transfer reaction first, and eventually adopts a conformation that resembles the phosphorylated, product form of the protein.

3.2. Introduction

Two Component System (TCS) signal transduction is the predominant mechanism in bacteria for sensing environmental conditions such as temperature, nutrients, osmopressure etc (Casino et al., 2010; Gao and Stock, 2009). Signal transduction in a TCS involves a membrane bound histidine kinase (HK) and a cognate response regulator (RR) protein. It is mediated by phosphorylation of the RR by the HK, which induces the active, phosphorylated conformation of RR upon the phosphoryl transfer (Bourret, 2010). In RR468 from *Thermotoga maritima*, crystal structure conformations for free and Beryllium trifluoride (BeF_3^-) bound forms highlight the structural difference between inactive and active conformations (Casino et al., 2009). Beryllium trifluoride, covalently attached to an aspartate residue mimics the phosphorylated aspartate (D53) at the site of phosphorylation in phosphorylated RR468 (P~RR468) (Casino et al., 2009).

Protein conformational dynamics play an important role in many biochemical processes and their regulation. While structural methods such as X-ray crystallography and NMR can be used to determine the structures of the stable states of a protein, the atomic details of a conformational transition are difficult to obtain. In this regard, molecular dynamics (MD) simulations are very useful to fill in intermediate states, especially when initial and final structures are available from experiment. A conformational transition is a local unfolding-folding event, which is often obscured in equilibrium simulations due to high energy barriers separating experimentally observed states. In previous work (Banerjee et al., 2014), we have shown that the conformations of RR468 remain close to their respective starting conformations when the MD simulation is started from the inactive or the active form crystal structures. The process of conformational transition in a different response regulator protein was investigated earlier using equilibrium and nonequilibrium simulation techniques (Formanek et al., 2006; Knaggs et al., 2007; Lei et al.,

2009; Roche et al., 2002; Tatsis et al., 2009). Transition path sampling using a combination of MD and Monte Carlo simulations was found to be useful in gaining mechanistic insight into the conformational dynamics at a distant point due to phosphorylation in chemotaxis Y (CheY) signaling protein (Ma and Cui, 2007). A strategy based on conformational clustering followed by establishment of the kinetic relationship among the clusters led to identification of the intermediates conformational states along the Src kinase activation pathway (Yang et al., 2009). To provide conformational sampling spanning experimentally observed states, nonequilibrium simulations followed by multiple short equilibrium MD simulations have proven useful (Ma and Cui, 2007; Yang et al., 2009). In another recent work, Markov state models built from MD simulation data in the absence of any ligand revealed a variety of conformations including ligand-bound configurations that implies a role for conformational selection in ligand binding (Bowman and Geissler, 2012).

The conformational differences between phosphorylated (active) and unphosphorylated (inactive) states of RR468 are mainly concentrated in two loop regions, the $\beta 3\alpha 3$ and $\beta 4\alpha 4$ loops, while the remaining core structure remains quite invariant in these two forms (Banerjee et al., 2014; Casino et al., 2009). The most significant conformational difference is in the $\beta 3\alpha 3$ loop, which is adjacent to the site of phosphorylation, the conserved aspartate residue, D53. The conformational differences between the inactive and active forms at the $\beta 4\alpha 4$ loop region are due to the $\alpha 4$ helix, which is slanted outward by $\sim 10^\circ$ at the N-terminus of the helix in the active conformation with respect to that in the inactive conformation.

While conformational changes in previous (Banerjee et al., 2014) equilibrium simulations started from the inactive and active conformations were unable to provide insight into the process of conformational transition connecting these two states, these simulations provided a great deal of

understanding on the conformational plasticity associated with the active and inactive states of the protein (Banerjee et al., 2014). We also used targeted molecular dynamics (TMD) simulations to investigate the critical interactions associated with the conformational transition of RR468 from its inactive to active form, and identified non native interactions that can play a crucial role for the conformational transition (Banerjee et al., 2014). In the present work, conformational snapshots obtained from a large number of equilibrium simulations, starting from various points along the TMD trajectory, are used to sample intermediate states. However, the complexity of the MD data, due to the stochastic motions of the myriad atoms that constitute the protein, makes the analysis of such data extremely challenging. A Markov state model (MSM) clustering approach is very useful in identifying metastable states and obtaining a description of both the thermodynamics and kinetic pathways of the conformational transitions (Beauchamp et al., 2011; Bowman et al., 2009; Pande et al., 2010).

In this study, the conformational transition of RR468 has been investigated by extensive MD simulations without any bound ligand, followed by MSM clustering analysis to identify metastable states along possible transition paths. Transition paths spanning inactive and active conformations were obtained from transition path theory using the metastable states identified by the MSM clustering, as parameterized by committor or splitting probabilities (Berezhevskii et al., 2009; Metzner et al., 2009). Conformations are found that closely resemble the active and inactive forms of RR468, and metastable intermediates are identified, based on an extensive set of inter-residue distances. The transition path analysis provides orders of events in terms of the local rearrangements of various structural elements, prior to the phosphoryl transfer reaction. It was found that the transition proceeds through a few critical metastable intermediate states, where the ligand binding site gradually becomes isolated from bulk water. Two metastable

intermediates were identified in which the active site is partially open. A completely closed active site conformation was obtained from the ligand free simulation followed by MSM analysis that resembles the phosphorylated form of RR468 (P~RR468), where the active site adopts a closed conformation to protect the high-energy phospho-aspartate residue (Banerjee et al., 2014)

3.3. Methods

3.3.1. Data Accumulation.

A targeted molecular dynamics (TMD) simulation was performed that pulls the protein from an inactive-like to the active-like conformation using a root-mean-square-deviation (RMSD) restraint. The details of the TMD simulation are described elsewhere (Banerjee et al., 2014). A total of 101 snapshots were selected along this TMD trajectory. For each of these snapshots, restraining potentials in the form of positional restraints on all protein heavy atoms were introduced first and then relaxed to zero in a stepwise manner over 2 ns. Then, unbiased simulations were performed for 15 ns, resulting in a total of 1.515 μ s of simulation trajectory. Amber10 (Case et al., 2008) was used for all simulations. In addition to protein, there were 7068 water molecules of solvation and 2 Na^+ ions to neutralize the system (already present from the TMD step) (Banerjee et al., 2014). A time step of 2 fs was used with the SHAKE option on all bonds containing H-atoms. Langevin dynamics was used for temperature control during constant pressure simulation at 300K. A conformational snapshot was saved every 10 ps to accumulate the data for subsequent MSM analysis (151500 snapshots).

3.3.2. Markov State Modeling. Markov state models can be constructed based on kinetic transitions between discrete states, which represent a partitioning of phase space into metastable states (Pande et al., 2010). MSM building techniques include some clustering method in the first place to divide the conformational space into clusters with a very stringent criterion, referred to as microstates. Here, the high degree of structural similarity within microstates ensures relatively easy exchanges. The numbers of transitions among different microstates over an interval of length τ , the lag time, provides a count matrix $\mathbf{C}(\tau)$ that can be related to a transition probability matrix, $\mathbf{T}(\tau)$. The two matrices are related via the diagonal matrix of equilibrium

probabilities, $\mathbf{P}^{\text{eq}}$, as $\mathbf{C}(\tau) = \mathbf{P}^{\text{eq}} \bullet \mathbf{T}(\tau)$. Construction of the transition matrix establishes the kinetic relationship among the microstates. If a suitable range of lag times can be found where the dominant eigenvalues of $\mathbf{T}(\tau)$ are independent of τ , the model should satisfy the Markov property. Relaxation timescales for transitions between different microstates are a function of the eigenvalues calculated after diagonalization of the transition matrix, $\mathbf{T}(\tau)$. The Markov property can be established by plotting the relaxation timescales for the slowest processes, i.e. based on the dominant eigenvalues, against the lag time. Microstates are lumped into macrostates based on kinetics, whereby there are fast exchanges between microstate pairs within a macrostate, and slow exchange between microstates in different macrostates. With the transition matrix and equilibrium probabilities available, transition path theory can be used to investigate the sequence of intermediates that the protein follows in transiting between given end point states (Metzner et al., 2009).

3.3.3. Clustering and Microstates. The MSM analysis was performed using MSMBuilder2.6.0 (Beauchamp et al., 2011; Bowman et al., 2009). The 151500 protein configurations from the 101 unbiased trajectories, as described in the Section 3.3.1, were clustered using a kcenters algorithm as implemented in MSMBuilder2. From the crystal structures of RR468, it is evident that the conformation of active and inactive forms differ most in the residue ranges D53-D60 ($\beta 3\alpha 3$) and T83-E91 ($\beta 4\alpha 4$ and part of $\alpha 4$). From the equilibrium simulations starting from the inactive and active form crystal structures, it was found that the root-mean-square-fluctuation (RMSF) is relatively high in residues in the same ranges (Banerjee et al., 2014). Backbone heavy atoms (N, C α , C, and O) and C β atoms of all the residues except in the range D53-D60 and T83-E91 were used to align all the conformations for clustering. Differences in the active and inactive states are

most conspicuous in the $\beta 3\alpha 3$ loop region. Also, our previous work suggested that the charge-charge repulsions among a set of ionized residues spanning E88, E89, D90, and E91 at the N-terminus of the $\alpha 4$ helix produce a somewhat unstable region in the protein. Backbone heavy atoms, N, C α , C, O and C β atoms of residues only in the range D53-D60 ($\beta 3\alpha 3$) were used for clustering using RMSD as the distance metric with the clustering cutoff 1.0 Å (maximum cophenetic distance).

Lag times 0.5 ns, 1.0 ns, 1.5 ns, 2.0 ns, and 3.0 ns were chosen for MSM model building to estimate the count ($\mathbf{C}(\tau)$) and transition ($\mathbf{T}(\tau)$) matrices, and equilibrium populations ($\mathbf{P}^{eq}$) of the microstates. The relaxation timescale versus lag time, τ , plot became relatively invariant to the choice of lag time in the range between 1 to 2 ns (Figure 3.1 A). For a choice of lag time more than 2 ns, count matrix elements ($\mathbf{C}(\tau)$) approach their limiting behavior and relaxation time estimation becomes unreliable (Banerjee and Cukier, 2014). The model is Markovian for the choice of lag times within the range between 1-2 ns. Reversible models that satisfy detailed balance were calculated from the estimation of the most likely reversible matrices using Maximum Likelihood Estimation (MLE) as implemented in MSMBuilder2. A microstate model with lag time (1 ns) was chosen for subsequent macrostate model generation for three reasons: (1) the microstate model is Markovian at this selection of lag time, i.e. the relaxation timescales become invariant for the choice of lag time at 1 ns, (2) it contains the maximum possible data compared to models built from larger lag time microstate models, (3) each individual simulation (15 ns) is sufficiently longer than the chosen lag time (1 ns).

3.3.4. Lumping of Microstates into Macrostates. The PCCA+ algorithm (Deuffhard and Weber, 2005) as implemented in MSMBuilder2 was used to lump microstates into models with

8, 12, 16, 20, 24, 28, and 32 macrostates. Relaxation timescales for the macrostate model with 24 states follow that of the microstate model (Figure 3.1 B). The first 10 relaxation times as a function of lag time calculated from 24 macrostates also identified roughly same number of processes that are distinct in terms of the timescale. The population probabilities of each macrostate were calculated by adding the equilibrium population probabilities of all the constituent microstates (Figure 3.1 C and D). Randomly chosen sets of 1000 conformations from each macrostate were used for subsequent structural characterization using Ambertools (Wang, 2008) and Wordom (Seeber et al., 2010). For a given set of beginning and end states, transition paths, committers (splitting probabilities), and net fluxes were calculated using transition path theory as implemented in MSMBuilder2 (Berezhevskii et al., 2009; Metzner et al., 2009).

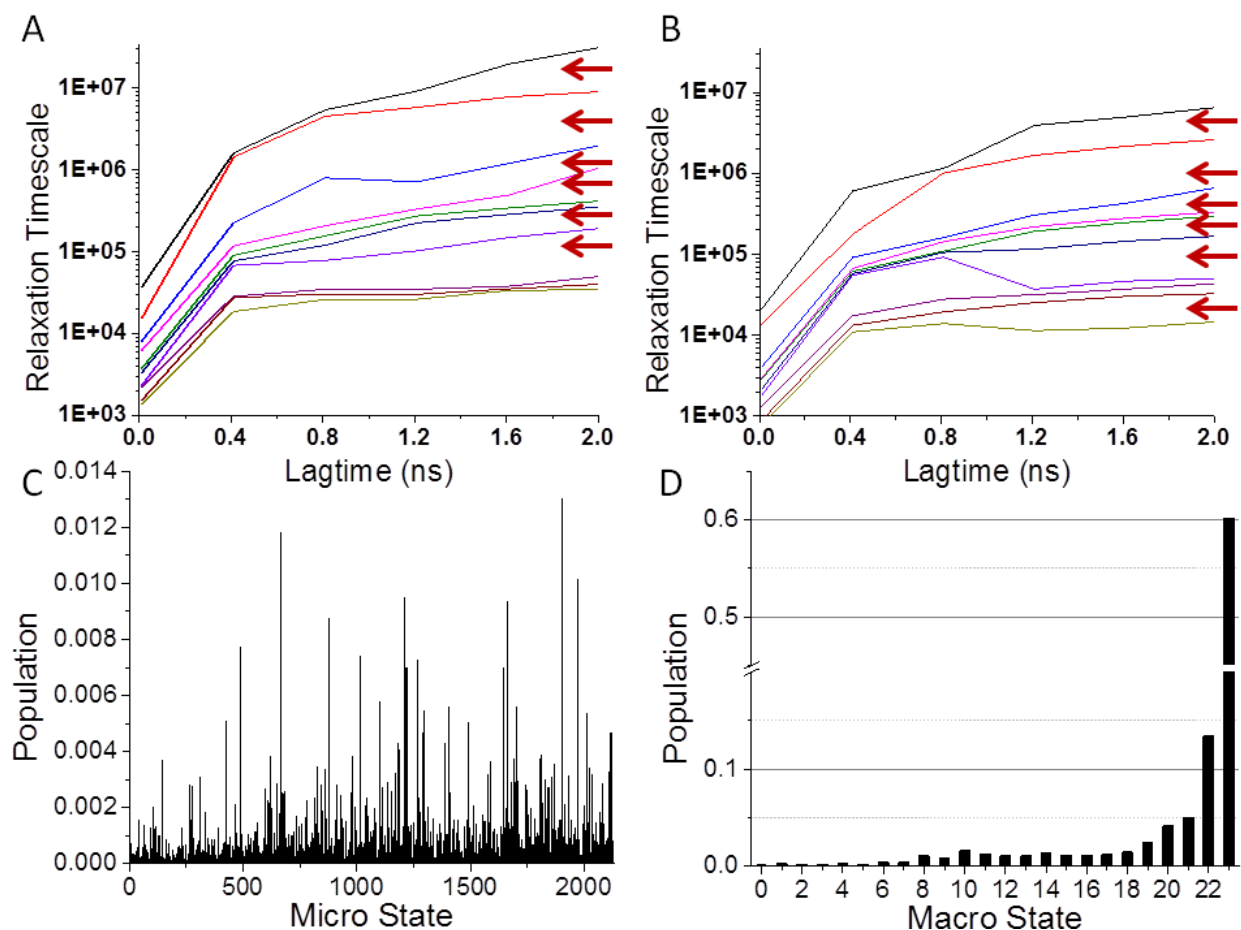


Figure 3.1. **Relaxation timescale and population.** Relaxation timescale as a function of lag time derived from (A) microstate and (B) macrostate transition matrix. (C) Equilibrium population of microstates lumped into (D) 24 macrostates, indexed as 0-23. First 10 relaxation times as a function of lag time shows that there are 6 modes that are characteristically different in terms of the timescale in the range between 1-2 ns, as evident from the number of gaps in (A), shown with arrows. A macrostate model with 24 metastable states appears to identify six slowest states in the same range as in (B).

3.3.5. Free energy landscapes. Projections of the free energy landscape on reaction coordinates that describe the $\beta 3\alpha 3$ loop and $\alpha 4$ helix conformations were done with the entire pool of data used for the MSM analysis (Rhee and Pande, 2003; Zhou et al., 2001). The $C\alpha$ - $C\alpha$ distance between a $\beta 3\alpha 3$ loop residue and a residue with minimal fluctuation in the protein was used to describe the $\beta 3\alpha 3$ loop conformational fluctuations. Similarly, the $C\alpha$ - $C\alpha$ distance between a $\beta 4\alpha 4$ loop or $\alpha 4$ helix residue and a $\beta 3\alpha 3$ residue was used to describe the relative orientation of the $\beta 3\alpha 3$ loop with respect to the $\beta 4\alpha 4$ loop or $\alpha 4$ helix. The 2D frequency count to obtain the free energy landscape was done with bin size 0.5 Å.

3.3.6. Insertion of Mg^{2+} for selected macrostates. A Mg^{2+} ion was docked into the active site of four randomly selected conformations for macrostates 23, 22, 20, and 16 with the help of the crystal structure of RR468 in the presence of Mg^{2+} (PDB id 3GL9). The carboxylate group of D10 points outward due to repulsion with that of D53 in the absence of the divalent cation in our simulations, resulting in a $\sim 180^\circ$ rotation of the χ_1 angle of D10. The D10 side-chain orientation was corrected by the rotation of the χ_1 torsion to bring the carboxylate close to the docked metal ion. All the conformations with a docked Mg^{2+} ion were refined using 10 ns unbiased simulations.

3.4. Results

3.4.1. Microstate and macrostate clustering of the trajectory. The first step of a discrete state-space MSM construction involves dividing the trajectory data into a set of microstates, based on structural similarity, with a fine-grained cluster cutoff. As noted in Section 3.2.2, diagonalization of the transition probability matrix obtained from the microstates as a function of the lag time, will yield relaxation timescales for the various modes of $\mathbf{T}(\tau)$. These relaxation timescales correspond to the timescales for transitions among the microstates. They show a strong dependence at short lag times but should become invariant at longer lag times (Bowman et al., 2009). The relaxation timescales for multiple lag times were examined to confirm that the relaxation times become independent of τ , which occurs at about 1 ns (Figure 3.1 A), i.e. the states are Markovian on this time scale. There are 2134 states in the microstate model with lag time 1 ns. Then, these microstates were lumped into varied numbers of macrostates using the PCCA+ algorithm (Deuffhard and Weber, 2005), and their relaxation timescales vs. lag time plotted for various numbers of macrostates. The relaxation timescales vs. lag time plot with 24 macrostates (Figure 3.1 B) follows that obtained from the microstate model. Equilibrium populations of the constituent microstates were added to get the macrostate populations (Figure 3.1 C and D). The equilibrium population of macrostate 23 outnumbers those of all other states (accounting for ~60% of the total population). The macrostates were constructed by lumping the microstates obtained from the clustering based on D53-D60 ($\beta 3\alpha 3$) conformers.

3.4.2. Residue distance based structural characterization of macrostates. A connectivity map among the macrostates was derived from the microscopic transition matrix, $\mathbf{T}$, after replacing the microstates with their corresponding macrostate. Matrix element $T_{ij} > 0$ implies that the macrostate i and j are dynamically connected in the network map (Figure 3.2). In order

to structurally characterize these macrostates, various inter-residue distances involving the $\beta 3\alpha 3$ and $\beta 4\alpha 4$ loops, and the $\alpha 4$ helix were examined to find a set that provides significant differences in their values in the 100 ns equilibrium simulations starting from the inactive and active form crystal structures. To this end, we chose 12 C α atom pairs where one atom is from a $\beta 3\alpha 3$ loop residue and another atom is from the stable part of the protein. Stable regions in the protein were determined from the residue-wise RMSFs in our previously equilibrium simulations starting from the active and inactive conformations. All these C α -C α distances satisfy the following criteria:

$$\begin{aligned} \sigma_A^{ij} - 2\delta_A^{ij} - (\sigma_I^{ij} + 2\delta_I^{ij}) &> 0 \text{ when } \sigma_A^{ij} > \sigma_I^{ij} \\ \sigma_I^{ij} - 2\delta_I^{ij} - (\sigma_A^{ij} + 2\delta_A^{ij}) &> 0 \text{ when } \sigma_I^{ij} > \sigma_A^{ij} \end{aligned} \quad (1)$$

where the σ_A^{ij} , σ_I^{ij} are the averages and the δ_A^{ij} , δ_I^{ij} are the standard deviations of the distances between C α atoms of the i^{th} and j^{th} residues during the equilibrium simulations starting from the active and inactive form crystal structures, respectively. This condition ensures that there is no overlap in the distance distribution as calculated from the active and inactive form simulations. Then, the average distance, $d_{\beta 3\alpha 3}^M(i, j)$, between the i^{th} and j^{th} C α atoms was calculated for a given macrostate, M. A new parameter was defined to monitor whether the macrostate is similar to the active/inactive like form based on these 12 distances that characterizes the $\beta 3\alpha 3$ conformation:

$$\Delta Q_{\beta 3\alpha 3}^M = \sum_{ij} f_{\beta 3\alpha 3}^M(i, j) \quad (2)$$

where

$$\begin{aligned}
f_{\beta 3 \alpha 3}^M(i, j) &= +1 \text{ when } \sigma_A^{ij} - 2\delta_A^{ij} < d_{\beta 3 \alpha 3}^M(i, j) < \sigma_A^{ij} + 2\delta_A^{ij} \\
&= -1 \text{ when } \sigma_I^{ij} - 2\delta_I^{ij} < d_{\beta 3 \alpha 3}^M(i, j) < \sigma_I^{ij} + 2\delta_I^{ij} \\
&= 0 \text{ otherwise}
\end{aligned} \tag{3}$$

In the $\Delta Q_{\beta 3 \alpha 3}^M$ scale an exactly inactive-like or active-like conformation will adopt the value -12 or +12, respectively. Similarly, another scale, $\Delta Q_{\alpha 4}^M$, was defined to characterize the $\beta 4 \alpha 4$ loop or $\alpha 4$ helix orientation with respect to the $\beta 3 \alpha 3$ loop. Note that differences in the $\beta 4 \alpha 4$ loop and/or the $\alpha 4$ helix are essentially equivalent. Residue pairs and C α -C α distances from the previous equilibrium trajectories starting with active and inactive form crystal structures are shown in Table B1, Appendix B.

