

COMPUTATIONAL AND EXPERIMENTAL INVESTIGATIONS INTO THE
MECHANISM OF FUNCTIONAL MOTION IN BIOMOLECULAR SYSTEMS

by

Ayşe Burcu Aykaç Fas

B.S., Chemical Engineering, Boğaziçi University, 1998

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

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

Graduate Program in Chemical Engineering
Boğaziçi University

2016

ACKNOWLEDGEMENTS

Firstly, I would like to express my gratitude to my thesis advisor, Prof. Türkan Halilođlu for her guidance and for sharing her scientific knowledge with me throughout my studies. Special thanks to Assoc. Prof. Hamdi Torun for his guidance/support in our Atomic Force Microscopy studies and for adding the experimental touch to our research.

I'm grateful to the members of my thesis committee Prof. Pemra Doruker Turgut, Assoc. Prof. Ebru Demet Akdođan, Assoc. Prof. Hamdi Torun and Assist. Prof. Özge Kürkçüođlu Levitas, who have read and commented on my thesis.

I am truly indebted to Prof. Zeynep İlsen Önsan for her guidance and presence in my life and the Chemical Engineering Department, Bođaziçi University. I am so lucky to be surrounded by my lifelong ChE family.

I would like to thank my PRC companions, Burçin Acar, Erge Akbař, Gökçehan Kara, Dr. Ece Acuner Özbabacan, Burcu Özden, Serdar Özsezen, Dr. Pemra Özbek Sarıca, Dr. Zeynep Kürkçüođlu Soner, Dr. Seren Soner, Dr. Fidan Sümbül and Dr. Arzu Uyar for their scientific/nonscientific support and friendship along my PhD journey. I would also like to thank Canan Dedeođlu and Yüksel Gülten for their kindness and positiveness in PRC.

It was a great opportunity for me that my PhD research has been supported by the projects Bođaziçi University B.A.P. (8424/14A05M6), TÜBİTAK 1001 (112T569), COSBIOM (FP6-2004-ACC-5SA-2) and Betil Fund.

And last but not least, I am thankful to my family for always being there for me with endless love, encouragement and support. Special thanks go to my husband Genco and my precious son Rüzgar, for being an inspiration to go on. This thesis is dedicated to my mother who I wish had lived to celebrate this accomplishment. I

thank my mother, for being an accomplished scientist, a loving mom-grandmother and a perfect role model for many around her in all aspects of life.

ABSTRACT

COMPUTATIONAL AND EXPERIMENTAL INVESTIGATIONS INTO THE MECHANISM OF FUNCTIONAL MOTION IN BIOMOLECULAR SYSTEMS

The seemingly random fluctuations that proteins exhibit in their native states actually harbor major contributions from the global/cooperative motions they undergo during functional transitions. Computationally efficient, robust approaches like elastic network models (ENM) that exploit the intrinsic dynamics of proteins are of great value in building up hybrid approaches bridging experimental and computational studies and extending each application's range. ENM global modes show functional significance and compares well with functional conformational changes. The purpose of this study is to build upon currently available computational techniques that generate transition pathways connecting end states. In this thesis, a practical hybrid approach, anisotropic network model (ANM) guided Langevin dynamics (LD) simulations (ANM-LD) has been developed, where the dominant ANM low-frequency modes are utilized to drive LD simulations. In ANM-LD, the initial conformation is moved along the selected ANM mode at each cycle, taking advantage of the evaluation of the interactions and energetics of the system via short cycles of all-atom implicit LD simulations. A detailed assessment of the method, in terms of creating physically meaningful pathways and approaching the final state, are carried out for a set of proteins. As exemplary cases, the analysis and results for adenylate kinase and maltose transporter is presented. ANM-LD was also applied to heat shock protein 90 (Hsp90) in apo, ATP-bound and Geldanamycin (GDM)-bound states, the latter combined with AFM experiments, to study the effect of temperature on the Hsp90-GDM binding dynamics. Lastly, conventional molecular dynamics simulations, ENM and ANM-LD are utilized to investigate CRP dynamics.

ÖZET

BİYOMOLEKÜLER SİSTEMLERDE FONKSİYONEL HAREKET MEKANİZMASI ÜZERİNE HESAPLAMALI VE DENEYSEL ARAŞTIRMALAR

Proteinlerin doğal hallerinde sergiledikleri görünürde rastgele dalgalanmalar, işlevsel geçişlerde etkisinde oldukları global/kooperatif hareketlerden önemli katkılar barındırır. Proteinlerin iç dinamiğini kullanan elastik ağyapı modelleri (ENM) gibi hesap açısından verimli ve sağlam yaklaşımlar, deneysel ve hesaplamalı çalışmaları birleştiren hibrit yaklaşımlara altyapı oluşturması açısından çok değerlidir, ve her bir uygulama alanının kapsamını genişletirler. ENM global modları fonksiyonel önem gösterir ve işlevsel konformasyonel değişikliklerle iyi örtüşürler. Bu tezin amacı, kararlı formları birbirine bağlayan geçiş patikaları üreten mevcut hesaplama tekniklerini ileriye götüren pratik bir platform oluşturmaktır. Bu çalışmada, anizotropik ağyapı modeli (ANM) düşük frekans modlarının Langevin dinamiği (LD) simülasyonlarını yönlendirme için kullanıldığı hibrit bir yaklaşım olan, ANM rehberli LD simülasyonları (ANM-LD) yöntemi geliştirilmiştir. ANM-LD’de, başlangıç konformasyonu her döngüde hedef yapı ile örtüşen ANM modu yönünde deforme edilir, gerçekleştirilen kısa atomik LD simülasyonları ile sistemin etkileşimlerinin ve enerjisinin düzeltilmesinden faydalanarak bu şekilde simülasyonlarda hedef yapıya yaklaşılır. Yöntemin, fiziksel olarak anlamlı geçiş patikaları oluşturma ve son haline yaklaşma açısından ayrıntılı bir değerlendirmesi bir grup protein için gerçekleştirilmiştir. Örnek sistem olarak, adenilat kinaz ve maltoz taşıyıcı proteini için analiz ve sonuçlar sunulmuştur. Apo, ATP- ve Geldanamisin (GDM)- bağlı ısı şok proteini 90 (Hsp90) sistemlerinin ANM-LD simülasyonları, Hsp90-GDM AFM deneyleri ile birleştirilerek, sıcaklığın Hsp90-GDM bağlanma dinamikleri üzerindeki etkisi incelenmiştir. Son olarak, klasik moleküler dinamik simülasyonları, ENM ve ANM-LD yöntemleri, CRP dinamiğini araştırmak için kullanılmıştır.

TABLE OF CONTENTS

| | |
|---|------|
| ACKNOWLEDGEMENTS | iii |
| ABSTRACT | v |
| ÖZET | vi |
| LIST OF FIGURES | x |
| LIST OF TABLES | xvi |
| LIST OF SYMBOLS | xvii |
| LIST OF ACRONYMS/ABBREVIATIONS | xx |
| 1. INTRODUCTION | 1 |
| 1.1. Protein Dynamics and Allostery | 1 |
| 1.2. Protein Dynamics by Experimental Methods | 3 |
| 1.2.1. Atomic Force Microscopy | 4 |
| 1.3. Protein Dynamics by Computational Methods | 5 |
| 1.3.1. Normal Mode Analysis and Elastic Network Models | 5 |
| 1.3.2. Molecular Simulations | 6 |
| 1.3.2.1. Molecular (and Langevin) Dynamics | 6 |
| 1.3.3. Advanced/Hybrid Molecular Simulation Methodologies | 7 |
| 1.4. Specific Aims and Outline of the Thesis | 16 |
| 2. MATERIALS AND METHODS | 18 |
| 2.1. Elastic Network Models | 18 |
| 2.1.1. Gaussian Network Model (GNM) | 18 |
| 2.1.2. Anisotropic Network Model (ANM) | 19 |
| 2.2. Molecular Simulations | 20 |
| 2.2.1. Molecular Dynamics | 20 |
| 2.2.2. Langevin Dynamics | 24 |
| 2.3. ANM-LD Methodology | 25 |
| 2.3.1. Parameters | 29 |
| 2.3.2. Computational Details and the Protein Dataset | 29 |
| 2.4. Characterization of conformational states by GNM eigenvalues | 30 |
| 3. CONFORMATIONAL TRANSITION PATHWAYS BY COLLECTIVE MODES | |

| | |
|--|----|
| DRIVEN MOLECULAR SIMULATIONS | 33 |
| 3.1. Introduction | 34 |
| 3.2. Materials and Methods | 36 |
| 3.2.1. Structures | 36 |
| 3.2.2. ANM-LD Simulation Protocol | 36 |
| 3.2.3. Adjustment of Parameters | 36 |
| 3.3. Results and Discussion | 37 |
| 3.3.1. Model Case: Adenylate Kinase | 37 |
| 3.3.1.1. Analysis of Pathways | 39 |
| 3.3.2. Case Study: MalFGK2 | 60 |
| 3.4. Conclusions | 66 |
| 4. CONFORMATIONAL TRANSITION AND THE EFFECT OF TEMPERA- TURE ON BINDING ENERGY LANDSCAPE OF HSP90–GELDANAMYCIN COMPLEX | 69 |
| 4.1. Introduction | 69 |
| 4.2. Materials and Methods | 73 |
| 4.2.1. Preparation of Molecules and AFM Setup | 73 |
| 4.2.2. Immobilization of Hsp90 Molecules and Functionalization of AFM Tips with Geldanamycin | 74 |
| 4.2.3. AFM Force Measurements and Analysis | 75 |
| 4.2.4. Structures | 76 |
| 4.2.5. ANM-LD Simulation Protocol | 79 |
| 4.3. Results and Discussion | 79 |
| 4.3.1. Single-Molecule Force Measurements: The Effect of Temperature on the Energy Landscape | 81 |
| 4.3.2. Binding Energy Landscape of Hsp90-Geldanamycin Complex | 81 |
| 4.3.3. Global Dynamics of Hsp90 Conformational Transition | 83 |
| 4.4. Conclusions | 93 |
| 5. DYNAMIC SWITCH MECHANISM IN APO CYCLIC AMP RECEPTOR PROTEIN | 98 |
| 5.1. Introduction | 98 |

| | |
|---|-----|
| 5.2. Materials and Methods | 102 |
| 5.2.1. Molecular Dynamics Simulations | 102 |
| 5.2.2. ANM-LD Simulation Protocol | 104 |
| 5.3. Results and Discussion | 105 |
| 5.3.1. Fluctuations and Correlations by MD Simulations | 105 |
| 5.3.2. Fluctuations and Correlations by GNM | 106 |
| 5.3.3. The Inter-domain Hinge as Allosteric Switch in CRP | 108 |
| 5.3.4. Internal Fluctuations and Allostery | 110 |
| 5.3.5. Global Dynamics by ANM-LD Simulations | 113 |
| 5.4. Conclusions | 116 |
| 6. CONCLUSIONS | 118 |
| 7. FUTURE WORK | 122 |
| REFERENCES | 123 |

LIST OF FIGURES

| | | |
|-------------|---|----|
| Figure 1.1. | Schematic representation of atomic force microscopy and its operating modes. | 4 |
| Figure 1.2. | Characteristic time scales for protein motions. | 8 |
| Figure 2.1. | Flowchart of the algorithm of ANM Driven Langevin Dynamics (ANM-LD) Simulations. | 32 |
| Figure 3.1. | Adenylate kinase structures. | 39 |
| Figure 3.2. | RMSD plots-AdK forward runs (open-closed) for different DF values. | 42 |
| Figure 3.3. | GNM eigenvalue of first eigenvector (eV1) versus GNM eigenvalue of second eigenvector (eV2) plots – AdK forward runs (open-closed) for different DF values. | 42 |
| Figure 3.4. | RMSD plots–AdK reverse runs (closed-open) for different DF values. | 43 |
| Figure 3.5. | GNM eigenvalue of first eigenvector (eV1) versus GNM eigenvalue of second eigenvector (eV2) plots–AdK reverse runs (closed-open) for different DF value. | 43 |
| Figure 3.6. | GNM eigenvalue of first eigenvector (eV1) versus GNM eigenvalue of second eigenvalue (eV2) plots for AdK forward and reverse runs. | 44 |
| Figure 3.7. | Overlap/collectivity values and RMSD plots with the selected slow modes for AdK unrestricted forward simulations with $DF=0.4 \text{ \AA}$ | 45 |

| | | |
|--------------|--|----|
| Figure 3.8. | Overlap/collectivity values and RMSD plots with the selected slow modes for AdK restricted forward simulations with $DF=0.4 \text{ \AA}$. . . | 46 |
| Figure 3.9. | Overlap/collectivity values and RMSD plots with the selected slow modes for AdK unrestricted reverse simulations with $DF=0.4 \text{ \AA}$. . . | 47 |
| Figure 3.10. | Overlap/collectivity values and RMSD plots with the selected slow modes for AdK restricted reverse simulations with $DF=0.4 \text{ \AA}$. . . | 48 |
| Figure 3.11. | GNM first eigenvalue (eV1) histograms for AdK forward runs. . . | 49 |
| Figure 3.12. | GNM first eigenvalue (eV1) histograms for AdK reverse runs. . . . | 49 |
| Figure 3.13. | GNM first eigenvalue (eV1) histograms for AdK forward runs with $DF=0.4 \text{ \AA}$ | 50 |
| Figure 3.14. | GNM first eigenvalue (eV1) histograms for AdK reverse runs with $DF=0.4 \text{ \AA}$ | 50 |
| Figure 3.15. | NMP-CORE angle versus LID-CORE angle for AdK forward runs. | 51 |
| Figure 3.16. | NMP-CORE angle versus LID-CORE angle for AdK forward runs with $DF=0.4 \text{ \AA}$ | 51 |
| Figure 3.17. | NMP-CORE angle versus LID-CORE angle for AdK reverse runs. | 52 |
| Figure 3.18. | NMP-CORE angle versus LID-CORE angle for AdK reverse runs with $DF=0.4 \text{ \AA}$ | 52 |
| Figure 3.19. | ANM-LD forward/reverse pathways on the potential of mean force (PMF) data provided by Beckstein Lab on their website. | 53 |

| | | |
|--------------|--|----|
| Figure 3.20. | Cross-Correlation Plots for AdK forward runs. | 54 |
| Figure 3.21. | Cross-Correlation Plots for AdK reverse runs. | 55 |
| Figure 3.22. | Cross-Correlation Plots for AdK forward runs: Unrestricted versus restricted case. | 57 |
| Figure 3.23. | Cross-Correlation Plots for AdK reverse runs: Unrestricted versus restricted case. | 58 |
| Figure 3.24. | GNM eV1 versus eV2 and LID-CORE/NMP-CORE angle plots of mutated AdK (Y171W and P177A) ANM-LD runs. | 59 |
| Figure 3.25. | Functional sites of MalFGK2 are presented on the OF structure. . | 62 |
| Figure 3.26. | MalFGK2 IF→OF transition case run RMSD plot and transition cross-correlation map. | 63 |
| Figure 3.27. | MalFGK2 IF→OF transition cross-correlations color coded on the ribbon structure. | 65 |
| Figure 3.28. | MalFGK2 OF→IF transition case run RMSD plot and transition cross-correlation map. Correlation of maltose binding sites and ATP-binding sites with the rest of the structure. | 66 |
| Figure 3.29. | MalFGK2 OF→IF transition cross-correlations color coded on the ribbon structure. Correlation of maltose binding sites and ATP- binding sites with the rest of the structure. | 67 |
| Figure 4.1. | Hsp90 domain architecture for yeast Hsp90. | 71 |

| | | |
|--------------|---|----|
| Figure 4.2. | Different conformational states of Hsp90. | 73 |
| Figure 4.3. | The molecular formula, geometry and the ligand conformations in the Hsp90 binding pocket. | 77 |
| Figure 4.4. | Schematic representation of Hsp90-Geldanamycin AFM pulling experiment setup, force curves and binding energy landscapes. | 78 |
| Figure 4.5. | Adhesion force histograms and most probable rupture forces obtained at different loading rates of Hsp90-Geldanamycin AFM pulling experiments at 300, 310 K. | 82 |
| Figure 4.6. | The dynamic adhesion force spectra of Hsp90-Geldanamycin complex depending on the loading rate obtained at 300, 310 K. | 84 |
| Figure 4.7. | RMSD plots for Hsp90 forward (open-closed) runs with different ligands bound and for different <i>modemax</i> and <i>T</i> values. | 86 |
| Figure 4.8. | Overlap/collectivity values and RMSD plots with the selected slow modes for Hsp90 forward case runs (Apo/ATP-bound/GDM-bound) and the reverse case run. | 87 |
| Figure 4.9. | GNM first eigenvector (eV1) versus second eigenvector (eV2) plots for Hsp90 forward runs (Apo/ATP-bound/GDM-bound). | 88 |
| Figure 4.10. | GNM first eigenvalue (eV1) histograms for Apo Hsp90 forward runs. | 89 |
| Figure 4.11. | GNM first eigenvalue (eV1) histograms for Apo Hsp90 reverse runs. | 89 |
| Figure 4.12. | GNM first eigenvalue (eV1) histograms for Hsp90-ATP forward runs. | 90 |

| | |
|---|-----|
| Figure 4.13. GNM first eigenvalue (eV1) histograms for Hsp90-GDM forward runs. | 90 |
| Figure 4.14. Cross-correlation plots for Apo Hsp90 forward runs. | 92 |
| Figure 4.15. Cross-correlation plots for Hsp90-ATP forward runs. | 92 |
| Figure 4.16. Cross-correlation plots for Hsp90-GDM forward runs. | 93 |
| Figure 4.17. Cross-correlation difference plots for Hsp90-ATP and Hsp90-GDM forward runs with reference to Apo Hsp90. | 94 |
| Figure 4.18. MSF difference plots for Hsp90-GDM forward runs. | 95 |
| Figure 5.1. Holo CRP crystal structure and functional sites. | 100 |
| Figure 5.2. Dynamics network of apo CRP monomer/dimer and holo CRP by MD simulations. | 107 |
| Figure 5.3. GNM cross-correlations of apo/holo CRP and the switch behavior. | 109 |
| Figure 5.4. RMSD plots for CRP apo-holo (open-closed) forward runs. | 113 |
| Figure 5.5. Overlap/collectivity values and RMSD plots with the selected slow modes for CRP Model 10 apo-holo simulations | 114 |
| Figure 5.6. GNM eV1 versus eV2 plots for CRP Model 10 apo-holo simulations. | 115 |
| Figure 5.7. GNM eV1 histograms for CRP apo, holo MD simulations, and CRP Model 10 apo-holo ANM-LD simulations. | 116 |

| | |
|---|-----|
| Figure 5.8. Cross-correlation plots for CRP Model 10 apo-holo runs. | 117 |
|---|-----|

LIST OF TABLES

| | | |
|------------|---|-----|
| Table 2.1. | Details of the proteins/models studied. | 30 |
| Table 3.1. | Summary of ANM-LD simulation results. | 38 |
| Table 4.1. | Hsp90: Summary of ANM-LD simulation results. | 80 |
| Table 4.2. | Parameters of the Hsp90-Geldanamycin interaction obtained from the Bell model. | 83 |
| Table 5.1. | CRP: Details of the simulated systems. | 103 |
| Table 5.2. | CRP: Summary of ANM-LD simulation results. | 105 |

LIST OF SYMBOLS

| | |
|-------------------------------|---|
| a_i | Acceleration of particle i |
| A_{ij} | van der Waals repulsion parameter |
| B_i | Debye-Waller or B-factor |
| B_{ij} | van der Waals attraction parameter |
| C_{ij} | Cross-correlation value between the fluctuations of residues i and j |
| $C\alpha$ | Alpha carbon coordinates |
| DF | Deformation factor |
| d_i | Distance between the coordinates of atom i in two structures |
| \mathbf{D}_i | Distance difference vector |
| f | Force |
| f^* | Most probable rupture force |
| F_i | Force exerted on particle i |
| \mathbf{H} | Hessian matrix |
| \mathbf{I}_i | Initial conformation coordinates in cycle i |
| $\mathbf{I}_{i,m}$ | Perturbed initial conformation obtained by deforming \mathbf{I}_i along the mostly overlapping mode m_i in cycle i |
| $\mathbf{I}_{i,m,\min}$ | Perturbed initial conformation obtained by energy minimizing $\mathbf{I}_{i,m}$ for a <i>stepmin</i> number of steps in cycle i |
| $\mathbf{I}_{i,m,\text{sim}}$ | Perturbed initial conformation obtained by simulating $\mathbf{I}_{i,m,\min}$ for a <i>stepsim</i> number of steps in cycle i |
| k | Mode |
| k_B | Boltzmann constant |
| k_l | Force constant for the bond |
| k_{off} | Unbinding rate constant |
| k_θ | Force constant for the bond angle |
| l_{eq} | Equilibrium bond length |
| l_i | Bond length between atoms $i - 1$ and i |
| m_i | Mostly overlapping ANM mode (selected mode) for cycle i |

| | |
|----------------------|--|
| M_i | Mass of particle i |
| $modemax$, MM | Maximum number of modes to choose among |
| N | Number of atoms |
| $O_{i,j}$ | Overlap value of j th eigenvector, \mathbf{u}_j , with the difference vector, \mathbf{D}_i , for cycle i |
| $O_{i,max}$ | Maximum overlap value for cycle i |
| p | Overall momentum |
| P | Pressure |
| q_i | Partial atomic charge of atom i |
| r_f | Loading rate |
| r_0 | The distance of the activated state from the bound state |
| r_{ij} | Interparticle distance |
| R_{cut} | Cut-off radius |
| \mathbf{R}_i | Position vector of residue i |
| R_{ij} | Instantaneous distance between residues i and j |
| R_{ij}^0 | Equilibrium distance between residues i and j |
| $\Delta\mathbf{R}_i$ | Fluctuation of residue i |
| s | Sensitivity |
| $stepmin$ | Number of minimization steps |
| $stepsim$ | Number of simulation steps |
| t | Time |
| T | Absolute temperature |
| \mathbf{T} | Target conformation coordinates |
| \mathbf{u}_k | k th eigenvector |
| \mathbf{U} | Matrix of eigenvectors |
| V | Potential function |
| V_{ANM} | ANM potential function |
| \AA | Angstrom |
| α | Alpha helix |
| β | Beta strand |

| | |
|---------------------|---|
| γ_f | Force constant for harmonic spring |
| γ | Damping constant of Langevin dynamics simulations |
| ϵ | Dielectric constant |
| ζ | Friction (damping) coefficient |
| θ_i | Instantaneous bond angle of atom i |
| θ_{eq} | Equilibrium bond angle |
| κ | The degree of collectivity |
| κ_m | The degree of collectivity of the mostly overlapping ANM mode m |
| λ_k | k th eigenvalue |
| ν_i | Velocity of atom i |
| π | Pi number |
| ϕ_i, φ_i | Dihedral (torsional) angle of bond i |
| $\mathbf{\Gamma}$ | Kirchhoff (or connectivity) matrix |
| Λ | Diagonal matrix of eigenvalues |

LIST OF ACRONYMS/ABBREVIATIONS

| | |
|----------|--|
| 3D | Three dimensional |
| aANM | Adaptive ANM |
| ABC | ATP-Binding Cassette |
| AdK | Adenylate Kinase |
| ADP | Adenosine Diphosphate |
| AFM | Atomic Force Microscopy |
| Amber | Assisted Model Building with Energy Refinement |
| AMD | Accelerated Molecular Dynamics |
| AMP | Adenosine Monophosphate |
| ANM | Anisotropic Network Model |
| ANM-LD | Anisotropic Network Model Guided Langevin Dynamics |
| ANM-MC | Anisotropic Network Model-Monte Carlo |
| ATP | Adenosine Triphosphate |
| BD | Brownian Dynamics |
| BEL | Binding Energy Landscape |
| BtuCD | Vitamin B_{12} Transporter |
| cAMP | Cyclic Adenosine Mono Phosphate |
| coMD | Collective Molecular Dynamics |
| cMD | Conventional Molecular Dynamics |
| C | Carboxyl Terminus |
| CaATPase | Calcium ATPase |
| CAP | Catabolite (Gene) Activator Protein |
| CL | Charged Linker Region |
| CRP | cAMP Receptor Protein |
| c-Src | Cellular Src Kinase |
| CTD | C-terminal Domain |
| CV | Collective Variable |
| DFS | Dynamic Force Spectroscopy |

| | |
|---------|--|
| DMSO | Dimethyl-Sulfoxide |
| DNA | Deoxyribonucleic Acid |
| eV | Eigenvalue |
| ENM | Elastic Network Model |
| fs | Femtosecond |
| FEL | Free Energy Landscape |
| FimD | Fimbrial Usher Protein |
| FRET | Förster/Fluorescence Resonance Energy Transfer |
| GDM | Geldanamycin |
| GNM | Gaussian Network Model |
| Hsp90 | Heat Shock Protein 90 |
| IF | Inward-Facing State |
| kDa | Kilodalton |
| LD | Langevin Dynamics |
| MalFGK2 | Maltose Transporter Protein |
| MBP | Maltose Binding Protein |
| MC | Monte Carlo |
| Md | Middle Domain |
| MD | Molecular Dynamics |
| ms | Millisecond |
| MSF | Mean Square Fluctuations |
| μ s | Microsecond |
| nN | Nanonewton |
| ns | Nanosecond |
| N | Amino Terminus |
| NBD | Nucleotide Binding Domain |
| NMA | Normal Mode Analysis |
| NMR | Nuclear Magnetic Resonance |
| NTD | N-terminal Domain |
| OF | Outward-Facing State |
| pN | Piconewton |

| | |
|--------|--|
| ps | Picosecond |
| PBS | Phosphate Buffered Saline |
| PCA | Principal Component Analysis |
| PEG | Polyethylene Glycol |
| PDB | Protein Data Bank |
| PIPES | Piperazine-N,N'-bis(2-ethanesulfonic Acid) |
| REMD | Replica-Exchange Molecular Dynamics |
| RMSD | Root-Mean-Square Deviation |
| RNAp | RNA Polymerase |
| Sander | Simulated Annealing with NMR-Derived Energy Restraints |
| SAXS | Small-Angle X-ray Scattering |
| TMD | Transmembrane Domain |
| TMD | Targeted Molecular Dynamics |
| VMD | Visual Molecular Dynamics |
| wt | Wild Type |

1. INTRODUCTION

1.1. Protein Dynamics and Allostery

Proteins undergo small/local vibrations (fast, ps) and large/global structural rearrangements (slow, ns) in order to function in the living cell, where the understanding of functioning mechanism is central in differentiating healthy and diseased cells. Proteins exist as an ensemble of conformational states, interconverting between these states intrinsically in their isolated states, in the presence of their effector molecules or partner molecules in the environment. Structural information on conformational states from the ensemble of structures could often be available through the experimental methods, X-ray crystallography, Nuclear Magnetic Resonance (NMR), and now more popularly cryo-electron microscopy or small-angle X-ray scattering (SAXS), individually or as a combination of suitable methods. However, the quasi-static structural information reveals much less about time-dependent and even equilibrium dynamic behavior. Another challenge is the experimental characterization of conformational transition paths connecting meta-stable states at the molecular level due to the limitations of each experimental technique, highlighting the importance of developing alternative/complementary computational approaches making it possible to sample the conformational (transition) space at the atomic level [1].

While molecular dynamics (MD)-based computational techniques potentially provide high-resolution time-dependent structural information, conventional atomistic simulations reach at most millisecond timescales with the current computational power [2]. It is thus possible to mostly overcome only small barriers with small structural changes but not large barriers to reach physiologically relevant timescales of milliseconds to seconds. Studying protein motions in microsecond/millisecond timescale is indispensable to investigate slower long-range concerted motions, which can be biologically important to carry signals over long distances potentially underlying allosteric mechanisms. Further elaboration of what constitutes these long-range correlations also requires slow-time dynamics to understand whether cumulative local couplings or other more complex

network of interactions cause them.

Thus, enhanced sampling simulation techniques are of significant importance to be able to explore the long time behavior with large conformational transitions and transition pathways. Obtaining the details of structural transitions between (known) conformational states is the way to understand and engineer solutions to disorders caused by any disease causing/transition-modifying mutation or any defect in functional regulation mechanisms. This need has also led to the development of hybrid molecular simulation approaches integrated with normal mode analysis/elastic network model (NMA/ENM)-based methods [3–6] to investigate functional transitions along low-frequency global modes without direct time-scale information on the transition [1, 7, 8]. Additionally, diverse approaches that utilize experimental data to guide computational explorations of the conformational space can be seen in recent studies, like NMR-derived structural restraint guided molecular simulations [9] or a conformation generation/scoring approach estimating the flexibility of high-resolution molecular models based on coarse-grained NMA, improving the model’s agreement with experimental SAXS profile [10].

Allosteric regulation of proteins by mechanical forces acting as perturbations for a change in the protein’s conformation and/or dynamics are increasingly studied in addition to more familiar perturbations like environmental changes, ligand binding or post-translational modifications (PTM) [11]. In that, force-based experiments, such as atomic force microscopy, coupled to computational techniques have been of interest to study forces acting on the molecules.

Whether it is a chemical or mechanical perturbation, adaptation to the deformation caused by perturbation results in local and/or global structural transitions leading to biological signaling. The understanding of the underlying phenomena of biological signal flow has clinical implications considering that the signals are mostly transferred from an allosteric site far from the active site involving direct interaction with a perturbation. Allosteric sites are thus the most effective drug targets due to the decreased selectivity of the active site drugs (orthosteric drugs) as a result of structural similarity

of the active sites across multiple proteins with the same function or origin. Also, targeting active sites decreases the effectiveness of the drug due to escape mutations as a mechanism to drug resistance [12, 13].

1.2. Protein Dynamics by Experimental Methods

Experimental techniques are advancing each day in their use to characterize macromolecular structure and dynamics. The limitations are in their description of equilibrium structure and dynamics due to the ensemble-averaging problem, as well as the problem of timescales probed in the wet laboratory. Experimentally measured quantities are Boltzmann-weighted averages over substates populated in the conformational ensemble at equilibrium. In the recent years, single-molecule techniques, including Förster/fluorescence resonance energy transfer (FRET)-based fluorescence microscopies, optical tweezers and atomic force microscopy, have made great advances, owing to their success in bypassing the ensemble averaging problem. Indeed, with these single-molecule techniques, it became possible to measure the properties of the weakly-populated/rare conformational states, which would be averaged in an ensemble dominated by the well-populated states.

Conformational transitions between different structural substates, and the well-populated intermediate conformations along the transition could be resolved by recent experimental techniques, with limited information about the short-lived transition states due to timescales involved in the transition. In order to fully elaborate the equilibrium dynamics of a macromolecule, these state-of-the-art single-molecule techniques have to be combined with their complementing computational approaches. These integrated experimental-computational methods hold promise to contribute to the efforts in closing the spatial and temporal gaps between the macroscopic experimental data and the microscopic simulation data (with adequately sampled conformational space) [1, 14].

1.2.1. Atomic Force Microscopy

Atomic force microscopy (AFM) [15] has become an important scientific tool to study macromolecules at a single-molecule level. AFM is a unique tool with force measurement, high resolution topographic imaging and molecule/cell level manipulation capabilities. It is mainly used as a single-molecule force-spectroscopy method to study intramolecular and intermolecular forces controlling protein-ligand/protein-protein interactions, enabling the measurement of 10-100 piconewton (pN) force levels. It was originally emerged as a visualization tool [16] and its high-speed form (HS-AFM) is directly used to obtain high-resolution images of molecules and recently live cells under physiological conditions with minimal intervention [17]. HS-AFM is referred to as the only nanotool that allows the acquisition of images and measure mechanical properties at the same time [18].

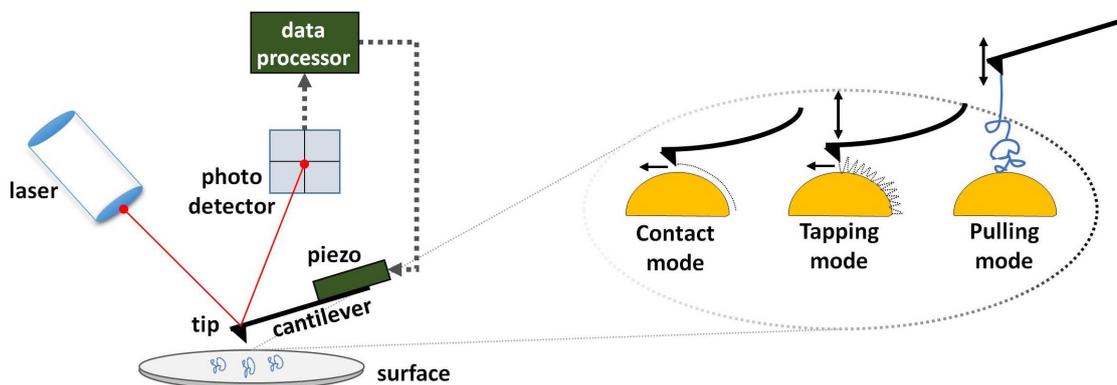


Figure 1.1. Schematic representation of atomic force microscopy and its operating modes.

An important potential application of AFM is in the study of signal transfer mechanisms initiated by a perturbation applied to the system. When a force is applied to a biological system, it results in new intramolecular/intermolecular interactions, finally leading to the functioning of the molecule. Adaptation to the deformation caused by the applied force results in conformational transitions, resulting in biological signaling. In other words, the factors that cause conformational change appear as a perturbation in the intrinsic force field of the molecule. Before AFM, it was not possible

to experimentally characterize the forces at the molecular level. Thus, by measuring the molecular forces it could be possible to understand how these signals emerge from the distribution of these forces.

The standard AFM setup consists of a movable piezo sample stage with the sample deposited on a mica on top of it; a flexible cantilever with a tip at the end and a photo detector to detect the laser beam reflected off the cantilever (See Figure 1.1). The common modes of operation of AFM are contact, tapping and pulling modes. The contact and tapping (dynamic/intermittent contact) modes are grouped as scanning modes, which are used for topographic imaging of the sample surface. In the tapping mode the soft biological sample is less damaged/scratched by the decreased contact time of the tip tapping on it and the image resolution is higher. In the pulling mode (also called as force spectroscopy/modulation), the deflection of the cantilever (which may have attached functionalized molecules) is used to measure the force being exerted on the molecule/surface to record a force-distance curve (See Figure 4.4 in Chapter 4 as an example) [14]. These force measurements provide detailed information about the unfolding/unbinding forces and folding/binding energy landscapes of the molecules [19–21] (See Section 4.2.3 of Chapter 4 for detailed theory).

Here in this thesis, AFM experiments were performed for the study of the effect of temperature on Hsp90-Geldanamycin binding/unbinding behavior. The computational work is carried out using the newly developed methodology ANM-LD (elaborated in Section 2.3 and Chapter 3) along with the relevant AFM experiments. The results are presented in Chapter 4.

1.3. Protein Dynamics by Computational Methods

1.3.1. Normal Mode Analysis and Elastic Network Models

The seemingly random fluctuations that proteins exhibit in their native states actually harbor major contributions from the cooperative/collective/global motions that they undergo during their functional transitions. The most collective motions are

usually the ones at the low frequency end of the mode spectrum. Principal component analysis (PCA) based approaches such as normal mode analysis (NMA) were proven to be useful in extracting information about these collective motions from the equilibrium topology of the protein in question, expressing the dynamics in terms of a superposition of normal mode coordinates [22]. The shape of the modes were determined to be coded by the inter-residue contact network constituting the protein structure, therefore coarse-graining the model were found to produce the same results, owing to the insensitivity of the global modes to the structural and energetic details. Instead of a detailed force-field, using a simple single-parameter harmonic potential [23] was a pioneering move and inspired the development of elastic network models (ENM), such as Gaussian network model (GNM) [24,25] and anisotropic network model (ANM) [26,27].

Global modes elucidated by these methods show functional significance and compares well with the conformational changes (upon ligand binding) observed in experimentally determined structures [28]. Computationally efficient, robust models like ENMs that exploit the intrinsic dynamics of proteins, combined with experimental and/or computational approaches, are of great value in building up hybrid approaches for bridging experimental and computational studies and extending application range [29,30].

1.3.2. Molecular Simulations

1.3.2.1. Molecular (and Langevin) Dynamics. Molecular Dynamics (MD) has become a universal simulation protocol for investigating the relationship between structure, function and dynamics since its first successful application to proteins, which was reported in the 1970s [31]. With continuous efforts to improve the potential energy function, simulation protocol and algorithm used, MD has been advancing continuously. There are still remaining issues that must be overcome to make use of its full potential. While most of the progress in MD is directly related to the vast advances in computing technology, developments in theory and methodology also contribute significantly [32,33]. Today microsecond and even millisecond MD simulations of moderate biomolecular systems are feasible in several days/weeks of computing [2,34–36].

However conventional MD simulations are not feasible for many systems of interest because of their large size and/or timescales of large amplitude motions/structural arrangements. In order to reach the physiologically relevant timescales of milliseconds to seconds with enhanced sampling of the thermally accessible conformational space, a vast range of hybrid approaches have been developed and applied to molecular systems, using experimental and computational methods in a complementary fashion.

1.3.3. Advanced/Hybrid Molecular Simulation Methodologies

MD simulations have shown to be successful in extracting useful information on the structure and dynamics of macromolecules with recent advances on force-fields and computer technology. Biologically important phenomena nevertheless would only emerge by simulating slow-time dynamics that is limited by the size of the system and the expected time of the phenomenon to be observed [1, 33]. Types of protein motions and corresponding timescales are shown on a sketch of a high-dimensional energy landscape (See Figure 1.2), which describes the complex dynamics of a protein [37, 38]. In order to overcome the time-scale limitation and to simulate for much longer times with enhanced sampling efficiency, different approaches are used such as restraining the potential energy function as in targeted MD (TMD) or/and coarse-graining the molecular representation of the system of question.

Selected modifications to molecular dynamics method to enhance sampling will be presented here in four broad classes; modified potentials, modified sampling/dynamics, more complex modifications to the algorithms like global reformulations [32, 33], and a hybrid class where the methods are combined without intervening with the algorithms, with some overlap between classes.

Modifying the potential energy function explicitly is the simplest approach for altering the sampling of conformational space. The basic principle in restrained potentials is to reduce the residence time of the simulated system in a local energy minimum by lowering the tendency of potential energy wells acting as conformational traps and forcing the system to visit other available local minima. Methods that modify the

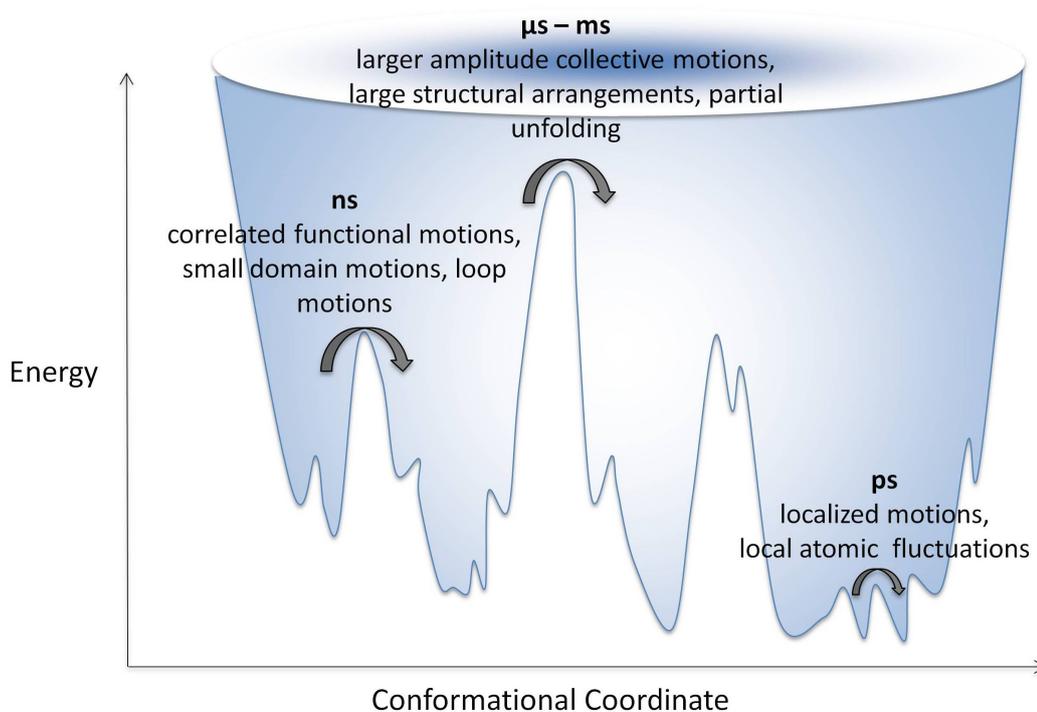


Figure 1.2. Characteristic time scales for protein motions.

potential energy surface include targeted MD [39], restricted perturbation TMD (RP-TMD) [40, 41], multiple-basin model [42], “switching Gō model” [43], umbrella sampling [44], conformational flooding [45, 46], hyperdynamics [47] and accelerated MD (AMD) [48], adaptive AMD [49]. In the following paragraphs some of them will be briefly explained.

In targeted MD (TMD), an artificial restraint potential term is added to the potential-energy function to force the system to evolve from the initial state toward a target state in a specified number of steps. However, the actual trajectory obtained using this restraint may lead to irreversible pathways. The generated pathway is not necessarily the lowest energy one and may show unphysical transition dynamics. TMD can thus be used to explore conformational space in a preliminary manner and could be used complementarily with other algorithms to improve the transition pathways. A successful application of TMD gave insights about the gating mechanism of *Escherichia coli* mechanosensitive channel, suggesting a four-step transition process that was not predicted by the previous studies [32, 50].

Another method is the multiple-basin model [42], which is developed to describe large conformational transitions, even for large systems, with known endpoints of the conformational change. Considering only a single dominant minimum, small amplitude fluctuations are well approximated by quasi-harmonic dynamics. However, large amplitude motions/conformational transitions go beyond the quasi-harmonic approximations, requiring multiple basins to be taken into account. The multi-basin model is developed to simulate very large-scale motions, such as transitions between ligand-bound (2)/ unbound states (1) and applied recently to study allostery by MD [42, 51]. In these studies they explicitly build up an energy landscape encoding multiple near degenerate (similar energies) basins. First, two independent single-basin potentials are constructed as $V(\mathbf{R}|\mathbf{R}_1)$ and $V(\mathbf{R}|\mathbf{R}_2) + \Delta V$. ΔV modulates the relative stability. The coupling constant Δ between the two potentials modifies the energy barrier directly and smoothly connects the basins to make a double-well energy landscape represented by a smoothed double basin potential V_{MB} . The smoothing potential (V_{MB}) is used as the eigenvalue of the characteristic equation in analogy with the electron transfer in quantum mechanics, (c_1, c_2) being the eigenvectors:

$$\begin{pmatrix} V(\mathbf{R}|\mathbf{R}_1) & \Delta \\ \Delta & V(\mathbf{R}|\mathbf{R}_2) + \Delta V \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = V_{MB} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \quad (1.1)$$

Solving this characteristic equation (Equation 1.1) leads to a secular equation (Equation 1.2) with the lower energy solution represented by (Equation 1.3):

$$\begin{vmatrix} V(\mathbf{R}|\mathbf{R}_1) - V_{MB} & \Delta \\ \Delta & V(\mathbf{R}|\mathbf{R}_2) + \Delta V - V_{MB} \end{vmatrix} = 0 \quad (1.2)$$

$$V_{MB} = \frac{V(\mathbf{R}|\mathbf{R}_1) + V(\mathbf{R}|\mathbf{R}_2) + \Delta V}{2} - \sqrt{\left(\frac{V(\mathbf{R}|\mathbf{R}_1) - V(\mathbf{R}|\mathbf{R}_2) - \Delta V}{2}\right)^2 + \Delta^2} \quad (1.3)$$

where \mathbf{R} collectively represents the coordinates of the protein structure and \mathbf{R}_1 and \mathbf{R}_2 correspond to the coordinates of the two reference structures supplied by X-ray crys-

tallography. V_{MB} is continuous and differentiable, which can directly be used for MD simulations of very large systems. This method is similar to the plastic network model suggested by [52] in which individual basins are approximated by the Tirion harmonic model [23] and are then smoothly connected by the secular equation formulation.

