

Microscopic Insight To Protein Functions

Thesis submitted for the degree of
Doctor of Philosophy (Science)

in

Physics (Theoretical)

by

Sutapa Dutta

Department of Physics

University of Calcutta

July, 2018

Dedicated to my family & friends.....

Acknowledgements

With deep regards and profound respect, I take this opportunity to convey my deepest gratitude to my supervisor Prof. Dr. Jaydeb Chakrabarti for introducing me to such an interesting and advanced field of research. His invaluable guidance, advice, technical discussion, constructive criticism, extensive support, constant encouragement and valuable suggestions kept me inspired throughout this research work. He has been always my friend, philosopher and guide, all moulded into one person. I found joy and enthusiasm for his research and dedication towards work. I am also very much grateful to my joint supervisor Dr. Mahua Ghosh for her extensive encouragement and spending her valuable time towards the academic discussions during my research work. ‘A good teacher knows how to bring out the best in student’, this completely goes well with her who is the pioneer of my journey to the beautiful yet mysterious world of bio-molecules.

I would also express my gratitude towards Dr. Santasabuj Das, Scientist E of National Institute of Cholera and Enteric Diseases, Kolkata for his fruitful discussion towards flagellin project. Thanks also to my thesis and PhD committee members, Dr. Rajib Kr. Mitra, Prof. Priya Mahadevan of S. N. Bose National Centre for Basic Sciences (SNBNCBS) and Prof. Kankan Bhattacharyya of IISER Bhopal for their helpful comments and suggestions. A special acknowledgement goes to my senior Dr. Amit Das for his guidance and suggestions during early days of my PhD carrier.

It is a pleasure to pay tribute also to my fellow group members who helped me during different phases of my works in their own unique ways. I am truly thankful to Dr. Manas Mondal, Dr. Laxmi Maganti, Dr. Mausumi Ray, Suman Dutta, Piya Patra, Camelia Manna, Arunava Adak, Abhik Ghoshmoulick, Rahul Karmakar, Sasthi

Charan Mandal and Edwin Tendong. Especially, I would like to express my heartiest gratitude to my seniors, Dr. Samapan Sikdar and Dr. Paramita Saha who assisted me a lot to finish this journey. Their support, knowledge and love always motivated me.

My special regards to all my teachers at primary, high school, undergraduate and postgraduate level who not only enriched my knowledge but also inspired me to grow as a human being. I extend my vote of thanks to Gouri didibhai, Shyamal sir, Purnendu sir, Subrata sir, AKB sir, Anathbandhu sir, Mousumi didibhai and Shyamal Adhikari sir. I gratefully acknowledge Saraswati Shishu Mandir, Barlow Girls' High School and Jadavpur University for providing me the perfect ambience and educational support.

I have been blessed with many good friends and this journey would not be possible without each of them. Special thanks to all of them particularly to Bijita, Suravi, Ashis, Bithika di, Sukarna, Sewli and Snehasish. I am especially grateful to Sankar for his moral support, encouragement and affection. I owe you a debt of gratitude for all that you have done for me. I extend my word of thanks to my seniors, Sreemoyee Di, Abhijit da, Gourab da, Suman da and Chiranjit da who helped me a lot during my days at SNBNCBS. I would like to thank staffs of SNBNCBS Mess for all their help and service.

I gratefully acknowledge S. N. Bose. National Centre for Basic Sciences for providing me wonderful research facilities. I am also thankful to all the faculty members and staffs of this centre for their sincere cooperation and help. Finally, special thanks to DST, Govt. of India for providing me INSPIRE Fellowship.

Last but not the least, I express my deep sense of gratitude to my parents Nilkanta Dutta and Priti Dutta and my sister Sandita Dutta for their love, dream,

encouragement and moral support over the years of my study. Without their sacrifices, moral supports and blessings, the thesis would not have taken its shape.

Once again, I acknowledge my deep gratitude to everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one. Many Thanks to all.

Sutapa Dutta

Department of Chemical, Biological and Macro-Molecular Sciences,

S. N. Bose National Centre for Basic Sciences,

Salt lake, Kolkata-700106, India.

Abstract

The thesis is based on understanding static and dynamic fluctuations of protein dihedrals and to correlate the observations towards proteins function. This is illustrated for four different aspects that can modulate protein functions: (a) long distant communication between residues; (b) electrostatically heterogeneous surface of proteins; (c) changes in conformational thermodynamics of protein in different medium and (d) balance of forces leading to protein aggregation.

Ligand bindings at long distant sites are often associated with protein functions. Here, we propose an approach to probe causal connection between distant binding residues of a small protein named ubiquitin (Ub). This is based on time dependent dihedral cross correlation function over μ s long all-atom Molecular Dynamics (MD) simulation and from mathematical modeling. We have found that dihedrals of these functionally important yet spatially distant residues possess significant amount of temporal correlations with biologically relevant time scales.

The different natures of hydrophilic residues lead to heterogeneous distributions of charges over protein surface. We address whether this heterogeneity leaves imprint to the motions of the metal ions, which play pivotal roles in protein functions. Using mean squared displacement and self-van Hove function of the metal ions from μ s long all-atom MD simulated trajectories; we find that the metal ions undergo anomalous diffusion due to trapping at different sites. However, if the strength of trapping is reduced, then the ions reclaim the normal Fickian diffusive profiles.

Conformational fluctuations are another factor that can tune proteins functionality. Here, we compute the changes in thermodynamics free energy and entropy of a bacterial protein, namely, flagellin in lipid bilayer with respect to

aqueous media. We use histogram based method on the equilibrium fluctuations of dihedral angles. The observations suggest that bilayer induced larger conformational stability to flagellin than water, and thus might play an important role in corresponding immune responses activated within the host cell upon bacterial adhesion.

Self-assembly of misfolded proteins is important in neurodegenerative disorders. We have studied protein aggregation using a model system having charge at the core but the solvophobic surface using Monte-Carlo (MC) simulation and mean field analytical treatment. The system exhibits aggregation primarily in presence of hydrophobicity mediated attraction. Electrostatic repulsion controls stability of finite size clusters or aggregates.

Table of Contents

Title Page	I
Dedication	II
Acknowledgement	III
Abstract	VI
CHAPTER 1 Introduction	1
1.1. Protein, a complex system	1
1.2. Proteins structure dynamics	3
1.3. Long distance correlations in proteins	4
1.4. Diffusion dynamics of metal ions in presence of protein	5
1.5. Conformational thermodynamics of protein in different medium	6
1.6. Model studies on protein aggregation	8
1.7. References	9
CHAPTER 2 Details of methods	13
2.1. MD simulation algorithm	13
2.2. Force-field for MD simulation of bio-molecules	14
2.3. Periodic boundary condition and minimum image convention	15
2.4. Particle Mesh Ewald (PME) methods	16
2.5. Simulation of protein	16
2.6. MC simulation algorithm	18
2.7. References	20

CHAPTER 3 Long distance temporal correlations in proteins	22
3.1. Introduction	22
3.2. Methods	27
3.2.1. MD simulation	
3.2.2. Computation of two-point correlation function (TDCF)	27
3.2.3. Computation of four-point correlation function	28
3.2.4. Computation of transfer entropy	28
3.2.5. Computation of conformational entropy	29
3.3. Simulation results	29
3.3.1. Correlated and anti-correlated dihedral fluctuations	30
3.3.2. Convergence in TDCF over MD simulated time scales	30
3.3.3. TDCFs among functional residues	32
3.3.4. TDCF map and dynamically correlated path in Ub	36
3.3.5. Mechanistic view of TDCF	38
3.3.6. TDCF for allosteric path in CaM	39
3.4. Mathematical model	41
3.5. Discussions	42
3.6. Conclusions	46
Appendix I	48
3.7. References	51
CHAPTER 4 Anomalous dynamics of metal ions in presence of a protein	54
4.1. Introduction	54
4.2. Methods	57
4.2.1. Details of MD simulation	57

4.2.2. Details of self-vHf calculation	59
4.2.3. MSD calculation	59
4.2.4. Density of the ions around the residues	60
4.3. Results	60
4.3.1. The self-vHf for Zn ²⁺ ions	61
4.3.2. MSD of the ions	66
4.3.3. Dynamic heterogeneity	66
4.3.4. Underlying mechanisms of heterogeneous diffusion	69
4.4. Discussions	71
4.5. Conclusions	74
4.6. References	74
CHAPTER 5 Microscopic studies on flagellin-receptor complexes	78
5.1. Introduction	78
5.2. Methods	84
5.2.1. System preparations	84
(I) fliC	84
(II) flaD	84
(III) TLR5	85
(IV) fliC-TLR5 complex	86
(V) flad-TLR5 complex	86
(VI) Preparation of bilayer	87
5.2.2. MD simulations	87
5.2.2.1. Simulation in water	87
5.2.2.2. Simulations in presence of membrane	88
5.2.3. Conformational thermodynamics and HBM	89

5.2.4. Docking of coreceptors with flagellin-TLR5 complexes	90
5.3. Results and discussions	92
5.3.1. Conformational stability of fliC	92
5.3.2. Conformational stability of flaD	96
5.3.3. Binding interface between flagellin and TLR5 receptor	99
5.3.4. Conformational stability of flagellin-TLR5 complexes	102
5.3.5. Coreceptor binding	109
5.4. Conclusions	113
5.5. References	113
CHAPTER 6 Model studies on aggregation of misfolded proteins	117
6.1. Introduction	117
6.2. Methods	121
6.2.1. Details of potential function	121
6.2.2. Details of MC simulation	122
6.2.3. Cluster-size analysis	122
6.2.4. Computation of radial distribution functions	123
6.2.5. Mean field analytical treatment	123
6.3. Results	123
6.3.1. Cluster-size analysis	124
6.3.2. Finite correlation in system	126
6.3.3. Analysis from mean field treatment	127
6.4. Discussions	129
6.5. Conclusions	131
Appendix I	132

6.6. References	137
CHAPTER 7 Conclusion & future direction	141
References	142
List of Publications	144

CHAPTER 1

Introduction

1.1. Protein, a complex system:

A complex system¹ consists of various interconnected parts which exhibit collective or emergent behaviors, different from that of the individuals. Studies of complex system are increasingly recognized in variety of interdisciplinary fields² ranging from financial market, school of fish, ecosystems, non-linear fluids, immune system, granular media to bio-macromolecules including protein, DNA^{1, 2}. Protein, is a complex bio-molecular system which performs variety of functions in living cell including enzyme catalysis, transportation of therapeutic molecules, responding to external stimuli³⁻⁵ to name a few.

The basic building block of protein, this sophisticated bio-molecule, is twenty essential amino acids³⁻⁵, often called residues, which have similar backbone chemical composition but different side chains. The polypeptide chain formed by peptide bond ($-CONH_2$) between different residues is known as primary sequence of a protein³⁻⁵. In general, proteins need to fold into three dimensional compact structures starting from the primary sequence of amino acids in order to be functionally active. The three dimensional arrangement of atoms in a folded protein is known as protein conformation, containing local secondary structures like alpha helices and beta sheets^{6, 7} (Fig.1.1.(a)). Difference in H-bond patterns between backbones, electrostatic interactions⁸ between polar and hydrophilic side-chains are the primary governing factors^{9, 10} of protein conformation. In folded conformations, proteins expose their

hydrophilic groups to aqueous environment⁸, while hydrophobic parts remain buried¹¹.

Proteins undergo series of conformational fluctuations due to intrinsic thermal motions^{12, 13}. The complete microscopic description of protein structures in different conformations is difficult to obtain from biophysical and biochemical studies mainly due to involvement of huge degrees of freedom. Moreover, having a wide range (femto-second (fs)-mili-second (ms)) of fluctuation time scales¹³⁻¹⁵ makes proteins dynamically complex to address. Local motions like backbone exhibits dynamics¹⁶ over micro-second (μ s) to nano-second (ns) regime, whereas side chain fluctuations¹⁷ typically lie within pico-second (ps) to ns time window. For global reorientation, domain motions of protein occur at much slower time scales¹⁸ (μ s-ms). However, these spontaneous fluctuations of protein in large ensemble of states around an average structure^{14, 19}, separated by an energy barrier are largely associated with protein functionality. Thus, microscopic understanding of protein functions in spatio-temporal paradigm presents major challenges due to variety of interactions as well as for fluctuations between different conformations. However, such studies are not only relevant to understand biological systems but also explore possibilities of novel materials of desired functionalities based on conformational fluctuations.

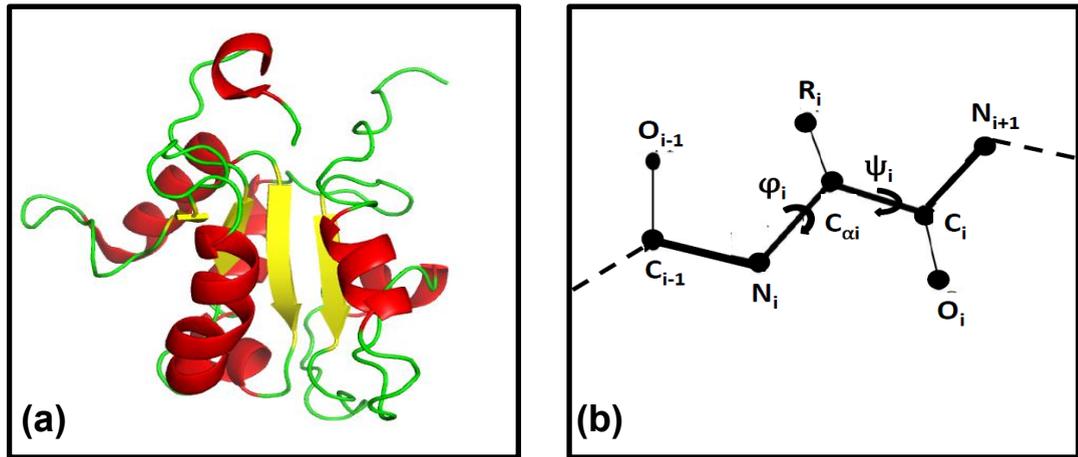


Fig.1.1.(a) Different secondary structures of protein. (b) Schematic representation of protein dihedrals.

1.2. Proteins structure dynamics:

Several experimental techniques including X-ray crystallography, NMR are capable of producing elaborate information on protein three dimensional conformations. However these experiments have their own limitations^{12, 13}. X-ray crystallography^{13, 20} is possible only if the protein can be crystallized. Moreover, the data give information on a frozen conformation, accompanied by B factor²¹ that gives deviations of atomic coordinates about their mean positions, generated due to temperature driven fluctuations. B factor²¹ also contains model error, invalid restrains along with lattice imperfections. In case of NMR^{13,22-24}, nuclear overhauser effect (NOE) provides bond vector constraints, chemical shift is associated with dihedral constraints, residual dipolar coupling (RDC) is related to orientation of bond vector and order parameter derived from spin relaxations can provide local motions from ps-ns time scales. However, NMR signals^{13,22-24} are not well resolved in particular for large proteins and proteins lacking in stable secondary structures.

Complementary to experiments Molecular Dynamics (MD) simulations²⁵⁻²⁷ can provide microscopic details about motions in proteins starting from ps to μ s time window. Classical all atom-MD²⁸ can mimic physical motions of protein atoms, comparable to actual experimental conditions by solving Newtonian equation of motion. The dihedrals are chosen as conformational degrees of freedom or microscopic variable²⁹ to illustrate protein conformation and dynamics. Proteins have three backbone dihedrals (φ, ψ, ω) and up to five side chain dihedrals ($\chi_i, i = 1..5$) depending on the amino acid types. Dihedral angle is the intersecting angle between two adjacent planes for four consecutive atoms. In case of polypeptide backbone for $C_{i-1} - N_i - C_{\alpha,i} - C_i$ (Fig.1.1.(b)), the angle between $C_{i-1} - N_i - C_{\alpha,i}$ and $N_i - C_{\alpha,i} - C_i$ planes is known as dihedral φ_i . Similarly angle between $N_i - C_{\alpha,i} - C_i$ and $C_{\alpha,i} - C_i - N_{i+1}$ is dihedral ψ_i where i is residue index. The side-chain dihedral, χ_1 defines the angle between $N_i - C_{\alpha,i} - C_{\beta,i}$ and $C_{\alpha,i} - C_{\beta,i} - C_{\gamma,i}$. The thermodynamic properties of a system of proteins^{30, 31} can also be computed through Monte Carlo (MC) simulation over a coarse-grained system interacting via model potentials. In this thesis, we consider examples of folded protein namely, ubiquitin, flagellin in order to probe static and dynamic fluctuations to relate their functionalities with underlying microscopic degrees of freedom. We also consider case of misfolded proteins which tend to aggregate through model calculations.

1.3. Long distance correlations in proteins:

Quite often proteins undergo series of ligand bindings at different sites. Such binding events are not only associated with numerous cellular processes, but also form the basis of technological applications³². The communication among distant binding sites in proteins is one of the fundamental questions which still remain largely

obscure. Recent simulation studies emphasize on existence of weak static correlation between dihedrals of such long distant residues³³⁻³⁹. However, the correlation between such residue pairs must persist till the binding partners reorient. Thus the correlation time scales should be comparable to the rotational correlation times (\sim tens of ns⁴⁰) of binding partners.

Here we propose a theoretical approach to probe causal connection between distant binding residues of a small protein named ubiquitin⁴¹ (Ub) using time dependent dihedral cross correlation functions (TDCFs) over μ s long all-atom MD simulation and mathematical modeling⁴². TDCFs or two point correlation functions²⁸ between two dihedrals of two different residues indicate how long the disturbance coming from one particular dihedral affects the other. The magnitude of TDCF on the other hand, implies the strength of correlation between two distant residues over time domain. We find that distant yet functionally important residues are connected by significant amount of dynamic correlations over biologically relevant time scales⁴². We apply this method to understand allosteric effect in protein, where ligand binding at one binding site affects that in another site called the effector site. We consider in particular allostery in Calmodulin⁴³, important for Ca^{2+} dependent signaling in cellular processes⁴⁴.

1.4. Diffusion dynamics of metal ions in presence of protein:

In a native structure, most proteins expose their hydrophilic residues to water. This leads electrostatically heterogeneous surface which can trap charged moieties to perform numerous functions in living cell⁴⁵. The metal ions act as electrophilic and can bind with acidic and polar residues of proteins. However, identifying ion trapping surface patch of a protein is not straightforward due to limitations of experiments⁴⁶⁻⁴⁹.

Here we propose that dynamics of metal ions can indicate the metal ion binding sites on protein surface.

We consider ubiquitin that is known to bind Zn^{2+} ions by some of its surface exposed acidic residues^{50, 51}. We examine the mean squared displacement (MSD) and the self-van Hove function⁵² (self-vHf) of the ions in presence of the protein and compare the behavior in its free state, where protein is not present in the water box. The averaging is done over five different μs long all-atom MD simulated trajectories. In order to probe the relation between trapping propensity and diffusive dynamics of the ions, we mutate some of the Zn^{2+} binding acidic residues of ubiquitin to hydrophobic ones and then address the diffusive profiles of Zn^{2+} .

In absence of the protein the ions show normal Fickian diffusion. We observe that the Zn^{2+} ions undergo heterogeneous diffusion with a linear tail in self-vHf due to strong trapping at different binding sites over ubiquitin. However, if we reduce the interaction strength between ions and protein, then the self-vHf shows double Gaussian dependence. Thus, we conclude that this diffusion analysis, based on scattering technique can be considered as an approach to identify metal ion binding residues over heterogeneous surface of a protein.

1.5. Conformational thermodynamics of protein in different medium:

In addition to dynamic features, we also use static features to understand protein functions. Proteins need to recognize numerous ligands or peptide in order to perform various functions in living cell. There are several mechanisms behind this specific molecular recognition process, like lock and key, induced fit, conformational selection to name a few. Recent experiments⁵³⁻⁵⁶ suggest that the changes in entropy of a protein in different conformations play important role in ligand binding.

Conformational entropy⁵⁷⁻⁵⁹ of a protein emphasizes on how the thermal fluctuations between different states of the protein around a mean structure affect its functionality. In recent studies^{92,93}, it has been shown that the difference in conformational thermodynamics between two states of a protein can be extracted from equilibrium fluctuations of the dihedral angles using Histogram based method (HBM). Equilibrium conformational changes in free energy is defined by $\Delta G(\varphi) = -K_B T \ln[(H_{bound}^{max}(\varphi)) / (H_{free}^{max}(\varphi))]$ where $H_{bound}^{max}(\varphi)$ and $H_{free}^{max}(\varphi)$ signify peak value of normalized probability distribution of protein dihedral φ in two different states; one is ligand bound state ($H_{bound}^{max}(\varphi)$) and the other is ligand free state ($H_{free}^{max}(\varphi)$), whereas K_B is the Boltzmann's Constant. Conformational entropy change associated with a particular dihedral φ is calculated using $\Delta S = S_{bound}(\varphi) - S_{free}(\varphi)$ where $S_{bound}(\varphi) = -K_B T \sum_i H_{bound,i}(\varphi) \ln H_{bound,i}(\varphi)$, $S_{free}(\varphi) = -K_B T \sum_i H_{free,i}(\varphi) \ln H_{free,i}(\varphi)$ and the sums are taken over histogram bins.

We consider a protein, named flagellin that is responsible for bacterial virulence⁶⁰⁻⁶³. Flagellum is an organelle of bacteria involved in several biological functions, like chemo taxis, protein export, biofilm formation as well as adhesion and invasion⁶⁰⁻⁶⁵ to the host cells and is composed of 20,000 subunits of a protein, known as flagellin. Toll like receptor (TLR5)^{66, 67} is the host cell surface receptor protein that can recognize external virulent flagellin and activate signaling within host cells. Although potential role of flagellin for bacterial adhesion and invasion⁶⁸ has been emphasized experimentally, the underlying microscopic mechanism remains elusive.

Here, we study relative conformational stability of flagellin from different organisms in lipid bilayer compared to aqueous media. We consider flagellin from both invasive (fliC) and non-invasive bacteria (flaD). We find that both the flagellins are thermodynamically and structurally more stable in bilayer than water. We also

observe that TLR5 binding interface between fliC-TLR5 and flaD-TLR5 are not similar. This could be the reason for different signaling pathway in these two complex systems.

1.6. Model studies on protein aggregation:

Mutation, physiological stress and changes in physio-chemical conditions, changes in amino acid sequences or changes in nature of intra and inter molecular attraction can lead proteins to deviate from native structures, exposing the hydrophobic parts to water^{69, 70}. This leads to self-assembly of peptide amphiphiles which are relevant in not only several neurodegenerative disorders^{69, 71} but also in nano-bio technology^{72, 73}. The complete understandings of driving forces that lead aggregations are difficult from experiments and all-atom MD simulations⁷⁴⁻⁷⁶ due to involvement of large time and length scales.

We study protein aggregation using a coarse-grained model system having charge at the core but hydrophobic surface, using MC simulation and mean field analytical treatment. We observe that the system exhibits aggregation primarily in presence of strong hydrophobic interactions. However, for comparatively weaker hydrophobic strength, particles form finite size clusters. Electrostatic interaction on the other hand, controls stability of these finite size clusters.

The thesis contains following chapters: Chapter 2 explains the details method of MD and MC simulations. Long distance correlations in protein are explained from time dependent dihedral cross correlation function analysis and mathematical modeling in Chapter 3. Chapter 4 includes details of diffusion dynamics of divalent metal ions in presence of electrostatically heterogeneous surface of protein. Conformational thermodynamics for flagellin in water as well as in bilayer is reported

in Chapter 5. MC simulation along with mean field analysis for protein aggregation is discussed in Chapter 6. The future outlooks based on our works are mentioned in Chapter 7.

1.7. References:

1. M. E. J. Newman, *Am. J. Phys.*, 2011, **79**, 800-810.
2. J. Ross and A. P. Arkin, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 6433-6434.
3. B. JM, T. JL and S. L, *Biochemistry. 5th edition*, W H Freeman, New York, 2002.
4. C. G. M, *The Cell: A Molecular Approach, 2nd edition*, Sinauer Associates, Sunderland (MA), 2000.
5. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell, 4th edition*, Garland Science, New York, 2007.
6. N. Bhattacharjee and P. Biswas, *Protein Peptide Lett.*, 2009, **16**, 953-960.
7. G. D. Rose, P. J. Fleming, J. R. Banavar and A. Maritan, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 16623-16633.
8. I. Gitlin, J. D. Carbeck and G. M. Whitesides, *Angew. Chem. Int. Ed.*, 2006, **45**, 3022-3060.
9. K. A. Dill and J. L. MacCallum, *Science*, 2012, **338**, 1042-1046.
10. K. A. Dill, S. B. Ozkan, M. S. Shell and T. R. Weikl, *Annu Rev Biophys*, 2008, **37**, 289-316.
11. C. Camilloni, D. Bonetti, A. Morrone, R. Giri, C. M. Dobson, M. Brunori, S. Gianni and M. Vendruscolo, *Sci. Rep.*, 2016, **6**, 28285.
12. X. Salvatella, *Protein Conformational Dynamics*, 2014, **805**, 67-85.
13. A. Ramanathan, A. Savol, V. Burger, C. S. Chennubhotla and P. K. Agarwal, *Acc. Chem. Res.*, 2014, **47**, 149-156.
14. K. Henzler-Wildman and D. Kern, *Nature*, 2007, **450**, 964-972.
15. B. Gipson, D. Hsu, L. E. Kaviraki and J. C. Latombe, *Annu. Rev. Anal. Chem.*, 2012, **5**, 273-291.
16. J. M. Lamley, M. J. Lougher, H. J. Sass, M. Rogowski, S. Grzesiek and J. R. Lewandowski, *Phys. Chem. Chem. Phys.*, 2015, **17**, 21997-22008.
17. K. H. Dubay, J. P. Bothma and P. L. Geissler, *PLoS Comput Biol.*, 2011, **7**.

18. M. Gur, E. Zomot and I. Bahar, *J. Chem. Phys.*, 2013, **139**.
19. J. O. Wrabl, J. Gu, T. Liu, T. P. Schrank, S. T. Whitten and V. J. Hilser, *Biophys. Chem.*, 2011, **159**, 129-141.
20. M. Jamroz, A. Kolinski and D. Kihara, *Proteins*, 2012, **80**, 1425-1435.
21. R. B. Fenwick, H. van den Bedem, J. S. Fraser and P. E. Wright, *Proc. Natl. Acad. Sci. U.S.A.*, 2014, **111**, E445-E454.
22. L. E. Kay, *J. Mol. Biol.*, 2016, **428**, 323-331.
23. A. F. Angyan and Z. Gaspari, *Molecules*, 2013, **18**, 10548-10567.
24. L. Salmon, G. Bouvignies, P. Markwick and M. Blackledge, *Biochemistry*, 2011, **50**, 2735-2747.
25. A. B. Guliaev, S. Cheng and B. Hang, *World J Methodol*, 2012, **2**, 42-49.
26. G. G. Dodson, D. P. Lane and C. S. Verma, *EMBO Rep.*, 2008, **9**, 144-150.
27. K. Lindorff-Larsen, P. Maragakis, S. Piana and D. E. Shaw, *J. Phys. Chem. B*, 2016, **120**, 8313-8320.
28. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids, 2nd edition*, Oxford Science Publications, New York, 1989.
29. A. Altis, P. H. Nguyen, R. Hegger and G. Stock, *J. Chem. Phys.*, 2007, **126**, 244111.
30. E. Rossinsky and S. Srebnik, *Biopolymers*, 2005, **79**, 259-268.
31. A. Sikorski and J. Skolnick, *Proc. Natl. Acad. Sci. U.S.A.*, 1989, **86**, 2668-2672.
32. A. Vallee-Belisle and K. W. Plaxco, *Curr. Opin. Struct. Biol.*, 2010, **20**, 518-526.
33. R. B. Fenwick, S. Esteban-Martin, B. Richter, D. Lee, K. F. Walter, D. Milovanovic, S. Becker, N. A. Lakomek, C. Griesinger and X. Salvatella, *J. Am. Chem. Soc.*, 2011, **133**, 10336-10339.
34. D. Long and R. Bruschweiler, *J. Phys. Chem. Lett.*, 2012, **3**, 1722-1726.
35. R. B. Fenwick, L. Orellana, S. Esteban-Martin, M. Orozco and X. Salvatella, *Nat. Commun.*, 2014, **5**, 1-9.
36. K. Kasahara, I. Fukuda and H. Nakamura, *Plos One*, 2014, **9**, e112419.
37. D. W. Li, D. Meng and R. Bruschweiler, *J. Am. Chem. Soc.*, 2009, **131**, 14610-14611.
38. A. Black Pyrkosz, J. Eargle, A. Sethi and Z. Luthey-Schulten, *J. Mol. Biol.*, 2010, **397**, 1350-1371.

39. H. Kamberaj and A. van der Vaart, *Biophys. J.*, 2009, **97**, 1747-1755.
40. Y. Ryabov, *Proteins*, 2015, **83**, 1571-1581.
41. S. Vijaykumar, C. E. Bugg and W. J. Cook, *J. Mol. Biol.*, 1987, **194**, 531-544.
42. S. Dutta, M. Ghosh and J. Chakrabarti, *Sci. Rep.*, 2017, **7**, 40439.
43. Z. G. Xia and D. R. Storm, *Nat. Rev. Neurosci.*, 2005, **6**, 267-276.
44. L. Maganti, S. Dutta, M. Ghosh and J. Chakrabarti, *J. Biomol. Struct. Dyn.*, 2018, 1-10.
45. K. H. Sippel and F. A. Quioco, *Protein Sci.*, 2015, **24**, 1040-1046.
46. H. Zheng, K. B. Handing, M. D. Zimmerman, I. G. Shabalin, S. C. Almo and W. Minor, *Expert Opin Drug Discov*, 2015, **10**, 975-989.
47. Y. Li and C. Kang, *Molecules*, 2017, **22**.
48. W. Shi, M. Punta, J. Bohon, J. M. Sauder, R. D'Mello, M. Sullivan, J. Toomey, D. Abel, M. Lippi, A. Passerini, P. Frasconi, S. K. Burley, B. Rost and M. R. Chance, *Genome Res.*, 2011, **21**, 898-907.
49. A. Passerini, M. Lippi and P. Frasconi, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 2012, **9**, 203-213.
50. A. Camara-Artigas, M. Plaza-Garrido, S. Martinez-Rodriguez and J. Bacarizo, *Acta Crystallogr F*, 2016, **72**, 29-35.
51. F. Arnesano, B. D. Belviso, R. Caliandro, G. Falini, S. Fermani, G. Natile and D. Siliqi, *Chemistry*, 2011, **17**, 1569-1578.
52. J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids, Fourth Edition*, Elsevier, 2013.
53. M. S. Marlow, J. Dogan, K. K. Frederick, K. G. Valentine and A. J. Wand, *Nat. Chem. Biol.*, 2010, **6**, 352-358.
54. S. Sikdar, J. Chakrabarti and M. Ghosh, *Mol. BioSyst.*, 2014, **10**, 3280-3289.
55. A. Das, J. Chakrabarti and M. Ghosh, *Biophys. J.*, 2013, **104**, 1274-1284.
56. R. D. Brox, M. M. Lopez, H. J. Vogel and G. I. Makhatadze, *J. Biol. Chem.*, 2001, **276**, 14083-14091.
57. M. S. Marlow, J. Dogan, K. K. Frederick, K. G. Valentine and A. J. Wand, *Nat. Chem. Biol.*, 2010, **6**, 352-358.
58. A. J. Wand, V. R. Moorman and K. W. Harpole, *Dynamics in Enzyme Catalysis*, 2013, **337**, 69-94.