Using these criteria, the macrostates most similar to the active and inactive forms can be identified. Macrostate 23 is identical to the inactive conformation in the sense that all the distances fall within the same distance distribution as in the inactive form simulation. Macrostate 16 is most similar to the active form, where both $\Delta Q_{\beta 3 \alpha 3}^M$ and $\Delta Q_{\alpha 4}^M$ adopts the value +6. Side-chain atom distances that characterize critical differences between the active and inactive forms further support the selection of macrostates 23 and 16 as inactive and active-like conformations, respectively (Figure B1, Appendix B). The other macrostates are likewise classified and displayed in a color coded connectivity map, shown in Figure 3.2. The macrostates were colored from red to violet according to the value in the $\beta 3 \alpha 3$ loop dependent parameter, $\Delta Q_{\beta 3 \alpha 3}^M$ and the $\alpha 4$ helix dependent parameter, $\Delta Q_{\alpha 4}^M$.

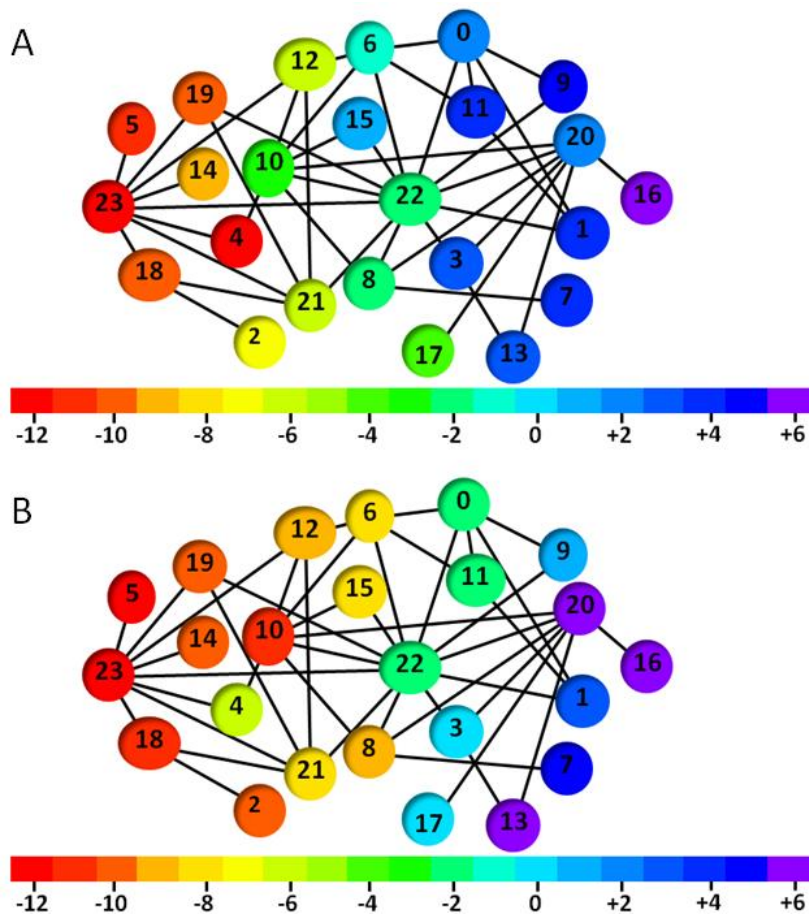


Figure 3.2. **Connectivity map.** Macrostates, color coded according to the (A) $\beta 3\alpha 3$ loop dependent parameter, $\Delta Q^M_{\beta 3\alpha 3}$ and (B) $\alpha 4$ helix dependent parameter, $\Delta Q^M_{\alpha 4}$. The connectivity between the macrostates was derived from the transition matrix.

3.4.3. Transition path theory (TPT) analysis. With the transition matrix and equilibrium probabilities available, transition path theory (TPT) can be used to investigate the sequence of intermediates that the protein follows in transiting between given end point states. Derivations of TPT have been outlined (Noé et al., 2009) and detailed elsewhere (E and Vanden-Eijnden, 2006; Metzner et al., 2009). It provides pathways, parameterized by the committor (splitting probability) values, spanning the transition from a beginning to an end state (Berezhtkovskii et al., 2009). The pathways are ordered in decreasing strength, with successive pathways defined by eliminating the bottleneck (rate limiting step) from the current pathway (Berezhtkovskii et al., 2009; Voelz et al., 2010). The highest population microstates for macrostates 23 and 16 were chosen, respectively, as initial and final states for the transition path calculation. Transition paths were initially obtained in terms of the microstates, and later mapped and replaced by the assigned macrostates for each microstate. There are 22 different paths found for the chosen end states. Paths and corresponding reactive fluxes are shown in Table B4, Appendix B. Note that only 12 out of the 24 macrostates are involved in these pathways (Figure 3.3). Averages over the committor values for all constituent microstates were calculated and assigned to the corresponding macrostate (Table B5, Appendix B).

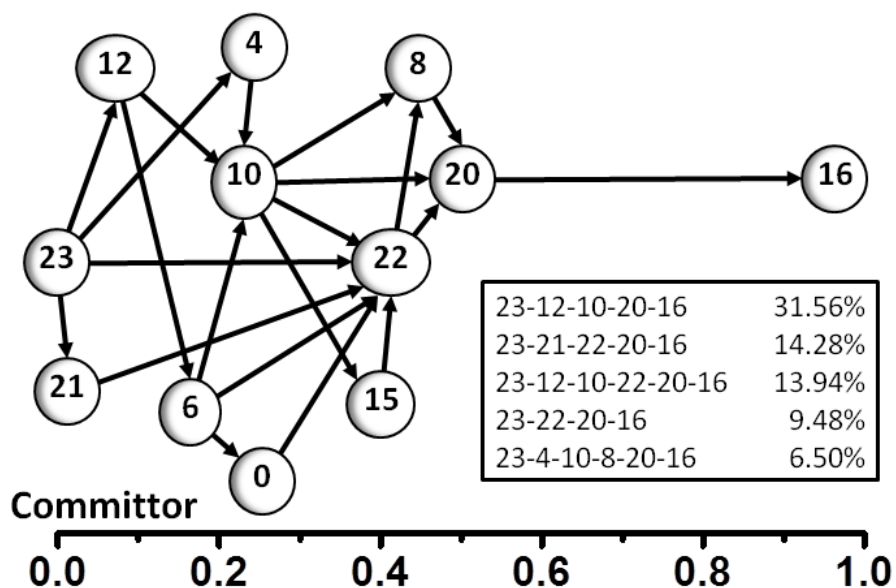


Figure 3.3. **Transition path.** Pathways connecting macrostates 23 (inactive like) and 16 (active like). There are a total of 22 pathways that involve 10 intermediate metastable states; 75% of the total transitional flux is through the top five pathways listed here. Commitor probabilities for all the macrostates were calculated by transition path theory.

3.5. Discussion

3.5.1. Conformational sampling by nonequilibrium simulation followed by multiple equilibrium simulations. High resolution crystal structures of RR468 characterize the conformational difference between the inactive and active states (Casino et al., 2009) of the protein, where a phosphoryl mimic BeF_3 , attached to the site of phosphorylation (D53) and a Mg^{2+} ion stabilizes the active conformation. The inactive conformation and the active conformation without the BeF_3 and Mg^{2+} ion were used as starting conformations for the equilibrium MD simulations. The conformations are trapped close to their respective starting structures in these simulations; hence provide no insight into the process of conformational transition. However, these simulations provide a great deal of understanding on the conformational plasticity associated with the active and inactive states of the protein (Banerjee et al., 2014). To connect these endpoints, we were able to pull the inactive-like conformation to the active-like one in a TMD simulation, using RMSD-based restraints on all heavy atoms of the protein (Banerjee et al., 2014). To provide conformational sampling spanning experimentally observed states, some nonequilibrium simulations followed by multiple short equilibrium MD simulations have proven useful (Ma and Cui, 2007; Yang et al., 2009). We accumulated conformational snapshots from multiple, unbiased, all-atom MD trajectories from 101 points along the TMD trajectory to sample the conformational states of RR468, as detailed in Section 3.3.1. Thus, the conformational sampling is free from any external potential as the biasing potential was eliminated before the production simulation to accumulate data for our MSM clustering and TPT analysis.

3.5.2. Macrostates 23 and 16, respectively, most resemble the inactive and active like conformations. Macrostate 23 closely resembles the inactive conformation. All the critical $\text{C}\alpha$ -

C α distances that distinguish the inactive conformation from the active one during the equilibrium simulations with crystal structures as the starting conformation match those in macrostate 23. The system, RR468 without a phosphoryl group (or a phosphoryl mimic, BeF₃) as in our simulation, should favor the inactive like conformation. So, it is not surprising that the population (~60%) of macrostate 23 outnumbers all other macrostates. The conformational state observed in the crystal structure of the active conformation is stabilized by phosphoryl-mimicking BeF₃ and Mg²⁺. It is a fair assumption that without these stabilizing factors the active conformation is a high energy conformation. It is indeed reflected in the population of the active-like macrostate, 16 (1.14%), that is much smaller than the equilibrium population of the inactive-like macrostate, 23. Moreover, all the critical C α -C α distances in the macrostate 16 do not match those in the active form simulation. A truly active like conformation should adopt the value +12 in both the $\Delta Q_{\beta 3 \alpha 3}^M$ and $\Delta Q_{\alpha 4}^M$ scales but macrostate 16 adopts +6 for both of these parameters.

3.5.3. Critical macrostate intermediates along the transition path can be parameterized by

$\Delta Q_{\beta 3 \alpha 3}^M$ and $\Delta Q_{\alpha 4}^M$ values. We have investigated the network map and the transition paths to identify critical intermediates in the conformational transition. Macrostates 21, 10, 22, 20 were chosen as hub-like centers based on the features of having large numbers of connections in the network map (Figure 3.2) and large numbers of inward and/or outward fluxes in the transition path (Figure 3.3). These states, along with the beginning and ending states, 23 and 16, respectively, were investigated for their structural features and critical interactions. The parameters $\Delta Q_{\beta 3 \alpha 3}^M$ and $\Delta Q_{\alpha 4}^M$ defined in the Section 3.4.2, indicate conformational states based on two critical structural elements, i.e. the $\beta 3 \alpha 3$ loop and $\alpha 4$ helix. These parameters are very useful to assign whether any of these structural elements in a given macrostate look more like the

active or inactive conformations. A similar strategy was proven useful in a study on Src kinase (Yang et al., 2009). Apart from the two extremes of $\Delta Q_{\beta 3\alpha 3}^M$ or $\Delta Q_{\alpha 4}^M$, their values close to zero mean that the conformation does not resemble the active or inactive-like conformation. Values of $\Delta Q_{\beta 3\alpha 3}^M$ and $\Delta Q_{\alpha 4}^M$ in a given macrostate are not strongly correlated. It appears that these two structural elements move rather independently of each other, hence they provide two reaction coordinates to characterize the conformation of the intermediate states (Table B5, Appendix B). It is also evident from the crystal structures of the inactive and active forms that there is no strong interaction between these two structural elements, except for the highly conserved T83 (adjacent to the $\beta 4\alpha 4$ loop) that interacts with the phosphoryl mimic group, BeF_3 , in the active form crystal structure. It is believed that highly conserved T83 residue plays an important role in neutralizing excess negative charge associated with the phosphoryl group and provides stability to the phosphorylated aspartate residue (Casino et al., 2009).

3.5.4. Macrostate 22 is the most stable intermediate where the $\beta 3\alpha 3$ loop conformation is significantly different from the inactive or active state. Macrostate 22 adopts “-2” values in both the $\Delta Q_{\beta 3\alpha 3}^M$ and $\Delta Q_{\alpha 4}^M$ scales. It is to be noted here that macrostate 22 is the second largest macrostate, with 13.4% population after macrostate 23 (Table B5, Appendix B). The $\beta 3\alpha 3$ loop backbone in macrostate 22 is significantly different from that of macrostate 23. As a result the backbone carbonyl group of M55 from the $\beta 3\alpha 3$ loop becomes close to three highly conserved carboxylate residues, D9, D10 and D53, in the active site. Rotation in the backbone torsion angle, Ψ of I54 accounts for the change in the $\beta 3\alpha 3$ loop backbone (Figure 3.4). A divalent cation is necessary for the phosphorylation and dephosphorylation activity of RR (Bourret, 2010). In RR468, the metal ion Mg^{2+} is present in the active form crystal structure along with

phosphoryl mimic BeF_3^- (Casino et al., 2009). The six coordination positions of the Mg^{2+} ion are occupied in part by the three conserved aspartate residues, D9, D10, and D53. One of these aspartates, D9 interacts through a water molecule. The backbone carbonyl group of M55 from the $\beta 3\alpha 3$ loop and two water molecules are also within the coordination sphere of the Mg^{2+} . It has been proposed (Bourret, 2010) that one of these water molecules is replaced by the phosphoryl group during the phosphoryl transfer reaction to yield a conformation similar to that observed in the active form crystal structure. This kind of active site architecture is very common among other RR systems as well (Bourret, 2010). During the simulation of RR468 in the absence of a Mg^{2+} ion, all the conserved aspartates in the active site remain close to each other with the exception of the carboxylate group of D10 that points outward by 180° rotation of side-chain χ_1 , due to repulsion with that of residue D53. The $\beta 3\alpha 3$ loop backbone, especially the backbone carbonyl group of M55, comes close to the active site, comprised of the three conserved aspartate residues D9, D10 and D53 in macrostate 22 and remains there in macrostates 20 and 16, unlike in macrostates 23, 21 and 10. It is clear that the $\beta 3\alpha 3$ loop in macrostates 22, 20, and 16 is close to the active site, unlike that in macrostates 23, 21, and 10, as a result of the change in the I54 backbone torsion angle (Figure 3.4).

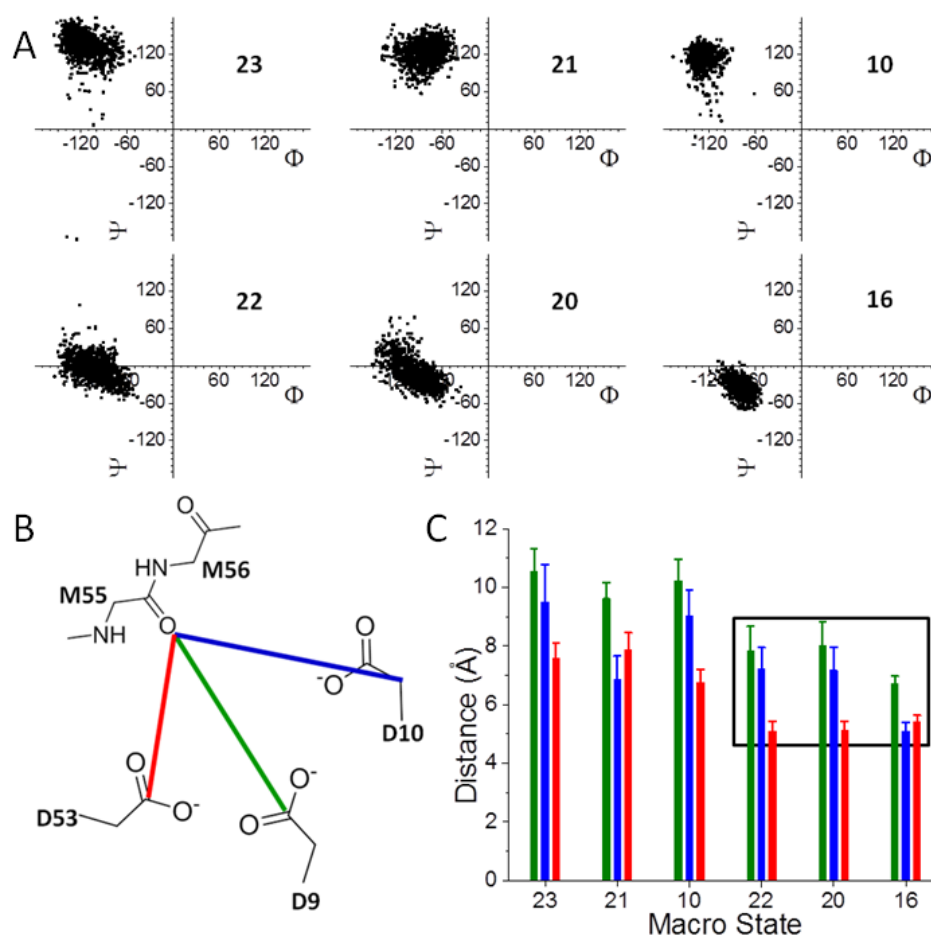


Figure 3.4. **Conformational change in macrostate 22.** (A) Change in backbone torsion angle of I54 along the transition path. The I54 Ψ angle approximates the inactive form value in states 23, 21 and 10. As a result, the M55 carbonyl group is away from the active site in these states. In states 22, 20 and 16, the changes in the I54 Ψ angle brings the M55 carbonyl group close to other active site residues. (B) The $\beta 3\alpha 3$ loop backbone comes close to the active site, comprised of the three conserved aspartate residues D9, D10 and D53 in macrostates 22, 20 and 16, unlike in macrostates 23, 21 and 10. (C) The average distance between the M55 carbonyl group and side-chain atoms of D9, D10 and D53 are shown in green, blue and red, respectively.

3.5.5. Critical intermediates preceding macrostate 22. There are two highly connected macrostates in the connectivity map (Figure 3.2) that appear before macrostate 22 along the committor axis (Figure 3.3); namely, macrostates 21 and 10. In macrostates 21 and 10, there are variable amounts of displacement in two structural elements, the $\beta 3\alpha 3$ loop and $\alpha 4$ helix. Macrostate 21 and 10 adopt (-6, -8) and (-3, -11) values, respectively, for the $\Delta Q_{\beta 3\alpha 3}^M$, $\Delta Q_{\alpha 4}^M$ parameters. In terms of the $\beta 3\alpha 3$ loop, macrostate 10 is very similar to macrostate 22 ($\Delta Q_{\beta 3\alpha 3}^M = -3$ in macrostate 10 compared to $\Delta Q_{\beta 3\alpha 3}^M = -2$ in macrostate state 22). But in terms of the $\alpha 4$ helix conformation, macrostate 10 remains similar to the inactive like state, 23 ($\Delta Q_{\alpha 4}^M = -11$ in macrostate 10 compared to $\Delta Q_{\alpha 4}^M = -12$ in macrostate 23). Conformations of the $\beta 3\alpha 3$ loop and $\alpha 4$ helix, in macrostate 21 ($\Delta Q_{\beta 3\alpha 3}^M = -6, \Delta Q_{\alpha 4}^M = -8$) are about halfway between macrostates 23 ($\Delta Q_{\beta 3\alpha 3}^M = -12, \Delta Q_{\alpha 4}^M = -12$) and 22 ($\Delta Q_{\beta 3\alpha 3}^M = -2, \Delta Q_{\alpha 4}^M = -2$). We have examined the $\Delta Q_{\beta 3\alpha 3}^M$, $\Delta Q_{\alpha 4}^M$ parameters in other macrostates as well that appear before macrostate 22 along the committor axis (macrostates 0, 4, 6, 12, 15). It appears that in macrostates 0 (+2, -2), 4 (-12, -6), 6 (-1, -8), 12 (-6, -9) and 15 (+1, -8) there is no concerted motion in the $\beta 3\alpha 3$ loop and $\alpha 4$ helix as these parameters (shown in parenthesis) do not show any trend.