For large-amplitude conformational motions of biomolecular machines, a computational framework -“switching Gō model”- was formulated by Koga and Takada [43]. Gō models accounting for both small fluctuations around the native basin and large fluctuations that involve local unfolding are able to describe large-amplitude conformational dynamics. With available experimental data and simulation results, the rotary motion of F1-ATPase was obtained. The results by this folding-based model opened an avenue of simulating dynamics of large biomolecular complexes undergoing large-amplitude conformational changes on the millisecond timescales.

In umbrella sampling (US), an experimentally based compensating function, known as an umbrella potential, is added to the potential-energy function to bias the sampling. Construction of the umbrella function requires a priori knowledge of the conformational states. In a recent study [53], they have guided the US application focusing on specific regions of phase space by using a restraining potential formulated using the distances by FRET experiments. They have compared the efficiency of the guided versus unguided US and observed improvement with respect to computational efficiency and accuracy in sampling intermediate structures.

The methods described so far require some prior knowledge about the system, either the geometry of the product or target state. In Grubmüller’s conformation flooding [45, 46, 54] collective motions (vibrational eigenvectors) of the system are extracted from a standard MD trajectory using the covariance matrix, and then the system is destabilised along these collective coordinates in order to drive the system from one conformational state to another by adding a non negative bias potential (the flooding potential, V_{fl}) along a pre-defined collective vibrational coordinate. The flooding energy E_{fl} determines the extent of flooding. Here, solely the energy landscape within the minimum -that they are interested in- is modified, but not the barriers and all

the other minima. This method thus induces unbiased transitions, which would be observed even without the flooding potential, but at a much longer timescale.

In a similar but alternative method, accelerated MD (AMD) [48], a non-negative bias potential is added to the potential energy to boost the potential and to allow escape from minima. A reference energy called the boost energy E is defined. If the potential of the system lies above E , the potential is unaltered, if the potential lies below E , a continuous non-negative bias potential (ΔV) is added to form a modified potential energy surface (Equation 1.4). This results in a flattening of the potential energy surface. In low free energy regions of the conformational space a large bias potential and in high energy regions, a small bias potential is added to the real potential $V(\mathbf{r})$. New forces, $V^*(\mathbf{r})$, are calculated on the modified potential using the term α as a factor that determines inversely the strength with which the boost is applied:

$$V^*(\mathbf{r}) = \begin{cases} V(\mathbf{r}), & V(\mathbf{r}) \geq E \\ V(\mathbf{r}) + \Delta V(\mathbf{r}), & V(\mathbf{r}) < E \end{cases} \quad \text{where } \Delta V(\mathbf{r}) = \frac{(E - V(\mathbf{r}))^2}{\alpha + (E - V(\mathbf{r}))} \quad (1.4)$$

Voter's hyperdynamics method (hyper-MD) [47] is a transition state theory (TST)-based method for extending the MD time scale. The method does not require advanced knowledge of the end-states of the system or the transition states separating them. The potential energy surface $V(\mathbf{r})$ is modified by adding a bias potential $\Delta V(\mathbf{r})$ to the true potential. The bias raises the potential surfaces near the minima. The surfaces near the barrier or saddle points are left unaffected. The bias potential is constructed from the smallest eigenvalue of the Hessian matrix, $\mathbf{H} \{H_{ij} \equiv [\partial^2 V(\mathbf{r}) / \partial x_i \partial x_j]\}$. Since the Hessian matrix is diagonalized at each step, this method comes with a computational cost when used for large systems.

Second class of modified MD methods include those that modify the simulation protocol to alter the sampling space, such as locally enhanced sampling [55], replica exchange MD [56] and self-guided MD methods.

Locally enhanced sampling (LES) technique relies on simulating multiple, noninteracting simulations that differ only locally. The method can be used for calculating free energy differences and only applicable to systems undergoing local modifications, like point mutations [55]. Similarly, replica exchange MD (REMD) utilizes a series of simultaneous and noninteracting simulations/copies (replicas) at different temperatures to accelerate crossings of energy barriers. The temperatures of these simulations may be exchanged (swapped) with a Monte Carlo-like transition probability at particular intervals to maintain each temperature’s equilibrium ensemble distribution in the canonical ensemble. As a result, those systems that are making better progress are retained [56]. REMD has been widely applied for smaller biomolecules’ folding/unfolding simulations, because it has been found extremely computationally expensive when applied to large complexes [33]. REMD has been increasing in popularity and many REMD combinations have been reported recently in the literature [32]. Some examples are, REMD combined with a coarse-grained implicit solvent model to determine equilibrium ensembles of proteins [57] and REMD combined with transition interface sampling (TIS) [58] to compute the free energy profile when high barriers are involved. These combined methodologies are promising ways to further enhance sampling.

Also, it is possible to enhance the sampling of “slow” degrees of freedom at the expense of “fast” degrees of freedom. Integration of the equation of motion may be performed faster by separating the faster components of the dynamic propagator from the slower components or the high-frequency motions may be completely reduced or eliminated. The stability and resonance limitations of MD integrators mainly depend on the fast modes and to the coupling between them. Therefore, many integrators are developed in order to improve both the stability and resonance problems, like multiple-timestep integrators combined with stochastic dynamics. Stochastic methods like Langevin dynamics (LD) and Brownian dynamics (BD) are promising since coarse-grained characteristics of these approaches allow the study of large systems on long time scales, where it is also easier to prove ergodicity.

Based on this principle, many algorithms have been developed including the widely (almost in every conventional MD simulations) used SHAKE algorithm [59],

normal mode multiple time stepping Langevin dynamics integrator (NML) [60], driven molecular dynamics (DMD) [61], amplified-collective-motion (ACM) method [62] and steered/interactive MD (SMD/IMD) [63–65].

In normal mode multiple time stepping Langevin dynamics integrator (NML) approach, the kinetics/thermodynamics of the molecule is approximated by separating the dynamics into slow, functionally relevant modes (essential subspace), and fast modes (near-constraints coordinates), based on a normal mode decomposition of the dynamical space. The reduced model is based on the subspace of low frequency dynamical modes. To dampen resonance effects, Langevin dynamics is used instead of Newtonian. The low frequency modes are propagated according to a Langevin integrator. The high frequency modes are dampened near their energy minimum and propagated using a less costly accelerated BD propagator to keep the fast vibrations around their equilibrium values [60].

The amplified collective motion (ACM) MD method has been utilized for coupling low-frequency modes (essential subspace) derived from the ANM to a higher temperature by using a weak coupling method. Local interactions and high frequency motions are kept at normal temperature. The method attempts to excite motions along the essential subspace to accommodate the large amplitude motions, resulting in enhanced sampling [62].

The methods of interest in the third category, global/algorithmic reformulations, are the methods that aim to explore the free energy landscape (FEL), like metadynamics [66] and milestoning [67]. Metadynamics involve exploring the FEL by following non-Markovian dynamics determined by the reformulation of the equations of motion in terms of a few collective variables. Milestoning is a procedure that coarse-grains the temporal and spatial description of the system by reducing the dynamics into a succession of independent transition events between intermediates (the milestones), which are hypersurfaces in the conformational space of the system. The procedure computes the local kinetic information to describe these transitions via short MD runs between the milestones. The systems of interest are complicated processes involving barrier

crossing events or long diffusive transitions between known states. In the study by Vanden-Eijnden *et al.* [67], the validity of the assumptions (statistical independence of transition events between milestones and time lags between the transitions) underlying milestoning is analyzed.

For the fourth class of methods, the hybrid simulation protocols combining coarse-grained mechanistic models (NMA, ANM, etc.) with simulation approaches such as MD and Monte-Carlo (MC) are presented. In this class, the functional motions along the low-frequency modes are utilized to guide the simulations towards a target structure or in a nontargeted manner only along the direction of slow modes. The methods are combined without intervening with the algorithms. Many approaches have been developed including ANM-restrained MD [3], adaptive ANM (aANM) [68], collective molecular dynamics (coMD) [4], ANM-MC [6, 69] and ANMPathway [5].

ANM-restrained MD protocol uses an ensemble of ANM modes as harmonic restraints in MD runs in an iterative scheme. ANM modes calculated are used to identify the directions along which steered MD forces will act, in other words, the system is biased along its most robust modes of motion as calculated by the ANM analysis. The underlying assumption in this protocol is that ANM-derived restraints drive the sampling of the molecule toward a direction that would otherwise be explored in much longer timescales [3].

Adaptive ANM is an iterative procedure recruiting a small subset of global modes (selected according to the eigenvalue distribution and collectivity) at each cycle, to create a series of intermediate conformers between two end structures until a predetermined root-mean-square-deviation (RMSD) value between the two pathways is attained [68]. Built upon aANM, collective MD (coMD) follows a similar route, differing in the selection of the modes with a Metropolis algorithm among a complete set of ANM modes, allowing the system to explore also other low energy paths that diverge from the shortest path [4].

In ANM-MC, the collective deformations obtained from ANM are incorporated into a Monte Carlo simulation [6, 69]. At each iteration, the structure is deformed along the slow mode that overlaps with the conformational transition direction to the target, followed by a brief MC simulation performed to minimize the conformational energy. This approach is improved to predict closed states using the initial structure and a predicted radius of gyration (Rg) value of the closed/target state [6].

In ANMPathway, the protein with two end states are represented by a two-state potential combining two ENM representations, where a cusp hypersurface is assumed to exist with equal energies for both of the ENMs. The minimum energy structure on the cusp hypersurface is set as the pseudo-transition state and it is used as a starting point for an iterative procedure to connect this point to the end-states by two separate steepest descent minimization steps. This approach yields physically meaningful pathways (between known experimental end-states) and intermediate states that could be tested experimentally [5].

In depth understanding of the above-discussed methodologies is a first step on the way to the ultimate goal of constructing the Free Energy Landscape (FEL). Energy landscape theory provides a common language for experimentalists and theorists to describe structure, dynamics and function of complex biomolecular systems. FELs of complex systems have many local minima, lowest one being the global minimum (the deepest basin). It remains still as a challenge of this century to develop methodologies to accurately sample functional landscapes of widely differing biomolecular complexes using explicit/implicit solvent molecular simulations in a biologically relevant time frame and deduce reliable functional information. Creating a representation of a protein's FEL allows the subsequent study of specific questions like allostery, large conformational changes, coupled motions or pathways of folding.

While building a free energy profile between two conformational states of a protein, it is necessary to select appropriate collective coordinates (order parameter) characterizing functional motions. Lowest frequency normal modes, or a linear combination of a few low frequency modes, could be suitable for this purpose, ideally based on an

experimental/computational/theoretical evidence to tell whether a particular collective variable is suitable. A collective variable might be suitable to describe the conformational transition of interest, if it changes slowly and in a correlated manner with the progression of the transition [70].

It is with a very low probability to observe large conformational changes between two stable states of a protein or to obtain intermediate structures by solely using experiments or conventional simulation methods, which makes hybrid simulation methodologies employing enhanced sampling techniques popular in recent literature [1,7,8,71]. Exploring the dynamics of large proteins and large conformational changes by hybrid methods integrating coarse-grained ENMs and all-atom molecular simulations, also with a possibility to incorporate experimental data, are of great value and there is still need for the development of new computationally efficient approaches that may produce biologically relevant reactive transition pathways for large systems.

We propose a simple and efficient method, belonging to the fourth class of methods previously defined, called anisotropic network model-Langevin Dynamics (ANM-LD), and apply the method to understand conformational transitions of several important globular and membrane proteins. Our methodology is a hybrid platform combining the intrinsic dynamics information obtained from ANM with all-atom LD simulations to create pseudo-equilibrium transition pathways between two distinct conformational states and sample the large conformational changes in between. ANM-LD protocol enables us to observe the cooperative changes underlying the functional transitions of proteins, control the generation of the alternative pathways by the restriction of certain modes and assess differences in the pathways further by GNM.

1.4. Specific Aims and Outline of the Thesis

The purpose of this study is to build upon and enhance currently available computational techniques to create conformational transition pathways connecting end states and to simulate large-scale conformational changes of small/large biomolecular complex structures, biomolecular machines. The association of functional large conformational

changes with the structure's intrinsic elastic modes has been demonstrated in several previous studies as outlined in the previous section. Nevertheless, the methodology developed here is unique in its entirety as a handy platform making it possible to create parallel transition pathways using the intrinsic dynamics information obtained from ANM by taking advantage of all-atom molecular simulations, controlled by many parameters that could easily be utilized to change the system conditions or incorporating any simulation methodology to the protocol by changing only a few parameters.

In this thesis, a practical computationally efficient hybrid approach, ANM-LD, has been developed, where the dominant low-frequency modes of fluctuations of a given biomolecular structure are utilized to drive all atom Langevin dynamics simulations. A detailed assessment of the method, in terms of creating physically meaningful pathways and approaching the final state, for a set of proteins was carried out and presented for the exemplary cases adenylylate kinase, and maltose transporter in Chapter 3.

ANM-LD method was also applied to heat shock protein 90 (Hsp90) in apo, adenosine triphosphate (ATP)-bound and geldanamycin (GDM)-bound states, the latter combined with AFM experiments, to study the effect of temperature on Hsp90-GDM binding/unbinding behavior, where this forms Chapter 4 of the thesis.

Finally, conventional molecular dynamics simulations, elastic network models and ANM-LD simulations are utilized to investigate the dynamics of CRP. This work is presented in Chapter 5.

The remaining chapters are Chapter 1, in which the available/relevant computational methods are summarized for protein dynamics and conformational transitions between well-defined states; Chapter 2, in which the computational methodologies and the dataset used in this thesis are introduced. Finally, the conclusions and recommendations are given in Chapter 6.

2. MATERIALS AND METHODS

2.1. Elastic Network Models

2.1.1. Gaussian Network Model (GNM)

The Gaussian Network Model (GNM) [24,25] is a one dimensional elastic network model (ENM). In GNM the residues represented by their backbone alpha carbon atoms (αC) are accepted to interact if they are within a certain cut-off radius (R_{cut} , 7Å, 10 Å etc.). αC residues denote nodes in the network and they are assumed to undergo Gaussian fluctuations about their equilibrium positions. The interactions between all residue pairs in the network are represented by the Kirchhoff (or connectivity) matrix $\mathbf{\Gamma}$. The correlation between the residue fluctuations of $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ is given in Equation 2.1,

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = \frac{3k_B T}{\gamma_f} [\mathbf{\Gamma}^{-1}]_{ij} \quad (2.1)$$

where k_B is the Boltzmann constant, T is the absolute temperature in degrees Kelvin and γ_f is the force constant of the elastic potential function. The correlation matrix in Equation 2.1 can also be expressed as the linear superimposition of N-1 eigenmodes as

$$[\mathbf{\Gamma}^{-1}]_{ij} = [\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T]_{ij} = [\lambda_k^{-1}\mathbf{u}_k\mathbf{u}_k^T]_{ij} \quad (2.2)$$

\mathbf{U} is the matrix of eigenvectors \mathbf{u}_k , where k refers to the k th eigenvector that gives the displacements of the residues along the k th mode. The k th eigenvalue, λ_k , is proportional to the frequency of the k th mode of motion. The normalized cross-correlation values (C) of residue fluctuations vary in the range $[-1, 1]$, referring to the limits of the negatively correlated ($C_{ij} < 0$) and positively correlated ($C_{ij} > 0$) pairs in their fluctuations, respectively.

2.1.2. Anisotropic Network Model (ANM)

The anisotropic network model (ANM) [26, 27] is an extension of the Gaussian Network Model (GNM) [24, 25]. In ANM, the residue fluctuations are anisotropic that incorporates the X, Y, and Z components of the position vector, \mathbf{R}_i , independently. The overall potential of the system is a sum of harmonic potentials as given in Equation 2.3,

$$V_{ANM} = \frac{1}{2} \sum_{ij} \gamma_{f,ij} (R_{ij} - R_{ij}^0) \quad (2.3)$$

where γ_f is the force constant. R_{ij}^0 and R_{ij} are the equilibrium and instantaneous separation distances between residues i and j , respectively. ANM involves the inversion of Hessian matrix (\mathbf{H}), which is a $3N \times 3N$ symmetric matrix and composed of $N \times N$ super elements H_{ij} each of size 3×3 , given by the second derivatives of V with respect to \mathbf{R}_i and \mathbf{R}_j of α C-atoms of respective i th and j th residues. The correlation between the fluctuations of the residues i and j from their equilibrium positions, $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle$, are decomposed into $3N - 6$ modes as contributed by the movement along mode k is given in terms of eigenvalue (λ_k) of \mathbf{H} and the corresponding eigenvector (\mathbf{u}_k):

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \frac{3k_B T}{\gamma_f} \text{tr} [\mathbf{H}^{-1}] = \frac{3k_B T}{\gamma_f} \sum_{k=1}^N \text{tr} [\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ij} \quad (2.4)$$

Here, the knowledge of fluctuation vectors lets us view the alternative conformational space that could possibly be sampled by the individual (or combination of) modes, simply by adding the fluctuation vectors $\pm \Delta \mathbf{R}_i$ to the equilibrium position vectors (\mathbf{R}_i^0) in the respective modes.

2.2. Molecular Simulations

2.2.1. Molecular Dynamics

Molecular dynamics (MD) simulations involve the iterative numerical calculation of instantaneous forces present in a molecular system and start with the knowledge of the energy of the system as a function of the atomic coordinates. The forces acting on the atoms of the system, which are related to the first derivatives of the potential with respect to the atom positions, can be used to calculate the dynamic behavior of the system by solving Newton's equations of motion for the atoms as a function of time. The classical Newtonian equations of motion (conventional MD) are adequate for the biomolecules in general, except for some cases where quantum corrections should be considered [72].

The potential energy function or force field, V , is a function of the Cartesian coordinate set for N atoms, (r_1, r_2, \dots, r_N) , and is composed of bonding terms representing bond lengths, bond angles, and torsional angles, and non-bonding terms consisting of van der Waals and electrostatic interactions as seen in Equation 2.5.

$$\begin{aligned}
 V(r_1, r_2, \dots, r_N) = & \sum_{\text{bonds}} \frac{1}{2} k_l (l_i - l_{eq})^2 + \sum_{\text{bond angles}} \frac{1}{2} k_\theta (\theta_i - \theta_{eq})^2 \\
 & + \sum_{\text{torsions}} \frac{1}{2} V_n [1 + \cos(n\phi_i - \gamma_f)] \\
 & + \sum_{i=1}^N \sum_{j=i+1}^N \left\{ 4\epsilon_{ij} \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^6 \right] \right\} \\
 & \qquad \qquad \qquad \text{Lennard Jones} \\
 & + \sum_{i=1}^N \sum_{j=i+1}^N \left(\frac{q_i q_j}{\epsilon r_{ij}} \right) \\
 & \qquad \qquad \qquad \text{electrostatic}
 \end{aligned} \tag{2.5}$$

The first term in Equation 2.5 represents instantaneous displacements of bond length l_i (length of the bond between atoms $i - 1$ and i) from the equilibrium bond

length, l_{eq} , by a Hooke's law (harmonic) potential. Harmonic potential is the first approximation to the energy of a bond as a function of its length. The bond force constant k_l determines the flexibility of the bond. The second term is the energy associated with alteration of bond angles, is also represented by a harmonic potential. k_θ is the force constant for the bond angles. θ_i and θ_{eq} are bond angle and equilibrium bond angle, respectively. The third term is the torsion angle potential function for rotations about bonds. It is assumed to be periodic and modeled by a cosine or sum over cosine functions. ϕ_i is the dihedral angle. V_n is the corresponding force constant; the phase angle γ_f takes values of either 0° or 180° .

The nonbonded part of the potential is the last two terms, represented by van der Waals (A_{ij}) term, London dispersion term (B_{ij}) and interactions between partial atomic charges (q_i and q_j). ϵ is the dielectric constant which describes the medium that is not explicitly represented. In a typical explicit solvent solvated environment ϵ is taken as 1.0. The nonbonded terms are calculated for all atom pairs that are not bonded or separated by more than three bonds [72].

The internal coordinates used in the potential function; bond lengths, l_i (Equation 2.6), bond angles, θ_i (Equation 2.7), dihedral angles, ϕ_i (Equation 2.8), and interparticle distances, r_{ij} (Equation 2.9) are calculated from the Cartesian atomic coordinates.

$$l_i = |r_i - r_{i-1}|; 1 \leq i \leq n \quad (2.6)$$

$$\cos \theta_i = \frac{\mathbf{l}_i \cdot \mathbf{l}_{i+1}}{l_i l_{i+1}}; 1 \leq i \leq n - 1 \quad (2.7)$$

$$\cos \phi_i = \frac{\mathbf{l}_{i-1} \times \mathbf{l}_i}{|\mathbf{l}_{i-1} \times \mathbf{l}_i|} \cdot \frac{\mathbf{l}_i \times \mathbf{l}_{i+1}}{|\mathbf{l}_i \times \mathbf{l}_{i+1}|}; 2 \leq i \leq n - 1 \quad (2.8)$$

$$r_{ij} = |r_j - r_i| \quad (2.9)$$

Newton's equations of motion gives the most exact and detailed information when solved for the atoms of the system and any surrounding solvent to generate successive conformations of the system [72]. Newton's equation of motion is given by

$$F_i = m_i a_i \quad (2.10)$$

where F_i is the force exerted on particle i , M_i is the mass of particle i and a_i is the acceleration of particle i . When a_i is expressed in terms of the second derivative of position of the particle r_i with respect to time, the equation becomes:

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{M_i} \quad (2.11)$$

The force can also be written as the gradient of the potential energy:

$$F_i = -gradV_i = -\nabla V_i \quad (2.12)$$

When the Equations 2.11 and 2.12 are combined in Equation 2.13, Newton's equation of motion can relate the derivative of the potential to the change in the position of the particle i as a function of time:

$$\frac{d^2 r_i}{dt^2} = -\frac{1}{M_i} \frac{\partial V(r_1, r_2, \dots, r_N)}{\partial r_i} = a_i \quad (2.13)$$

To calculate an MD trajectory the requirements are: The initial positions of the atoms, an initial distribution of velocities and the acceleration. The initial positions of the atoms can be obtained from experimental structures, such as the X-ray crystal structure or NMR spectroscopy solution structure of the protein. The initial distribution of velocities are calculated using the Maxwell-Boltzmann distribution (Equation 2.14), which gives the probability that an atom i has a velocity v_x in the x-direction at

a temperature T , and corrected so that there is no overall momentum (Equation 2.15). Initial distribution of the acceleration is determined by the gradient of the potential energy function (Equation 2.16).

$$p(v_{ix}) = \left(\frac{M_i}{2\pi k_B T} \right)^{1/2} \exp \left[-\frac{1}{2} \frac{M_i v_{ix}^2}{k_B T} \right] \quad (2.14)$$

$$p = \sum_{i=1}^N M_i v_i = 0 \quad (2.15)$$

$$a_i = -\frac{1}{M_i} \frac{\partial V(r_1, r_2, \dots, r_N)}{\partial r_i} \quad (2.16)$$

The temperature is related to the kinetic energy of the system by the relation, where N is the number of atoms in the system, $\langle v_i^2 \rangle$ is the average velocity squared of the i th atom and k_B is the Boltzmann constant:

$$T = \frac{2}{3Nk_B} \sum_{i=1}^N \frac{1}{2} M_i \langle v_i^2 \rangle \quad (2.17)$$

Numerous algorithms exist for integrating the last form of the equation of motion (Equation 2.13). Many of the integration algorithms are finite difference methods in which the integration is partitioned into small time steps, separated by Δt . The continuous potentials describing atomic interactions preclude an analytical solution. In the simple Verlet algorithm [73], the new positions $t + \Delta t$ are determined using the atomic positions and accelerations at time t and the positions from the prior step, $r(t - \Delta t)$. Computational demands of the integration method are important but insignificant compared to the calculation of all the forces acting within the system. Therefore it is useful to limit the number of force calculations required during the simulation.

2.2.2. Langevin Dynamics

In stochastic simulation approaches the influence of particles on the solute is incorporated to the equation of motion through stochastic terms. This way the effects of degrees of freedom that are neglected in Newtonian approach are approximately introduced into the simulations. The basic equation of motion for stochastic dynamics is the Langevin equation, which is

$$M_i \frac{d^2 r_i}{dt^2} = -\frac{\partial V}{\partial r_i} - M_i \gamma \frac{dr_i}{dt} + R_i(t) \quad (2.18)$$

for the motion of an atom, i , which can be rewritten in a more compact form as

$$M_i \ddot{r}_i = -\nabla_i V - \zeta_i \dot{r}_i + R_i \quad (2.19)$$

$\nabla_i V$ is the gradient of the usual potential function V (Equation 2.5) used in Newtonian (classical) MD and accounts for molecular interactions. The other two terms are unique to stochastic dynamics, which are, ζ_i , the friction (damping) coefficient, that is intended to represent the fictional drag force reflecting friction due to solvent that is implicitly simulated, and R_i , the random force representing stochastic collisions between solvent molecules and solute. R_i is a "white noise" vector with zero mean. The stochastic force introduces energy into the system, whereas the friction force removes energy. When $\zeta_i = 0$, LD is equivalent to conventional MD. When $\zeta_i > 0$, the random force acting on the system can assist in barrier-crossing motions, which may improve conformational sampling compared to conventional MD. The temperature of the simulated system is maintained by a relationship between ζ_i and R_i [33, 74, 75].

In the limit of small friction coefficient, ζ_i , the motion described by Langevin equation is termed "inertial", and in the limit of large ζ_i , it is "diffusive" or "Brownian". Brownian dynamics (BD) is a simplified version of Langevin dynamics in the limit of

no average acceleration. Langevin equation under high friction conditions becomes:

$$M_i \frac{d^2 r_i}{dt^2} = 0 = -\frac{\partial V}{\partial r_i} - M_i \gamma \frac{dr_i}{dt} + R_i(t) \quad (2.20)$$

$$\frac{dr_i}{dt} = -\frac{1}{M_i \gamma} \frac{\partial V}{\partial r_i} + \frac{1}{M_i \gamma} R_i(t) \quad (2.21)$$

With high effective viscosity solvents, the damping effect of the solvent will overcome any inertial effects, resulting in a random walk motion of the molecule. BD has other names like “overdamped Langevin dynamics” or “Langevin dynamics without inertia”. Due to extra terms added to force in the equation of motion for LD, particle acceleration is no longer a time derivative of particle velocity as it is defined in Newtonian dynamics. A popular solution to Langevin equation is the BBK (Brünger, Brooks, Karplus) integrator [76, 77]. The accuracy of BBK integration approaches its theoretical limit when using smaller friction coefficient γ and integration time step Δt .

It is common for proteins to be treated as rigid bodies in BD simulations, but also all-atom BD simulations can be performed. The choice of γ is related to the hydrodynamic theory and is a trade-off between “approximating the true dynamics” and “efficient sampling”. Examples of biological processes which are suitable to study by BD include diffusion-controlled reactions, floppy motions of proteins, and ionic diffusion under the influence of an electrostatic field. The choice of MD versus LD versus BD needs to be carefully considered depending upon which contributions are thought to dominate in the motion of interest [33, 74, 75].

2.3. ANM-LD Methodology

The developed hybrid methodology implements the predicted intrinsic functional dynamics by ANM into an all-atom LD simulation protocol. The initial ($\mathbf{I} = \mathbf{I}_0$) and end state ($\mathbf{T} = \mathbf{T}_0$) coordinates of a conformational transition of a protein are needed

to start the procedure. At each iteration (cycle i) ANM is carried out for the current initial structure (\mathbf{I}_i for cycle i) by constructing the Hessian matrix composed of the second derivatives of the ANM potential function given in Equation 2.3 previously. This step is followed by the calculation of the overlap values ($O_{i,j}$, Equation 2.22) between the 3N-dimensional eigenvector corresponding to the j th collective mode (\mathbf{u}_j) with the 3N-dimensional vector representing the difference between the aligned initial and end state coordinates (difference vector, \mathbf{D}_i , Equation 2.23), calculated by their inner products.

$$O_{i,j} = \frac{|\mathbf{u}_j \cdot \mathbf{D}_i|}{|\mathbf{u}_j| \cdot |\mathbf{D}_i|} \quad (2.22)$$

$$\mathbf{D}_i = \mathbf{T} - \mathbf{I}_i \quad (2.23)$$

Overlap measure given here, shows the similarity between the direction of global conformational displacement (here given by \mathbf{D}_i) and the direction of a single ANM mode vector (\mathbf{u}_j). The ANM mode vector (chosen among a predefined number of first modes, *modemax*) that presents the largest overlap value with \mathbf{D}_i in cycle i is given by Equation 2.24 and denoted by $\mathbf{u}_{i,m}$.

$$O_{i,max} = \max(O_{i,j}) = \frac{|\mathbf{u}_{i,m} \cdot \mathbf{D}_i|}{|\mathbf{u}_{i,m}| \cdot |\mathbf{D}_i|} \quad (2.24)$$

The mostly overlapping ANM mode (selected mode) for cycle i is denoted by m_i where necessary. A new conformation is generated for cycle i , by deforming the protein structure along the direction of slow mode m_i . $\mathbf{u}_{i,m}$ is multiplied by the square root of the number of residues (N) and a deformation factor (DF , in Å) that corresponds to the RMSD between the new and the old conformations (Equation 2.25). The same deformation value that is applied to the α C coordinates are used for the other atoms

that belong to that residue, to convert to all-atom coordinates.

$$\mathbf{I}_{i,m} = \mathbf{I}_i \pm \mathbf{u}_{i,m} \sqrt{N}(DF) \quad (2.25)$$

This new conformation generated, $\mathbf{I}_{i,m}$, will be the initial conformation for the next cycle following the energy minimization and LD simulation by a predefined number of steps. This iterative procedure is continued until RMSD to target reaches a plateau, where Equation 2.26 gives RMSD,

$$RMSD = \sqrt{\frac{\sum_{i=1}^N d_i^2}{N}} \quad (2.26)$$

Here, N is the number of atoms over which the RMSD is measured and d_i is the distance between the coordinates of atom i in the two superimposed structures. RMSD is a widely used measure of similarity between two protein structures. The algorithm consists of the following steps:

- (i) Energy minimization of initial (\mathbf{I}) and target (\mathbf{T}) state conformations
- (ii) ANM calculations on the energy-minimized initial conformation (\mathbf{I}) in α C coordinates
- (iii) Find the mostly overlapping ANM mode (m_i) with the difference vector D_i between \mathbf{I} and \mathbf{T} (Equations 2.22-2.24)
- (iv) Generate the perturbed conformer of \mathbf{I} along the mostly overlapping ANM mode m with a predefined deformation factor (DF) value and get a new conformation $\mathbf{I}_{i,m}$ (Equation 2.25). Convert α C coordinates to all-atom coordinates
- (v) All-atom energy minimization of the perturbed structure $\mathbf{I}_{i,m}$ for a predefined number of steps and get conformation $\mathbf{I}_{i,m,\min}$
- (vi) All-atom Langevin Dynamics simulation of $\mathbf{I}_{i,m,\min}$ for a predefined number of steps and get a new conformation $\mathbf{I}_{i,m,\text{sim}}$
- (vii) Calculate the RMSD of $\mathbf{I}_{i,m,\text{sim}}$ from the target conformation \mathbf{T} (Equation 2.26)
- (viii) If the RMSD value has converged go to step (ix) if not go to the next cycle (step

- (ii). The next cycle of ANM-LD simulations starts again with the last recorded frame of the simulation ($\mathbf{I}_{i,m,\text{sim}}$), the conformation from the end of the last cycle. The procedure from (ii) to (vii) is carried out repeatedly until the RMSD to target reaches a plateau
- (ix) Obtain the transition pathway trajectory for the conformational transition ($\mathbf{I} \rightarrow \mathbf{T}$) and the list of the selected (mostly overlapping ANM mode at each cycle) ANM modes.

The outline for ANM-LD protocol that disposes large conformational transitions and multiple transition pathways between given protein functional states of is summarized in Figure 2.1.

Along the simulations, κ_m (the degree of collectivity of the mostly overlapping ANM mode m) is calculated and recorded for each cycle i , for the best three overlapping modes. The progress of κ_m is monitored and its plots are given along with the overlap values in the analysis sections of ANM-LD simulations.

The degree of collectivity (κ) of a motion is originally proposed by Brüschweiler (1995). Tama and Sanejouand have used it to estimate the degree of collectivity of the conformational change considered [28], reflecting the number of atoms or residues that is affected by the conformational change. It is defined as [78],

$$\kappa = \frac{1}{N_{res}} \exp \left(- \sum_i^{N_{res}} \alpha \Delta R_i^2 \ln(\alpha \Delta R_i^2) \right) \quad (2.27)$$

where the summation is performed over αC coordinates, N_{res} is the number of residues and α is the normalization factor. The normalization factor α is chosen such that $\sum_i^{N_{res}} \alpha \Delta R_i^2 = 1$. κ is defined as proportional to the exponential of ‘the information entropy’ that is embedded in the ΔR_i^2 term; in other words, κ gives the effective number of non-zero ΔR_i^2 . If the conformational change is maximally collective, all ΔR_i^2 are identical and κ approaches 1. For the extreme local motion, where the conformational change involves very little number of residues, κ is minimal.

In another study, the degree of collectivity is used as a measure to determine the plausability of complex formation according to the increase in degree collectivity of chains A and B upon complex formation [79]. Here, the degree of collectivity (κ_m), of a motion by any individual mode m is defined as proportional to the exponential of ‘the information entropy’ of the eigenvector m . The degree of collectivity by any individual mode m is defined as [78, 79]:

$$\kappa_m = \frac{1}{N_{res}} \exp \left(- \sum_i^{N_{res}} \alpha \Delta R_i^2 \ln(\alpha \Delta R_i^2) \right)_m \quad (2.28)$$

2.3.1. Parameters

The simulation parameters for the ANM part are the cutoff radius (R_{cut}), maximum number of modes to choose among ($modemax$) and the deformation factor (DF) to deform the conformations. The parameters changed in the simulation part are the temperature (T), the number of energy minimization steps ($stepmin$), the number of Langevin dynamics simulation steps ($stepsim$) and the damping constant (γ) of LD simulations. The other parameters that are used are given in Section 2.3.2.

2.3.2. Computational Details and the Protein Dataset

Langevin dynamics simulations were performed using the Sander program of Amber 11 and the force field parameter sets Amber ff03 and ff10 [80, 81]. The models used as the starting configurations are the unbound forms. These structures are constructed by the removal of the ligands from the X-ray structures (with an exception of Hsp90, which was also simulated in the presence of ATP and the drug geldanamycin, given in Chapter 4). The details of the models are summarized in Table 2.1. Protonation states and pK of charged residues are predicted using PDB2PQR and H++ servers [82, 83]. The generalized Born implicit solvent model (igb1) is used in all simulations.

Table 2.1. Details of the proteins/models studied.

| Model Protein | System Size | RMSD (Å) | Based on PDB Entry |
|------------------|---------------------|-------------|----------------------------|
| Adenylate Kinase | 214 res., monomer | 6.80 | 4AKE (open), 1AKE (closed) |
| MalFGK2 | 1369 res., tetramer | 6.85 | 3FH6 (IF), 3PUY (OF) |
| Hsp90 | 1352 res., dimer | 24.00 | 2CG9 (open), 2IOQ (closed) |
| CRP | 394 res., dimer | 4.44 | 2WC2 (apo), 1G6N (holo) |
| CaATPase | 994 res., monomer | 13.90 | 1SU4 (open), 1IWO (closed) |
| GroEL | 524 res., monomer | 11.98 | 1GRU (open), 1KP8 (closed) |
| FimD | 518 res., monomer | 17.10 | 4J3O (open), 3OHN (closed) |
| c-Src Kinase | 457 res., monomer | 23.31 | 2SRC (open), 1Y57 (closed) |
| BtuCD | 1144 res., tetramer | 3.72 | 1L7V (OF), 4R9U (IF) |

Energy minimization was performed using 50 cycles of steepest descent algorithm, followed by 500 and 1000 steps of conjugate gradient method, for small and large systems, respectively. Initial velocities were selected at random from the Maxwell-Boltzmann distribution. The Langevin thermostat [84] and Berendsen barostat [85] was used for the NPT ensemble (310 K, P=1 bar) with a time step of 2 fs. The SHAKE algorithm [59] was used as the bond constraints for all bonds involving hydrogens to eliminate the high frequency bond vibrations. A cutoff distance of 99 Å was used for the nonbonded interactions.

2.4. Characterization of conformational states by GNM eigenvalues

Combination of MD and GNM could be powerful in characterizing different conformational dynamic states along MD trajectories, as also utilized in a novel study [86]. How to characterize different conformational states is still an open issue. Clustering is one way to reduce the conformational space into a subset of conformations. In the latter study, it was discussed that using GNM with MD is advantageous over solely distance based metrics for the analysis such as PCA and clustering alone, which may

be subject to bias like large loop motions.

Here, the idea of plotting the frequency distributions of the eigenvalue of the first (or first few) GNM eigenvector of a series of snapshots is exploited to compare the transition pathways and characterize different conformational dynamic states that the protein assume during the transition. The frequency distributions are shown to be sensitive to the different conformational states of the protein that are assumed in the presence of a ligand or a mutation. The scatter plot of the first and second eigenvalues of GNM is used to visualize each conformational transition pathway generated by the ANM-LD method.

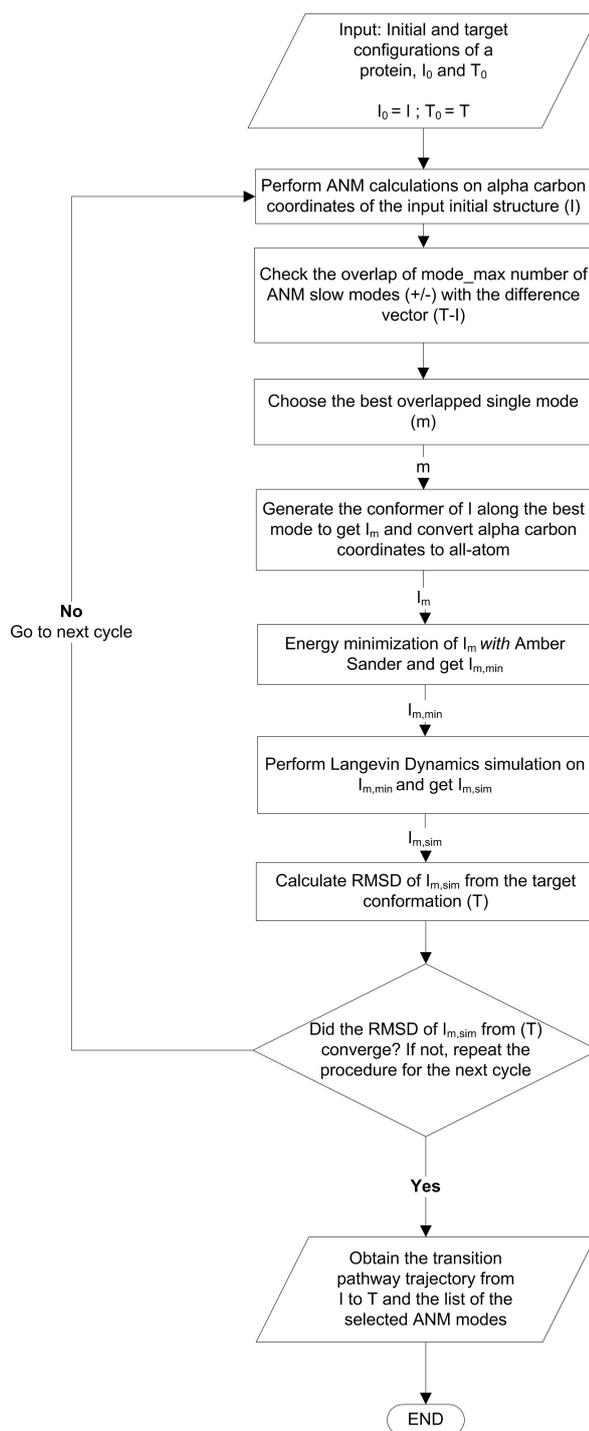


Figure 2.1. Flowchart of the algorithm of ANM Driven Langevin Dynamics (ANM-LD) Simulations.

3. CONFORMATIONAL TRANSITION PATHWAYS BY COLLECTIVE MODES DRIVEN MOLECULAR SIMULATIONS

Proteins dynamics and its correlation to functional mechanism still remain elusive and become more complex with large protein complexes or membrane proteins. Proteins sample many conformations and undergo conformational changes to realize their function. Each protein has its own intrinsic conformational landscape that discloses thermodynamically stable states and the transitions in between for a functional trait. Deciphering the functioning mechanism underlying how biological macromolecules act for the principles to regulate their function is of significant interest. Here, we propose an all-atom molecular simulation methodology that integrates anisotropic network model with stochastic Langevin Dynamics simulations, ANM-LD, for generating, rendering and controlling biologically plausible conformational transition pathways between two given functional states. ANM-LD is applied for a set of proteins and/or complexes, where transitions starting from initial RMSDs as high as 14 Å or 24 Å ended up with 2 Å RMSD from the target state. As exemplary cases, the analysis and results for adenylate kinase (AdK) and a membrane protein, maltose transporter (MalFGK2), is elaborately presented. Intrinsic dynamics enables conformational transitions towards the target from an RMSD of 6.8 Å to 0.8 Å for AdK and from 6.85 Å to 3.38 Å for MalFGK2. Certain global modes are repeatedly selected and the hindrance of these highly sampled cooperative modes lead to the evolution of alternative pathways with yet similar cooperativity or diminishing cooperativity to the target. To this end, our methodology is unique in its entirety as a handy platform making it possible to create parallel transition pathways using the intrinsic dynamics information obtained from ANM by taking advantage of its all-atom molecular simulation partner Amber-Sander, controlled by many parameters that could easily be utilized to change the system conditions or incorporating any simulation methodology, such as explicit solvent MD, AMD, self-guided LD or umbrella sampling, by changing only a few parameters. Also, it is possible to incorporate experimental or other restraints into the protocol by modifying

the developed code. Controlling the generation of the pathways by the restriction of certain modes or assessing the difference of the analyzed pathways in terms of sampling different conformational states by GNM eigenvalues are this work's other novelties.