59. J. A. Caro, K. W. Harpole, V. Kasinath, J. Lim, J. Granja, K. G. Valentine, K. A. Sharp and A. J. Wand, *Proc. Natl. Acad. Sci. U.S.A.*, 2017, **114**, 6563-6568.
60. J. Haiko and B. Westerlund-Wikstrom, *Biology*, 2013, **2**, 1242-1267.
61. R. S. Friedlander, N. Vogel and J. Aizenberg, *Langmuir*, 2015, **31**, 6137-6144.
62. Y. Rossez, E. B. Wolfson, A. Holmes, D. L. Gally and N. J. Holden, *PLOS Pathog.*, 2015, **11**.
63. Q. Duan, M. Zhou, L. Zhu and G. Zhu, *J. Basic Microbiol.*, 2013, **53**, 1-8.
64. S. A. Beatson, T. Minamino and M. J. Pallen, *Trends Microbiol.*, 2006, **14**, 151-155.
65. A. S. Utada, R. R. Bennett, J. C. N. Fong, M. L. Gibiansky, F. H. Yildiz, R. Golestanian and G. C. L. Wong, *Nat. Commun.*, 2014, **5**.
66. J. I. Godfroy, 3rd, M. Roostan, Y. S. Moroz, I. V. Korendovych and H. Yin, *Plos One*, 2012, **7**, e48875.
67. B. M. Buchholz and A. J. Bauer, *Neurogastroent Motil*, 2010, **22**, 232-245.
68. M. Zhou, Q. Duan, Y. Li, Y. Yang, P. R. Hardwidge and G. Zhu, *Archives of microbiology*, 2015, **197**, 797-803.
69. V. Vetri and V. Fodera, *Febs Lett*, 2015, **589**, 2448-2463.
70. J. Beerten, J. Schymkowitz and F. Rousseau, *Curr. Top. Med. Chem.*, 2012, **12**, 2470-2478.
71. T. K. Karamanos, A. P. Kalverda, G. S. Thompson and S. E. Radford, *Progress in nuclear magnetic resonance spectroscopy*, 2015, **88-89**, 86-104.
72. A. Dehsorkhi, V. Castelletto and I. W. Hamley, *J. Pept. Sci.*, 2014, **20**, 453-467.
73. Y. Chen, H. X. Gan and Y. W. Tong, *Macromolecules*, 2015, **48**, 2647-2653.
74. A. Morriss-Andrews and J. E. Shea, *J. Phys. Chem. Lett.*, 2014, **5**, 1899-1908.
75. M. A. Blanco, E. Sahin, A. S. Robinson and C. J. Roberts, *J. Phys. Chem. B*, 2013, **117**, 16013-16028.
76. M. Carballo-Pacheco and B. Strodel, *J. Phys. Chem. B*, 2016, **120**, 2991-2999.

CHAPTER 2

Details of methods

We use Molecular Dynamics (MD) simulations^{1, 2} as a primary tool to understand conformational rearrangement and the interactions of protein with other molecules. The thermodynamic properties of misfolded proteins^{3, 4} have been studied via a model system using the Monte Carlo (MC) simulation trajectories.

2.1. MD simulation algorithm:

MD simulation^{5, 6} is a powerful tool to calculate particle trajectory numerically. A three-dimensional system containing N number of particles with position $\bar{r}_i = (\bar{r}_{i1}, \bar{r}_{i2}, \dots, \bar{r}_{iN})$ and momentum $\bar{p}_i = (\bar{p}_{i1}, \bar{p}_{i2}, \dots, \bar{p}_{iN})$, is considered to interact via a pair potential $V(r_{ij})$, where $r_{ij} = \bar{r}_i - \bar{r}_j$, $i \neq j$, and i, j are the particle indices. The force \bar{F}_i acting on the i th particle by all other j th particles can be obtained from the gradient of $V(r_{ij})$, so the mathematical expression is written as $\bar{F}_i = \sum_{j=1}^N -\nabla V(r_{ij})$. Thus, the time evolution of this classically interacting system can be achieved by integrating Newton's equation of motion, $\bar{F}_i = m_i \bar{a}_i$, where m_i and \bar{a}_i are mass and acceleration of the i th particle respectively. The most widely used integrator for MD simulation is Verlet⁵, based on central difference algorithm. Using Taylor Series expansion, the updated position \mathbf{r} of a particle at a time $t + \Delta t$ is obtained by computing acceleration (\mathbf{a}) of the particle according the equation.2.1. as,

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 \quad (2.1).$$

Velocity is calculated using equation (2.2),

$$\mathbf{v}(t) = [\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)]/2\Delta t \quad (2.2).$$

Thus, the intrinsic error in position calculation is $\sim \Delta t^4$ and that for velocity is $\sim \Delta t^2$.

The MD trajectories can be used for both dynamics and static quantities of a system.

2.2. Force-field for MD simulation of bio-molecules:

NAMD⁷ and GROMACS⁸ are some standard packages that are used for parallel computation in order to simulate bio-molecules. In this thesis, simulation is performed only for the protein. One of the important step in simulating the protein is to choose appropriate force field (CHARMM⁹, AMBER¹⁰, and GROMOS¹¹), which is essentially parameterization of potential energy surface^{12, 13} of proteins based on semi-empirical quantum mechanical calculations or experimental data. All common force fields^{9, 10} generally include bonded interactions like bond stretching, bond rotations, torsional dihedrals via simple harmonic oscillations. Non-bonded interactions between transient dipoles are taken into account^{36, 37} using Van der Waals interaction; whereas contributions from permanent dipoles are manifested though electrostatic Coulomb interaction.

The expression for potential energy is:

$$\begin{aligned} V = & \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi [1 + \cos(n\phi - \\ & \delta)] + \sum_{impropers} k_w (w - w_0)^2 + \sum_{Urey-Bradley} k_u (u - u_0)^2 + \\ & \sum_{nonbonded} 4\epsilon \left[\left(\frac{R_{minij}}{r_{ij}} \right)^{12} - \left(\frac{R_{minij}}{r_{ij}} \right)^6 \right] + \sum_{ij} \frac{q_i q_j}{4\pi\epsilon_r \epsilon_0 r_{ij}} \end{aligned} \quad (2.3).$$

In equation (2.3), the first term describes energy cost associated with bond stretching.

k_b is the bond stretching constant, b_0 , the equilibrium bond length, whereas the

deviation at bond length is $(b - b_0)$. Similarly, the second term is the cost of the changes in angle from the equilibrium value θ_0 with force constant k_θ . k_θ is the dihedral force constant, n , the number of multiplicity of the torsion angle and δ is the phase shift. The fourth term stands for improper or outer plane bending. Here, k_w is the corresponding force constant and $(w - w_0)$ implies deviation of outer plane angle from its equilibrium value (w_0) . For the cross-term associated with the angle bending, 1,3 non bonded interaction is considered. The change in distance between the 1, 3 atoms $(u - u_0)$ is given in terms of a force constant k_u . In case of non-bonded interactions, R_{minij} is the distance for which Van der Waals potential is minimum, ϵ is the width of potential well. For the Columbic interactions, q_i and q_j are charges of the particles i, j respectively; ϵ_0 is permittivity of vacuum and ϵ_r , the dielectric constant of the medium. However, for the classical calculations the atoms have been assigned fixed partial charges^{36, 37} which lead deviations from proper polarization effect.

2.3. Periodic boundary condition and minimum image convention:

In computer simulation, the system is kept in a central box and periodic boundary condition (PBC)⁵ is implemented in order to mimic the real infinite bulk system. The box of length L can induce surface effect, such that the particles close to the surface experience different forces than the bulk. To minimize the surface effects, it is assumed that the central box is surrounded by infinite replica, so that if any atom leaves the central box in course of simulation, its image enters the box from opposite face. The PBCs are accompanied with minimum image convention, where the interaction for one particular atom is considered only with the nearest image of another atom among all the boxes. In order to avoid interaction between an atom in

central box with its mirror image in other replica boxes, a cut off ($r_c \sim L/2$) is used for truncating long-ranged interaction. This also makes simulation less expensive.

2.4. Particle Mesh Ewald (PME) methods:

The long ranged contributions of electrostatic interactions are treated using Particle Mesh Ewald (PME) method¹⁴, where the interaction potential is divided into two parts; the summation in short range part is computed in real space and the long ranged one is estimated in Fourier Space. The range of interaction originated by each point charge in the system is truncated by placing Gaussian charge distribution of opposite sign around that given charge. This screens the effective contribution of the point charge over all neighboring charges. Then the total screened short ranged potential is computed from sum over all atoms in the central box as well as over their images in replica boxes. Next a cancelling charge distribution with same shape and same sign of original point charge is added, so that the overall potential reduces to that of generated due to original charge.

2.5. Simulation of protein:

First, the initial structure of the protein is considered from protein data bank (PDB)¹⁵ depository or from homology modeling, which is a technique used to predict structure of an unknown protein using a known experimentally determined structure of a homologous protein¹⁶. A topology file containing values of the force constant and a position restraint file are required in order to keep protein structure intact throughout the simulation. Most of the biochemical processes in cellular conditions occur in aqueous environment, thus the protein molecule is solvated within a box either by explicit water model^{9, 10}, containing minimized structure of water lattice or by implicit

model where solvent effect is modeled as continuum dielectric. The total charge of the system is nullified^{9, 10} by adding proper number of ions to avoid polarization of ensemble. A snapshot of protein with ions, solvated in water box is shown in Fig.2.1.(a).

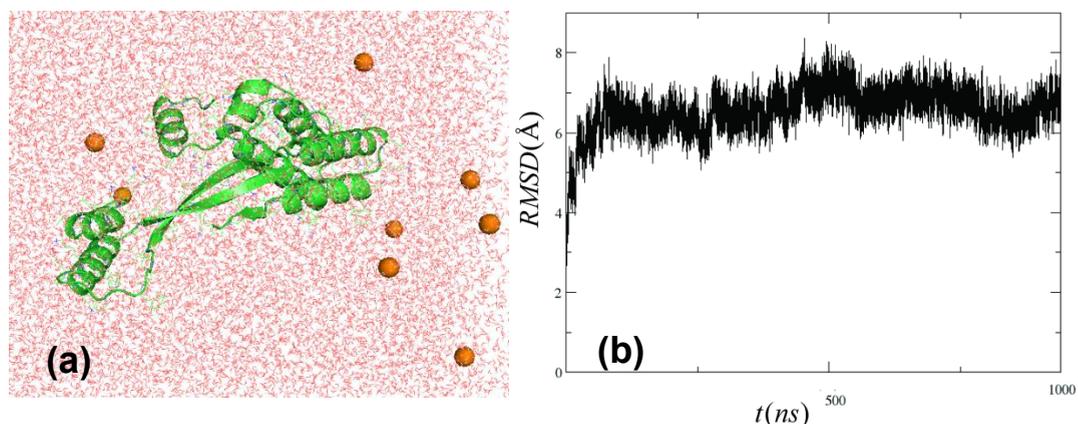


Fig.2.1.(a) Snapshot of protein solvated in water box along with ions. (b) Sample representation of RMSD of a protein over MD trajectories.

After that we need to perform energy minimization^{9, 10}, which removes bad contacts or inappropriate geometry of the model structure and leads to physically more reliable starting conformation, corresponding to any local minimum of the protein energy surface. In order to solve the equations numerically, entire simulation trajectories have been discretized⁵ into n number of steps and an integrator is used to proceed over associated time step (Δt) for each n steps. Very small value of Δt can lead numerical accuracy but at a cost of expansive simulation, whereas large value of Δt can generate unstable simulation⁵ due to large fluctuations of energy. Generally $\Delta t \sim$ fs, comparable to the mean collision time of the particles.

Next step is equilibration⁵, where temperature of the system is brought around the desired temperature by coupling all atoms to a thermal bath. This initial equilibration is important in order to achieve well sampled trajectories, so that the ergodicity can be applied. At thermal equilibrium, temperature of the system is

directly connected to its kinetic energy through degrees of freedom. So, at the first phase of equilibration, in isothermal-isochoric (NVT) ensembles, the velocities are scaled⁵ by a factor $\sqrt{T_B/T(t)}$, where T_B is the desired temperature and $T(t)$ is the instantaneous temperature at time t . During the equilibration, the deviation in temperature decays exponentially with a time constant τ . In second phase of equilibration, pressure is stabilized in isothermal-isobaric (NPT) ensemble, which closely resembles the experimental conditions. For the pressure coupling, an algorithm is used to rescale coordinates and box vectors of the system in every step. This has equivalent effect of a first-order kinetic relaxation of the generated pressure towards a reference pressure. The final step is production run, equilibration is ensured by plotting root mean square deviation (RMSD)^{9, 10} of the protein with respect to starting configuration over simulated trajectories. A sample diagram for RMSD is represented in Fig.2.2.(b). Thermodynamic variables, structural and dynamic properties of the protein can be calculated^{9, 10} using MD trajectories.

2.6. MC simulation algorithm:

MC simulation^{5, 6} is the common method to get outcome of stochastic processes using random numbers and probability statistics. It is actually the equilibrium sampling done over statistical ensemble using random walk algorithm. The partition function^{5, 6} Q for a system of N number of particles is defined as $Q = c \int d\mathbf{p}^N d\mathbf{r}^N \exp\left[-\frac{H(\mathbf{r}^N, \mathbf{p}^N)}{K_B T}\right]$, where \mathbf{r}^N is the coordinates of all particles, \mathbf{p}^N associated momentum, Hamiltonian $H(\mathbf{r}^N, \mathbf{p}^N)$, the total energy ($k + u$) of the system, K_B Boltzmann's constant, T the temperature and $c = 1/h^{3N}n!$ is proportionality constant. The expectation value of any dynamical variable A is defined as $\langle A \rangle =$

$$\int d\mathbf{p}^N d\mathbf{r}^N A(\mathbf{r}^N, \mathbf{p}^N) \exp[-H(\mathbf{r}^N, \mathbf{p}^N)/K_B T] / \int d\mathbf{p}^N d\mathbf{r}^N \exp[-H(\mathbf{r}^N, \mathbf{p}^N)/K_B T].$$

As kinetic energy, k is solely the function of momentum, the above integration can be solved analytically for the momentum term.

Instead of directly computing the integral $\int d\mathbf{r}^N A(\mathbf{r}^N) \exp[-u(\mathbf{r}^N)/K_B T]$, it is computed as; $\int d\mathbf{r}^N A(\mathbf{r}^N) \exp[-u(\mathbf{r}^N)/K_B T] / \int d\mathbf{r}^N \exp[-u(\mathbf{r}^N)/K_B T]$. The term $\frac{\exp[-u(\mathbf{r}^N)/K_B T]}{\int d\mathbf{r}^N \exp[-u(\mathbf{r}^N)/K_B T]}$ gives the probability density of finding the system in a configuration space around \mathbf{r}^N . So, instead of looking into the absolute probability, the relative probability of visiting different points in configuration space is taken into account.

This sampling of probable configuration space is done using Metropolis algorithm; firstly a random particle at a random position is considered and the corresponding potential energy $u(\mathbf{r}^N)$ is computed. Next, a random displacement Δ is given to the particle such that $r' = r + \Delta$, the potential energy at new configuration is represented as $u(\mathbf{r}'^N)$. If the changes in potential energy in new configuration with respect to the old one ($u(\mathbf{r}'^N) - u(\mathbf{r}^N) < 0$) is negative, then the new configuration is chosen as the updated one. The probability of finding the particle in new position, $p_n \propto \exp(-u(\mathbf{r}'^N)/K_B T)$ and that for the initial one is $p_i \propto \exp(-u(\mathbf{r}^N)/K_B T)$, if the ratio $(p_n/p_i) >$ a random number generated between 0 and 1, then also the new position is accepted as the updated coordinate of the particle. The process is repeated for all the particles present in the system.

Periodic boundary condition along with minimum image convention and a cut off (\sim half of the box length) for interaction are implemented. In each MC cycle, the total number of trial moves for displacing all of the N number of particles is N_{trial}

and the numbers for which new position is accepted is written as N_{accept} . If the amount of random displacement given to the particle is too small, then almost all moves will be accepted but the phase space will not be sampled properly. Similarly, if it is too large then almost every moves will be rejected, leading to poor sampling. So, during the course of simulation the value of maximum displacement is adjusted such that the acceptance ratio (N_{accept}/N_{trial}) reaches an optimal value ~ 0.5 ^{5, 6}, which ensures the proper sampling of equilibrium phase space. The structural properties and thermodynamics of the system are computed over the simulated trajectories.

2.7. References:

1. A. Hospital, J. R. Goni, M. Orozco and J. L. Gelpi, *Adv Appl Bioinform Chem*, 2015, **8**, 37-47.
2. J. D. Durrant and J. A. McCammon, *BMC Biol*, 2011, **9**, 71.
3. E. Rossinsky and S. Srebnik, *Biopolymers*, 2005, **79**, 259-268.
4. A. Sikorski and J. Skolnick, *Proc. Natl. Acad. Sci. U.S.A.*, 1989, **86**, 2668-2672.
5. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids, 2nd edition*, Oxford Science Publications, New York, 1989.
6. D. Frenkel and B. Smit, *Understanding Molecular Simulation From Algorithms to Applications, 2nd edition*, Academic Press, 2002.
7. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781-1802.
8. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701-1718.
9. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586-3616.

10. J. W. Ponder and D. A. Case, *Adv. Protein Chem.*, 2003, **66**, 27-85.
11. C. Oostenbrink, A. Villa, A. E. Mark and W. F. van Gunsteren, *J. Comput. Chem.*, 2004, **25**, 1656-1676.
12. F. R. Salsbury, *Curr Opin Pharmacol*, 2010, **10**, 738-744.
13. A. Ahmadi, C. McBride, J. J. Freire, A. Kajetanowicz, J. Czaban and K. Grela, *J. Phys. Chem. A*, 2011, **115**, 12017-12024.
14. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
15. S. Parasuraman, *J Pharmacol Pharmacother*, 2012, **3**, 351-352.
16. V. K. Vyas, R. D. Ukawala, M. Ghate and C. Chintha, *Indian J Pharm Sci.*, 2012, **74**, 1-17.

CHAPTER 3

Long distance temporal correlations in proteins

3.1. Introduction:

Quite often bio-macromolecules undergo cascade of ligand bindings at different sites. Such binding events not only control cellular processes, but also lie at the heart of technological applications with bio-molecules as scaffold¹. The microscopic basis of communication among the binding sites in bio-macromolecules is one of the fundamental questions which have drawn considerable attention, but still remains largely obscure. Motivated by this, in this chapter we make an attempt to understand the communication among functional residues in proteins, incorporating information of microscopic motions.

Although experimental probes are limited^{2, 3}, recent simulation works in this direction emphasize on the existence of covariance (Pearson Correlation Coefficient) between the instantaneous values of dihedral angles of such long distant residue pairs in a protein⁴⁻¹⁰. Non-zero but very small values of Pearson Correlation Coefficients have been observed among the dihedral angles of functional but spatially distant residues in a protein named ubiquitin, (Ub)⁴. However, information provided by Pearson Correlations is far from complete. Macromolecular binding takes place typically by rotational diffusion ranging in timescales of tens of ns¹¹, so that the binding surfaces are mutually exposed. This means that the changes at sites upon first ligand binding must affect the downstream binding sites till this time. Such temporal information are absent in Pearson Correlation Coefficient. The information entropy transfer^{10, 12} approaches has been proposed to causally connect residues depending

upon the history of their correlated fluctuations which includes all non-linear coupling between the fluctuating variables. Not only that the computation of information entropy transfer is quite involved, but also the physical mechanisms leading to the non-linear couplings is not understood, primarily due to lack of experimental probes. Moreover, time is introduced in the formalism on an ad-hoc basis¹².

Alternatively, temporal cross-correlations of fluctuations of two physical quantities $A(t)$ and $B(t')$ at times t and t' with respect to their mean values, also known as two-point correlations functions^{13, 14}, are used to describe time scales of correlated stochastic processes¹⁵. The equal time correlation function, $t = t'$, is the statistical Pearson Correlation. Temporal cross-correlation functions can be thought of generalization of Pearson correlation in time domain. Two-point correlation functions do not contain information on non-linear coupling. One major advantage of two-point correlation functions is that they are experimentally accessible by scattering techniques. Interestingly, temporal cross-correlations between fluorescence intensities show asymmetry with inversion in time which has been used to study co-localization of proteins¹⁶. This observation seems to suggest that temporally causal connections can be extracted from the two-point correlation functions. The power of two-point cross-correlation functions has not been exploited for in-depth understanding of bio-molecular phenomena. Time dependent dihedral cross-correlation functions (TDCF) have been employed for correlation between protein residues only up to a few hundred ps¹⁷, far too low compared to the bio-molecular binding time scales to have that functional relevance. We have established in an earlier work¹⁸ from much longer simulations that TDCF can relate large scale changes in a protein upon ligand binding.

With this backdrop we examine in this chapter TDCF by long computer simulations and mathematical modeling to understand functional co-ordination among

residues. We consider the case of a small protein, ubiquitin¹⁹ (PDB id: 1UBQ, Fig.3.1.(a)) involved in ubiquitination^{20, 21}, a process ubiquitous among the eukaryotes by which ubiquitin attaches with a target protein to degrade the latter. The process is initiated by covalent attachment of Adenosine mono-phosphate (AMP) to C-terminal Glycine; G76 of Ub. Following this, ubiquitin activation enzyme-E1 binds at different residues of Ub²⁰⁻²³, (Figs.3.1.(b)-(d)). The question is: How do the spatially distant (~1.5 nm) residues get temporally correlated so that the binding information at one site at a given time affects the binding at other sites at a later time?

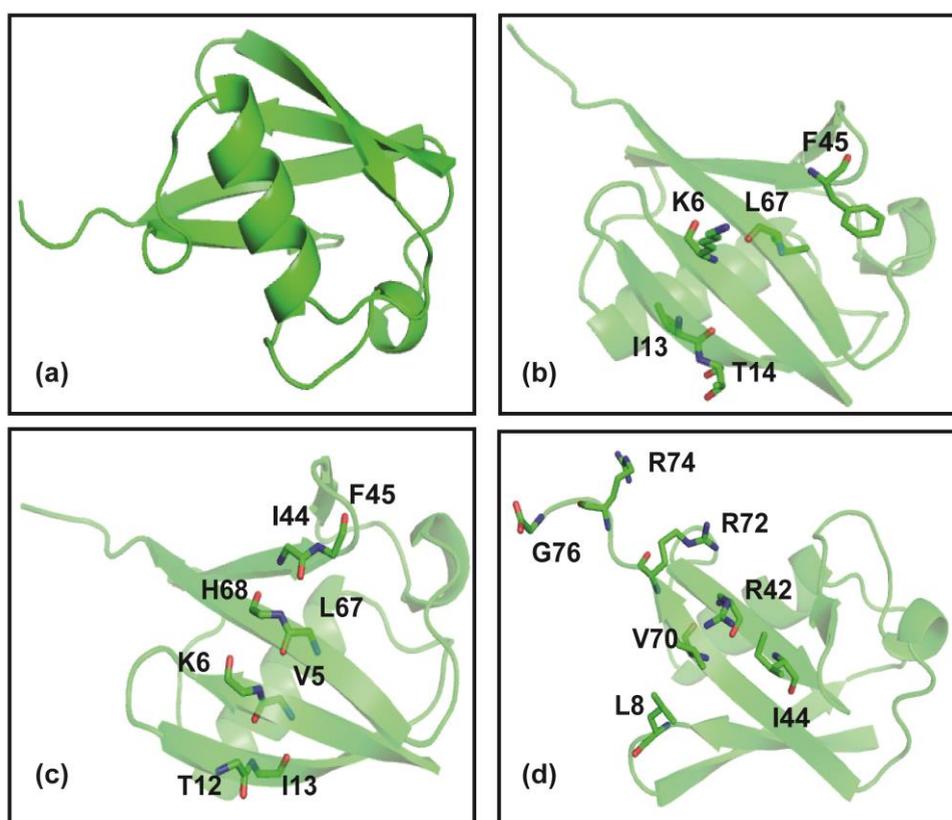


Fig.3.1.(a) Cartoon representation of the crystal structure of ubiquitin (PDB id: 1UBQ). (b) Correlated residues belonging to binding surface patch of ubiquitin. (c) H-bonded residues of that surface patch. (d) Residues participating in binding with E1 enzyme, involved in ubiquitination.

We show that the TDCF can explain the causal connection between the functional residues of Ub in the time scale of tens of ns. We explain the qualitative

features of TDCF using simple mathematical model. Interesting aspect of our result is that TDCFs contain temporal information which may be useful to understand biological processes which are orders of magnitude slower than atomic motions without invoking the non-linear effects. The simulated trajectories give us unique opportunity to examine typical non-linear terms which we find to be less dominant than the TDCFs themselves. Moreover, we propose a dynamically correlated path of residues along which the perturbation starting from G76 upon AMP initiation can manifest in ubiquitin.

We also correlate dynamic aspects of TDCF with equilibrium thermodynamic responses of ubiquitin using statistical fluctuations of dihedrals. Thermodynamics stability between different conformations of a protein²⁴⁻²⁷ plays critical role in governing protein functions. In case of Ub, ubiquitination is initiated at terminal G76 residue. In order to probe thermodynamic changes before and after AMP attachment to Ub, we simulate another system where G76 remains fixed through the entire simulation time. By freezing G76, we try to capture the perturbation given by AMP to G76 in real bio-chemical process. We find that the residues that are disordered or more flexible in perturbed state with respect to their free state, have weak correlations in time domain along with faster decaying time scales.

The connection between TDCF and dihedrals equilibrium fluctuations have been addressed for another protein, known as Calmodulin (CaM)²⁸. This protein is important²⁸ for Ca²⁺ dependent signaling, involved in numerous cellular processes. The protein has²⁹ two EF-hand motifs at each of the terminals, which are connected by a flexible linker helix (Fig.3.2.(a)). In order to maintain cellular regulation, CaM interacts with Orai1³⁰, a plasma membrane protein that influx Ca²⁺ from extra cellular environment into cytosol. It is found³¹ that a path of residues (Isoleucine; I100-

Aspartic Acid; D93-Phenylalanine; F89-Glutamic acid; E83-Threonine; T79-Lysine; K75-Methionine; M71-E67-Valine; V55-M51-Glutamine; Q49-I44, in Fig.3.2.(b)) starting from the C-terminal domain of CaM and those can bind with Orai1, provide long distance connection to some residues belong to the N terminal domain of CaM. The residues belong to the path are destabilized and disordered in Orai1 bound state with respect to Orai1 free state and show prominent changes primarily in side chain dihedral distributions in two states³¹. We find that these residues are dynamically correlated with a significant amount of temporal correlations in Orai1 bound state³¹. Thus, the dynamic information obtained from TDCF calculation fit well with statistical observations.

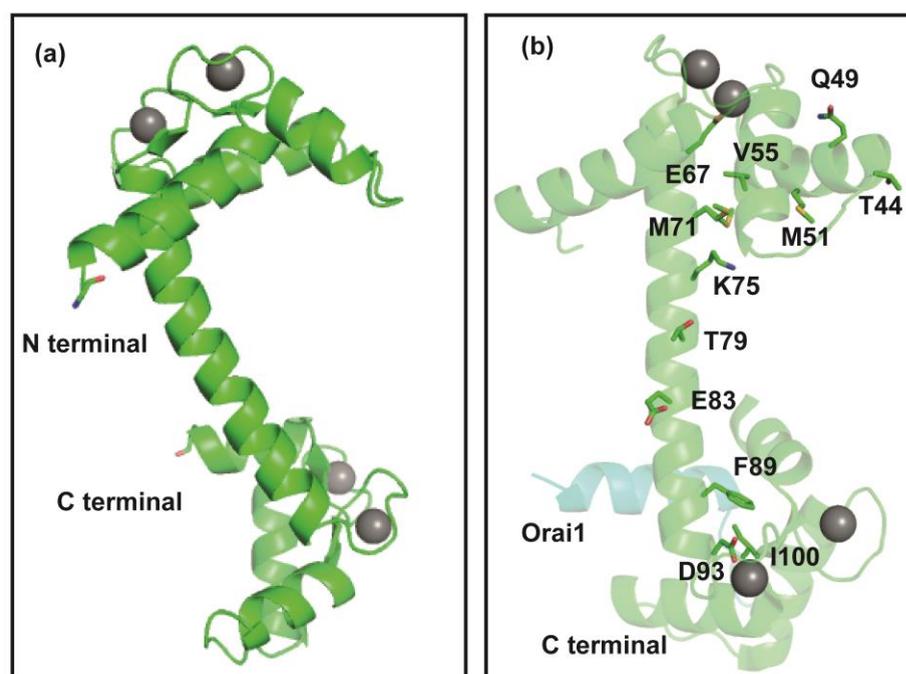


Fig.3.2.(a) Cartoon representation of the crystal structure of calcium bound Calmodulin (PDB id: 1CLL). (b) Path of correlated residues of Calmodulin in presence of Orai1, which connects C terminal domain to N terminal.

The rest of the chapter is organized as follows: we describe the methods in section 3.2. Section 3.3. describes detailed results on TDCF analysis. Mathematical

modeling to address TDCF qualitatively is elaborated in section 3.4. Section 3.5. contains discussions, overall conclusion along with novelty of our work is mentioned in section 3.6.

3.2. Methods:

We perform 1.05 μ s long all atom MD simulation of ubiquitin in NAMD³² parallel package and the calculations are performed over equilibrated trajectories. Details of MD simulation protocol, computation of two-point as well as four-point correlation functions, calculation details of transfer entropy and conformational entropy are explained in this section.

3.2.1. MD simulation:

We perform MD simulation using NAMD³² at 310 K and 1atm pressure, following standard protocols for NPT ensemble. We use TIP3P water model, periodic boundary condition and CHARMM27³³ force field with 1 fs time step. Electronutrality is maintained by adding proper number of mono-valent ions Na⁺ and Cl⁻. Long ranged electrostatic interaction is included by PME³⁴ method. Energy minimization was done for first 10,000 steps and simulation was performed for 1.05 μ s. Equilibration is ensured by RMSD plot over entire simulation time.

In order to address disturbance given by AMP to terminal G76 of ubiquitin, we simulate another system following the same protocol that is used for unperturbed ubiquitin. Here, the terminal G76 remains frozen throughout the entire simulation trajectories.

3.2.2. Computation of two-point correlation function (TDCF) :

We have computed dihedral angles from the specified atomic positions obtained from the simulated MD trajectory. The computation of TDCF from the MD trajectory

has been done as follows: The series of the conformations is ordered in time with a given choice of initial condition. For any given time-difference $\Delta t = |t_2 - t_1|$, we compute the product between fluctuations of dihedral θ of residue i and θ' of residue j for the l -th observation for a given Δt as $D_{i,j}^{(l)}(\theta\theta'; \Delta t) = (\sin[\theta_{i,l}(t_2)] - \langle \sin \theta_i \rangle)(\sin[\theta'_{j,l}(t_1)] - \langle \sin \theta'_j \rangle)$. Here, the angular bracket signifies ensemble average or mean of the respective quantity over simulation trajectory. The TDCF is given by $C_{i,j}^{(2)}(\theta\theta'; \Delta t) = \frac{1}{N_l} \sum_{l=1}^{N_l} C_{i,j}^{(2,l)}(\theta\theta'; \Delta t)$, where $C_{i,j}^{(2,l)}(\theta\theta'; \Delta t) = D_{i,j}^{(l)}(\theta\theta'; \Delta t) / \sqrt{\text{var}(\sin \theta_i) \text{var}(\sin \theta'_j)}$ and var denoting the variance of quantity within parenthesis. N_l is the number of observations corresponding to given Δt . For instance, for N number of observations, $N_0 = N$. Similarly, $N_1 = N - 1$ which correspond to the data set $(t_1 = 0, t_2 = 1), (t_1 = 1, t_2 = 2), \dots (t_1 = N - 1, t_2 = N)$. The computation is done for sufficiently large Δt until $C_{i,j}^{(2)}(\theta\theta'; \Delta t)$ approaches zero. We extract time scale by computing numerically Laplace Transform of TDCF as,

$$F_{i,j}(\theta\theta'; s) = \int_0^{t/2} e^{-st} C_{i,j}^{(2)}(\theta\theta'; \Delta t) dt.$$

3.2.3. Computation of four-point correlation function:

For a particular time interval Δt we compute four-point correlation function $C_{i,j}^{(4)}(\theta\theta'; \Delta t)$ between θ dihedral of residue i and θ' dihedral of residue j as,

$$C_{i,j}^{(4)}(\theta\theta'; \Delta t) = \frac{1}{N_l} \sum_{l=1}^{N_l} (D_{i,j}^{(l)}(\theta\theta'; \Delta t) - \langle D_{i,j}(\theta\theta'; \Delta t) \rangle)^2, \text{ where}$$

$\langle D_{i,j}(\theta\theta'; \Delta t) \rangle$ denotes the average value over N_l observations.