3.5.6. Macrostate 20 exhibits enhanced flexibility of the $\beta 3\alpha 3$ loop. The $\beta 3\alpha 3$ loop is implicated in the phosphoryl transfer reaction (Bourret, 2010; Casino et al., 2009). This loop appears to be crucial in two possible ways. Firstly, the M55 (from the $\beta 3\alpha 3$ loop) backbone carbonyl group interacts with the Mg^{2+} ion that is essential for the phosphorylation reaction. Secondly, movement in the loop can affect the exposure of the site of phosphorylation. Unwinding of structural elements adjacent to the loop may significantly affect the flexibility of

the loop. From the equilibrium simulations it was found that D60 is part of the $\alpha 3$ helix in the inactive conformation but it is part of the $\beta 3\alpha 3$ loop in the active conformation. Loss of backbone interaction between D60 and V64 at the early stage of the TMD simulation also points towards the weakening of the N-terminus of the $\alpha 3$ helix (Banerjee et al., 2014) The number of occurrences of the characteristic backbone hydrogen bonds between D60 and V64, F62 and K66, and T63 and K67 at the N-terminus of the $\alpha 3$ helix were calculated for all the crucial macrostates. It was found that the N-terminus of the $\alpha 3$ helix is significantly weaker in macrostate 20 as compared to the other macrostates (Figure 3.5 A, B).

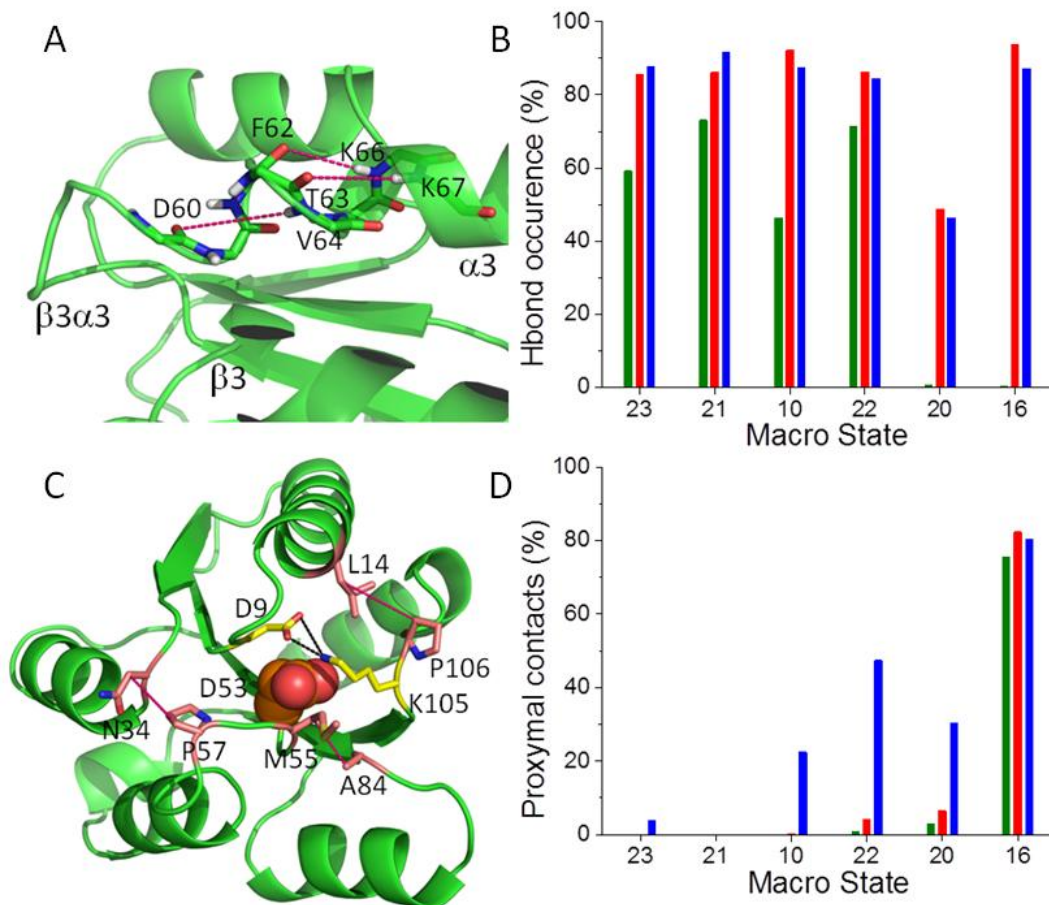


Figure 3.5. **Conformational change in macrostate 20 and 16.** (A) Characteristic helix backbone hydrogen bonds between D60 and V64, F62 and K66, T63 and K67, were counted in each critical macrostate. Distance and angular cutoffs of 3.3 Å and 130°, respectively, were used as hydrogen bond criteria in Wordom. (B) Occurrence (%) of helix backbone hydrogen bonds between D60 and V64, F62 and K66, and T63 and K67, at the N-terminus of the $\alpha 3$ helix are shown in green, red and blue, respectively. (C) Interactions among selected proximal residues were counted in critical macrostates. The number of snapshots where residue pairs L14 ($\alpha 1$ helix) and P106 ($\beta 5\alpha 5$ loop), N34 ($\beta 2\alpha 2$ loop) and P57 ($\beta 3\alpha 3$), M55 ($\beta 3\alpha 3$) and A84 ($\beta 4\alpha 4$) have van der Waals contact were counted in each critical macrostate using Wordom. Residues L14, P106, N34, P57, M55, and A84 are shown in pink.

Figure 3.5 (cont'd)

Site of phosphorylation, D53, is shown in CPK. A salt bridge between K105 and D9 (in yellow) is another important feature of the active conformation. The average distance between the N ζ and O δ atoms of the K105 and D53 residues during the active and inactive form simulations, and in all critical macrostates, are listed in Table 3.1. (D) Occurrence (%) of inter-loop contacts between residues L14 and P106, N34 and P57, and M5 and A84 in critical macrostates are shown in green, red, and blue, respectively.

3.5.7. Macrostate 16 has a stable $\beta 3\alpha 3$ loop due to proximal residue interactions. The backbone hydrogen bonds between F62 and K66, and T63 and K67, at the N-terminus of the $\alpha 3$ helix are restored in macrostate 16, unlike that in macrostate 20 (Figure 3.5 A, B). However, the D60 residue remains as part of the $\beta 3\alpha 3$ loop in macrostate 16. Salient features of macrostate 16 are the explicit contacts among the proximal residues. A count of van der Waals contacts among inter loop residue pairs shows that the residues from the proximal side close down around the site of phosphorylation, D53, (Figure 3.5 C, D) in macrostate 16. Most of the residues that form the proximal contacts are hydrophobic in nature. These hydrophobic interactions, involving L14, V18, L22, L39, M59 and F107, were observed during the equilibrium simulation starting from the active conformation as well (Banerjee et al., 2014). A salt bridge between the conserved K105 ($\beta 5\alpha 5$ loop) and D9 ($\beta 1$) residues from a distant part of the protein is observed in macrostate 16. This salt bridge also is a critical feature of the active form, as observed from the equilibrium simulation, and was not present at all in the simulation with the inactive form as the starting conformation (Banerjee et al., 2014). The salt bridge between the residues from the analogous position is closely related to the stability of the phosphorylated RR in a related response regulator protein, CheY (Lukat et al., 1991).

3.5.8. Proximal contacts control exposure of the D53 side-chain to bulk water. The solvent accessible surface (ASA) of D53 was calculated from the equilibrium simulation trajectories starting from the active and inactive form crystal structures (Table 3.1). It was found that the active conformation is a closed conformation where the site of phosphorylation, D53, is less exposed to water, as compared to that in the inactive conformation. Proximal interactions including the salt bridge and the hydrophobic contacts stabilize the closed conformation and isolate the site of phosphorylation from bulk solvent. In the phosphorylated form of RR, the

phosphoryl group on D53 is protected due to the proximal interactions. In the absence of the proximal interactions the active site is more exposed and water can act as a strong nucleophile to attack the phosphorus atoms that leads to dephosphorylation of the phosphorylated aspartate. In terms of solvent accessibility of the D53 residue, macrostate 16 closely resembles the active form conformation, where a strong salt bridge between K105 and D9 facilitates interactions among other proximal residues (Table 3.1). Ligand free simulation followed by Markov state model analysis was able to identify a conformation similar to that in the ligand bound states in three different proteins reported in a recent report.(Bowman and Geissler, 2012)

Table 3.1. **Accessible surface area (ASA) calculated for D53 and the distance between the K105 and D9 side-chains.** Distances were calculated for simulations with inactive and active starting conformations in comparison with those in the six critical macrostates.

Conformation	ASA of D53 ($\text{\AA}^2$)	K105 N ζ –O δ D53 ($\text{\AA}$)
Inactive (100 ns equilibrium trajectory)	22.24 $\pm$ 8.09	5.44 $\pm$ 1.42
Active (100 ns equilibrium trajectory)	6.63 $\pm$ 2.78	3.08 $\pm$ 0.29
Macrostate 23	22.72 $\pm$ 8.71	5.95 $\pm$ 1.19
Macrostate 21	14.85 $\pm$ 6.22	6.13 $\pm$ 0.83
Macrostate 12	22.87 $\pm$ 5.37	5.83 $\pm$ 1.09
Macrostate 10	24.16 $\pm$ 7.31	3.20 $\pm$ 0.82
Macrostate 22	11.98 $\pm$ 4.77	4.31 $\pm$ 1.28
Macrostate 20	12.28 $\pm$ 6.49	4.44 $\pm$ 1.48
Macrostate 16	1.87 $\pm$ 1.35	2.76 $\pm$ 0.22

3.5.9. Macrostate 20 is a transition state separating inactive and active-like states.

Macrostate 16 is found to be an isolated state from the connectivity map, and from the transition paths (Figures 2 and 3). All the transition paths that connect the inactive like macrostate 23 to the active like macrostate 16 are through state 20. However, on the committor axis, macrostate 20 is not very close to macrostate 16. A committor by definition is a measure of the kinetic distance between an intermediate conformation and the beginning and end states (Du et al., 1998). From the definition of committor, any intermediate state with committor probability more than 0.5 is likely to first reach the final state before reaching the initial state, and is therefore kinetically closer to the final state. A similar argument holds for committor probability less than 0.5. Thus, a state with committor probability 0.5 is as likely to reach the initial and final states. It is therefore reasonable to define such a state as a transition state (Du et al., 1998). The committor probability for macrostate 20 is 0.49 (Figure 3.2 and Table B5, Appendix B) and may therefore be viewed as a transition state connecting beginning and final states, i.e. macrostates 23 and 16, respectively. In macrostate 20, the characteristic helix backbone interaction at the N-terminus of the $\alpha 3$ helix is significantly weak, as discussed in Section 3.5.6. As a result, the $\beta 3\alpha 3$ loop orientation with respect to the rest of the protein can change significantly in macrostate 20, as compared with other macrostates. However, the $\beta 3\alpha 3$ loop in macrostate 20 does not have as many contacts with its surroundings proximal residues as in the following, final state, macrostate 16, as discussed in Section 3.5.7. Consequently, macrostate 20 appears to be a transition state where the flexibility in the $\beta 3\alpha 3$ loop facilitates protein-protein interactions, and can be stabilized only in the presence of its cognate histidine kinase, HK853. In contrast, macrostate 16 closely resembles the active-like, closed conformation as observed in the presence of phosphoryl group and Mg^{2+} .

3.5.10. Free energy landscapes can suggest a most stable intermediate. Projections of the free energy landscape onto a few reaction coordinates can provide free energy barriers associated with the process of conformational transition. Such projections are meaningful if an appropriate reaction coordinate is chosen; however, this kind of projections will be quite misleading with the wrong choice of the reaction coordinates. Conformational differences among active and inactive-like conformations are mainly concentrated in the $\beta 3\alpha 3$ loop and the $\alpha 4$ helix. Residue-distance-based structural characterization of all the macrostates proved very useful, as discussed in Section 3.4.2, where the distances were chosen that either describe the absolute orientation of the $\beta 3\alpha 3$ loop or relative orientation of the $\alpha 4$ helix with respect to the $\beta 3\alpha 3$ loop (all the residue pairs and their average distances are given in Tables B2 and B3, Appendix B). Two pairs of such distances were chosen as reaction coordinates to construct free energy surfaces from the entire pool of MD data used for the MSM analysis. For example, the distance between M56 ($\beta 3\alpha 3$) and A32 is a descriptor for the orientation of the $\beta 3\alpha 3$ loop with respect to a core residue, A32, the latter having minimal fluctuation during the simulation. Similarly, the distance between P57 ($\beta 3\alpha 3$) and D90 ($\alpha 4$) describes the relative orientation of these two structural elements. Each of these pairs of distances make good two-dimensional reaction coordinates because we found that, in each pair, the distances vary quite independently of each other. The projection of the free energy landscape on these reaction coordinates (Figure 3.6 A), where one characterizes the $\beta 3\alpha 3$ conformation and other characterizes the $\alpha 4$ conformation, enabled us to locate macrostate 22 as a metastable intermediate state, from the characteristic distances in macrostate 22 (Tables S2 and S3). We also examined the projection of the free energy landscape on a different pair of reaction coordinates, so as to not be misled by any arbitrary choice of the reaction coordinates (Figure 3.6

B). In both projections the end states, macrostate 23 and 16, can be located as stable states, along with the stable intermediate, macrostate 22.

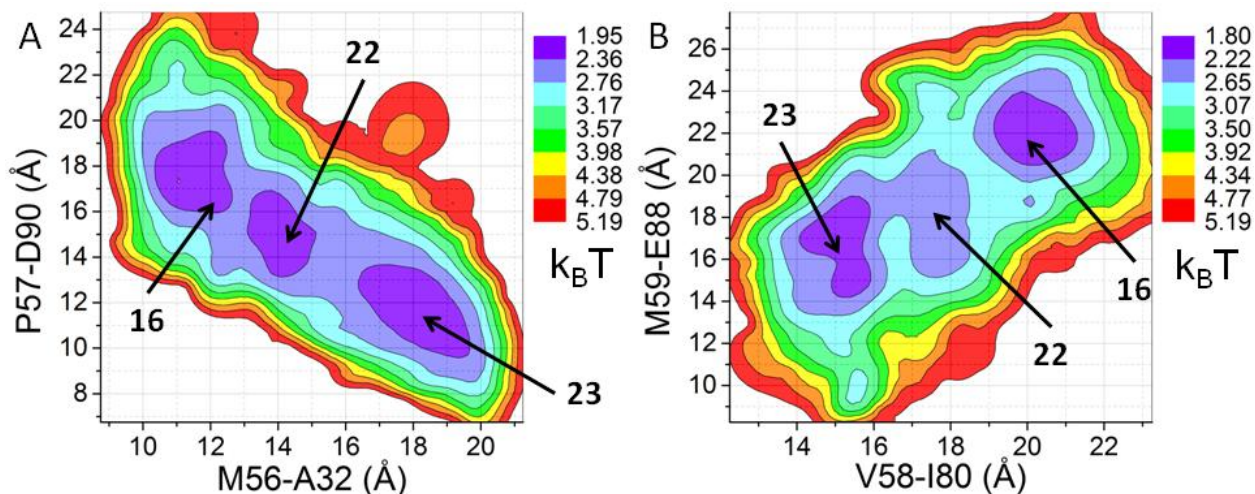


Figure 3.6. **Free energy landscape.** Projection of the free energy landscape on two reaction coordinate pairs. In each case, one coordinate for the $\beta 3\alpha 3$ loop conformation (on the X-axis) and the other for the $\beta 4\alpha 4$ loop/ $\alpha 4$ -heix conformation (on the Y-axis). The residues A32 and I80 are from the core region of the protein and have minimal fluctuation during equilibrium simulations. Hence, the distances between M56 ($\beta 3\alpha 3$) and A32 in (A) or V58 ($\beta 3\alpha 3$) and I80 in (B) are descriptors for the $\beta 3\alpha 3$ loop orientation. Similarly, the distances between P57 ($\beta 3\alpha 3$) and D90 ($\alpha 4$) in (A) or M59 ($\beta 3\alpha 3$) and E88 ($\alpha 4$) in (B) are descriptors for the relative orientation of the $\beta 3\alpha 3$ loop and $\alpha 4$ helix. The two dimensional free energy surfaces constructed from the simulation data were able to capture macrostate 22 as distinct from macrostate 23 (inactive like) and macrostate 16 (active like). These average distances in all the crucial macrostates are given in Tables S2 and S3.

3.5.11. Active site architecture in the presence of a Mg^{2+} ion. The phosphoryl transfer reaction from HK to RR requires a divalent cation, which is either Mg^{2+} or Mn^{2+} in most cases (Bourret, 2010). According to the proposed phosphoryl transfer mechanism, before phosphorylation the six coordination positions of the metal ion bound to RR are occupied by the three conserved aspartate residues (one acts through a water molecule), a backbone carbonyl group, and two water molecules, as discussed in Section 3.5.4. During phosphorylation, an oxygen atom in the phosphoryl group replaces one of the waters in the metal ion coordination sphere. (Bourret, 2010) In the active form crystal structure of RR468 that represents the phosphorylated, or the product form, the Mg^{2+} ion interacts with the three conserved aspartate side-chains (D9, D10, and D53), the M55 backbone carbonyl group and the phosphoryl-mimic BeF_3 , attached to D53.

To investigate this active site architecture before phosphorylation, an Mg^{2+} ion was docked in four randomly selected apo protein conformations, for each of the macrostates, 23, 22, 20 and 16. The states were chosen for the following reasons. Macrostates 23 and 16 represent, respectively, inactive and active-like states, and macrostates 22 and 20 represent intermediate conformations significantly different from macrostates 23 and 16, in both the $\Delta Q_{\beta 3 \alpha 3}^M$ and $\Delta Q_{\alpha 4}^M$ scales. The active site architecture and the Mg^{2+} ion coordination were investigated after structural refinement with 10 ns MD simulations (Figure 3.7). In all the macrostates, the Mg^{2+} ion interacts with the three conserved aspartates (D9, D10, and D53). Unlike in macrostate 23, in macrostates 22, 20, and 16, the $\beta 3 \alpha 3$ loop backbone at M55 is close to the active site, i.e. the three conserved aspartates, as discussed in Section 3.5.4. In macrostates 22, 20, and 16, the carbonyl group of M55 is capable of interaction with the Mg^{2+} , due to a large conformational change that brings the M55 backbone close to the conserved aspartates in the active site (Figure 3.4). Binding of Mg^{2+} to the RR before phosphorylation was reported to drive a large conformational change in the

$\beta 4\alpha 4$ loop and the $\alpha 4$ helix in chemotaxis protein CheY (Bellolell et al., 1994). In the crystal structure of the NarL family response regulator, Spr1814, in the presence of only Mg^{2+} ion (but not BeF_3 or any equivalent group) it was observed that the E55 (equivalent to M55 in RR468) backbone interacts with the metal ion (PDB id 4E7O). In this crystal structure also, the N-terminus of the $\alpha 3$ helix was found to be very stable, as observed for macrostate 22 along the RR468 transition path, as discussed in Section 3.5.6.

In the proposed reaction mechanism, one water molecule in the coordination sphere of the Mg^{2+} ion is replaced by the oxygen atom of the phosphoryl group during the phosphoryl transfer reaction (Bourret, 2010). From this consideration, the conformation in macrostates 22, 20, and 16 could qualify as a suitable substrate for the phosphoryl transfer reaction. However we rule out macrostate 16 based on the fact that the Mg^{2+} binding in macrostate 16 is too tight, involving the backbone carbonyl groups of M55 and M56. This leaves hardly any room for interaction with HK, where interacting groups (D53 from RR468 and H260 from cognate HK853) can line up for a phosphoryl transfer reaction. It is evident from the accessible surface area calculation for D53 as well that macrostate 16 represents a completely closed conformation. We suggest that macrostate 16 mimics the phosphorylated form of the protein where, in absence of the phosphoryl group as in our simulation conditions, the M56 backbone carbonyl is attracted toward the Mg^{2+} ion. We select macrostate 20 as the transition state over macrostate 22, based on the transition path analysis, and subsequent structural characterization based on $\beta 3\alpha 3$ loop flexibility, as discussed in Section 3.5.6.