3.1. Introduction

Exploring the dynamics of large proteins and large conformational changes by hybrid methods integrating coarse-grained elastic network models and all-atom molecular simulations are of great value. Conventional MD simulations alone suffer from sampling inaccuracies and most probable accessible motions by current MD simulations are local conformational changes, fluctuations near a global minimum and some loop orientations (up to microseconds). Combining various types of computational approaches (such as coarse-grained NMA/ENM based approaches with different types of all-atom molecular simulations and experiments) is valuable for bridging experimental/computational studies and may help consolidate the data obtained from each method.

Hybrid methodologies employing enhanced sampling techniques have the potential to plot large conformational changes between functional states and in between intermediate states that cannot solely be accessible by experiments or conventional simulation methods [7, 8, 71]. A summary of these hybrid methodologies is given in Section 1.3.3.

The methodology proposed in this study combines ANM with all-atom LD simulations, ANM-LD, to enhance stochastic simulations by the intrinsic dynamics of the protein to observe large conformational changes between two stable functional states (e.g. open/closed, inactive/active, etc.) and obtain intermediates along (parallel/alternative) transition pathways sampled in between these states. In this methodology, the initial conformation is moved along the elastic modes of motion encoded by its fold towards the target conformation, taking advantage of the evaluation of the interactions and energetics of the system via short cycles of all-atom implicit LD simulations. Consecutive distortions driven by single low-frequency modes, i.e. the disposition of

intrinsic dynamic behavior adopted by the topology of the structure, can rationalize the functional conformational change along the transition pathway generated using ANM-LD simulations.

ANM-LD runs were performed for the diverse data set of proteins described in Section 2.3.2, focusing on the prototype protein AdK and the membrane protein MalFGK2 as the exemplary cases. The method was developed and tested by AdK, which has known and distinct open/closed structures. AdK is well characterized, but with open issues, such as the order of binding of the ligands ATP, AMP and correlation dynamics of the domains, whether they move concertedly or independently [87, 88]. Additionally, the transition pathways and the coupling dynamics in the exemplary case MalFGK2 are explored with the present methodology. The conformational dynamics and open/closed transition of the molecular chaperon Hsp90, which show very large conformational changes, is explored as another case along with AFM experimental work and given in Chapter 4.

The methodology is tested with its ability to sample the experimentally available (intermediate state) structures for obtaining biological and reactive conformational transition pathways (those by which transition from one to the other state actually take place). The parameter set for each case was adjusted in order to increase the computational efficiency and the success of simulations. Also, different strategies have been explored to generate alternative pathways, such as controlling the generation of the pathways by the restriction of certain modes. Differences in the analyzed pathways are assessed in terms of sampling different conformational states by GNM eigenvalues. Our method is computationally efficient and applicable to larger proteins/protein complexes like other coarse-grained approaches, with the use of intrinsic dynamic information and implicit solvent LD simulations.

Developing this methodology has led to a framework where the effect of perturbations, such as temperature change, restriction of certain modes/movements or *in silico* mutations on the preferred conformational pathways could be studied. In order to permit the system to diverge from the shortest path and explore pathways involving higher

energy barriers, the random selection of the modes is also possible in the algorithm. In this way, not all motions toward the target are enforced [4].

3.2. Materials and Methods

3.2.1. Structures

ANM-LD runs are performed for open/closed states of AdK (PDB IDs: 4AKE/1AKE) [89,90] and inward-facing (IF)/outward-facing (OF) states of MalFGK2 (PDB IDs: 3FH6/3PUY) [91,92]. PDB structures used and the system sizes for all cases are given in Table 2.1. Missing residues are modeled using ModWeb/ModLoop servers [93,94]. While modeling, steric clash problems encountered are solved by refining the protein structure on Chiron server [95].

3.2.2. ANM-LD Simulation Protocol

We have performed ANM-LD conformational transition simulations between two known conformational states, where the details of the methodology was given in Section 2.3. ANM-LD hybrid methodology simulates the conformational transition starting from an initial state towards a target state by biasing the motion of the initial structure along the ANM modes mostly overlapping with the initial-target coordinate difference vector followed by short cycles of the energy minimization and implicit solvent LD simulations. The cycles are repeated until the RMSD from the target conformation reaches a plateau. Sander of Amber 11 [81] biomolecular simulation program was used for the energy minimization and LD simulations. The computational details for the simulation part of the method are given in Section 2.3.2.

3.2.3. Adjustment of Parameters

The effect of parameters are investigated and optimized in order to approach the target conformation more closely in both forward and reverse runs. Parameters, R_{cut} , DF , $modemax$, $stepmin$ and $stepsim$ are changed systematically to determine

the optimum set for each case (See Section 2.3.1 for the description of the parameters). For the prototype case, AdK, ANM-LD runs for different deformation values ($DF=0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1 \text{ \AA}$) were performed for forward and reverse directions, by keeping the other parameters constant ($R_{cut} = 13 \text{ \AA}$, $modemax=30$ (for $DF=0.4 \text{ \AA}$ only), 100 , $T=310 \text{ K}$, $stepmin=500$, $stepsim=100$, $\gamma = 5ps^{-1}$). RMSD values for these forward and reverse runs are plotted in Figures 3.2 and 3.4. $DF=0.4 \text{ \AA}$ - $modemax=100$ is the most successful in terms of approaching the target ($RMSD_{final} \leq 0.9 \text{ \AA}$). Thus, the results for AdK will be presented for the chosen parameter set. Lowest $RMSD_{final}$ values obtained using the combination of these parameters for all runs are listed in Table 3.1.

The effect of maximum mode number on assumed pathways is explored by using different number of modes ($modemax=2, 5, 10, 30, 100, 200, \text{all modes}$). Converged RMSD values have improved further when the maximum number of modes was changed from 30 to 100. For $modemax=2$, the convergence to the target structure was not good enough, considering that large and complex conformational transitions require more than five slow modes (changing according to the system studied). Further systematic trials have been performed in order to optimize parameters and to generate different transition pathways assessed by the similarity between predicted intermediate and experimental conformations. The final parameter set obtained for each protein (using the developed approach for pathway exploration) comes out as an intrinsic property of each case studied. $RMSD_{initial}$ and $RMSD_{final}$ values for the forward and reverse (where available) runs for the studied cases are listed in Table 3.1.

3.3. Results and Discussion

3.3.1. Model Case: Adenylate Kinase

AdK protein consists of 214 residues and three domains (LID, CORE, NMP). These domains are shown on the open and closed X-ray crystal structures in Figure 3.1. LID and NMP domains are in the open state in the open structure and in the closed form in the closed structure. The closure of both LID and NMP domains is observed

Table 3.1. Summary of ANM-LD simulation results.

| Protein-Direction | DF (Å) | RMSD initial (Å) | RMSD final* (Å) |
|--------------------------------------|-------------------|---------------------------------|--------------------------------|
| Adenylate kinase (Open-Closed) | 0.4 | 6.80 | 0.87 |
| Adenylate kinase (Closed-Open) | 0.4 | 6.80 | 0.78 |
| Adenylate kinase Y171W (Open-Closed) | 0.4 | 6.80 | 0.91 |
| Adenylate kinase P177A (Open-Closed) | 0.4 | 6.80 | 0.90 |
| MalFGK2/IF-OF | 1 | 6.85 | 3.53 |
| MalFGK2/OF-IF | 1 | 6.85 | 3.38 |
| Hsp90/Open-Closed | 1 | 24.0 | 2.81 |
| Hsp90/Closed-Open | 1 | 24.0 | 4.76 |
| Hsp90-ATP/Open-Closed | 1 | 24.4 | 2.88 |
| Hsp90-GDM/Open-Closed | 1 | 24.4 | 2.88 |
| CRP/Inactive-Active | 0.6 | 4.44 | 1.89 |
| CaATPase/Open-Closed | 1 | 13.97 | 2.81 |
| CaATPase/Closed-Open | 1 | 13.97 | 2.79 |
| GroEL/Open-Closed | 0.4 | 11.98 | 1.24 |
| FimD/Open-Closed | 1 | 17.53 | 5.02 |
| FimD/Closed-Open | 1 | 17.53 | 3.83 |
| c-Src Kinase/Open-Closed | 0.8 | 23.31 | 2.85 |
| c-Src Kinase/Closed-Open | 0.8 | 23.31 | 3.24 |
| BtuCD/OF-IF | 0.2 | 3.72 | 1.54 |

* The best result observed is given for each system for the parameter set: ($R_{cut} = 13 \text{ \AA}$, $modemax = 100$ (30 for FimD and BtuCD), $stepmin=500$ (1000 for Hsp90), $stepsim=100$, $\gamma=5 \text{ ps}^{-1}$ (50 ps^{-1} for Hsp90)).

during the open-closed transition [96]. This situation is an example of the presence of parallel pathways.

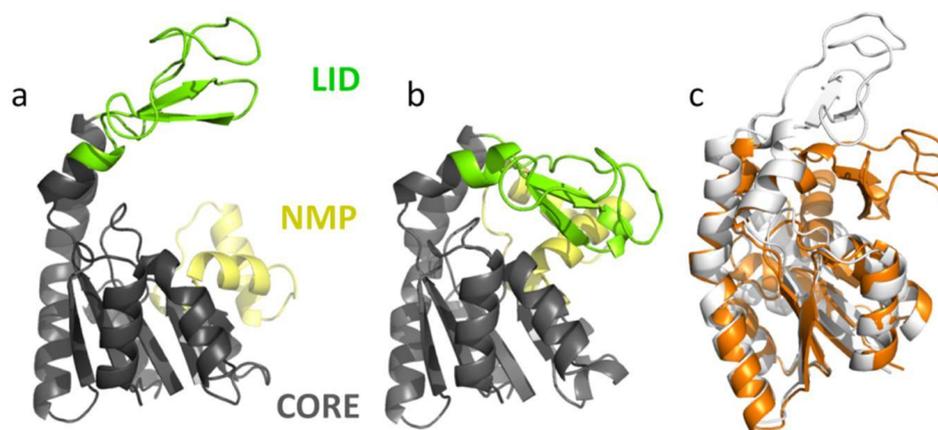


Figure 3.1. Adenylate kinase (a) open (PDB ID: 4AKE) and (b) closed (PDB ID: 1AKE) conformations are shown with LID, CORE and NMP domains colored as green, gray and yellow, respectively. (c) The open (white) and closed (orange) forms are given superimposed on each other (RMSD = 6.8 Å).

The structural transition of AdK is defined by two angles, NMP-CORE angle and the LID-CORE angle, representing the closure of the NMP and LID domains on the CORE, respectively. NMP-CORE angle is defined by the representative axes formed by the centers of geometry of α C atoms of residues 115-125, 90-100 and 35-55, whereas the LID-CORE angle is defined by residues 179-185, 115-125 and 125-153 [71, 97]. These two angles are widely used as informative low-dimensional descriptors/collective variables in the recent literature for describing the extent of the sampling of the conformational transition by different methods and providing a common ground for comparisons.

3.3.1.1. Analysis of Pathways. For the prototype protein AdK, conformational transition pathways generated by ANM-LD forward and reverse directions were analyzed with respect to experimentally known PDB structures for their biological relevance. RMSD plots of ANM-LD runs performed for different deformation values ($DF=0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1$ Å) in forward and reverse directions are presented in Figures

3.2 and 3.4. $DF=0.4$ Å- $modemax=100$ is the most successful in terms of approaching the target state ($RMSD_{final} \leq 0.9$ Å). Also, RMSD values are plotted separately in Figures 3.6c-d, 3.7 and 3.9 for the selected parameter set ($R_{cut} = 13$ Å, $modemax=30$, 100 , $DF=0.4$, 1 Å, $T=310$ K, $stepmin=500$, $stepsim=100$, $\gamma = 5ps^{-1}$), with the corresponding selected modes, overlap, and degree of collectivity values for each cycle. The maximum number of modes was observed to change the course of the transition. The converged RMSD to target values improves when the mode selection is allowed to be among 100 slow modes.

The nature of the mostly selected modes was explored. The effect of mode restriction on the course of the transition and alternative pathways emerged were elaborated. In forward runs slow modes 1 and 2 were selected frequently. Restriction of these modes hindered the transition. $RMSD_{final}$ stuck at around 3.89-4 Å in all four parallel runs (See Figure 3.8), which has a minimum $RMSD_{final}$ value of 0.86 Å in its unrestricted counterpart. Mode restriction appears to result in alternative pathways verified by the overlap of the appearing conformations on the path with available AdK experimental crystal structures. The restriction of the mostly selected modes (slow modes 1 and 5) in reverse runs (See Figure 3.10) ended up with an $RMSD_{final}$ of 1.27 and 1.53 Å, where the unrestricted transition had an $RMSD_{final}$ of 0.78 Å. Although selected modes did not totally hinder the transition, it was observed to change the course of the transition significantly. This possibly implies appearance of alternative pathways, which indeed overlap with some available experimental crystal structures. Figures 3.5, 3.6 and 3.15 through 3.18 show these overlaps.

GNM eV1 vs eV2 scatter plots (See Section 2.4 for the description of GNM eVs distinguishing conformational states) exhibit the differences of the pathways. Here, the path with the lowest $RMSD_{final}$ for both forward (Figures 3.3) and reverse (Figures 3.5) cases, are for $DF=0.4$ Å. The size of the applied deformation along the selected mode observed to change the transition pathways, as well as the selected mode of the next cycle. All unrestricted/restricted paths are separately drawn for forward and reverse runs in Figures 3.6c and 3.6d, respectively. Overlaps of the pathways with the experimental structures are observed in especially the reverse restricted runs (Figure

3.6d).

Histogram plots of GNM eV1 are given in Figures 3.11 through 3.14 for all studied runs. This way, the appearance of possible intermediate states may be spotted in the bimodally distributed histograms, such as the forward runs with $DF=0.8$ and 1 \AA (Figure 3.11).

In forward simulations, LID closure precedes NMP closure (See Figure 3.15 and 3.16), consistent with the ligated coarse-grained approaches [98] modeling the AdK transition, and opposing the non-ligated [99]. Although the lack of the ligands (ATP and AMP) may decrease the probability of attaining stability and slow down the path to reach the fully closed state, intrinsic dynamics of AdK has allowed the appearance of the same sequence of events. In all forward restricted runs the transition is hindered (RMSD reaches a plateau at $3.89\text{-}4.0 \text{ \AA}$) and also the sequence of events have changed (See Figure 3.16), i.e., NMP closure came before the LID closure.

For the reverse case, biologically, the transition does not require the presence of the ligands and it should complete fastly. In a coarse-grained path modeling study [52] it is stated that, again the LID domain motion precedes the NMP domain, agreeing the findings of the reverse unrestricted runs of this study. Nevertheless, in another recent enhanced simulation study it is stated that NMP domain reaches the open state before the LID domain [87], which is observed in all reverse restricted runs (See Figure 3.17 and 3.18).

Seyler *et al.* [71] has compiled a summary plot, comparing the path-sampling approaches for the AdK closed-open transition with each other on the NMP-CORE/LID-CORE angle plot. Precedence of the LID domain motion is observed in both ENM-based approaches (iENM and ANMPathway). Also, the generated ANM-LD forward/reverse pathways are superimposed on the free energy landscape/potential mean force (PMF) data provided by Beckstein *et al.* [97] (See Figure 3.19). The created ANM-LD pathways are observed to bypass high penalty energy regions, implying their biological relevance.

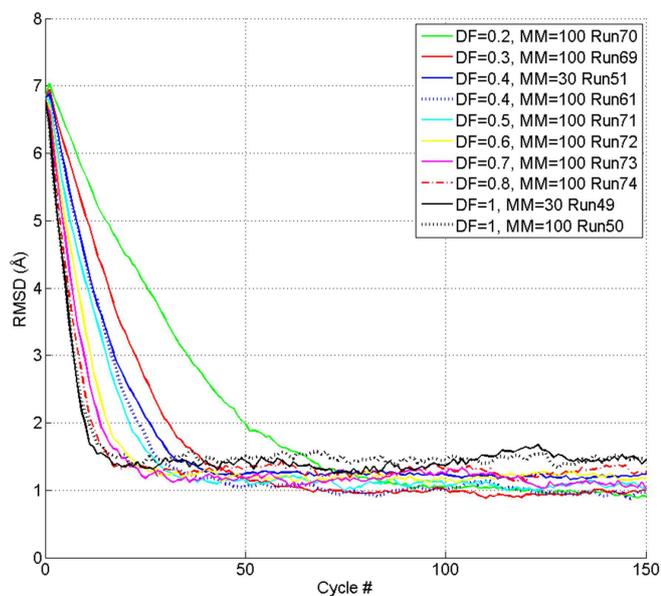


Figure 3.2. RMSD plots-AdK forward runs (open-closed) for different DF values and $modemax$, MM=30, 100 ($T=310$ K, ANM $R_{cut}=13$ Å).

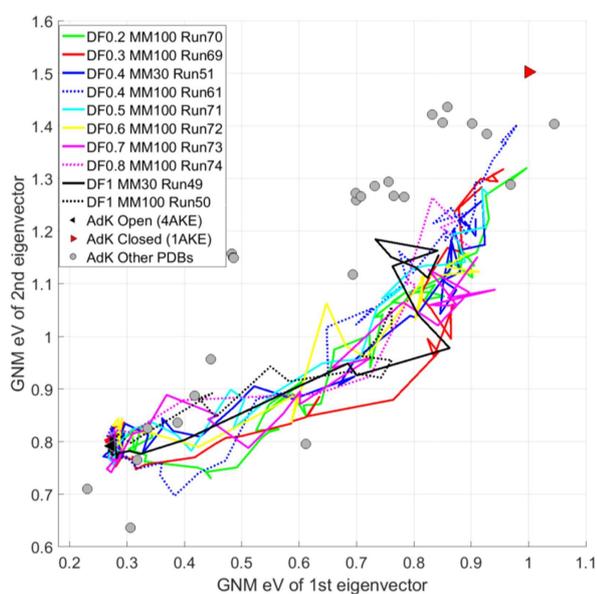


Figure 3.3. GNM eigenvalue of first eigenvector (eV1) versus GNM eigenvalue of second eigenvector (eV2) plots – AdK forward runs (open-closed) for different DF values and $modemax$, MM=30, 100 ($T=310$ K, ANM $R_{cut}=13$ Å). Eigenvalues for the AdK open (black triangle) and closed states (red triangle) are shown in the eV plots.

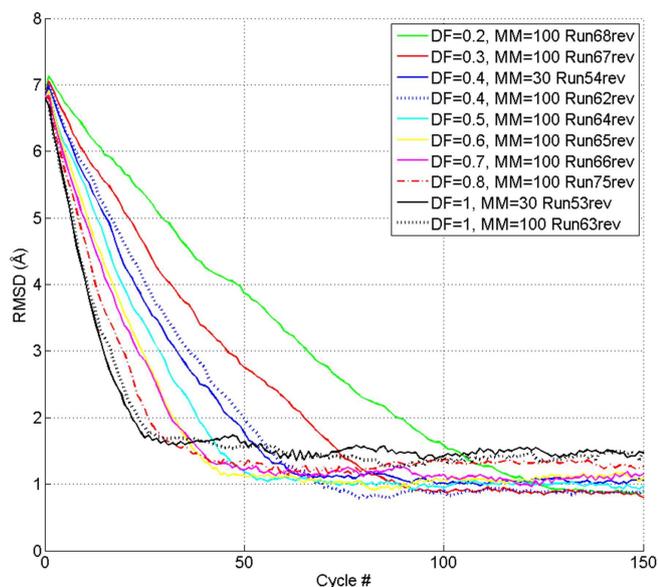


Figure 3.4. RMSD plots–AdK reverse runs (closed-open) for different DF values and $modemax$, MM=30, 100 ($T=310$ K, ANM $R_{cut}=13$ Å).

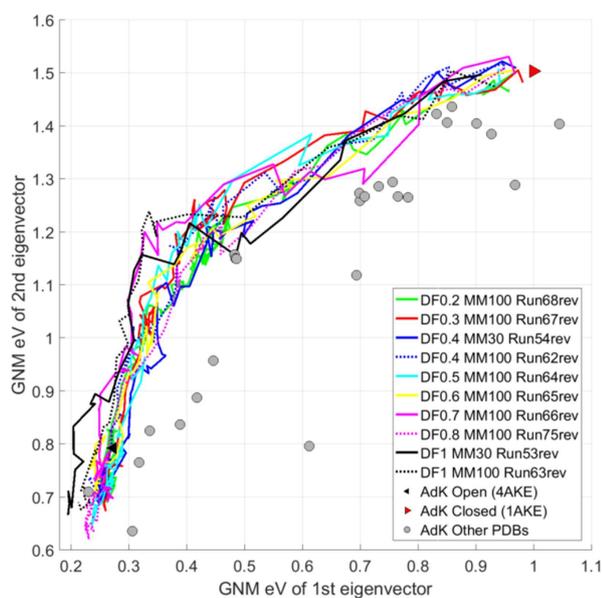


Figure 3.5. GNM eigenvalue of first eigenvector (eV1) versus GNM eigenvalue of second eigenvector (eV2) plots–AdK reverse runs (closed-open) for different DF values and $modemax$, MM=30, 100 ($T=310$ K, ANM $R_{cut}=13$ Å). Eigenvalues for the AdK open (black triangle) and closed states (red triangle) are shown in the eV plots.

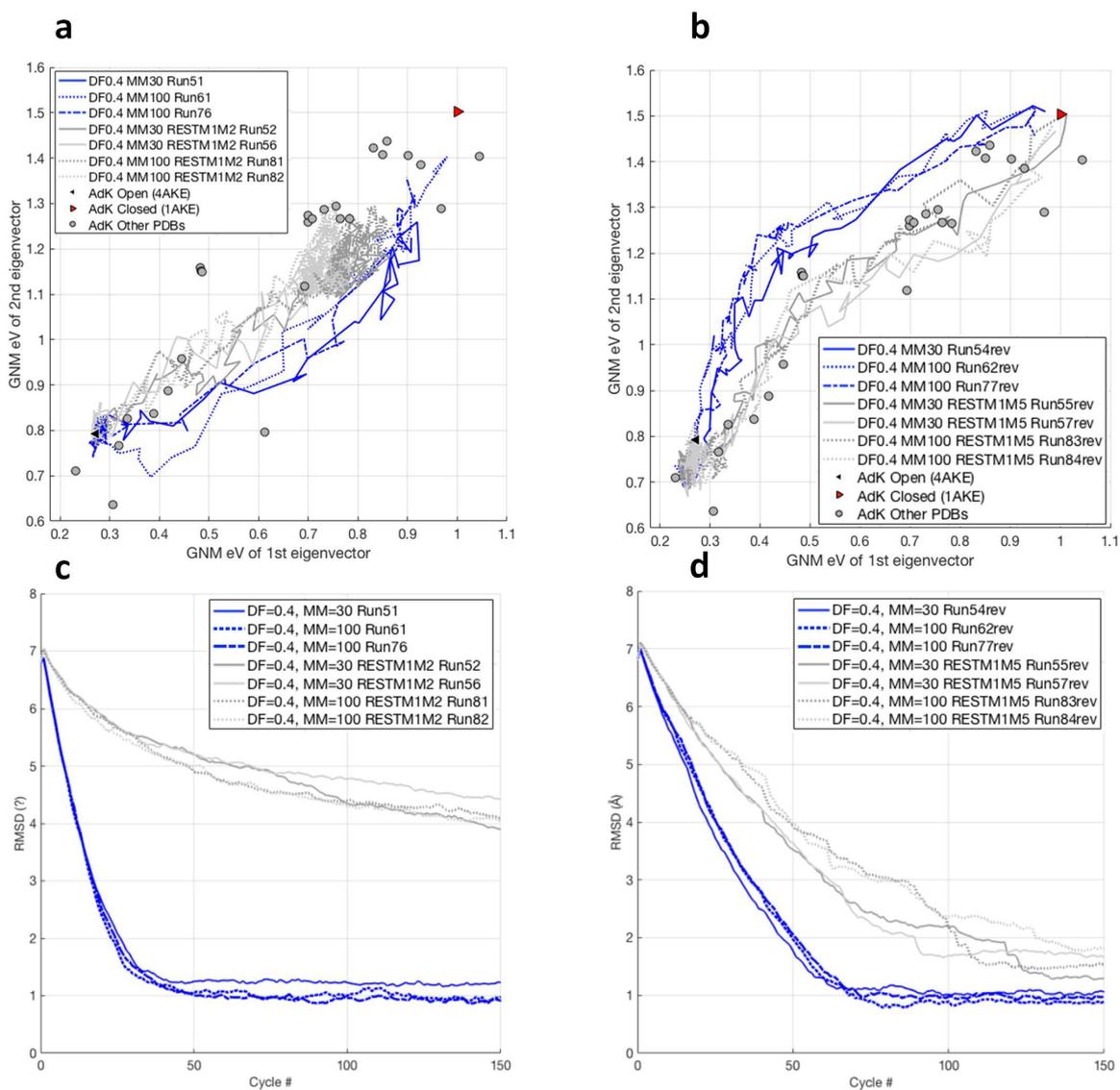


Figure 3.6. GNM eigenvalue of first eigenvector (eV1) versus GNM eigenvalue of second eigenvalue (eV2) plots for AdK forward (a) and reverse runs (b), with parameter set ($T=310$ K, ANM $R_{cut}=13$ Å, $DF=0.4$ Å, *modemax*, MM=30, 100). Eigenvalues for the AdK open (black triangle) and closed states (red triangle) are shown in the eV plots. ANM 1st and 2nd modes are restricted to block the open-closed transition, whereas ANM 1st and 5th modes are restricted to block the closed-open transition in the restricted runs. The RMSD plots for the forward and reverse runs are given in (c) and (d), respectively.

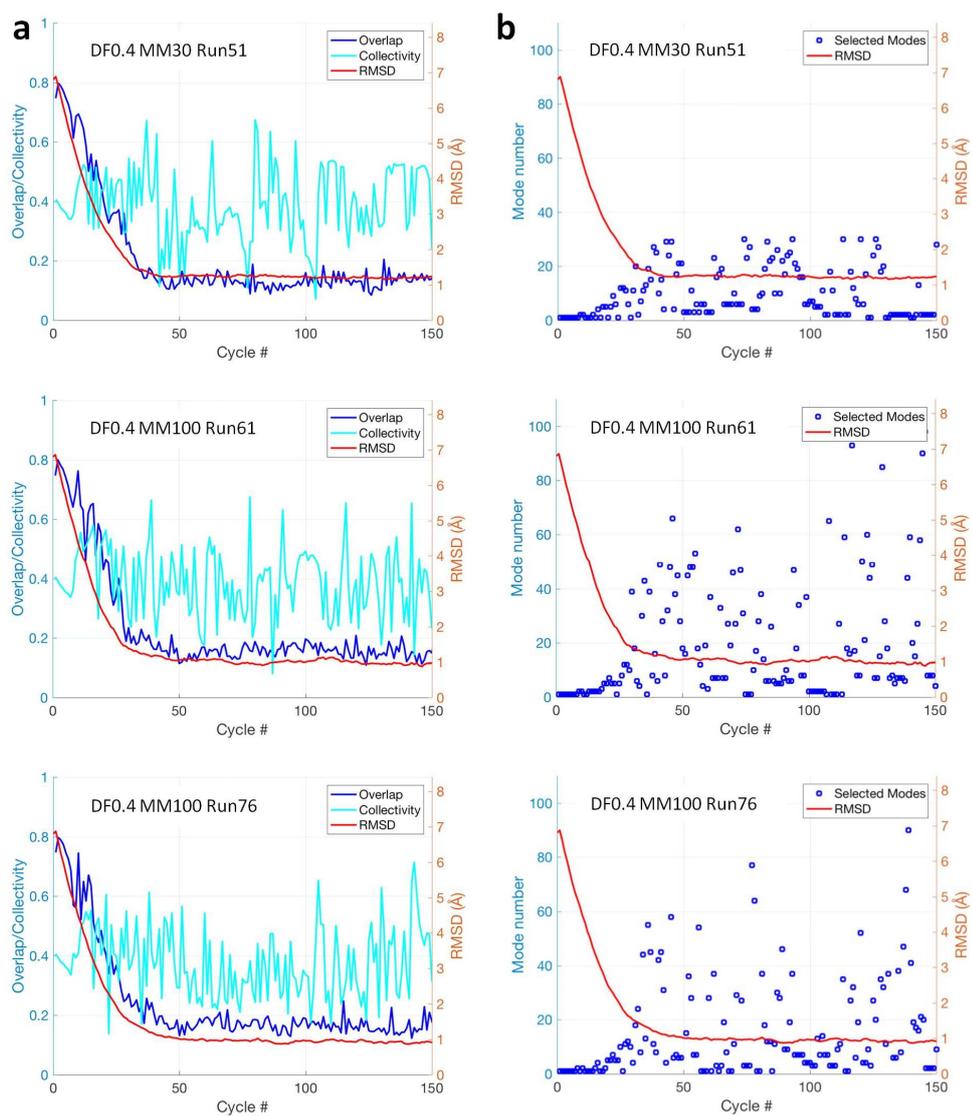


Figure 3.7. (a) Overlap/collectivity values of the selected slow modes and (b) the selected slow modes are plotted with RMSD values for AdK (unrestricted) forward simulations with $DF=0.4 \text{ \AA}$ ($T=310 \text{ K}$, $modemax=30, 100$, ANM $R_{cut}=13 \text{ \AA}$).

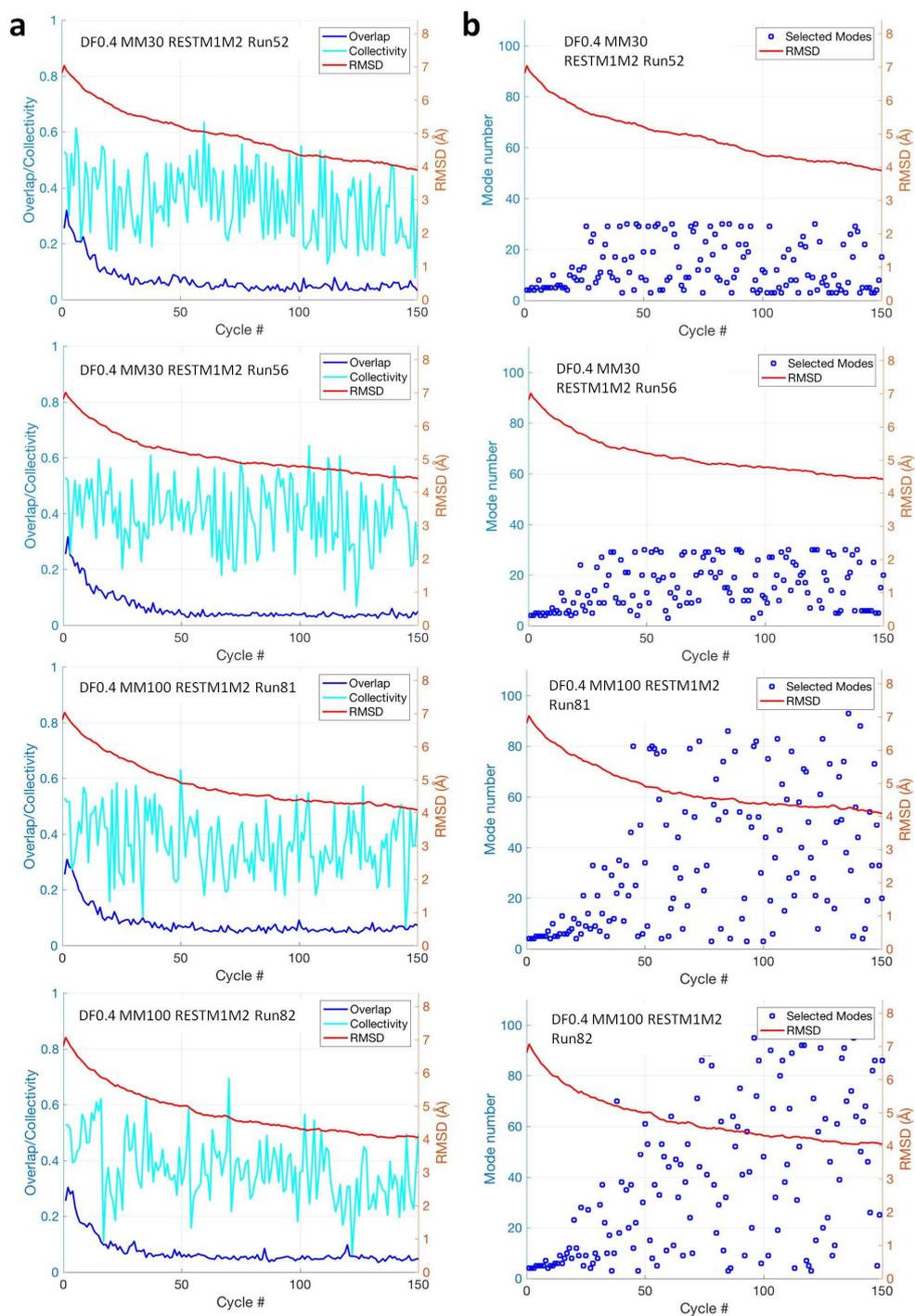


Figure 3.8. (a) Overlap/collectivity values of the selected slow modes and (b) the selected slow modes are plotted with RMSD values for AdK restricted forward simulations with $DF=0.4 \text{ \AA}$ ($T=310 \text{ K}$, $modemax=30, 100$, ANM $R_{cut}=13 \text{ \AA}$).

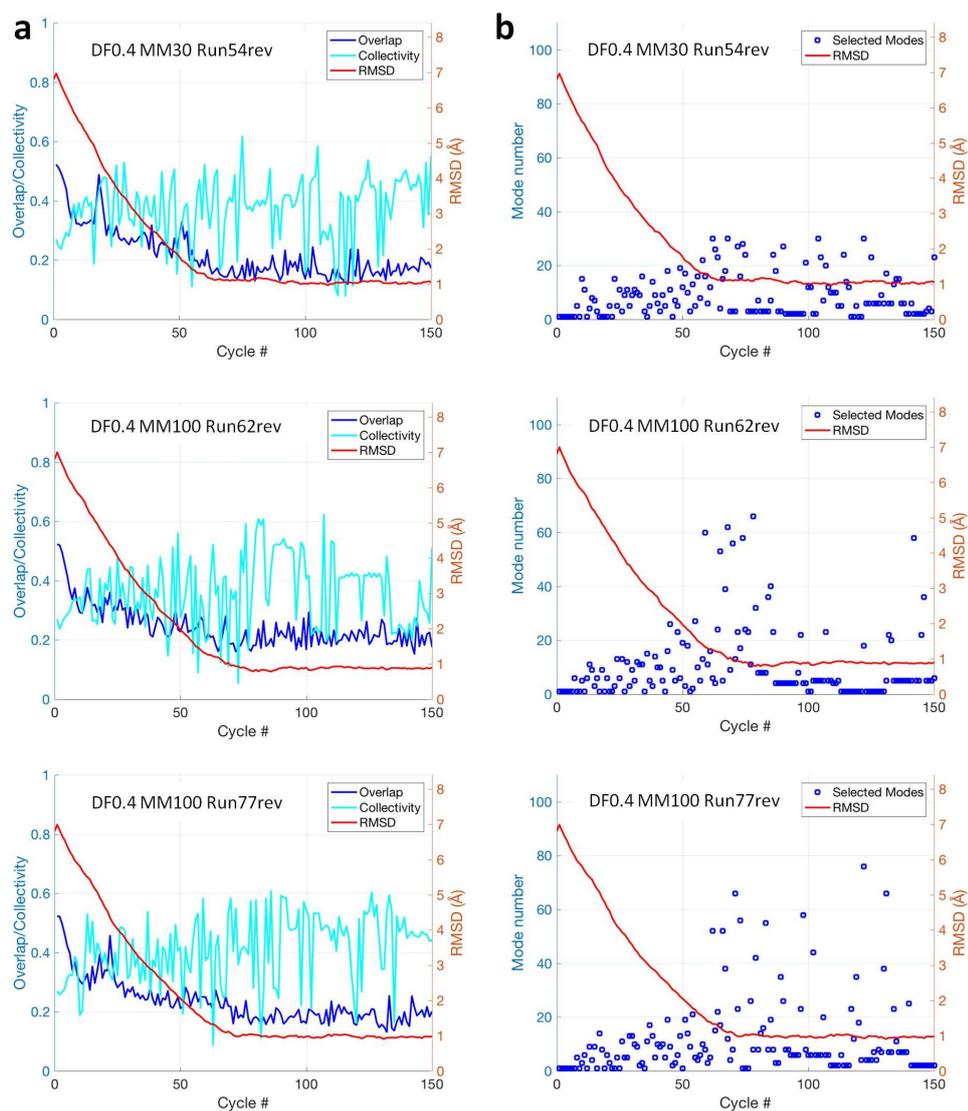


Figure 3.9. (a) Overlap/collectivity values of the selected slow modes and (b) the selected slow modes are plotted with RMSD values for AdK (unrestricted) reverse simulations with $DF=0.4 \text{ \AA}$ ($T=310 \text{ K}$, $modemax=30, 100$, ANM $R_{cut}=13 \text{ \AA}$).

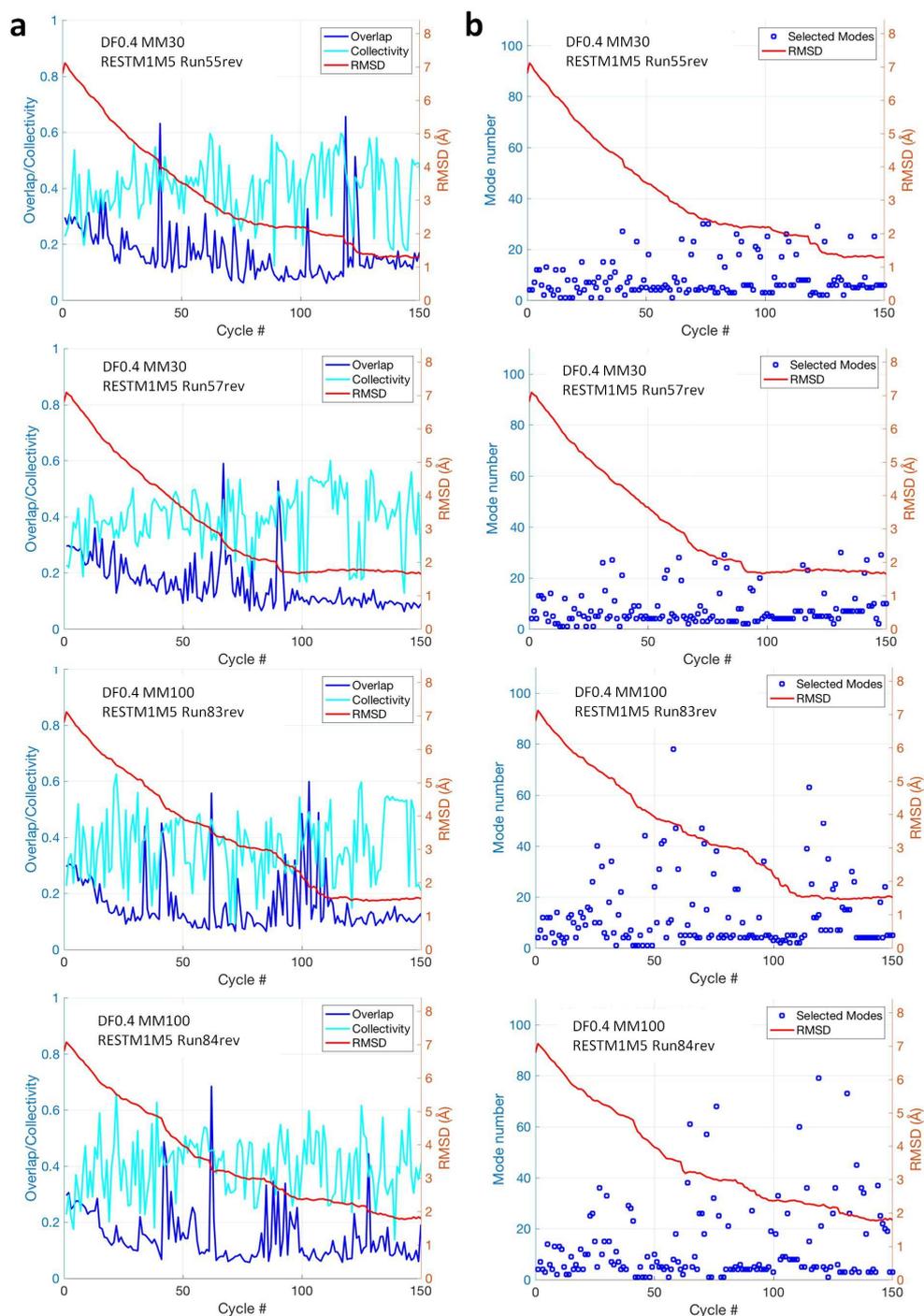


Figure 3.10. (a) Overlap/collectivity values of the selected slow modes and (b) the selected slow modes are plotted with RMSD values for AdK restricted reverse simulations with $DF=0.4 \text{ \AA}$ ($T=310 \text{ K}$, $modemax=30, 100$, ANM $R_{cut}=13 \text{ \AA}$).

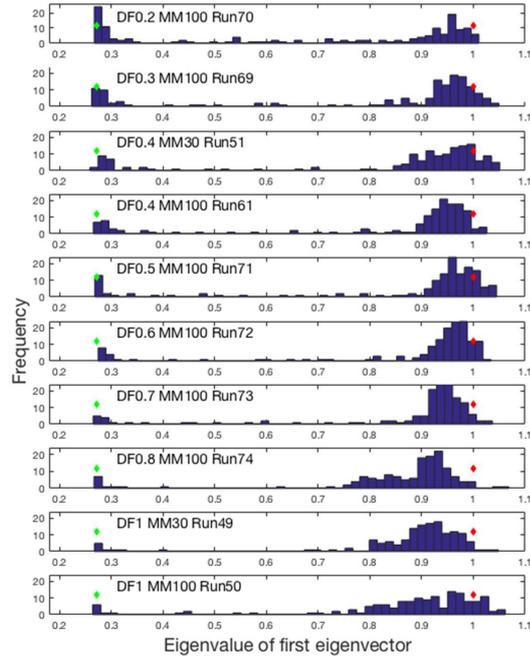


Figure 3.11. GNM first eigenvalue (eV1) histograms for AdK forward runs.