3.2.4. Computation of transfer entropy:

We use standard methodology for the computation of transfer entropy¹⁵. First we calculate the minimal embedding dimension $(m)^{35}$ for two time series by the false

nearest neighbors method³⁶. Next we compute mutual information (MI) using the TISEAN package³⁵ for fluctuations of pairs of dihedral angles for different residues belonging to the path. We identify the embedding time intervals (τ) for which MI is optimum^{10, 35}. We compute the transfer entropy¹⁰ between two degrees of freedom from the Shannon entropy and joint entropy using the MIToolbox³⁷ for fluctuations of two time series using m and τ .

3.2.5. Computation of conformational entropy:

We simulate another system where G76 of ubiquitin remains fixed throughout the 1.05 μ s long trajectories. The conformational changes in entropy associated with a particular dihedral φ in perturbed state (p) with respect to unperturbed free state (f) of ubiquitin is defined by $T\Delta S = T(S_p(\varphi) - S_f(\varphi))$. $S_p(\varphi)$ and $S_f(\varphi)$ can be obtained using Gibbs entropy formula $S(\varphi) = -K_B T \sum_i H_i(\varphi) \ln H_i(\varphi)$, $H_i(\varphi)$ is normalized probability distribution in each i th bins and sum is taken over histogram bins. K_B is the Boltzmann's Constant.

3.3. Simulation results:

In this section, we explain trivial nature of the TDCFs, correlation strength and relaxation time scales shown by the correlated and anti-correlated dihedrals of functional yet distant residue pairs of ubiquitin. Followed by the convergence of TDCF over simulation time scale, TDCF map, dynamically correlated path of ubiquitin and the mechanistic aspects of TDCFs. Finally, we demonstrate TDCFs for the residues belong to allosteric path of a protein, named Calmodulin.

3.3.1. Correlated and anti-correlated dihedral fluctuations:

We perform 1.05 μ s long all-atom MD simulations³² for Ub with initial input from crystal structure¹⁹ in explicit water. We analyze data using the portion of the simulated trajectory where the root mean squared deviation (RMSD) of the backbone atoms is saturated (Fig.3.3.).

We calculate the dihedral angles for backbone (φ, ψ) and side-chain (χ_1) of the residues of the protein. We plot the dihedrals χ_1 for residue pairs I13 and F45 denoted by $\chi_{1_{I13}}$ and $\chi_{1_{F45}}$ respectively as functions of time in Fig.3.4.(a). I13 and F45 have backbone distance ($d_{\alpha-\alpha}$), given by that of their C_α atoms as large as 1.5 nm. Despite that both $\chi_{1_{I13}}$ and $\chi_{1_{F45}}$ exhibit correlated behavior: The increase in one is coupled to increase in the other till very long time. Similarly, for Histidine; H68 and I44 with $d_{\alpha-\alpha} \sim 0.5$ nm, the plots of $\chi_{1_{H68}}$ of H68 and $\chi_{1_{I44}}$ of I44 as functions of time (Fig.3.4.(b)), reveal anti-correlated behavior even at long times.

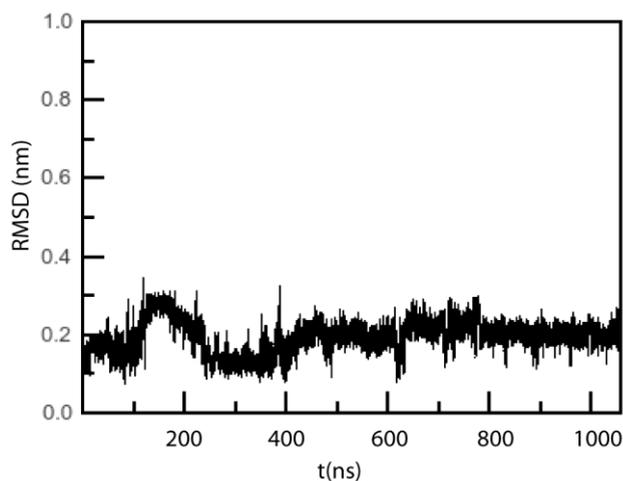


Fig.3.3. RMSD plot of ubiquitin over 1.05 μ s simulated trajectory.

3.3.2. Convergence in TDCF over MD simulated time scales:

Now we proceed to quantify temporal correlations between these time series. We extract the TDCF between dihedral θ of residue i and θ' of residue j in time

interval Δt from equilibrated trajectory. TDCF is denoted as $C_{i,j}^{(2)}(\theta\theta'; \Delta t)$, for a time interval $\Delta t = t_2 - t_1$. We compute TDCF from MD trajectory for three different sets of maximum time up to $t = 500$ ns, 950 ns and 1.05 μ s of the simulated trajectory. Since the correlation functions are computed from the dihedral values at two time intervals over trajectory, averaging at larger time interval gets better with longer time trajectory. We show $C_{I13,F45}^{(2)}(\chi_1\chi_1; \Delta t)$ (Fig.3.4.(c)) and $C_{H68,I44}^{(2)}(\chi_1\chi_1; \Delta t)$ (Fig.3.4.(d)) for three different cases. Figs.3.4.(c)-(d) show that data for trajectory up to $t = 500$ ns have differences with respect to larger time trajectories. However, data with trajectory up to $t = 950$ ns and 1.05 μ s are comparable, indicating saturation in the temporal behavior of the TDCFs.

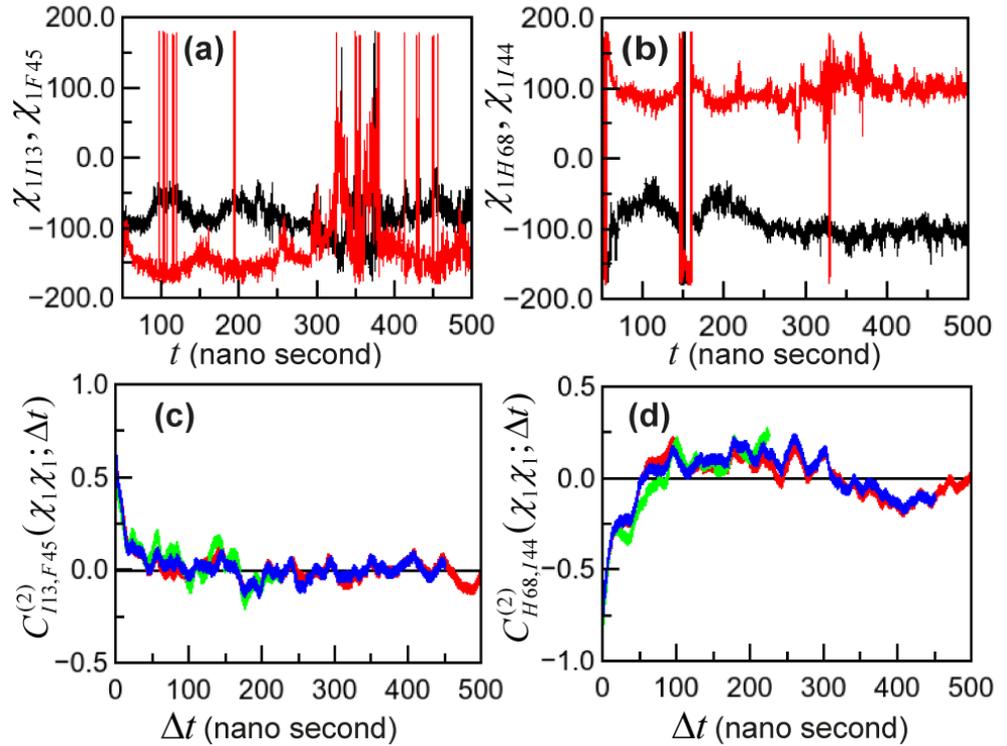


Fig.3.4.(a) Dihedral angles as functions of time t ; $\chi_{1_{I13}}$ (black) and $\chi_{1_{F45}}$ (red). **(b)** $\chi_{1_{H68}}$ (black) and $\chi_{1_{I44}}$ (red). **(c)** Convergence of TDCFs for three different $t = 500$ ns (green), 950ns (blue) and 1.05 μ s (red); $C_{I13,F45}^{(2)}(\chi_1\chi_1; \Delta t)$ and **(d)** $C_{H68,I44}^{(2)}(\chi_1\chi_1; \Delta t)$ as functions of Δt .

3.3.3. TDCFs among functional residues:

We further report our analysis based on all the data for the longest trajectory, $t = 1.05 \mu\text{s}$. Figs.3.5.(a) and (b) bring out further non-trivial aspect of the TDCFs, exhibited by several dihedral pairs, despite large separation between the residues. We show TDCFs for both forward and reverse direction, obtained by interchanging i and j and θ and θ' for the longest trajectory. $C_{I13,F45}^{(2)}(\chi_1\chi_1; \Delta t)$ shows statistical Pearson Correlation, $C_{I13,F45}^{(2)}(\chi_1\chi_1; 0) > 0$ at $\Delta t = 0$ as indicated in Fig.3.5.(a). The function further decays with increasing time interval. Similarly, $C_{H68,I44}^{(2)}(\chi_1\chi_1; \Delta t)$, in Fig.3.5.(b) shows initial statistical Pearson anti-correlation, $C_{H68,I44}^{(2)}(\chi_1\chi_1; 0) < 0$ at $\Delta t = 0$ following which it decays to zero for large Δt . We also observe in Figs.3.5.(a) and (b) that $C_{i,j}^{(2)}(\theta\theta'; \Delta t)$ is different in forward and reverse directions, the decay time scales being different indicating asymmetry in TDCFs.

Fig.3.5.(c) shows representative cases of Laplace Transform $F_{i,j}(\theta\theta'; s)$ of $C_{i,j}^{(2)}(\theta\theta'; \Delta t)$, the correlation function. For small s , $F_{I13,F45}(\chi_1\chi_1; s)$ has a maximum where there is a statistical correlation (Fig.3.5.(a)), while $F_{H68,I44}(\chi_1\chi_1; s) < 0$ and having a minimum in case of statistical anti-correlation (Fig.3.5.(b)). The asymmetry in $F_{i,j}(\theta\theta'; s)$ under interchanges of i and j and θ and θ' is evident from Fig.3.5.(c). The peak value of $F_{i,j}(\theta\theta'; s)$ in s , $F_{i,j}^{max}(\theta\theta')$ is a measure of the strength, and the inverse of peak position gives a characteristic time scale $\tau_{i,j}^{\theta\theta'}$ of correlation. These time scales (Table.3.1.) are tens of ns, in the regime of rotational diffusion time much larger than atomic fluctuation time scales. The log-log plots in Fig.3.5.(d) show decaying tail ($s^{-\kappa}$) with exponent κ for large s . The values of κ

(Table.3.1.) show quasi-universality ($\kappa \sim 1.0$). Additional cases for $F_{i,j}(\theta\theta';s)$ are shown in Figs.3.6.(a)-(f)).

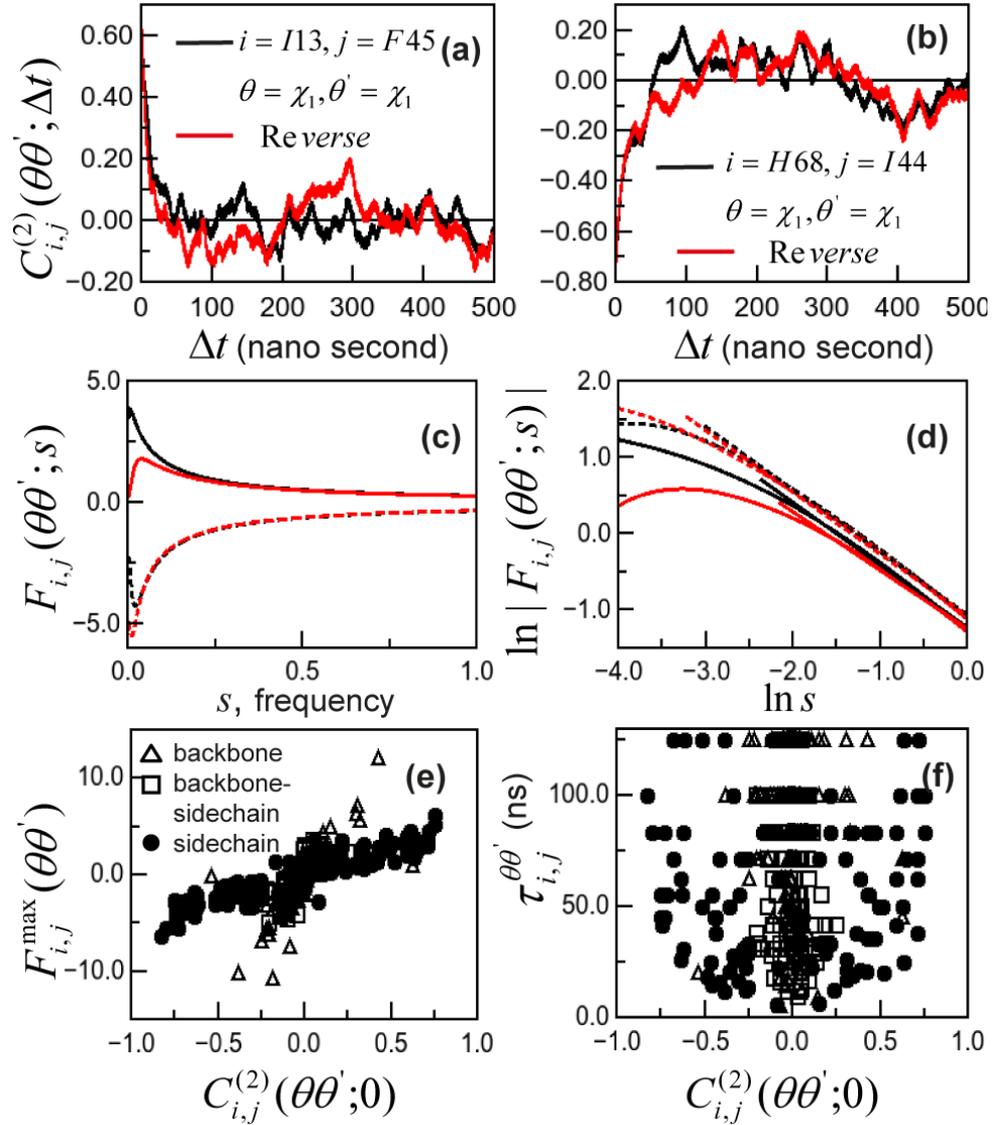


Fig.3.5. TDCFs between various dihedrals of ubiquitin, (Black: forward direction, red: reverse direction); (a) $C_{I13,F45}^{(2)}(\chi_1\chi_1; \Delta t)$ and (b) $C_{H68,I44}^{(2)}(\chi_1\chi_1; \Delta t)$ as functions of Δt . (c) Laplace Transform $F_{i,j}(\theta\theta'; s)$ of correlation function, $F_{I13,F45}(\chi_1\chi_1; s)$ (solid line) and $F_{H68,I44}(\chi_1\chi_1; s)$ (dashed line) versus s plots. (d) $\ln|F_{I13,F45}(\chi_1\chi_1; s)|$ versus $\ln s$ (solid line) and $\ln|F_{H68,I44}(\chi_1\chi_1; s)|$ versus $\ln s$ (dashed line) plots showing algebraic tails. (e) Correlations plot between functionally important residues of ubiquitin; $F_{i,j}(\theta\theta'; s)$ versus $C_{i,j}^{(2)}(\theta\theta'; 0)$. (f) $\tau_{i,j}^{\theta\theta'}$ versus $C_{i,j}^{(2)}(\theta\theta'; 0)$ for similar residues. The symbols have the same meaning in (e) and (f).

Table.3.1. Strength of TDCFs, correlation time scales and the exponents in larger frequency regimes for residue pairs having functional importance.

Residue-pair	Dihedral-pair	Exponents(κ)	$ F_{i,j}^{max}(\theta\theta') $	$\tau_{i,j}^{\theta\theta'}$ (ns)
I13-V5	$\chi_1\chi_1$	0.7	2.84	45.5
T14-F45	$\chi_1\chi_1$	0.7	2.08	71.4
L69-H68	$\psi\chi_1$	0.7	0.26	125.0
H68-I44	$\chi_1\chi_1$	0.8	4.25	45.5
K6-F45	$\chi_1\chi_1$	0.8	2.82	50.0
I44-F45	$\chi_1\chi_1$	0.8	3.63	55.6
K6-H68	$\chi_1\chi_1$	0.8	5.27	83.3
I13-F45	$\chi_1\chi_1$	0.8	3.83	125.0
I13-L67	$\chi_1\chi_1$	0.8	2.68	125.0
R72-V70	$\varphi\psi$	0.9	2.74	100.0
G76-R74	$\varphi\varphi$	0.9	7.13	125.0
K6-T12	$\chi_1\varphi$	0.9	1.15	125.0
H68-L67	$\psi\chi_1$	1.0	0.31	41.7
R74-R72	$\varphi\varphi$	1.0	10.25	100.0
V70-L69	$\psi\varphi$	1.1	0.35	71.4
V5-K6	$\chi_1\varphi$	1.3	0.15	55.6
L67-V5	$\psi\chi_1$	1.3	2.49	71.4
K6-L67	$\chi_1\varphi$	1.6	0.25	83.3

We next examine the TDCFs for functionally relevant residues⁴ of ubiquitin. In recent simulation studies⁴ statistical Pearson correlations are observed between backbone dihedrals of residue pairs (I13-F45, T14-F45, K6-F45, I13-Leucine; L67, I13-V5, V5-K6, K6-H68, H68-I44 and I44-F45) of ubiquitin, which belong to the binding surface patch of Ub. We plot $F_{i,j}^{max}(\theta\theta')$ and $\tau_{i,j}^{\theta\theta'}$ for all these residue pairs as functions of $C_{i,j}^{(2)}(\theta\theta'; 0)$ in Figs.3.5.(e) and (f) respectively. We observe that a strong

correlation exists between peak values of TDCFs and $C_{i,j}^{(2)}(\theta\theta';0)$. However, the timescales are not correlated to $C_{i,j}^{(2)}(\theta\theta';0)$. This is not surprising, for statistical correlation coefficients do not contain temporal information.

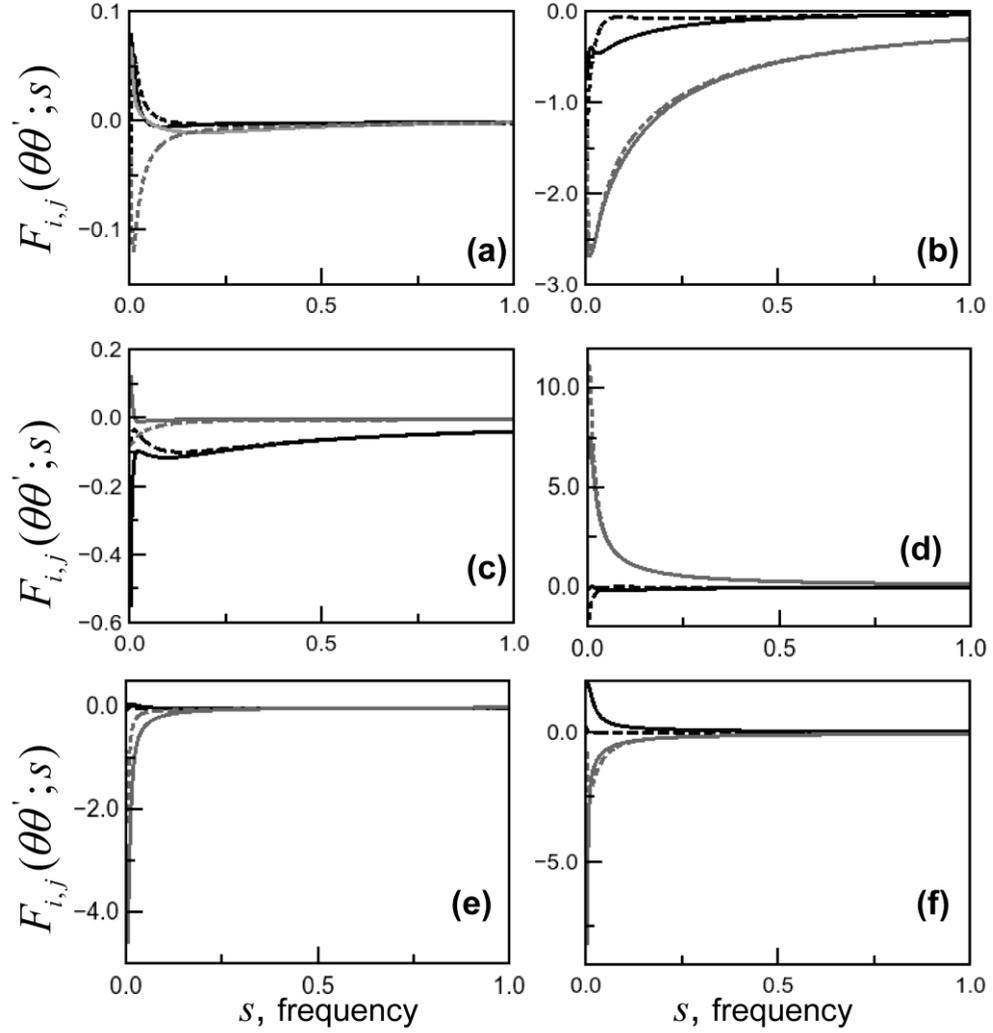


Fig.3.6. Laplace transform of TDCFs between functionally important residues (Solid line: forward direction, dashed: reverse direction): (a) $F_{I13,L67}(\varphi\varphi;s)$ (black) and $F_{I13,L67}(\varphi\psi;s)$ (grey). (b) $F_{I13,L67}(\psi\chi_1;s)$ (black) and $F_{I13,L67}(\chi_1\chi_1;s)$ (grey). (c) $F_{I13,V5}(\psi\Phi;s)$ (black) and $F_{H68,I44}(\psi\varphi;s)$ (grey). (d) $F_{I13,V5}(\psi\chi_1;s)$ (black) and $F_{G76,R74}(\varphi\varphi;s)$ (grey). (e) $F_{V70,L8}(\varphi\psi;s)$ (black) and $F_{R42,R74}(\psi\psi;s)$ (grey). (f) $F_{V70,L8}(\varphi\chi_1;s)$ (black) and $F_{R42,R74}(\chi_1\psi;s)$ (grey).

3.3.4. TDCF map and dynamically correlated path in Ub:

Our analysis yields a detailed map of correlated residue pairs R_i and R_j , as shown in Fig.3.7.(a). For a particular residue pair R_i and R_j we compute $C_{R_i,R_j}^{(2)}(\theta\theta';\Delta t)$ for every possible pairs of degrees of freedom (dof), like $\varphi\varphi', \varphi\psi', \varphi\chi_1', \psi\psi', \psi\psi', \psi\chi_1', \chi_1\varphi', \chi_1\psi', \chi_1\chi_1'$. The dof pairs for which $|F_{R_i,R_j}^{max}(\theta\theta')|$ is maximum is considered to determine the direction of correlation in time domain, namely, if any perturbation at R_i affects R_j at a later time or vice versa. We generate a 76×76 matrix by noting $|F_{R_i,R_j}^{max}(\theta\theta')|$ for all of the 76 residues of ubiquitin. By applying the condition of directionality we obtain the upper triangular matrix showing detailed TDCF map.

This map can be used to understand correlated path among the residues. Let us consider the terminal residue G76 which binds to AMP during Ub activation in ubiquitination. The dihedral φ of G76, φ_{G76} is correlated to Arginine; R74 by φ_{R74} , ψ_{R74} and χ_{1R74} both in forward and reverse direction. However, among all these correlated dihedrals $F_{G76,R74}^{max}(\varphi\varphi)$ is the largest which we take as an indication that G76 is downstream correlated to R74 via dihedral φ of both the residues. Similarly G76 is downstream correlated to other set of residues, like L73, L67, Q62, Tyrosine; Y59, L56, R54, D52, K48, F45, L43, Q41, Q40, D39, Proline: P38, K33, I30, K27, V26, E24, I23, D21, T12, T7 and V5. Among all the downstream correlated residues to G76 we find that R74 is having the shortest $d_{\alpha-\alpha}$, which is the mean distance between C_α atoms over the entire trajectory. Similarly, the closest downstream correlated residue to R74 is R72. In this way we construct the path of downstream correlated dihedrals of different residues, $\varphi_{G76} - \varphi_{R74} - \varphi_{R72} - \psi_{V70} - \varphi_{L69} -$

$\chi_{1_{H68}} - \chi_{1_{L67}} - \chi_{1_{V5}}$, as shown in a snapshot of ubiquitin obtained from simulation (Fig.3.7.(b)). Among these temporally correlated residues G76, R74 and R72 belong to the C terminal loop region. The residues V70, L69, H68 and L67 belong to β_5 , while V5 belongs to β_1 in β strands of the crystal structure¹⁹.

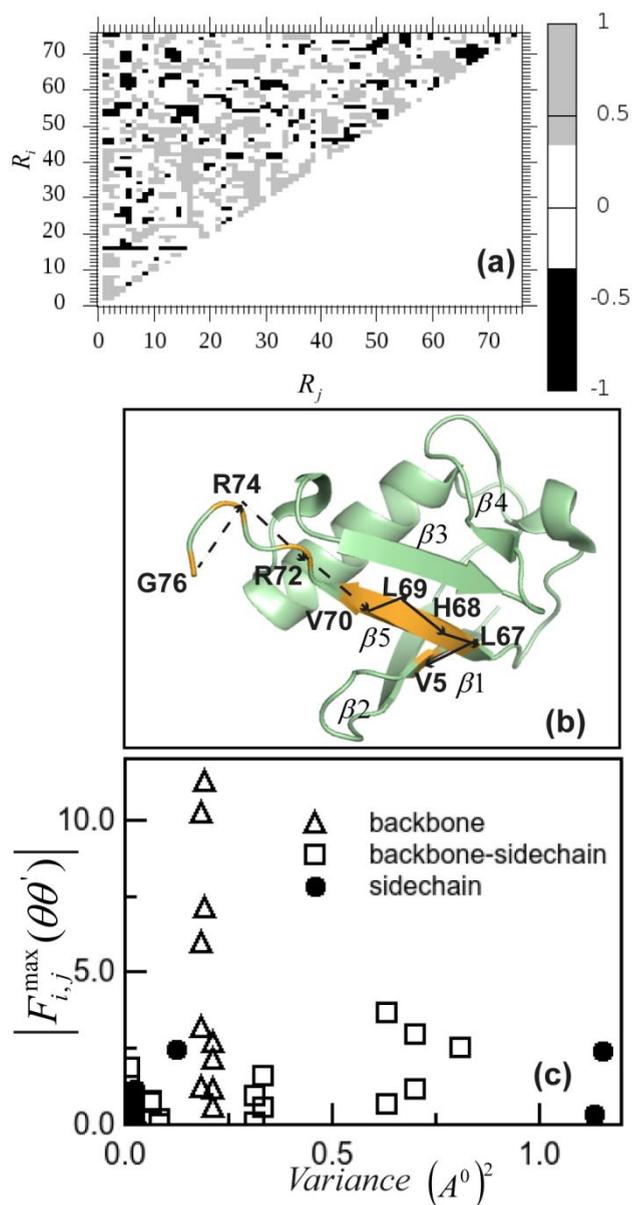


Fig.3.7.(a) TDCF map for any two residue pair in ubiquitin; Black represents downstream and Grey represents upstream TDCFs. (b) Residues belong to dynamically correlated path of ubiquitin. Solid line connects the residue pairs belong to β -sheets, dashed line connects the pairs belong to the loop region. (c) Correlation peak versus distance fluctuations of residue pairs belong to temporally correlated path.

Crystal structure of ubiquitin activation enzyme-E1 loaded with Ub molecules indicates^{21, 22} that the hydrophobic surface patch of Ub including L8, I44, V70 and C terminal tail of Ub (G76, R74, R72) interact with the activation enzyme. The temporally correlated path with G76 contains many of the residues, like R74, R72 and V70. Besides, the slowest time scale in this path is that between G76 and R74, around 125 ns. This time scale is comparable to the rotational time scale of the enzyme which is about 90 ns obtained using the Stokes-Einstein³⁸ equation. Thus the path obtained using TDCF analysis is functionally relevant.

3.3.5. Mechanistic view of TDCF:

In order to get mechanistic view of long distance correlations, we calculate variance of the distances between residues belonging to the temporally correlated path. We generate average structure of ubiquitin over simulated trajectory. We superimpose $\phi - \psi$ correlation plots for all residues of the protein in crystal structure as well as of in average structure (Fig.3.8.). We find that there are no changes in secondary structural element.

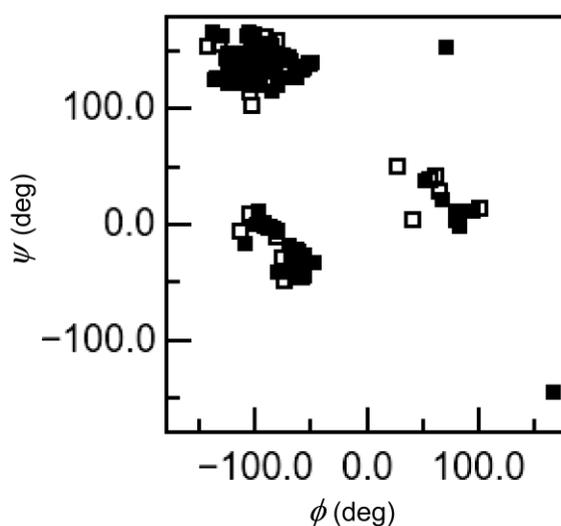


Fig.3.8. $\phi - \psi$ Correlation plot of residues in ubiquitin; filled rectangle represents the crystal structure and hollow rectangle shows the simulated average structure of ubiquitin obtained from equilibrated trajectory.

Next, we compute $var(d_{\alpha-\alpha})$ which represents variance of $d_{\alpha-\alpha}$. Similarly for backbone-side chain distances, we calculate $var(d_{\alpha-\beta})$, where $d_{\alpha-\beta}$ denotes distance between C_α and C_β atoms of the residue pair. For side chain dihedrals, we compute $var(d_{\beta-\beta})$, the variance of distance between C_β atoms of the correlated pairs. We plot $|F_{i,j}^{max}(\theta\theta')|$ versus these variances in Fig.3.7.(c). We observe that the large correlation amplitudes are clustered near smaller values of the variance. This indicates that dihedral dynamical correlations are destroyed by large fluctuations.

3.3.6. TDCF for allosteric path in CaM:

CaM, a Ca^{2+} binding protein is known to interact with a protein, known as Orai1³⁰ to regulate cellular functions in eukaryotes. There are some residues of C terminal domain of CaM which are reported to interact with Orai1. These residues of CaM show noticeable changes in side chain dihedral distributions in Orai1-bound state with respect to Orai1-free state. These residues are considered to form a path to connect N terminal domain of CaM with its C terminal region (Fig.3.2.(b)). Thus, any perturbation given at a particular residue belong to the C domain can modulate functionality of N domain. This type of long distance communication between two distant yet functionally connected residues generates allosteric^{4, 39-41} effect within a protein.

Here, we probe TDCFs between each pair of the residues belong to the allosteric path of CaM. We show two such representative cases for TDCFs in Figs.3.9.(a)-(b). We observe that the TDCF between χ_1 of E67 and χ_1 of V55, $C_{E67,V55}^{(2)}(\chi_1\chi_1;\Delta t)$ (Fig.3.9.(a)) is initially correlated with a strong statistical correlation coefficient ~ 0.69 . However, with increasing time interval an anti-correlation (negative values) develops and finally the function decays to zero at larger time interval. On the other

hand, in Fig.3.9.(b) $C_{M51,Q49}^{(2)}(\chi_1\chi_1; \Delta t)$ exhibits initial anti-correlation, then small amount of correlation grows and gradually diminishes at larger time interval. We quantify the amplitude of the correlations from the Laplace transforms. We find that for the initially correlated pair, $F_{E67,V55}(\chi_1\chi_1; s)$ develops a peak for lower s (Fig.3.9.(c)) along with maximum value ~ 3.36 . Similarly for the initially anti-correlated pair, $F_{M51,E49}(\chi_1\chi_1; s)$ has a dip in lower frequency regime (Fig.3.9.(d)) along with a minimum value ~ 0.90 . We take the magnitude of the maximum or minimum as the strength of dynamic correlations, as shown in Table.3.2. between all the residue pairs. All our data represent that the residue pairs which belong to the path generated from static analysis are also correlated in time domain.