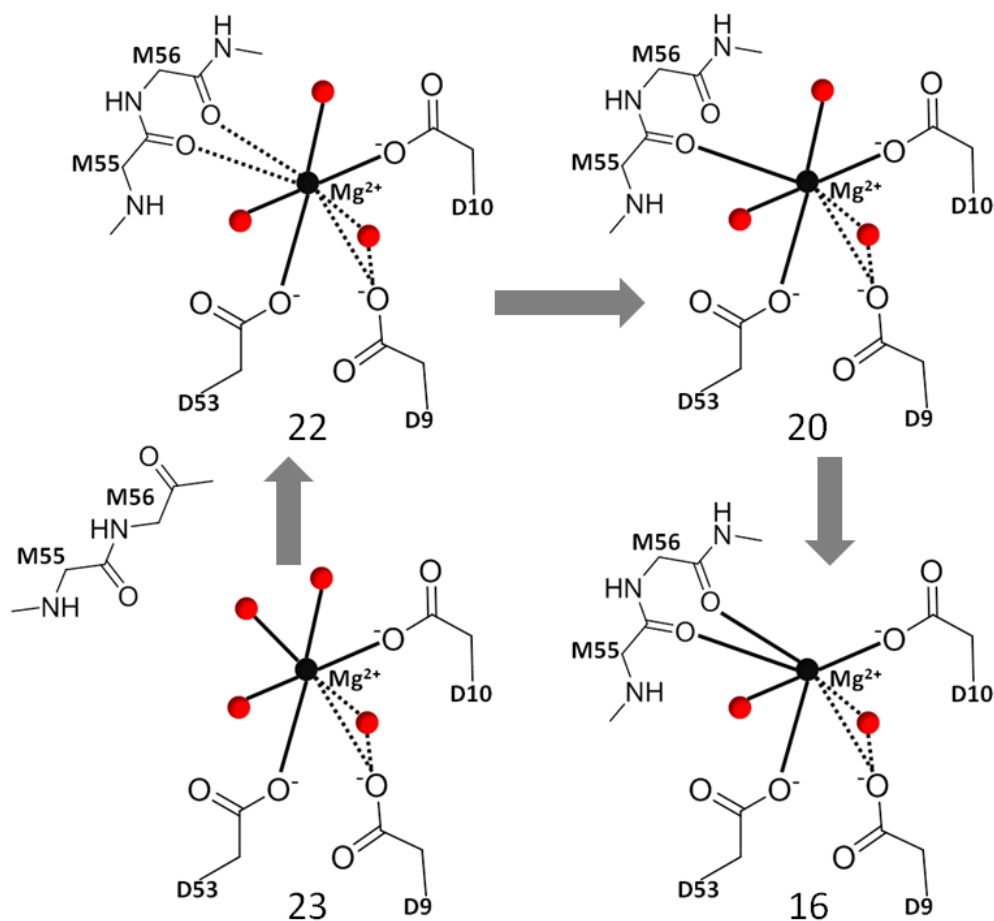


Figure 3.7. **Active site architecture in presence of Mg^{2+} .** Active site architecture after docking an Mg^{2+} ion in the active site followed by 10 ns equilibration simulation for refinement. Four randomly selected conformations from each state were used for this purpose. Metal ion interactions observed in all conformations for a given state are shown in solid line. There are instances found after refinement where the interaction between D9 and the Mg^{2+} ion is either direct or water mediated. Similarly, for macrostate 22, either M55 or M56 has an interaction with Mg^{2+} , depending on the starting conformation.

3.6. Conclusions

The inactive conformation of RR468 becomes phosphorylated in the phosphoryl transfer reaction by its cognate histidine kinase, HK853. In other words, the inactive conformation of RR468 acts as the substrate for the phosphoryl transfer reaction. RR systems in general are known to be the active participants in the phosphoryl transfer reaction and can be phosphorylated by various small molecule phosphodonors (Bourret, 2010; Wolfe, 2010). A completely exposed active site as observed in the inactive (unphosphorylated) form crystal structure (and during MD of the inactive form) may not make a good substrate for the phosphoryl transfer reaction, where phosphorylated aspartate will immediately get dephosphorylated due to nucleophilic attack by water molecules. Furthermore, the $\beta 3\alpha 3$ loop, as observed in the inactive form crystal structure, is far from the active site and unlikely to have any interaction with the divalent ion if present. Thus, it is likely that RR468 should undergo a conformational transition in order to act as a substrate for the phosphoryl transfer reaction. To investigate this possibility and propose a mechanism for this conformational transition, we have used nonequilibrium molecular dynamics simulation followed by multiple equilibrium simulation with the protein alone (without the bound ligand). These trajectories provide sample conformations along the transition path and, by discretizing the resulting conformations in kinetically related clusters, and using a committor analysis, a set of metastable states were identified to provide a potential reaction path. Apart from the inactive-like state that resembles the unphosphorylated form crystal structure, we were able to identify a small population of the active like state that resembles the phosphorylated form crystal structure, from inter-residue distance dependent parameters. Two critical intermediates were identified along major transition paths. One of these intermediates (macrostate 20) can be categorized as a high energy transition state as obtained from the transition path analysis. The

other intermediate (macrostate 22) is a stable precursor to this transition state. The structural features of the inactive and active-like states along with the intermediates are summarized in Table 3.2.

As a general strategy we propose here a very effective way to deal with the huge amount of conformational sampling data obtained through the MD simulations. We started with a set of ~150 000 MD snapshots that are clustered first by using a structural metric. We ended up with 2134 microstates that represent the kinetics of the system with a coarse graining, lag time. However, this large number of microstates is not suited to obtaining insight in the process of a conformational transition. We lumped these microstates into 24 macrostates based on the rates of transitions among these states. Transition path analysis led us to identify only 12 states that take part in the process of conformational transition among the specified end states. We were able to identify 6 states that represent critical hub points along the transition path and the connectivity network. All the 6 critical states were characterized for their structural features. There are 4 among these 6 states that are totally distinct in terms of two structural elements in the system, the $\beta 3\alpha 3$ loop and $\alpha 4$ helix. The other two states represent inactive-like states, where the change along the $\beta 3\alpha 3$ loop and $\alpha 4$ helix appear to occur in a different order or extent. The active site architecture in four of these states was tested in the presence of a divalent cation, essential for the function of the protein. Two intermediate states were identified based on structural features and committor analysis, where one of these states represents the transition state and the other represents a stable precursor to the transition state. The two dimensional free energy landscapes also appropriately identified the stable precursor to the transition state, along with the stable inactive and active-like conformational states of the protein.

Table 3.2. **Structural features of four critical macrostates identified from the MSM and TPT analysis, and subsequent docking and structural refinement.**

Structural feature	23	22	20	16
Position of M55 backbone carbonyl with respect to D9, D10, and D53 side-chains	Far	Close	Close	Close
M55 backbone interaction with docked Mg^{2+}	No	Yes	Yes	Yes
Secondary structure of D60	$\alpha 3$	$\alpha 3$	$\beta 3 \alpha 3$	$\beta 3 \alpha 3$
Helix backbone hydrogen bond of F62, T63	Strong	Strong	Weak	Strong
Proximal contacts	Weak	Moderate	Moderate	Strong
Accessibility of the D53 to the bulk water	Highly accessible	Moderately accessible	Moderately accessible	Not accessible

APPENDIX

Table B1. **Critical C α - C α distances.** C α - C α distances between residues that were used to characterize the $\beta 3\alpha 3$ and $\beta 4\alpha 4$ loop/ $\alpha 4$ helix conformation/orientation, i.e. for the calculation of $\Delta Q_{\beta 3\alpha 3}^M$, $\Delta Q_{\alpha 4}^M$ parameters. The average distances were calculated over the 100 ns equilibrium simulations ACT and IAC, starting from the active and inactive form crystal structures, respectively.

Index	$\beta 3\alpha 3$ loop				$\beta 4\alpha 4$ loop/ $\alpha 4$ helix			
	Res1	Res2	ACT dist (Å)	IAC dist (Å)	Res1	Res2	ACT dist (Å)	IAC dist (Å)
1	M56	V8	9.09±0.65	15.49±0.77	E88	M56	19.44±1.04	11.20±3.05
2	M56	A32	11.17±0.65	18.78±1.03	S92	V58	23.01±0.80	12.19±1.27
3	M56	L42	17.26±0.71	24.05±0.97	E88	P57	22.30±1.17	11.80±3.11
4	M56	L52	9.27±0.60	13.36±0.49	E89	P57	21.12±1.42	11.18±2.59
5	P57	A32	12.57±0.95	20.14±1.25	D90	P57	17.76±1.19	9.45±2.69
6	P57	L42	18.98±1.06	24.70±1.04	E91	P57	19.23±0.93	11.84±2.38
7	P57	A94	17.69±0.61	12.11±1.26	A84	V58	16.88±0.99	9.80±1.43
8	V58	F62	12.33±0.32	6.89±0.85	E88	V58	24.13±1.30	11.89±2.27
9	V58	I80	20.71±0.62	15.81±1.25	E89	V58	22.48±1.68	11.02±1.64
10	V58	V81	17.97±0.48	12.15±1.25	D90	V58	19.26±1.34	8.49±2.10
11	V58	T83	16.45±0.70	9.54±1.42	E91	V58	20.99±0.94	10.44±1.79
12	V58	A94	18.82±0.56	9.40±1.45	E88	M59	22.05±1.16	14.64±2.12

Table B2. **Critical $\beta 3\alpha 3$ loop distances in macrostates.** Critical distances to characterize the $\beta 3\alpha 3$ loop conformation in critical macrostates. Average distances were measured in 6 critical macrostates. Distances M56-A32 and V58-I80 (shown in bold) were used to project the free energy landscape. On the free energy landscape the critical macrostates were located based on their representative distance (shown in bold).

Index	Res1	Res2	23 (Å)	21 (Å)	12 (Å)	10 (Å)	22 (Å)	20 (Å)	16 (Å)
1	M56	V8	15.20	13.20	12.89	12.56	11.75	11.83	10.41
2	M56	A32	18.22	15.04	14.96	14.71	14.08	14.11	12.26
3	M56	L42	23.74	21.01	21.20	20.80	20.99	21.00	19.55
4	M56	L52	13.39	12.76	12.42	12.09	11.59	11.59	11.09
5	P57	A32	19.37	16.34	16.13	15.17	13.48	13.58	11.81
6	P57	L42	23.92	21.07	21.04	20.44	19.77	20.40	18.69
7	P57	A94	12.55	13.82	14.57	14.94	15.05	15.91	17.37
8	V58	F62	6.36	6.36	6.51	8.34	7.79	8.88	10.68
9	V58	I80	14.97	15.74	15.86	18.03	17.45	19.25	20.01
10	V58	V81	11.30	12.28	12.15	14.43	13.89	15.80	16.81
11	V58	T83	8.51	10.64	8.71	11.60	11.29	13.60	14.32
12	V58	A94	9.02	10.07	11.00	12.54	12.82	14.66	17.10

Table B3. **Critical $\beta 4\alpha 4$ loop distances in macrostates.** Critical distances to characterize the $\beta 4\alpha 4$ loop/ $\alpha 4$ helix conformation in critical macrostates. Average distances were measured in 6 critical macrostates. Distances D90-P57 and E88-M59 (shown in bold) were used to project the free energy landscape. On the free energy landscape the critical macrostates were located based on their representative distance (shown in bold).

Index	Res1	Res2	23 (Å)	21 (Å)	12 (Å)	10 (Å)	22 (Å)	20 (Å)	16 (Å)
1	E88	M56	15.86	18.14	18.29	16.60	18.20	18.74	19.44
2	S92	V58	12.66	13.79	14.17	14.61	15.82	17.69	20.09
3	E88	P57	16.47	19.08	18.22	18.11	19.81	21.18	21.90
4	E89	P57	14.86	16.93	15.14	15.94	18.29	19.11	19.78
5	D90	P57	11.76	13.90	13.16	13.11	14.85	16.01	16.62
6	E91	P57	14.03	16.20	16.04	15.76	16.53	17.69	18.96
7	A84	V58	8.80	10.57	7.72	11.81	11.06	13.21	13.33
8	E88	V58	14.43	16.09	14.55	15.08	17.18	19.70	21.15
9	E89	V58	12.90	13.98	11.56	12.56	15.25	17.27	18.52
10	D90	V58	9.52	10.78	9.43	10.05	12.02	14.49	15.80
11	E91	V58	11.11	12.71	12.33	13.08	14.17	16.55	18.65
12	E88	M59	16.21	17.78	15.28	16.05	18.25	21.37	21.59

Table B4. **Transition path and reactive flux between macrostate 23 and 16.** The relative strength of a given path was found by dividing the reactive flux in that path by the total flux between macrostates 23 and 16.

Path rank	Path	Reactive Flux	Relative Strength (%)
1	23-12-10-20-16	9.15E-06	31.56
2	23-21-22-20-16	4.14E-06	14.28
3	23-12-10-22-20-16	4.04E-06	13.94
4	23-22-20-16	2.75E-06	9.48
5	23-4-10-8-20-16	1.88E-06	6.50
6	23-12-6-10-15-22-20-16	1.32E-06	4.54
7	23-12-6-22-20-16	1.22E-06	4.20
8	23-12-10-22-8-20-16	1.14E-06	3.94
9	23-4-10-22-8-20-16	8.15E-07	2.81
10	23-21-22-8-20-16	8.02E-07	2.77
11	23-4-10-20-16	5.49E-07	1.89
12	23-4-10-22-20-16	2.40E-07	0.83
13	23-19-23-4-10-8-20-16	2.35E-07	0.81
14	23-12-6-22-9-22-20-16	2.18E-07	0.75
15	23-12-6-0-22-20-16	1.52E-07	0.53
16	23-12-10-20-8-20-16	1.42E-07	0.49
17	23-12-6-0-22-8-20-16	1.11E-07	0.38
18	23-12-6-22-8-20-16	2.35E-08	0.08

Table B4 (cont'd)

Path rank	Path	Reactive Flux	Relative Strength (%)
19	23-22-8-20-16	2.30E-08	0.08
20	23-21-22-20-8-20-16	1.91E-08	0.07
21	23-12-10-20-22-20-16	1.69E-08	0.06
22	23-12-6-10-15-22-8-20-16	2.52E-09	0.01

Table B5. **Characterization of macrostates.** The population, number of microstates, committor (splitting) probabilities, and $\Delta Q_{\beta 3 \alpha 3}^M$, $\Delta Q_{\alpha 4}^M$ parameter values for the macrostates. Macrostates shown in bold appear in the transition path (between macrostate 23 and 16) calculated using TPT.

Macrostate	Population	Number of microstates	Committor probability	$\Delta Q_{\beta 3 \alpha 3}^M$	$\Delta Q_{\alpha 4}^M$
0	8.48E-04	24	0.2503±0.0111	+2	-2
1	2.45E-03	40	0.2203±0.0101	+4	+3
2	1.54E-03	21	0.0033±8.01E-07	-7	-10
3	1.26E-03	7	0.4965±0.0012	+3	0
4	2.35E-03	21	0.2420±0.0554	-12	-6
5	1.76E-03	10	0.0020±1.26E-05	-11	-12
6	4.07E-03	27	0.1637±0.0030	-1	-8
7	3.63E-03	14	0.4501±9.43E-07	+4	+5
8	1.04E-02	89	0.4449±0.0046	-2	-9
9	8.21E-03	34	0.3894±0.0025	+5	+1
10	1.58E-02	64	0.2328±0.0655	-3	-11
11	1.27E-02	60	0.1834±0.0015	+4	-2
12	1.08E-02	30	0.0669±0.0264	-6	-9
13	1.05E-02	8	0.5029±6.63E-05	+3	+8
14	1.30E-02	42	0.0021±1.35E-06	-9	-10
15	1.16E-02	7	0.4030±0.0006	+1	-8

Table B5 (cont'd)

Macrostate	Population	Number of microstates	Committer probability	$\Delta Q_{\beta 3 \alpha 3}^M$	$\Delta Q_{\alpha 4}^M$
16	1.14E-02	12	0.9589±0.0614	+6	+6
17	1.17E-02	30	0.5015±9.01E-16	-4	0
18	1.44E-02	129	0.0032±6.73E-05	-10	-11
19	2.42E-02	68	0.0133±0.0050	-10	-10
20	4.13E-02	176	0.4979±0.0174	2	6
21	5.03E-02	125	0.0038±0.0027	-6	-8
22	1.34E-01	276	0.4079±0.0212	-2	-2
23	6.02E-01	820	0.0024±0.0015	-12	-12

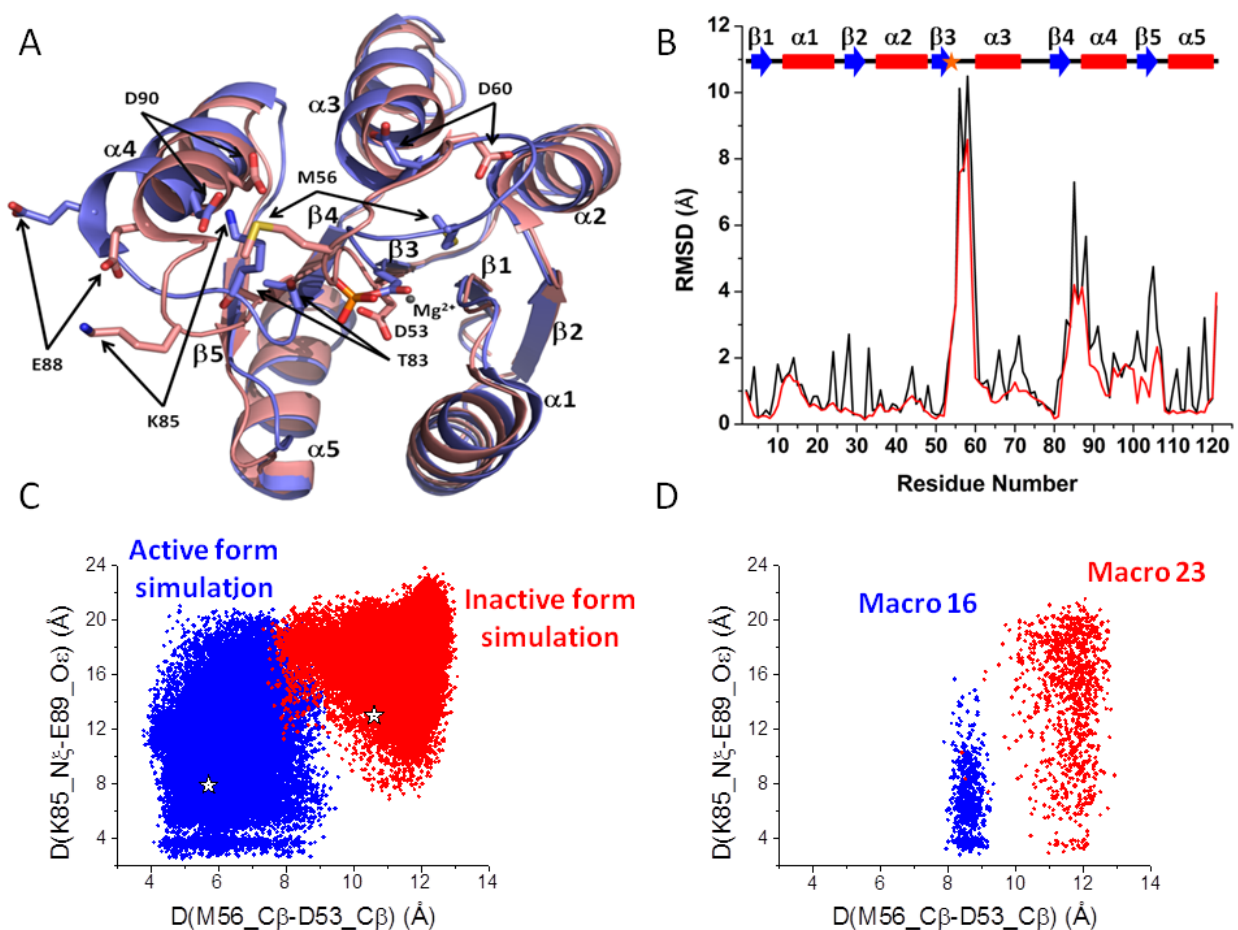


Figure B1. **Characterization of macrostates based on critical side chain orientation.** (A) Difference between the active and inactive conformations of RR468. (B) RMSD of backbone atoms (red) and all heavy atoms (black) calculated for the inactive conformation with respect to the active conformation. Orientation of the M56 and K85 side-chains are significantly different in the two forms. Some critical sidechain distances involving M56 and K85 are shown along the X and Y axes from the equilibrium simulation (C) and from the macrostate assignments of 23 and 16 (D).

REFERENCES

REFERENCES

- Banerjee, R., and Cukier, R.I. (2014). Transition paths of met-enkephalin from markov state modeling of a molecular dynamics trajectory. *J Phys Chem B* *118*, 2883-2895.
- Banerjee, R., Yan, H.G., and Cukier, R.I. (2014). Conformational Transition of Response Regulator RR468 in a Two-Component System Signal Transduction Process. *J Phys Chem B* *118*, 4727-4742.
- Beauchamp, K.A., Bowman, G.R., Lane, T.J., Maibaum, L., Haque, I.S., and Pande, V.S. (2011). MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J Chem Theory Comput* *7*, 3412-3419.
- Bellsolell, L., Prieto, J., Serrano, L., and Coll, M. (1994). Magnesium Binding to the Bacterial Chemotaxis Protein CheY Results in Large Conformational-Changes Involving Its Functional Surface. *J Mol Biol* *238*, 489-495.
- Berezhevskii, A., Hummer, G., and Szabo, A. (2009). Reactive flux and folding pathways in network models of coarse-grained protein dynamics. *J Chem Phys* *130*.
- Bourret, R.B. (2010). Receiver domain structure and function in response regulator proteins. *Curr Opin Microbiol* *13*, 142-149.
- Bowman, G.R., and Geissler, P.L. (2012). Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc Natl Acad Sci USA* *109*, 11681-11686.
- Bowman, G.R., Huang, X.H., and Pande, V.S. (2009). Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* *49*, 197-201.
- Case, D.A., Darden, T.A., III, T.E.C., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Crowley, M., Walker, R.C., Zhang, W., *et al.* (2008). AMBER 10 (University of California, San Francisco.).
- Casino, P., Rubio, V., and Marina, A. (2009). Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell* *139*, 325-336.
- Casino, P., Rubio, V., and Marina, A. (2010). The mechanism of signal transduction by two-component systems. *Curr Opin Struct Biol* *20*, 763-771.
- Deuffhard, P., and Weber, M. (2005). Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl* *398*, 161-184.
- Du, R., Pande, V.S., Grosberg, A.Y., Tanaka, T., and Shakhnovich, E.S. (1998). On the transition coordinate for protein folding. *J Chem Phys* *108*, 334-350.