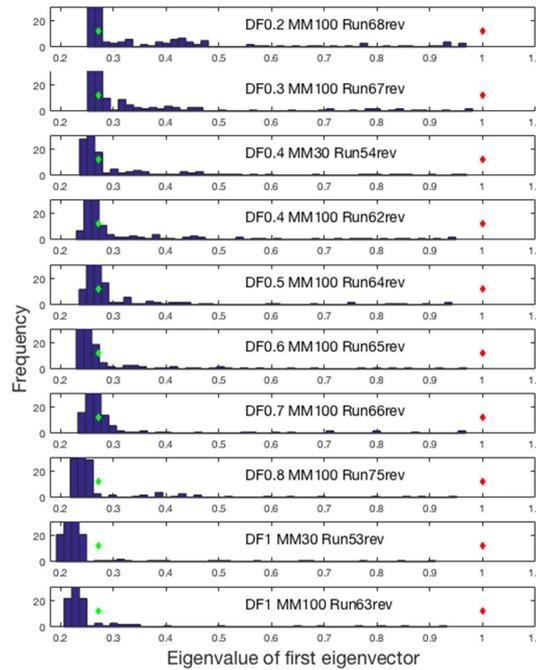


Figure 3.12. GNM first eigenvalue (eV1) histograms for AdK reverse runs.

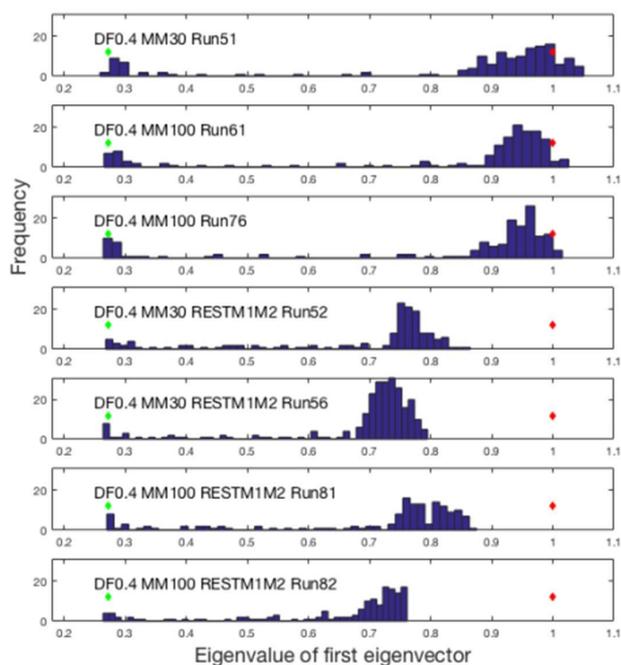


Figure 3.13. GNM first eigenvalue (eV1) histograms for AdK forward runs with $DF=0.4 \text{ \AA}$.

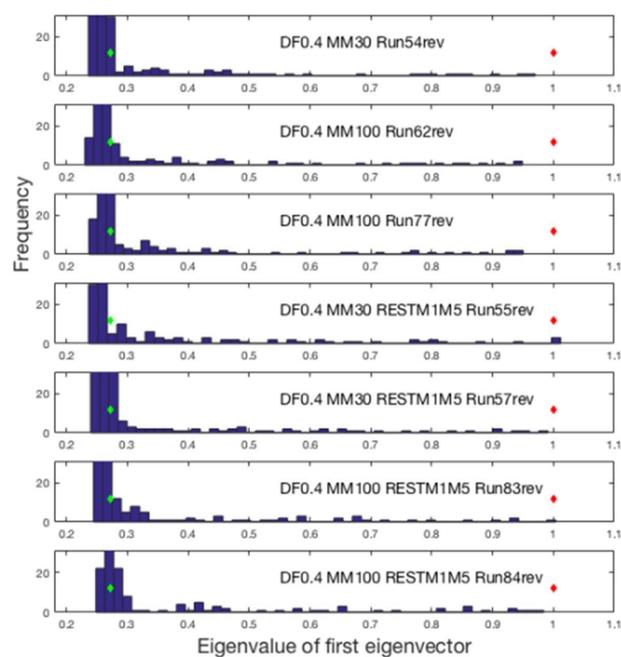


Figure 3.14. GNM first eigenvalue (eV1) histograms for AdK reverse runs with $DF=0.4 \text{ \AA}$.

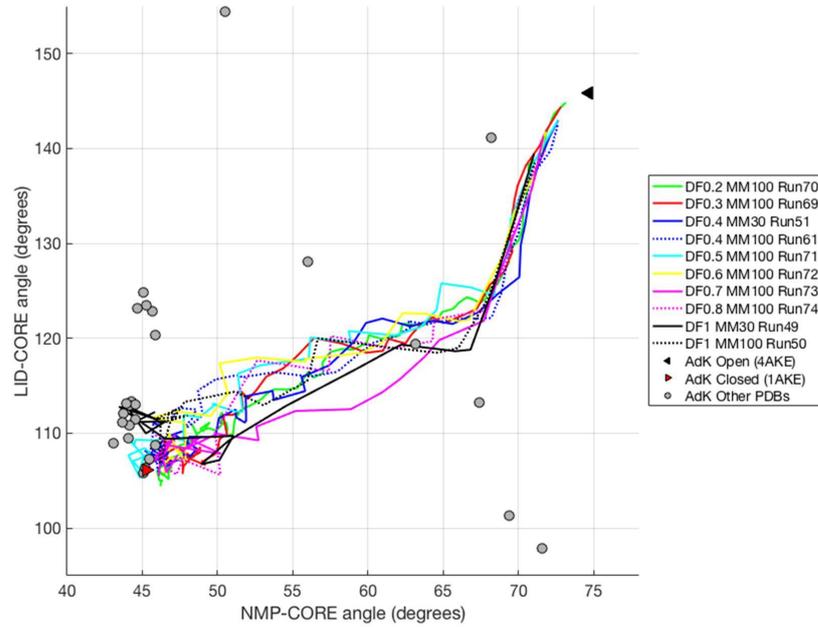


Figure 3.15. NMP-CORE angle versus LID-CORE angle for AdK forward runs.

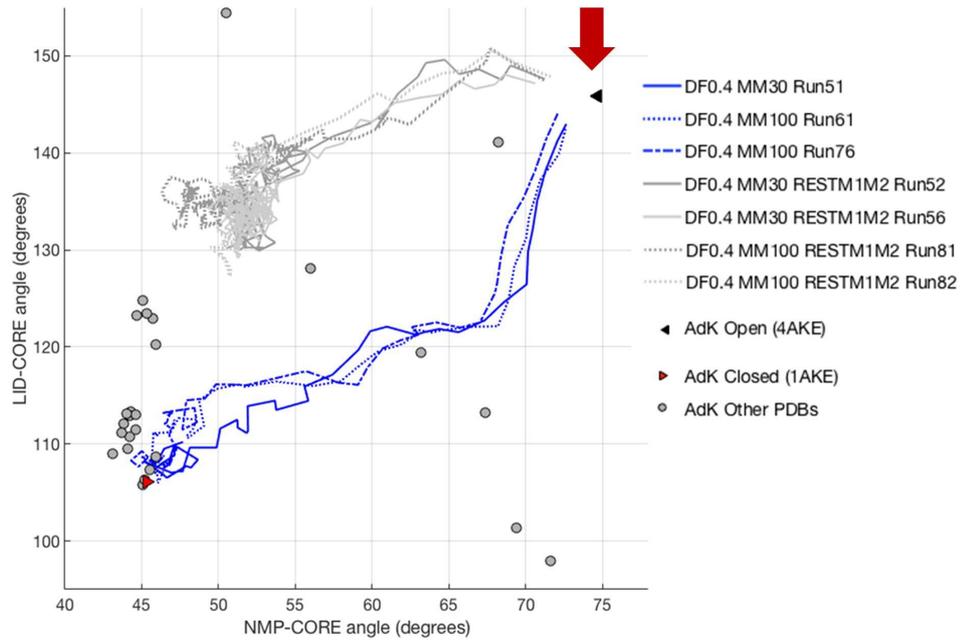


Figure 3.16. NMP-CORE angle versus LID-CORE angle for AdK forward runs with $DF=0.4 \text{ \AA}$.

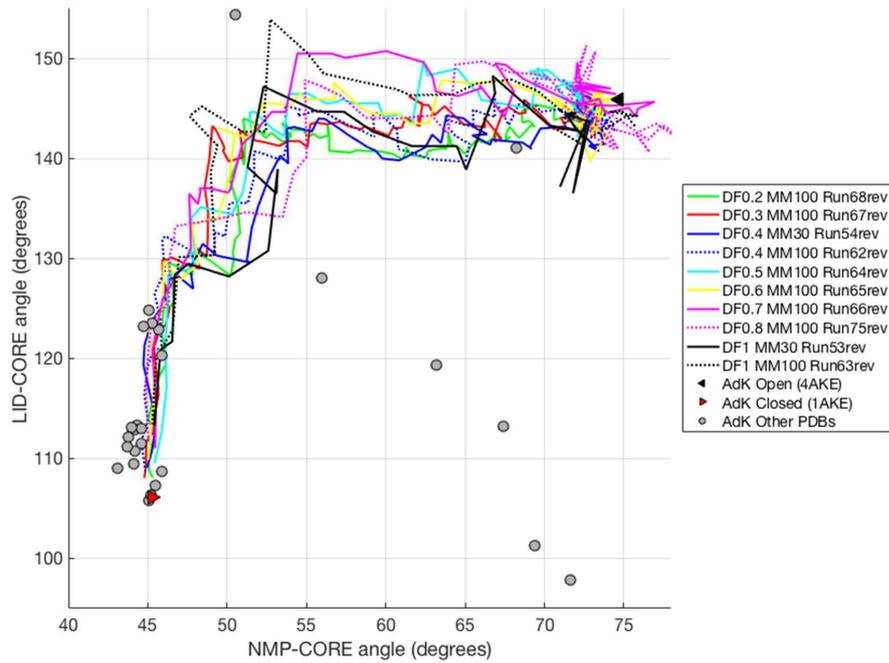


Figure 3.17. NMP-CORE angle versus LID-CORE angle for AdK reverse runs.

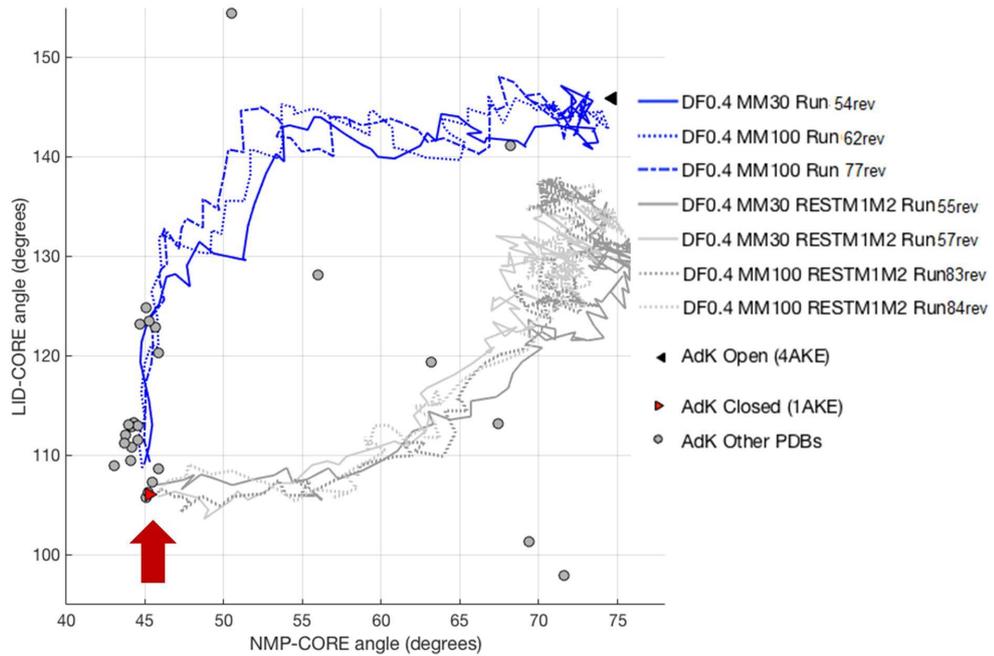


Figure 3.18. NMP-CORE angle versus LID-CORE angle for AdK reverse runs with $DF=0.4 \text{ \AA}$.

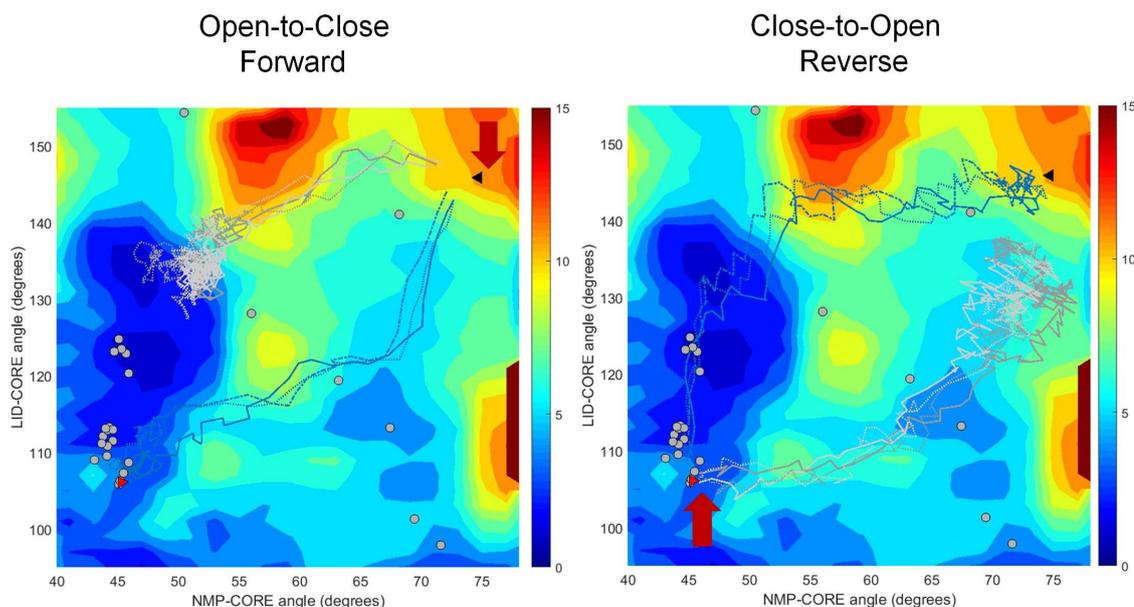


Figure 3.19. ANM-LD forward/reverse pathways on the potential of mean force (PMF) data provided by Beckstein Lab on their website [100].

Cross-correlations (See Figure 3.20) for the forward unrestricted and both M1 M2 restricted runs, show that, with the restriction of the slow modes 1 and 2 the domains lose their structured/well-defined pattern and cross maps become blurry, probably this is the reason underlying the hindered transition. Moreover, new strongly correlated regions (residues 140 vs 70-100 region) appear in the maps (more apparent for the first-half of the transition part, figures on the right).

In the cross maps of the reverse restricted runs, the domains are still well-defined just as the normal unrestricted transition maps. Explaining the failure in hindering the transition (See Figure 3.21). Here, in the restricted runs strong correlations between the residues 10 vs 55-60 region, disappear in the restricted runs, including the parallel cases.

In the forward transition, regions 85-110 and 170-190 of the core region coordinates the transition. For the unrestricted case core region is correlated with the LID-domain (and is not correlated with the NMP-domain), and LID motion precedes

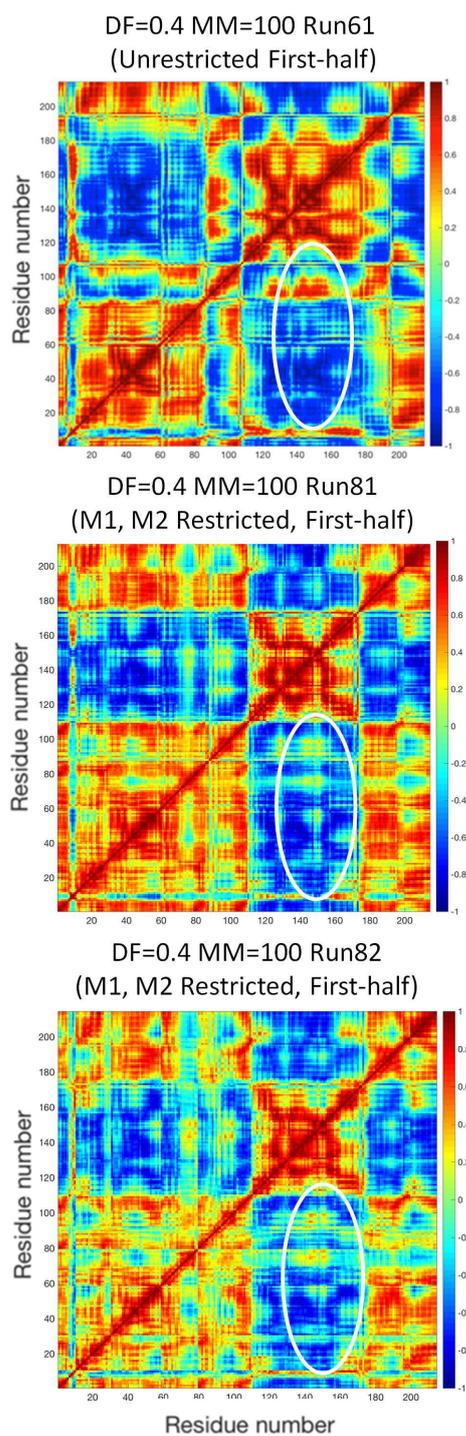


Figure 3.20. Cross-Correlation Plots for AdK forward runs. First-half of the transition parts of the trajectories (just before RMSD to target reaches a plateau) are considered for the calculations.

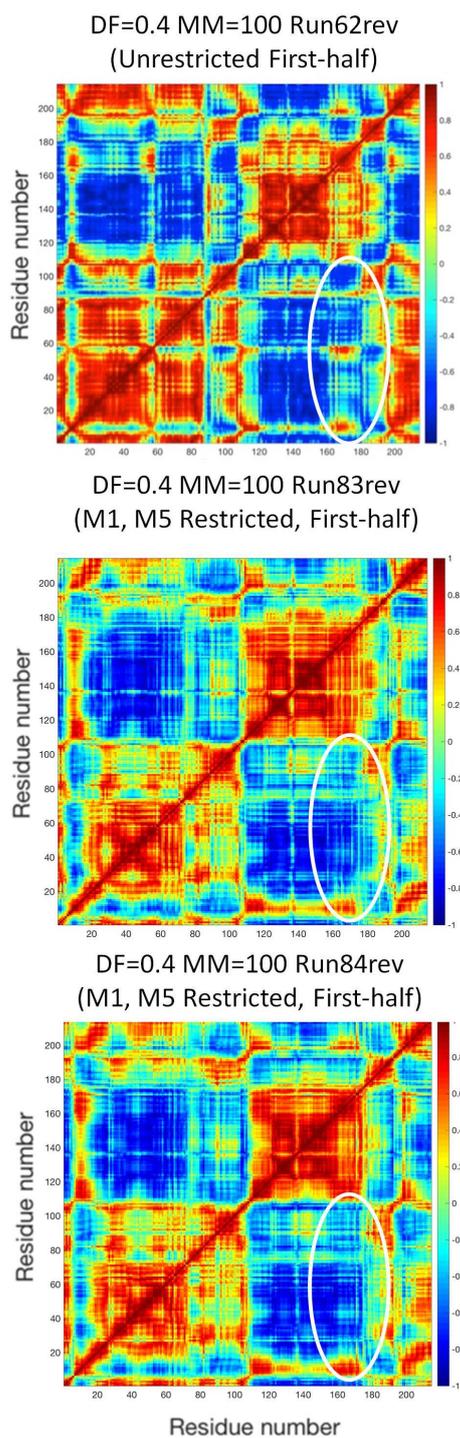


Figure 3.21. Cross-Correlation Plots for AdK reverse runs. First-half of the transition parts of the trajectories (just before RMSD to target reaches a plateau) are considered for the calculations.

the NMP motion in the transition. With the restriction of the first two slowest modes, the correlation of the core region with the LID domain disappears, and the core starts to correlate with the NMP-domain resulting in the precedence of the NMP-motion in the conformational transition (See Figure 3.22). For the reverse case (See Figure 3.23), with the restriction of two slow modes, the correlation of the core region with the LID domain lessens, and the correlation of the core with the NMP-domain increases. The precedence of the domain motions in the transition also changed with the restriction, but the correlations of the core region with LID and NMP domains did not switch that obviously (as observed in forward case). Maybe if we could have restricted the modes mainly responsible for the transition, we would observe a similar switching of the correlations like the forward case.

Restricting the mostly selected modes resulted in a change in the correlation dynamics in both forward (open to closed) and reverse (closed to open) transitions for AdK. Representing the progression of the transition pathways as a function of collective coordinates (if there is any for the studied system) defining the transition is a good way to monitor the transition. In addition to this, GNM eV1-eV2 scatter plots/distributions provide a practical means to view the differences in the pathways, being representative reaction coordinates, complementing the views from the specific collective coordinates. Moreover, where there are no collective coordinates identified to define the transition, GNM eV scatter plots/distributions could provide a means to compare the paths.

Although for the forward restricted case, the complete transition cycle is hindered (an RMSD plateau of 4 Å instead of 0.8 Å of the unrestricted case), for the reverse case the reached RMSD is within 1.3-1.7 Å (compared to an RMSD of 0.8 Å of the unrestricted case). By the ANM-LD method we can identify the main collective modes and regions coordinating the transition process. The procedure allows us to modulate the progression (tune the correlation dynamics) of the transition process by restricting certain modes, either hindering the transition or end up in the generation of alternative transition pathways.

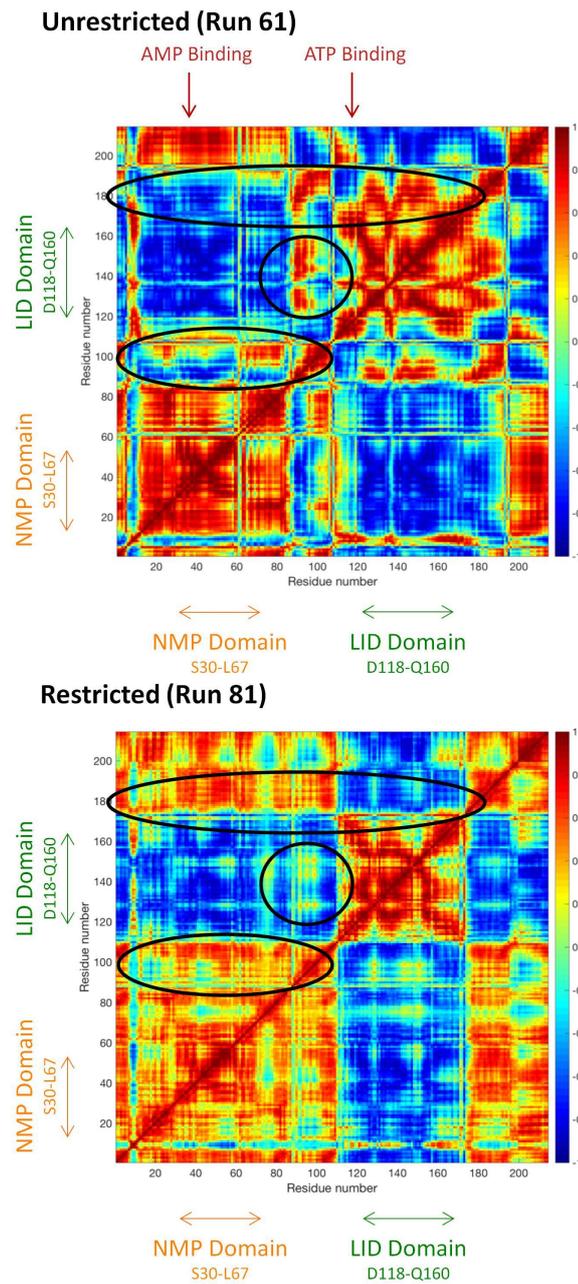


Figure 3.22. Cross-Correlation Plots for AdK forward runs: Unrestricted versus restricted case.

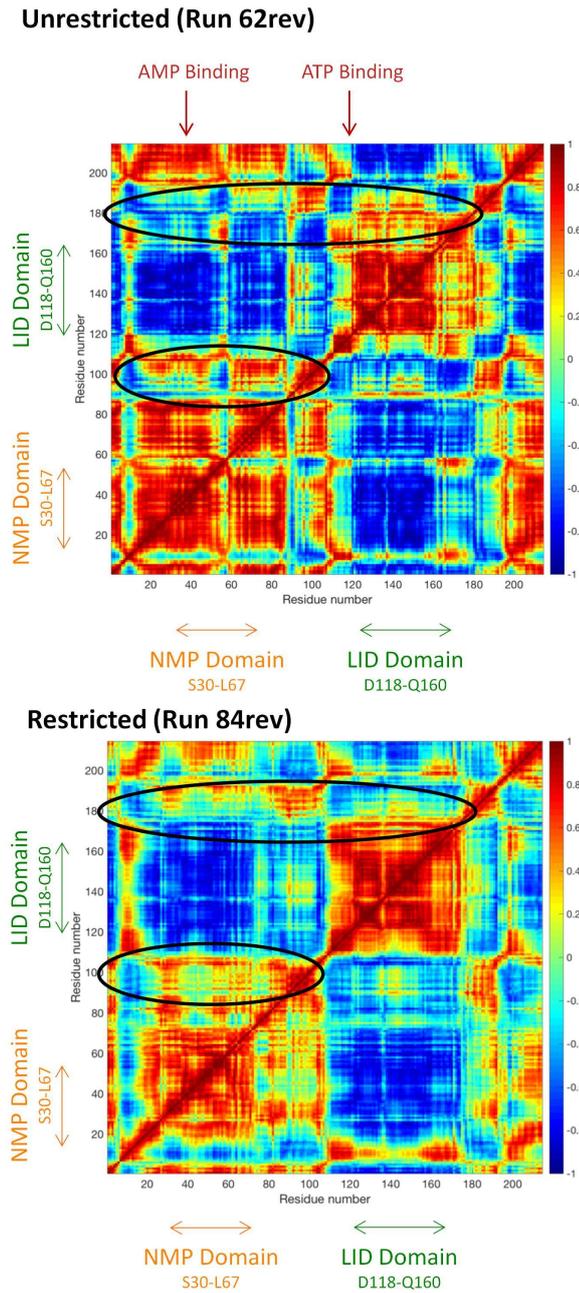


Figure 3.23. Cross-Correlation Plots for AdK reverse runs: Unrestricted versus restricted case.

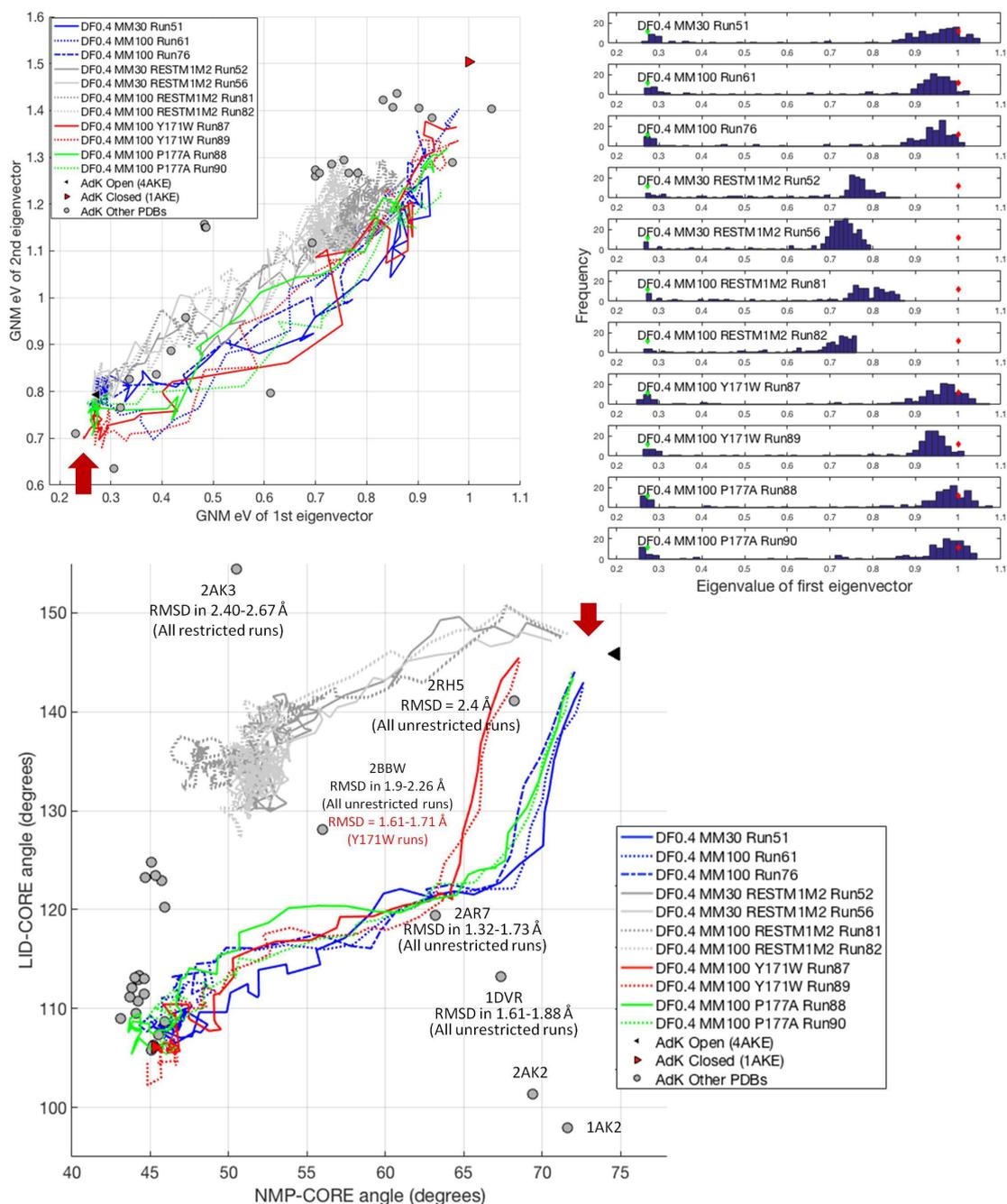


Figure 3.24. GNM eV1 versus eV2 and LID-CORE/NMP-CORE angle plots of mutated AdK (Y171W and P177A) ANM-LD runs. The red lines represent the paths for the two parallel runs of the AdK structure with Y171W mutation (4X8M.pdb). The green lines represent the paths for the two parallel runs of the AdK structure with P177A mutation (4X8H.pdb). with parameter set ($T=310$ K, ANM $R_{cut}=13$ Å, $DF=0.4$ Å, $modemax$, MM=30, 100).

The biological relevance of the generated pathways are validated by the overlap of the created ensembles with the experimental AdK structures in different states, defined by the degree of closeness of LID and NMP domains, and also with the AdK from different species. For example, the experimental structures clustered in one corner of the LID-CORE angle-NMP-CORE angle plot, are observed to be spreaded along the transition pathway according to GNM eV1-eV2 scatter plots, such as the mesophilic and thermophilic AdK homologues (PDB entries 1S3G and 1P3J) [101], which is worth investigating further for the functional relevance.

In Figure 3.24, the GNM eV and LID-CORE/NMP-CORE angle plots are shown including the mutated AdK ANM-LD runs [102]. The red lines represent the paths for the two parallel runs of the AdK structure with Y171W mutation (4X8M.pdb). The green lines represent the paths for the two parallel runs of the AdK structure with P177A mutation (4X8H.pdb). The results show that the assumed transition pathways for the mutated runs show similar traits with the wt runs. This may be due to the quality of the *in silico* mutated target structures, which could be improved by performing moderately long conventional MD simulations before performing ANM-LD runs. The structure of baker's yeast AdK (trapped with an ATP analogue; PDB ID: 1DVR) matches conformations from the unrestricted trajectories. This structure clearly shows that the LID domain is closed and the NMP domain is open. Also, all of the unrestricted trajectories pass close to the human AdK (PDB ID: 2AR7). Here, the Y171W mutated AdK pathway comes closer to 2BBW within an RMSD of 1.61 Å, a value lower than all of the other runs.

3.3.2. Case Study: MalFGK2

Here, ANM-LD simulations are conducted to characterize the large conformational changes and biologically relevant transition pathways in between the metastable states of the ABC transporter protein, maltose transporter, MalFGK2, described by low-frequency modes and the corresponding correlation dynamics. It is observed that a couple of single low-frequency modes are adequate to describe the required conformational changes to go from the inward-facing (IF) state of the protein to the

outward-facing (OF) state.

The maltose transporter system of *Escherichia coli* has long been the well characterized member/prototype for ATP-binding cassette (ABC) transporter studies. Maltose binding protein (MBP) binds maltose in the periplasm and the closed/stabilized form of MBP interacts with the resting state MalFGK2 (IF state, 3FH6.pdb) [91] bringing NBDs closer (an intermediate state, called pre-translocation state, 3PV0.pdb) [92]. This is followed by ATP binding/interaction leading the closure of MalK, opening of MBP orienting the transmembrane domains (TMd) toward the periplasm to release the maltose to the TMd maltose-binding site and positioning ATP sites for hydrolysis (OF state, 3PUY.pdb) [92]. After the hydrolysis of ATP, the TMd will likely reorient toward the cytoplasm and release maltose into the cell, since ADP is not expected to stabilize the closed dimer alone according to isolated NBD studies (post-hydrolysis state with ADP, unknown structure), summarizing the alternating access model [92]. The functional sites of MalFGK2 are presented on the OF structure in Figure 3.25. Since the nucleotide binding domain (NBD) is a homodimer, the functional motifs involved in the binding of one ATP (binding chain A) is shown (Figure 3.25a). The motifs for the chain B are the symmetric of given sites. Since the TMd of maltose transporter is a heterodimer, the functional sites are given for both subunits F and G. (Figure 3.25b) ATP Binding Site 1 and maltose binding sites of MalFGK2 are presented on the structure with colors described in (Figure 3.25a).

In this study, several parallel IF-OF and OF-IF transition simulations using X-ray crystal structures, 3FH6.pdb (IF) [91] and 3PUY.pdb (OF) [92] are performed, including both TMds and NBDs, excluding ATPs and maltose. Among the many parallel ANM-LD runs performed, the best final RMSD values obtained as a result of the conformational transitions are 3.83 and 3.38 Å for IF-OF and OF-IF cases, respectively, the initial RMSD being 6.85 Å between the IF and OF states. Although, many parallel runs have been performed, here the results of the case runs selected for each transition are presented (Figures 3.26-3.29). The slowest ANM modes 3 and 8 are the mostly overlapping modes for the IF-OF transition, whereas for the OF-IF case modes 4 and 8 are the mostly overlapping modes with the initial-target state difference

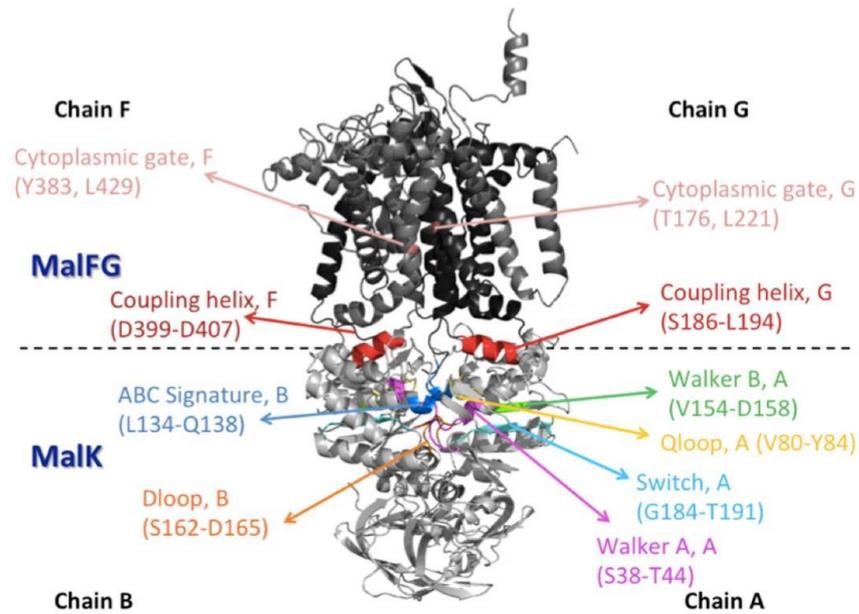
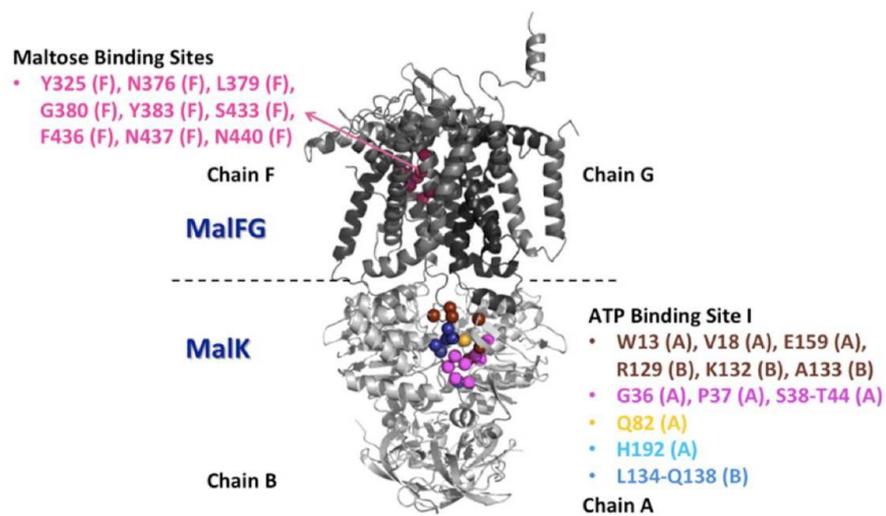
a**b**

Figure 3.25. Functional sites of MalFGK2 are presented on the OF structure (PDB code: 3PUY).

vector. GNM correspondents of mostly selected ANM modes were determined based on the mode shape match between ANM and GNM modes to find hinge residues with potential importance for the transition behavior. The corresponding GNM modes are slowest 4,6, and 3,4 for the IF-OF and OF-IF cases, respectively. For the IF-OF case, contribution to hinge axis in both slowest GNM modes 4 and 6, any hinge mutation associated with these modes could be considered to affect the transition. Slow mode hinges of GNM modes 2 (selected along the transition part, but not so popular) and GNM modes 3, 4 (popular modes) that were predicted to be used in the transition process. Cytoplasmic gate is located on the hinge axis of GNM mode 3 (OF state) with maltose binding site, coupling helices, Walker A/B, Q loop, ABC Signature, D loop, ATP-binding sites I/II and MalE binding sites on OF structure, excluding the switch region. Coupling helices share the hinge axes with the other functional regions for GNM modes 3 and 6, for the OF and IF states, respectively. For both cases, most of the ATP-binding sites are included in the hinge axes.

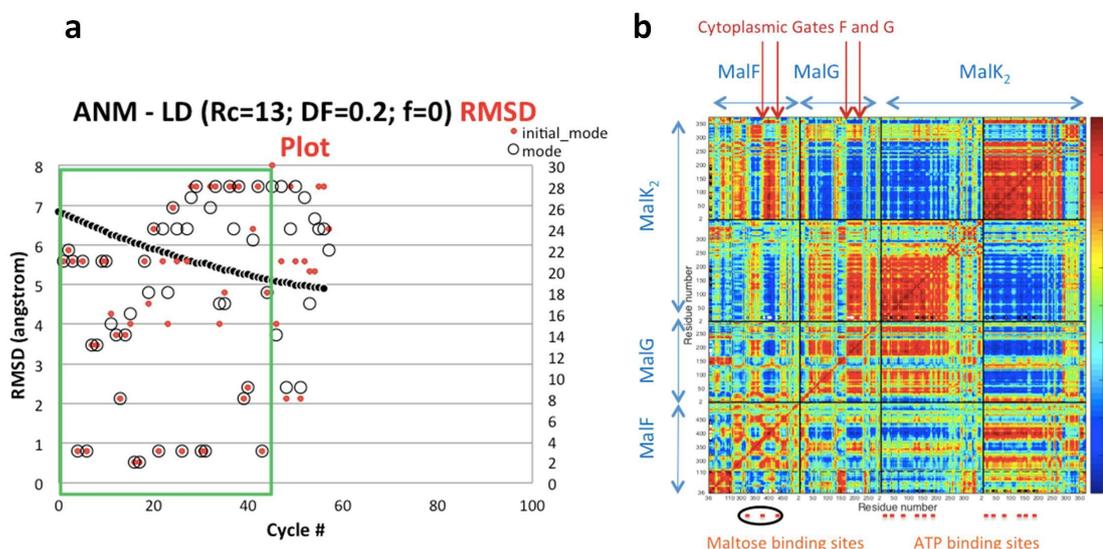


Figure 3.26. MalFGK2 IF→OF transition case run RMSD plot (a) and transition cross-correlation map (b) of first 40 cycles.

In Figure 3.26a RMSDs between the conformation sampled at each cycle and final structure OF from IF→OF transition trajectory of case run. Hollow circles indicate

the selected ANM modes at each cycle while pink filled circles are the corresponding mode in the initial IF structure. Green rectangle corresponds to the first 40 cycles of the transition that is further analyzed.

In Figure 3.27a, MalFGK2 IF structure (3FH6) is colored accordingly with the average cross-correlations of the maltose binding sites (Chain F Y325, N376, L379, G380, Y383, S433, F436, N437, N440) with the rest of the structure for the first part of the IF→OF transitions; Runs 19, 20, 25, 29 and 31. In Figure 3.27b MalFGK2 IF structure (3FH6) is colored accordingly with the average cross-correlations of W13, V18 (A) of ATP Binding Site 1 (left) and W13, V18 (B) of ATP Binding Site 2 (right) with the rest of the structure for the first parts of the IF→OF transition run 29 (case run). ATP Binding Site1 W13, V18 (A) is correlated with right part of the Chain G and MaleE binding sites. ATP Binding Site2 W13, V18 (B) is correlated with left part of the Chain F (maltose binding sites) and Chain G and some of the MaleE binding sites.

In Figure 3.28a RMSDs between the conformation sampled at each cycle and final structure IF from OF→IF transition trajectory of case run. Hollow circles indicate the selected ANM modes at each cycle while pink filled circles are the corresponding mode in the initial IF structure. Green rectangle corresponds to the first 20 cycles of the transition that is further analyzed.

In Figure 3.29a, MalFGK2 OF structure (3PUY) is colored accordingly with the average cross-correlations of the maltose binding sites (Chain F Y325, N376, L379, G380, Y383, S433, F436, N437, N440) with the rest of the structure for the first part of the OF→IF transitions; runs 21, 27, 28, 30 and 32. In Figure 3.29b MalFGK2 OF structure (3PUY) is colored accordingly with the average cross-correlations of W13, V18 (A) of ATP Binding Site 1 (left) and W13, V18 (B) of ATP Binding Site 2 (right) with the rest of the structure for the first parts of the OF→IF transition run 27 (case run). ATP Binding Site1 W13, V18 (A) is correlated with right part of the Chain G and MaleE binding sites on the left part. ATP Binding Site2 W13, V18 (B) is correlated with right part of Chain F and Chain G, and MaleE binding sites on right part.