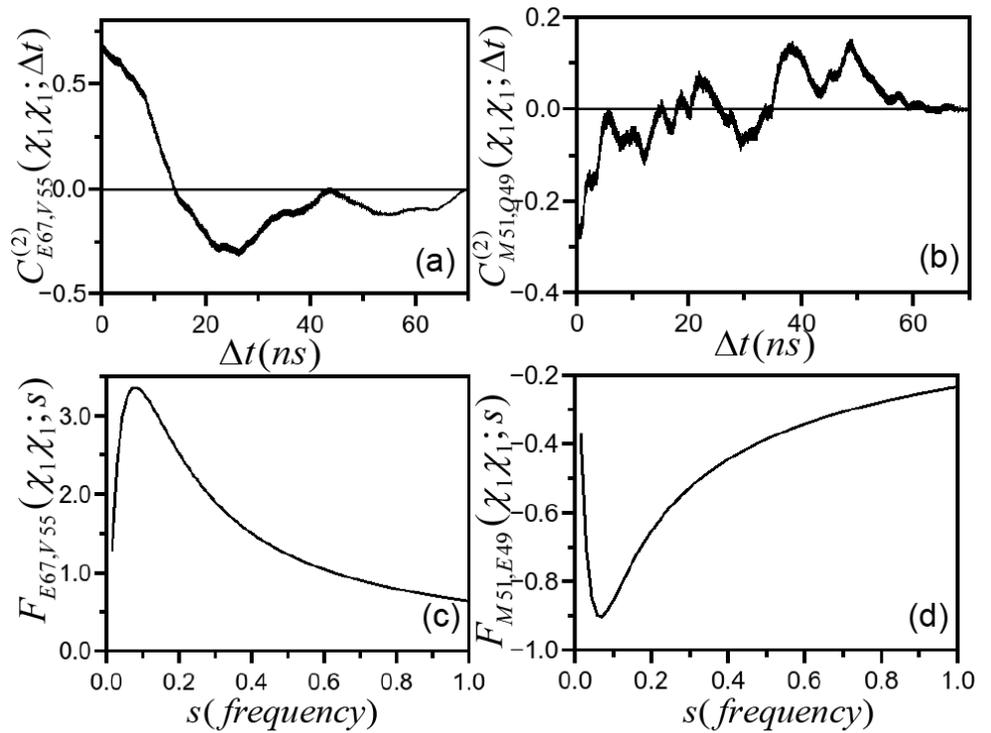


Fig.3.9. Temporal correlations as functions of time interval Δt between residue pairs belonging to the path: (a) $C_{E67,V55}^{(2)}(\chi_1\chi_1; \Delta t)$ and (b) $C_{M51,Q49}^{(2)}(\chi_1\chi_1; \Delta t)$. (c) Laplace Transform of correlation function versus frequency plots: $F_{E67,V55}(\chi_1\chi_1; s)$ and (d) $F_{M51,E49}(\chi_1\chi_1; s)$.

Table.3.2. Amplitudes of temporal correlations between side chains dihedrals of residue pairs belong to the path.

Residue Pairs	Amplitude
I100-D93	0.41
D93-F89	0.38
F89-E83	0.91
E83-T79	2.57
T79-K75	0.38
K75-M71	0.48
M71-E67	0.13
E67-V55	3.36
V55-M51	0.36
M51-Q49	0.90
Q49-T44	0.93

3.4. Mathematical model:

We have used equation of motion of over damped coupled classical harmonic oscillators to mathematically model TDCFs between two different dihedrals.

$$-\Gamma_i \dot{\theta}_i(t) + \omega_i^2 \theta_i(t) + \alpha' \theta_j(t) = 0 \quad (3.1)$$

$$-\Gamma_j \dot{\theta}_j(t) + \omega_j^2 \theta_j(t) + \beta' \theta_i(t) = 0 \quad (3.2).$$

Using Laplace Transform of equations (3.1) and (3.2) we model behavior of TDCF in frequency, s domain. Details are in Appendix I, Chapter 3.

We model the qualitative behaviors of the TDCF in terms of equations of motion of two dihedrals $\theta_i(t)$ and $\theta_j(t)$ which are coupled to each other with strengths α' and β' respectively. Let the characteristic frequencies associated with them be ω_i and ω_j . They perform motions in a solvent experiencing drags proportional to $-\Gamma_i\dot{\theta}_i$ and $-\Gamma_j\dot{\theta}_j$ respectively¹³. We calculate the Laplace transformed correlation function from the equations of motion, $\langle \theta_i(s)\theta_j(s) \rangle$ in frequency, s by averaging over initial conditions on the variables. We find that in $s \rightarrow 0$ limit, $\langle \theta_i(s)\theta_j(s) \rangle \sim -C_{i,j}^{(2)}(\theta\theta'; 0) \left(\Gamma_i^2 \Gamma_j^2 / (\omega_i^2 \omega_j^2)^2 \right) s^2$, where $C_{i,j}^{(2)}(\theta\theta'; 0)$ being the statistical correlation coefficient. There is thus maximum in the low s limit if the TDCF shows statistical correlation, $C_{i,j}^{(2)}(\theta\theta'; 0) > 0$, while a minimum for statistically anti-correlated TDCF with $C_{i,j}^{(2)}(\theta\theta'; 0) < 0$. These are qualitatively similar to low s behaviour of the simulated TDCFs. We get an algebraic tail for large s , $\langle \theta_i(s)\theta_j(s) \rangle \sim s^{-2}$ where the exponent is universal and independent of the parameters in the model. This universality is revealed by the simulated TDCF, albeit with exponent 1.0. The difference in the exponent may be due to simplicity of the model equations of motion where all effects are neglected except solvent drag and mutual coupling. Moreover, we find that so far as $\alpha' \neq \beta'$, $\langle \theta_i(s)\theta_j(s) \rangle \neq \langle \theta_j(s)\theta_i(s) \rangle$ as seen in simulations.

3.5. Discussions:

Direct probe of dynamical correlation among the dihedrals is difficult due to limitation of probes. However, our analysis suggests an indirect way of probing the dynamical correlations. Our analysis shows that the residues, like R74, R72 and V70

lie in dynamically correlated path with G76 where ubiquitination initiates. We expect these residues to play an important role in the process which can be tested experimentally. R72 is experimentally known to give specificity to the ubiquitin activation enzyme-E1 binding^{21, 22}. The role of the other residues needs to be looked into.

The success of two-point correlation functions to bring out the spatio-temporal correlations lead us to examine the higher order correlation. We compute typical four-point correlation functions between dihedral θ of residue i and θ' of residue j for a fixed Δt , denoted by $C_{i,j}^{(4)}(\theta\theta'; \Delta t)$ which measures the variance of the correlated fluctuations of the dihedral angles for a given time window Δt . $C_{I13,F45}^{(4)}(\chi_1\chi_1; \Delta t)$, $C_{H68,I44}^{(4)}(\chi_1\chi_1; \Delta t)$ are shown in Fig.3.10. Using two-point correlation functions we observe that $|C_{I13,F45}^{(2)}(\chi_1\chi_1; 0)|$ is about 0.63 (Fig.3.4.(a)) and $|C_{H68,I44}^{(4)}(\chi_1\chi_1; 0)|$ lies around 0.73 (Fig.3.4.(b)). Whereas $|C_{I13,F45}^{(4)}(\chi_1\chi_1; 0)|$ is around 0.15 and $|C_{H68,I44}^{(4)}(\chi_1\chi_1; 0)|$ lies around 0.25 as for all Δt (Fig.3.10.). Thus the four-point correlation functions are smaller in the magnitude than the corresponding two point functions in the time domain of biological relevance. The higher correlations may be important for very large time scales involving large scale domain motions.

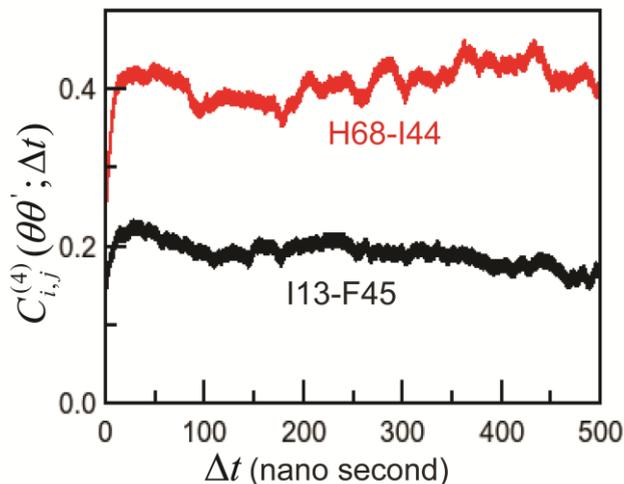


Fig.3.10. Four-point correlation functions as functions of Δt , $C_{I13,F45}^{(4)}(\chi_1\chi_1; \Delta t)$ (black) and $C_{H68,I44}^{(4)}(\chi_1\chi_1; \Delta t)$ (red).

Transfer entropy is another method to probe connection or information transfer between two variables. Here, we compute transfer entropy from mutual information³⁵ between maximally correlated pairs of fluctuating degrees of freedom which constitute the functionally relevant path. We assign the directionality of entropy transfer between two residues by the larger magnitude of the transfer entropy for the correlated pairs in forward and reverse directions. For instance, in case of $C_{G76,R74}^{(2)}(\theta\theta'; \Delta t)$, φ_{G76} and φ_{R74} are the maximally correlated dof both in forward and reverse directions. We find that mutual transfer entropy -0.17 for φ_{R74} to φ_{G76} and that for the reverse direction is 0.55, indicating that the transfer of information takes place from φ_{G76} to φ_{R74} similar to experimental observations. We construct in similar way the direction of entropy transfer and time scales of the correlated dof over the path, as given in Table. 3.3. It is clear from the table that the directionality of path is not maintained between R74 and R72 where the entropy transfer takes place from R72 to R74. Moreover, the time scale of optimum mutual information is in sub-ns range, orders of magnitude shorter than biologically relevant time scales. Thus the TDCF describes the functionally relevant path in more reliably.

Table.3.3. Embedding dimension (m), optimal time interval (τ in ps) and transfer entropy for the residues belonging to temporally correlated path, both in forward and reverse direction.

Forward direction	(m, τ) of 1 st residue	(m, τ) of 2 nd residue	Transfer entropy in forward direction	Reverse direction	(m, τ) of 1 st residue	(m, τ) of 2 nd residue	Transfer entropy in reverse direction
φ_{G76} - φ_{R74}	5, 5	5, 18	0.55	φ_{R74} - φ_{G76}	5, 18	5, 5	-0.17
φ_{R74} - φ_{R72}	5, 18	5, 5	0.52	φ_{R72} - χ_{1R74}	5, 5	5, 20	0.76
φ_{R72} - ψ_{V70}	5, 5	5, 11	0.62	χ_{1V70} - χ_{1R72}	5, 20	5, 17	0.65

ψ_{V70} - φ_{L69}	5, 11	5, 4	-0.07	φ_{L69} - ψ_{V70}	5, 4	5, 11	0.53
φ_{L69} - χ_{1H68}	5, 4	5, 5	0.14	ψ_{H68} - χ_{1L69}	5, 2	5, 4	0.70
χ_{1H68} - χ_{1L67}	5, 5	5, 9	-0.35	χ_{1L67} - φ_{H68}	5, 9	5, 5	-0.27
χ_{1L67} - χ_{1V5}	5, 9	5, 10	0.66	φ_{V5} - χ_{1L67}	5, 3	5, 9	0.51

Next, we address connection between dynamic responses of ubiquitin with respect to its conformational thermodynamics. In earlier studies^{25, 26}, it has been reported that one can address changes in free energy (ΔG) and entropy ($T\Delta S$) of a protein in a given conformation with respect to a reference state using Histogram Based Method on dihedral equilibrium fluctuations. The destabilized ($\Delta G > 0$) and disordered ($T\Delta S > 0$) residues of a particular state are found to participate in further binding activities in order to reduce free energy. Ubiquitin functions upon interaction with AMP at its terminal residue G76. Thus, in order to capture changes in conformational thermodynamics of ubiquitin before and after bonding with AMP, we further simulate ubiquitin by keeping G76 fixed over the simulation. We calculate TDCFs between G76 and the residues belonging to the dynamically correlated path as obtained in the unperturbed conformation. Next, we estimate changes in conformational entropy of those particular residues in perturbed state with respect to the unperturbed one.

We extract strength of dynamic correlations and time scales between φ dihedral of G76 and all other residues belong to temporally correlated path of ubiquitin, for each possible dofs ($\varphi\varphi$, $\varphi\psi$, $\varphi\chi_1$). We compute $T\Delta S(\theta')$ for each θ' dofs of the

residues belong to dynamically correlated path in perturbed state with respect to their unperturbed conformation. Next, we plot $F_{i,j}^{max}(\theta\theta')$ and $\tau_{i,j}^{\theta\theta'}$ separately with respect to $T\Delta S(\theta')$ in Figs.3.11.(a)-(b) respectively. We observe that $F_{i,j}^{max}(\theta\theta')$ exhibits a strong anti-correlation with respect to changes in conformational entropy along with a correlation coefficient ~ 0.72 . Similarly correlation time scales also show anti-correlated behavior towards $T\Delta S(\theta')$ with correlation coefficient ~ 0.50 . Dotted lines represent linear fitting. Increase in conformational entropy signifies enhance in flexibility of that particular residue. Thus if a residue fluctuates more its correlation with G76 tends to decrease fast with lower strength of dynamic correlation.

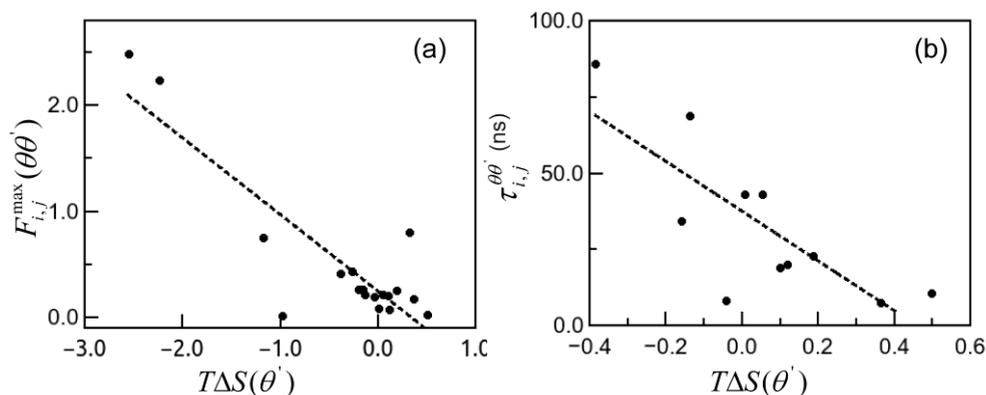


Fig.3.11. Anti-correlation between TDCFs and changes in conformational entropy for G76 and the residues belong to dynamically correlated path. (a) $F_{i,j}^{max}(\theta\theta')$ vs $T\Delta S(\theta')$ and (b) $\tau_{i,j}^{\theta\theta'}$ vs $T\Delta S(\theta')$.

3.6. Conclusions:

To summarize, we show with long molecular simulations and mathematical modeling that TDCFs explain the causal connection between binding sites in Ub in the biologically relevant temporal regime⁴². More importantly, our studies indicate that non-linearities are not the primary deciding factor for causal connections between functional sites. Although the simulations are illustrated for ubiquitin activation

enzyme-E1 binding to Ub and Orail binding to CaM, the generality of our mathematical analysis shows that qualitative features of TDCF can be extended to any microscopic degrees of freedom. On a wider perspective, two point cross-correlation functions between relevant microscopic variables may provide a correct description of bio-molecular function. The related kinetics in terms of underlying microscopic dynamics can be explained without invoking the non-linear effects.

We probe connectivity between changes in conformational entropy of ubiquitination in two different states with the dynamic response manifested by TDCFs. Moreover, we observe that in case of calmodulin, the residues that belong to allosteric path, show noticeable changes in dihedral distributions along with significant TDCFs. The correlated behaviour between TDCFs and conformational entropy suggest that TDCF might be considered as an alternative theoretical approach to address static and dynamic fluctuations of proteins to associated thermodynamics.

Appendix I

Over damped equations of motion in the long time limit:

The over damped equations of motion in long time limit are:

$$-\Gamma_i \dot{\theta}_i(t) + \omega_i^2 \theta_i(t) + \alpha' \theta_j(t) = 0 \quad (1),$$

$$-\Gamma_j \dot{\theta}_j(t) + \omega_j^2 \theta_j(t) + \beta' \theta_i(t) = 0 \quad (2).$$

Using Laplace transforms of equations (1) and (2) we obtain,

$$-s\Gamma_i \theta_i(s) + \omega_i^2 \theta_i(s) + \alpha' \theta_j(s) = -\Gamma_i \theta_i(0) \quad (3),$$

$$-s\Gamma_j \theta_j(s) + \omega_j^2 \theta_j(s) + \beta' \theta_i(s) = -\Gamma_j \theta_j(0) \quad (4),$$

or writing equation (3) and (4) in matrix form:

$$\begin{pmatrix} \omega_i^2 - s\Gamma_i & \alpha' \\ \beta' & \omega_j^2 - s\Gamma_j \end{pmatrix} \begin{pmatrix} \theta_i(s) \\ \theta_j(s) \end{pmatrix} = - \begin{pmatrix} \Gamma_i \theta_i(0) \\ \Gamma_j \theta_j(0) \end{pmatrix} \quad (5).$$

Denoting,

$$A = \begin{pmatrix} \omega_i^2 - s\Gamma_i & \alpha' \\ \beta' & \omega_j^2 - s\Gamma_j \end{pmatrix} \text{ and } \text{Det}A = \begin{vmatrix} \omega_i^2 - s\Gamma_i & \alpha' \\ \beta' & \omega_j^2 - s\Gamma_j \end{vmatrix} \text{ we get,}$$

$$\begin{pmatrix} \theta_i(s) \\ \theta_j(s) \end{pmatrix} = - \begin{pmatrix} \omega_i^2 - s\Gamma_i & \alpha' \\ \beta' & \omega_j^2 - s\Gamma_j \end{pmatrix}^{-1} \begin{pmatrix} \Gamma_i \theta_i(0) \\ \Gamma_j \theta_j(0) \end{pmatrix},$$

$$\begin{pmatrix} \theta_i(s) \\ \theta_j(s) \end{pmatrix} = - \frac{1}{\text{Det}A} \begin{pmatrix} \omega_j^2 - s\Gamma_j & -\alpha' \\ -\beta' & \omega_i^2 - s\Gamma_i \end{pmatrix} \begin{pmatrix} \Gamma_i \theta_i(0) \\ \Gamma_j \theta_j(0) \end{pmatrix}.$$

Equation (5) yields the following expressions:

$$\theta_i(s) = - \frac{1}{\text{Det}A} [\Gamma_i \theta_i(0)(\omega_j^2 - s\Gamma_j) - \alpha' \Gamma_j \theta_j(0)] \quad (6),$$

$$\theta_j(s) = -\frac{1}{\text{Det}A} [\Gamma_j \theta_j(0)(\omega_i^2 - s\Gamma_i) - \beta' \Gamma_i \theta_i(0)] \quad (7).$$

We construct the product between $\theta_i(s)$ and $\theta_j(s)$ is using equation (6) and (7)

and taking average over initial condition to get the TDCF

$$\begin{aligned} \langle (\theta_i(s)\theta_j(s)) \rangle &= \frac{1}{(\text{Det}A)^2} \times [-\Gamma_i^2(\omega_j^2 - s\Gamma_j)\beta' \langle \theta_i(0)\theta_i(0) \rangle \\ &+ \alpha'\beta'\Gamma_i\Gamma_j \langle \theta_j(0)\theta_i(0) \rangle + (\omega_j^2 - s\Gamma_j)(\omega_i^2 - s\Gamma_i)\Gamma_i\Gamma_j \langle \theta_i(0)\theta_j(0) \rangle \\ &- \alpha'\Gamma_j^2(\omega_i^2 - s\Gamma_i) \langle \theta_j(0)\theta_j(0) \rangle] \quad (8). \end{aligned}$$

Here, $\langle \theta_i(0)\theta_i(0) \rangle = C_{ii} = 1$, $\langle \theta_j(0)\theta_j(0) \rangle = C_{jj} = 1$ and

$$\langle \theta_i(0)\theta_j(0) \rangle = C_{i,j}^{(2)}(\theta\theta'; 0) = \langle \theta_j(0)\theta_i(0) \rangle.$$

Introducing these in equation (8),

$$\begin{aligned} \langle (\theta_i(s)\theta_j(s)) \rangle &= \frac{1}{(\text{Det}A)^2} \times [\Gamma_i^2\Gamma_j^2 C_{i,j}^{(2)}(\theta\theta'; 0)s^2 + s\Gamma_i\Gamma_j [(\beta'\Gamma_i + \\ &\alpha'\Gamma_j) - C_{i,j}^{(2)}(\theta\theta'; 0)(\omega_j^2\Gamma_i + \omega_i^2\Gamma_j)] + [(\Gamma_i\Gamma_j C_{i,j}^{(2)}(\theta\theta'; 0)(\omega_i^2\omega_j^2 + \\ &\alpha'\beta')] - (\Gamma_i^2\omega_j^2\beta' + \alpha'\omega_i^2\Gamma_j^2)] \quad (9). \end{aligned}$$

One can find from equation (9) that for large s , the leading term in numerator is s^2

and that in denominator is s^4 . Thus $\langle (\theta_i(s)\theta_j(s)) \rangle \sim s^{-2}$. Let us now consider

$s \rightarrow 0$ limit. Expanding the numerator and denominator in equation (9),

$$\begin{aligned}
\langle \theta_i(s)\theta_j(s) \rangle &= \frac{1}{(\omega_i^2\omega_j^2 - \alpha'\beta')^2} \times \left(\Gamma_i^2\Gamma_j^2 C_{i,j}^{(2)}(\theta\theta'; 0) s^2 + s\Gamma_i\Gamma_j \left[(\beta'\Gamma_i + \right. \right. \\
&\alpha'\Gamma_j) - C_{i,j}^{(2)}(\theta\theta'; 0)(\omega_j^2\Gamma_i\omega_i^2\Gamma_j) \left. \right] + \left[(\Gamma_i\Gamma_j C_{i,j}^{(2)}(\theta\theta'; 0)(\omega_i^2\omega_j^2 + \alpha'\beta') - \right. \\
&\left. \left. (\Gamma_i^2\omega_j^2\beta' + \alpha'\omega_i^2\Gamma_j^2) \right] \right) \times \left(1 + \frac{2s(\omega_j^2\Gamma_i + \omega_i^2\Gamma_j)}{(\omega_i^2\omega_j^2 - \alpha'\beta')} - s^2 \left[\frac{2\Gamma_i\Gamma_j}{(\omega_i^2\omega_j^2 - \alpha'\beta')} + \frac{(\omega_j^2\Gamma_i + \omega_i^2\Gamma_j)^2}{(\omega_i^2\omega_j^2 - \alpha'\beta')^2} \right] \right)
\end{aligned}
\tag{10}.$$

From equation (10) we obtain coefficient of s^2 ,

$$\begin{aligned}
&\frac{\Gamma_i^2\Gamma_j^2 C_{i,j}^{(2)}(\theta\theta'; 0)}{(\omega_i^2\omega_j^2 - \alpha'\beta')^2} + \left[\frac{2(\omega_j^2\Gamma_i + \omega_i^2\Gamma_j)\Gamma_i\Gamma_j}{(\omega_i^2\omega_j^2 - \alpha'\beta')^3} \left\{ (\beta'\Gamma_i + \alpha'\Gamma_j) - C_{i,j}^{(2)}(\theta\theta'; 0)(\omega_j^2\Gamma_i + \omega_i^2\Gamma_j) \right\} \right] - \\
&\left[\left\{ \frac{2\Gamma_i\Gamma_j}{(\omega_i^2\omega_j^2 - \alpha'\beta')^3} + \frac{(\omega_j^2\Gamma_i + \omega_i^2\Gamma_j)^2}{(\omega_i^2\omega_j^2 - \alpha'\beta')^4} \right\} \left\{ (\Gamma_i\Gamma_j C_{i,j}^{(2)}(\theta\theta'; 0)(\omega_i^2\omega_j^2 + \alpha'\beta')) - \right. \right. \\
&\left. \left. (\Gamma_i^2\omega_j^2\beta' + \alpha'\Gamma_j^2\omega_i^2) \right\} \right].
\end{aligned}$$

Assuming $\omega_i^2\omega_j^2 \gg \alpha'\beta'$, consider the terms having highest order of $\Gamma_i\Gamma_j$ that is independent of $\alpha'\beta'$ and have $C_{i,j}^{(2)}(\theta\theta'; 0)$ dependence. In this limit the coefficient of

s^2 is $\sim -C_{i,j}^{(2)}(\theta\theta'; 0) \frac{\Gamma_i^2\Gamma_j^2}{(\omega_i^2\omega_j^2)^2}$. Thus for statistical anticorrelation ($C_{i,j}^{(2)}(\theta\theta'; 0) <$

0) coefficient of s^2 becomes +ve quantity which implies minimum value of

$F_{i,j}(\theta\theta'; s)$ for low s . For initial correlation ($C_{i,j}^{(2)}(\theta\theta'; 0) > 0$) coefficients of s^2

remain -ve, implying a maximum of $F_{i,j}(\theta\theta'; s)$ for small s .

Interchanging i and j and α' and β' in equation (1) and (2), it is easy to check that

asymmetry in TDCF that $\langle \theta_i(s)\theta_j(s) \rangle \neq \langle \theta_j(s)\theta_i(s) \rangle$.

3.7. References:

1. A. Vallee-Belisle and K. W. Plaxco, *Curr. Opin. Struct. Biol.*, 2010, **20**, 518-526.
2. M. Wachsmuth, C. Conrad, J. Bulkescher, B. Koch, R. Mahen, M. Isokane, R. Pepperkok and J. Ellenberg, *Nat. Biotechnol.*, 2015, **33**, 384-U391.
3. B. Vogeli, *J. Chem. Phys.*, 2010, **133**, 014501.
4. R. B. Fenwick, S. Esteban-Martin, B. Richter, D. Lee, K. F. Walter, D. Milovanovic, S. Becker, N. A. Lakomek, C. Griesinger and X. Salvatella, *J. Am. Chem. Soc.*, 2011, **133**, 10336-10339.
5. D. Long and R. Bruschweiler, *J. Phys. Chem. Lett.*, 2012, **3**, 1722-1726.
6. R. B. Fenwick, L. Orellana, S. Esteban-Martin, M. Orozco and X. Salvatella, *Nat. Commun.*, 2014, **5**, 1-9.
7. K. Kasahara, I. Fukuda and H. Nakamura, *Plos One*, 2014, **9**, e112419.
8. D. W. Li, D. Meng and R. Bruschweiler, *J. Am. Chem. Soc.*, 2009, **131**, 14610-14611.
9. A. Black Pyrkosz, J. Eargle, A. Sethi and Z. Luthey-Schulten, *J. Mol. Biol.*, 2010, **397**, 1350-1371.
10. H. Kamberaj and A. van der Vaart, *Biophys. J.*, 2009, **97**, 1747-1755.
11. Y. Ryabov, *Proteins*, 2015, **83**, 1571-1581.
12. T. Schreiber, *Phys. Rev. Lett.*, 2000, **85**, 461-464.
13. Lubensky. T. C and Chaikin. P. M, *Principles of condensed matter physics*, Cambridge University Press, Cambridge, 1995.
14. J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids, Fourth Edition*, Elsevier, 2013.
15. N. G. V. Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier Science, Amsterdam, 1992.
16. D. R. Sisan, D. Yarar, C. M. Waterman and J. S. Urbach, *Biophys. J.*, 2010, **98**, 2432-2441.
17. N. Garnier, D. Genest and M. Genest, *Biophys. Chem.*, 1996, **58**, 225-237.
18. M. G. Amit Das, and J. Chakrabarti, *Chem. Phys. Lett.*, 2016, **645**, 200-204.
19. S. Vijaykumar, C. E. Bugg and W. J. Cook, *J. Mol. Biol.*, 1987, **194**, 531-544.
20. B. A. Schulman and J. W. Harper, *Nat. Rev. Mol. Cell. Biol.*, 2009, **10**, 319-331.
21. A. Schafer, M. Kuhn and H. Schindelin, *Acta Crystallogr D*, 2014, **70**, 1311-1320.

22. I. Lee and H. Schindelin, *Cell*, 2008, **134**, 268-278.
23. J. M. Winget and T. Mayor, *Mol. Cell*, 2010, **38**, 627-635.
24. M. S. Marlow, J. Dogan, K. K. Frederick, K. G. Valentine and A. J. Wand, *Nat. Chem. Biol.*, 2010, **6**, 352-358.
25. S. Sikdar, J. Chakrabarti and M. Ghosh, *Mol. BioSyst.*, 2014, **10**, 3280-3289.
26. A. Das, J. Chakrabarti and M. Ghosh, *Biophys. J.*, 2013, **104**, 1274-1284.
27. R. D. Brokx, M. M. Lopez, H. J. Vogel and G. I. Makhatadze, *J. Biol. Chem.*, 2001, **276**, 14083-14091.
28. Z. G. Xia and D. R. Storm, *Nat. Rev. Neurosci.*, 2005, **6**, 267-276.
29. Y. S. Babu, C. E. Bugg and W. J. Cook, *J. Mol. Biol.*, 1988, **204**, 191-204.
30. F. M. Mullins, C. Y. Park, R. E. Dolmetsch and R. S. Lewis, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 15495-15500.
31. L. Maganti, S. Dutta, M. Ghosh and J. Chakrabarti, *J. Biomol. Struct. Dyn.*, 2018, 1-10.
32. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781-1802.
33. B. R. Brooks, C. L. Brooks, 3rd, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545-1614.
34. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
35. R. Hegger, H. Kantz and T. Schreiber, *Chaos*, 1999, **9**, 413-435.
36. M. B. Kennel, R. Brown and H. D. Abarbanel, *Phys. Rev. A*, 1992, **45**, 3403-3411.
37. G. Brown, A. Pocock, M. J. Zhao and M. Lujan, *J Mach Learn Res*, 2012, **13**, 27-66.
38. K. L. Linegar, A. E. Adeniran, A. F. Kostko and M. A. Anisimov, *Colloid. J.*, 2010, **72**, 279-281.
39. R. B. Fenwick, L. Orellana, S. Esteban-Martín, M. Orozco and X. Salvatella, *Nat. Commun.*, 2014, **5**, 1-9.

40. J. J. Guo and H. X. Zhou, *Chem. Rev.*, 2016, **116**, 6503-6515.
41. J. Liu and R. Nussinov, *PLOS Comput. Biol.*, 2016, **12**, e1004966.
42. S. Dutta, M. Ghosh and J. Chakrabarti, *Sci. Rep.*, 2017, **7**, 40439.

CHAPTER 4

Anomalous dynamics of metal ions in presence of a protein

4.1. Introduction:

The proteins functionality involving ligand binding are mediated primarily via surface interactions^{1, 2}. Apart from the obvious importance in cellular processes, ligand attached proteins have large potential for bio-nanotechnological applications³⁻⁵ due to their easy tenability by external conditions, like temperature, pH and so on. In this chapter, we report diffusion profiles of the metal ions in presence of electrostatically heterogeneous protein surface and connect the observations towards proteins functionality.