E, W., and Vanden-Eijnden, E. (2006). Towards a theory of transition paths. *J Stat Phys* 123, 503-523.

Formanek, M.S., Ma, L., and Cui, Q. (2006). Reconciling the "old" and "new" views of protein allostery: A molecular simulation study of chemotaxis Y protein (CheY). *Proteins* 63, 846-867.

Gao, R., and Stock, A.M. (2009). Biological Insights from Structures of Two-Component Proteins. *Annu Rev Microbiol* 63, 133-154.

Knaggs, M.H., Salsbury, F.R., Edgell, M.H., and Fetrow, J.S. (2007). Insights into correlated motions and long-range interactions in CheY derived from molecular dynamics simulations. *Biophys J* 92, 2062-2079.

Lei, M., Velos, J., Gardino, A., Kivenson, A., Karplus, M., and Kern, D. (2009). Segmented Transition Pathway of the Signaling Protein Nitrogen Regulatory Protein C. *J Mol Biol* 392, 823-836.

Lukat, G.S., Lee, B.H., Mottonen, J.M., Stock, A.M., and Stock, J.B. (1991). Roles of the Highly Conserved Aspartate and Lysine Residues in the Response Regulator of Bacterial Chemotaxis. *J Biol Chem* 266, 8348-8354.

Ma, L., and Cui, Q. (2007). Activation mechanism of a signaling protein at atomic resolution from advanced computations. *J Am Chem Soc* 129, 10261-10268.

Metzner, P., Schutte, C., and Vanden-Eijnden, E. (2009). Transition Path Theory for Markov Jump Processes. *Multiscale Model Sim* 7, 1192-1219.

Noé, F., Schütte, C., Vanden-Eijnden, E., Reich, L., and Weikl, T.R. (2009). Constructing the Equilibrium Ensemble of Folding Pathways From Short Off-Equilibrium Simulations. *Proc Natl Acad Sci USA* 106, 19011-19016.

Pande, V.S., Beauchamp, K., and Bowman, G.R. (2010). Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* 52, 99-105.

Rhee, Y.M., and Pande, V.S. (2003). Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys J* 84, 775-786.

Roche, P., Mouawad, L., Perahia, D., Samama, J.P., and Kahn, D. (2002). Molecular dynamics of the FixJ receiver domain: movement of the beta 4-alpha 4 loop correlates with the in and out flip of Phe101. *Protein Sci* 11, 2622-2630.

Seeber, M., Felling, A., Raimondi, F., Muff, S., Friedman, R., Rao, F., Caflisch, A., and Fanelli, F. (2010). Wordom: A User-Friendly Program for the Analysis of Molecular Structures, Trajectories, and Free Energy Surfaces. *J Comput Chem* 32, 1183-1194.

Tatsis, V.A., Tsoulos, I.G., Krinas, C.S., Alexopoulos, C., and Stavrakoudis, A. (2009). Insights into the structure of the PmrD protein with molecular dynamics simulations. *Int J Biol Macromol* 44, 393-399.

Voelz, V.A., Bowman, G.R., Beauchamp, K., and Pande, V.S. (2010). Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1-39). *J Am Chem Soc* *132*, 1526-1528.

Wang, J. (2008). AmberTools Users' Manual.

Wolfe, A.J. (2010). Physiologically relevant small phosphodonors link metabolism to signal transduction. *Curr Opin Microbiol* *13*, 204-209.

Yang, S., Banavali, N.K., and Roux, B. (2009). Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories. *Proc Natl Acad Sci USA* *106*, 3776-3781.

Zhou, R.H., Berne, B.J., and Germain, R. (2001). The free energy landscape for beta hairpin folding in explicit water. *Proc Natl Acad Sci USA* *98*, 14931-14936.

CHAPTER 4

Conformational Dynamics of the ATP-lid Loop in the Histidine Kinase HK853 in Two-component System Signal Transduction

4.1. Abstract

The catalytic (CA) domain of histidine kinase (HK) binds MgATP and plays a critical role in autophosphorylation in two-component system signal transduction. The CA domain can assume at least three dominant conformations during the catalytic cycle of HK: the first is adopted by the apo form, i.e., before the binding of MgATP; the second by the MgATP-bound form, i.e., after the binding of MgATP; and the third by the ADP-bound form, i.e., after the phosphoryl transfer. The conformational transitions between the three conformations are largely unknown. A highly flexible ATP-lid loop in the CA domain is implicated in the autophosphorylation reaction in various histidine kinases. Conformational dynamics associated with the CA domain of a histidine kinase from a TCS in *Thermotoga maritima*, is investigated in this work by using a computational approach that integrates homology modeling, protein-ligand docking, molecular dynamics and targeted molecular dynamics simulations. Molecular dynamics simulations of the ligand free and ATP bound conformations of the HK853 CA domain were performed in which the ATP-lid loop adopts two distinct conformations. Critical interactions that stabilize the ATP-lid orientation are proposed based on this study. Atomistic details of the ATP-lid interactions with the bound nucleotide are proposed, and the possible role of the ATP-lid loop residues in the autophosphorylation reaction and stabilization of the protein ligand complex is suggested.

4.2. Introduction

Two-component system (TCS) signal transduction is the predominant mechanism in bacteria to sense environmental growth factors such as temperature, pH, nutrients etc (Stock et al., 2000). The TCS proteins are promising targets for future therapeutic interventions to treat bacterial infection, mainly due to two reasons. Firstly, the gene clusters contributing to processes such as cell growth and pathogenicity are often controlled by this mode of signal transduction in the microbial systems (Gotoh et al., 2010). Secondly, this mode of signal transduction is completely absent in higher organisms, including human (Wuichet et al., 2010). Integral to the TCS signaling pathways is an autophosphorylation reaction in a membrane-bound histidine kinase (HK) in response to the external stimuli followed by a phosphoryl transfer reaction from the HK to a cognate response regulator (RR) protein (Casino et al., 2010; Gao and Stock, 2009). The HK is typically the input component of the pathway with an extracellular sensor domain, designed to sense extracellular stimuli (Gao and Stock, 2009). In a prototypical HK, the cytoplasmic portion consists of two distinct domains: a dimerization and histidine-containing phosphotransfer domain (DHp) and a C-terminal catalytic and ATP binding (CA) domain (Gao and Stock, 2009).

The extracellular region of HK recognizes the external stimulus in TCS signal transduction (Gao and Stock, 2009). A conserved histidine residue that usually belongs to the DHp domain in the intracellular region of the HK gets autophosphorylated in response to the stimulus, by the CA domain of the HK (Gao and Stock, 2009). Histidine autophosphorylation is the first step in signal propagation through a TCS, after signal recognition by the extracellular sensory domain(s) of the HK (Gao and Stock, 2009; Mascher et al., 2006). The CA domain exists as monomer and it hosts one ATP molecule situated between a flexible loop i.e. ATP-lid loop, and a central helix (Gao and Stock, 2009). A highly conserved ATP binding cavity in the CA domain is defined by a few

characteristic sequence motifs; namely, the N, G1, F, and G2 boxes (Gao and Stock, 2009). A divalent cation is present in the ATP-binding pocket of several HKs that seems to neutralize the negatively charged phosphate oxygen atoms in the ATP. It also binds to conserved asparagines in the central helix, the N-box motif (Bilwes et al., 2001; Marina et al., 2001). The ATP-lid loop is a long flexible loop that connects a short helix F-box and a G2 motif in the CA domain, which serves as a lid, as the name suggests, to the ATP-binding pocket (Gao and Stock, 2009). The γ -phosphate of the ATP is attacked by the conserved His of the DHp domain in the autophosphorylation reaction, resulting in a phosphorylated histidine in the DHp domain, and leaving ADP in the ATP-binding pocket of the CA domain (Casino et al., 2010). Despite being a very common sensor kinase domain, the atomistic details of the autophosphorylation reaction in HK remain poorly understood (Casino et al., 2014). The highly dynamic nature of the CA domains, especially the ATP-lid loop region, makes it difficult to elucidate a functional role of the ATP-lid residues in nucleotide binding and catalysis. In a previous work, a conformational change in the ATP-lid loop and the bound ATP molecule was studied using molecular dynamics simulation starting with a modeled ATP-lid conformation of the CheA HK (Zhang et al., 2005). It was proposed that conformational switching from an extended to folded conformation, or vice versa, in the ATP molecule triggers the conformational change in the ATP-lid loop.

In this work, the histidine kinase, HK853, in the Two-component system (Casino et al., 2007) from *Thermotoga maritima*, with HK853 and RR468 as the cognate pair of HK and RR, respectively, is investigated to elucidate the conformational dynamics associated with the CA domain, with a focus on the ATP-lid loop. Structural information on the CA domain of HK853 is found in two crystal structures available for the cytoplasmic portion of HK853 (Casino et al., 2009; Marina et al., 2005). The first structure (Marina et al., 2005) reported was the crystal

structure of the entire cytoplasmic domains of HK853 in the presence of ADP (PDB id 2C2A). The HK853 was crystallized in the presence of the putative non-hydrolysable ATP analog AMPPNP and MgCl_2 (Marina et al., 2005). But, it was found that the AMPPNP nucleotide got hydrolyzed in the crystallization buffer. As a result, no electron density was observed for the γ -phosphate of the ATP analogue and the Mg^{2+} ion (Marina et al., 2005). Nine residues in the ATP-lid loop of the HK853 were unresolved in this study, and this phenomenon was attributed to the higher mobility of the ATP-lid loop, due to lack of interaction with the ATP analogue (Marina et al., 2005). In contrast, the ATP analogue was stable under similar conditions for crystallization of histidine kinases, PhoQ (Marina et al., 2001) and CheA (Bilwes et al., 2001), where the ATP-lid loop was well defined due to direct interactions with the γ -phosphate of the ATP analogue. The second structure (Casino et al., 2009) reported is a complex of HK853 with its cognate response regulator protein RR468 (PDB id 3DGE). The complex was prepared in presence of Mg^{2+} and the ATP analogue AMPPNP (Casino et al., 2009). However, in this structure as well, the bound ligand in the active site of the CA domain is ADP, due to hydrolysis of the AMPPNP.

As found in many other TCSs (Stock et al., 2000), dephosphorylation of the phosphorylated RR468 is catalyzed in presence of the HK853 (Casino et al., 2009). Structural evidence suggests that the RR468 conformation in the complex represents the phosphorylated form of RR468 (Casino et al., 2009). Based on this fact it is believed that the structure in the complex form represents the product complex subsequent to the phosphoryl transfer reaction from HK853 to RR468 (Casino et al., 2009). One of the major differences between the two crystal structures of HK853 is in the ATP-lid loop. The ATP-lid loop of HK853 is well defined in the presence of RR468 and the loop folds back on the CA domain, completely covering the adenosine base and

the pentose sugar moiety of the bound nucleotide, ADP. In the other structure, residues 433–441 of the ATP lid loop are highly disordered in the absence of the γ -phosphate and the Mg^{2+} , resulting in the lack of electron density for the residues in the range 433–441 (Marina et al., 2005). However, the part of the ATP-lid loop that is resolved through the X-ray diffraction (E426-D432 in the N-terminus and T442-G445 in the C-terminus of the ATP-lid loop) clearly indicate that the loop orientation in these two structures is very different (Casino et al., 2009).

The ATP-lid is relatively ordered and covers the ATP binding site if the nucleotide is present, whereas it is highly mobile in nucleotide-free structures of CheA (Bilwes et al., 1999) and NtrB (Song et al., 2004). The conformational changes of the ATP-lid loop are proposed to couple with the ATP binding to alter the interactions between the DHp and the CA domains (Gao and Stock, 2009). Both existing crystal structures present the HK853 in the product form after the phosphoryl transfer reaction i.e. after dephosphorylation of the phosphorylated histidine. Enhanced mobility of the ATP-lid loop in the absence of the γ -phosphate of the ATP and the Mg^{2+} ion suggest some role of the ATP γ -phosphate or the Mg^{2+} ion or both for the stabilization of the ATP-lid loop prior to the autophosphorylation reaction. Unlike HK853, the ATP analogue does not undergo hydrolysis in the PhoQ (Marina et al., 2001) and CheA (Bilwes et al., 2001) histidine kinases and the ATP-lid loop has several polar interactions with the γ -phosphate of the ATP analogue in those HKs. From existing crystal structures of HK853 the role of the ATP-lid loop in autophosphorylation is not clear due to lack of mutational studies focusing on the residues in the ATP-lid loop. Existing mutational studies suggest a critical role of an arginine residue in the ATP-lid loop on the autophosphorylation activity in the related histidine kinase, PhoQ (Marina et al., 2001). The ATP-lid conformation is implicated in HK autophosphorylation reaction in two related HKs, PhoQ (Marina et al., 2001) and CheA (Bilwes et al., 2001), for its

ability to make favorable contact with the DHP domain. Molecular dynamics simulation with the CA domain of CheA suggested that a conformational switch of the bound ATP between a folded and extended conformation triggers the ATP-lid opening (Zhang et al., 2005). In this work, we have used the structure of the CA domain of HK853 to study conformational dynamics associated with the ATP-lid loop, in the presence or absence of the nucleotide and the Mg^{2+} ion, where the entire ATP-lid was modeled using the PhoQ (Marina et al., 2001) conformation. The ATP-lid conformation modeled using PhoQ depicts another closed conformation that is different from that observed in the HK853-RR468 complex crystal structure. There is no structural evidence for the open ATP-lid conformation in HK853. Simulation starting from the modeled ATP-lid loop results in a conformation where the ATP-lid loop forms an angle $\sim 70^\circ$ compared to that less than 40° in the closed conformation from the HK853-RR468 complex with respect to a centrally located stable β -strand. In presence of a bound nucleotide the loop open-close movement was observed for the ATP-lid loop within 100 ns simulation time. However in absence of the bound nucleotide the ATP-lid remains open. The other closed conformation of the ATP-lid loop as observed in the HK853-RR468 complex is influenced by the presence of the RR468 and cannot undergo complete transformation in absence of a bound nucleotide.

4.3. Methods

4.3.1. Structural models for ATP-Mg²⁺ bound form of the HK853 CA domain. The CA domain of HK853 in presence of MgATP was simulated starting from two models of the protein-ligand complexes, where the ATP-lid loop adopts distinct orientations. The first model was based on the crystal structure of HK853 in the absence of RR468 (Marina et al., 2005). The missing part of the ATP-lid loop was modeled from a CA domain crystal structure of a related histidine kinase, PhoQ (Marina et al., 2001) using homology modeling. The second model was based on the CA domain conformation as observed in the HK853-RR468 complex structure (Casino et al., 2009) also with ADP in the ATP-binding cleft of the CA domain, where the ATP-lid is well defined. Here, the ATP-Mg²⁺ was docked from the PhoQ (Marina et al., 2001) crystal structure, after structural alignment of the protein backbone of the HK853 CA domain against that in PhoQ. Acronyms of all the simulations along with the details of their starting conformations are provided in the Table 4.1.

4.3.2. Equilibrium MD simulations. All simulations were performed using Amber 10. The ff99SB force field was used for the simulations. Water molecules in the crystal structure were removed, hydrogen atoms were added, and the systems were neutralized by addition of an appropriate number of Na⁺ ions. Each system was solvated with a ≥ 10 Å layer of TIP3P waters in a rectangular water box. The system was minimized first using a combination of steepest descent and conjugated gradient methods for 8,000 steps. A weak positional restraint (2 kcal/mol/Å²) was applied on all protein atoms during this minimization step. The whole system was minimized again using a combination of steepest descent and conjugated gradient methods for 20,000 steps without any positional restraint on any atoms. A time step of 2 fs was used for all subsequent heating, equilibration and production runs with the SHAKE option on all bonds

containing H-atoms. Langevin dynamics was used for temperature control. The minimized system was heated from 0 to 300 K in 1 ns. Weak positional restraint ($2 \text{ kcal/mol/\AA}^2$) on all protein atoms were applied during the heating cycle. Constant volume equilibration was performed at 300 K for 5 ns without any positional restraints. Finally, a trajectory of 100 ns was generated using constant pressure simulation. Force field parameters and charges for ATP were obtained from Bryce et al. (<http://www.pharmacy.manchester.ac.uk/bryce/amber/>). Another set of simulations was performed using the same two starting conformation without MgATP (APO2C2A and APO3DGE) to understand the conformational dynamics of the free form of the CA domain.

All analyses of the MD trajectories were performed using the Ambertools 10 (Cheatham, 2008; Roe and Cheatham, 2013) ptraj module and Wordom (Seeber et al., 2010). All individual frames were aligned for RMSD and RMSF calculations with the C α atoms of the protein, excluding the ATP-lid loop and N- and C-terminal residues, which are very mobile during simulation. A distance cutoff of 3.5 Å between the heavy atoms and angular cutoff of 120° was used for hydrogen bond analysis in ptraj.

Table 4.1. **Summary of structural models.** Structural model considered as the starting conformations for MD simulations along with the acronyms used in the manuscript to refer to the corresponding MD trajectory.

Acronym	Protein structural model		Ligand interaction based on		Orientation Angle
	CA853 except ATP-lid	ATP-lid loop	Method	Ligand	
COM2C2A	X-ray (2C2A)	Homology modeling [#]	Docking [#]	MgATP	54.1
APO2C2A	X-ray (2C2A)	Homology modeling [#]	-	-	54.1
COM3DGE	X-ray (3DGE)	X-ray (3DGE)	Docking [#]	MgATP	29.9
APO3DGE	X-ray (3DGE)	X-ray (3DGE)	-	-	29.9

[#]Homology model of ATP-lid loop is based on the PhoQ crystal structure (PDB id 1ID0). The same PhoQ crystal structure was used as a template to dock the MgATP complex.

4.3.3. Targeted MD simulations. Targeted MD (TMD) simulations were performed to simulate transition paths connecting two distinct conformations of the ATP-lid in the absence of ATP and Mg^{2+} . In this method, a harmonic restraint in the mass-weighted RMSD between the initial and target structures is introduced that can, over a suitably slow transformation, provide a reasonable set of intermediate conformations between the initial and target structures (Case et al., 2008). All $\text{C}\alpha$ atoms were best fit on the target reference structure and the RMSD was calculated for all heavy atoms in the protein. A force constant of $8 \text{ kcal/mol/\AA}^2$ was used for the TMD. The simulation length was 5 ns for TMD trajectories in either direction.

4.3.4. Pairwise interaction analysis. A Protein Structure Network (PSN) analysis (Brinda and Vishveshwara, 2005) was done on each MD and TMD trajectory using Wordom (Seeber et al., 2010). Pairwise interactions were compared in different simulations to identify characteristic interactions in the conformational state associated with the simulation. The interaction strength (I_{ij}) between residue pair i and j , is defined as

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} 100 \quad (1)$$

where n_{ij} is the number of side-chain atom pairs within a given distance cutoff ($4.5\text{\AA}$ as a default), N_i and N_j are, respectively, the normalization factors for residues i and j , that take into account the differences in size of the different residues as bigger residues are likely to make more contact pairs, n_{ij} .