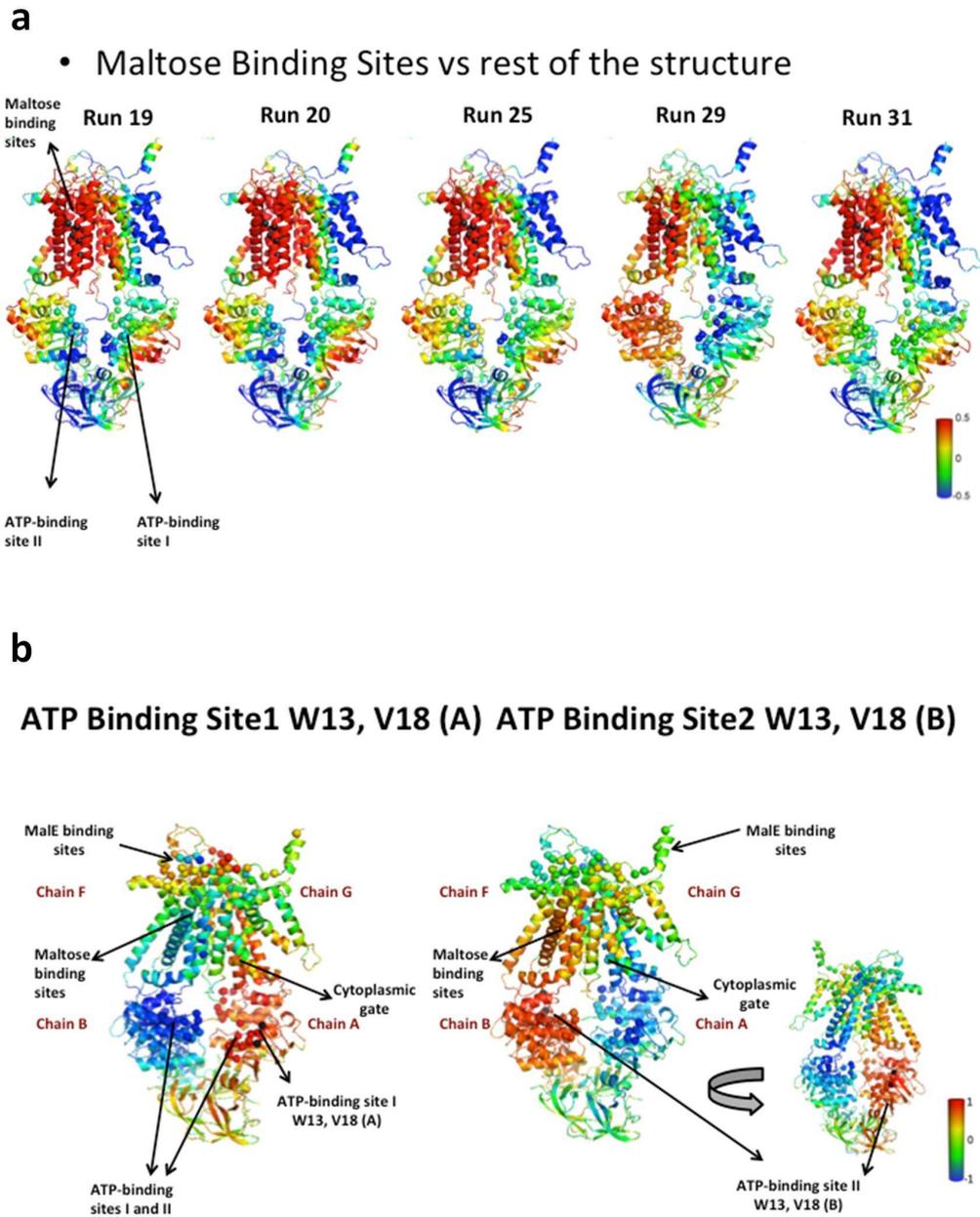


Figure 3.27. MalFGK2 IF→OF transition cross-correlations (depicted on Figure 3.26b) color coded on the ribbon structure. Correlation of maltose binding sites (a) and ATP-binding sites (b) with the rest of MalFGK2 structure are shown on ribbon structures.

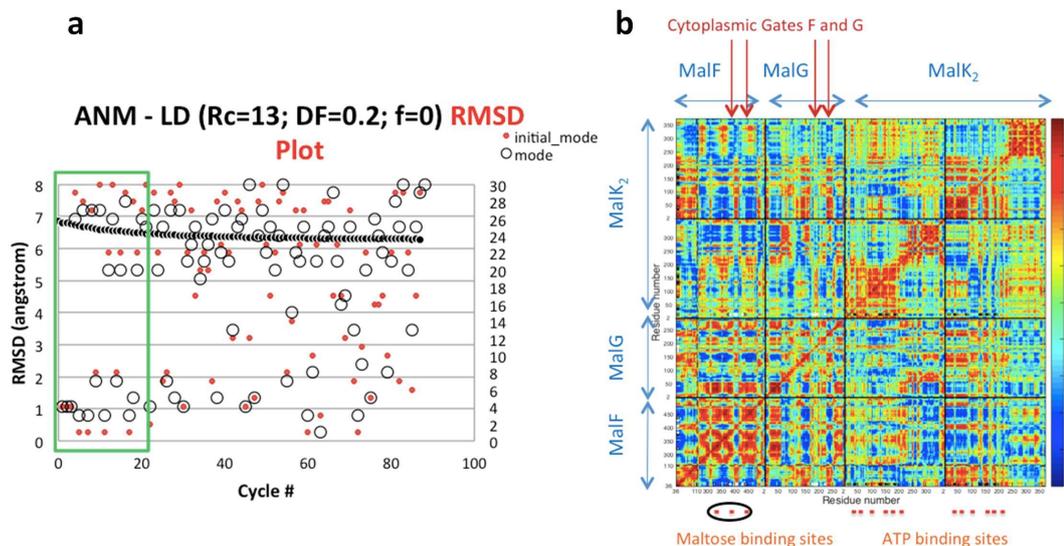


Figure 3.28. MalFGK2 OF→IF transition case run RMSD plot (a) and transition cross-correlation map (b) of first 20 cycles.

3.4. Conclusions

Sampling the conformational transition and the reactive pathways between given functional states (initial/target) of a protein is a challenging problem. Utilizing the intrinsic dynamics information coded in the 3D fold of proteins is valuable information in terms of their functional dynamics. Elastic Network Models, GNM and ANM provide this information in a computationally efficient way. Here, we have developed a hybrid methodology where we have utilized collective ANM modes of motion encoded by its fold to simulate the conformational transition from an initial towards a target conformation, taking advantage of the evaluation of the interactions and energetics of the system via short cycles of all-atom implicit LD simulations in between. The methodology is tested for a set of proteins, but the results are presented for the prototype protein AdK where we have obtained successful results when we assessed its ability to sample the experimentally available structures and the converged RMSD values. Also, the results for the membrane protein MalFGK2 are presented, where the global dynamics underlying the activation of the protein were explored.

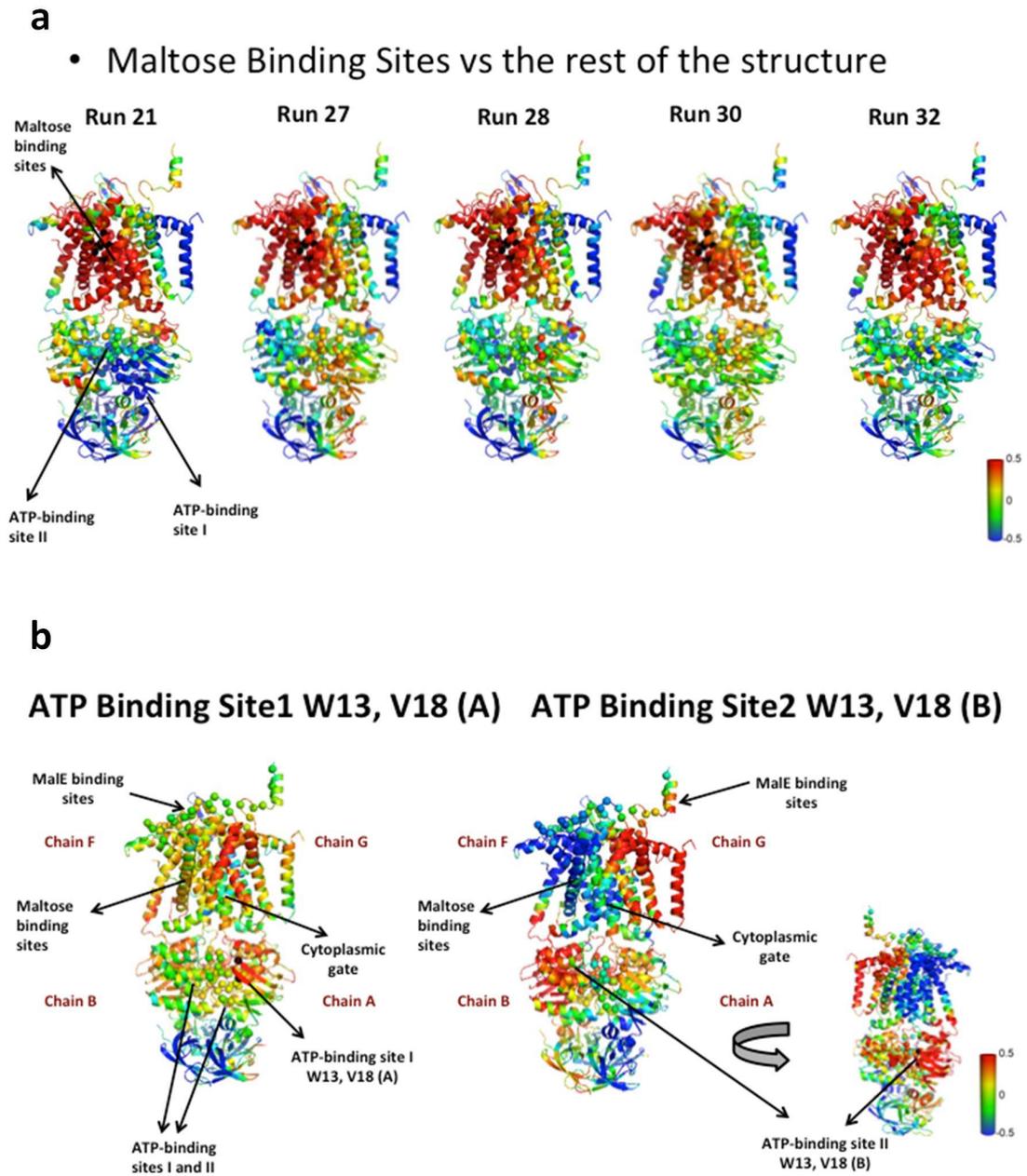


Figure 3.29. MalFGK2 IF→OF transition cross-correlations (depicted on Figure 3.28b) color coded on the ribbon structure. Correlation of maltose binding sites (a) and ATP-binding sites (b) with the rest of MalFGK2 structure are shown on ribbon structures.

Here, we have identified the mechanistic determinants/underpinnings in analyzing and controlling the transition pathways. The developed methodology would allow the investigation of any perturbation application, such as added external force, in silico mutations or ligand binding, which has been exemplified with the study of Hsp90 in the presence of ATP and the inhibitor Geldanamycin in Chapter 4.

Possible future investigations would be exploring the effect of mutations and different ligands on the transition pathways. Comparing the evolution of the hinges obtained along the transition pathway by changing conditions could determine the crucial residues taking part in the functional transition. Changelessness or change in the hinge residues at each cycle would have implications on the mechanism of conformational change leading to the target state, especially in the investigations of the important residues controlling the transition and the effect of mutations.

4. CONFORMATIONAL TRANSITION AND THE EFFECT OF TEMPERATURE ON BINDING ENERGY LANDSCAPE OF HSP90–GELDANAMYCIN COMPLEX

The open and closed X-ray crystal structures of the molecular chaperone Hsp90 suggest that the functioning of the molecule involves large conformational changes coupled to the ATPase cycle. Conventional simulation methods or experimental methods alone are of limited use in the study of these large conformational changes due to the inadequacy of the current methodologies. In this work we have combined enhanced molecular simulations, hybrid Anisotropic Network Model-Langevin Dynamics (ANM-LD), with dynamic force spectroscopy approach using atomic force microscopy (AFM-DFS) to characterize the ensembles in the transition conformational pathways between functional states of Hsp90 and the binding dynamics of Hsp90 with the inhibitor geldanamycin. Here, we have investigated the binding/unbinding behavior of the Hsp90 inhibitor geldanamycin-GDM and the effect of temperature on this interaction. Global binding behaviour was observed to couple to the local binding behavior of the ATP/GDM, which bind to the same cavity of Hsp90, but with totally different residue contacts. The effects of this local binding site were observed to couple to the global dynamics behavior and appeared to change with temperature, in agreement with the dynamic correlation maps of the assumed conformational transition pathways of the ANM-LD simulations.

4.1. Introduction

Hsp90 makes up 1-2% of the total cellular protein and its expression was found to be critical for cell survival under most conditions in eukaryotes. Hsp90 is a dimeric molecular chaperone involved in the activation and stabilization of numerous client proteins including kinases, transcription factors and steroid receptors. It is a highly dynamic and flexible molecule that can adopt a wide variety of structurally distinct states by the intrinsic ATPase activity as the driving force. Unlike other chaperones,

ATP binding and hydrolysis does not dictate the specific conformational states, but shift the equilibria between pre-existing set of conformational states. For bacterial, yeast and human Hsp90 there is a conserved three-state (Apo-ATP-ADP) conformational cycle, however the equilibria between states are species specific. In eukaryotes, co-chaperones as well as post-translational modifications regulate the *in vivo* dynamic behavior of Hsp90 by shifting conformational equilibria, affecting the kinetics of structural changes and ATP hydrolysis. Regulation by co-chaperones is an absent mechanism in the bacterial analogue (HtpG) [103]. Hsp90 can achieve broad specificity with client diversity in a size range of 14-290 kDa with no shared sequence or structural characteristics, likely due to its conformational flexibility. Although ATPase function of Hsp90 plays a major role in recognizing/activating its clients, the determinants of client recognition and required conformational changes associated with Hsp90 function remains unclear. What is the role of the intrinsic conformational flexibility of Hsp90 in this molecular recognition? What are the contributions of co-chaperons? [104]

There is 40% sequence identity between human and *E. coli* (HtpG) and 60% between human and yeast. There are four human isoforms, two cytosolic (Hsp90 α , Hsp90 β), one mitochondrial (Trap1) and one specific to the endoplasmic reticulum (Grp94). Hsp90 exists as a homodimer with three domains (N-terminal nucleotide binding domain-NTD, middle domain-Md, C-terminal dimerization domain-CTD) per monomer. Some of the other important functional regions are the (-) charged linker (residues 210-265) connecting NTD with Md and the C-terminal MEEVD motif present in eukaryotic cytosolic Hsp90s (See Figure 4.1 for residue numbers for the domains). Jahn *et al.* has used a combination of single-molecule techniques and biochemical methods in order to study the unresolved charged linker regions of Hsp90 protein. These linkers are known to regulate the movement of the three domains of the protein during the chaperone cycle [105].

The newly discovered mitochondrial Hsp90 (TRAP1) crystal structure contains a strap region -that is only present in higher eukaryotes- which stabilizes the closed state and seem to have a role in the protein's unusual temperature dependence in ATPase rates. Also, a thermally sensitive kinetic barrier is discovered to exist between the

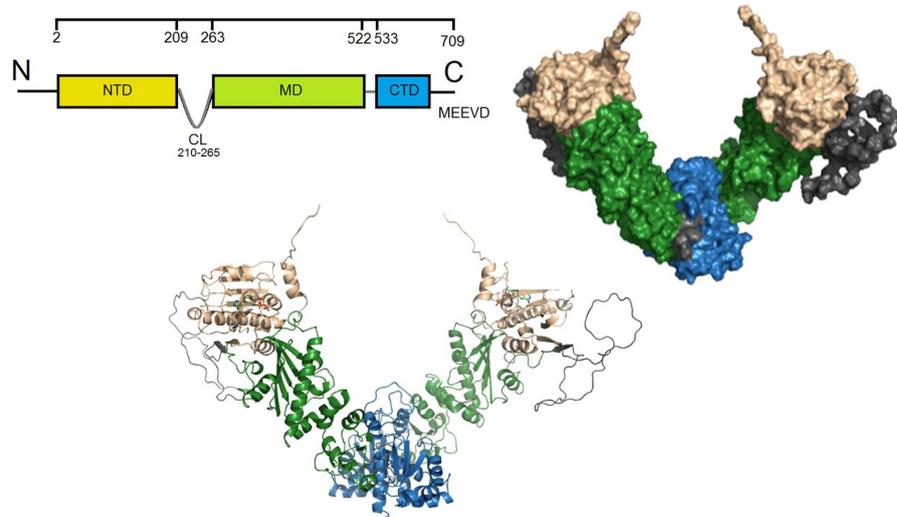


Figure 4.1. Hsp90 domain architecture for yeast Hsp90. Nucleotide binding domain (NBD), middle domain (Md) and C-terminal domain (CTD) are denoted by yellow, green and blue, respectively. The linkers connecting the domains are shown with gray.

apo/open and holo/closed conformations [106].

Full-length Hsp90 crystal structures are difficult to determine due to its structural flexibility. There is a highly open apo conformation, as well as a more closed apo conformation (resembling Grp94) populated according to the pH of the environment. These apo conformations largely differ by rigid body rotation at the interface between Md and CTD, causing NTDs to undergo 50 Å translation and 50° rotation. This extreme apo state conformational heterogeneity is universal to all homologs, implying that the flexibility of Hsp90 is functionally important. Conformational equilibrium of Hsp90 plays a central role in client interactions, differing according to the client requirements of each organism. The large size of the structure in combination with its conformational diversity implies that it can adopt a combinatorial set of binding interfaces/surfaces and chaperone geometries in order to interact with its structurally diverse clients and co-chaperones. The effect of nucleotide binding and hydrolysis adds a layer of complexity.

In Morra *et al.*'s molecular dynamics study [107], they show that ATP- and ADP-bound Hsp90 has different sets of residues in their communication network between NTD and CTD. These different pathways may give information about the different conformations of Hsp90 in the presence of ATP or ADP. Mutational studies could help in the elucidation of each network's role in establishing Hsp90's conformation.

The role of Hsp90 in diverse diseases like cancer, Alzheimer's, vascular disease and viral diseases, makes it a promising drug target, but the full utility of Hsp90 inhibitors is limited by the lack of molecular level information about co-chaperone/client binding and the role of ATP hydrolysis.

Investigators established Hsp90's druggability using the natural products geldanamycin (which is a benzoquinone ansamycin) and radicicol, which mimic the unusual ATP structure adopted (bergerat fold) in the chaperone's N-termini nucleotide binding pocket that cause;

- (i) potent and selective blockade of ATP binding and hydrolysis,
- (ii) inhibit chaperone function and,
- (iii) deplete oncogenic clients through ubiquitin-mediated proteosomal degradation.

17 agents are in clinical trials that avoid some of the drawbacks of the first-generation inhibitors. There are approaches to enhance the effectiveness of the Hsp90 inhibitors, which will progress by both increased understanding of the basic structure-function biology of Hsp90 and translating this knowledge into anticancer drug therapies. Future success will most likely come from the use of chemically optimized forms of the current Hsp90 inhibitors [108].

A second druggable site (which could represent a previously unrecognized ADP-ATP binding site) has been identified in the C-termini of the protein. Coumarin antibiotics such as novobiocin, are the prototypic inhibitors that interact with this site. To date, no Hsp90 inhibitor has received marketing approval.

In order to fulfill its function, Hsp90 undergoes large conformational changes. In the inactive form of Hsp90, the arms (the monomers interact at their C-termini to form a V-shape) are open (Figure 4.2a). These flexible arms are closed when ATP binds the structure, transforming it to the active form (Figure 4.2b). Hsp90 analogs belonging to several organisms has semi-closed intermediate forms in the PDB.

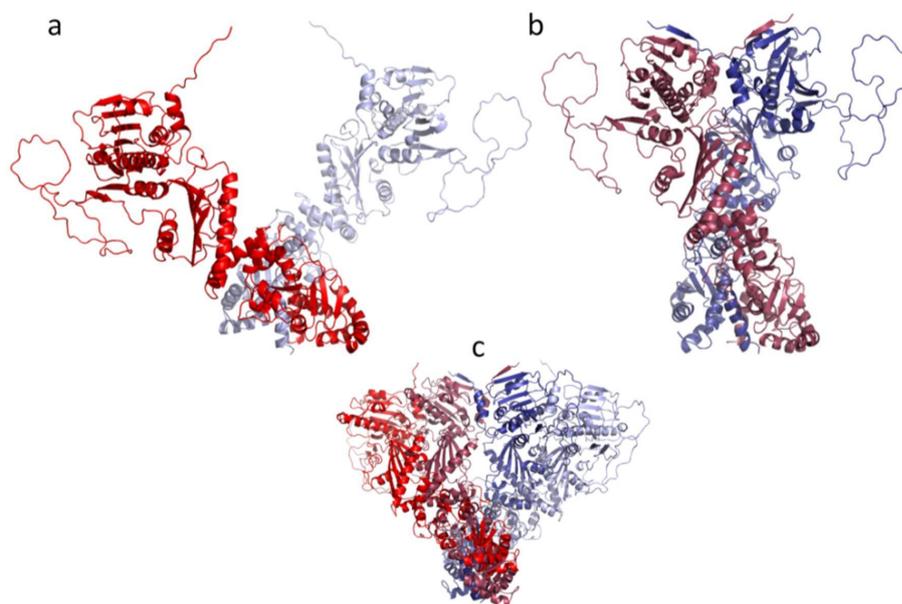


Figure 4.2. (a) The open conformation of Hsp90 modeled using *E. coli* open, inactive (PDB ID: 2IOQ) and yeast closed, active (PDB ID: 2CG9) conformations, (b) yeast closed X-ray crystal structure (PDB ID: 2CG9). (c) two structures in (a) and (b) are shown as superposed (RMSD=24 Å).

4.2. Materials and Methods

4.2.1. Preparation of Molecules and AFM Setup

Human Hsp90 molecules (produced in HeLa cells) were purchased as aqueous solution with a concentration of 0.73 mg/ml in 50mM TRIS-HCl at pH 7.5, containing 1.0mM DTT and 100mM NaCl solution from Enzo Life Sciences. Aqueous solution of Hsp90 was diluted to 100 μ g/ml with sufficient amount of phosphate buffered saline PBS. Biotinylated polyethylene glycol (PEG) attached geldanamycin (GDM) molecules

were purchased as lyophilized crystalline solid from Enzo Life Sciences and reconstituted in dimethyl-sulfoxide (DMSO) to have final concentration of 100 $\mu\text{g}/\text{ml}$. Streptavidin molecules were purchased as lyophilized powder from Sigma-Aldrich (St. Louis, MO, USA) and reconstituted in PBS to have final concentration of 100 $\mu\text{g}/\text{ml}$. Mica surface was purchased from Ted Pella, Inc.. Biotinylated PEG attached Si₃N₄ AFM probes (MLCT. BIO, Novascan, Ames, IA, USA) were used in the experiments. The nominal spring constant of the cantilever used in the experiment is 0.01 N/m. A commercial AFM system (Dimension Edge, Bruker Nano, Santa Barbara, CA, USA) was used in the experiments. Closed Loop Temperature Controller, (12mW, ThorLabs) was used to set and also keep the temperature at desired values.

4.2.2. Immobilization of Hsp90 Molecules and Functionalization of AFM Tips with Geldanamycin

Prior to AFM data collection, Hsp90 molecules were immobilized on a mica surface via physical adsorption by incubating 200 μl of previously prepared Hsp90 aqueous solution on mica surface for 20 minutes. Then the mica surface was gently washed with deionized water to remove loosely adhered molecules. The mica surface with immobilized molecules was kept hydrated via PBS prior to data collection.

To functionalize biotinylated Si₃N₄ AFM tips with biotinylated PEG attached GDM molecules, streptavidin was used as mediator. Biotinylated cantilevers were first incubated with 20 μl streptavidin (100 $\mu\text{g}/\text{ml}$ in PBS) for 20 minutes at room temperature. The streptavidin functionalized tips were then dipped into the 60 μl of GDM (100 $\mu\text{g}/\text{ml}$ in DMSO, purged with helium) and were incubated for 15 minutes at room temperature. The GDM functionalized AFM tips were used immediately in force measurements after rinsing with PBS to remove the loosely bound GDM molecules. Figure 4.4a shows the multi-step functionalization of AFM tip and surface in this experiment.

4.2.3. AFM Force Measurements and Analysis

The functionalized cantilevers were calibrated using the thermal noise method [109]. We performed the experiment by actuating the AFM cantilever repeatedly on top of a Hsp90 functionalized mica surface using the piezotube actuator of the AFM system. The deflection signal was recorded during the experiment. We actuated the piezotube using triangular waves at six different translational speeds (0.5, 1, 2.5, 5, 10, 25 $\mu\text{m/s}$) during the experiment. At each speed, we have collected 477 and 353 force curves to obtain statistically significant data at 27 °C (300 K) and 37 °C (310 K), respectively. Figure 4.4b shows a typical force curve (representing both approach and retract) indicating an unbinding event of 75 pN between a single pair of molecules and Figure 4.4c shows another force curve with no adhesion/rupture event.

The effective loading rates were estimated based on the slope of the force-time curve immediately before the rupture point in each approach/retract cycle. The effective loading rates obtained from the experiment were equally quantized in the logarithmic space to get uniform number of most probable rupture force versus logarithm of loading rates relation. All the detected rupture forces were grouped according to the quantized loading rate intervals. The most probable rupture force of the Hsp90-GDM complex at a given loading rate interval was determined by calculating the median of the force histograms (Figure 4.5a and b). The dependence between the unbinding force and the loading rate can yield the estimation of basic parameters (the unbinding rate constant k_{off} when the bond is not loaded by an external force and the distance from the bound state of the activated state r_o) of one-dimensional binding energy landscape (BEL) of protein complexes [19–21, 110] as illustrated schematically in Figure 4.4d for single global minimum and Figure 4.4e for one global, one local minimum according to the following Bell’s model:

$$f^* = \frac{k_B T}{r_o} \ln \left(\frac{r_f}{k_B T / r_o \times k_{off}} \right) \quad (4.1)$$

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

The most probable rupture force, f^* , versus the natural logarithm of the loading rate, $\ln(r_f)$, is plotted and the Bell parameters (k_{off} and r_0), which are characteristic values of the investigated molecular pair, were estimated from the slope, $\frac{k_B T}{r_0}$, and the intercept, $\frac{k_B T}{r_0} k_{off}$, of the fitted linear curves (See Equation 4.2).

$$f^* = \frac{k_B T}{r_0} \left(\ln r_f - \ln \frac{k_B T}{r_0} k_{off} \right) \quad (4.2)$$

4.2.4. Structures

Since both closed and open conformations of a single organism did not exist in the PDB, the open conformation of Hsp90 modeled using *E. coli* open/inactive (PDB ID: 2IOQ) [111] and yeast closed ATP bound chaperone complex (PDB ID: 2CG9) [112] conformations (See Figure 4.2). The modeled open conformation is named as 2CG9o throughout the text. The PDB structure of the closed form is in complex with co-chaperone p23/Sba1 and ATP molecule, they are removed for the simulations. The missing residues V217-K261, D330-K338 and L598-K610 in the structures are modeled using ModWeb/ModLoop servers [93,94]. While modeling the open form (with missing long stretches of loops) of this large molecule, the steric clash problem encountered is solved by refining the protein structure on Chiron server [95]. ATP binding sites are listed in PDBsum as follows: E33, N37, A38, A41, D79, M84, N92, S99, G100, G118, Q119, F120, G121, V122, G123, F124, T171, R380. GDM binding sites are listed in PDBsum as follows: N51, S52, D54, A55, K58, D93, I96, M98, D102, N106, L107, K112, G135, V136, G137, F138, T184 [113]. There are no common binding sites for ATP and GDM.

The ATP- and GDM-bound Hsp90 complexes are modeled using 2CG9o and the GDM-bound N-terminal domain structure of Hsp90 (PDB ID: 1YET) [114]. The Amber parameter file for the drug molecule GDM is prepared using the Gaussian 09 program [115], and the ATP parameter file is obtained from an Amber parameter database [116] (See Figure 4.3).

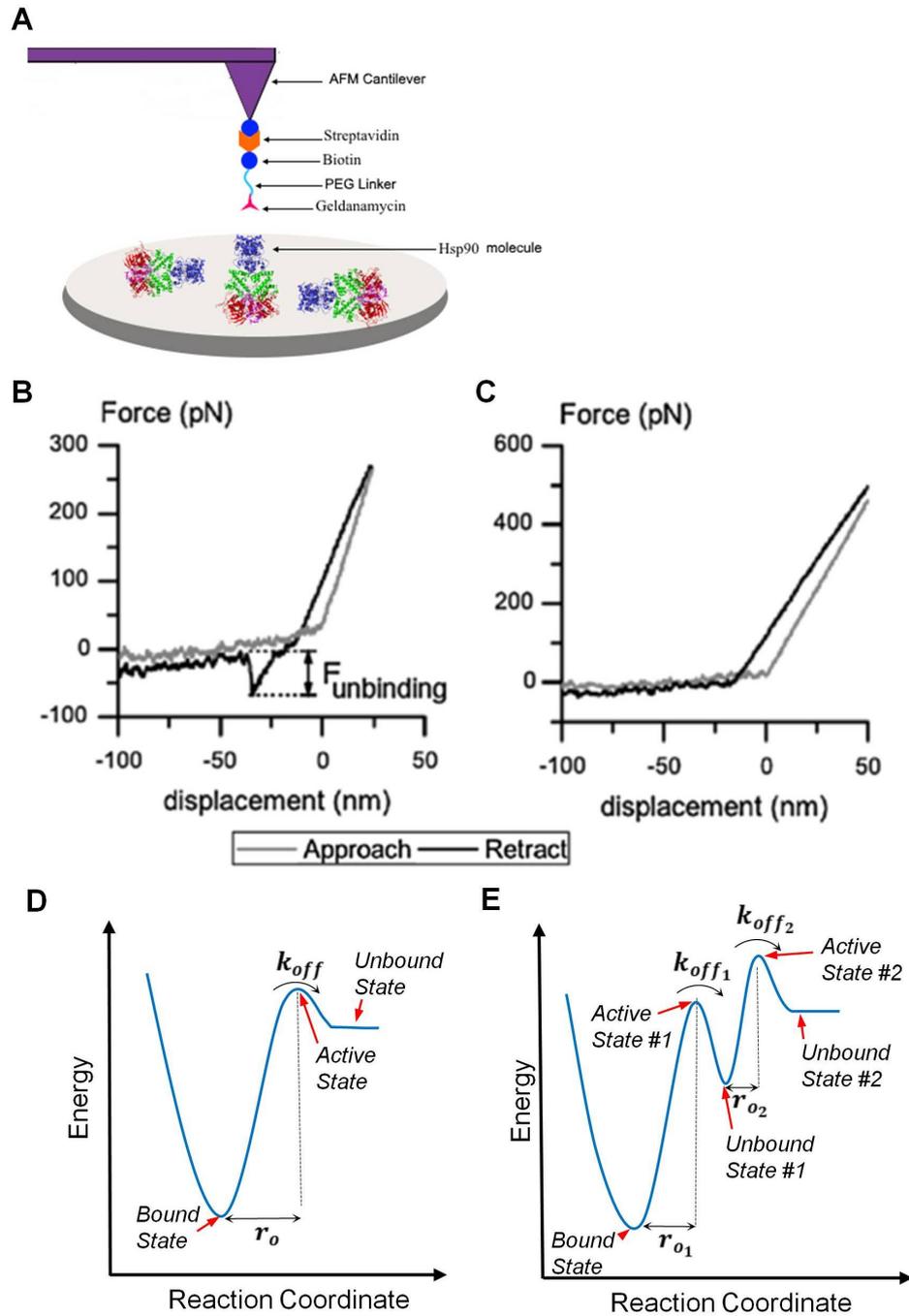


Figure 4.4. (a) Schematic representation of Hsp90-Geldanamycin AFM pulling experiment setup. (b) A typical force curve exhibiting an unbinding event with a force strength ($F_{unbinding}$). (c) Another force curve indicating no adhesion/rupture events. (d) Schematic representation of one-dimensional binding energy landscape with single global minimum. (e) Schematic representation of one-dimensional binding energy landscape with one global and one local minimum.

4.2.5. ANM-LD Simulation Protocol

We have performed several parallel open-closed and closed-open conformational transition simulations using the modeled open structure 2CG9o [111,112] and the closed X-ray crystal structure 2CG9.pdb [112], by ANM-LD methodology at 300 K and 310 K. ANM-LD hybrid methodology simulates the conformational transition of a structure starting from a given initial state to a target one by distorting the initial structure along the ANM modes mostly overlapping with the initial-target coordinate difference vector followed by short cycles of minimization and implicit LD simulations. Details of the ANM-LD methodology can be found in Chapter 2. The simulated systems that are presented in this chapter are given in Table 4.1.

Amber 11 [81] biomolecular simulation programs were used for the minimization runs and LD simulations. The Amber ff03 [80] force field parameter set was used for the proteins/ions. Initially, energy minimization was performed using 50 cycles of steepest descent algorithm, followed by 1000 steps of conjugate gradient method. Initial velocities were selected at random from the Maxwell-Boltzmann distribution at temperatures, 300 K and 310 K. The Langevin thermostat [84] and Berendsen barostat [85] was used for the NPT ensemble ($T=300$ or 310 K, $P=1$ bar) with a time step of 2 fs. The SHAKE algorithm [59] was used as the bond constraints for all bonds involving hydrogens to eliminate the high frequency bond vibrations. A cutoff distance of 99 Å was used for the nonbonded interactions.

4.3. Results and Discussion

Geldanamycin is a known inhibitor of Hsp90 protein which bounds to Hsp90 and prevents ATP binding, however, the detailed dynamics of how geldanamycin obstructs the transition is unknown. Therefore, we have conducted AFM pulling experiments to measure the magnitude of binding force, to determine the force distribution/localization and the binding dynamics of the Hsp90-GDM complex, as well as the effect of temperature on their association. Along, ANM-LD conformational transition simulations were performed to investigate the effect of absence and presence of ATP, as well as

Table 4.1. Hsp90: Summary of ANM-LD simulation results.

| Direction | T (K) | Mode Max | RMSD initial (Å) | RMSD final* (Å) | Run Alias |
|-------------------|-------|-------------|------------------------|-----------------------|-----------|
| Hsp90 Apo Forward | 300 | 100 | 24.0 | 2.81 | Run11 |
| Hsp90 Apo Forward | 310 | 100 | 24.0 | 2.91 | Run12 |
| Hsp90 Apo Forward | 300 | 30 | 24.0 | 3.86 | Run18 |
| Hsp90 Apo Forward | 310 | 30 | 24.0 | 4.28 | Run19 |
| Hsp90 Apo Reverse | 300 | 100 | 24.0 | 5.54 | Run15rev |
| Hsp90 Apo Reverse | 310 | 100 | 24.0 | 5.69 | Run14rev |
| Hsp90 Apo Reverse | 300 | 30 | 24.0 | 4.76 | Run17rev |
| Hsp90 Apo Reverse | 310 | 30 | 24.0 | 6.34 | Run16rev |
| Hsp90 ATP Forward | 300 | 100 | 24.4 | 2.88 | ATP Run12 |
| Hsp90 ATP Forward | 310 | 100 | 24.4 | 2.97 | ATP Run13 |
| Hsp90 ATP Forward | 300 | 30 | 24.4 | 4.28 | ATP Run1 |
| Hsp90 ATP Forward | 310 | 30 | 24.4 | 4.16 | ATP Run14 |
| Hsp90 GDM Forward | 300 | 100 | 24.4 | 2.92 | GDM Run4 |
| Hsp90 GDM Forward | 310 | 100 | 24.4 | 2.88 | GDM Run5 |
| Hsp90 GDM Forward | 300 | 30 | 24.4 | 4.28 | GDM Run1 |
| Hsp90 GDM Forward | 310 | 30 | 24.4 | 4.27 | GDM Run6 |

* The best result observed is given for each system for the parameter set: (R_{cut} = 13 Å, $stepmin$ =1000, $stepsim$ =100, γ =50 ps^{-1}).

the inhibitor GDM, on the global dynamics of Hsp90 at the two temperatures that we have conducted the AFM pulling experiments.

4.3.1. Single-Molecule Force Measurements: The Effect of Temperature on the Energy Landscape

To investigate the effect of temperature on the interaction forces between Hsp90 and geldanamycin, as well as the parameters of kinetics and binding energy landscape, force measurements were carried out at two different temperatures. The adhesion force between Hsp90 and geldanamycin inhibitor was extracted from the collected data. Figure 4.5a and b show the force histograms of Hsp90-Geldanamycin bond strength at different loading rates at room (27 ° C, 300 K) and body (37 ° C, 310 K) temperatures. We observed a large portion of moderate adhesive interactions ranging from 0.01 to 0.4 nN in both temperatures. The distribution of the unbinding forces observed at 300 K are similar to 310 K, as well as the extracted values of the most probable rupture forces. This implies that the temperature does not affect the unbinding forces of Hsp90-Geldanamycin complex.

Hsp90-Geldanamycin dissociation reaction exhibits bimodal distribution in both temperatures indicating the existence of two distinct stable binding states of the complex (a low-strength and a high-strength binding). This two state behavior may imply two stable states in the conformational ensemble of Hsp90.

4.3.2. Binding Energy Landscape of Hsp90-Geldanamycin Complex

The most probable rupture force vs natural logarithm of loading rate exhibited single linear regime for low- and high-strength states at both 300 K and 310 K (Figure 4.6a and b). These findings indicate that, at both temperatures the disassociation process of Hsp90-Geldanamycin complex passes a single energy barrier while going from the bound state to the unbound state, in both conformational ensembles. To obtain the disassociation rate and the position of the energy barrier for Hsp90-Geldanamycin interaction, each of the observed regimes were fitted by the Bell's model as described

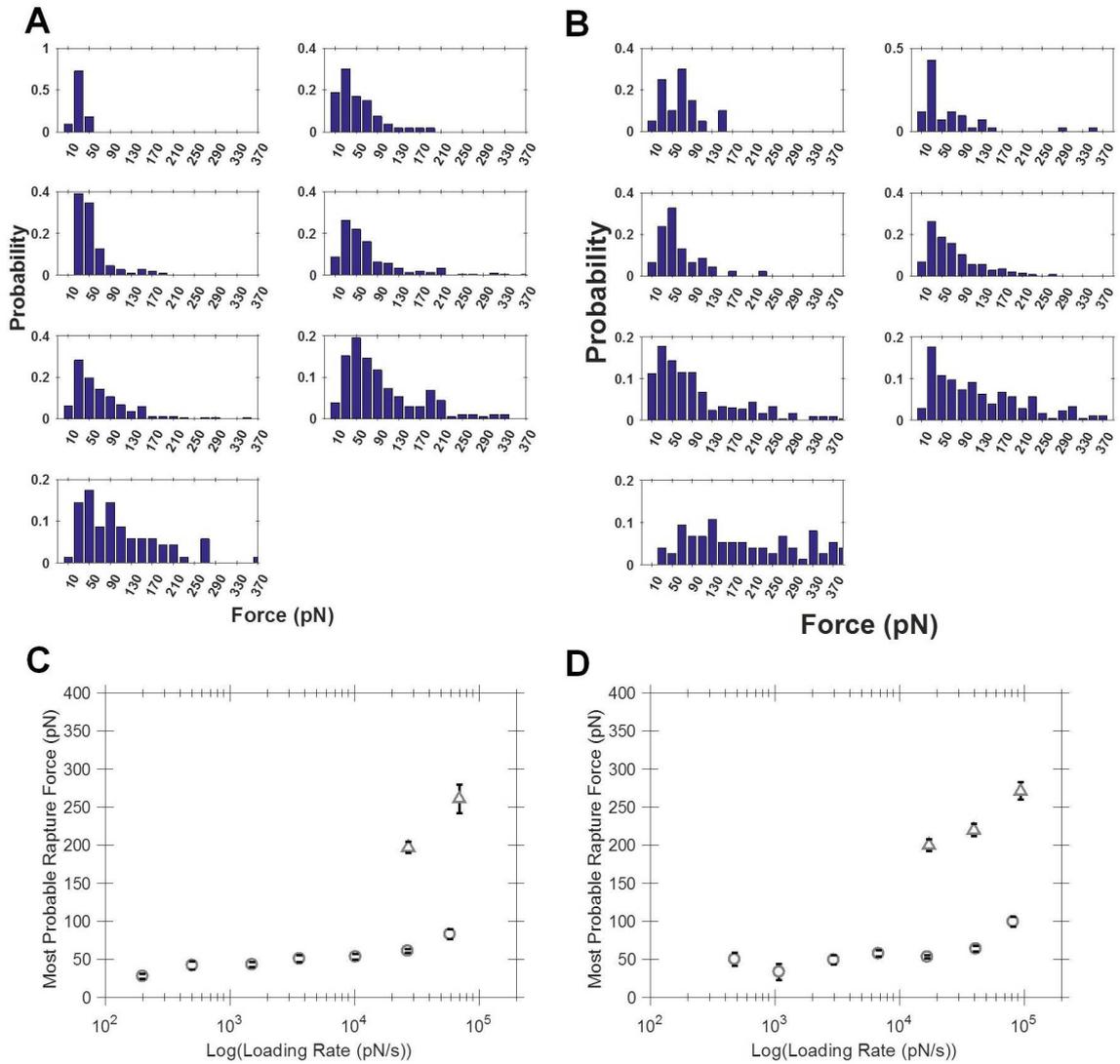


Figure 4.5. Adhesion force histograms obtained at different loading rates of Hsp90-Geldanamycin AFM pulling experiments at (a) 300 K, (b) 310 K. Probability values are calculated from total number of observed forces at each loading rates. The most probable rupture forces vs loading rates in semilog scale for (c) 300 K, (d) 310 K. The most probable rupture forces are extracted from the median values of the distribution of adhesion forces at each loading rate. The low- and high-strength states are shown in circles and triangles, respectively.

in Equation 4.2 as illustrated with straight linear lines in Figure 4.6a and b. The kinetic parameters of the interaction obtained from Bell's model at each temperature are summarized at Table 4.2, which are directly related to the binding energy landscape of the molecules. The disassociation rate of the low-strength state complex is lower at 310 K, whereas it is remarkably higher for the high-strength state. However, the position of the energy barrier is the opposite. Since there is an inverse relation between the dissociation rate and the natural logarithm of the activation energy barrier, the results exhibit that increase in temperature decreases the activation energy barrier of the low-strength state but extremely increases the barrier at high-strength state.