The microscopic basis of ligand-protein surface interaction is far from understood. Recent bioinformatics studies⁶ indicate that ion-dipole interaction over protein surface provide useful insights to functional activities of protein. Protein surfaces are quite intriguing⁷: Firstly, water soluble proteins expose their hydrophilic groups, both polar and charged to aqueous environment, leading to electrostatically heterogeneous surface composition⁸. Such surface heterogeneities are probed by Single-Molecule tracking experiments⁹. Secondly, the protein surface fluctuates due to internal motions of the atoms, studied extensively in the past both experimentally and theoretically¹⁰⁻¹³.

Ligand-protein binding has been investigated through static perspectives primarily via x-ray crystallography¹⁴ and nuclear magnetic resonances¹⁵. Both methods albeit being capable of producing microscopically the most elaborate

information, have their own limitations. Metal ions binding sites can be predicted using solution NMR¹⁵, although results depend on protein stability and spectral quality. NMR relaxation can also predict metal ion bindings but this experiment is suitable only for the paramagnetic ions¹⁶. In case of X-ray¹⁴, crystallization condition and modeling technique can affect protein conformation, ligand position and may lead to wrong ligand binding site. The spectroscopic analysis for characterizing metal associated proteins, undergo potential loss of native metal ions due to protein expression protocols¹⁷. High-throughput X-ray absorption spectroscopy can predict metal type but cannot locate the binding sites¹⁸ which is retrieved by prior knowledge of bioinformatics and computational model¹⁹.

In this study, we examine theoretically the possibility of using dynamic information towards prediction of metal ions binding sites over a protein surface. In order to illustrate this we consider dynamic properties of metal ions in the vicinity of a protein in aqueous environment. We address how the heterogeneity of electrostatic interaction on dynamic protein surface leaves imprint on ionic motion. Metal ions are present in trace amount (\sim mM concentration²⁰) in cellular environment but can control folding path and function of a number of proteins²¹. We choose a small protein, named, ubiquitin that plays an important role in numerous cellular processes like transcription, translation, cell trafficking to protein degradation and is known to bind different divalent metal ions (Cd^{2+} , Pt^{2+} , Zn^{2+} , Mg^{2+} , Ca^{2+}). The crystal structure (2XK5.PDB)^{22, 23} shows that the surface exposed acidic residues including E16, E18, D21, D32 and E64 of ubiquitin can bind numerous Zn^{2+} ions (Fig.4.1.(a)). The heterogeneity of electrostatic surface of ubiquitin is depicted in Fig.4.1.(b).

The dynamical features in condensed matter systems are typically extracted from the self-van Hove function (self-vHf) $G_s(\Delta r, \Delta t)$, the probability

distribution of displacement Δr of particles in time interval Δt ; and the MSD, the mean squared displacements of a tagged particle from an initial position²⁴. The self-vHf and MSD can be measured by neutron scattering and DLS experiments²⁵⁻²⁸ respectively. We compute the self-vHf and MSD of Zn^{2+} ions in presence of ubiquitin from all-atom molecular dynamics (MD) trajectories²⁹ in an electro-neutral aqueous medium at low salt (ZnCl_2) concentration (~ 20 mM). We observe that Zn^{2+} ions have non-Gaussian self-vHf along with exponentially decaying tail and MSD increasing linearly in time. This is quite unlike normal diffusive motion of free Zn^{2+} ions without any protein, where the self-vHf is Gaussian and MSD, linear in time. This heterogeneous diffusion is due to strong trapping of the ions on the heterogeneous protein surface. In order to probe the connection between diffusion dynamics of Zn^{2+} and the trapping, we study two more systems; in one case we mutate the acidic residue E18 to hydrophobic Alanine and for the second one, we mutate both the E18 and D21 into Alanine. We observe that for the mutated cases, the self-vHf of Zn^{2+} exhibits Gaussian profile, albeit double Gaussian nature. By mutating acidic residues into hydrophobic ones, the interaction between Zn^{2+} and the acidic residues gets significantly reduced. As a consequence, Zn^{2+} ions get attached to isolated residues intermittently, leading to double Gaussian profile.

The rest of the chapter is organized as follows; we describe methods in details in section 4.2. Results are elaborated in section 4.3. We discuss an application for an uncharacterized protein in section 4.4. and the conclusion is given in section 4.5.

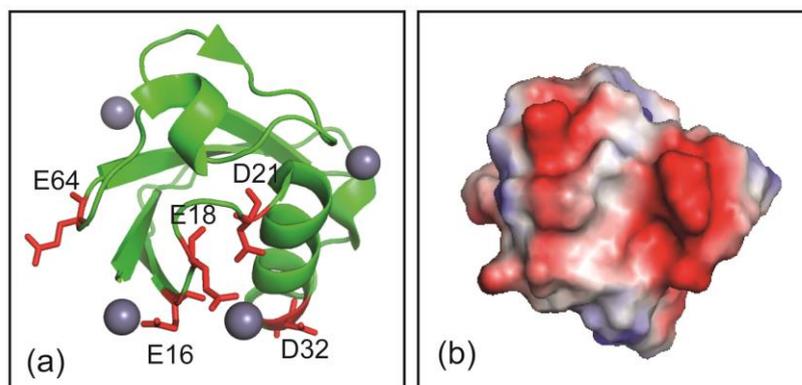


Fig.4.1.(a) Crystal structure of ubiquitin in presence of Zn^{2+} , showing the binding residues. (b) Electrostatic surface of ubiquitin, (red; -ve charge and blue; +ve charge distribution).

4.2. Methods:

In this section we provide details of MD simulation with analysis over the trajectories.

4.2.1. Details of MD simulation:

We consider the initial conformation of monomeric ubiquitin in presence of seven Zn^{2+} ions from the available crystal structure and perform 0.5 μs long all-atom MD simulation. The salt concentration for $ZnCl_2$ is ~ 20 mM. We use the standard protocol for isothermal isobaric ensemble (NPT) with 310K and 1 atm pressure using GROMACS³⁰ package. Electro-neutrality is achieved by adding proper number of Na^+ and Cl^- ions and periodic boundary conditions are implemented. The protein is solvated with water in a cubic box of dimension ~ 80.0 Å. The interactions between different atoms have been modeled using the GROMOS9653a6 force field³¹ along with SPC 3-point water model for water molecules. Long ranged electrostatic interaction is estimated by particle mesh Ewald³² method. Simulation time step is 1 fs.

The dynamical quantities are computed over unfolded trajectories. In order to get better average, we perform five different simulations using different initial configurations. The equilibration is ensured from the root mean squared deviation (RMSD) (Fig.3.2.(a)) of the C_α atom, which is the first carbon atom attached to the functional group of each amino acid. The analysis has been carried out for the last 200 ns for each trajectory of total time up to 500 ns. Data for length is scaled by diameter of carbon atom in our analysis. We follow the same protocol to simulate Zn^{2+} in presence of mutated structures of ubiquitin (RMSDs shown in Figs.4.2.(b)-(c)) and also for the free Zn^{2+} diffusing within a water box.

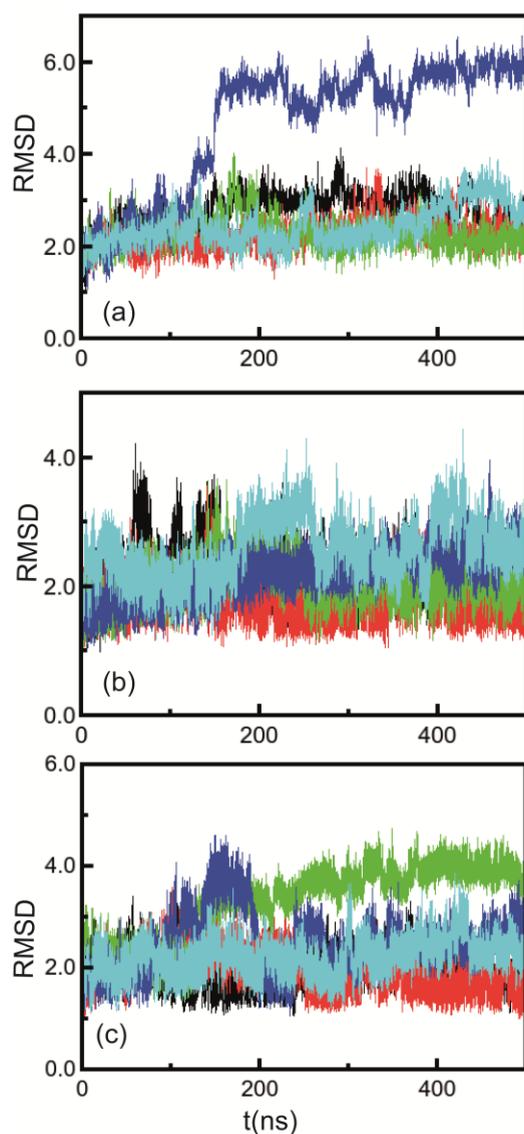


Fig.4.2. RMSD for the C_α atom of each residues of the protein shown for five different trajectories for three different cases of ubiquitin; (a) Crystal structure of ubiquitin in presence of Zn^{2+} , (b) For the one mutation case of ubiquitin in presence of Zn^{2+} and that for (c) the two mutation case.

4.2.2. Details of self-vHf calculation:

We generate self-vHf for the ions using³³, $G_{s,\xi}(\Delta r, \Delta t) = \frac{1}{N\rho} \times \langle \sum_{i=1}^N \delta(\Delta r - |\mathbf{r}_{\xi,i}(\mathbf{t}) - \mathbf{r}_{\xi,i}(\mathbf{t} + \Delta \mathbf{t})|) \rangle$. Here, $\mathbf{r}_{\xi,i}(\mathbf{t} + \Delta \mathbf{t})$ denotes the coordinates of the i th ion of a given ionic species ξ after time interval Δt starting from initial position, $\mathbf{r}_{\xi,i}(\mathbf{t})$. N is total number of ions of a particular species in the system and ρ the number density of the ions. The angular brackets denote average over choices of $\{\mathbf{r}_{\xi,i}(\mathbf{t})\}$ associated with each $\Delta \mathbf{t}$ over the five different trajectories. Further averaging is considered by taking the equal length of equilibrated trajectory with 21 different origins chosen between 300-320 ns at an interval of 1.0 ns for each single trajectory. This time interval is comparable to typical time scale of protein backbone motion³⁴.

4.2.3. MSD calculation:

The diffusion dynamics investigated in terms of MSD³⁵ plot gives displacement of the particle over time from its initial position. Two well-defined regimes³⁵ on MSD have been reported theoretically as well as experimentally, ballistic regime at shorter time interval where MSD has square dependence on time (t) and diffusive regime where MSD linearly depends on t at larger time interval. However, sub-diffusive motion³⁶ along with $\text{MSD} \sim t^\beta$ where $\beta < 1$ or plateau region can exist in MSD over t , for diffusion in crowded medium. This indicates caging³⁷ effect, the trapping of the particle by neighboring environment on intermittent time scale.

We compute MSD of the ions of a given species ξ as $\delta r_\xi^2(t)$ over equilibrated trajectories using³⁸: $\delta r_\xi^2(t) = \frac{1}{N} \langle \sum_{i=1}^N |\mathbf{r}_{\xi,i}(\mathbf{t}) - \mathbf{r}_{\xi,i}(\mathbf{0})|^2 \rangle$, with ensemble average over five different trajectories along with 21 different origins. The

translational diffusion coefficients (D) are obtained from slope of MSD vs t plot in the long time limit²⁴.

4.2.4. Density of the ions around the residues:

We compute the density of ions about the known Zn^{2+} binding residues of ubiquitin as $\rho_{\xi}(R) = \langle \frac{n_{\xi}(R, \Delta R)}{N_1 N(N_1 + N) 4\pi R^2 \Delta R} \rangle$. Here R is the separation between an ion of the ξ species and C_{α} atoms of a residue. $n_{\xi}(R, \Delta R)$ denotes the number of ions in a spherical shell of radius R and $R + \Delta R$. N is the total number of ions, whereas N_2 implies number of residues taken into account. The angular brackets define average on structures over five different trajectories.

4.3. Results:

The self-vHf of a system is the density-density auto correlation function. In this section, we elaborate different nature of self-vHfs for Zn^{2+} ions in four different circumstances. Firstly, we study the diffusion dynamics of free Zn^{2+} , secondly, in presence of ubiquitin, where the ions are trapped to surface residues. We next analyze how the diffusivity of the ions is changed if we reduce the strength of interaction by mutating the hydrophilic residues to hydrophobic ones. The self-vHfs are computed from the coordinates of Zn^{2+} ions at different time intervals and averaged over different trajectories. The time intervals for these spatio temporal distributions of Zn^{2+} ions are computed from $\Delta t = 0.5$ ns to 3.0 ns. Beyond this, the magnitude of the probability function is reduced significantly and thus it leads to large statistical error in data interpretation. Moreover, the Gaussian nature of self-vHf is reclaimed at 3.0 ns for each case. This suggests that this time interval is sufficient enough to capture the entire diffusion dynamics of the ions over the simulated trajectories.

4.3.1. The self-vHf for Zn^{2+} ions:

We show the behavior of self-vHf for Zn^{2+} ions over different Δt , for four different circumstances: case (I). Zn^{2+} moving freely in water box ($G_{s,\text{Zn}}^{(f)}(\Delta r, \Delta t)$); case (II). Zn^{2+} in presence of non-mutated binding residues of ubiquitin ($G_{s,\text{Zn}}^{(0)}(\Delta r, \Delta t)$); case (III). Zn^{2+} in a system, where one acidic residue E18 is mutated to hydrophobic Alanine ($G_{s,\text{Zn}}^{(1)}(\Delta r, \Delta t)$) and finally case (IV). in presence of double mutations, both E18 and D21 are mutated to Alanine ($G_{s,\text{Zn}}^{(2)}(\Delta r, \Delta t)$). The logarithmic of self-vHfs vs Δr plots for various cases are shown in Figs.4.3.(a)-(f) and in Figs.4.4.(a)-(f). The data for $\ln G_{s,\text{Zn}}^{(f)}(\Delta r, \Delta t)$ at Δt (= 0.5 ns) is given in Fig.4.3.(a). In order to extract functional dependence of $G_{s,\text{Zn}}^{(f)}(\Delta r, \Delta t)$ over Δr , we fit this using square dependence as indicated by the dotted line. The goodness of the fit (R^2) along with the fitting errors (χ^2) are given in Table.4.1. Here, we observe that $G_{s,\text{Zn}}^{(f)}(\Delta r, \Delta t)$ is Gaussian in nature. The same nature is observed for $\Delta t = 2.0$ ns (Fig.4.3.(b)) as well. Finally, $G_{s,\text{Zn}}^{(f)}(\Delta r, \Delta t)$ shows square dependence over larger Δr at further larger $\Delta t = 3.0$ ns (Fig.4.3.(c)).

We follow the same analysis for $G_{s,\text{Zn}}^{(0)}(\Delta r, \Delta t)$ in presence of ubiquitin in Figs.4.3.(d)-(f). We observe that unlike the free case, $G_{s,\text{Zn}}^{(0)}(\Delta r, \Delta t)$ shows deviation from square dependence over Δr at Δt (= 0.5 ns) in Fig.4.3.(d). It exhibits square dependence up to a threshold value ($r_c \sim 1.90$) of Δr . Beyond that r_c , it fits better to a linear dependence, suggesting existence of long exponential tail in self-vHf. This linear tail in self-vHf persists up to $\Delta t = 2.0$ ns (Fig.4.3.(e)). Although, in this case the square dependence is observed up to a larger value of ($r_c \sim 4.00$) than the previous

one. Finally, at a further larger Δt ($= 3.0$ ns), $G_{s,Zn}^{(0)}(\Delta r, \Delta t)$ becomes Gaussian (Fig.4.3.(f)).

We next consider self-vHf for Zn^{2+} ions in mutated cases (Fig.4.4.(a)-(f)). Fig.4.4.(a) indicates that in presence of singly mutated ubiquitin, there are two different square dependences in $G_{s,Zn}^{(1)}(\Delta r, \Delta t)$ for two different regimes of Δr . The first Gaussian dependence persists up to a certain value of $r_c \sim 4.00$, after that r_c , $G_{s,Zn}^{(1)}(\Delta r, \Delta t)$ shows a different square dependence. The profile remains similar at $\Delta t = 2.0$ ns (Fig.4.4.(b)), along with slightly increasing range of $r_c \sim 4.10$. We get back the Fickian Gaussian profile of self-vHf at $\Delta t = 3.0$ ns (Fig.4.4.(c)) for larger Δr . We also observe that the nature of $G_{s,Zn}^{(2)}(\Delta r, \Delta t)$ remains similar to $G_{s,Zn}^{(1)}(\Delta r, \Delta t)$ for various Δt (Figs.4.4.(d)-(f)). We find that unlike the free case, in presence of strong electrostatic attraction, exerted by the acidic residues of ubiquitin, Zn^{2+} ions show linear tail in corresponding self-vHf. However, even by single mutation of the acidic residue reduces the electrostatic attraction that leads to noticeable changes in self-vHfs of the ions

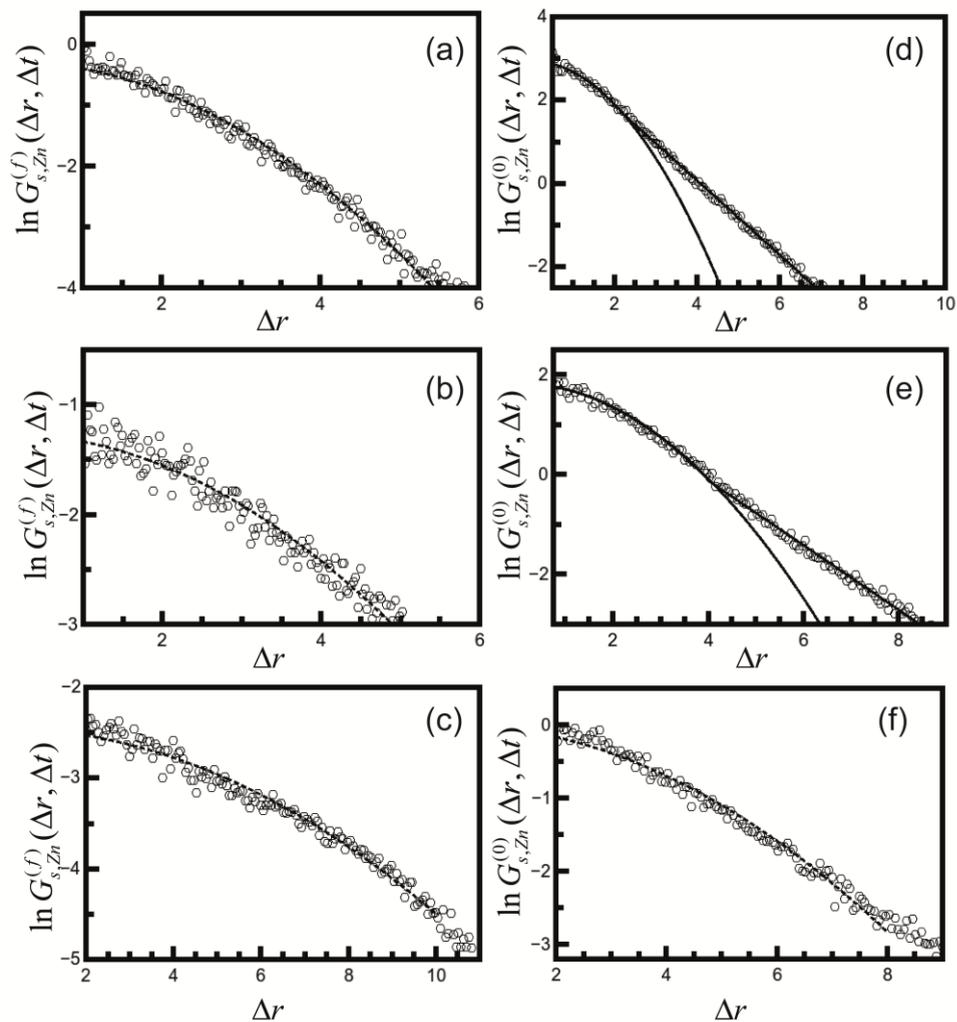


Fig.4.3. $\ln G_{s,Zn}^{(f)}(\Delta r, \Delta t)$ and $\ln G_{s,Zn}^{(0)}(\Delta r, \Delta t)$ vs Δr plots for three different time intervals. (a) Normal Gaussian diffusive profile for the free Zn^{2+} ions at $\Delta t = 0.5$ ns, that for (b) $\Delta t = 2.0$ ns and (c) $\Delta t = 3.0$ ns. (d) Deviation from Gaussian dependence of self-vHf of the ions in presence of ubiquitin. Square dependence is prominent for small Δr and linear tail at larger Δr as indicated by the solid lines. (e) $\ln G_{s,Zn}^{(0)}(\Delta r, \Delta t)$ at $\Delta t = 2.0$ ns. (f) Gaussian nature of $G_{s,Zn}^{(0)}(\Delta r, \Delta t)$ is retained at $\Delta t = 3.0$ ns.

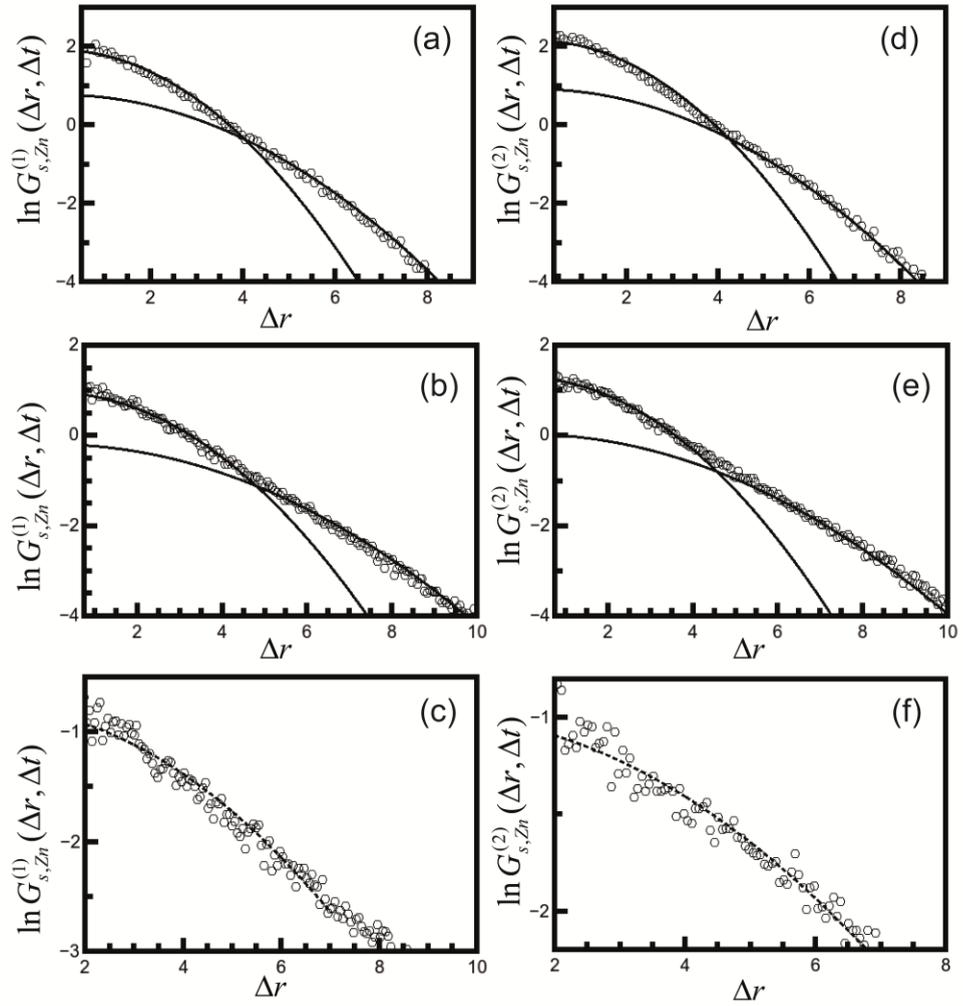


Fig.4.4. $\ln G_{s,Zn}^{(1)}(\Delta r, \Delta t)$ vs Δr plots for three different time intervals. (a) Deviation from normal single Gaussian dependence, double Gaussian diffusive profile for the Zn^{2+} ions at $\Delta t = 0.5$ ns, that for (b) $\Delta t = 2.0$ ns and (c) Gaussian nature of $G_{s,Zn}^{(1)}(\Delta r, \Delta t)$ is retained at $\Delta t = 3.0$ ns. (d) Double Gaussian behavior for $G_{s,Zn}^{(2)}(\Delta r, \Delta t)$ at $\Delta t = 0.5$ ns and (b) $\Delta t = 2.0$ ns. (e) Fickian diffusion is obtained at $\Delta t = 3.0$ ns.

Table.4.1. Fitting parameters for self-vHf of Zn²⁺ ions in different surroundings.

System	Δt (ns)	Fitting	Range (Δr)	R^2	χ^2
Zn-free	0.5	square	1.00-5.50	0.99	0.002
Zn-free	2.0	square	1.00-5.00	0.94	0.002
Zn-free	3.0	square	2.00-10.00	0.98	0.009
Zn in no mutation	0.5	square	0.60-1.90	0.95	0.003
Zn in no mutation	0.5	linear	2.20-7.00	0.99	0.004
Zn in no mutation	2.0	square	0.50-4.00	0.98	0.007
Zn in no mutation	2.0	linear	4.00-9.00	0.99	0.009
Zn in no mutation	3.0	square	2.00-8.00	0.94	0.008
Zn in one mutation	0.5	square	0.50-4.00	0.98	0.008
Zn in one mutation	0.5	square	4.50-9.00	0.98	0.008
Zn in one mutation	2.0	square	0.75-4.10	0.98	0.009
Zn in one mutation	2.0	linear	5.00-10.00	0.98	0.009
Zn in one mutation	3.0	square	2.00-7.00	0.97	0.009
Zn in two mutation	0.5	square	0.50-4.00	0.98	0.007
Zn in two mutation	0.5	square	4.90-9.00	0.98	0.007
Zn in two mutation	2.0	square	0.75-4.10	0.98	0.004
Zn in two mutation	2.0	square	5.00-10.00	0.99	0.007
Zn in two mutation	3.0	square	2.00-7.00	0.96	0.006

4.3.2. MSD of the ions:

Next, we compute MSD of the ions in all of the four different circumstances. Fig.4.5. shows that $\delta r_{Zn}^2(t)$ in each case depends linearly on t . The translational diffusion coefficient of Zn^{2+} (D_{Mg}) in free case is obtained as $\sim 0.65 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ and that in presence of ubiquitin is $\sim 0.34 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$, whereas for the one mutation system it lies $\sim 0.42 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ and that for two mutation $\sim 0.48 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$. The D values for free Zn^{2+} in aqueous solution³⁹ at room temperature is reported to be around $0.70 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$. However, in presence of ubiquitin, Zn^{2+} ions diffuse at slower rate in mutated as well as in non-mutated case. Although, the rate of slowing down of diffusion coefficient is much more prominent in presence of unmutated ubiquitin.

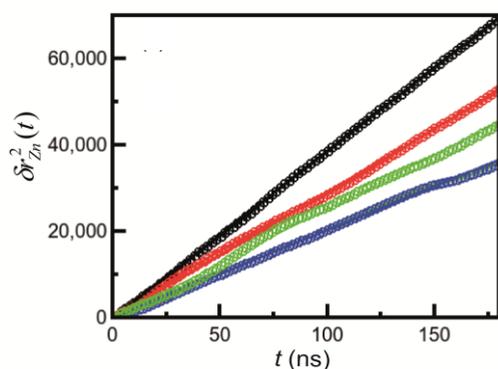


Fig.4.5. MSD; $\delta r_{Zn}^2(t)$ as a function of time t , black: free case, red: in presence of two mutation, green: in presence of one mutation and blue: for nonmutated structure of ubiquitin.

4.3.3. Dynamic heterogeneity:

The Gaussian self-vHf and linear dependence of MSD over t are characteristics of Fickian diffusion dynamics in a normal liquid²⁴. However, linear MSD but self-vHf with exponential tail has been observed in many soft matter systems⁴⁰. The normal Gaussian self-vHf is retrieved only at large time. This behavior is known as non-Fickian diffusion⁴⁰⁻⁴⁵ and can be observed if the particles show signature of trapping²⁵. In general, slowly varying heterogeneous fluctuations of

environment comparable to the fast jiggly dynamics of diffusive particles lead to Brownian yet non-Gaussian diffusion profile^{40, 45, 46}. The non-Fickian dynamics has been phenomenologically described in terms of distributions of diffusion coefficients^{40-45, 47}, where instantaneous diffusion coefficient of individual tracer particles changes gradually. At times less than diffusivity correlation times, motion of each particles form local domain of given diffusivity, which leads to spatial heterogeneity⁴¹⁻⁴³. However at larger time scales Gaussian dynamics is restored along with an overall mean diffusion coefficient. This is a signature of dynamic heterogeneity so that there are particles in the system having characteristic time scales faster or slower than the average⁴⁰⁻⁴⁵. Our data suggest non-Fickian diffusion dynamics of metal ions in presence of strong interaction of the protein surface residues. Similarly, existence of double Gaussian parameterization of self-vHf^{48, 49} can correspond to two regions with two different diffusion coefficients. Although, trapping effect is not associated here, some of the ions still can undergo the effect of weak interactions, leading differences in mobility.

In order to probe dynamic heterogeneity⁴⁰⁻⁴⁵ microscopically, we follow motions of the individual ions in all four cases. We generate their MSD plots for different Brownian trajectories. In Fig.4.6.(a), we show MSD of different Zn²⁺ ions in presence of ubiquitin. We observe that $\delta r_{Zn,(i)}^2(t)$ for $i = 1$ (first Zn²⁺ ion) shows different slopes in different time intervals, like between 0-50 ns, and 100-150 ns. But in case of $i = 3$, MSD exhibits overall linear dependence on t . Thus the motions of different Zn²⁺ ions are different. Next we consider MSD of $i = 1$ Zn²⁺ ion in different trajectories. We observe that $\delta r_{Zn,(1)}^2(t)$ (Fig.4.6.(b)) shows linear dependence of MSD on t with locally varying slopes in one trajectory, while exhibiting complete

linear dependence on t in another trajectory. However, this kind of heterogeneity is not observed in MSD plots of the ions in presence of mutated ubiquitin (Fig.4.6. (c)).

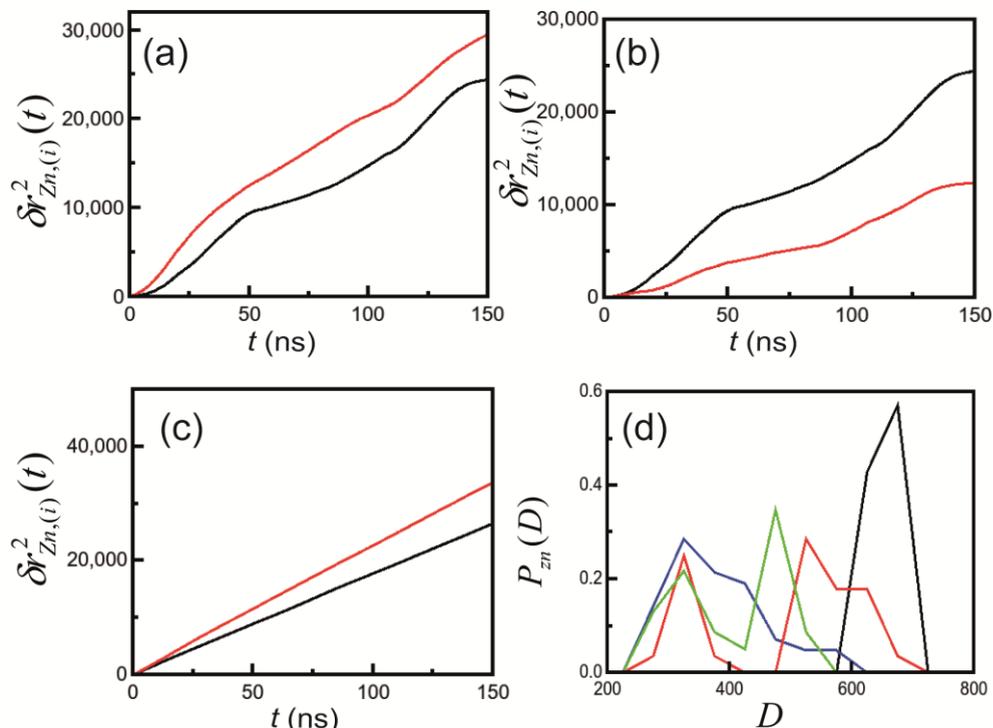


Fig.4.6.(a) MSD, $\delta r_{Zn(i)}^2(t)$ vs t plots of different Zn^{2+} ions for unperturbed ubiquitin in one particular trajectory and (b) $\delta r_{Zn(1)}^2(t)$ vs t plots in different trajectories indicate dynamic heterogeneity. (c) MSD, $\delta r_{Zn(i)}^2(t)$ vs t plots of different Zn^{2+} ions for doubly mutated ubiquitin in one particular trajectory indicating no such strong signature of heterogeneous diffusion. (d) Distributions of diffusion coefficients of the ions ($P_{Zn}(D)$) in three different cases; black: free case, red: in presence of two mutation, green: in presence of one mutation and blue: for nonmutated structure of ubiquitin. The scaling of x axis is represented as $D \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$.