4.4. Results

4.4.1. ATP binding models of the CA domain of HK853. We have built two structural models of the CA domain of the HK853 in the presence of ATP and Mg^{2+} . The first model, COM2C2A was based on the crystal structure of HK853 (Marina et al., 2005) by Marina et al. (PDB id 2C2A). The ATP-lid in the model was built using the CA domain structure of the PhoQ histidine kinase (1ID0) using homology modeling. The second model, COM3DGE was built based on the HK853-RR468 complex structure (Casino et al., 2009) (PDB id 3DGE). The DHp domain was removed in both cases to obtain only the CA domain of HK853 for the simulations. Docking of the ATP molecule and Mg^{2+} ion were performed based on the PhoQ crystal structure in both cases. Interactions involving the adenine base and the pentose sugar moieties of the ATP molecule are consistent with the existing ADP-bound crystal structures of HK853 (Casino et al., 2009; Marina et al., 2005). The ATP and Mg^{2+} interactions involve structural motifs, N, G1, F, G2 (Figure 4.1) in our structural models, which are consistent with crystal structures of related HKs as well in presence of ATP analogue and metal ion (Bilwes et al., 2001; Marina et al., 2001). In both ATP-bound structural models, the adenosine ring of the ATP interacts with a shallow hydrophobic pocket comprised of L377, I414, I416, I424, L446, and F472 in one face, where residues I424 and I416 belong to the G1 motif of the HK853 CA domain. On the other face, the adenosine ring is isolated from the bulk solvent due to π -stacking interactions with Y384 from the N-box motif. The Y384 also forms a hydrogen bond with the ATP O_γ . Another N-box residue, K383, interacts with the ATP γ -phosphate, which helps to neutralize the negative charge associated with the phosphate end of the ATP molecule. Two asparagines from the N-box are known to chelate with the metal ion in the related crystal structures of PhoQ and CheA (Bilwes et al., 2001; Marina et al., 2001). The residues, N376 and N380 are at the equivalent

positions in the N-box of HK853. The asparagines were brought close to the metal ion using a brief minimization cycle with distance restraints applied between O δ atoms of N376 and N380, and the docked Mg²⁺ ion. The Mg²⁺ ion interacts with the γ , β , and α -phosphate oxygen atoms, and one position within the coordination sphere of Mg²⁺ is occupied by a water molecule. The residue, S470, forms a hydrogen bond with the N1 atom in the adenine base in both structural models. The differences in the protein-ligand interaction in the two structural models that exist involve only the ATP-lid loop and neighboring G2 motif. In the COM3DGE starting structure, the T442 from the G2 motif is shown to have an interaction with the ATP γ -phosphate, whereas in the COM2C2A model it is R430 from the ATP-lid. Structural motifs and residues in the HK853 CA domain that play a crucial role in ATP binding with a few critical residues are illustrated in Figure 4.1.

The fates of the hydrogen bond and metal chelating interactions involving ATP and Mg²⁺ during MD simulation, starting from the COM2C2A and COM3DGE models, are summarized in Table 4.2. Interactions involving the Mg²⁺ ion are very stable in both COM2C2A and COM3DGE simulations. Common polar interactions are more or less stable in these simulations. The hydrogen bond between the ATP γ -phosphate and R430 as observed in the COM2C2A simulation, is more stable than that involving T442 as observed in the COM3DGE simulation.

Table 4.2. **Polar and ionic interactions of ATP with Mg²⁺ and CA domain residues in COM2C2A and COM3DGE simulations.**

Atom	COM2C2A		COM3DGE	
	Res/Atom	Occur (%)	Res/Atom	Occur (%)
O3G	K383 NZ	97.75	K383 NZ	91.35
O2G	Mg ²⁺	100.0	Mg ²⁺	100.0
O1G	Y384 OH	94.81	Y384 OH	83.23
O2B	Mg ²⁺	100.0	Mg ²⁺	100.0
O1B	R430 NH2	74.74	T442 OG1	43.84
O2A	Mg ²⁺	100.0	Mg ²⁺	100.0
	N380 ND2	90.49	N380 ND2	90.88
	N376 ND2	6.16	N376 ND2	51.93
O1A	N376 ND2	83.86	N376 ND2	80.10
N7	N380 ND2	37.82	N380 ND2	77.66
N6	S470 OG	19.69	S470 OG	49.11
	D411 OD1 OD2	12.62	-	-
N1	S470 OG	73.38	S470 OG	17.62

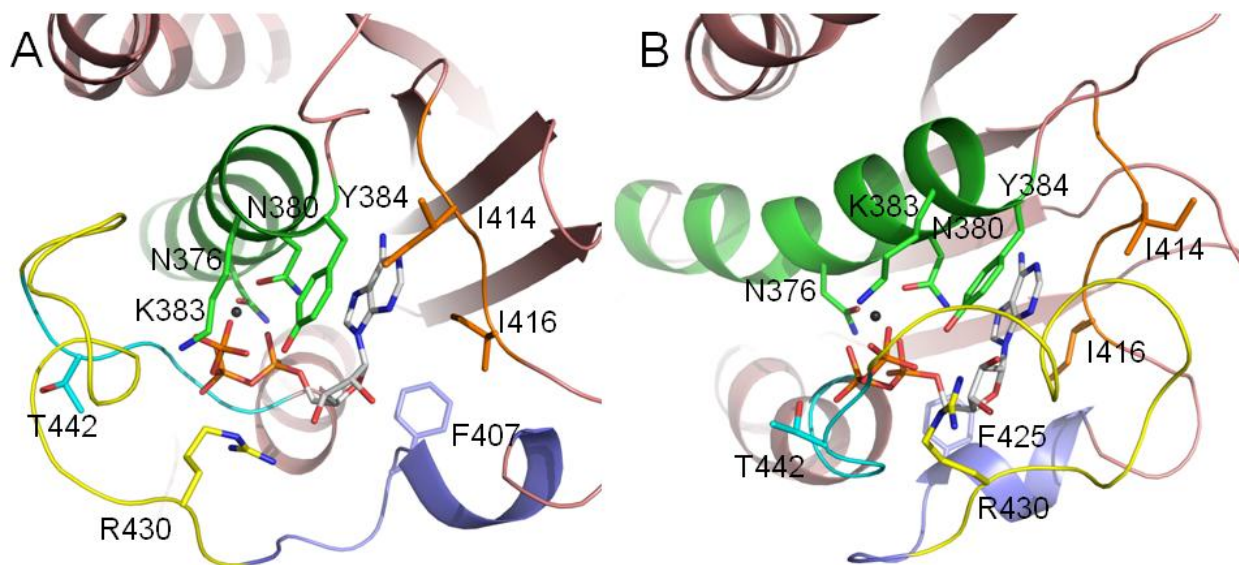


Figure 4.1. **Binding models of MgATP in the CA domain of HK853.** The ATP molecule is shown with the carbon atoms colored in gray and the Mg^{2+} ion as a black sphere. The residues in the structural motifs, N, G1, F, and G2 structural motifs are shown in green, orange, blue and cyan, respectively. The ATP-lid loop is shown in yellow, which has distinct conformation in (A) the model built using PhoQ catalytic domain as template and (B) the crystal structure of the HK853 in complex with the RR468. The Mg^{2+} ion is coordinated to triphosphate of the ATP, and the side chains of N376 and N380 from the N-motif helix. The adenine base binds to a shallow hydrophobic pocket consisting of I414, I416 (G1-motif), and several other residues. Y384 is stacked with the adenine base, and F407 (F-motif) is the gatekeeper residue in the other entry point for the adenine binding pocket. K383 from the N-motif forms a salt bridge with the γ -phosphate of ATP. Interactions of R430 (ATP-lid) and T442 differentiate the two structural models. In the model built using the PhoQ crystal structure as the template, R430 forms a salt bridge with the γ -phosphate of the ATP molecule. This interaction mimics the interaction of R434 in the PhoQ crystal structure in presence of an ATP analogue (PDB id 1ID0).

Figure 4.1 (cont'd)

In the second model, derived from the HK853-RR468 crystal structure by docking of ATP and Mg^{2+} , this interaction is not there. The T442 (G2-motif) forms a hydrogen bond with the γO of the ATP molecule in the second model.

4.4.2. The ATP-lid is the most flexible region of the CA domain. Conformations of the CA-domain were stable during simulation irrespective of the ATP-lid conformation or presence of the ligand (Figure 4.2). Fluctuations are observed mainly in the ATP-lid loop during all equilibrium MD simulations. Moreover, the fluctuations of the ATP-lid loop are similar in the COM2C2A and COM3DGE simulations (Figure 4.2 C). However, the ATP-lid loop fluctuates more in the absence of ATP and Mg^{2+} if the ATP-lid conformation is the one observed in the COM3DGE model (Figure 4.2 D). The residues in a small stretch, 386-395 fluctuate more in presence of ATP and Mg^{2+} relative to the corresponding apo conformation simulation. This region is a loop following the N-box motif.

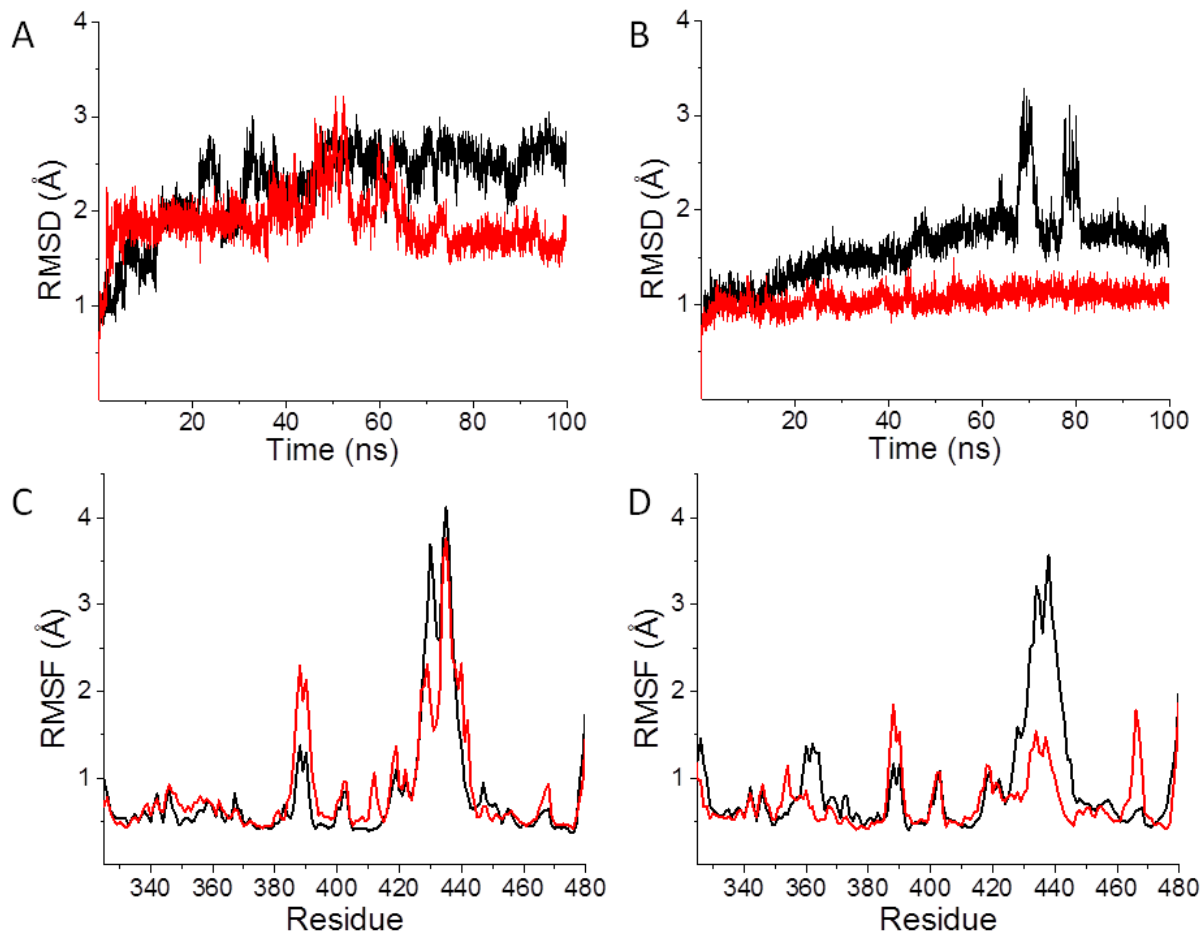


Figure 4.2. **RMSDs and RMSFs of the MD simulations of HK853 CA domain.** A, RMSDs of APO2C2A (black) and COM2C2A (red); B, RMSDs of APO3DGE (black) and COM3DGE (red); C, RMSFs of APO2C2A (black) and COM2C2A (red); D, RMSFs of APO3DGE (black) and COM3DGE (red).

4.4.3. Orientation angle: the angle K468-I476-S434 provides a reaction coordinate for the ATP-lid orientation. The ATP-lid orientation with respect to the rest of the protein is the only major difference between the two starting models, COM2C2A and COM3DGE. The S434 is roughly the middle residue in the ATP-lid loop. The angle formed by the C α atoms of K468, I476 and S434 provides an excellent reaction coordinate for understanding the ATP-lid loop orientation during the simulations. The line formed by the C α atoms of K468 and I476 is fixed, as these two residues are among the residues in the protein with minimal fluctuation during simulations starting with either structural models, with or without ATP and Mg²⁺. The other line of the angle is formed by the C α atoms of K468 and S434 and its orientation changes as the ATP-lid moves relative to the rest of the protein (Figure 4.3). The orientation angles in the starting structures of the COM2C2A (or APO2C2A) and COM3DGE (or APO3DGE) simulations are 54.1° and 29.9°, respectively. The value of this angle varies roughly between 40-80° in the COM2C2A or APO2C2A simulations whereas in COM3DGE or APO3DGE simulations the angle is less than 40° and stays close to the initial value, 29.9° during the simulation (Figure 4.4).

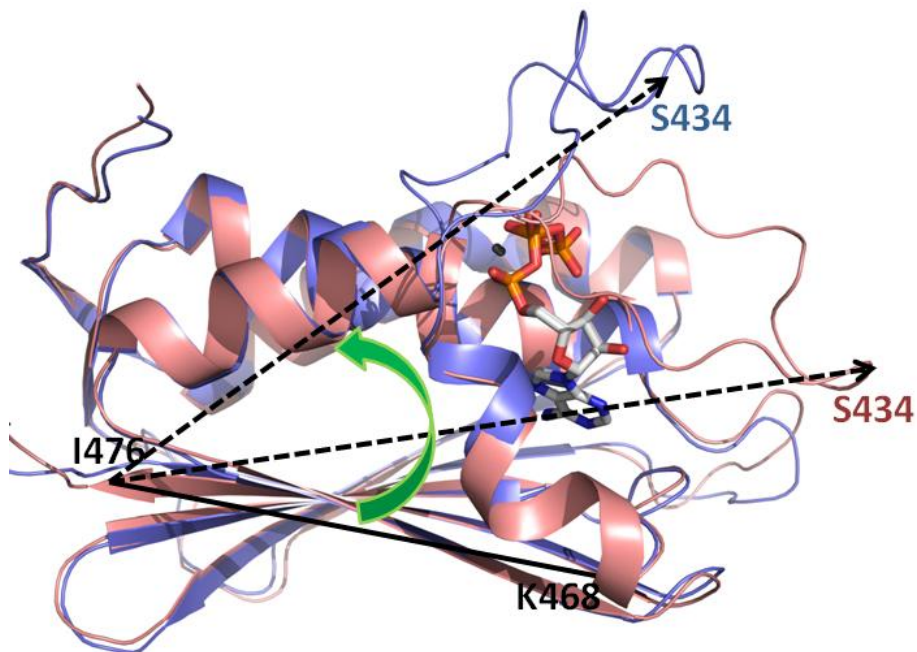


Figure 4.3. **Definition of orientation angle.** The orientations with respect to the rest of the protein are defined by the angle formed by C α atoms of K468, I476, and S434. The orientation angle in the starting structures of the COM2C2A (blue) and COM3DGE (red) simulations are 54.1 and 29.9 respectively. The C-terminus residues, K468 and I476, have minimal fluctuation as evident from the RMS fluctuation of all four simulation trajectories, where S434 is roughly the middle residue in the ATP-lid.

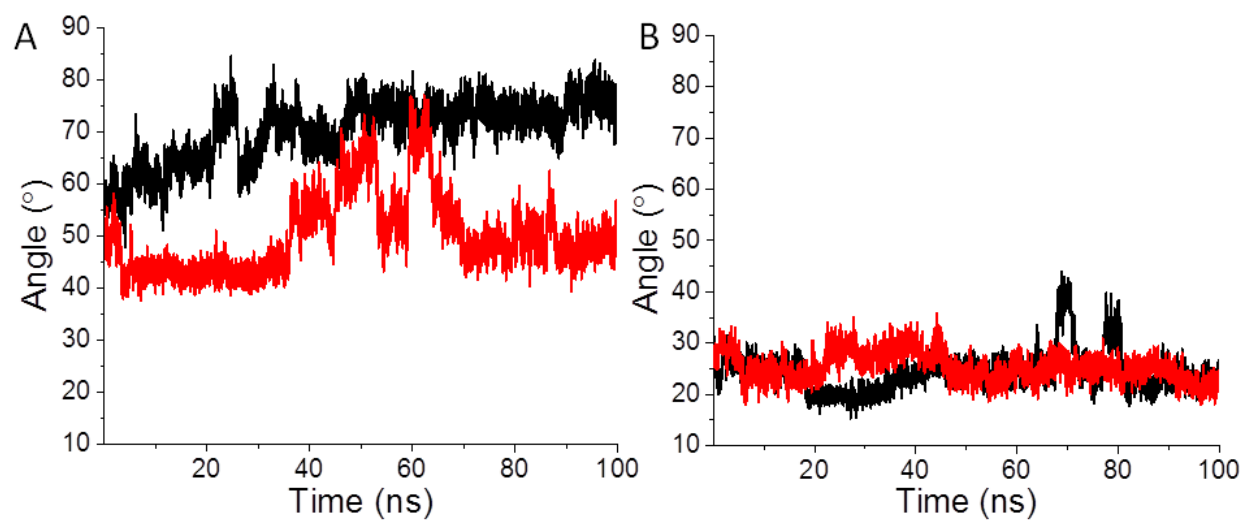


Figure 4.4. **Change in the ATP-lid orientation angle in the MD simulations.** A, APO2C2A (black) and COM2C2A (red); B, APO3DGE (black) and COM3DGE (red).

4.4.4. ATP-lid dihedral angles show a characteristic difference between the two starting structures. Backbone torsion angles of the ATP-lid and the neighboring residues were investigated to identify the dihedral angles that sample different conformational states in the COM2C2A and COM3DGE simulations. Backbone dihedral Φ and Ψ angles were measured for all the residues in the ATP-lid and neighboring residues from the F-box and G2 motifs. There are only eight residues in the ATP-lid loop and neighboring regions that sample distinct backbone torsion angles in the COM2C2A and COM3DGE simulations. There is almost no overlap in the Ramachandran plot for these residues F428, S433, S434, L435, Y437, Q438, G443 and G445 where F428 is the last residue in the F-box and G443 and G445 belongs to G2-box. It is postulated that the difference in backbone torsion angles of some or all of these eight residues are the reason for the observed difference in the orientation angle (Figure 4.5) as shown in the COM2C2A or COM3DGE simulations. Unlike in COM2C2A, the residues S434, L435, T436, Y437, and E438 form a 3_{10} helix during COM3DGE simulation.

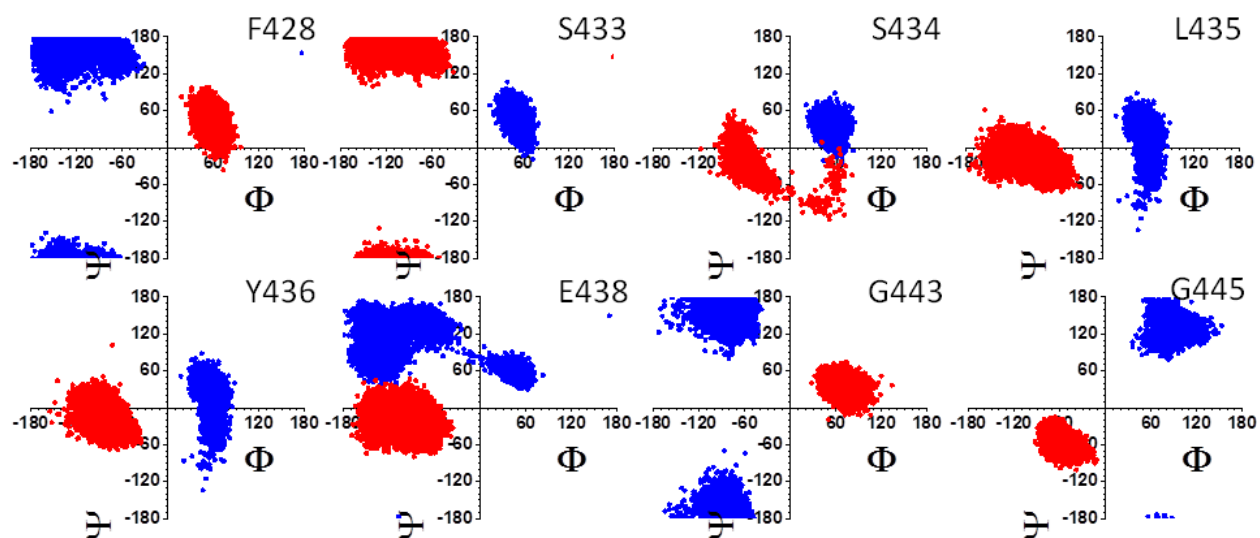


Figure 4.5. **Ramachandran plot analysis of ATP-lid loop residues.** Ramachandran plots for residues F428, S433, S434, L435, Y436, E438, G443, and G445. Backbone torsion angles of eight residues from the ATP-lid loop and the neighboring regions (F and G2 motifs) show distinctive values in the COM2C2A (blue) and COM3DGE (red) simulations. F428 is the last residue of the F-box and G443 and G445 belongs to the G2-box motif.