Table 4.2. Parameters of the Hsp90-Geldanamycin interaction obtained from the Bell model.

| | Low-Strength State | | High-Strength State | |
|--|--------------------|---------|---------------------|---------|
| | T=300 K | T=310 K | T=300 K | T=310 K |
| r_0 (nm) | 0,522 | 0,480 | 0,061 | 0,098 |
| k_{off} (s-1) | 0,660 | 0,871 | 21,787 | 3,976 |
| $\Delta\Delta G$ (kBT)* | -0,277 | | 1,701 | |
| * $\Delta\Delta G$ is relative to T=27°C Hsp90-Geldanamycin binding energy | | | | |

4.3.3. Global Dynamics of Hsp90 Conformational Transition

ANM-LD simulations were performed to investigate the functional dynamics of Hsp90 open-closed conformational transition at temperatures 27 ° C (300 K) and 37 ° C (310 K), as summarized in Table 4.1. Figure 4.7 shows the evolution of the RMSD values for each cycle for all the performed transition simulations. MM=100 runs are the most successful in terms of approaching the target state for all of the forward runs (Apo, ATP- and GDM- bound), with ($RMSD_{\text{final}} \leq 3 \text{ \AA}$) (See Table 4.1). Also, RMSD values are plotted separately in Figure 4.8 for the selected parameter set ($R_{\text{cut}}=13 \text{ \AA}$, $modemax=100$, $DF=1 \text{ \AA}$, $T=300, 310 \text{ K}$, $stepmin=1000$, $stepsim=100$, $\gamma=50 \text{ ps}^{-1}$), with the corresponding selected modes, overlap, and degree of collectivity values

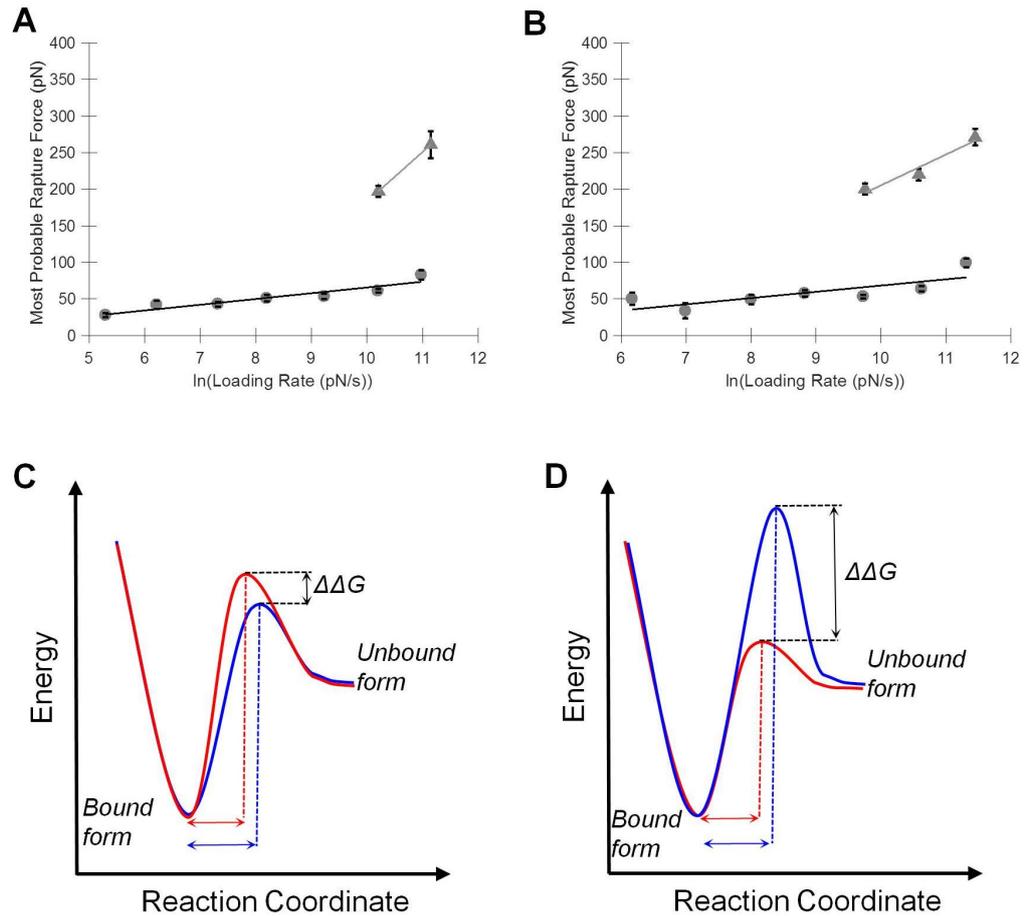


Figure 4.6. The dynamic adhesion force spectra of Hsp90-Geldanamycin complex depending on the loading rate obtained at (a) 300 K, (b) 310 K. The mean adhesion forces vs. the natural logarithm of the loading rate are shown as straight lines. A smooth diagram of the binding energy landscape of the complex at 300 K (red) and 310 K (blue) at low-strength state (c) and high-strength state (d).

for each cycle. The maximum number of modes employed was observed to effect the course of the transition, as also observed in the prototype case AdK.

Mostly selected modes significantly differ between the GDM-bound and ATP-free/ATP-bound cases, slow modes 8 and mode 5 being the most popular modes for these cases, respectively. Presence of the inhibitor GDM, seem to prevent the occurrence of the functional modes reaching the target for the assumed transition pathway, as illustrated in the GNM eV1 vs eV2 scatter plots of Figure 4.9 (See Section 2.4 for the description of GNM eVs distinguishing conformational states). ATP-bound case (green path) shows a shortest path kind of approach to the target/closed structure, such as the presence of ATP allows the appearance of the modes/motions that is required for the transition, whereas presence of the inhibitor GDM prevents the selection of the mode that is required to reach the active state (red path). ATP and GDM binding sites are different (no common residues, given in Section 4.2.4), although they bind the same cavity with a similar orientation. Since the results show significant differences in Apo, ATP-bound and GDM-bound runs, this implies that the developed method is able to simulate the presence/effect of the ligands on the created pathways. Starting from the ATP and GDM's contacting residues, the signal propagated differently, affecting the communication network and resulted in different pathways.

When we compare the GNM eV1 histogram plots for the ATP-bound and the GDM-bound cases for 300 K (lowT) and 310 K (highT), the conformational states sampled does not differ for the lowT, but for the highT GDM-bound case appears to sample two major conformational states, since the distribution is bimodal. This bimodal distribution is not observed in the Apo and ATP-bound cases appearing with one state (Figures 4.10 through 4.13). This representation also distinguishes clearly the inadequate sampling of the target state with the available 30 slow modes, for all cases.

For the major transition part of the pathways, the cross-correlation maps are constructed using the generated conformations in order to characterize the coupling dynamics of the conformational transition. In the Apo and ATP-bound correlation

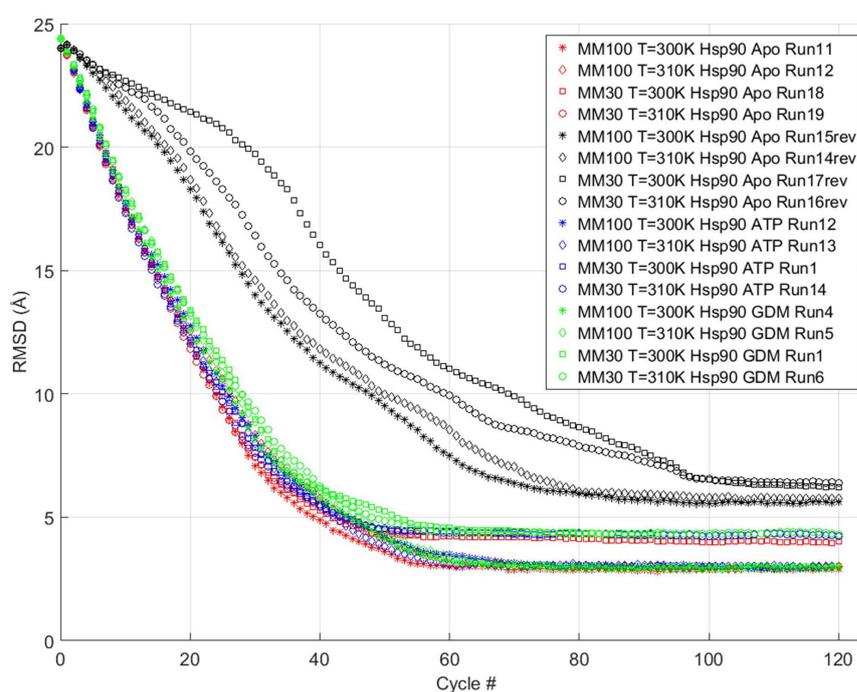


Figure 4.7. RMSD plots for Hsp90 forward (open-closed) runs with different ligands bound and for different *modemax* and T values ($DF=1$ Å; *modemax*, MM=30, 100; $T=300, 310$ K; ANM $R_{cut}=13$ Å).

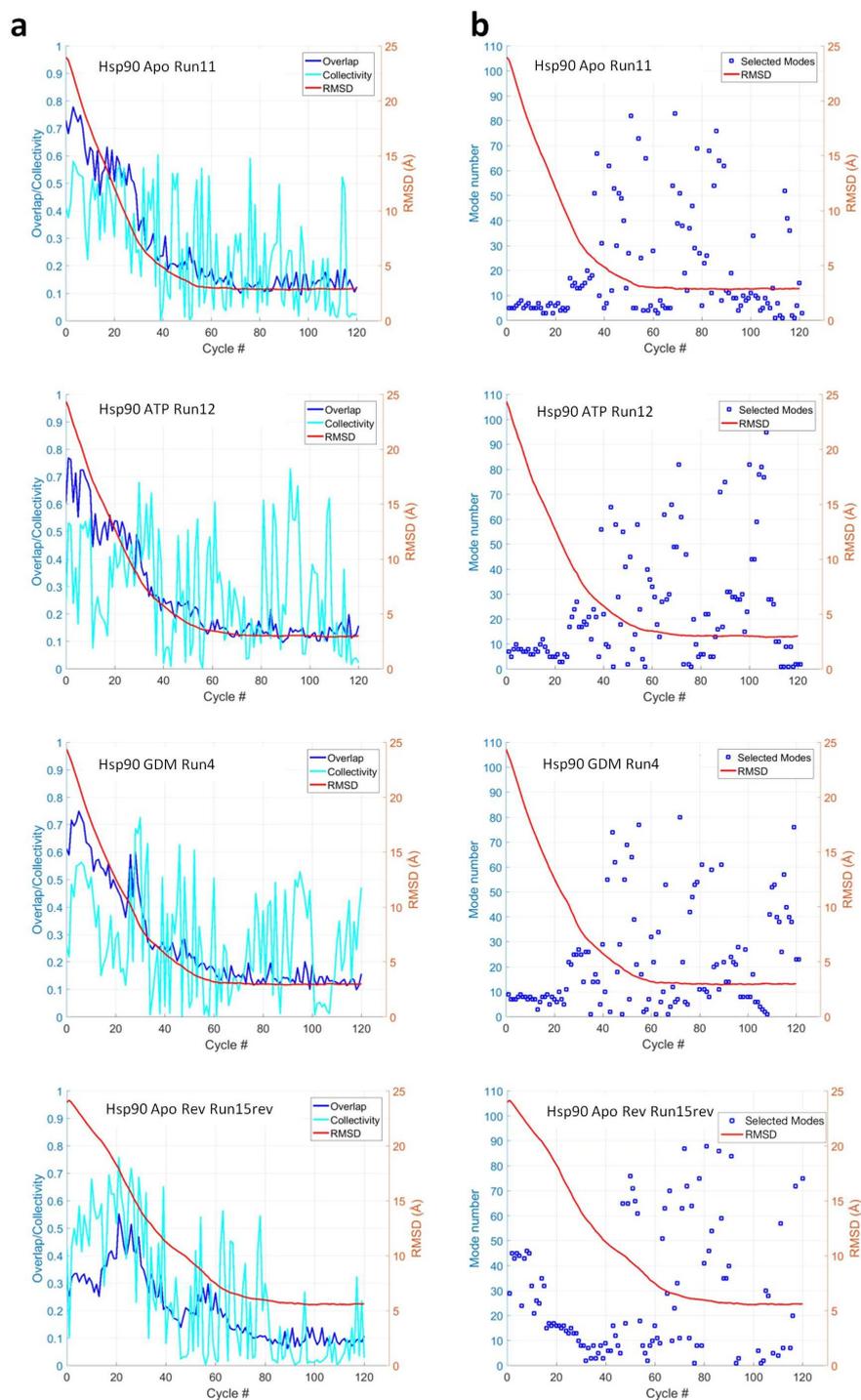


Figure 4.8. (a) Overlap/collectivity values of the selected slow modes and (b) the selected slow modes are plotted with RMSD values for the Hsp90 forward case runs (Apo/ATP-bound/GDM-bound) and the reverse case run, for the parameter set ($DF=1 \text{ \AA}$; $modemax$, $MM=100$; $T=300 \text{ K}$, $ANM R_{cut}=13 \text{ \AA}$).

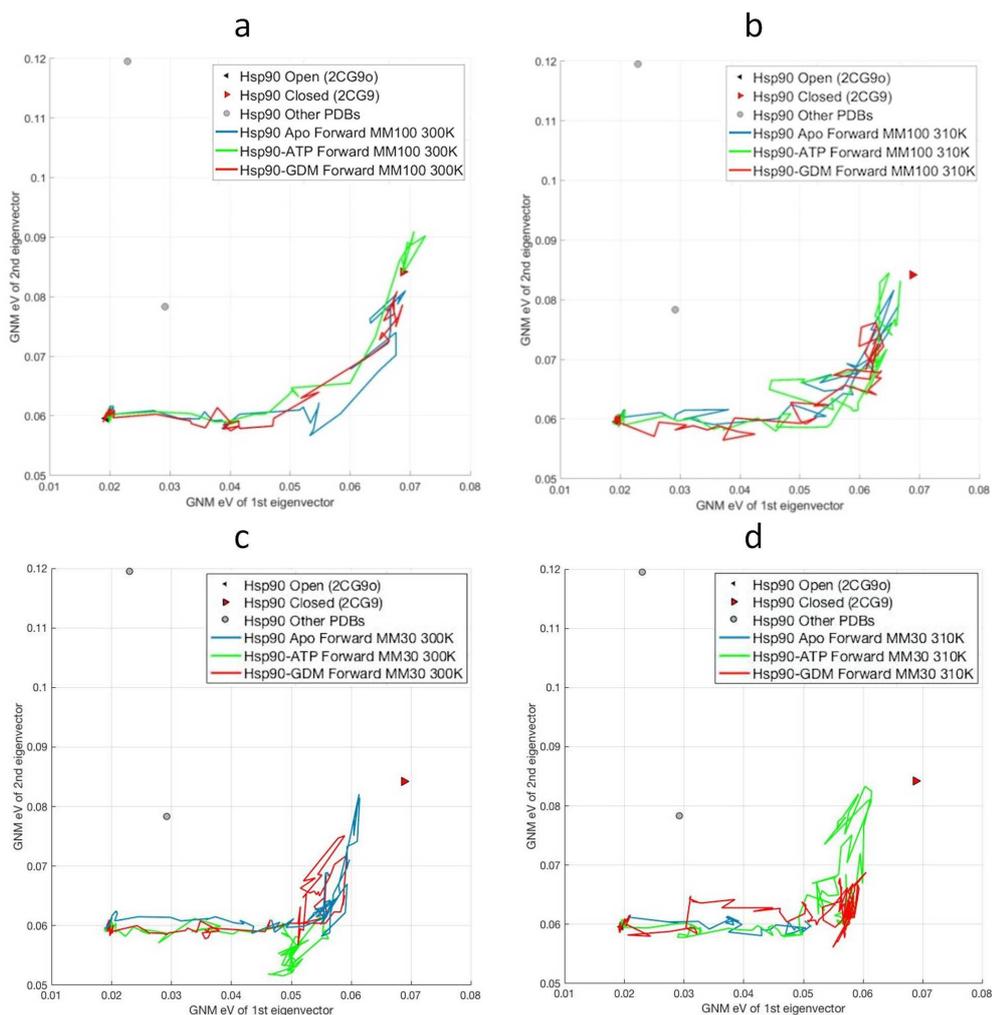


Figure 4.9. GNM first eigenvector (eV1) versus second eigenvector (eV2) plots for Hsp90 forward runs (Apo/ATP-bound/GDM-bound) for *modemax*, MM=100 and (a) $T=300$ K (b) $T=310$ K, for *modemax*, MM=30 (c) $T=300$ K and (d) $T=310$ K (Fixed parameters are $DF=1$ Å, ANM $R_{cut}=13$ Å). Eigenvalues for the Hsp90 open (black triangle) and closed states (red triangle) are shown in the eV plots.

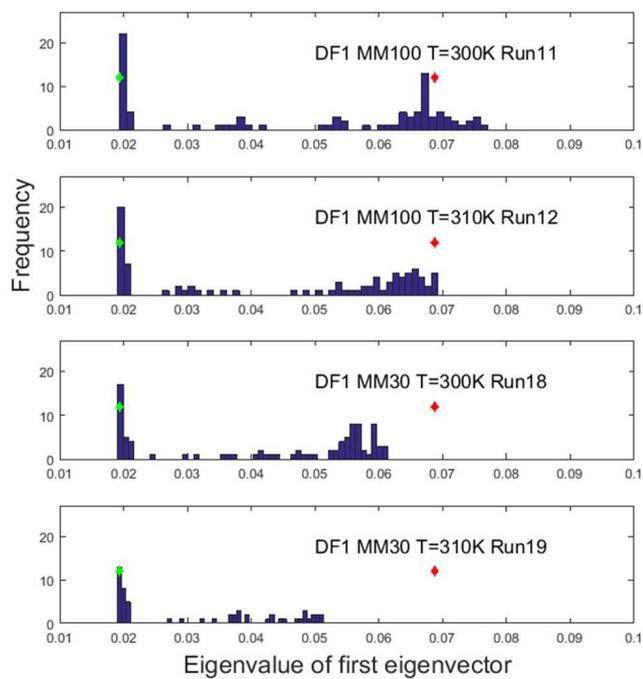


Figure 4.10. GNM first eigenvalue (eV1) histograms for Apo Hsp90 forward runs.

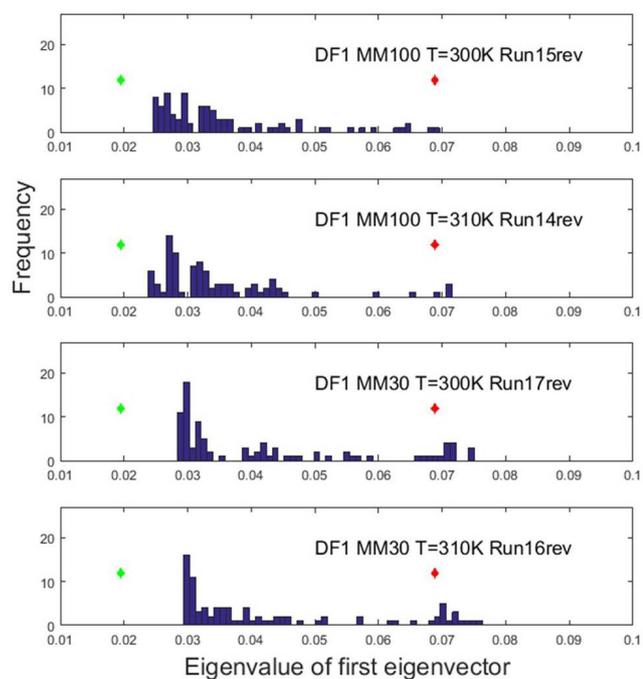


Figure 4.11. GNM first eigenvalue (eV1) histograms for Apo Hsp90 reverse runs.

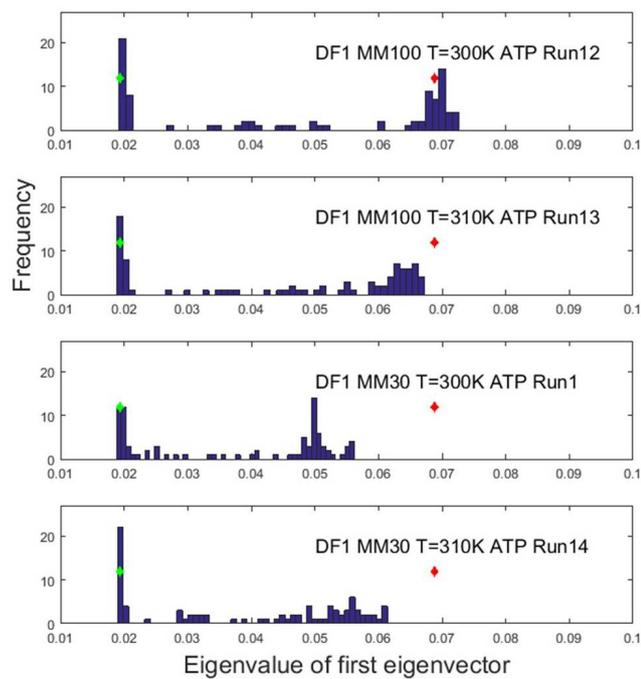


Figure 4.12. GNM first eigenvalue (eV1) histograms for Hsp90-ATP forward runs.

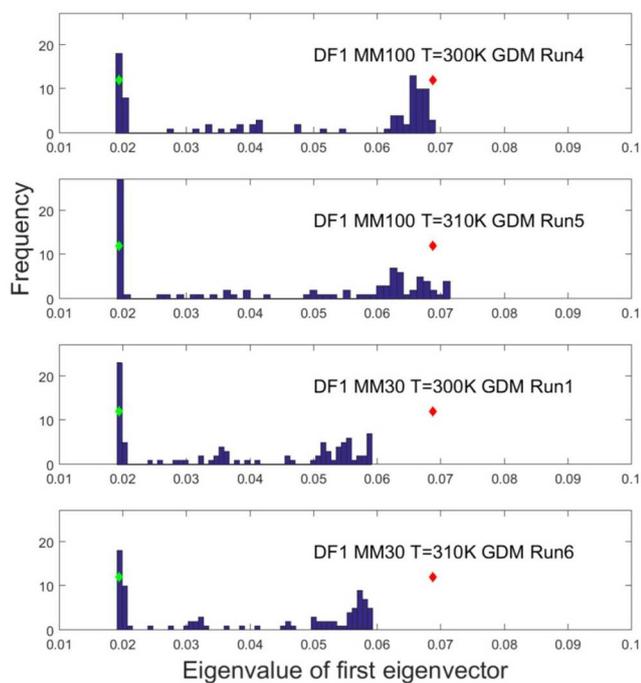


Figure 4.13. GNM first eigenvalue (eV1) histograms for Hsp90-GDM forward runs.

maps, the residues E222 and D248 and their vicinity are observed to correlate with the rest of the structure (both intra- and inter- chain), forming the two hinges in the CL region (T210-E265) for both lowT and highT cases, where it is stronger for the highT one (Figures 4.14 and 4.15). These hinges mostly fade away in the GDM-bound lowT map. For the highT case the change in the strength of the correlation is even more when the difference of the GDM-bound map (Figure 4.16) from the apo map (Figure 4.14) is considered (See the difference maps in Figure 4.17).

Stronger correlations observed in the Apo and ATP-bound state implies the separation of the movements of the domains NTD and Md. This extra flexibility of NTD (caused by the undocked CL) has a crucial role in the open-close transition, by the closing of the NTDs together on the top of the molecule (Figures 4.14 and 4.15). The lack of these stronger correlations in the GDM-bound highT map, is a sign of decreased flexibility of NTD together with the disappearance of the hinges, in this case the transition should be slower (Figure 4.16).

For the Apo highT case, the NTD (including the ATP binding sites listed in Structures section) highly correlates with the D560-K610 region of the CTD in both chains (both intra- and inter-chain), which are over 70 Å distant from each other (Figure 4.14). For the ATP-bound case, the said ATP binding sites-CTD correlation gets stronger within the each subunit (cross-subunit counterpart weakens) (Figure 4.15). GDM-bound case resembles the apo form in the way ATP binding sites and CTD correlates, but much weaker (Figure 4.16).

Here, the role of the aforementioned hinges is apparent in the conformational transition simulated. This is in agreement with the study stating the importance of the CL region in the progression of the conformational transition. Docked CL fixes the NTD to the Md, resulting in decreased flexibility and slower transition, whereas undocked CL comes with some flexibility and results in faster transition [105].

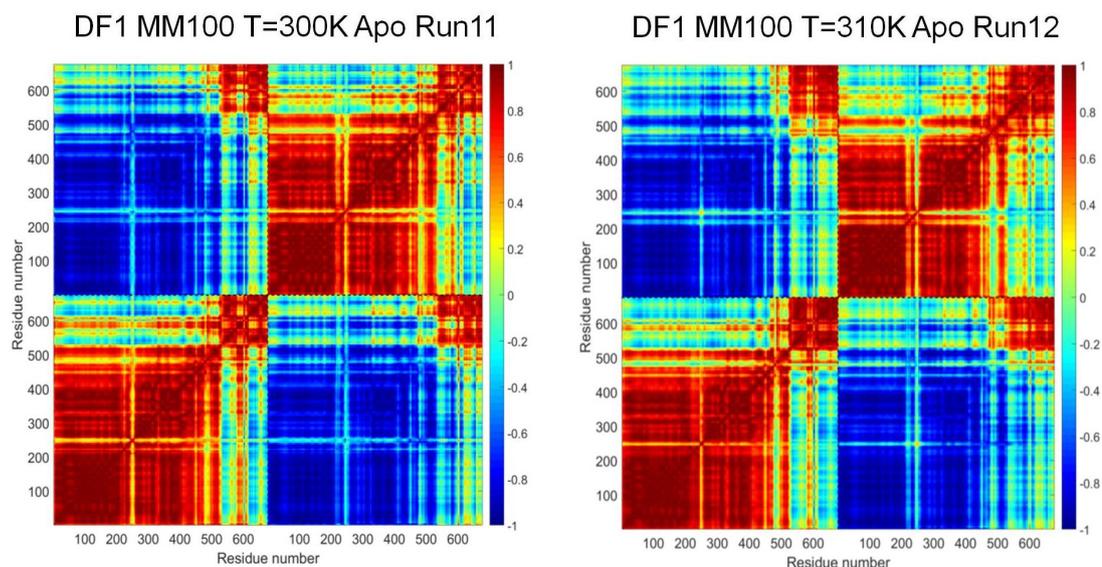


Figure 4.14. Cross-correlation plots for Apo Hsp90 forward runs. The transition parts of the trajectories (just before RMSD to target reaches a plateau) are considered for the calculations.

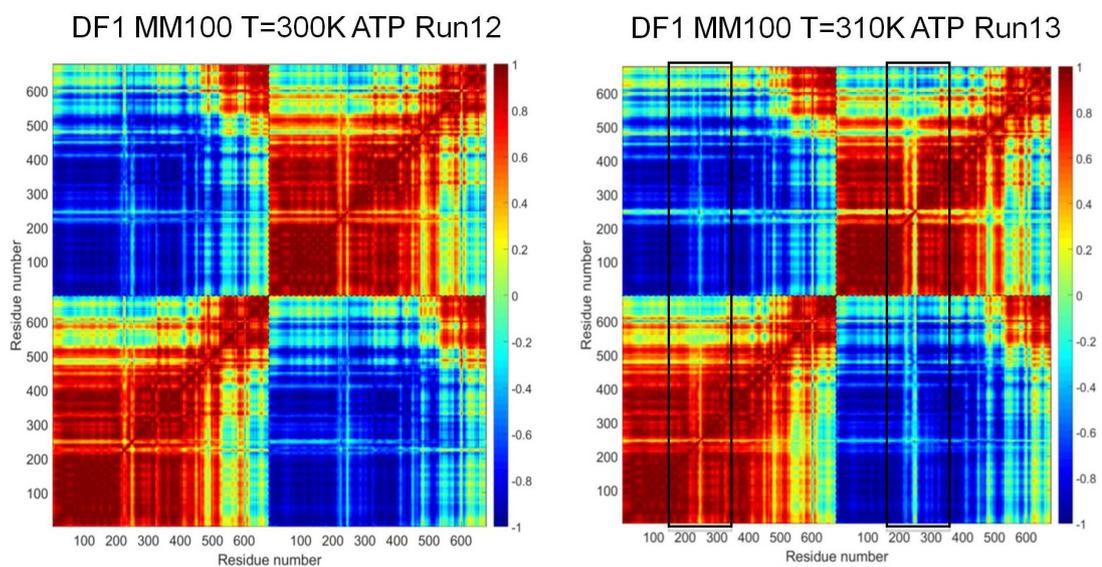


Figure 4.15. Cross-correlation plots for Hsp90-ATP forward runs. The transition parts of the trajectories (just before RMSD to target reaches a plateau) are considered for the calculations.

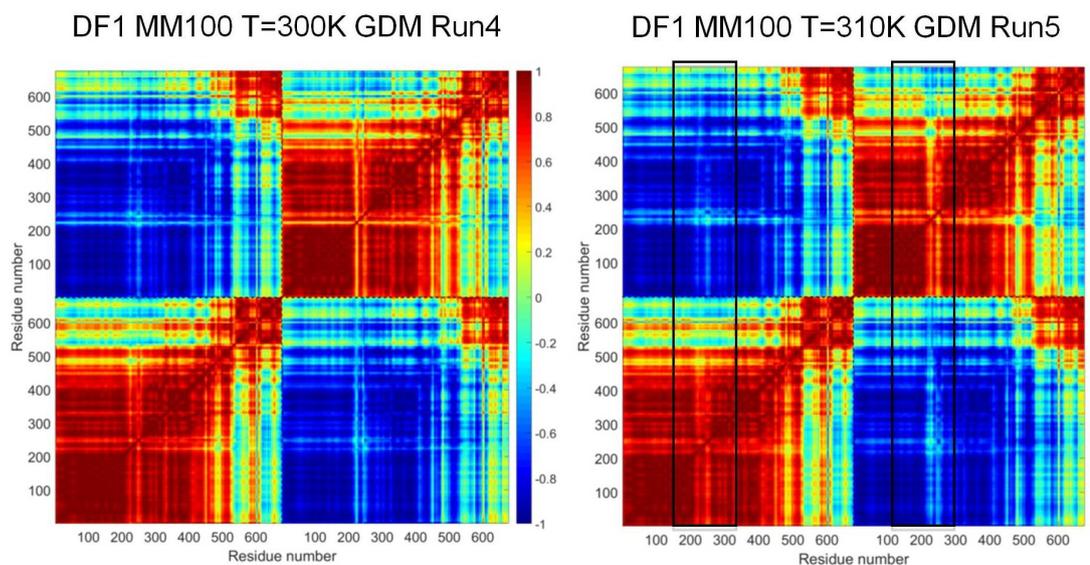


Figure 4.16. Cross-correlation plots for Hsp90-GDM forward runs. The transition parts of the trajectories (just before RMSD to target reaches a plateau) are considered for the calculations.

4.4. Conclusions

In this study, the binding/unbinding behavior of the Hsp90 inhibitor geldanamycin, GDM, and the effect of temperature on this interaction are explored. Coupling of the global conformational change to the local binding behavior and dependence of these on the temperature is explored with experiments and simulations.

The unbinding forces, and the kinetics of the unbinding process as well as the parameters of the energy landscape of the Hsp90-GDM complex were investigated by AFM pulling experiments. To this end, AFM dynamic force spectroscopy experiment of Hsp90-GDM complex is performed at 27 °C (300 K) and 37 °C (310 K) to investigate the effect of temperature on Hsp90-GDM binding dynamics. As the temperature increases, the unbinding forces also increase over the entire range of observed loading rates, which indicates that the interaction between Hsp90 and GDM may originate from the temperature-dependent interactions.

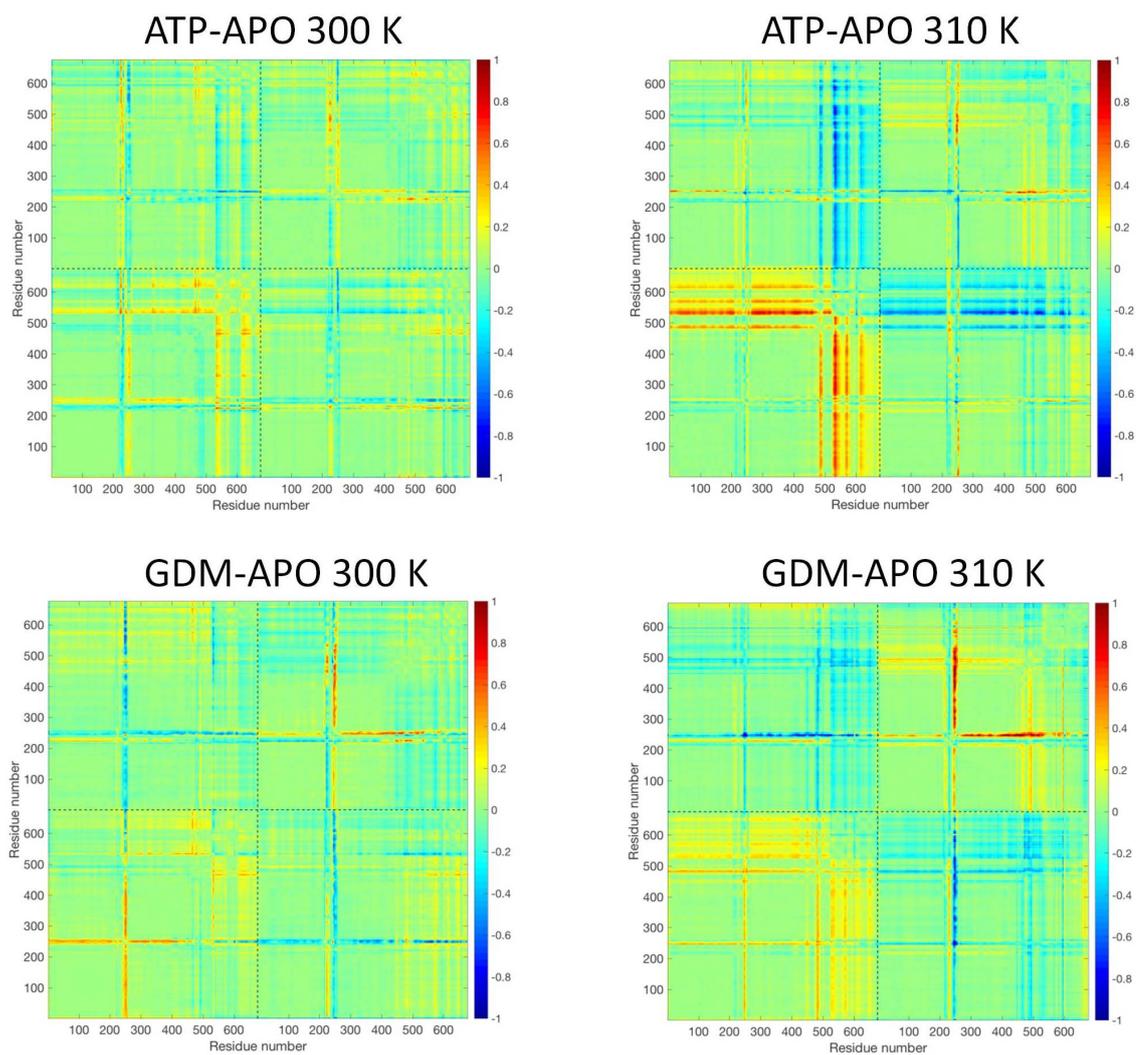


Figure 4.17. Cross-correlation difference plots for Hsp90-ATP and Hsp90-GDM forward runs from the Apo Hsp90 cross-correlation maps at the corresponding temperature. The transition parts of the trajectories (just before RMSD to target reaches a plateau) are considered for the calculations.

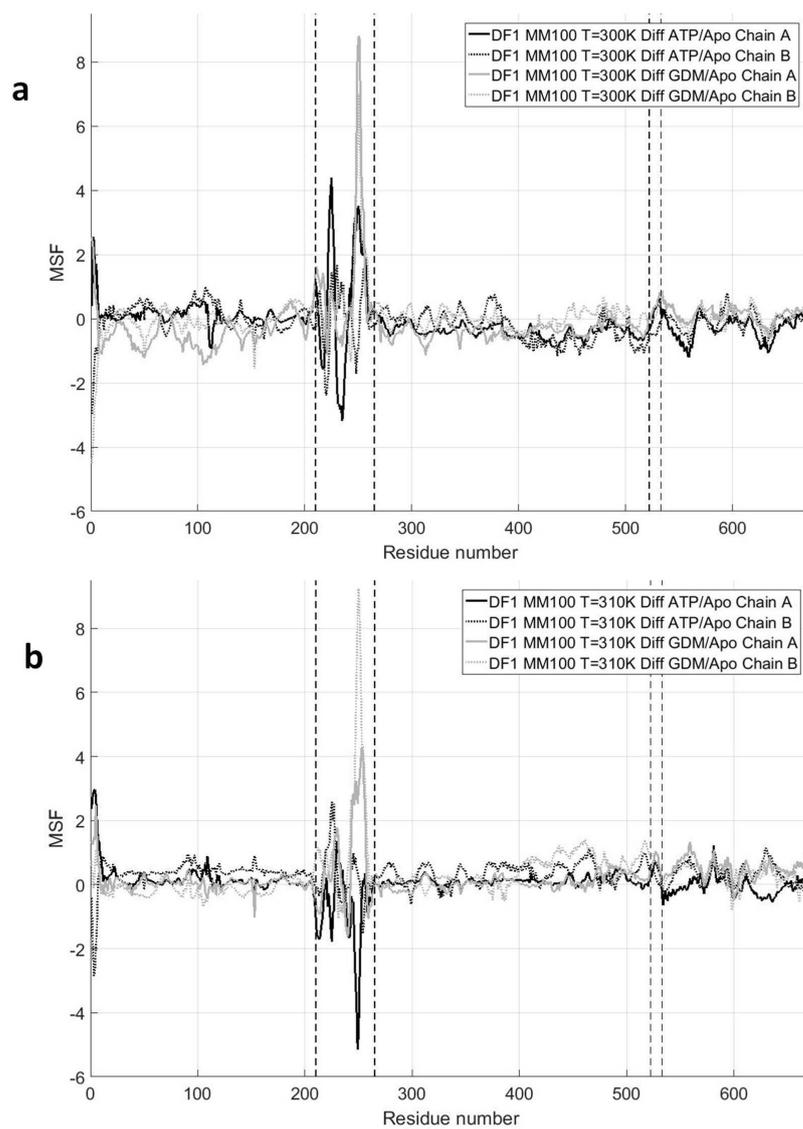


Figure 4.18. MSF difference plots for Hsp90-GDM forward runs (MM=100) for (a) T=300 K (b) T=310 K. The transition parts of the trajectories (just before RMSD to target reaches a plateau) are considered for the calculations.

The kinetic parameters of the interaction obtained from Bell's model at each temperature are directly related to the binding energy landscape of the molecules. The disassociation rate of the low-strength state complex is lower at 310 K, whereas it is remarkably higher for the high-strength state. However, the position of the energy barrier is the opposite. Since there is an inverse relation between the dissociation rate and the natural logarithm of the activation energy barrier, the results exhibit that increase in temperature decreases the activation energy barrier of the low-strength state but extremely increases the barrier at high-strength state.

In order to investigate how does inhibitor bound states disrupt the conformational dynamics of Hsp90 –with temperature dependence–, the lowT and highT cases for the ATP-free, ATP-bound and GDM-bound ANM-LD simulations were performed. Global binding behaviour was observed to couple to the local binding behavior of the ATP/GDM, where both bind to the same cavity of Hsp90, but with totally different residue contacts. The effects of this local binding site were observed to couple to the global dynamics behavior and appeared in the cross-correlation maps of the assumed conformational transition pathways of the ANM-LD simulations.

Stronger correlations observed in the Apo and ATP-bound state open-closed transition cross-correlation maps implies the separation of the movements of the domains NTD and Md. This extra flexibility of NTD (caused by the undocked CL) has a crucial role in the open-close transition, by the closing of the NTDs together on the top of the molecule. The lack of these stronger correlations in the GDM-bound highT map, is a sign of decreased flexibility of NTD together with the disappearance of the hinges, in this case the transition should be slower. For the Apo highT case, the NTD (including the ATP binding sites listed in Structures section) highly correlates with the D560-K610 region of the CTD in both chains (both intra- and inter-chain), which are over 70 Å distant from each other. For the ATP-bound case, the said ATP binding sites-CTD correlation gets stronger within the each subunit (cross-subunit counterpart weakens). Here, the role of the aforementioned hinges is apparent in the conformational transition simulated. This is in agreement with a recent single-molecule experimental study stating the importance of the CL region in the progression of the conformational

transition. Docked CL fixes the NTD to the Md, resulting in decreased flexibility and slower transition, whereas undocked CL comes with some flexibility and results in faster transition [105].

GDM-bound dynamics of Hsp90 compared to the apo form dynamics observed to be dependent on the temperature, especially at highT, which is also the case for the experiments. Simulations also revealed the controlling role of the CL on the course of the transition.

The results of this study, providing hints for the functional/inhibition dynamics of GDM, could contribute to the research on the design of the new line of Hsp90 inhibitors or increasing the specificity of the current ones.

5. DYNAMIC SWITCH MECHANISM IN APO CYCLIC AMP RECEPTOR PROTEIN

Escherichia coli cyclic AMP Receptor Protein (CRP) undergoes conformational changes with cAMP binding and allosterically promotes CRP to bind specifically to the DNA. In that, the structural and dynamic properties of apo CRP prior to cAMP binding, holo CRP and the conformational transition in-between are of interest for the comprehension of the activation mechanism. Here, the dynamics of apo CRP monomer/dimer (*off-state*) and holo CRP dimer (*on-state*) were studied by Molecular Dynamics (MD) simulations and Gaussian Network Model (GNM) and the conformational transition between these two states were explored by ANM-LD simulations. The hinge at L134-D138 by a shift to the N-terminus displays intra- and inter-subunit coupled fluctuations with the cAMP and DNA binding domains and leads to the emergence of stronger coupled fluctuations between the two domains and describes an *off-* to *on-state* transition. The flexible regions at K52-E58, P154/D155 and I175 corroborate with the L134-D138 hinge and maintain the dynamic coupling of the two domains. The eigenvalue distribution distinguishes the dynamic states sampled in different equilibrium and pseudo-equilibrium trajectories. We present a mechanistic view on how the structural dynamic units are hierarchically built and employed during the conformational transitions that underlie the allosteric functional mechanism.

5.1. Introduction

The *Escherichia coli* cAMP Receptor Protein (CRP) (also known as Catabolite Activator Protein, CAP) activates the transcription of more than 150 genes. Upon binding of cAMP (cyclic AMP), the transcriptional activity of CRP is altered resulting in a change in the affinity for its target CRP-dependent promoter region on the DNA. CRP is then able to recruit RNA polymerase (RNAP) for the transcription activation to begin.