We calculate slopes in MSD plots for each ion in different trajectories and time windows to estimate the diffusion coefficients. For example, for the first Zn^{2+} ion ($i = 1$ in Fig.4.6.(a)), we compute slopes of MSD for two different time intervals (0-50 ns, and 100-150 ns) to obtain the diffusion coefficients in those time intervals. From the slopes of different regions of MSD we obtain the distributions of the

diffusion coefficients, denoted by $P_{Zn}(D)$ (Fig.4.6. (d)). We observe that in free case, the distribution $P_{Zn}^{(f)}(D)$ is unimodal which indicates a single value of the diffusion coefficient. In presence of surface residues of ubiquitin, $P_{Zn}^{(0)}(D)$ shows broad width, signifying wide range of diffusion coefficients. For the single mutation, $P_{Zn}^{(1)}(D)$ and for the two mutation case, $P_{Zn}^{(2)}(D)$ exhibit bimodal distribution with two distinct diffusion coefficients. Mean diffusion coefficients obtained from the distributions for Zn^{2+} are comparable to those estimated from the MSD data in Fig.4.5.

4.3.4. Underlying mechanisms of heterogeneous diffusion:

The existence of decaying tail in self-vHf along with the broad range of diffusivity arises due to strong caging of the ions in the vicinity of trapping residues. On the other hand, double Gaussian self-vHf along with two distinct value of diffusivity can be attributed to the weak transient localization of some of the ions. We relate heterogeneous dynamics to the underlying interaction between protein residues and ions. We generate the density plots of the ions around the known surface exposed acidic residues of the ubiquitin ($\rho_{Zn}(R)$) over the simulated trajectories. Fig.4.7.(a) shows that in unperturbed state of ubiquitin, $\rho_{Zn}^{(0)}(R)$ has a strong peak which indicates strong localization of the ions around the residues over sufficiently long amount of time. Whereas, for one mutation, the peak height of $\rho_{Zn}^{(1)}(R)$ is comparatively lower than that in native structure. Similarly, $\rho_{Zn}^{(2)}(R)$ in two mutations has lowest magnitude. This signifies that Zn^{2+} experiences less electrostatic interaction upon mutation, causing weaker localization of the ions around the residues with increasing mutations.

The connections between the dynamics of the ions and their interactions with the binding residues are observed in distance variation ($R(t)$) for the Zn^{2+} ions from the C_α atom of such residues over time t in each equilibrated trajectory. Some of the representative cases for one such residue, E16 in all three cases are shown in Figs.4.7.(b)-(f). We observe that, in case of unperturbed ubiquitin, the fluctuations in $R^{(0)}(t)$ (Fig.4.7.(b)) for the first Zn^{2+} ($i = 1$) remain negligible over the entire time, indicating that the ion is trapped. The third Zn^{2+} ($i = 3$) undergoes large fluctuations along with higher magnitude. This indicates that the third ion is not trapped. The difference in strength of trapping for the same Zn^{2+} ($i = 1$) ion is also observed in two different trajectories as indicated in Fig.4.7.(c). In case of one mutation, $R^{(1)}(t)$ in Fig.4.7.(d) shows that the first Zn^{2+} comes in proximity of E16, but the fluctuations are larger than the no mutation case. The third ion is not trapped here as well. Similarly, Fig.4.7.(e) shows that for double mutations, fluctuations in $R^{(2)}(t)$ for the first ion are larger in comparison with two previous cases, while third ion behaves identically. Trajectory dependent variations in fluctuations of distance for the same Zn^{2+} ($i = 1$) ion is also observed (Fig.4.7.(f)) for double mutation. Although the trajectory dependent difference in distance variation is less prominent than that is observed for the unperturbed ubiquitin. These data indicate that for all three cases, Zn^{2+} ions possess two distinct type of mobility; some are getting closer to protein surface, while some are moving freely. However, as we mutate the residues; the strength of trapping is decreased; this is manifested in the corresponding diffusive dynamics of the ions.

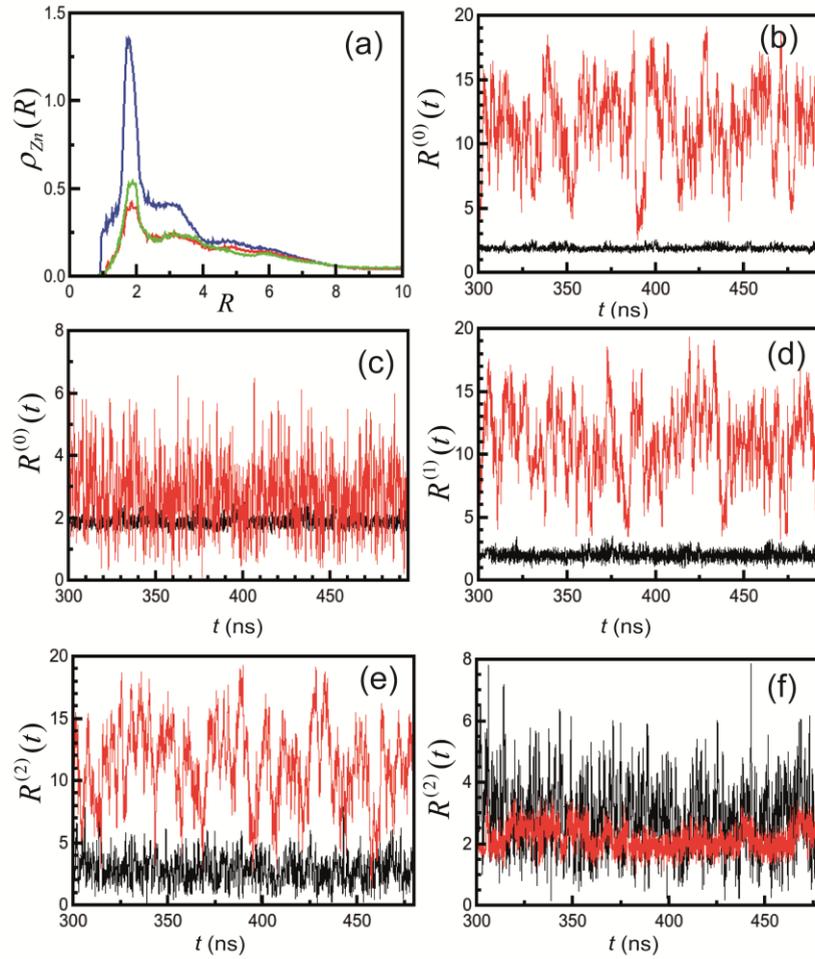


Fig.4.7.(a) $\rho_{Mg}(R)$ plots indicating strong localization Zn^{2+} in unperturbed ubiquitin (blue), strength of trapping is gradually reduced in mutations, (green) one mutation and (red) two mutations. (b) $R^{(0)}(t)$ between Zn^{2+} and E16 as a function of time t ; 1st Zn^{2+} (black) shows trapping, 2nd Zn^{2+} (red) shows no such signature of trapping. (c) $R^{(0)}(t)$ between 1st Zn^{2+} and E16, showing variation in distance in different trajectories. (d) $R^{(1)}(t)$ between Zn^{2+} and E16 as a function of time t showing decreasing in trapping propensity of the 1st ion and (d) $R^{(2)}(t)$ indicates significant decrease in trapping of the 1st Zn^{2+} around E16. (f) Variation in $R^{(2)}(t)$ in two different trajectories.

4.4. Discussions:

Our studies suggest that metal ion binding sites over a protein surface can be identified from dynamical information of the ions. We use this to predict metal ion

binding propensity of a bacterial protein, namely STY3178⁵⁰⁻⁵³ of *Salmonella* Typhi, which infects typhoid in humans. This protein belongs to a class of uncharacterized protein called yfdX that occurs in many other virulent bacteria⁵⁰⁻⁵³. Recent studies⁵⁰⁻⁵³ on STY3178 suggests that the protein may have connection to drug response of the bacteria. The model structure of the protein is shown in Fig.4.8. The protein has a net negative charge suggesting that the protein might interact with positively charged ligands. We consider dynamics of the most abundant metal ions, Mg²⁺ and Ca²⁺ with ~ 20 mM concentration in presence of STY3178 in cellular environment. The dynamic quantities are calculated by averaging over five different trajectories for both the ions.

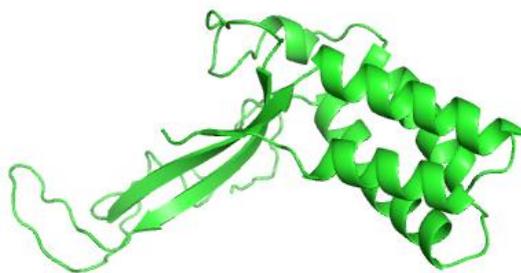


Fig.4.8. Model Structure of STY3178.

We find that the diffusion profiles of Mg²⁺ ions are similar to those shown by the Zn²⁺ in presence of mutated structures of ubiquitin. The $\ln G_{s,Mg}(\Delta r, \Delta t)$ vs Δr plots (Figs.4.9.(a)-(b)) represent that the self-vHf of Mg²⁺ exhibit double Gaussian dependence for $\Delta t = 0.5, 1.0$ ns. Similarly, the self-vHf for Ca²⁺, $G_{s,Ca}(\Delta r, \Delta t)$ (Figs.4.9.(c)-(d)) indicate that the double Gaussian profile persists at $\Delta t = 0.5$ and 1.0 ns. The normal Fickian diffusive nature for both the ions are retained at larger $\Delta t = 10.0$ ns for larger Δr . The corresponding fitting parameters (R^2, χ^2) for two differently mobile domains at different time intervals are tabulated in Table.4.2. This

indicates that although both the metal ions come in proximity of surface exposed residues of STY3178, they do not get trapped there.

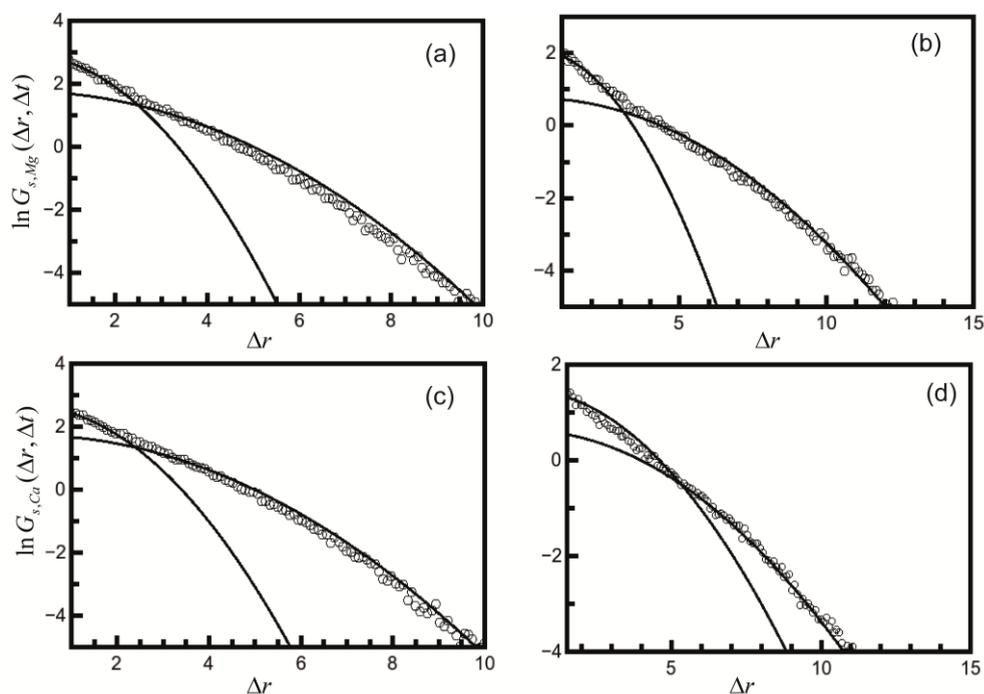


Fig.4.9.(a) $\ln G_{s,Mg}(\Delta r, \Delta t)$ vs Δr plots at different time intervals; Deviation from normal diffusive profile of the ions in vicinity of STY3178 and the double Gaussian dependence is observed. $\Delta t = 0.5$ ns and (b) $\Delta t = 1.0$ ns, (c) Double Gaussian nature of $G_{s,Ca}(\Delta r, \Delta t)$ at $\Delta t = 0.5$ ns, (d) at $\Delta t = 1.0$ ns.

Table.4.2. Fitting parameters for $\ln G_{s,Mg}(\Delta r, \Delta t)$ and $\ln G_{s,Ca}(\Delta r, \Delta t)$ vs Δr plots at different time intervals.

System	Δt (ns)	Fitting	Range (Δr)	R^2	χ^2
Mg ²⁺ in presence of STY3178	0.5	square	1.00-2.00	0.97	0.002
Mg ²⁺ in presence of STY3178	0.5	square	2.00-8.00	0.99	0.008
Mg ²⁺ in presence of STY3178	1.0	square	1.00-2.50	0.95	0.004
Mg ²⁺ in presence of STY3178	1.0	square	2.70-12.00	0.99	0.008

Ca ²⁺ in presence of STY3178	0.5	square	1.00-2.00	0.94	0.002
Ca ²⁺ in presence of STY3178	0.5	square	2.15-10.00	0.99	0.008
Ca ²⁺ in presence of STY3178	1.0	square	1.00-5.00	0.98	0.008
Ca ²⁺ in presence of STY3178	1.0	square	5.00-11.00	0.99	0.007

4.5. Conclusions:

We observe that the strong electrostatic interactions exerted by the residues over a protein surface can trap metal ions for sufficiently large time. This leads to anomalous dynamic behavior of the trapped ions. Earlier studies also show that anomaly in dynamics may arise due to localization of different particles at different spatial points^{25, 35, 38, 46}. The dynamical heterogeneity of a physiochemical system has been reported earlier in bio-molecular systems^{40, 54}. However, we for the first time to the best of our knowledge show here an alternative approach based on scattering measurements²⁵⁻²⁸ for finding metal ion binding sites over a protein surface. Due to specific location of the binding pockets, the ligand attached protein may act as directed template for heterogeneous structure at nanometer length scale⁵⁵.

4.6. References:

1. X. Du, Y. Li, Y. L. Xia, S. M. Ai, J. Liang, P. Sang, X. L. Ji and S. Q. Liu, *Int. J. Mol. Sci.*, 2016, **17**.
2. S. Z. Grinter and X. Zou, *Molecules*, 2014, **19**, 10150-10176.
3. T. Govindarajan and R. Shandas, *Polymers.*, 2014, **6**, 2309-2331.
4. L. Bocquet and E. Charlaix, *Chem. Soc. Rev.*, 2010, **39**, 1073-1095.

5. H. Assender, V. Bliznyuk and K. Porfyraakis, *Science*, 2002, **297**, 973-976.
6. K. H. Sippel and F. A. Quioco, *Protein Sci.*, 2015, **24**, 1040-1046.
7. H. P. Erickson, *Biol Proced Online*, 2009, **11**, 32-51.
8. I. Gitlin, J. D. Carbeck and G. M. Whitesides, *Angew. Chem. Int. Ed.*, 2006, **45**, 3022-3060.
9. M. Kastantin, R. Walder and D. K. Schwartz, *Langmuir.*, 2012, **28**, 12443-12456.
10. S. Khodadadi and A. P. Sokolov, *Soft Matter*, 2015, **11**, 4984-4998.
11. Y. Matsunaga, C. B. Li and T. Komatsuzaki, *Phys. Rev. E.*, 2010, **82**.
12. Q. Y. Tang, Y. Y. Zhang, J. Wang, W. Wang and D. R. Chialvo, *Phys. Rev. Lett.*, 2017, **118**, 088102.
13. L. Q. Yang, P. Sang, Y. Tao, Y. X. Fu, K. Q. Zhang, Y. H. Xie and S. Q. Liu, *J. Biomol. Struct. Dyn.*, 2014, **32**, 372-393.
14. H. Zheng, K. B. Handing, M. D. Zimmerman, I. G. Shabalin, S. C. Almo and W. Minor, *Expert Opin. Drug Discov.*, 2015, **10**, 975-989.
15. Y. Li and C. Kang, *Molecules*, 2017, **22**.
16. M. R. Jensen, G. Petersen, C. Lauritzen, J. Pedersen and J. J. Led, *Biochemistry*, 2005, **44**, 11014-11023.
17. W. Shi, M. Punta, J. Bohon, J. M. Sauder, R. D'Mello, M. Sullivan, J. Toomey, D. Abel, M. Lippi, A. Passerini, P. Frasconi, S. K. Burley, B. Rost and M. R. Chance, *Genome Res.*, 2011, **21**, 898-907.
18. A. Passerini, M. Lippi and P. Frasconi, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 2012, **9**, 203-213.
19. G. B. Akcapinar and O. U. Sezerman, *Biosci. Rep.*, 2017, **37**.
20. J. Anastassopoulou and T. Theophanides, *The role of metal ions in biological systems and medicine*, Kluwer Academic Publishers, D.P. Kessissoglou (ed.), Bioinorganic Chemistry edn., 1995.
21. P. W.-S. Cláudio M. Gomes, *Protein Folding and Metal Ions: Mechanisms, Biology and Disease*, Taylor & Francis Group, 2011.
22. A. Camara-Artigas, M. Plaza-Garrido, S. Martinez-Rodriguez and J. Bacarizo, *Acta Crystallogr F*, 2016, **72**, 29-35.

23. F. Arnesano, B. D. Belviso, R. Caliandro, G. Falini, S. Fermani, G. Natile and D. Siliqi, *Chemistry*, 2011, **17**, 1569-1578.
24. J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids, Fourth Edition*, Elsevier, 2013.
25. J. Kim, C. Kim and B. J. Sung, *Phys. Rev. Lett.*, 2013, **110**, 047801.
26. L. Hong, D. C. Glass, J. D. Nickels, S. Perticaroli, Z. Yi, M. Tyagi, H. O'Neill, Q. Zhang, A. P. Sokolov and J. C. Smith, *Phys. Rev. Lett.*, 2013, **110**, 028104.
27. B. M. Braun and H. Weingartner, *J. Phys. Chem.*, 1988, **92**, 1342-1346.
28. C. Ridou, M. Rousseau, B. Pernot and J. Bouillot, *J. Phys. C: Solid State Phys.*, 1986, **19**, 4847.
29. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids, 2nd edition*, Oxford Science Publications, New York, 1989.
30. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701-1718.
31. C. Oostenbrink, A. Villa, A. E. Mark and W. F. van Gunsteren, *J. Comput. Chem.*, 2004, **25**, 1656-1676.
32. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
33. T. O. Skinner, S. K. Schnyder, D. G. Aarts, J. Horbach and R. P. Dullens, *Phys. Rev. Lett.*, 2013, **111**, 128301.
34. O. B. Okan, A. R. Atilgan and C. Atilgan, *Biophys. J.*, 2009, **97**, 2080-2088.
35. C. Yildirim, J. Y. Raty and M. Micoulaut, *J. Chem. Phys.*, 2016, **144**, 224503.
36. S. K. Ghosh, A. G. Cherstvy, D. S. Grebenkov and R. Metzler, *New J. Phys.*, 2016, **18**.
37. E. Flenner and G. Szamel, *Phys. Rev. E.*, 2005, **72**.
38. B. Shen, Z. Y. Wang, F. Dong, Y. R. Guo, R. J. Zhang, Y. X. Zheng, S. Y. Wang, C. Z. Wang, K. M. Ho and L. Y. Chen, *J. Phys. Chem. B.*, 2015, **119**, 14945-14951.
39. S. Kariuki and H. D. Dewald, *Electroanal.*, 1996, **8**, 307-313.
40. B. Wang, J. Kuo, S. C. Bae and S. Granick, *Nat. Mater.*, 2012, **11**, 481-485.
41. M. V. Chubynsky and G. W. Slater, *Phys. Rev. Lett.*, 2014, **113**, 098302.
42. V. A. Chechkin, F. Seno, R. Metzler and I. M. Sokolov, *Phys. Rev. X*, 2017, **7**, 021002.

43. R. Metzler, J. H. Jeon, A. G. Cherstvy and E. Barkai, *Phys. Chem. Chem. Phys.*, 2014, **16**, 24128-24164.
44. P. Chaudhuri, L. Berthier and W. Kob, *Phys. Rev. Lett.*, 2007, **99**, 060604.
45. R. Metzler, *Biophys. J.*, 2017, **112**, 413-415.
46. S. Acharya, U. K. Nandi and S. Maitra Bhattacharyya, *J. Chem. Phys.*, 2017, **146**, 134504.
47. S. Dutta and J. Chakrabarti, *Europhys. Lett.*, 2016, **116**.
48. S. Dutta and J. Chakrabarti, *Soft Matter*, 2018, **14**, 4477-4482.
49. M. S. Shell, P. G. Debenedetti and F. H. Stillinger, *J. Phys. Condens. Matter*, 2005, **17**, S4035-S4046.
50. P. Saha, C. Manna, J. Chakrabarti and M. Ghosh, *Sci. Rep.*, 2016, **6**.
51. P. Saha, C. Manna, S. Das and M. Ghosh, *Sci. Rep.*, 2016, **6**.
52. P. Saha, S. Sikdar, J. Chakrabarti and M. Ghosh, *RSC Adv.*, 2016, **6**, 91256-91264.
53. P. Saha, S. Sikdar, C. Manna, J. Chakrabarti and M. Ghosh, *RSC Adv.*, 2017, **7**, 6209-6214.
54. S. H. Chong and S. Ham, *J. Phys. Chem. Lett.*, 2016, **7**, 3967-3972.
55. S. Bhaskar and S. Lim, *Npg Asia Mater*, 2017, **9**.

CHAPTER 5

Microscopic studies on flagellin-receptor complexes

5.1. Introduction:

In the previous two chapters, we address dynamics of fluctuations of different degrees of freedom to understand protein functionality. In this chapter, we focus on static aspects of conformational fluctuations of a bacterial protein. Our aim is to provide insight to conformational stability and functions of the protein. We consider a bacterial protein; flagellin which forms the major subunit of flagellum, an organelle of bacteria involved in several biological functions, like protein export, biofilm formation as well as adhesion and invasion¹⁻⁶ to the host cells. When bacteria invade host cells, specific host cell surface receptor protein recognizes flagellin⁷⁻¹⁰ to trigger a chain of biochemical reactions within the host cell and the signal is transmitted from one cell to other. The pathway of these intracellular reactions is termed signal transduction cascades¹¹⁻¹⁸, which finally leads to immunological responses¹⁹. The interaction between host cell receptor and flagellin, although of vital importance in immune-response of the host cells, is poorly understood till date.

Depending on functions and organisms, flagellin proteins are named differently¹. We consider in particular bacterial species invading human guts. Flagellin in bacteria *Salmonella typhimurium* is known as fliC. In case of *Vibrio cholerae* there are five distinct flagellin (flaA to flaE)²⁰ among which flaC and flaD are most abundant²¹. Flagellin has a tendency to polymerize into filament, which evades crystallization. The only full length crystal structure²²⁻²⁴ of monomeric fliC is reported (1UCU.PDB), where terminal residues (D0 region) of the protein are

chopped off. The D0 region has been modeled using FEX-PLOR program^{22, 23}. According to this structure²²⁻²⁴ the shape of the protein is similar to Greek upper case letter Γ with four linearly connected domains (D0-D4) as shown in Fig.5.1.(a). Out of these domains, D0 and D1 are conserved across organisms, whereas D2 and D3 domains show diversity among organisms and are known as hyper-variable (HPV) domain^{1, 4}. Among the conserved domains, D1 region participates in different molecular interactions; while D0 is disordered and very little is known about this region.

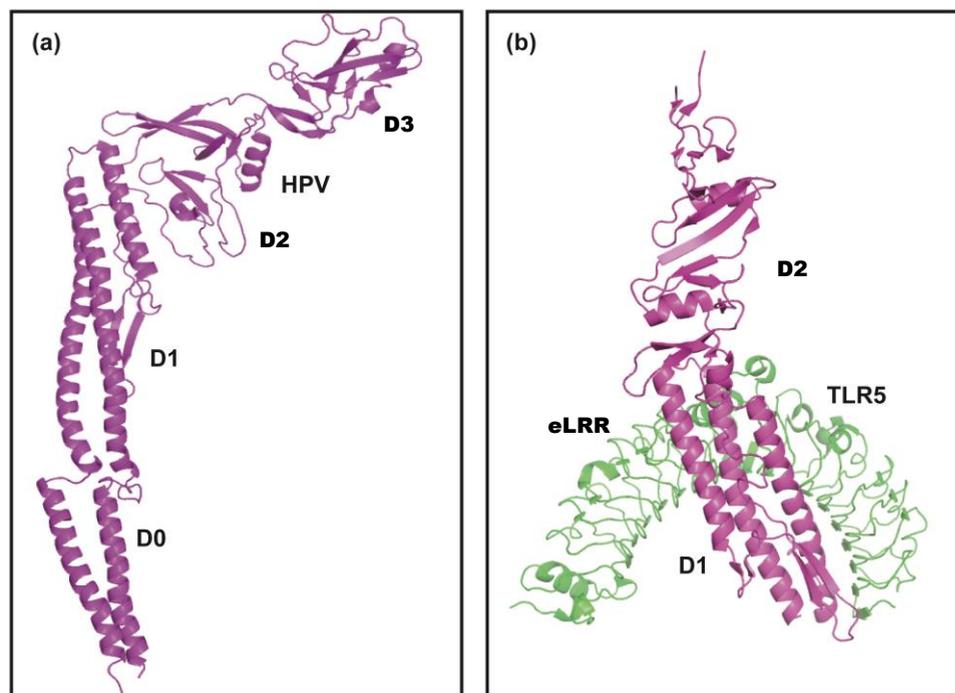


Fig.5.1.(a) Different domains of flagellin as indicated in cartoon representation of crystal structure of monomeric fliC. (b) Crystal structure of fliC-TLR5 complex.

Toll like receptor, TLR5^{25, 26} is the host cell surface receptor protein that interacts with flagellin and forms a dimeric flagellin-TLR5 complex. The membrane bound protein TLR5 is characterized²⁷ by a horse shoe shaped Leucine rich repeat extracellular domain (eLRR) and a conserved intracellular TIR region. The crystal structure (3V47.PDB) has been reported for eLRR region of TLR5 bound¹⁶ to D1

domain of fliC, along with the intact D2 region. The detection of pathogen associated molecular pattern of flagellin is done by the eLRR domain of TLR5 (Fig.5.1.(b)) by interacting with D1 region. The signal²⁸ is activated by TIR domain that releases inflammatory responses. Recent immunological studies^{29, 30} on fliC reveal that deletion of D0 significantly reduces TLR5 activation in immunological response without affecting binding between flagellin and TLR5. The structural data of TLR5-flaD is yet to be solved. However, immunological studies establish interaction between flaD and TLR5³¹⁻³³. Although TLR5 is known¹⁶ to interact with the conserved domain of flagellin of different bacteria, discrepancy is reported in associated signaling pathway for fliC-TLR5 and flaD-TLR5 systems. In recent immunological experiment³⁴ it is observed that in contrast to fliC-TLR5 complex, flaD-TLR5 induced activation is dependent on co-receptors, G protein coupled receptors (GPCRs)^{35, 36} in presence of lipid raft.

The complete understanding of flagellin-TLR5 interaction for different bacterial organisms is crucial for therapeutic development. This leads us to perform in-silico model studies on flagellin, receptor protein and their complexes. We address the problem in two steps: First, we focus on relative stability of the D0 domain of free flagellin in two different media, water and membrane bilayer. Then we address relative stability of flagellin in presence of TLR5 with respect to the free states and correlate the observations to coreceptor bindings.

The relative stability of protein conformations can be understood from conformational thermodynamics^{37, 38}, where the changes in thermodynamics free energy and entropy of a protein in a given conformation with respect to a reference conformation are computed from fluctuations of dihedral angles in the two states over simulated trajectories. If the changes in free energy and entropy of protein dihedrals in

a given conformation are positive in comparison to the reference conformation, then the protein is considered to be destabilized and disordered in the given conformation. In earlier studies^{37, 38}, it is reported that the destabilized and disordered residues of a protein participate in further binding activity in order to reduce their free energy. In our study, we apply conformational thermodynamics to understand stability of flagellin both in free and TLR5 bound states. Only the conserved domains of flagellin are considered here.

We have used experimentally available structure of fliC²²⁻²⁴ as initial structure for the simulation. In absence of structural information, we model the structure of conserved domains of flaD (Fig.5.2.(a)) using available crystal structure of fliC as a template and perform similar calculations. We perform 1.5 μ s long all-atom MD simulations on two systems: (1) W system, where the entire conserved domains D0 and D1 are kept in water (Fig.5.3.(a)) and (2) L system, where D0 region is immersed in a modeled dipalmitoyl phosphatidylcholine membrane bilayer (DPPC) while D1 is kept in aqueous medium as shown in Fig.5.3.(b). We generate histogram of dihedrals from equilibrated trajectories and compute changes in thermodynamics of conformational changes in L case compared to W case. We find that D0 of both fliC and flaD are thermodynamically and structurally more stable in bilayer than water. There are destabilized and disordered residues of D1 in both fliC and flaD. In case of fliC, these residues belong to the binding interface with TLR5 receptor. The nature of those destabilized residues is different in fliC and flaD. This suggests differences in binding mode by those two different bacterial flagellin proteins to TLR5.

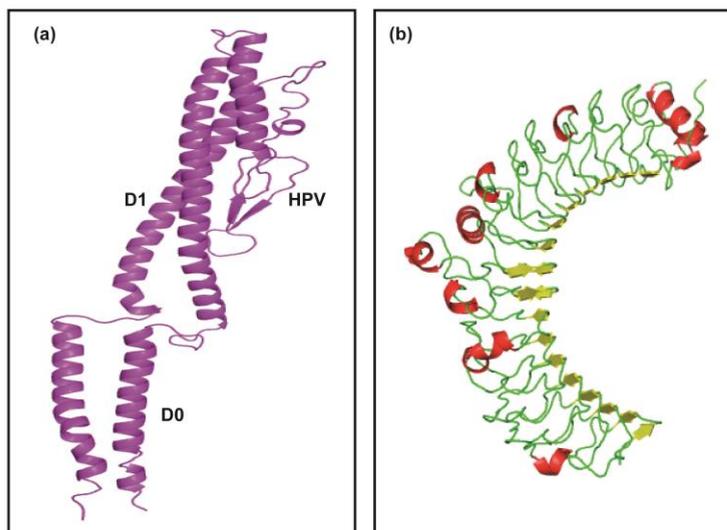


Fig.5.2.(a). Model structure of flaD indicating structured conserved domains, HPV regions are unstructured. **(b)** Crystal structure of extracellular part of TLR5 (3V44. PDB).