4.4.5. Conformational transition connects two distinct states of the ATP-lid loop in the absence of ATP and Mg^{2+} . The ATP-lid conformation remains similar to the respective starting structures in the apo protein simulations (APO2C2A and APO3DGE). The ATP-lid loop has similar fluctuations in the COM2C2A and APO2C2A simulations as evident from the RMSF plot (Figure 4.2 C). The ATP-lid orientation angles observed in the two simulations are in the same range, 50-80° in the APO2C2A simulation and 40-75° in the COM2C2A simulation (Figure 4.4 A). The ATP-lid loop fluctuates more in the APO3DGE simulation compared to that in the COM3DGE simulation (Figure 4.2 D). The orientation angle oscillates around 30°, as observed in the crystal structure of the HK853-RR468 complex, in the presence or absence of MgATP (Figure 4.4 B). For a short span of time around 70 ns and 80 ns in the APO3DGE simulation, the orientation angle jumps to 40°. However, in terms of the orientation angle, we did not observe the ATP-lid orientation angles in either APO2C2A or COM2C2A simulation reaching around 30°. We therefore performed Targeted MD simulations to obtain possible transition paths between the two distinct conformational states of the ATP-lid in the absence of ATP and Mg^{2+} . The TMD simulations were started from the equilibrated conformation of either the APO2C2A or the APO3DGE structural model and the conformation of the other model was used as the target conformation. The ATP-lid conformation starting with one model converged to the other during 5 ns TMD simulation as observed from the orientation angle plotted against time (Figure 4.6). Changes in interactions between residue pairs in the TMD trajectories were compared with those in the equilibrium simulations to understand the basis of conformational plasticity associated with the two apo form starting models.

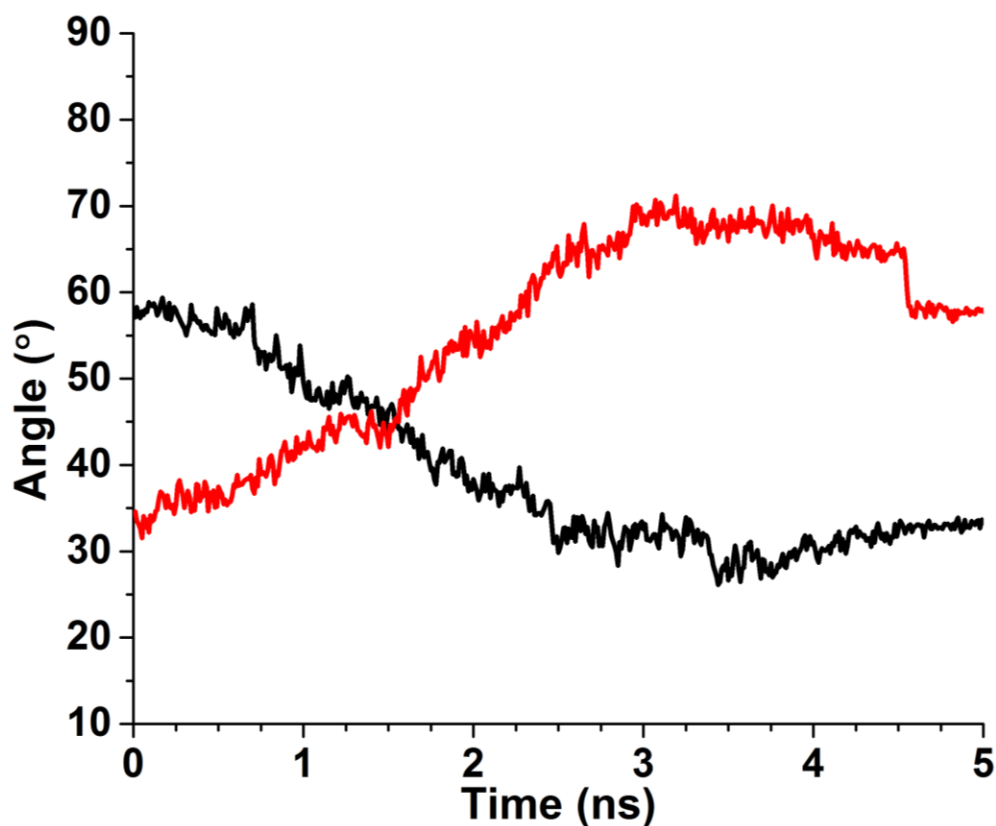


Figure 4.6. **Change in the ATP-lid orientation angle along the TMD simulations.** The first TMD simulation was started with the equilibrated conformation of the APO2C2A model and the equilibrated conformation of the APO3DGE model was used as the target conformation (in black). A possible transition path in the reverse direction was obtained by using equilibrated conformations of APO3DGE and APO2C2A as the starting and target conformations, respectively, in the other TMD simulation.

4.5. Discussion

4.5.1. The ATP-lid orientation in HK853 models compared with that in the CheA and

PhoQ histidine kinases. Structural alignment of the starting models of the HK853 with the kinase domain in PhoQ and CheA suggests that all the secondary structural elements align very well leaving only the ATP-lid loop region significantly different (Figure 4.7). The CheA CA domain has three extra helices, one after the N-box and two between the G1 and F-boxes connected by a loop (Bilwes et al., 2001). Also, part of the ATP-lid loop in CheA forms a helix as shown in the crystal structure in presence of the ATP analogue (PDB id 1I58) (Bilwes et al., 2001). The ATP-lid conformation of various CA domains is disordered in the absence of ATP or any ATP analogue and cannot be determined in the crystallographic experiments (Bilwes et al., 2001; Casino et al., 2009; Marina et al., 2001; Marina et al., 2005). Also, the transient nature of the ATP (or ATP analogues) and HKs (Dago et al., 2012) interactions have hampered the crystallization of a HK in the midst of the autophosphorylation reaction (Casino et al., 2014). The ATP-lid loop is known to be the most flexible region in the CA domain for other HKs (Bilwes et al., 2001; Marina et al., 2001) as well, even in the presence of ATP analogues, where the B-factor calculated in the crystallographic experiment for the ATP-lid loop is higher compared to the rest of the protein (Bilwes et al., 2001; Marina et al., 2001). The loop is anchored by the interaction between a conserved phenylalanine residue (F428 in HK853) from the F-box and a hydrophobic residue (L446 in HK853) next to the G2-box at the N and C-terminus respectively. In both HK853 structural models, F428 is docked into a hydrophobic cluster consisting of the hydrophobic residue (L446 in HK853) next to the G2-box among others and this feature is consistent with other HKs (Bilwes et al., 2001; Marina et al., 2001) as well. So it is obvious that the conformational differences between the two starting models are centered

within the range F428-L446 during simulations based on the two starting structures, and there is minimal difference in the backbone conformations beyond this range as observed from the dihedral analysis. All the eight residues in the ATP-lid and neighboring regions that sample different conformations during the simulations, starting from two models of the HK853, are within the range F428-L446. In other CA domain structures as well, the hydrophobic interactions between two ends of the ATP-lid loop make the rest of the protein fold unaffected due to any movement in the ATP-lid loop (Bilwes et al., 2001; Marina et al., 2001). As suggested in the literature, the ATP-lid forms a closed conformation in PhoQ, where the ATP-lid partly covers the α , β , γ -phosphate region of the ATP. However, in the CheA structure, the loop orientation completely exposes the adenosine binding site (Marina et al., 2001). Hence our first model (COM2C2A or APO2C2A) being built based on the ATP-lid loop conformation in the PhoQ crystal structure, represents a closed conformation. The orientation angle in the starting structure of the COM2C2A or APO2C2A model is 54.1°. In the absence of ATP the orientation angle stabilizes around 70° during simulation of the APO2C2A form, which signifies a conformation that is more open (Figure 4.4 A). In the presence of MgATP the orientation angle is stabilized around 50°, which is similar to the PhoQ (closed conformation). However the orientation angle can jump to an open state similar to the apo form, i.e., adopt a value ~70° (Figure 4.4 C).

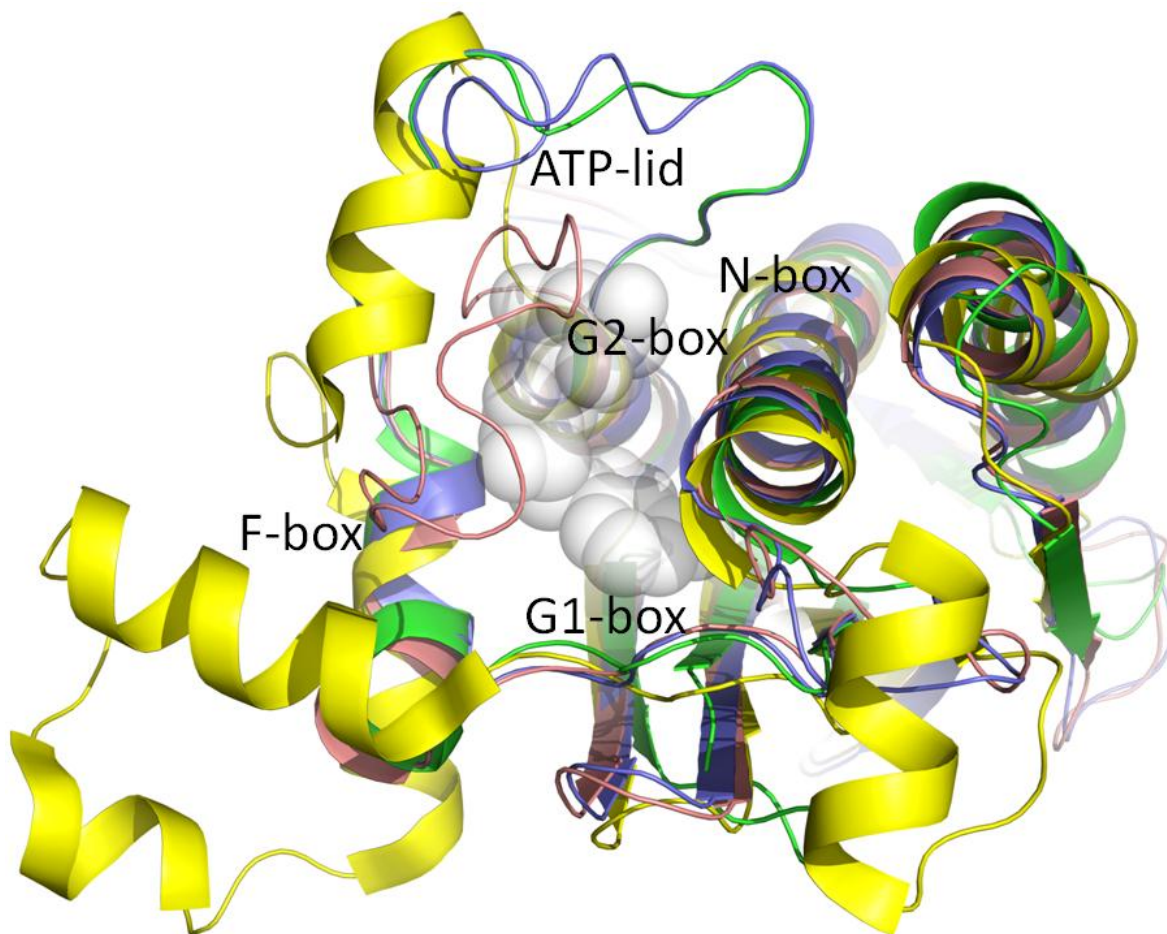


Figure 4.7. **Structural comparison of CA domain of HK853, PhoQ and CheA.** Structural alignment of two conformations of HK853, COM2C2A (blue) and COM3DGE (pink), with the catalytic domain of PhoQ (green) and CheA (yellow). The bound ATP molecule is shown with atoms in gray space filling spheres. The ATP-lid loop is defined by the connecting residues of the structural elements, F and G2-box. The ATP-lid loop in the COM2C2A model, being based on that in the PhoQ crystal structure, is similar except in the range S434-V439 where there are four insertions in the HK853 sequence compared to that of PhoQ. Besides the three additional helices (including a very short one) in CheA, ATP-lid forms a helix in CheA unlike that in the HK853 or PhoQ.

We have compared the orientation angle in the simulation trajectories with the NMR structural ensemble obtained from the solution state NMR experiments (unpublished data). The NMR structural refinement of the CA domain was done in the presence of ATP and an Mg^{2+} ion, docked into the active site, using the data obtained from the solution state NMR experiments. Residual Dipolar Coupling (RDC) data for the ATP-lid were used for the refinement of the loop conformation along with intramolecular Nuclear Overhauser Effect (NOE) and J-coupling data. The orientation angles in the 20 structures of the NMR ensemble vary between 62.0-83°, very similar to those observed in the APO2C2A and COM2C2A simulations (Figure 4.4 A, C).

4.5.2. Interaction between the ATP-lid and MgATP in HK853, CheA and PhoQ histidine kinases. Interactions between the bound MgATP and the conserved motifs in the CA domains of PhoQ and CheA can explain the basis of reduced mobility associated the ATP-lid loop in the presence of a bound nucleotide (Figure 4.8 A, B). On the other hand, the sequence comparison of the N, G1, F, ATP-lid and the G2 motifs reveal inherent conformational plasticity associated with the nucleotide binding site irrespective of a bound nucleotide (Figure 4.8 C). Mobility in the ATP-lid is facilitated by the presence of three conserved glycine residues (G441, G443, and G445 in case of HK853) of the G2 box. Sequence alignment of CA domains of HK853 with PhoQ and CheA suggests that the ATP-lid loop is the longest in HK853 (Figure 4.8 C). Nucleotide binding in PhoQ may arrest the mobility of the ATP-lid loop in presence of a nucleotide due to two reasons. Firstly, the ATP-lid in PhoQ is shorter by three residues compared to that of the HK853. That explains inherent flexibility associated with the ATP-lid loop in HK853 compared to that in PhoQ. Moreover, three loop residues in PhoQ, R434, R439, and Q442, make extensive contacts with the phosphate groups of the ATP and the Mg^{2+} ion (Figure 4.8 A). In CheA, the ATP-lid residues in the range S492-S501 form a short helix, and two

residues (K494 and S498) from the helix form strong polar interaction with the ATP γ -phosphate. In the absence of the γ -phosphate, the region E495-R503 in CheA becomes disordered as observed in the CheA structure in the presence of ADP (PDB id 1I5C) (Bilwes et al., 2001). It is clear that the ATP-lid loop in HK853 is the longest among these three HKs and has fewer interactions with the bound ATP molecule. It explains why the loop flexibility remains unaffected by the presence of any nucleotide, as observed from the RMS fluctuation of the ATP-lid loop residues (Figure 4.2 C) in the COM2C2A simulation.

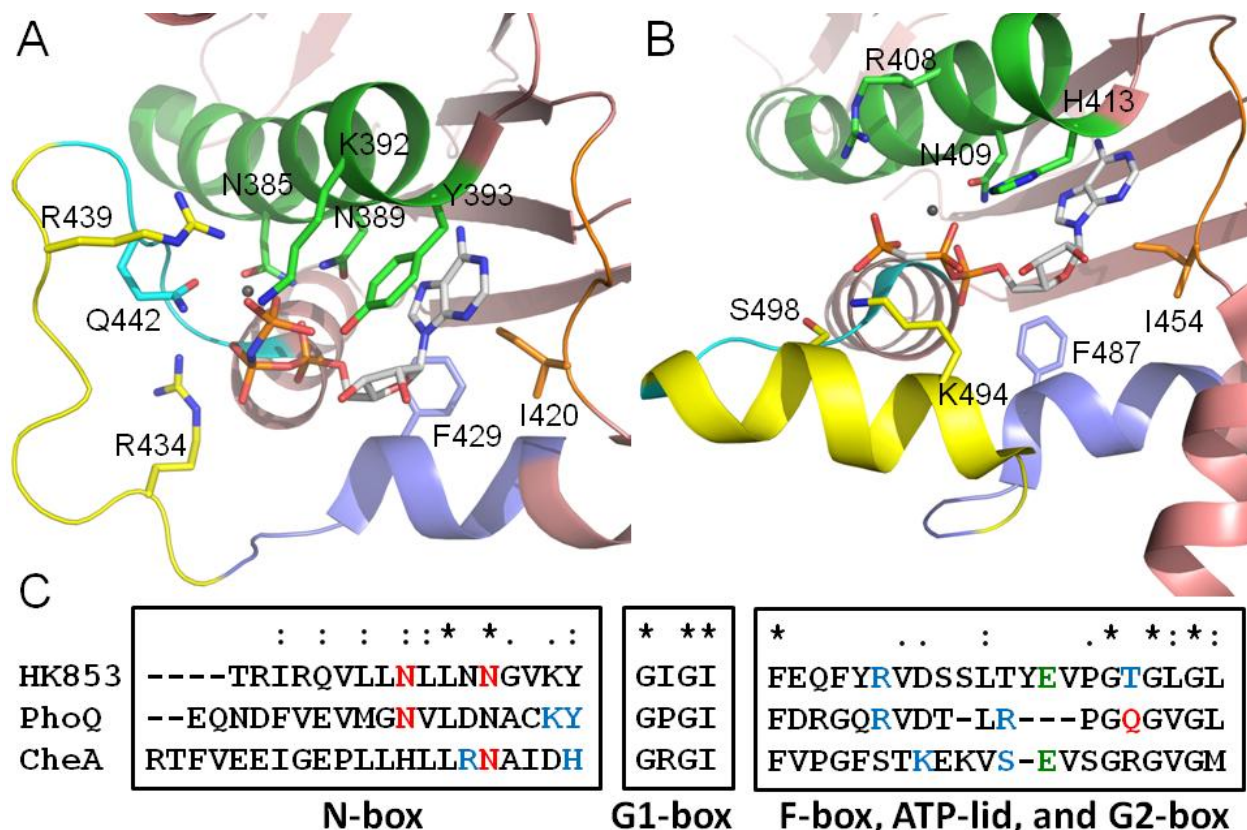


Figure 4.8. Active site in HK853, PhoQ and CheA. Catalytic and ATP-binding domains of (A) PhoQ (PDB id 1ID0), (B) CheA (PDB id 1I58) histidine kinase in presence of ATP analogues (shown in sticks with C-atoms in white) ADPNP and ADPCP, respectively. The C-atoms of the residues in the structural motifs N, G1, F, ATP-lid, and G2 structural motifs are shown in green, orange, blue, yellow, and cyan, respectively. (C) Sequence alignment of the HK853 with PhoQ and CheA. Residues hydrogen bonded to the ATP molecule are shown in blue. Residues chelated to the Mg^{2+} ion are shown in red. The residue E438 from the ATP-lid loop is shown in green that forms a salt bridge with K387 (not shown in the alignment). In the corresponding position in PhoQ there is a gap. In the corresponding position in CheA there is a glutamate (in green). In CheA, in the position equivalent to the K387 of HK853, there is a glutamate (alignment is not shown here).

Three basic residues that have direct interaction with the ATP analogue in the PhoQ crystal structures were subjected to kinetic analysis to understand their role in cleavage, and to transition state stabilization, during the autophosphorylation reaction (Marina et al., 2001). The residues, K392, R434 and R439 in PhoQ were individually mutated to alanine and kinetic parameters, K_m and k_{cat} were measured. It was suggested that K392 is critical for both nucleotide binding and catalysis. The R434 from the ATP-lid loop in PhoQ interacts with the ATP β -phosphate and is shown to be critical for catalysis, but not for binding. In spite of the fact that the R439 interacts with the ATP γ -phosphate, it was shown to have a modest effect in binding and catalysis. In the HK853 catalytic domain, R430 is the equivalent residue of R434 in PhoQ. In our first model (COM2C2A) of HK853, the R430 residue interacts with the β -phosphate of the ATP. Affinity for the non-hydrolysable ATP analogue AMPPNP (K_d 6.8 mM for the wild-type protein) was found to have moderate effects, due to the presence of R430 (K_d 28.3 mM for the R430A mutant) (Casino et al., 2014). But the R430A mutant was found to abrogate autophosphorylation activity in HK853 as well as in the presence of ATP, and this is very consistent with the finding in PhoQ case (Casino et al., 2014; Marina et al., 2001). The interaction between the ATP β -phosphate and R430 is fairly stable during simulation of the COM2C2A model (Figure 4.9 A), which represents a state more relevant for catalysis. During the 100 ns equilibrium simulation, this interaction is broken a number of times, resulting in the interacting atoms moving up to 15 Å from each other. This explains the observed only moderate effect of the R430A mutation on nucleotide binding. The interaction is not formed during the simulation of the other model, COM3DGE, although R430 does not move too far at any point during the simulation (Figure 4.9 B). In the HK853-RR468 crystal structure, the β and α -phosphate of the ADP is hydrogen bonded to the G2-box backbone. However, these interactions were not stable during simulations

starting from the COM3DGE model. Only the T442 hydrogen bond with the ATP γ -phosphate has moderate stability during simulation of this form. The ATP-lid orientation is similar to that in the HK853-RR468 complex, and may pave the way for the release of ADP through its phosphate end. Release of ADP prompts binding of ATP for a subsequent cycle of autophosphorylation. The ATP-lid in the COM3DGE model is folded toward the adenine base and the ribose sugar moiety, leaving the phosphate tail completely exposed. The fact that in the absence of nucleotide (ATP/ADP), as in APO3DGE, the ATP-lid loop remains close to the starting conformation, indicating there are interactions elsewhere that stabilize this product release form (Figure 4.2 B). Fluctuations observed for the ATP-lid loop residues in APO3DGE simulation are similar to those of the other form (APO2C2A or COM2C2A) in absence of any nucleotide present in the active site (Figure 4.2 C and D). However, in the presence of the nucleotide the fluctuations are minimal in COM3DGE (Figure 4.2 D).