CRP is a 47 kDa homodimer with 209 amino acid residues in each monomer where individual subunits fold into two domains [117]. The N-terminal domain, extending from V1 to N133, contains the primary cAMP binding site (anti cAMP) and mediates subunit-subunit interactions. This domain is formed by α helices A, B, C and eight β strands 1 to 8. The C-terminal domain extends from V139 to R209 with α helices D, E, F and four β strands 9 to 12, which has the helix–turn–helix (HTH) motif involved in the specific DNA and secondary cAMP binding (syn cAMP). Two domains are connected by the linker or hinge region L134-D138 [118]. The cAMP molecules in the primary cAMP binding sites interact mainly with residues G71, E72, R82, S83, T127 and S128 of the other subunit [119]. The active state that initiates transcription is accepted to be with two anti cAMPs bound form. The syn cAMP molecules can bind CRP only when the two anti cAMPs are bound. The syn cAMPs interact with the HTH motif (mainly R180) and β strand of the cAMP binding domain (E58), the DNA, and A135 of the other subunit [120]. Two cAMP molecules (anti and syn) bind to each subunit, making a total of four cAMPs per dimer. The binding of four cAMPs to CRP occurs at millimolar concentrations, yet the cAMP concentration is at micromolar levels *in vivo* [121]. The role of CRP-cAMP₄ complex in the transcription mechanism is yet unknown [118].

CRP dimer interacts with DNA by a two-fold symmetric consensus DNA sequence. R180, E181 and R185 are the key residues of the HTH motif involved in specific protein-DNA interactions [122], whereas the nonspecific DNA binding residues are R169, Q170 and S179 [123] with the other DNA binding sites [113] D138, V139, T168, C178, T182, G184, K188, H199, and G200. Next to the HTH motif, residues A156-Q164 (activating region 1, AR1) are responsible for the transcription activation of lac class I and class II CRP-dependent promoters. The photo-cross-linking experiments [124] indicate that the C-terminal domain of the RNAP α -subunit (α CTD) binds one of the CRP subunits contacting AR1, whereas the other CRP subunit makes contacts with the other parts of the RNA polymerase. Then α CTD interacts with the minor groove adjacent to the DNA site for CRP. Thus, the transcription activation at lac promoter involves both protein-protein and protein-DNA interactions [125–128]. In class II promoters, the interaction with RNAP is complemented by two more activating

regions of residues H19, H21, E96, and K101 (AR2) and K52–E55 and E58 (AR3) [129]. The X-ray crystal structure of holo CRP and the functional sites are shown in Figure 5.1a and b, respectively.

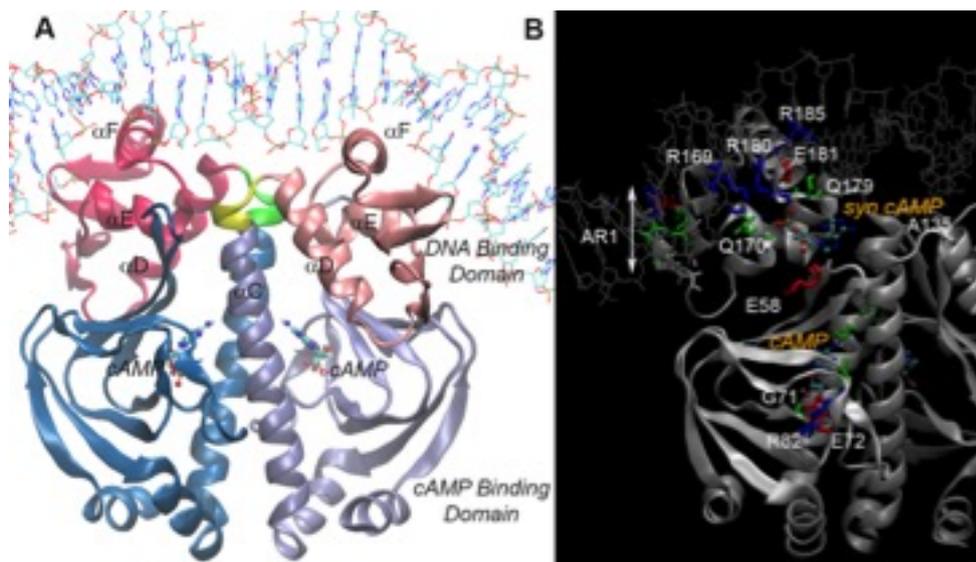


Figure 5.1. Holo CRP crystal structure and functional sites. (a) The CRP structure in complex with DNA and cAMP (PDB: 1CGP). (b) The N-terminal cAMP binding domains (V1-N133), the C-terminal DNA binding domains (V139-R209) and the hinge regions (L134-D138) connecting the two domains in subunits A/B are represented in cyan/ice blue, dark pink/light pink and green/yellow, respectively.

Within the sequence of events up to the transcription activation, the structural and dynamic characteristics of CRP and the activation mechanism through ligand binding can best be described by the comparison of the apo and holo states. The structures of holo CRP with and without the DNA have been elucidated since its first isolation from *E. Coli* in 1970 [130]. Several biochemical, genetic and biophysical experiments [118, 121, 128, 129, 131] along with computer simulations [132–134] have been performed to understand the cAMP allosteric switch mechanism from the inactive to active states. The three previous simulations [132–134] were based on the crystal structure of holo CRP dimer from which the apo state and the state with a single cAMP were also modeled. There are several recent studies and reviews addressing the allostery in CRP [118, 129, 134]. The NMR solution structure of apo CRP [135] and

X-ray crystal structures of holo CRP [136, 137] have contributed significantly to the understanding of allosteric mechanism in play [135, 138].

The basic mechanism of allosteric control is the transmission of the signal from the cAMP to DNA binding domains upon anti cAMP binding and the interaction of cAMP adenine base atoms with the side chain hydroxyls of residues T127 and S128. This induces a coil-to-helix transition and the elongation of C-terminus of C-helix by 11 residues. The change in the secondary structure content with a rearrangement of the hinge [129, 138–140] directs the domain movements and triggers the rotation and translation of the DNA binding domain placing the F-helices into the required orientation to enter into the major grooves of the DNA. The shift in the inter-domain hinge was first observed by a structural inspection at the secondary structure level by NMR [139]. The hinge region appears to play a key role by modulating the inter-domain interactions and stabilizing the altered domain movements leading to the transcriptional activation [118, 129, 135]. The two states of the coiled-coil and the transition towards the ordered form coupled to the ligand binding functions as a regulatory switch. This ensures the precise allosteric control during the protein's functioning [135]. These studies suggest that the rearrangement of the hinge with the ligand binding is a critically conserved feature that controls the global allosteric transformation in CRP-family structures [129]. Thus, the majority of the CRP* (cAMP independent mutant of CRP) mutations in apo CRP are found to achieve proper allosteric transition by modifying the hinge-mediated inter-domain network pre-existing in apo CRP [129, 141, 142]. Although many X-ray crystal/NMR structures of different states of CRP are available, further structural information on the naturally occurring functional states of CRP, like CRP-cAMP1, CRP-cAMP-RNAP, as well as conformational transitions in-between is required for a complete understanding of CRP's allosteric mechanism and the functional implication of each state.

Further to the structural understanding of apo and holo states, allostery was also put forward as a dynamic relationship [143, 144]. Besides a mechanical view [145, 146] of structural changes, the dynamics is also of interest in the allosteric regulation of protein activity. With the changes in the protein motion, an allosteric communication may

evolve [144]. Here, the dynamics of apo CRP monomer/dimer and holo CRP are explored computationally by extensive Molecular Dynamics (MD) simulations combined with the Gaussian Network Model (GNM) [24, 25] analysis of MD sampled conformations. The GNM analysis was performed on apo CRP NMR solution structures and the holo CRP crystal structure. Additionally, several parallel apo-holo conformational transition simulations starting from the apo CRP NMR structure 2WC2 Models 10 and 11 to the target holo CRP crystal structure 1G6N have been performed by the ANM-LD methodology.

We mainly focus on how the fluctuation dynamics provide the basis for an allosteric communication and describe a conformational switch with its off and on states and how this behavior is evolved from the dynamics of apo CRP monomer to apo-to-holo CRP dimer. The dynamic infrastructure for the coordination between the effector and DNA binding domains is largely observed in apo CRP monomer. This is an example of the pre-existence of functional dynamic states in the smaller subunits of the structure and also the fact that the inter-domain hinges are not simply linkers that connect the two domains but coordinate the global structural motion, where the key dynamic states are pre-encoded [147].

5.2. Materials and Methods

5.2.1. Molecular Dynamics Simulations

The MD simulations were performed for the dimer (subunits A and B) and the unbound monomer (subunit A) of apo CRP for a simulation time of 150 ns each with the initial structure of apo CRP NMR solution structure (PDB: 2WC2, model 11) [135], as well as a 150 ns holo CRP dimer run with the initial holo CRP X-ray crystal structure (PDB: 1G6N) [137]. Two parallel runs of 75 ns were performed for each apo CRP monomer and dimer with different initial structures (PDB: 2WC2, models 2 and 10) [135]. The details of the simulated systems are given in Table 5.1.

Table 5.1. CRP: Details of the simulated systems.

| Model | No. of Amino Acids | Box Dim. (Å) | No. of Water Molecules | No. of Atoms | No. of Cl-Ions | Sim. Time (ns) |
|---------------------|---------------------------|---------------------|-------------------------------|---------------------|-----------------------|-----------------------|
| Apo CRP monomer (1) | 209 | 80.66 | 32,292 | 35,650 | 7 | 150 |
| Apo CRP monomer (2) | 209 | 73.72 | 27,519 | 30,877 | 7 | 75 |
| Apo CRP monomer (3) | 209 | 73.64 | 27,486 | 30,844 | 7 | 75 |
| Apo CRP dimer (1) | 418 | 89.18 | 48,072 | 54,788 | 14 | 150 |
| Apo CRP dimer (2) | 418 | 93.05 | 48,183 | 54,899 | 14 | 75 |
| Apo CRP dimer (3) | 418 | 90.42 | 43,728 | 50,444 | 14 | 75 |
| Holo CRP dimer | 401 | 82.51 | 43,578 | 50,069 | 2 | 150 |

Truncated octahedron periodic boundary conditions were applied to each system.

The Amber 8.0 [148, 149] and Amber 11 [81] biomolecular simulation programs were used in the MD simulations. The Amber ff03 [80] force field parameter set was used for the proteins/ions. Each system was solvated using TIP3P [150] water molecules in an octahedron periodic box. Histidine residues 17, 19, 21, 31, 159 and 199 were protonated for the states predicted by H++ server [151, 152]. Initially, energy minimization was performed using 50 cycles of steepest descent algorithm, followed by the conjugate gradient method. Initial velocities were selected at random from the Maxwell-Boltzmann distribution at a temperature of 10 K, which was gradually increased to 300 K. The Berendsen thermostat and barostat [85] was used for the NPT ensemble ($T = 300$ K, $P = 1$ bar) with a time step of 2 fs. The SHAKE algorithm [59] was used as the bond constraints for all bonds involving hydrogens to eliminate the high frequency bond vibrations. A cutoff distance of 9 Å was used for the nonbonded interactions. The long-range interactions in electrostatic terms were corrected using the particle mesh Ewald method [153]. The equilibration periods were taken as 5 ns and 50 ns for the dimer and monomer trajectories, respectively.

5.2.2. ANM-LD Simulation Protocol

We have performed several parallel apo-holo conformational transition simulations starting from the apo CRP NMR structure 2WC2 Models 10 and 11, and targeting the holo CRP crystal structure 1G6N, by ANM-LD methodology at 310 K. ANM-LD hybrid methodology simulates the conformational transition of a structure starting from a given initial state to a target one by distorting the initial structure along the ANM modes mostly overlapping with the initial-target coordinate difference vector followed by short cycles of minimization and implicit LD simulations. Details of the methodology can be found in Chapter 3. The simulated systems that are presented in this chapter are given in Table 5.2.

Table 5.2. CRP: Summary of ANM-LD simulation results (apo-holo transitions).

| Direction | T (K) | Mode Max | DF (Å) | RMSD initial (Å) | RMSD final* (Å) | Run Alias |
|--------------|-------|-------------|-----------|------------------------|-----------------------|--------------|
| CRP Model 11 | 310 | 30 | 0.4 | 4.40 | 2.33 | Run1 |
| CRP Model 11 | 310 | 100 | 0.4 | 4.40 | 1.97 | Run2 |
| CRP Model 11 | 310 | 100 | 0.6 | 4.40 | 1.89 | Run3 |
| CRP Model 11 | 310 | 100 | 0.8 | 4.40 | 2.13 | Run4 |
| CRP Model 11 | 310 | 100 | 0.2 | 4.40 | 2.21 | Run5 |
| CRP Model 10 | 310 | 100 | 0.6 | 4.12 | 1.93 | Run6 |
| CRP Model 10 | 310 | 100 | 0.4 | 4.12 | 1.85 | Run7 |

* The best result observed is given for each system for the parameter set: ($R_{cut}=13$ Å, $stepmin=500$, $stepsim=100$, $\gamma=5$ ps⁻¹).

5.3. Results and Discussion

5.3.1. Fluctuations and Correlations by MD Simulations

Dimerization stabilizes the C-helix at the interface and decreases the mobility of the primary cAMP binding (T127 and S128) and DNA binding residues. Moreover, the binding of the cAMPs further stabilizes the inter-domain hinge and the DNA binding residues.

The correlations between residue fluctuations are also of interest here to understand the cooperativity in residues' motion. Figure 5.2a-c display the correlations between residue fluctuations based on the first ten essential modes for apo CRP monomer/dimer and holo CRP dimer, respectively.

5.3.2. Fluctuations and Correlations by GNM

The correlations between residue fluctuations are further evaluated below through the GNM analysis of the MD sampled conformational states of apo CRP monomer, dimer and holo CRP. The hinge sites are the key mechanical sites for the cooperative motion that we observe in the correlation maps.

Dynamic network in apo CRP NMR and holo CRP crystal structures (Figure 5.3) displays the cross-correlation maps based on the average ten slowest modes and the ribbon diagrams color coded with the correlation values of the L134-D138 hinge of subunits A and B, with all other residues in a selected model of the apo CRP NMR solution structures (PDB: 2WC2, model 1) [135] and in the holo CRP X-ray crystal structure with the two cAMPs bound (PDB: 1G6N) [137]. A major difference in the internal dynamics between the apo and holo states is related to the dynamic behavior of the inter-domain hinge L134-D138, K52-E58 of the cAMP binding domain, and G173-V176 of the DNA binding domain. L134-D138, while being correlated with only the cAMP and DNA binding domains of its own subunit in the former, correlates with the cAMP and DNA binding domains of both subunits in the latter. The same behavior is observed in the majority of apo NMR model structures. A holo-like correlation dynamics is observed only in some NMR models (Models 5 and 10) as a weak population in the apo state ensemble.

The L134-D138 hinge coordinates the relative motion of the two domains and both CRP subunits in the holo state. There is yet a shift in the boundaries of the two domains, when L134-D138 becomes a part of the dynamic DNA binding domain in the apo state. L134-D138 loses its global hinge behavior and thus the coordinated motion of the domains and subunits weakens. This is observed with the weakening of the interactions between L134-D138 hinges of both subunits and between K52-E58 of the cAMP binding domain and the DNA binding domain (particularly with G173-V176). The latter is clearly observed with the difference-correlation values of the average correlation map of all NMR models and the holo crystal structure color-coded (Figure 5.3). The relatively high positive values in the DNA binding domains indicate

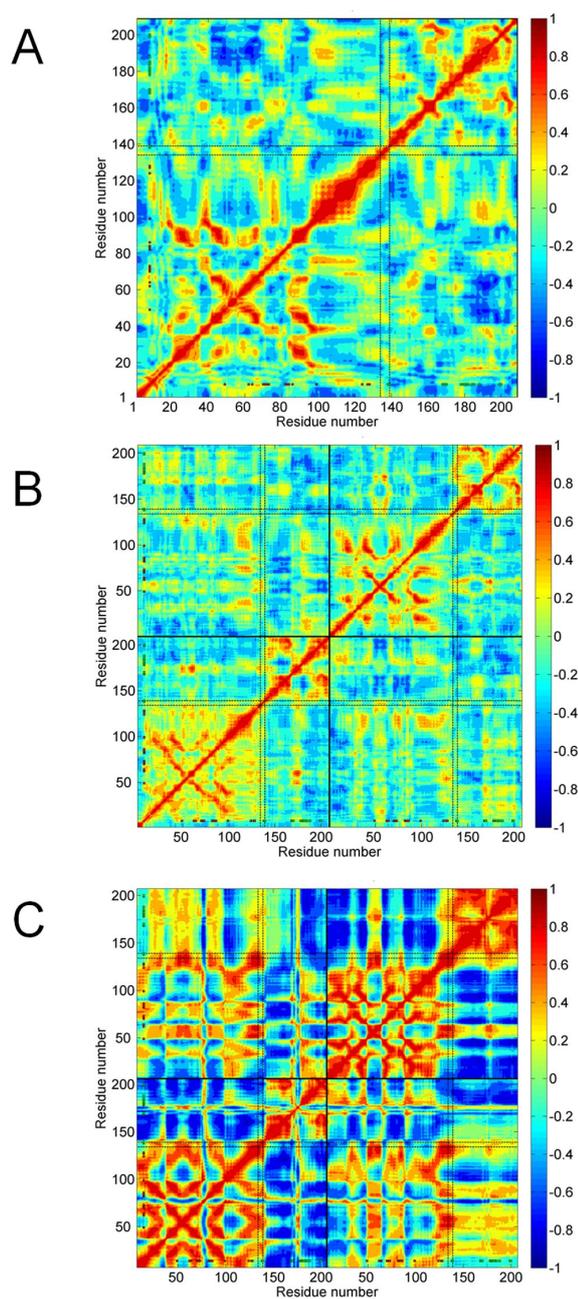


Figure 5.2. Dynamics network of apo CRP monomer/dimer and holo CRP by MD simulations. The correlation between residue fluctuations based on first ten essential modes is calculated as an average over three parallel MD runs for apo CRP monomer/dimer and holo CRP MD trajectories. The correlation maps are given for: (a) apo CRP monomer AVG, (b) apo CRP dimer AVG, and (c) holo CRP (single run).

that the L134-D138 hinges display higher correlation values with the DNA binding domains of their own subunits in the apo state. The relative more negative values in the L134-D138 hinge, K52-E58 of the cAMP binding domain, and G173-V176 of the DNA binding domain of the other subunit show that the L134-D138 hinge displays stronger correlations with the latter regions in the holo state. Along, the relatively high negative values between K52-E58 and G173-V176 (and whole DNA binding domain) mean that the two domains are coupled stronger in the holo state or vice versa for the apo state.

A very similar difference-correlation pattern is observed if the average cross-correlation map of the holo MD cluster best members is used in place of the holo crystal structure. All holo conformations show the same correlation behavior, including the two conformations from a previous holo CRP MD simulation study [133].

5.3.3. The Inter-domain Hinge as Allosteric Switch in CRP

GNM cross-correlation maps of apo NMR structures, holo CRP crystal structure and the conformations from the apo CRP monomer/dimer and holo CRP dimer MD simulations suggest a conformational switch mechanism mediated by the L134-D138 hinge in the allosteric communication of the DNA and cAMP binding domains. The network of cooperative fluctuations describes the off and on states of this switch: In the on state, the L134-D138 hinge displaying stronger fluctuations with the L134-D138 hinge of the neighboring subunit and the K52-E58 and G173-V176 regions in the cAMP and DNA binding domains, respectively, of both subunits, is able to coordinate the movement of the domains and the interactions in between. The coordinated behavior of the two DNA binding domains should provide the DNA binding residues with a more favorable dynamics for the DNA binding. On the other hand, the L134-D138 hinge is no longer a global hinge and only correlated with the DNA binding residues of its own subunit in the off state, when the coordination between the key functional elements of the structure is weakened. The on state and the off state are as well strongly coupled with the conformational preference of the C terminus of C helix. This preference is apparently correlated with the position of K52-E58 with respect to C helix.

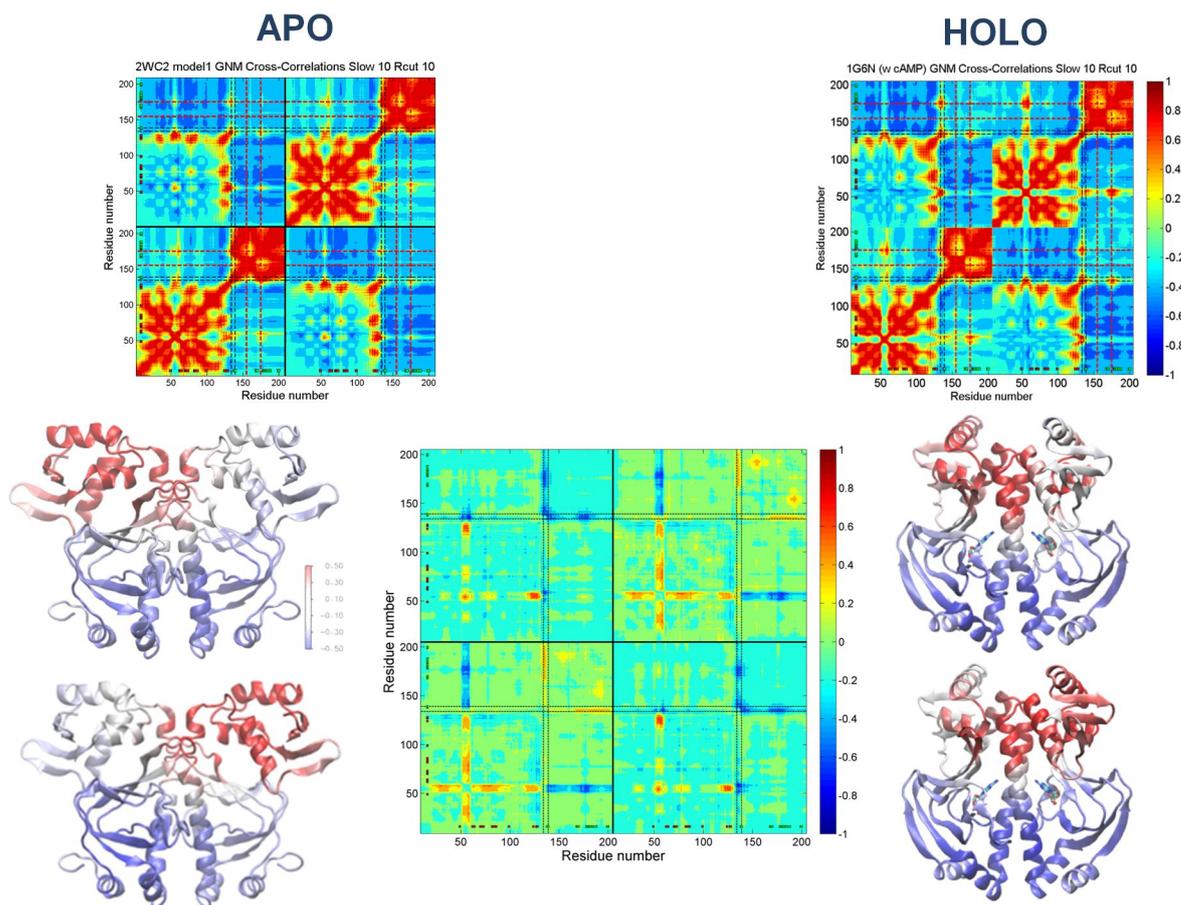


Figure 5.3. GNM cross-correlations of apo/holo CRP and the switch behavior. Comparison of the correlations between residue fluctuations in average ten slowest modes for the apo CRP NMR solution structure (PDB: 2WC2, model 1) and holo CRP X-ray crystal structure with the two cAMPs bound (PDB: 1G6N). Below, the ribbon diagrams color coded with the correlation values of the L134-D138 hinge (average) with the rest of the structure are given. The difference-correlation map of the two correlation maps is given in the middle.

As noted in the literature, upon binding of cAMP W85 is expelled into solvent and $\beta 4$ - $\beta 5$ flap (K52-E58) moves towards C helix, resulting in hydrophobic interactions of I51, K57, M59, L61 with F136 [135]. The $\beta 4$ - $\beta 5$ flap interacts with F136 of the hinge upon cAMP binding and stabilizes the L134-D138 inter-domain hinge. This, in other words, locks the on state with the coupled fluctuations of K52-E58 with the L134-D138, which in return reassumes its global hinge behavior for the inter-domain and then inter-subunit interactions; i.e., shifts the hinge region towards C terminus. The latter refers to the two end states of the allosteric transition pathway between apo and holo structures. Nevertheless, the GNM analysis of both NMR model structures and mainly MD sampled conformations shows that the on state and the off state with respect to both subunits is rare within the time window of simulations.

The dynamic infrastructure for the allosteric communication pre-exists in the apo state, yet the fluctuations are not fully organized for a proper communication of the two domains and the two subunits. Binding of the cAMPs organizes the couplings and elicits proper communication. When we look at the dynamics of the unbound monomer, the elements of a plausible allosteric mechanism is still observed: The fluctuations of the position of the L134-D138 hinge to N terminus, the unstable correlated fluctuations of L134-D138 with K52-E58, and unstable correlated fluctuations between the DNA binding (P154-A156 and G173-V176) and cAMP binding domains. To this end, it is plausible to state the key dynamic elements of the allosteric functional mechanism is hierarchically built up but yet stabilized to the fully functional state with the association of the structural units; the dimerization and cAMP binding.

5.3.4. Internal Fluctuations and Allostery

GNM assumes an ensemble of conformations around a given protein structure topology and predicts residue fluctuations. GNM may expand the MD sampled space through the predictions on mainly MD sampled conformations. Clustering is one way to reduce the MD conformational space into a subset of conformational states, where the conformational ensemble could be enlarged from each. The cluster best members could be considered as some energetically favorable conformational states. As alterna-

tive to distance based metric such as clustering and PCA, GNM was previously used to characterize different conformational states and dynamics along MD trajectories [86]. We have also used this idea of plotting the frequency distributions of the eigenvalue of the first (or first few) GNM eigenvector of a series of MD snapshots to characterize different conformational states visited during the simulations. The frequency distributions were seen to be sensitive to the states of CRP. The frequency values for the cluster best members were distributed homogeneously over the frequency values of all conformations (data not shown), which also show that the cluster best members could capture possible differences in the topology and dynamics of CRP.

The general patterns of the correlation maps are captured by the predicted correlations. Nevertheless, it is observed that, the larger the number of clusters for the MD conformations, the greater the difference between the MD/GNM cross-correlations of MD cluster best members. When there are large conformational changes, given the GNM calculations performed on mainly sampled conformations, the expansion in the conformational space could be larger compared to the relatively constrained cases. Also, for the MD cross correlations, averaging over whole trajectory might hide the dynamic behavior of less dominating conformational states.

GNM helps elaborating the dynamics assumed in each of these conformational states. Harmonic motion assumption and non-specificity in the underlying potential function are GNM's limitations together with the dependency on the given structure. On the other hand, PCA-based analyses of MD trajectories provide the dominant motion suggested by the MD time window, but it may change from one sampling window to another [154, 155]. Although usefulness of PCA analysis on insufficiently sampled MD trajectories may still be enhanced through multiple MD trajectories [156], the dominant dynamic behavior might have contributions from several conformational states through the trajectory and may not uncover the individual states well. The identified conformational states structurally might look similar, yet the dynamics can be affected with differences in some contacts if involving some key mechanical sites. The cooperative residue fluctuations may allow the propagation of the allosteric signal with the minimal structural changes in the mean conformation [157–160]. CRP variants,

although having structurally poised DNA interfaces, was seen to bind to DNA with different binding affinities; whereas the S62F CRP mutant, although the DNA binding domain is not reoriented to the active conformation, could show strong DNA binding affinity [143, 144, 159]. This suggests that binding may entirely be driven by the conformational entropy change. To this end, the dynamic analysis here demonstrates that the apo state has predisposed dynamics for both on and off state of the allosteric switch mechanism without undergoing the major conformational changes.

It has previously been demonstrated that the sequence evolution correlates with structural dynamics [161, 162]. It is expected that the key residues that mediate cooperative fluctuations could be conserved or assume correlated mutations. These are basically hinge sites that have been also suggested to overlap the regions of maximum frustrations that have a role in the emergence of allosteric interactions [163]. Allosteric functional motion and the cooperative modes are closely related. The robustness of the low frequency cooperative modes should depend on their nearly invariant nature, placing their foundations on the core network of residues responsible for transmitting signals as suggested by [164]. The local perturbations could be coupled to these modes which possibly transmit the signal by inducing conformational and/or dynamic changes encoded in the structure's topology [154, 165]. Here, as a contribution to the understanding of the allosteric mechanism, we used GNM and MD simulations combined to suggest the dynamic infrastructure of a possible conformational switch mechanism from apo CRP monomer to dimer, then to holo CRP dimer. The key features of the allosteric dynamics are encoded in apo CRP dimer as well as in apo CRP monomer, providing a basis to elicit the transmission of a signal from the cAMP binding site to DNA binding domains. The use of MD sampled conformations along GNM allows having more than one conformational state to be used in the GNM analysis, which could particularly be important for the cases where there are several conformational states accessible with some topological differences. For the details of this study, see the publication [166].

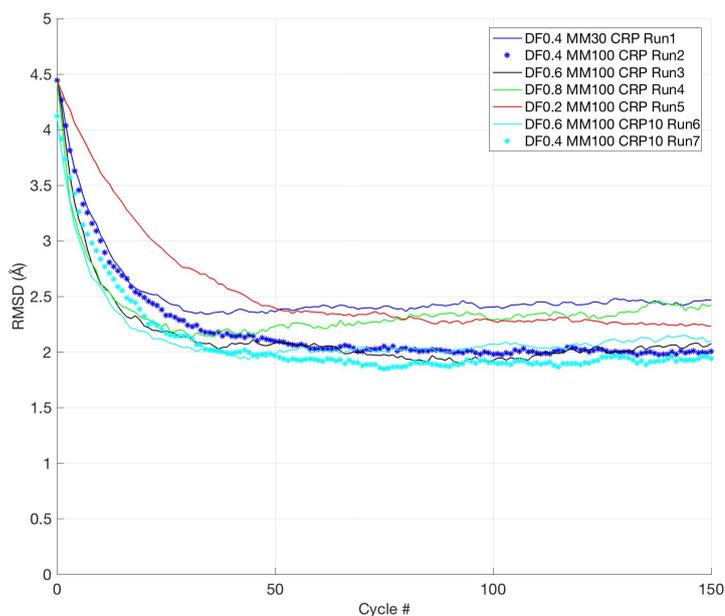


Figure 5.4. RMSD plots for CRP apo-holo forward runs (open-closed) for different DF values and $modemax$, $MM=30, 100$ ($T=310$ K, ANM $R_{cut}=13$ Å).

5.3.5. Global Dynamics by ANM-LD Simulations

ANM-LD simulations were performed to investigate the functional dynamics of CRP apo-holo conformational transition. Figure 5.4 shows the evolution of the RMSD values for each cycle for all the performed transition simulations. $MM=100$ and $DF=0.4$ and 0.6 Å runs are the most successful in terms of approaching the target state ($RMSD_{final} \leq 1.9$ Å) (See Table 5.2). Also, RMSD values are plotted separately in Figure 5.5 for the best parameter set with the corresponding selected modes, overlap, and degree of collectivity values for each cycle. Mostly selected modes are the slow modes 5 and 6 in most of the runs.

The GNM eV1 vs eV2 scatter plots given in Figure 5.6, exhibit the apo and holo MD conformations and the ANM-LD apo-holo transition pathways on top of each other. It is observed that for the assumed transition pathways, GNM eigenvalues are good at distinguishing the two distinct conformational states. Here, in ANM-LD simulations,

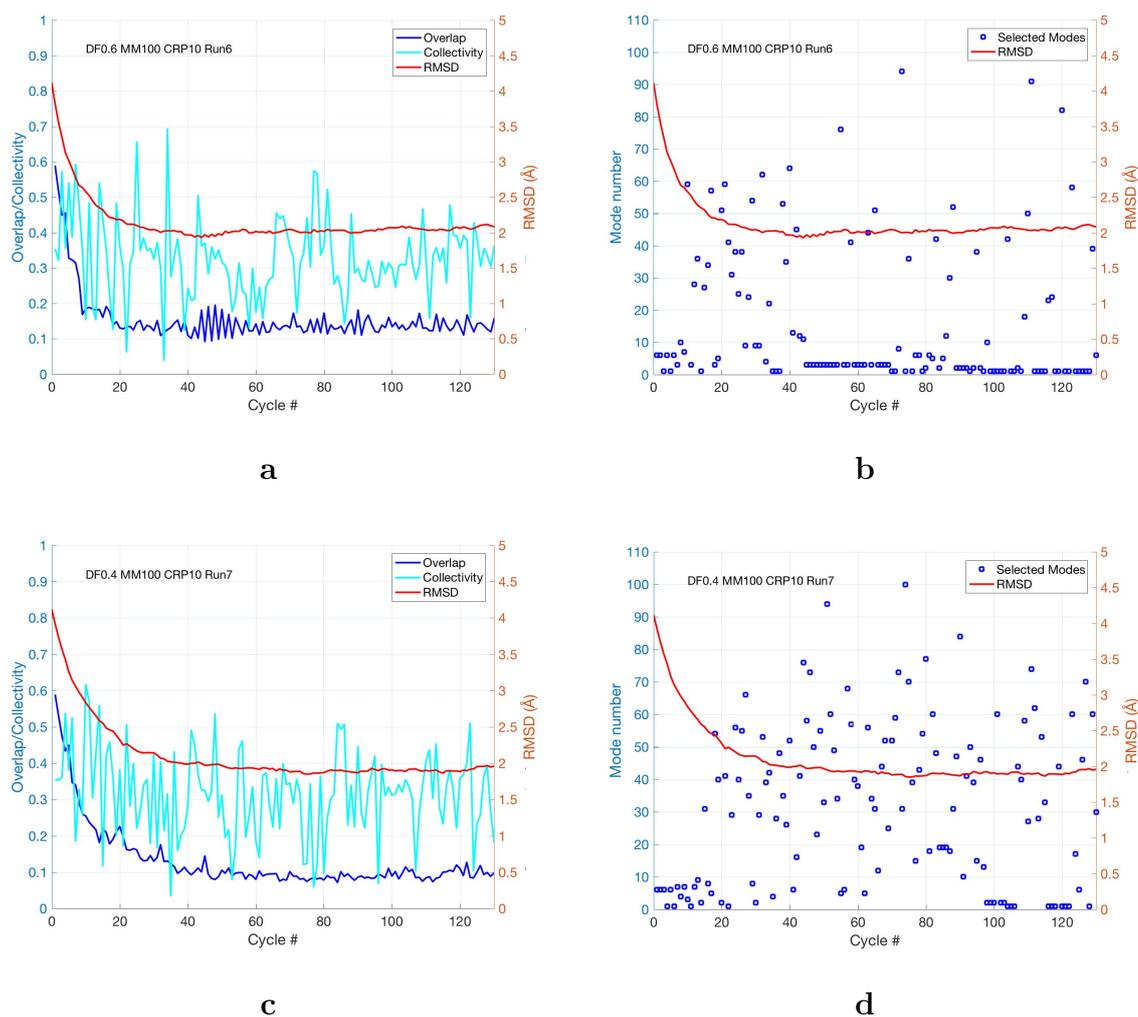


Figure 5.5. CRP Model 10 apo-holo simulations ($T=310$ K, *modemax*, MM=100, ANM $R_{cut}=13$ Å) overlap/collectivity plots for the selected slow modes. $DF=0.6$ Å Run (a), $DF=0.4$ Å Run (c), RMSD plots for the selected slow modes for $DF=0.6$ Å Run (b), $DF=0.4$ Å Run (d).

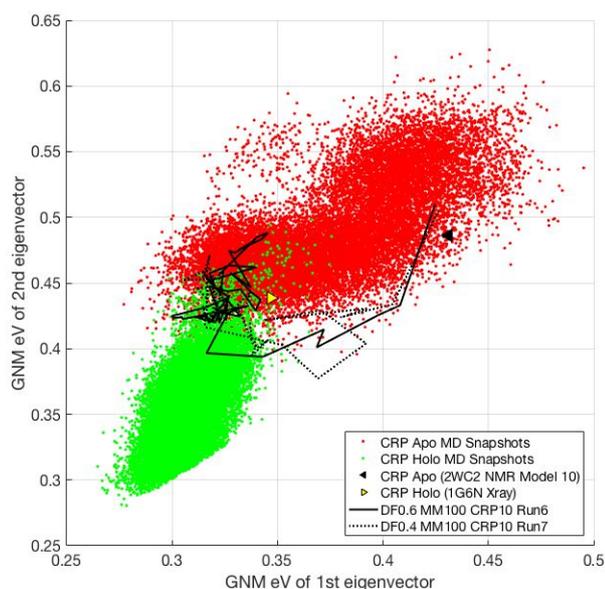


Figure 5.6. GNM eV1 versus eV2 plots for CRP Model 10 apo-holo simulations ($T=310$ K, ANM $R_{cut}=13$ Å, *modemax*, MM =100).

the presence of the ligand cAMP, which is essential for the activation of the protein, would effect the course of the transition pathways, as was observed for the Hsp90 case (See Chapter 4).

When we compare the GNM eV1 histogram plots (See Figure 5.7), the apo and holo conformational states sampled display the distinct behaviors of the two states. In the apo form the bimodal distribution exhibits the presence of the two-state ensemble of conformations, i.e., the apo form intrinsically have the ability to sample holo state. Upon binding of the ligand, the shift of the ensemble of conformations in favor of the active state can be observed in the holo state's narrow distribution (simulations were performed in the presence of cAMP).

For the major transition part of the pathways, the cross-correlation maps are constructed using the generated conformations in order to characterize the coupling dynamics of the conformational transition (See Figure 5.8). Cross-correlations for the assumed apo-holo transition pathway, resemble the structured/well-defined pattern of

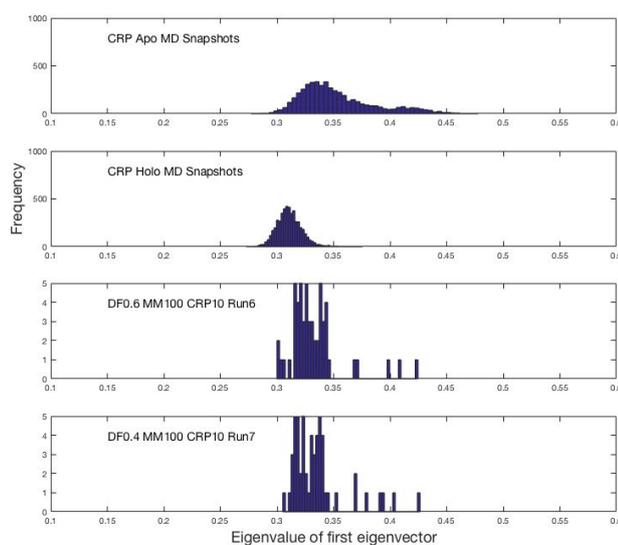


Figure 5.7. GNM eV1 histograms for CRP apo, holo MD simulations, and CRP Model 10 apo-holo ANM-LD simulations.

the holo state MD cross-correlations (See Figure 5.2C).

5.4. Conclusions

The MD simulations coupled with the GNM analysis of the apo and holo states of CRP, as well as the conformational transition between these two states were explored by ANM-LD simulations. The results have provided a mechanistic view on how the structural units are dynamically built up for a plausible allosteric functional mechanism; from apo CRP monomer to apo-to-holo CRP dimers and the transitions in between. The dimerization restricts the conformational states accessible to the structure, so does the cAMP binding, towards a favorable dynamics for the DNA binding. The key dynamic elements; the inter-domain hinge L134-D138 and the K52-E58, P154-A156 and G173-V176 sites, provide the dynamic infrastructure starting from the monomeric state and the orchestration of which leads to the allosteric communication between the cAMP and DNA binding sites/domains. A switch mechanism appears with the main role of the global hinge L134-D138; the on and off states are evidenced in apo CRP

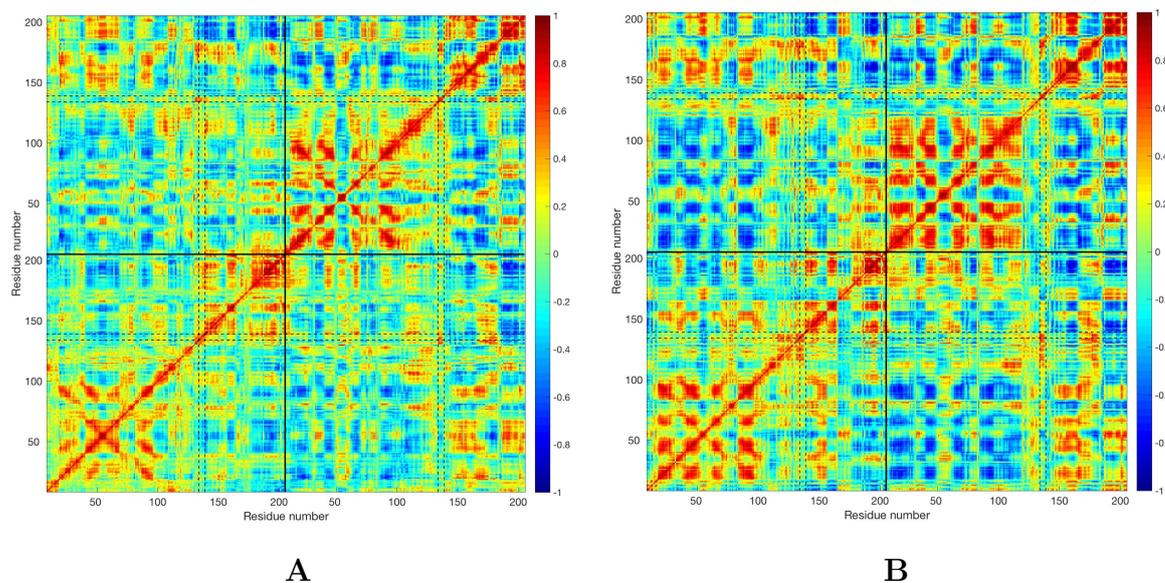


Figure 5.8. Cross-correlation plots for CRP Model 10 apo-holo runs, (a) DF=0.6 MM=100 Run 6 and (b) DF=0.4 MM=100 Run 7. The transition parts of the trajectories (just before RMSD to target reaches a plateau) are considered for the calculations.

dimer with the precursor dynamics as well observed in the monomeric form. Here the ANM-LD simulations could be performed in the presence of the ligand cAMP, which is essential for the activation of the CRP. In this way, the ligand's effect on the course of the transition pathways could be investigated.

6. CONCLUSIONS

In this thesis, a computationally efficient hybrid approach, ANM-LD, has been developed to generate conformational transition pathways between known end states of a protein and the coupling dynamics of the conformational transition. The methodology utilizes dominant low-frequency ANM modes to simulate the conformational transition from an initial towards a target conformation, taking advantage of the evaluation of the interactions and energetics of the system via short cycles of all-atom implicit LD simulations in between. The methodology is tested for a set of proteins, including proteins undergoing large conformational changes.