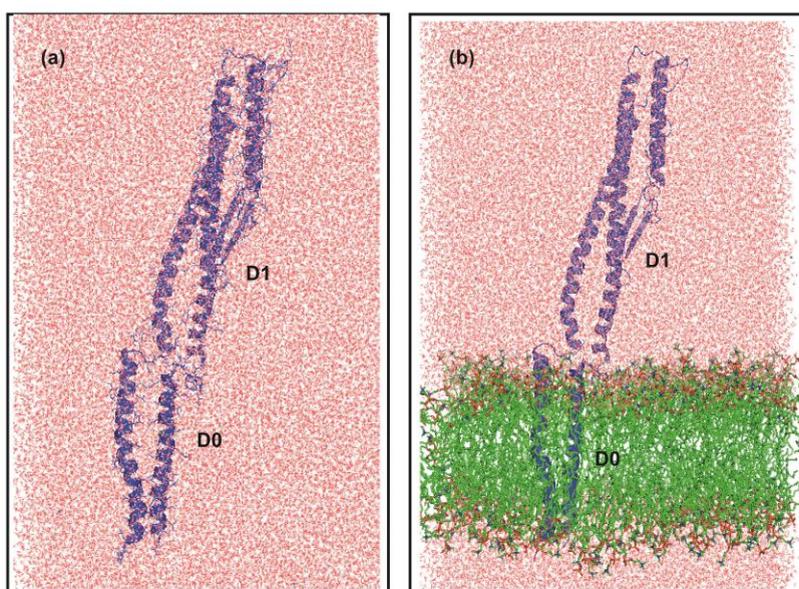


Fig.5.3. Snapshots of fliC simulated in different medium; **(a)** D0 and D1 are in water and **(b)** D0 in bilayer, D1 in water.

Next, we consider thermodynamics stability of fliC and flaD in presence of TLR5. We use the reported crystal structure of D1 domain of fliC bound to extracellular domain of TLR5¹⁶ as initial structure for simulation. Since no structural information is available for flaD-TLR5 complex, we propose a probable structure of flaD bound to

extracellular domain of TLR5. In both cases D0 is kept in DPPC and D1 along with extracellular domain of TLR5 are kept in water (Fig.5.4.(a)). We perform 1.0 μ s long all-atom MD simulations for both fliC-TLR5 and flaD-TLR5. We also consider free extracellular domain of TLR5 in water from the 3V44.PDB as the initial structure for simulation (Fig.5.4.(b)). We generate distributions of dihedrals of TLR5 bound flagellin over the equilibrated trajectories and compute changes in thermodynamics free energy and entropy of flagellin with respect to its free state, where D0 is kept in bilayer and D1 remains in water. We find that the destabilized and disordered residues of fliC in presence of TLR5 are different than those in flaD. Thus, the residues of fliC and flaD in TLR5 bound states which might participate in further binding processes are not the same, although the binding between flagellin and TLR5 takes place in D1 domain of flagellin in both the cases.

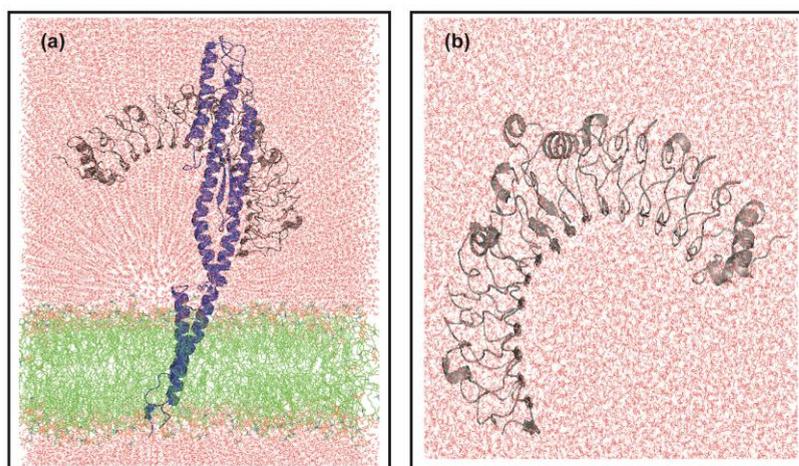


Fig.5.4.(a) Snapshot of simulated structure of fliC-TLR5; D0 is kept in bilayer and D1 along with TLR5 are immersed in water. (b) Free TLR5 simulated in water.

The rest of the chapter is organized as follows: The methods are explained in details in section 5.2. Section 5.3. contains all the results elaborately and the implications of our results. The conclusion is in section 5.4.

5.2. Methods:

The details of different systems, the protocol for MD simulation, docking for coreceptors are explained in this section. We also elaborate calculation of conformational thermodynamics.

5.2.1. System preparations:

We consider different systems to simulate in water and in bilayer. (I). Free fliC, (II). Free flaD, (III). Free TLR5, (IV). fliC–TLR5 complex, (V). flaD-TLR5 complex and (VI) DPPC bilayer.

(I) fliC:

We consider the crystal structure of monomeric fliC (1UCU.PDB)²²⁻²⁴ as an initial conformation in our calculation (Fig.5.1.(a)). The region from Q2 to Serine; S32, along with Alanine; A459 to R494 form helical D0 domain. The spoke region consists of S32 to A44 and E454 to A459 residues, which is a loop that connects D0 with D1. D1 domain is rod shaped, containing α helices and one unique β hairpin motif. A44 to K179 form N terminal D1, followed by a C terminal side (Asparagine; N406 to E454). The residue from K179 to A401 makes hyper-variable domains, containing randomly oriented β hairpin motif, β sheet and helices. We exclude hyper-variable regions of fliC in order to make our simulation less expansive and focus on the conserved domains only.

(II) flaD:

We generate model structure of flaD from Homology modeling, using experimentally determined structure of fliC as a template³⁹. The primary sequence of flaD⁴⁰ in FASTA format has been submitted to Protein Model Portal Database⁴¹. We obtain modeled structure from four different servers including ImHCE, SWISS-

MODEL⁴², IntFOLD2⁴³ and Phyre2⁴⁴. All these structures are minimized for 100000 steps using NAMD⁴⁵ as indicated in Fig.5.5. and the energetically favorable one is chosen for simulation. Final structure of flaD lacks of compact alpha helices in hyper-variable domain and mostly remains structured for the conserved region as depicted in Fig.5.2.(a). In analogy with fliC, in case of flaD we also primarily focus on the structured conserved domains. D0 domain mainly comprises of helices and the residues from M2 to S34, K343 to G377 belong to the domain. Whereas G35 to A44 along with K337 to A342 makes flexible spoke region. D1 consists of A45 to A179 together with G285 to K337, forming helices and β sheets.

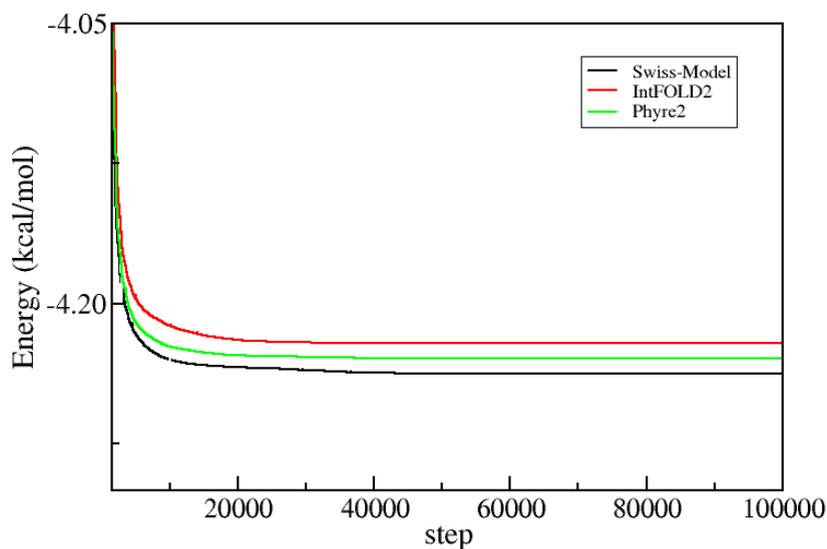


Fig.5.5. Energy minimization for the various modeled structures of flaD obtained from different modeling software.

(III) TLR5:

The full length crystal structure of TLR5 containing both the extracellular and the intracellular domains is not reported yet. The crystal structure (3V44.PDB) reveals that extracellular domain of TLR5 consists of Leucine rich repeat domain. Smooth curved β sheets of TLR5 (Fig.5.2.(b)) include residues from E64 to P464. Two anti-

parallel β strands and thirteen parallel β strands form concave surface, whereas helices form the irregular surface of the receptor.

(IV) fliC-TLR5 complex:

We consider crystal structure of fliC-TLR5 (3V47.PDB) as indicated in Fig.5.1.(b). The structure reveals that D1 domain interact with TLR5, forming primary interface. Interaction is mainly hydrophobic, salt bridges and H bonded are also observed.

(V) flaD-TLR5 complex:

As no crystal structure is available for flaD in presence of TLR5, we dock D1 domain of our model flaD to extracellular crystal structure of TLR5 using HADDOCK⁴⁶. We find the destabilized and disordered residues of D1 domain of flaD in bilayer with respect to its water embedded conformation. We chose these residues as active sites for TLR5 binding. For TLR5, we choose the known fliC binding residues as probable active sites. After rigid body energy minimization^{46, 47}, MD based refinement process by HADDOCK we get modeled structures. The top cluster is chosen on basis of Z-score^{46, 47}. Details of docking result are tabulated in Table.5.1.

Table.5.1. Docking study on flaD-TLR5 system.

HADDOCK score	-138.0 +/- 6.8
Cluster size	23
RMSD from the overall lowest-energy structure	0.7 +/- 0.4
Van der Waals energy	-81.6 +/- 8.3
Electrostatic energy	-420.4 +/- 54.0
Desolvation energy	26.6 +/- 4.1
Restraints violation energy	10.7 +/- 17.31
Buried Surface Area	2697.2 +/- 47.7
Z-Score	-2.9

(VI) Preparation of bilayer:

We keep pre-equilibrated model structure of DPPC bilayer, obtained from GROMACS⁴⁸ topology in XY plane and flagellin has been immersed in bilayer along z axis. We use inflateGRO methodology⁴⁹ to pack lipid molecules around the embedded protein. We apply a very strong position restraining force on protein heavy atoms in order to keep it fixed during energy minimization. Following standard protocol of GROMACS⁴⁸ we add topology of lipids⁵⁰ to the force-field documents. We use “Berger Lipids” in our simulation. The “Berger Lipids” parameters are somewhat of a hybrid between GROMOS atom types and OPLS partial charges. Since the long alkane chains are poorly represented by GROMOS bonded parameters, a Ryckaert-Bellemans dihedral potential is used with a scaling factor of 0.125 is being applied to Leenard-Jones 1-4 interactions.

5.2.2. MD simulations:

We prepare seven different systems to simulate for our analysis. (I). Free fliC, D0 and D1 are immersed in water, (II). Free fliC, D0 is immersed in DPPC, D1 remains in water, (III). Free flaD, both D0 and D1 are kept in water, (IV). Free flaD, D0 is embedded in DPPC, D1 is in water, (V). Free TLR5 simulated in water, (VI). D0 of fliC is kept in DPPC, whereas D1 along with TLR5 are placed within water and (VII). D1 of flaD with TLR5 are simulated in water, D0 remains in DPPC. Details of simulation protocols are explained in this section.

5.2.2.1. Simulation in water:

We use periodic boundary conditions, spc216 water model and GROMOS9353a6 force field for simulations in GROMACS package⁵¹. Electro-neutrality is maintained by adding mono-valent ions Na⁺ and Cl⁻. Long ranged columbic interactions are

considered using PME approach⁵² and cut of range for van der Waals interactions is 1nm. LINCS algorithm⁵³ is used to constraint the bonds and leap-frog integration was used to perform simulation. We choose the entire conserved domain of fliC and immerse it in water. We then perform an energy minimization for first 10,000 steps using steepest descent algorithm, and then run 1.5 μ s long isothermal isobaric (NPT) simulations with 1 fs time step to capture the protein dynamics. We use the similar protocol to simulate free flaD in aqueous environment. We ensure equilibration by RMSD plots as shown in Fig.5.6.(a). Calculations are performed over last 600 ns equilibrated trajectory files.

We also simulate free TLR5 in presence of aqueous environment (Fig.5.4.(b)). We run 1.0 μ s long simulations and RMSD plot is shown in Fig.5.6.(b). Calculations are performed over last 500 ns equilibrated trajectory files.

5.2.2.2. Simulations in presence of membrane:

For our simulation, we consider DPPC containing 392 number lipid molecules. We embed the D0 domain of fliC in bilayer, whereas D1 remains in water. Both the water and membrane along with the protein contain \sim 2,00,000 number of particles and are minimized for 10,000 steps in the absence of constraints. Following minimization, 1 ns of equilibration is carried out at a temperature of 323 K and pressure of 1 atm. We perform 1.5 μ s simulation step using standard protocol for isothermal isobaric (NPT) simulation and GROMOS9353a6 force field. We follow the same process in order to simulate flaD in bilayer. RMSDs are shown in Fig.5.6.(c) and analysis is done from the equilibrated trajectory over last 600 ns for both the cases.

We consider the initial conformation of fliC-TLR5 complex from the available crystal structure. Next we immerse D0 domain of fliC in DPPC, whereas D1 and

TLR5 are kept in water and perform 1.0 μs simulation. We follow the same process in order to simulate flaD in presence of TLR5 in the vicinity of bilayer. RMSDs for the equilibrations are shown in Fig.5.6.(d) and analysis is done from the equilibrated trajectory over last 500 ns for both the complexes.

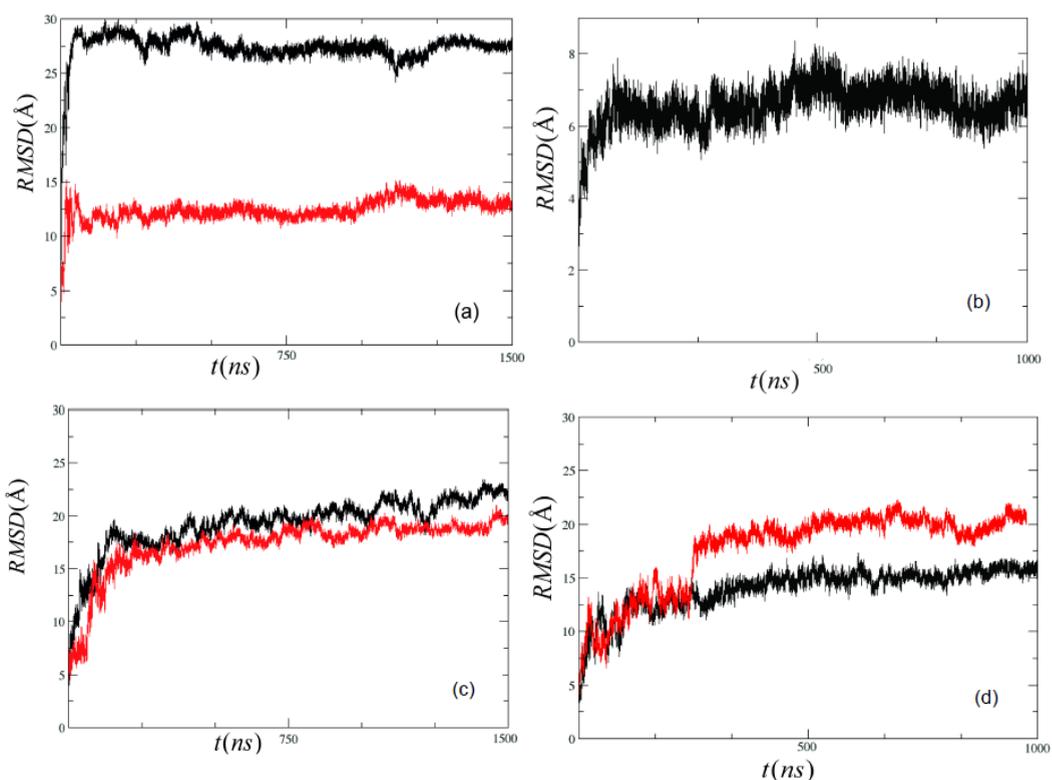


Fig.5.6.(a) RMSD of free flagellin as a function of simulation time, where both D0 and D1 are kept in aqueous medium; Black; fliC, Red; flaD and (b) RMSD of free TLR5 in water. (c) RMSD of free flagellin while D0 is kept in bilayer and D1 in water; color code is same as previous panel. (d) RMSD of fliC-TLR5 (black) and flaD-TLR5 complexes (red), D0 remains in bilayer, D1 along with TLR5 are kept in water.

5.2.3. Conformational thermodynamics and HBM:

Conformational thermodynamics changes for proteins between two different conformations are estimated properly using the Histogram Based Method (HBM)^{37, 38}.

Equilibrium conformational changes in free energy is defined by $\Delta G(\varphi) =$

$$-K_B T \ln \left[\frac{H_L(\varphi)}{H_W(\varphi)} \right] \text{ where } H_L(\varphi) \text{ and } H_W(\varphi) \text{ signify peak value of normalized}$$

probability distribution of protein dihedral φ in bilayer and water respectively, and K_B , the Boltzmann's Constant. Conformational entropy change associated with a particular dihedral φ at a temperature T is calculated using $T\Delta S(\varphi) = T(S_L(\varphi) - S_W(\varphi))$ where $S_L(\varphi)$ and $S_W(\varphi)$ can be obtained using Gibbs entropy formula $S(\varphi) = -K_B T \sum_i H_i(\varphi) \ln H_i(\varphi)$, sum is taken over histogram bins.

5.2.4. Docking of coreceptors with flagellin-TLR5 complexes:

We investigate the binding propensity between flagellin-TLR5 complexes with respect to different coreceptor proteins (A₂A and VPAC1), involved in downstream signaling for immune responses. The extracellular loop (eCL2) of A₂A^{54,55}, one of the coreceptor protein and extracellular domain (ecd) of VPAC1⁵⁶⁻⁵⁸, another coreceptor protein are reported to provide binding specificity to ligands. However, the crystal structure for extracellular loop of A₂A, is not available due to poor electron density. Similarly, the full length crystal structure of VPAC1 is not available till date. So, we consider the sequences for extracellular loop of A₂A and extracellular domain of VPAC1 and perform homology modeling using SWISS-MODEL⁴², IntFOLD2⁴³ and Phyre2⁴⁴. We minimized all the model structures as indicated in Figs.5.7.(a)-(b) and consider the one having lowest energy. Next, we dock them separately with fliC-TLR5 and flaD-TLR5 complexes using HADDOCK⁴⁶. The known binding hot spot residues of A₂A and VPAC1 are chosen as active sites for docking with flagellin-TLR5 complexes. The residues of flagellin showing significant amount of destabilization and disorder in presence of TLR5 are chosen as active sites to interact with these two coreceptors. Details of docking are shown in Table.5.2.

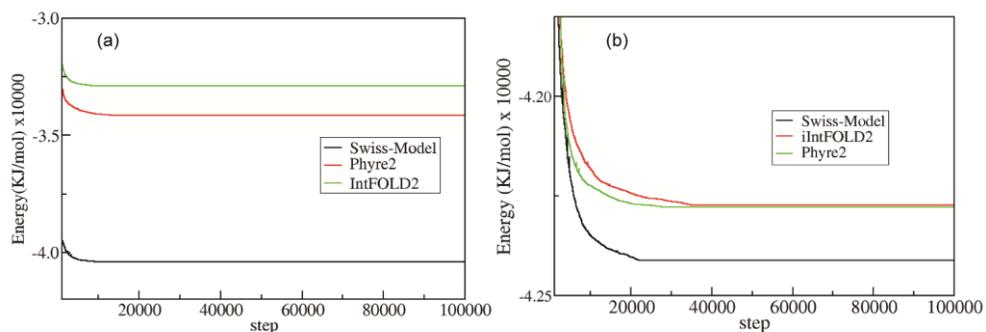


Fig.5.7. Energy minimization for the various modeled structures of extracellular part of GPCR proteins as obtained from different modeling software, (a) for A₂A and (b) that of VPAC1.

As extracellular parts of the coreceptors are known to provide ligand specificity, we study the coreceptors bindings with flagellin-TLR5 complexes by performing small simulations ~ 100 ns in aqueous medium for eight different systems; (I). D1 domain of fliC-TLR5 complex docked with ecL2 of A₂A, (II). Extracellular domain of VPC1 docked with D1 region of fliC-TLR5, (III). D1 of flaD-TLR5 docked with ecL2 of A₂A, (IV). Extracellular domain of VPC1 docked with D1 region of flaD-TLR5 complex, (V). D1 domain of fliC bound to TLR5, (VI). D1 region of flaD bound to TLR5, (VII). Extracellular loop2 of free A₂A and (VIII). Extracellular domain of free VPAC1.

Table.5.2. Docking studies on flagellin-TLR5 and GPCRs.

System	fliC-TLR5 -A ₂ A	fliC-TLR5- VPAC1	fliC-TLR5- A ₂ A	fliC-TLR5- VPAC1
HADDOCK score	-50.8 +/- 9.2	-61.2 +/- 13.8	-59.5 +/- 2.7	-62.4 +/- 6.8
Cluster size	25	10	8	12
RMSD from the overall lowest- energy	1.3 +/- 0.8	0.8 +/- 0.5	6.3 +/- 0.7	21.0 +/- 0.1

structure				
Van der Waals energy	-44.3 +/- 5.7	-51.1 +/- 7.7	-34.6 +/- 9.2	-53.5 +/- 3.5
Electrostatic energy	-36.2 +/- 13.2	-71.8 +/- 43.2	-145.4 +/- 47.6	-105.2 +/- 34.4
Desolvation energy	0.8 +/- 10.0	1.7 +/- 4.9	3.3 +/- 3.5	5.5 +/- 7.2
Restraints violation energy	0.1 +/- 0.13	25.7 +/- 14.04	9.3 +/- 16.05	66.0 +/- 26.70
Buried Surface Area	1186.8 +/- 152.3	1587.3 +/- 135.4	1259.5 +/- 220.1	1325.5 +/- 79.5
Z-Score	-2.4	-1.9	-2.0	-2.0

5.3. Results and discussions:

We address conformational stability of flagellin in different media, namely bilayer and water. Next, we show changes in conformational thermodynamics of flagellin-TLR5 complexes with respect to corresponding free states. These results are given in the following subsections.

5.3.1. Conformational stability of fliC:

The atomic model reported in the crystal structure of monomeric fliC²²⁻²⁴ shows that D0 domain of flagellin forms α coil structure, while D1 is rod shaped with three α helices and one unique β hairpin motif. We generate Ramachandran Plot⁵⁹ ($\varphi - \psi$) of the residues of D0 (Fig.5.8.(a)), over the average structure obtained from simulated trajectories. We observe that the helical domain of D0 gets unwound in aqueous medium. Around 37 % residues lose compact secondary structural conformation in water with respect to initial helical structure and majority of these unstructured

residues are hydrophobic in nature. On the other hand, in bilayer embedded conformation, alpha helices of D0 remains almost intact, causing structural loss for only ~ 21 % residues. That indicates bilayer induced structural stability to D0 domain. There is no noticeable secondary structural change for D1 domain in two systems (Fig.5.8.(b)).

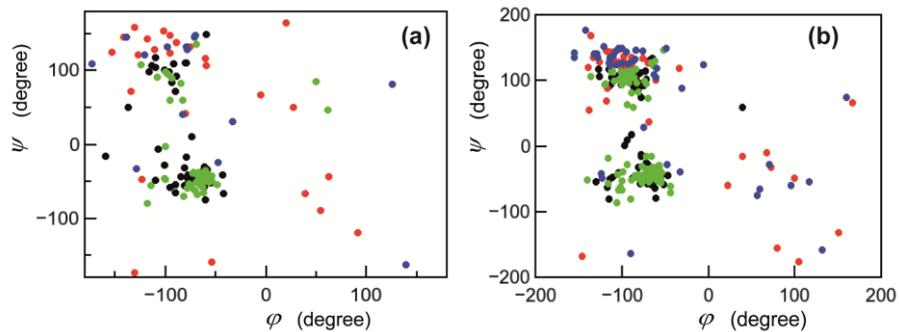


Fig.5.8. $\varphi - \psi$ correlation plot of residues of fliC for water and bilayer embedded conformations obtained from average structure of equilibrated trajectory, Black; residues that are structured in water, Red; the unstructured ones in water, while Green; structured residues in bilayer and Blue; unstructured residues in bilayer. (a) D0 region and (b) for D1 domain.

We compare the dihedral distributions of various residues of the entire conserved fliC over equilibrated trajectories for membrane and water embedded conformations. We observe that residues of D0 show prominent changes in backbone as well as side chain dihedrals distributions as illustrated by few representative cases in Figs.5.9.(a)-(b). We denote distribution of dihedral θ of fliC in water as $H_i^{flic,W}(\theta)$ and in bilayer as $H_i^{flic,L}(\theta)$, where dihedral θ belongs to residue i . We observe that (Fig.5.9.(a)) $H_{A23}^{flic,W}(\psi)$ exhibits bimodal distribution, whereas $H_{A23}^{flic,L}(\psi)$ is unimodal. Similarly $H_{L492}^{flic,L}(\chi_1)$ (Fig.5.9.(b)) remains sharply peaked unimodal but $H_{L492}^{flic,W}(\chi_1)$ shows flat distribution. Bimodal or multimodal distributions signify enhancement of dihedral flexibility where as sharp peaked distributions indicate loss of flexibility or compactness in secondary structure. Both A23 and L492 of fliC undergo loss of

secondary structures in water. But, in membrane they become more structured which leads to sharply peaked unimodal distribution. There are other residues (L17-A23 in N terminal D0 domain and R467-Q484 in C terminal D0 domain as indicated in Figs.5.10.(a)-(b)) which show complete loss of secondary structures in presence of water, showing significant changes in dihedral flexibility in bilayer than water. Thus D0 remains less flexible and more structured in membrane than water.

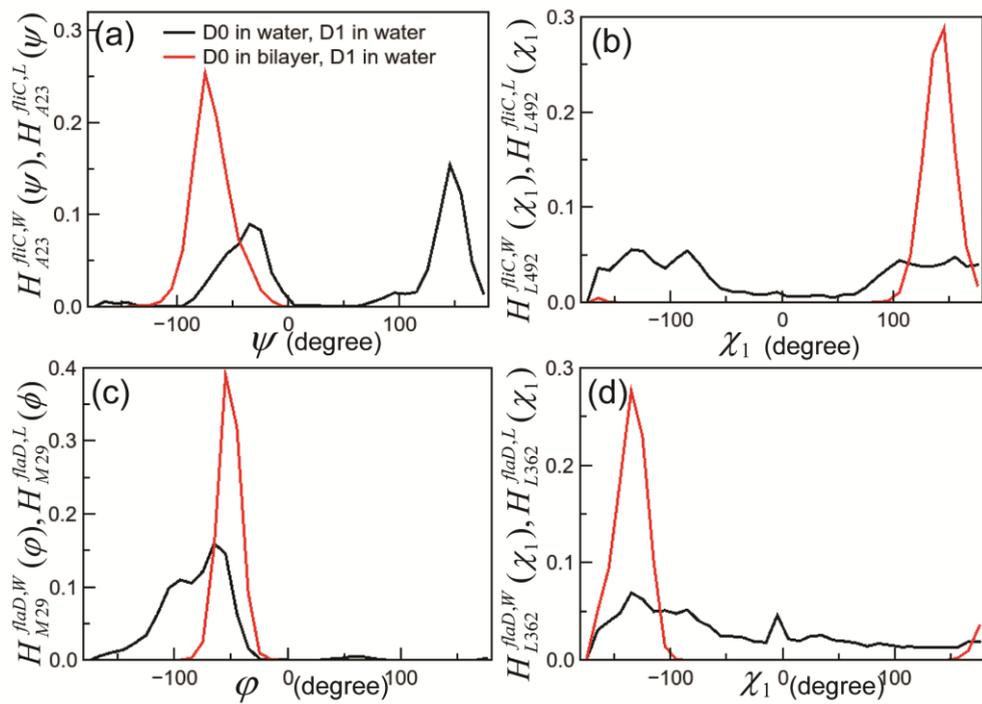


Fig.5.9. Equilibrium responses for the residues of flagellin which show significant changes in presence of bilayer with respect to aqueous medium: (a) Dihedral distributions of fliC: $H_{A23}^{fliC,W}(\psi)$ vs ψ ; $H_{A23}^{fliC,L}(\psi)$ vs ψ and (b) $H_{L492}^{fliC,W}(\chi_1)$ vs χ_1 ; $H_{L492}^{fliC,L}(\chi_1)$ vs χ_1 . (c) Dihedral distributions of flaD: $H_{M29}^{flaD,W}(\phi)$ vs ϕ ; $H_{M29}^{flaD,L}(\phi)$ vs ϕ and (d) $H_{L362}^{flaD,W}(\chi_1)$ vs χ_1 ; $H_{L362}^{flaD,L}(\chi_1)$ vs χ_1 .

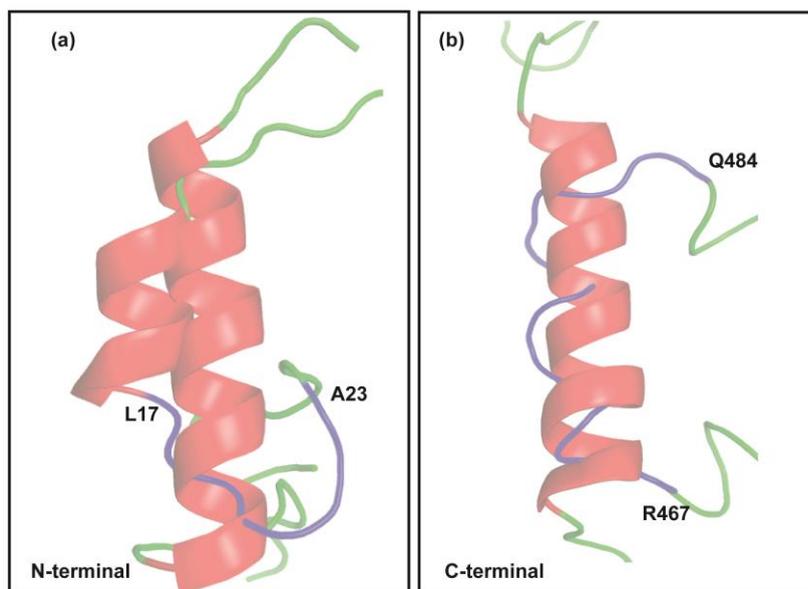


Fig.5.10. Residues of D0 domain of fliC, having loss of compact secondary structures and showing prominent changes in dihedral distributions in water than bilayer; (a) N terminal and (b) C terminal.

We account for changes in free energy and entropy of conformational changes of the L-system with respect to W-system from the distributions of the dihedral angles. The overall changes in conformational thermodynamics are obtained by adding all the dihedral contributions from all residues. The data are shown in Table.5.3. We observe that D0 remains energetically stabilized and ordered in membrane than water, (both ΔG_{flic}^{D0} and $T\Delta S_{flic}^{D0}$ negative). We also report changes in free energy and entropy separately by backbone ($\Delta G_{flic}^{bb,D0}$, $T\Delta S_{flic}^{bb,D0}$) and side chain dihedrals ($\Delta G_{flic}^{sc,D0}$, $T\Delta S_{flic}^{sc,D0}$) respectively. We find that the stability and order are primarily dominated by side chain fluctuations. This large ordering indicates reduced flexibility of D0 in presence of bilayer. D1 domain also undergoes stabilization and order as indicated by changes in free energy (ΔG_{flic}^{D1}) and entropy ($T\Delta S_{flic}^{D1}$) in Table.5.3. Ordering as well as stability in free energy are governed by side chain fluctuations. Besides, the total changes in free energy for the entire conserved domain of fliC (ΔG_{flic}) indicates that the side chain (ΔG_{flic}^{sc}) is the primary governing factor in providing stabilization to the

protein than backbone (ΔG_{flic}^{bb}). Similarly change in entropy for entire conserved domain ($T\Delta S_{flic}$) shows major contributions of side chain dihedrals ($T\Delta S_{flic}^{sc}$). Thus, we find that bilayer induces larger conformational stability and order to each individual domains of fliC primarily via side chain.

Table.5.3. Changes in free energy and entropy of different domains of fliC in presence of bilayer than water, indicating enhanced stability in bilayer.