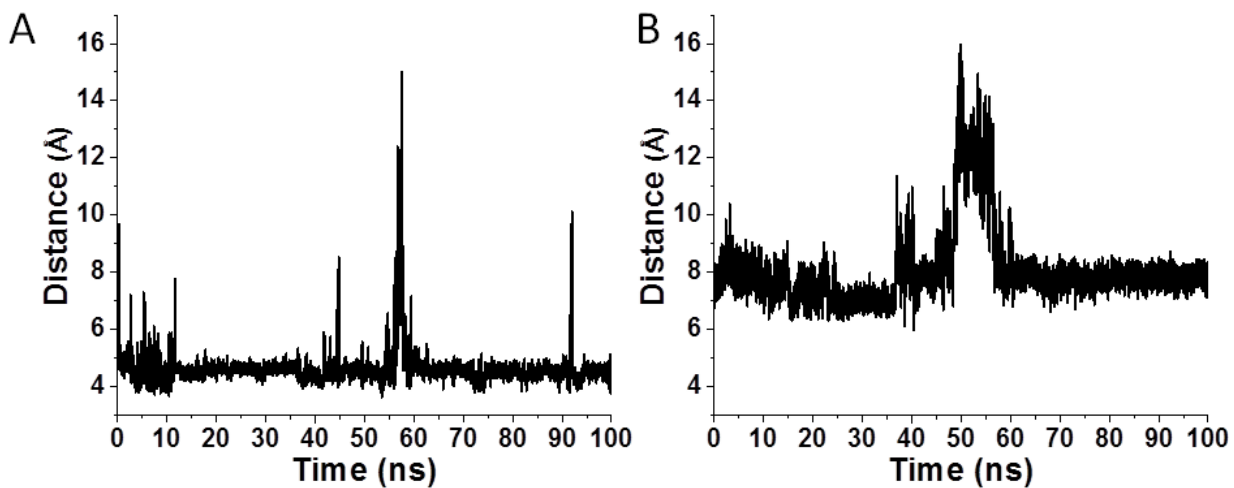


Figure 4.9. **Interaction of the ATP molecule with ATP-lid loop residue.** The distance between the R430 C ξ and ATP P γ atoms in the COM2C2A (A) and COM3DGE (B) simulation.

4.5.3. Role of a salt bridge in the conformational transition. The salt bridge between K387 and E438 locks the ATP-lid loop in the orientation observed in the HK853-RR468 complex structure. Residue K387 is from the loop following the N-box helix (D366-Y384) and E438 is from the ATP-lid loop that seems to significantly affect the ATP-lid orientation. The ATP-lid orientation angle changes when these two residues become solvated in any simulation. Changes in this orientation angle at ~70 ns and ~80 ns (Figure 4.4 B) during the simulation of the APO3DGE conformation are due to the dissociation of this salt bridge. This salt bridge was observed initially during simulation of the COM2C2A conformation in the presence of MgATP. As a result, the orientation angle remains in the range 40-50° for the first 30 ns of the simulation of the open ATP-lid conformation in the presence of ATP-Mg²⁺ (Figure 4.7). During the 35-70 ns interval the salt bridge between K387 and E438 is absent (Figure 4.9 B). As a result, the orientation angle adopts larger values and varies significantly in the same range (Figure 4.4 A in red).

4.5.4. Hydrophobic interactions of Val431 and Leu444 and their role in conformational plasticity. Hydrophobic interactions formed by Val431 in the COM3DGE of the ATP-lid are not observed in the other model, COM2C2A. The sidechain of Val431 is buried within a hydrophobic patch formed by Ile416, Pro417, Ala420, and Ile424 in the COM3DGE ATP-lid conformation. The strengths of the pairwise interaction involving Val431 were calculated using the Protein Structure Network analysis in Wordom (Table 4.3). There are relatively weak hydrophobic interactions formed at the other end of the ATP-lid loop (by Leu444) in the COM2C2A conformation of the ATP-lid with Ile448, and Ile452.

The salt bridge between K387 and E438 is a feature of the APO3DGE conformation, but this salt bridge is not observed in the APO2C2A simulation. This salt bridge seems to facilitate the

hydrophobic contacts formed by Val431 of the ATP-lid loop in the APO3DGE simulation. During the TMD simulation in the direction APO2C2A to APO3DGE-like conformation, the salt bridge forms before V431 makes hydrophobic contact with the Ile416, Pro417, Ala420, and Ile424 residues. Also, during the TMD the simulation in the reverse direction, the salt bridge breaks before the hydrophobic interactions of the V431. During the equilibrium simulation starting with the APO3DGE model, the salt bridge breaks briefly during the 70 and 80 ns and reforms, but the hydrophobic interactions of V431 remain intact (Figure 4.10 B). The salt bridge is observed several times during the 7-35 ns of the COM2C2A simulation (Figure 4.10 C). But every time it was short lived enough to not facilitate the V431 interactions to form that can orient the ATP-lid loop in a way similar to the conformation observed in the HK853-RR468 complex structure.

Table 4.3. **Hydrophobic interactions of Val431 and Leu444.**

Residue 1	Residue 2	Interaction Strength (initial conformation)		Average Interaction Strength (MD simulation)	
		COM2C2A	COM3DGE	COM2C2A	COM3DGE
Val431	Ile416	0.00	38.4	0.00	30.22
	Pro417	0.00	61.86	0.00	49.28
	Ala420	0.00	25.44	0.00	26.35
	Ile424	0.00	24.58	0.00	22.38
Leu444	Ile448	45.67	5.71	26.73	30.01
	Ile452	44.24	0.00	32.55	0.00

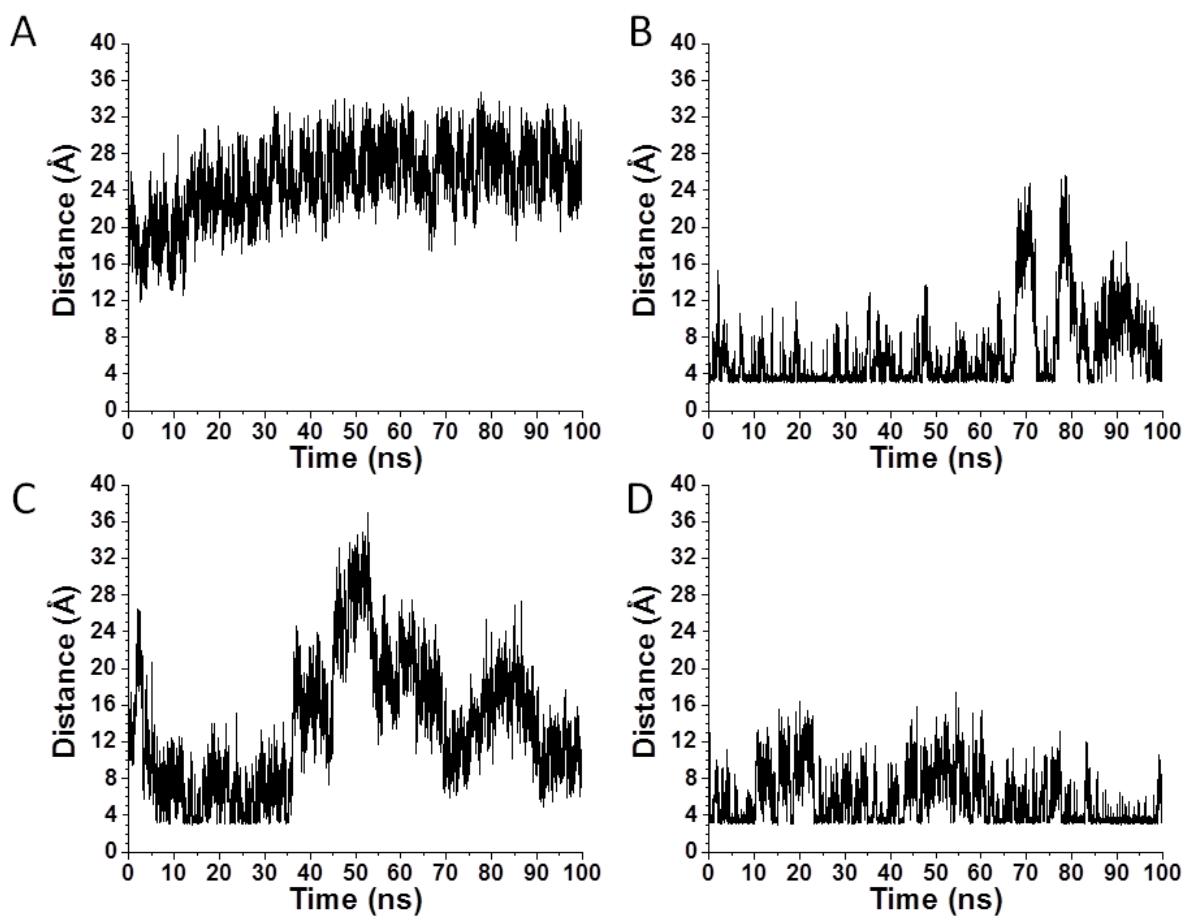


Figure 4.10. **Critical salt bridge interaction.** The distance between C δ of Glu438 and N ξ of Lys387 during CA-domain simulations (A) APO2C2A, (B) APO3DGE, (C) COM2C2A, and (D) COM3DGE.

4.6. Conclusion

In this work, two structural models have been built for the MgATP-bound complex of the HK853 CA domain, one based on the crystal structure of the CA domain (PDB code: 2C2A) in the absence of the response regulator RR468 and homology modeling of the ATP-lid loop, designated as COM2C2A, and the other based on the crystal structure of the CA domain (PDB code: 3DGE) in the presence of RR468, designated as COM3DGE. Molecular dynamics (MD) simulations have been performed on both models. Additional MD simulations have been performed on the two models with MgATP removed, designated as APO2C2A and APO3DGE, respectively. The four 100-ns MD simulations reveal three ensemble conformations, differing mainly in the conformation of the ATP-lid. The conformation of the ATP-lid can be described by an orientation angle defined by the line formed by the C α atoms of I476 and K468 and another line formed by the C α atoms of I476 and K468. One conformation is very similar to the starting conformation of COM2C2A, representing the conformation of the MgATP-bound form. Another conformation is very similar to the starting conformation of APO3DGE, representing the conformation after the phosphoryl transfer. The third conformation is different from both starting conformations, representing the conformation of the apo form of the CA domain. The ATP-lid orientation angles indicate that the ATP lid adopts the apo conformation during the APO2C2A simulation, whereas during the COM2C2A simulation, it adopts mainly the MgATP-bound conformation but can also sample the apo conformation. The ATP lid stays in the ADP-bound conformation during the COM3DGE simulation and largely remain in the the ADP-bound conformation during the APO3DGE simulation with occasional excursions to the MgATP-bound conformation. The MD simulations also reveal critical interactions that stabilize these conformations and their roles in the conformational transitions are further investigated by

targeted MD simulations. The histidine kinase HK853 has a longer ATP-lid loop compared to the other two HKs, CheA and PhoQ. In HK853, the interactions between ATP and ATP-lid residues are minimal compared to PhoQ and CheA HKs. That explains the greater flexibility of the ATP-lid loop in HK853. Unlike other HKs, the lack of order observed in the ATP-lid of HK853 may be independent of ligand binding. The opening-closing motion of the ATP-lid loop is strongly related to the salt bridge formed between an ATP-lid residue and another residue from a loop adjacent to the N-box helix. This salt bridge, along with the hydrophobic interaction of another ATP-lid residue, may restrict the exposure of the ATP-binding site and hence can affect the autophosphorylation reaction.

REFERENCES

REFERENCES

- Bilwes, A.M., Alex, L.A., Crane, B.R., and Simon, M.I. (1999). Structure of CheA, a signal-transducing histidine kinase. *Cell* 96, 131-141.
- Bilwes, A.M., Quezada, C.M., Croal, L.R., Crane, B.R., and Simon, M.I. (2001). Nucleotide binding by the histidine kinase CheA. *Nat Struct Biol* 8, 353-360.
- Brinda, K.V., and Vishveshwara, S. (2005). A network representation of protein structures: Implications for protein stability. *Biophys J* 89, 4159-4170.
- Case, D.A., Darden, T.A., III, T.E.C., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Crowley, M., Walker, R.C., Zhang, W., *et al.* (2008). AMBER 10 (University of California, San Francisco.).
- Casino, P., Fernandez-Alvarez, A., Alfonso, C., Rivas, G., and Marina, A. (2007). Identification of a novel two component system in *Thermotoga maritima*. Complex stoichiometry and crystallization. *Bba-Proteins Proteom* 1774, 603-609.
- Casino, P., Miguel-Romero, L., and Marina, A. (2014). Visualizing autophosphorylation in histidine kinases. *Nat Commun* 5.
- Casino, P., Rubio, V., and Marina, A. (2009). Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell* 139, 325-336.
- Casino, P., Rubio, V., and Marina, A. (2010). The mechanism of signal transduction by two-component systems. *Curr Opin Struct Biol* 20, 763-771.
- Cheatham, T.E.I. (2008). AmberTools Users' Manual.
- Dago, A.E., Schug, A., Procaccini, A., Hoch, J.A., Weigt, M., and Szurmant, H. (2012). Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis. *Proc Natl Acad Sci USA* 109, E1733-E1742.
- Gao, R., and Stock, A.M. (2009). Biological Insights from Structures of Two-Component Proteins. *Annu Rev Microbiol* 63, 133-154.
- Gotoh, Y., Eguchi, Y., Watanabe, T., Okamoto, S., Doi, A., and Utsumi, R. (2010). Two-component signal transduction as potential drug targets in pathogenic bacteria. *Curr Opin Microbiol* 13, 232-239.
- Marina, A., Mott, C., Auyzenberg, A., Hendrickson, W.A., and Waldburger, C.D. (2001). Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. Insight into the reaction mechanism. *J Biol Chem* 276, 41182-41190.
- Marina, A., Waldburger, C.D., and Hendrickson, W.A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J* 24, 4247-4259.

Mascher, T., Helmann, J.D., and Uden, G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev* 70, 910-+.

Roe, D.R., and Cheatham, T.E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* 9, 3084-3095.

Seeber, M., Felling, A., Raimondi, F., Muff, S., Friedman, R., Rao, F., Caflisch, A., and Fanelli, F. (2010). Wordom: A User-Friendly Program for the Analysis of Molecular Structures, Trajectories, and Free Energy Surfaces. *J Comput Chem* 32, 1183-1194.

Song, Y., Peisach, D., Pioszak, A.A., Xu, Z.H., and Ninfa, A.J. (2004). Crystal structure of the c-terminal domain of the two-component system W transmitter protein nitrogen regulator II (NRII; NtrB), regulator of nitrogen assimilation in *Escherichia coli*. *Biochemistry-Us* 43, 6670-6678.

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-component signal transduction. *Annu Rev Biochem* 69, 183-215.

Wuichet, K., Cantwell, B.J., and Zhulin, I.B. (2010). Evolution and phyletic distribution of two-component signal transduction systems. *Curr Opin Microbiol* 13, 219-225.

Zhang, J., Xu, Y.C., Shen, J.H., Luo, X.M., Chen, J.G., Chen, K.X., Zhu, W.L., and Jiang, H.L. (2005). Dynamic mechanism for the autophosphorylation of CheA histidine kinase: Molecular dynamics simulations. *J Am Chem Soc* 127, 11709-11719.

CHAPTER 5

Conclusions and Perspective

5.1. Understanding signal transduction in a Two-component system

The information flow in Two-component system signal transduction begins with detection of an environmental stimulus by the input or extracellular domain of a HK. Signal transmission in the structural sense is understood so far as it involves conserved domains of the sensor kinase and response regulator protein in the intracellular regions. Signal transmission through these domains takes place via kinase, phosphoryl transfer, and phosphatase reactions where the information is encoded through a covalently attached phosphoryl group with HK and RR. However, the molecular basis for susceptibility or specificity toward different stimuli in the extracellular sensor domains is poorly understood (Bourret and Silversmith, 2010). A number of crystal structures have been determined from representative sensor domains of transmembrane HKs that suggest the extracellular sensor domain(s) are extremely diverse (Zhang and Hendrickson, 2010). There have been several previous attempts to classify extracellular sensor domains using sequence-structure-function analysis. These studies are more or less inconclusive to this date, due to lack of information on the functional role of the different TCSs (Anantharaman and Aravind, 2000). In a previous study it was shown that the signal amplification, sensory adaptation, molecular memory and high sensitivity of the extracellular sensory domains can be associated with the oligomeric state of the protein (Hazelbauer and Lai, 2010). The molecular basis for signal transmission through the transmembrane region of the HKs is also poorly understood due to lack of structural information on the transmembrane region of the membrane bound sensor HKs. Signal transduction is mediated by phosphorylation of the RR by the HK, which induces or stabilizes the active conformation of the RR protein (Bourret, 2010). Apart from phosphoryl transfer activity, the HK in some TCSs consist of phosphatase activity as well. In several TCSs including HK853-RR468 cognate pair, the HK also can catalyze the

dephosphorylation of the phosphorylated RR (Stock et al., 2000). However, a molecular basis for the dual activity of the HK853 is still not clearly understood. A complete understanding of signal transduction in TCSs depends on the resolution of the issues mentioned above.

Structural information on conserved cytoplasmic domains of HK and RR paved the way to understand how signal transmission is encoded through a phosphoryl group associated with HK or RR. A conformational transition in RR is the key to generate the response against an environmental stimulus. Our work on RR468 shows possible pathways for how this conformational transition can be achieved, from inactive to active form or vice versa. Mechanisms for enzymatic activation are understood based on induced fit and population models (Koshland, 1958; Ma et al., 1999; Tsai et al., 1999). But, the truth may be somewhere in between. Using molecular dynamics and different statistical analysis techniques we have shown that there is some information embedded in the MD trajectory starting with one form about the direction of the conformational change required to sample the other, but other interactions may be required to complete the transition-a situation in between induced fit and conformational selection model.

5.2. Perspective

Understanding of metastable states during conformational dynamics of a macromolecule is the key to understanding the process of conformational transition in RR468, the model system of our interest. Identification of predominant conformational states for a given molecule and subsequent structural characterization of these intermediates may find broader application across the field. Structural discretization of resulting conformations using Markov state modeling is applied on some medically relevant systems to identify allosteric sites in recent literature (Bowman and Geissler, 2012). Identification of unique allosteric sites associated with one enzyme could help in

targeting that particular enzyme selectively against a series of related enzymes. Identification of unique intermediate states along the activation path could also be potentially used for drug design, as shown in recent literature (Shukla et al., 2014). Understanding of critical non-native interactions in the activation pathway of a protein can be tested experimentally using mutational studies. Subtle differences in the activation mechanisms could enable selective targeting of one particular protein (or enzyme) for the purpose of structure-based drug design.

REFERENCES

REFERENCES

- Anantharaman, V., and Aravind, L. (2000). Cache - a signaling domain common to animal Ca²⁺ channel subunits and a class of prokaryotic chemotaxis receptors. *Trends Biochem Sci* 25, 535-537.
- Bourret, R.B. (2010). Receiver domain structure and function in response regulator proteins. *Curr Opin Microbiol* 13, 142-149.
- Bourret, R.B., and Silversmith, R.E. (2010). Two-component signal transduction. *Curr Opin Microbiol* 13, 113-115.
- Bowman, G.R., and Geissler, P.L. (2012). Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc Natl Acad Sci USA* 109, 11681-11686.
- Hazelbauer, G.L., and Lai, W.C. (2010). Bacterial chemoreceptors: providing enhanced features to two-component signaling. *Curr Opin Microbiol* 13, 124-132.
- Koshland, D.E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci USA* 44, 98-104.
- Ma, B.Y., Kumar, S., Tsai, C.J., and Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein Eng* 12, 713-720.
- Shukla, D., Meng, Y.L., Roux, B., and Pande, V.S. (2014). Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat Commun* 5.
- Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-component signal transduction. *Annu Rev Biochem* 69, 183-215.
- Tsai, C.J., Kumar, S., Ma, B.Y., and Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Sci* 8, 1181-1190.
- Zhang, Z., and Hendrickson, W.A. (2010). Structural Characterization of the Predominant Family of Histidine Kinase Sensor Domains. *J Mol Biol* 400, 335-353.