A detailed assessment of the ANM-LD methodology was carried out for a diverse set of proteins, including nine different cases, focusing on elaborately presenting the results for the prototype protein adenylate kinase, AdK. Transitions starting from initial RMSDs as high as 14 Å or 24 Å ending up with 2 Å approach towards the target state, demonstrated the ability of the intrinsic elastic modes to guide large conformational transitions reasonably well. Intrinsic dynamics enable conformational transitions towards the target state from an RMSD of 6.8 Å to 0.8 Å for AdK and from 6.85 Å to 3.38 Å for MalFGK2, with different parameter sets, implying the need to fine-tune/optimize the parameters according to the different properties of the systems studied.

The results presented for AdK show that the methodology is able to create physically meaningful/reactive and parallel pathways which is firstly assessed by the RMSD of the pathway intermediates from the experimentally available structures and the extent of approach towards the final state. The biological relevance of the pathways are also validated by the overlap of the created conformations with the experimental AdK structures in different states, defined by the degree of closeness of LID and NMP domains, and also with the AdK from different species. The LID-CORE and NMP-CORE angles are widely used as informative low-dimensional descriptors/collective variables in the recent literature for describing the extent of sampling for AdK conformational

transition by different methods and providing a common ground for comparisons.

Representing the progression of the transition pathways as a function of collective coordinates (if there is any for the studied system) defining the transition is a good way to monitor the transition. In addition to this, GNM eV1-eV2 scatter plots/distributions provide a practical means to view the differences in the pathways, being representative reaction coordinates, complementing the views from the specific collective coordinates. Moreover, where there are no collective coordinates identified to define the transition, GNM eV scatter plots/distributions could provide a means to compare the paths.

Certain global modes are selected repeatedly and the hindrance of these highly sampled cooperative modes lead to the evolution of alternative pathways. The mechanistic determinants in analyzing and controlling the transition pathways are elaborated by restricting the mostly selected modes, which resulted in a change in the correlation dynamics in both forward and reverse transitions for AdK. By the ANM-LD method restriction procedure we identified the main collective modes and regions coordinating the transition process. The procedure allowed us to modulate the progression (tune the correlation dynamics) of the transition process by restricting certain modes, either hindering the transition or ending up with alternative transition pathways.

The developed methodology is able to create conformational transition pathways guided by ANM towards a target state and have the potential to simulate large-scale conformational changes of small/large biomolecular complex structures. The association of functional large conformational changes with the structure's intrinsic elastic modes has been demonstrated in several previous studies. Nevertheless, the methodology developed here is unique in its entirety as a handy platform making it possible to create parallel transition pathways using the intrinsic dynamics information obtained from ANM by taking advantage of atomistic molecular simulations, controlled by many parameters that could easily be utilized to change the system conditions or incorporating any simulation methodology to the protocol by changing only a few parameters.

Hsp90, coupling of the global conformational change to the local binding behavior and dependence of these on the temperature is explored with AFM experiments and ANM-LD simulations.

The Bell parameters of Hsp90-Geldanamycin (GDM) interaction at each temperature are directly related to the binding energy landscape of the molecules. The experiment carried out for Hsp90-GDM showed that the disassociation rate of the low-strength state complex is lower at 310 K, whereas it is remarkably higher for the high-strength state (rupture at relatively low and high unbinding forces correspond to low-strength and high-strength states, respectively). However, the position of the energy barrier is the opposite. Since there is an inverse relation between the dissociation rate and the natural logarithm of the activation energy barrier, the results exhibit that increase in temperature decreases the activation energy barrier of the low-strength state but extremely increases the barrier at high-strength state (implying a higher $\Delta\Delta G$).

In order to investigate how inhibitor bound states disrupt the conformational dynamics of Hsp90 –with temperature dependence–, the lowT and highT cases for the ATP-free, ATP-bound and GDM-bound ANM-LD simulations were performed. Global binding behavior was observed to couple to the local binding behavior of the ATP/GDM, where both bind to the same cavity of Hsp90, but with totally different residue contacts which triggers different transition pathways. The effects of this local binding site were observed to couple to the global dynamics behavior through the hinge at charged linker (CL) region and appeared in the cross-correlation maps of the assumed conformational transition pathways of the ANM-LD simulations.

Stronger correlations observed in the ATP-free (Apo) and ATP-bound state cross-correlation maps implies the separation of the movements of the domains NTD and Md. This extra flexibility of NTD (caused by the undocked CL) has a crucial role in the open-close transition, by the closing of the NTDs together on the top of the molecule. The lack of these stronger correlations in the GDM-bound highT map, is a sign of decreased flexibility of NTD together with the disappearance of the hinges, in this case the transition should be slower. For the apo highT case, the NTD (including the ATP

binding sites listed in Structures section) highly correlates with the D560-K610 region of the CTD in both chains (both intra- and inter-chain), which are over 70 Å distant from each other. For the ATP-bound case, the said ATP binding sites-CTD correlation gets stronger within the each subunit (cross-subunit counterpart weakens). Here, the role of the aforementioned hinges is apparent in the conformational transition simulated. This is in agreement with a recent single-molecule experimental study stating the importance of the CL region in the progression of the conformational transition. Docked CL fixes the NTD to the Md, resulting in decreased flexibility and slower transition, whereas undocked CL comes with some flexibility and results in faster transition [105].

Hsp90 GDM-bound dynamics is observed to be dependent on the temperature (compared to the apo form dynamics), especially at highT, which is also the case for the experiments.

7. FUTURE WORK

The ANM-LD methodology would allow the investigation of any perturbation application, such as added external force, *in silico* mutations (exemplified by AdK mutated runs) or ligand binding (exemplified with the study of Hsp90 in the presence of ATP and the inhibitor geldanamycin).

The methodology could be further improved to include the energy as a criterion in the creation of the transition pathway and most importantly will be developed further to work in a non-targeted manner. This way, AdK mutation runs could be performed without the need for the target structure (which we had used and *in silico* model of mutated Adk structures as target conformation). Also, liganded simulations will be performed for AdK and all the GNM eV1-eV2 scatter plots could be compared to the PCA first two principal direction's (PC1-PC2) scatter plots for AdK.

Moreover, different simulation schemes could be performed (such as explicit solvent or self-guided LD) to see if the simulation scheme has significant effects on the transition pathways.

Suggestions for future work for Hsp90 would be to conduct new AFM experiments with different inhibitors, such as, GDM's less toxic analogs (tanespimycin-17-AAG, the clinical drug NVP-AUY922 etc.), radicicol and novobiocin (binding at the ATP binding site located at the C-termini, while other inhibitors bind to the N-termini), to compare the force ranges in the apo, ATP-bound and inhibitor-bound Hsp90 structures. Finding the biotinylated forms of these other inhibitors was not possible; therefore these experiments could not be performed in the scope of this study.

REFERENCES

1. Maximova, T., R. Moffatt, B. Ma, R. Nussinov and A. Shehu, “Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics”, *PLoS Computational Biology*, Vol. 12, No. 4, 2016.
2. Lindorff-Larsen, K., P. Maragakis, S. Piana and D. E. Shaw, “Picosecond to Millisecond Structural Dynamics in Human Ubiquitin”, *The Journal of Physical Chemistry B*, Vol. 120, No. 33, pp. 8313–8320, 2016.
3. Isin, B., K. Schulten, E. Tajkhorshid and I. Bahar, “Mechanism of signal propagation upon retinal isomerization: insights from molecular dynamics simulations of rhodopsin restrained by normal modes”, *Biophysical Journal*, Vol. 95, 2008.
4. Gur, M., J. D. Madura and I. Bahar, “Global Transitions of Proteins Explored by a Multiscale Hybrid Methodology: Application to Adenylate Kinase”, *Biophysical Journal*, Vol. 105, No. 7, pp. 1643–1652, 2013.
5. Das, A., M. Gur, M. H. Cheng, S. Jo, I. Bahar and B. Roux, “Exploring the Conformational Transitions of Biomolecular Systems Using a Simple Two-State Anisotropic Network Model”, *PLoS Computational Biology*, Vol. 10, No. 4, 2014.
6. Uyar, A., N. Kantarci-Carsibasi, T. Haliloglu and P. Doruker, “Features of large hinge-bending conformational transitions. Prediction of closed structure from open state”, *Biophysical Journal*, Vol. 106, pp. 2656–2666, 2014.
7. Ho, B. K., D. Perahia and A. M. Buckle, “Hybrid approaches to molecular simulation”, *Current Opinion in Structural Biology*, Vol. 22, No. 3, pp. 386–393, 2012.
8. Bernardi, R. C., M. C. Melo and K. Schulten, “Enhanced sampling techniques in molecular dynamics simulations of biological systems”, *Biochimica et Biophysica*

- Acta (BBA)-General Subjects*, Vol. 1850, No. 5, pp. 872–877, 2015.
9. Boomsma, W., P. Tian, J. Frellsen, J. Ferkinghoff-Borg, T. Hamelryck, K. Lindorff-Larsen and M. Vendruscolo, “Equilibrium simulations of proteins using molecular fragment replacement and NMR chemical shifts”, *Proceedings of the National Academic Science of the United States of America*, Vol. 111, No. 38, pp. 13852–13857, 2014.
 10. Panjkovich, A. and D. I. Svergun, “Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis”, *Physical Chemistry Chemical Physics*, Vol. 18, No. 8, pp. 5707–5719, 2016.
 11. Seifert, C. and F. Grater, “Force Distribution Reveals Signal Transduction in *E. coli* Hsp90.”, *Biophysical Journal*, Vol. 103, pp. 2195–2202, 2012.
 12. Ghanakota, P. and H. A. Carlson, “Moving Beyond Active-Site Detection: MixMD Applied to Allosteric Systems”, *The Journal of Physical Chemistry B*, Vol. 120, No. 33, pp. 8685–8695, 2016.
 13. Nussinov, R. and C. J. Tsai, “The different ways through which specificity works in orthosteric and allosteric drugs”, *Current Pharmaceutical Design*, Vol. 18, No. 9, pp. 1311–1316, 2012.
 14. Deniz, A. A., S. Mukhopadhyay and E. A. Lemke, “Single-molecule biophysics: at the interface of biology, physics and chemistry”, *Journal of the Royal Society Interface*, Vol. 5, No. 18, pp. 15–45, 2008.
 15. Binnig, G., C. F. Quate and C. Gerber, “Atomic force microscope”, *Physical Review Letters*, Vol. 56, No. 9, p. 930, 1986.
 16. Binnig, G., H. Rohrer, C. Gerber and E. Weibel, “Surface studies by scanning tunneling microscopy”, *Physical Review Letters*, Vol. 49, No. 1, p. 57, 1982.

17. Ando, T., T. Uchihashi and N. Kodera, “High-speed AFM and applications to biomolecular systems”, *Annual Review of Biophysics*, Vol. 42, pp. 393–414, 2013.
18. Eghiaian, F., F. Rico, A. Colom, I. Casuso and S. Scheuring, “High-speed atomic force microscopy: Imaging and force spectroscopy”, *FEBS Letters*, Vol. 588, No. 19, pp. 3631–3638, 2014.
19. Bell, G. I., “Models for the specific adhesion of cells to cells”, *Science*, Vol. 200, pp. 618–627, 1978.
20. Evans, E. and K. Ritchie, “Dynamic strength of molecular adhesion bonds”, *Biophysical Journal*, Vol. 72, No. 4, pp. 1541–55, 1997.
21. Evans, E. and F. Ludwig, “Dynamic strengths of molecular anchoring and material cohesion in fluid biomembranes”, *Journal of Physics: Condensed Matter*, Vol. 12, No. 8A, p. A315, 2000.
22. Bahar, I., T. R. Lezon, A. Bakan and I. H. Shrivastava, “Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins”, *Chemical Reviews*, Vol. 110, No. 3, pp. 1463–1497, 2010.
23. Tirion, M. M., “Large amplitude elastic motions in proteins from a single-parameter, atomic analysis”, *Physical Review Letters*, Vol. 77, No. 9, p. 1905, 1996.
24. Bahar, I., A. R. Atilgan and B. Erman, “Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential”, *Folding and Design*, Vol. 2, No. 3, pp. 173–181, 1997.
25. Haliloglu, T., I. Bahar and B. Erman, “Gaussian Dynamics of Folded Proteins”, *Physical Review Letters*, Vol. 79, No. 16, pp. 3090–3093, 1997.
26. Doruker, P., A. R. Atilgan and I. Bahar, “Dynamics of Proteins Predicted

- by Molecular Dynamics Simulations and Analytical Approaches: Application to Alpha-Amylase Inhibitor”, *Proteins-Structure Function and Bioinformatics*, Vol. 40, No. 3, pp. 512–524, 2000.
27. Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I. Bahar, “Anisotropy of fluctuation dynamics of proteins with an elastic network model”, *Biophysical Journal*, Vol. 80, No. 1, pp. 505–515, 2001.
 28. Tama, F. and Y.-H. Sanejouand, “Conformational change of proteins arising from normal mode calculations”, *Protein Engineering*, Vol. 14, No. 1, pp. 1–6, 2001.
 29. Haliloglu, T. and I. Bahar, “Adaptability of protein structures to enable functional interactions and evolutionary implications”, *Current Opinion in structural biology*, Vol. 35, pp. 17–23, 2015.
 30. López-Blanco, J. R. and P. Chacón, “New generation of elastic network models”, *Current Opinion in Structural Biology*, Vol. 37, pp. 46–53, 2016.
 31. McCammon, J. A., B. R. Gelin and M. Karplus, “Dynamics of folded proteins”, *Nature*, Vol. 267, No. 5612, pp. 585–590, 1977.
 32. Schlick, T., “Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules”, *F1000 biology reports*, Vol. 1, pp. 1–9, 2009.
 33. Adcock, S. and A. McCammon, “Molecular dynamics: survey of methods for simulating the activity of proteins”, *Chemical Reviews*, Vol. 106, No. 5, pp. 1589–1615, 2006.
 34. Freddolino, P. L., F. Liu, M. Gruebele and K. Schulten, “Ten-microsecond molecular dynamics simulation of a fast-folding WW domain”, *Biophysical Journal*, Vol. 94, No. 10, pp. L75–77, 2008.

35. Bjelkmar, P., P. S. Niemelä, I. Vattulainen and E. Lindahl, “Conformational changes and slow dynamics through microsecond polarized atomistic molecular simulation of an integral Kv1.2 ion channel”, *PLoS Computational Biology*, Vol. 5, No. 2, p. e1000289, 2009.
36. Shaw, D. E., P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan *et al.*, “Atomic-level characterization of the structural dynamics of proteins”, *Science*, Vol. 330, No. 6002, pp. 341–346, 2010.
37. Frauenfelder, H., S. G. Sligar and P. G. Wolynes, “The Energy Landscape and Motions of Proteins”, *Science*, Vol. 254, No. 5038, pp. 1598–1603, 1991.
38. Henzler-Wildman, K. and D. Kern, “Dynamic personalities of proteins”, *Nature*, Vol. 450, pp. 964–972, 2007.
39. Schlitter, J., M. Engels, P. Krüger, E. Jacoby and A. Wollmer, “Targeted Molecular Dynamics Simulation of Conformational Change—Application to the T - R Transition in Insulin”, *Molecular Simulation*, Vol. 10, No. 2, pp. 291–308, 1993.
40. van der Vaart, A. and M. Karplus, “Minimum free energy pathways and free energy profiles for conformational transitions based on atomistic molecular dynamics simulations”, *The Journal of Chemical Physics*, Vol. 126, No. 16, p. 164106, 2007.
41. Ovchinnikov, V. and M. Karplus, “Analysis and elimination of a bias in targeted molecular dynamics simulations of conformational transitions: Application to calmodulin”, *Journal of Physical Chemistry B*, Vol. 116, pp. 8584–8603, 2012.
42. Okazaki, K., N. Koga, S. Takada, J. N. Onuchic and P. G. Wolynes, “Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations”, *Proceedings of the National Academic Science of the United States of America*, Vol. 103, No. 32, pp. 11844–11849, 2006.

43. Koga, N. and S. Takada, “Folding-based molecular simulations reveal mechanisms of the rotary motor F1-ATPase”, *Proceedings of the National Academic Science of the United States of America*, Vol. 103, No. 14, pp. 5367–72, 2006.
44. Torrie, G. and J. Valleau, “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”, *Journal of Computational Physics*, Vol. 23, No. 2, pp. 187–199, 1977.
45. Grubmüller, H. and P. Tavan, “Molecular dynamics of conformational substates for a simplified protein model”, *Journal of Chemical Physics*, Vol. 101, No. 6, pp. 5047–5057, 1994.
46. Grubmüller, H., “Predicting slow structural transitions in macromolecular systems: Conformational flooding”, *Physical Review E*, Vol. 52, No. 3, pp. 2893–2906, 1995.
47. Voter, A. F., “Hyperdynamics: Accelerated molecular dynamics of infrequent events”, *Physical Review Letters*, Vol. 78, No. 20, pp. 3908–3911, 1997.
48. Hamelberg, D., J. Mongan and J. A. McCammon, “Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules”, *The Journal of Chemical Physics*, Vol. 120, No. 24, pp. 11919–29, 2004.
49. Markwick, P. R., L. C. Pierce, D. B. Goodin and J. A. McCammon, “Adaptive Accelerated Molecular Dynamics (Ad-AMD) Revealing the Molecular Plasticity of P450cam”, *The Journal of Physical Chemistry Letters*, pp. 158–164, 2011.
50. Kong, Y., Y. Shen, T. E. Warth and J. Ma, “Conformational pathways in the gating of Escherichia coli mechanosensitive channel”, *Proceedings of the National Academic Science of the United States of America*, Vol. 99, No. 9, pp. 5999–6004, 2002.
51. Okazaki, K. and S. Takada, “Dynamic energy landscape view of coupled binding

- and protein conformational change: induced-fit versus population-shift mechanisms”, *Proceedings of the National Academic Science of the United States of America*, Vol. 105, No. 32, pp. 11182–11187, 2008.
52. Maragakis, P. and M. Karplus, “Large Amplitude Conformational Change in Proteins Explored with a Plastic Network Model: Adenylate Kinase”, *Journal of Molecular Biology*, Vol. 352, No. 4, pp. 807–822, 2005.
53. Mills, M. and I. Andricioaei, “An experimentally guided umbrella sampling protocol for biomolecules”, *The Journal of Chemical physics*, Vol. 129, No. 11, p. 114101, 2008.
54. Schulze, B. G., H. Grubmüller and J. D. Evanseck, “Functional Significance of Hierarchical Tiers in Carbonmonoxy Myoglobin: Conformational Substates and Transitions Studied by Conformational Flooding Simulations”, *Journal of the American Chemical Society*, Vol. 122, No. 36, pp. 8700–8711, 2000.
55. Roitberg, A. and R. Elber, “Modeling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations”, *The Journal of Chemical Physics*, Vol. 95, No. 12, p. 9277, 1991.
56. Sugita, Y. and Y. Okamoto, “Replica-exchange molecular dynamics method for protein folding”, *Chemical Physics Letters*, Vol. 314, No. 1-2, pp. 141–151, 1999.
57. Chebaro, Y., X. Dong, R. Laghaei, P. Derreumaux and N. Mousseau, “Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent”, *The Journal of Physical Chemistry.B*, Vol. 113, No. 1, pp. 267–274, 2009.
58. Bolhuis, P. G., “Rare events via multiple reaction channels sampled by path replica exchange”, *The Journal of Chemical Physics*, Vol. 129, No. 11, p. 114108, 2008.

59. Ryckaert, J., G. Ciccotti and H. Berendsen, “Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes”, *Journal of Computational Physics*, Vol. 23, pp. 327–341, 1977.
60. Sweet, C. R., P. Petrone, V. S. Pande and J. A. Izaguirre, “Normal mode partitioning of Langevin dynamics for biomolecules.”, *The Journal of Chemical Physics*, Vol. 128, No. 14, p. 145101, 2008.
61. Kaledin, M., A. L. Kaledin, A. Brown and J. M. Bowman, “Driven Molecular Dynamics for Normal Modes of Biomolecules without the Hessian, and Beyond”, Q. Cui and I. Bahar (Editors), *Normal Mode Analysis: Theory and applications to biological and chemical systems*, chap. 14, pp. 281–300, CRC press, 2005.
62. Zhang, Z. Y., Y. Y. Shi and H. Y. Liu, “Molecular Dynamics Simulations of Peptides and Proteins with Amplified Collective Motions”, *Biophysical Journal*, Vol. 84, No. 6, pp. 3583–3593, 2003.
63. Isralewitz, B., J. Baudry, J. Gullingsrud, D. Kosztin and K. Schulten, “Steered molecular dynamics investigations of protein function.”, *Journal of Molecular Graphics and Modelling*, Vol. 19, No. 1, pp. 13–25, 2001.
64. Sotomayor, M. and K. Schulten, “Single-molecule experiments in vitro and in silico”, *Science*, Vol. 316, No. 5828, pp. 1144–1148, 2007.
65. Stone, J. E., J. Gullingsrud and K. Schulten, “A system for interactive molecular dynamics simulation”, *Proceedings of the 2001 symposium on Interactive 3D graphics*, pp. 191–194, ACM, 2001.
66. Leone, V., F. Marinelli, P. Carloni and M. Parrinello, “Targeting biomolecular flexibility with metadynamics”, *Current Opinion in Structural Biology*, Vol. 20, No. 2, pp. 148–154, 2010.
67. Vanden-Eijnden, E., M. Venturoli, G. Ciccotti and R. Elber, “On the assumptions

- underlying milestoning”, *The Journal of Chemical Physics*, Vol. 129, No. 17, p. 174102, 2008.
68. Yang, Z., P. Majek and I. Bahar, “Allosteric Transitions of Supramolecular Systems Explored by Network Models: Application to Chaperonin GroEL”, *PloS Computational Biology*, Vol. 5, No. 4, p. e1000360, 2009.
69. Kantarci-Carsibasi, N., T. Haliloglu and P. Doruker, “Conformational Transition Pathways Explored by Monte Carlo Simulation Integrated with Collective Modes”, *Biophysical Journal*, Vol. 95, No. 12, pp. 5862–5873, 2008.
70. Zhuravlev, P. I. and G. Papoian, “Protein functional landscapes, dynamics, allostery: A tortuous path towards a universal theoretical framework.”, *Quarterly Reviews of Biophysics*, Vol. 43, No. 3, pp. 295–332, 2010.
71. Seyler, S. L., A. Kumar, M. F. Thorpe and O. Beckstein, “Path Similarity Analysis: A Method for Quantifying Macromolecular Pathways”, *PLoS Computational Biology*, Vol. 11, No. 10, p. e1004568, 2015.
72. Karplus, M. and G. A. Petsko, “Molecular dynamics simulations in biology”, *Nature*, Vol. 347, No. 6294, pp. 631–639, 1990.
73. Verlet, L., “Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”, *Physical Review*, Vol. 159, No. 1, pp. 98–103, 1967.
74. Becker, O. M., A. D. MacKerell Jr, B. Roux and M. Watanabe (Editors), *Computational Biochemistry and Biophysics*, CRC Press, 2001.
75. Schlick, T., *Molecular Modeling and Simulation: An interdisciplinary guide*, Springer, 2010.
76. Brünger, A., C. L. I. Brooks and M. Karplus, “Stochastic boundary conditions

- for molecular dynamics simulations of ST2 water”, *Chemical Physics Letters*, Vol. 105, No. 5, pp. 495–500, 1984.
77. Brooks, C. . B., A. . Karplus and M., “Active site dynamics in protein molecules: A stochastic boundary molecular-dynamics approach”, *Biopolymers*, Vol. 24, pp. 843–865, 1985.
78. Brüschweiler, R., “Collective protein dynamics and nuclear spin relaxation”, *The Journal of Chemical Physics*, Vol. 102, No. 8, pp. 3396–3403, 1995.
79. Fleishman, S. J., T. A. Whitehead, E.-M. Strauch, J. E. Corn, S. Qin, H.-X. Zhou, J. C. Mitchell, O. N. Demerdash, M. Takeda-Shitaka, G. Terashi *et al.*, “Community-wide assessment of protein-interface modeling suggests improvements to design methodology”, *Journal of Molecular Biology*, Vol. 414, No. 2, pp. 289–302, 2011.
80. Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, “A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations”, *Journal of Computational Chemistry*, Vol. 24, pp. 1999–2012, 2003.
81. Case, D. A., T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman, *AMBER 11*, University of California, San Francisco, 2010.
82. Dolinsky, T. J., P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe and N. A. Baker, *PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations*, *Nucleic acids research*, 35(suppl 2), W522-W525, 2007.

83. Anandakrishnan, R., B. Aguilar and A. V. Onufriev, “H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations”, *Nucleic Acids Research*, Vol. 40, No. W1, pp. W537–W541, 2012.
84. Loncharich, R. J., B. R. Brooks and R. W. Pastor, “Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide”, *Biopolymers*, Vol. 32, No. 5, pp. 523–535, 1992.
85. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, “Molecular dynamics with coupling to an external bath”, *The Journal of Chemical Physics*, Vol. 81, pp. 3684–3690, 1984.
86. Hall, B. A., S. L. Kaye, A. Pang, R. Perera and P. C. Biggin, “Characterization of protein conformational states by normal-mode frequencies”, *Journal of the American Chemical Society*, Vol. 129, No. 37, pp. 11394–11401, 2007.
87. Formoso, E., V. Limongelli and M. Parrinello, “Energetics and Structural Characterization of the large-scale Functional Motion of Adenylate Kinase”, *Scientific Reports*, Vol. 5, No. 8425., 2015.
88. Kerns, S. J., R. V. Agafonov, Y. J. Cho, F. Pontiggia, R. Otten and . Pachov, D. V., “The energy landscape of adenylate kinase during catalysis”, *Nature Structural and Molecular Biology*, Vol. 22, No. 2, pp. 124–131, 2015.
89. Müller, C. W., G. J. Schlauderer, J. Reinstein and G. E. Schulz, “Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding”, *Structure*, Vol. 4, No. 2, pp. 147–156, 1996.
90. Müller, C. W. and G. E. Schulz, “Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution: A model for a catalytic transition state”, *Journal of Molecular Biology*, Vol. 224, No. 1, pp. 159–177, 1992.

91. Khare, D., M. L. Oldham, C. Orelle, A. L. Davidson and J. Chen, “Alternating access in maltose transporter mediated by rigid-body rotations”, *Molecular Cell*, Vol. 33, No. 4, pp. 528–536, 2009.
92. Oldham, M. L. and J. Chen, “Crystal structure of the maltose transporter in a pretranslocation intermediate state”, *Science*, Vol. 332, No. 6034, pp. 1202–1205, 2011.
93. Fiser, A., R. K. G. Do and A. Sali, “Modeling of loops in protein structures”, *Protein Science*, Vol. 9, No. 09, pp. 1753–1773, 2000.
94. Fiser, A. and A. Sali, “ModLoop: Automated modeling of loops in protein structures”, *Bioinformatics*, Vol. 19, No. 18, pp. 2500–2501, 2003.
95. Ramachandran, S., P. Kota, F. Ding and N. V. Dokholyan, “Automated minimization of steric clashes in protein structures”, *PROTEINS: Structure, Function and Bioinformatics*, Vol. 79, pp. 261–270, 2011.
96. Whitford, P. C., S. Gosavi and J. N. Onuchic, “Conformational transitions in adenylate kinase: Allosteric communication reduces misligation”, *Journal of Biological Chemistry*, Vol. 283, pp. 2042–2048, 2008.
97. Beckstein, O., E. J. Denning, J. R. Perilla and T. B. Woolf, “Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open-closed transitions”, *Journal of Molecular Biology*, Vol. 394, No. 1, pp. 160–176, 2009.
98. Whitford, P. C., O. Miyashita, Y. Levy and J. N. Onuchic, “Conformational Transitions of Adenylate Kinase: Switching by Cracking”, *Journal of Molecular Biology*, Vol. 366, pp. 1661–1671, 2007.
99. Daily, M. D., H. Yu, P. Jr, G. N. and Q. Cui, “Allosteric activation transitions in enzymes and biomolecular motors: insights from atomistic and coarse-grained simulations”, *Dynamics in Enzyme Catalysis*, pp. 139–164, 2013.

100. Beckstein, O., E. J. Denning, J. R. Perilla and T. B. Woolf, *Adenylate Kinase Potential of Mean Force*, 2009, <https://becksteinlab.physics.asu.edu/research/52/adk-apo-pmf>, accessed at Jul 2016.
101. Bae, E. and G. N. Phillips, “Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases”, *Journal of Biological Chemistry*, Vol. 279, No. 27, pp. 28202–28208, 2004.
102. Kovermann, M., J. Ådén, C. Grundström, A. E. Sauer-Eriksson, U. H. Sauer and M. Wolf-Watz, “Structural basis for catalytically restrictive dynamics of a high-energy enzyme state”, *Nature Communications*, Vol. 6, 2015.
103. Krukenberg, K. A., T. O. Street, L. A. Lavery and D. A. Agard, “Conformational dynamics of the molecular chaperone Hsp90”, *Quarterly Reviews of Biophysics*, Vol. 44, pp. 229–255, 2011.
104. Taipale, M., D. F. Jarosz and S. Lindquist, “HSP90 at the hub of protein homeostasis: emerging mechanistic insights”, *Nature Reviews. Molecular Cell Biology*, Vol. 11, pp. 515–528, 2010.
105. Jahn, M., A. Rehn, B. Pelz, B. Hellenkamp, K. Richter, M. Rief, J. Buchner and T. Hugel, “The charged linker of the molecular chaperone Hsp90 modulates domain contacts and biological function”, *Proceedings of the National Academic Science of the United States of America*, Vol. 111, No. 50, pp. 17881–17886, 2014.
106. Partridge, J. R., L. A. Lavery, D. Elnatan, N. Naber, R. Cooke and D. A. Agard, “A novel N-terminal extension in mitochondrial TRAP1 serves as a thermal regulator of chaperone activity”, *eLife*, Vol. 3, p. e03487, 2015.
107. Morra, G., G. Verkhivker and G. Colombo, “Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full-length dimer”, *PLoS Computational Biology*, Vol. 5, No. 3, p. e1000323,

- 2009.
108. Neckers, L. and P. Workman, “Hsp90 molecular chaperone inhibitors: Are we there yet?”, *Clinical Cancer Research*, Vol. 18, No. 1, pp. 64–76, 2012.
 109. Hutter, J. L. and J. Bechhoefer, “Calibration of atomic-force microscope tips”, *Review of Scientific Instruments*, Vol. 64, No. 7, pp. 1868–1873, 1993.
 110. Lee, C. K., Y. M. Wang, L. S. Huang and S. Lin, “Atomic force microscopy: determination of unbinding force, off rate and energy barrier for protein-ligand interaction”, *Micron*, Vol. 38, No. 5, pp. 446–461, 2007.
 111. Shiau, A. K., S. F. Harris, D. R. Southworth and D. A. Agard, “Structural analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements.”, *Cell*, Vol. 127, No. 2, pp. 329–340, 2006.
 112. Ali, M. M., S. M. Roe, C. K. Vaughan, P. Meyer, B. Panaretou, P. W. Piper, C. Prodromou and L. H. Pearl, “Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex”, *Nature*, Vol. 440, No. 7087, pp. 1013–1017, 2006.
 113. Laskowski, R. A., E. G. Hutchinson, A. D. Michie, A. C. Wallace, M. L. Jones *et al.*, “PDBsum: A web-based database of summaries and analyses of all PDB structures”, *Trends in Biochemical Sciences*, Vol. 22, pp. 488–490, 1997.
 114. Stebbins, C. E., A. A. Russo, C. Schneider, N. Rosen, F. U. Hartl and N. P. Pavletich, “Crystal structure of an Hsp90?geldanamycin complex: targeting of a protein chaperone by an antitumor agent”, *Cell*, Vol. 89, No. 2, pp. 239–250, 1997.
 115. Frisch, M. J. . T., G. W. . Schlegel, H. B. . Scuseria, G. E. . Robb, M. A. . Cheeseman, J. R. . Scalmani, G. . Barone, V. . Mennucci, B. . Petersson, G. A. . Nakatsuji, H. . Caricato, M. . Li, X. . Hratchian, H. P. . Izmaylov, A. F. . Bloino,

- J. . Zheng, G. . Sonnenberg, J. L. . Hada, M. . Ehara, M. . Toyota, K. . Fukuda, R. . Hasegawa, J. . Ishida, M. . Nakajima, T. . Honda, Y. . Kitao, O. . Nakai, H. . Vreven, T. . Montgomery and J. J. A., *Gaussian*, Wallingford CT, 2009.
116. Meagher, K. L., L. T. Redman and H. A. Carlson, “Development of polyphosphate parameters for use with the AMBER force field”, *Journal of Computational Chemistry*, Vol. 24, No. 9, pp. 1016–1025, 2003.
117. Aiba, H., S. Fujimoto and N. Ozaki, “Molecular cloning and nucleotide sequencing of the gene for E. coli cAMP receptor protein.”, *Nucleic Acids Research*, Vol. 10, pp. 1345–1361, 1982.
118. Harman, J. G., “Allosteric regulation of the cAMP receptor protein”, *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, Vol. 1547, pp. 1–17, 2001.
119. Weber, I. T. and T. A. Steitz, “Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution.”, *Journal of Molecular Biology*, Vol. 198, pp. 311–326, 1987.
120. Passner, J. M. and T. A. Steitz, “The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer”, *Proceedings of the National Academic Science of the United States of America*, Vol. 94, pp. 2843–2847, 1997.
121. Tutar, Y., “Syn-, anti, and finally both: Cyclic AMP conformation altered CRP dependent transcription initiation mechanism in E coli lac operon.”, *Cell Biochemistry and Function*, Vol. 26, pp. 399–405, 2008.
122. McKay, D. B. and T. A. Steitz, “Structure of catabolite gene activator protein at 2”, *9 Å resolution suggests binding to left-handed B-DNA. Nature*, Vol. 290, pp. 745–749, 1981.
123. Scott, S. P. and S. Jarjous, “Proposed structural mechanism of Escherichia coli

- cAMP receptor protein cAMP-dependent proteolytic cleavage protection and selective and nonselective DNA binding”, *Biochemistry*, Vol. 44, pp. 8370–8748, 2005.
124. Jin, R., K. Sharif and J. Krakow, “Evidence for contact between the cyclic AMP receptor protein and the subunit of Escherichia coli RNA polymerase”, *Journal of Biological Chemistry*, Vol. 270, pp. 19213–19216, 1995.
125. Williams, R. M., V. A. Rhodius, A. I. Bell, A. Kolb and S. Busby, “Orientation of functional activating regions in the Escherichia Coli CRP protein during transcription activation at class II promoters”, *Nucleic Acids Research*, Vol. 24, pp. 1112–1118, 1996.
126. Busby, S. and R. H. Ebright, “Transcription activation by catabolite activator protein (CAP)”, *Journal of Molecular Biology*, Vol. 293, pp. 199–213, 1999.
127. Benoff, B., H. Yang, C. L. Lawson, G. Parkinson, J. Liu *et al.*, “Structural basis of transcription activation: the CAP- α CTD-DNA complex”, *Science*, Vol. 297, pp. 1562–1566, 2002.
128. Lawson, C. L., D. Swigon, K. S. Murakami, S. A. Darst, H. M. Berman *et al.*, “Catabolite activator protein: DNA binding and transcription activation”, *Current Opinion in Structural Biology*, Vol. 14, pp. 10–20, 2004.
129. Won, H. S., Y. S. Lee, S. H. Lee and B. J. Lee, “Structural overview on the allosteric activation of cyclic AMP receptor protein”, *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, Vol. 1794, pp. 1299–1308, 2009.
130. Zubay, G., D. Schwartz and J. Beckwith, “Mechanism of activation of catabolite-sensitive genes: A positive control system”, *Proceedings of the National Academic Science of the United States of America*, Vol. 66, pp. 104–110, 1970.
131. Tomlinson, S. R., Y. Tutar and J. G. Harman, “CRP subunit association and

- hinge conformation changes in response to cAMP binding: Analysis of C-helix cysteine-substituted CRP”, *Biochemistry*, Vol. 45, pp. 13438–13446, 2006.
132. Garcia, A. and J. G. Harman, “Simulations of CRP:(cAMP) 2 in noncrystalline environments show a subunit transition from the open to the closed conformation”, *Protein Science*, Vol. 5, pp. 62–71, 1996.
 133. Berrera, M., S. Pantano and P. Carloni, “Catabolite activator protein in aqueous solution: A molecular simulation study”, *Journal of Physical Chemistry*, Vol. 111, pp. 1496–1501, 2007.
 134. Li, L., V. N. Uversky, A. K. Dunker and S. O. Meroueh, “A computational investigation of allostery in the catabolite activator protein”, *Journal of the American Chemical Society*, Vol. 129, pp. 15668–15676, 2007.
 135. Popovych, N., S. R. Tzeng, M. Tonelli, R. H. Ebright and C. G. Kalodimos, “Structural basis for cAMP-mediated allosteric transition in the catabolite activator protein”, *Proceedings of the National Academic Science of the United States of America*, Vol. 106, pp. 6927–6932, 2009.
 136. Schultz, S. C., G. Shields and T. A. Steitz, “Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees”, *Science*, Vol. 253, No. 5023, pp. 1001–1007, 1991.
 137. Passner, J. M., S. C. Schultz and T. A. Steitz, “Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 Å resolution.”, *Journal of Molecular Biology*, Vol. 304, pp. 847–859, 2000.
 138. Tzeng, S. R. and C. G. Kalodimos, “Dynamic activation of an allosteric regulatory protein”, *Nature*, Vol. 462, pp. 368–372, 2009.
 139. Won, H. Y., T. Yamazaki, T. W. Lee, M. K. Yoon, S. H. Park *et al.*, “Structural understanding of the allosteric conformational change of cyclic AMP receptor

- protein by cyclic AMP binding”, *Biochemistry*, Vol. 39, pp. 13953–13962, 2000.
140. Baker, C. H., S. R. Tomlinson, A. E. Garcia and J. G. Harman, “Amino acid substitution at position 99 affects the rate of CRP subunit exchange”, *Biochemistry*, Vol. 40, pp. 12329–12338, 2001.
141. Won, H. S., M. D. Seo, H. S. Ko, W. S. Choi and B. J. Lee, “Interdomain interaction of cyclic AMP receptor protein in the absence of cyclic AMP”, *Journal of Biochemistry*, Vol. 143, pp. 163–167, 2008.
142. Kim, J., S. Adhya and S. Garges, “Allosteric changes in the cAMP receptor protein of *Escherichia coli*: hinge reorientation”, *Proceedings of the National Academic Science of the United States of America*, Vol. 89, pp. 9700–9704, 1992.
143. Kalodimos, C. G., “Protein function and allostery: A dynamic relationship”, *Annals of the New York Academy of Sciences*, Vol. 1260, No. 1, pp. 81–86, 2012.
144. Tzeng, S. R. and C. G. Kalodimos, “Protein activity regulation by conformational entropy”, *Nature*, Vol. 488, pp. 236–240, 2012.
145. Changeux, J. P. and S. J. Edelstein, “Allosteric mechanisms of signal transduction”, *Science*, Vol. 308, pp. 1424–1428, 2005.
146. Changeux, J. P., “Allostery and the Monod-Wyman-Changeux model after 50 years”, *Annual Review of Biophysics*, Vol. 41, pp. 103–133, 2011.
147. Ma, B., C. J. Tsai, T. Haliloglu and R. Nussinov, “Dynamic allostery: Linkers are not merely flexible”, *Structure*, Vol. 19, pp. 907–917, 2011.
148. Case, D. A., T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, “The Amber biomolecular simulation programs”, *Journal of Computational Chemistry*, Vol. 26, No. 16, pp. 1668–1688, 2005.

149. Case, D. A., T. A. Darden, T. E. I. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, B. Wang and D. A. e. a. Pearlman, *AMBER 8*, University of California, San Francisco, 2004.
150. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, “Comparison of simple potential functions for simulating liquid water”, *Journal of Chemical Physics*, Vol. 79, pp. 926–935, 1983.
151. Gordon, J. C., J. B. Myers, T. Folta, V. Shoja, L. S. Heath *et al.*, “H++: A server for estimating pKas and adding missing hydrogens to macromolecules”, *Nucleic Acids Research*, Vol. 33, pp. W368–W371, 2005.
152. Bashford, D. and M. Karplus, “pKa’s of ionizable groups in proteins: Atomic detail from a continuum electrostatic model”, *Biochemistry*, Vol. 29, pp. 10219–10225, 1990.
153. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, “A smooth particle mesh Ewald method”, *Journal of Chemical Physics*, Vol. 103, pp. 8577–8593, 1995.
154. Meireles, L., M. Gur, A. Bakan and I. Bahar, “Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins”, *Protein Science*, Vol. 20, No. 10, pp. 1645–1658, 2011.
155. Balsera, M. A., W. Wriggers, Y. Oono and K. Schulten, “Principal component analysis and long time protein dynamics”, *The Journal of Physical Chemistry*, Vol. 100, No. 7, pp. 2567–2572, 1996.
156. Liu, L., A. M. Gronenborn and I. Bahar, “Longer simulations sample larger subspaces of conformations while maintaining robust mechanisms of motion”, *Proteins*, Vol. 80, pp. 616–625, 2011.
157. Tsai, C.-J., A. Del Sol and R. Nussinov, “Allostery: absence of a change in shape

- does not imply that allostery is not at play”, *Journal of Molecular Biology*, Vol. 378, No. 1, pp. 1–11, 2008.
158. Tzeng, S. R. and C. G. Kalodimos, “Protein dynamics and allostery: An NMR view”, *Current Opinion in Structural Biology*, Vol. 21, pp. 62–67, 2011.
159. Kalodimos, C. G., “NMR reveals novel mechanisms of protein activity regulation”, *Protein Science*, Vol. 20, pp. 773–782, 2011.
160. Popovych, N., S. Sun, R. H. Ebricht and C. G. Kalodimos, “Dynamically driven protein allostery”, *Nature Structural and Molecular Biology*, Vol. 13, pp. 831–838, 2006.
161. Liu, Y. and I. Bahar, “Sequence evolution correlates with structural dynamics”, *Molecular Biology and Evolution*, Vol. 29, No. 9, pp. 2253–2263, 2012.
162. Maguid, S., S. Fernandez-Alberti, G. Parisi and J. Echave, “Evolutionary conservation of protein vibrational dynamics”, *Gene*, Vol. 422, pp. 7–13, 2008.
163. Ferreiro, D. U., J. A. Hegler, E. A. Komives and P. G. Wolynes, “On the role of frustration in the energy landscapes of allosteric proteins”, *Proceedings of the National Academic Science of the United States of America*, Vol. 108, pp. 3499–3503, 2011.
164. Zheng, W., B. R. Brooks and D. Thirumalai, “Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations”, *Proceedings of the National Academic Science of the United States of America*, Vol. 103, pp. 7664–7669, 2006.
165. Cecchini, M., A. Houdusse and M. Karplus, “Allosteric communication in myosin V: From small conformational changes to large directed movements”, *PLoS Computational Biology*, Vol. 4, No. 8, p. e1000129, 2008.

166. Aykac Fas, B., Y. Tutar and T. Haliloglu, “Dynamic Fluctuations Provide the Basis of a Conformational Switch Mechanism in Apo Cyclic AMP Receptor Protein”, *PLoS Computational Biology*, Vol. 9, No. 7, 2013.