System	Change in free energy ($K_B T$)	Change in entropy ($K_B T$)
D0 of fliC	$\Delta G_{flic}^{D0} \sim - 24.0$	$T\Delta S_{flic}^{D0} \sim - 113.2$
Backbone of D0 of fliC	$\Delta G_{flic}^{bb,D0} \sim - 11.2$	$T\Delta S_{flic}^{bb,D0} \sim - 35.3$
Side chain of D0 of fliC	$\Delta G_{flic}^{sc,D0} \sim - 12.8$	$T\Delta S_{flic}^{sc,D0} \sim - 77.9$
D1 of fliC	$\Delta G_{flic}^{D1} \sim - 31.5$	$T\Delta S_{flic}^{D1} \sim - 203.0$
Backbone of D1 of fliC	$\Delta G_{flic}^{bb,D1} \sim 1.4$	$T\Delta S_{flic}^{bb,D1} \sim 0.4$
Side chain of D1 of fliC	$\Delta G_{flic}^{sc,D1} \sim - 32.9$	$T\Delta S_{flic}^{sc,D1} \sim - 203.4$
(D0+D1) of fliC	$\Delta G_{flic} \sim - 60.7$	$T\Delta S_{flic} \sim - 64.5$
Backbone for (D0+D1) of fliC	$\Delta G_{flic}^{bb} \sim - 10.9$	$T\Delta S_{flic}^{bb} \sim - 9.8$
Side chain for fliC	$\Delta G_{flic}^{sc} \sim - 49.8$	$T\Delta S_{flic}^{sc} \sim - 54.7$

5.3.2. Conformational stability of flaD:

We repeat the same analysis for simulated trajectories on model structure of flaD. We find that both D0 and D1 domain are helical in secondary structural assignment in the model structure. We generate Ramachandran plot for D0 domain, both in water and bilayer embedded states (Fig.5.11.) to address structural stability of D0 domain in

different medium. We find that $\sim 47.9\%$ residues of D0 domain undergo loss of alpha helices in aqueous environment with respect to modeled initial conformation. However, in presence of bilayer only $\sim 12.7\%$ residues of D0 become unstructured.

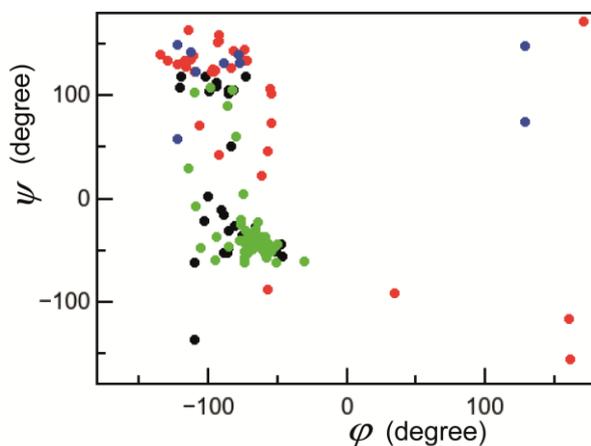


Fig.5.11. $\varphi - \psi$ correlation plot of D0 residues of modeled flaD for water and, Black; residues that are structured in water and Red; the unstructured ones in water, Green; structured residues in bilayer and Blue; unstructured residues in bilayer.

We generate dihedral distributions for different residues of D0 domain of flaD in presence of water and bilayer. We show here a few representative cases. We find that distribution of dihedral φ of the residue M29 in water ($H_{M29}^{flaD,W}(\varphi)$) is unimodal with a broad width; however, that in bilayer ($H_{M29}^{flaD,L}(\varphi)$) becomes unimodal with a sharp strong peak, as depicted in Fig.5.9.(c). Similarly, flat multimodal distribution of L362, $H_{L362}^{flaD,W}(\chi_1)$ in Fig.5.9.(d) converts into sharply peaked unimodal $H_{L362}^{flaD,L}(\chi_1)$. In analogy with fliC, there are other residues like G20-R31 and E344-L375 (Figs.5.12.(a)-(b)), which loss compact secondary structures in water and exhibit significant changes in dihedral distributions in bilayer. Thus, we find that bilayer induces structural stability to D0 domain.

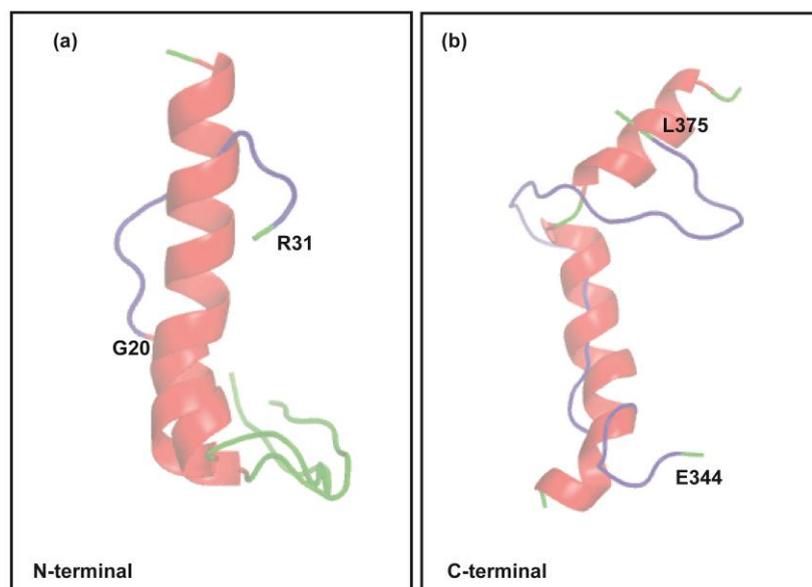


Fig.5.12. Residues of D0 domain of flaD, having loss of compact secondary structures and showing prominent changes in dihedral distributions in water than bilayer; (a) N terminal and (b) C terminal.

The data for changes in conformational thermodynamics are shown in Table.5.4. Data for D0 (ΔG_{flaD}^{D0} , $T\Delta S_{flaD}^{D0}$) and D1 (ΔG_{flaD}^{D1} , $T\Delta S_{flaD}^{D1}$) indicate that both domains are more stabilized and ordered in bilayer than water. We observe that in analogy with fliC, the stabilization and order are due to side chain dihedrals ($\Delta G_{flaD}^{sc,D0}$, $\Delta G_{flaD}^{sc,D1}$, $T\Delta S_{flaD}^{sc,D0}$ and $T\Delta S_{flaD}^{sc,D1}$). Moreover, the entire conserved domain of flaD gains conformational stability ($\Delta G_{flaD} < 0$) and order ($T\Delta S_{flaD} < 0$) in presence of bilayer owing to larger stability and ordering in side chain fluctuations.

Table.5.4. Changes in free energy and entropy of different domains of flaD in presence of bilayer than water, indicating enhanced stability in bilayer.

System	Change in free energy ($K_B T$)	Change in entropy ($K_B T$)
D0 of flaD	$\Delta G_{flaD}^{D0} \sim - 20.4$	$T\Delta S_{flaD}^{D0} \sim - 111.0$
Backbone of D0 of flaD	$\Delta G_{flaD}^{bb,D0} \sim - 8.2$	$T\Delta S_{flaD}^{bb,D0} \sim - 40.6$
Side chain of D0 of flaD	$\Delta G_{flaD}^{sc,D0} \sim - 12.2$	$T\Delta S_{flaD}^{sc,D0} \sim - 70.4$
D1 of flaD	$\Delta G_{flaD}^{D1} \sim - 21.5$	$T\Delta S_{flaD}^{D1} \sim - 183.6$
Backbone of D1 of flaD	$\Delta G_{flaD}^{bb,D1} \sim - 9.4$	$T\Delta S_{flaD}^{bb,D1} \sim - 24.7$
Side chain of D1 of flaD	$\Delta G_{flaD}^{sc,D1} \sim - 30.9$	$T\Delta S_{flaD}^{sc,D1} \sim - 208.3$
(D0+D1) of flaD	$\Delta G_{flaD} \sim - 45.9$	$T\Delta S_{flaD} \sim - 49.5$
Backbone for (D0+D1) of flaD	$\Delta G_{flaD}^{bb} \sim - 0.3$	$T\Delta S_{flaD}^{bb} \sim 0.5$
Side chain for (D0+D1) of flaD	$\Delta G_{flaD}^{sc} \sim - 46.2$	$T\Delta S_{flaD}^{sc} \sim - 50.0$

5.3.3. Binding interface between flagellin and TLR5 receptor:

Earlier studies^{37, 38} based on conformational thermodynamics suggest that the destabilized and disordered residues of a protein in a particular conformation are the functional ones in that state. We find that there are residues of D1 domain of fliC that are destabilized as well as disordered (thermodynamics data in Table.5.5.) as indicated in Fig.5.13.(a). We consider the available crystal structure of fliC-TLR5 complex and compare the binding interface to these residues. We find residues like

E93, V96, S110, I111, A113, G440, N441 and T447 belong to TLR5 binding interface, as shown in Fig.5.13.(b).

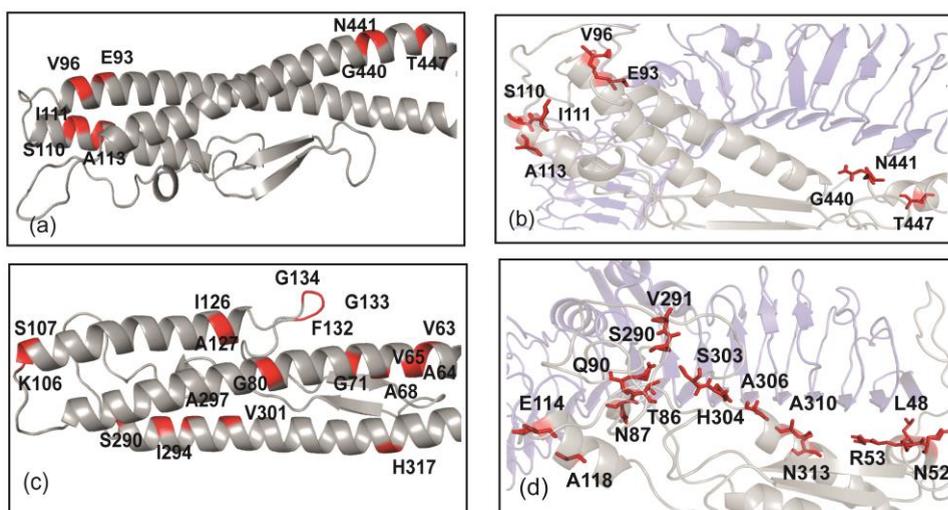


Fig.5.13.(a) Destabilized and disordered residues of fliC in presence of bilayer. **(b)** The destabilized and disordered residues of fliC that belong to known TLR5 binding interface as compared with the crystal structure. **(c)** Destabilized and disordered residues of flaD in presence of bilayer. **(d)** Residues of flaD, belonging to TLR5 binding interface as observed in average structure generated over simulated trajectories of flaD-TLR5 complex.

Table.5.5. Amount of destabilization (ΔG) and disorder ($T\Delta S$) of the residues of D1 domain of fliC in presence of bilayer with respect to water.

Residue	$\Delta G (K_B T)$	$T\Delta S (K_B T)$
E93	0.22	2.23
V96	0.04	1.09
S110	0.79	0.76
I111	0.16	0.22
A113	0.13	0.70
G440	0.24	1.52
N441	0.09	0.36
T447	0.03	0.37

In case of flaD, there are destabilized and disordered (Table.5.6.) residues in D1 (Fig.5.13.(c)) domain. Majority of these residues are hydrophobic in nature, quite different than those of fliC. We perform MD simulation on model structure of flaD-TLR5 complex using the GROMACS package⁵¹. We identify the residues like L48, N52, R53, T86, N87, Q90, E114, A118, S290, V291, S303, H304, A306, A310 and N113, which belong to the binding interface (Fig.5.13.(d)) based on the average simulated structure. We observe that the TLR5 binding interface of flaD is quite different sequence wise than that of fliC.

The site-directed mutagenesis analysis of D0 reveals that the D0 domain is mainly responsible for pro-inflammatory responses²⁷, associated with TLR5 binding. Moreover, from experiments based on protein fusion strategy²⁸, it is proposed that D0 has functional role in formation and activation of dimeric TLR5 receptor. It has been also suggested²⁸ that D0 region might interact with TLR5 in the process of receptor dimerization. Since, the full length crystal structure of TLR5 along with transmembrane and intracellular domains⁶⁰ is not reported yet, the microscopic basis of involvement of D0 region in TLR5 interaction is not understood. Our studies show that D0 of both fliC and flaD is conformationally more stable in bilayer than in water. Our observations suggest the possible role of D0 may be anchoring of flagellin in bilayer. This is likely to play a crucial role in maintaining orientation of D1 during interaction with TLR5.

Table.5.6. Amount of destabilization (ΔG) and disorder ($T\Delta S$) of the residues of D1 domain of flaD in presence of bilayer with respect to water.

Residue	$\Delta G (K_B T)$	$T\Delta S (K_B T)$
V63	0.94	0.97
A64	1.02	1.86
V65	0.36	1.90
A68	0.31	1.99
G71	0.34	0.81
G80	0.26	0.42
K106	0.43	0.76
S107	0.91	1.04
I126	0.41	1.03
A127	0.65	1.56
F132	0.29	1.49
G133	0.26	2.16
G134	0.22	1.13
S290	0.26	1.83
I294	0.39	2.11
A297	0.45	1.16
V301	0.88	1.01
H317	0.37	0.46

5.3.4. Conformational stability of flagellin-TLR5 complexes:

We generate distributions of dihedral θ for the i th residue of fliC in presence of TLR5 ($H_i^{flic,c}(\theta)$) over simulated trajectories and compare to those in the free state

($H_i^{fliC,L}(\theta)$). Some representative cases are shown in Figs.5.14.(a)-(b). We find that both $H_{Q146}^{fliC,L}(\psi)$ and $H_{L167}^{fliC,L}(\chi_1)$ are unimodal. However, in presence of TLR5, $H_{Q146}^{fliC,c}(\psi)$ along with $H_{L167}^{fliC,c}(\chi_1)$ exhibit bimodal distributions. This signifies that both Q146 and L167 gain flexibility in complex form. There are also residues of fliC which do not show significant changes in dihedral distribution in presence of TLR5. One such representative case is T116, where both $H_{T116}^{fliC,L}(\varphi)$ and $H_{T116}^{fliC,c}(\varphi)$ remain unimodal as indicated in Fig.5.14.(c).

Similarly, we calculate histogram for dihedral θ for the i th residues of TLR5 in presence of fliC ($H_i^{TLR5,fliC}(\theta)$) as well as in its free state ($H_i^{TLR5,f}(\theta)$). We observe that for some residues dihedrals show prominent changes in their distributions. Figs.5.15.(a)-(b) signify that $H_{T113}^{TLR5,f}(\psi)$ is bimodal in free state, whereas it loses flexibility and becomes unimodal ($H_{T113}^{TLR5,fliC}(\psi)$) in the complex. In analogy with T113, bimodal nature of $H_{N131}^{TLR5,f}(\chi_1)$ converts into unimodal $H_{N131}^{TLR5,fliC}(\chi_1)$. However, for some residues there are no noticeable changes in dihedral distributions as represented in Fig.5.15.(c) (unimodal $H_{I28}^{TLR5,f}(\varphi)$ and $H_{I28}^{TLR5,fliC}(\varphi)$).

Next, we calculate changes in conformational free energy ($\Delta G_{fliC,c}$) and entropy ($T\Delta S_{fliC,c}$) of fliC in presence of TLR5 with respect to the free state (Table.5.7.). We observe that fliC remains energetically destabilized as well as entropically disordered in the complex. Major contribution in destabilization and disorder comes from backbone. We find that for fliC-TLR5 complex, the residues of D1 domain like Q146 and L167 (Fig.5.16.(a)), have significant amount of destabilization ($> 1.0 K_B T$) and disorder ($> 1.0 K_B T$) in presence of TLR5. But, no residues of D0 show such significant destabilization in fliC-TLR5 system. We also calculate changes in conformational free energy (ΔG_{fliC}^{TLR5}) and entropy ($T\Delta S_{fliC}^{TLR5}$) of TLR5 in presence of

fliC with respect to the free state. TLR5 is energetically stabilized and entropically ordered (Table.5.7.) in the complex primarily due to side chain fluctuations. We estimate the changes in free energy ($\Delta G_{flic-TLR5}$) and entropy ($T\Delta S_{flic-TLR5}$) of the fliC-TLR5 complex by adding the contributions from fliC and TLR5. We observe that the complex is overall stabilized and ordered compared to the free state of fliC.

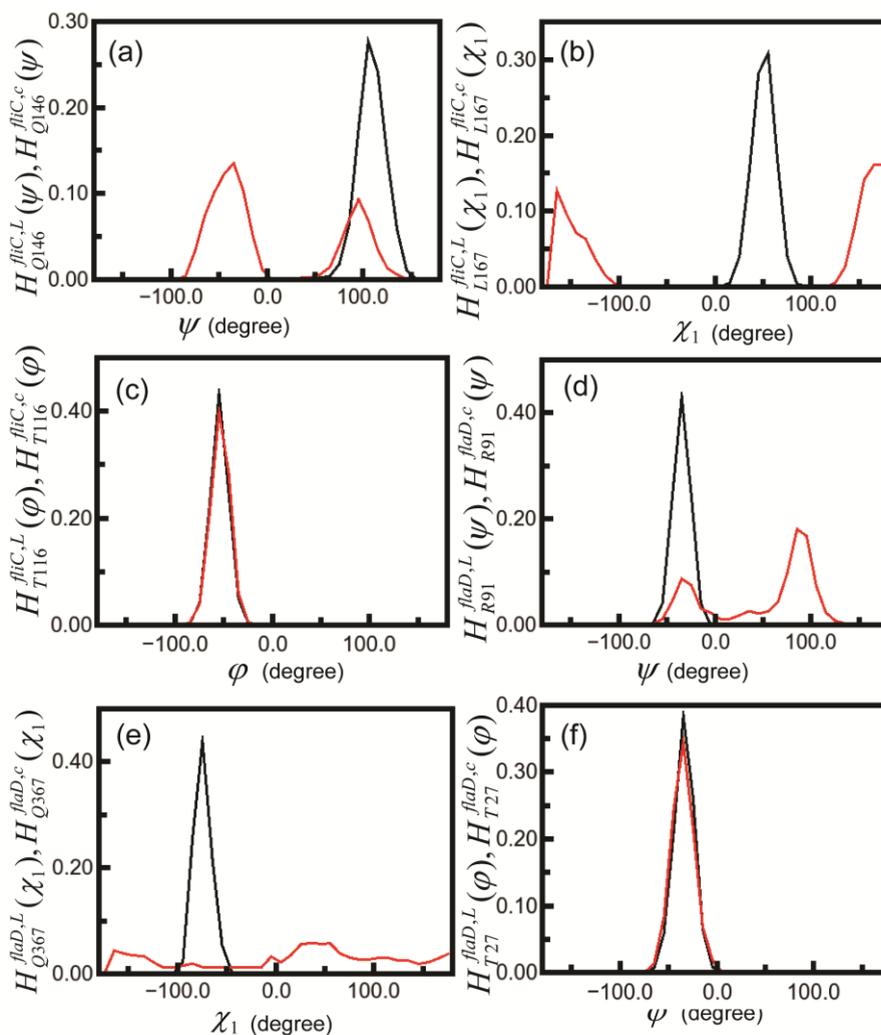


Fig.5.14. Equilibrium responses for the residues of flagellin in presence of TLR5 (red) with respect TLR5 free state (black): (a) Dihedral distributions of fliC: $H_{Q146}^{flic,c}(\psi)$ vs ψ ; $H_{Q146}^{flic,L}(\psi)$ vs ψ and (b) $H_{L167}^{flic,c}(\chi_1)$ vs χ_1 ; $H_{L167}^{flic,L}(\chi_1)$ vs χ_1 showing significant changes in presence of TLR5. (c) $H_{T116}^{flic,c}(\phi)$ vs ϕ and $H_{T116}^{flic,L}(\phi)$ vs ϕ do not exhibit noticeable changes. (d) Changes in dihedral distributions of flaD: $H_{R91}^{flaD,c}(\psi)$ vs ψ ; $H_{R91}^{flaD,L}(\psi)$ vs ψ and (e) $H_{Q367}^{flaD,c}(\chi_1)$ vs χ_1 ; $H_{Q367}^{flaD,L}(\chi_1)$ vs χ_1 . (f) $H_{T27}^{flaD,c}(\phi)$ vs ϕ and $H_{T27}^{flaD,L}(\phi)$ vs ϕ remain almost similar.

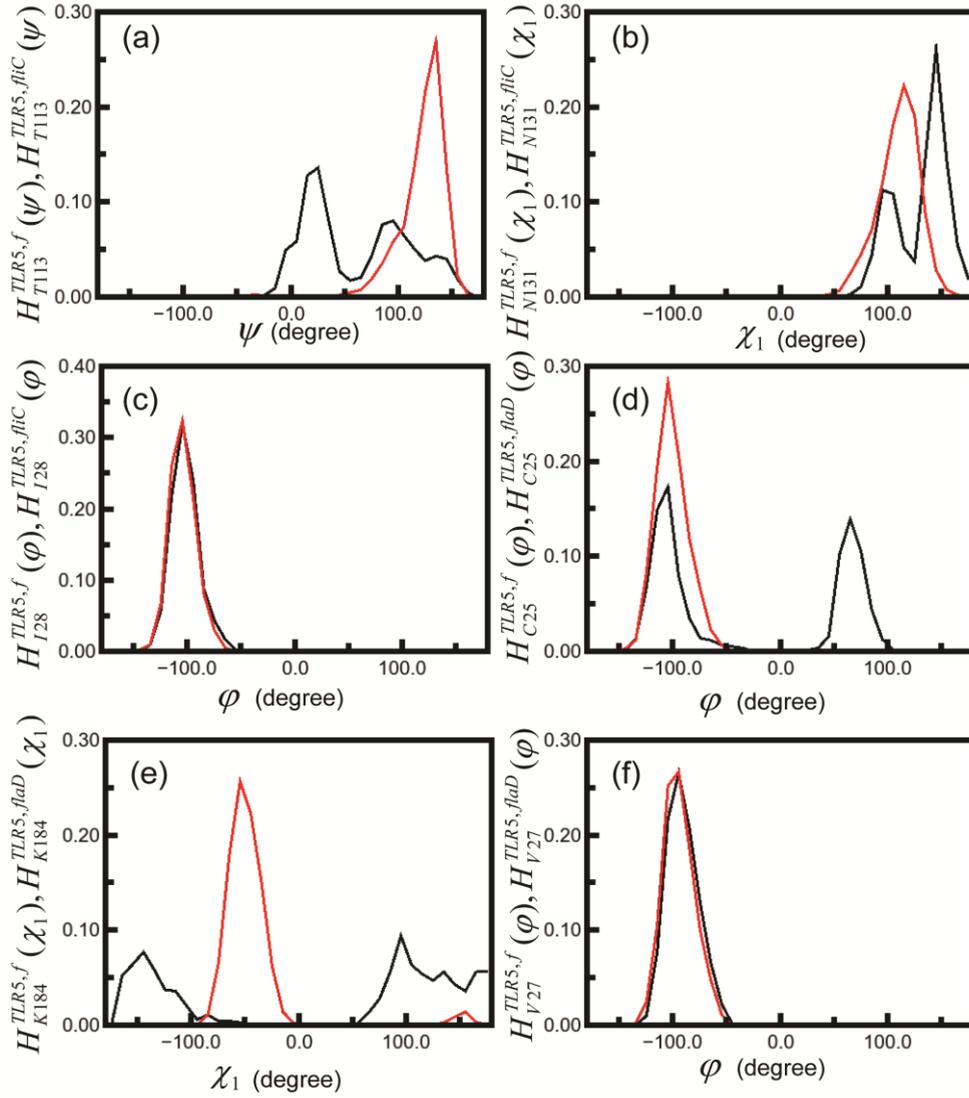


Fig.5.15. Equilibrium responses for the residues of TLR5 in presence of flagellin (red) with respect its free state (black): (a) $H_{T113}^{TLR5,flc}(\psi)$ vs ψ ; $H_{T113}^{TLR5,f}(\psi)$ vs ψ and (b) $H_{N131}^{TLR5,flc}(\chi_1)$ vs χ_1 ; $H_{N131}^{TLR5,f}(\chi_1)$ vs χ_1 showing significant changes in presence of fliC. (c) $H_{I28}^{TLR5,flc}(\varphi)$ vs φ and $H_{I28}^{TLR5,f}(\varphi)$ vs φ do not exhibit noticeable changes. (d) $H_{C25}^{TLR5,flaD}(\varphi)$ vs φ ; $H_{C25}^{TLR5,f}(\varphi)$ vs φ and (e) $H_{K184}^{TLR5,flaD}(\chi_1)$ vs χ_1 ; $H_{K184}^{TLR5,f}(\chi_1)$ vs χ_1 indicating changes in distributions. (f) $H_{V27}^{TLR5,flaD}(\varphi)$ vs φ and $H_{V27}^{TLR5,f}(\varphi)$ vs φ remain almost similar.

Table.5.7. Changes in free energy and entropy of fliC, TLR5 and the fliC-TLR5 complex.

System	Change in free energy ($K_B T$)	Change in entropy ($K_B T$)
fliC	$\Delta G_{flic,c} \sim 11.2$	$T\Delta S_{flic,c} \sim 57.7$
Backbone of fliC	$\Delta G_{flic,c}^{bb} \sim 7.9$	$T\Delta S_{flic,c}^{bb} \sim 34.8$
Side chain of fliC	$\Delta G_{flic,c}^{sc} \sim 3.3$	$T\Delta S_{flic,c}^{sc} \sim 22.9$
TLR5	$\Delta G_{flic}^{TLR5} \sim -20.2$	$T\Delta S_{flic}^{TLR5} \sim -145.4$
Backbone of TLR5	$\Delta G_{flic}^{TLR5,bb} \sim 27.8$	$T\Delta S_{flic}^{TLR5,bb} \sim 69.3$
Side chain of TLR5	$\Delta G_{flic}^{TLR5,sc} \sim -48.0$	$T\Delta S_{flic}^{TLR5,sc} \sim -214.7$
fliC-TLR5 complex	$\Delta G_{flic-TLR5} \sim -9.0$	$T\Delta S_{flic-TLR5} \sim -87.7$

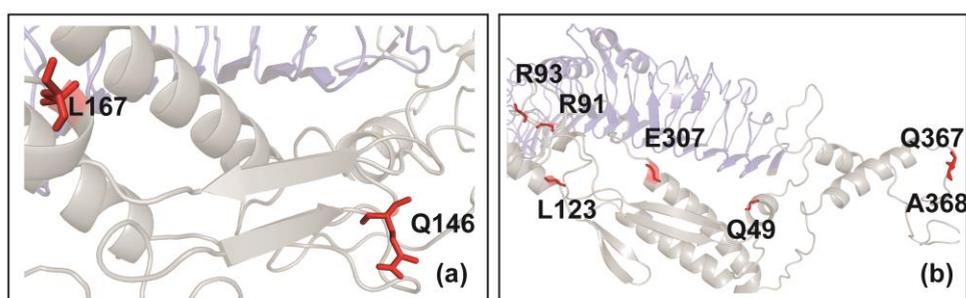


Fig.5.16.(a) Destabilized and disordered residues of fliC in presence of TLR5 and those of (b) flaD as obtained from conformational thermodynamics.

We further address distributions of dihedral θ for the i th residues of flaD in TLR5 bound form ($H_i^{flaD,c}(\theta)$) as well as in the free state ($H_i^{flaD,L}(\theta)$). We have shown

some representative cases in Figs.5.14.(d)-(f). We observe that $H_{R91}^{flaD,L}(\psi)$ exhibits unimodal distribution in TLR5 free state; whereas in presence of TLR5, $H_{R91}^{flaD,c}(\psi)$ becomes bimodal. Similarly, $H_{Q367}^{flaD,L}(\chi_1)$ is unimodal, but it shows uniform flat distribution in presence of TLR5. These observations indicate that those residues gain flexibility in presence of TLR5. But, there are some residues like T27 which do not exhibit changes in dihedral distribution, both $H_{T27}^{flaD,L}(\varphi)$ and $H_{T27}^{flaD,L}(\varphi)$ are almost identical (Fig.5.14.(f)).

For TLR5, dihedral distributions (Figs.5.15.(d)-(e)) indicate that both $H_{C25}^{TLR5,f}(\varphi)$ and $H_{K184}^{TLR5,f}(\chi_1)$ are bimodal in absence of flaD. However in presence of flaD, they become less flexible as indicated by the unimodal profile of $H_{C25}^{TLR5,flaD}(\varphi)$ and $H_{K184}^{TLR5,flaD}(\chi_1)$. Nevertheless, the unimodal nature of $H_{V27}^{TLR5,flaD}(\varphi)$ and $H_{V27}^{TLR5,f}(\varphi)$ (Fig.5.15.(f)) signify that V27 does not show any changes in presence of flaD.

The data for changes in conformational free energy ($\Delta G_{flaD,c}$) and entropy ($T\Delta S_{flaD,c}$) of flaD in presence of TLR5 indicate that flaD is destabilized as well as disordered in the complex (Table.5.8.). Backbone is mainly responsible for destabilization whereas side chain contributes to disorder. However, unlike fliC-TLR5 system, in case of flaD-TLR5 there are several residues of D1 domain like Q49, R91, R93, L123 and E307 that are strongly destabilized ($\Delta G > 1.0 K_B T$) and disordered ($T\Delta S > 1.0 K_B T$) (Fig.5.16.(b)) with respect to free flaD. We also identify that among such residues, there are two residues like Q367 and A368 that belong to D0. The amount of destabilization and disorder of such residues for both fliC and flaD are shown in Table.5.9. Next we calculate changes in conformational free energy (ΔG_{flaD}^{TLR5}) and entropy ($T\Delta S_{flaD}^{TLR5}$) of TLR5 in presence of flaD with respect to the free state. We find that TLR5 remains stabilized and ordered primarily due to side chain.

The changes in overall free energy ($\Delta G_{flaD-TLR5}$) and entropy ($T\Delta S_{flaD-TLR5}$) of flaD-TLR5 complex show that the overall complex is stabilized and ordered with respect to the free state of flaD.

Table.5.8. Changes in free energy and entropy of flaD, TLR5 and the flaD-TLR5 complex.

System	Change in free energy ($K_B T$)	Change in entropy ($K_B T$)
flaD	$\Delta G_{flaD,c} \sim 17.0$	$T\Delta S_{flaD,c} \sim 81.5$
Backbone of flaD	$\Delta G_{flaD,c}^{bb} \sim 8.6$	$T\Delta S_{flaD,c}^{bb} \sim 32.5$
Side chain of flaD	$\Delta G_{flaD,c}^{sc} \sim 8.4$	$T\Delta S_{flaD,c}^{sc} \sim 49.0$
TLR5	$\Delta G_{flaD}^{TLR5} \sim -40.3$	$T\Delta S_{flaD}^{TLR5} \sim -206.4$
Backbone of TLR5	$\Delta G_{flaD}^{TLR5,bb} \sim 24.5$	$T\Delta S_{flaD}^{TLR5,bb} \sim 68.6$
Side chain of TLR5	$\Delta G_{flaD}^{TLR5,sc} \sim -64.8$	$T\Delta S_{flaD}^{TLR5,sc} \sim -275.0$
flaD-TLR5 complex	$\Delta G_{flaD-TLR5} \sim -23.3$	$T\Delta S_{flaD-TLR5} \sim -193.5$

Table.5.9. Significant amount of destabilization ($\Delta G > 1.0 K_B T$) and disorder ($T\Delta S > 1.0 K_B T$) of the residues of fliC and flaD in presence of TLR5 with respect to their free states.

Flagellin	Residue	$\Delta G (K_B T)$	$T\Delta S (K_B T)$
fliC	Q146	1.02	1.97
fliC	L167	1.42	1.51
flaD	Q49	1.12	3.24
flaD	R91	1.58	4.29
flaD	R93	1.50	5.94
flaD	L123	1.17	3.58

flaD	E307	1.30	1.81
flaD	Q367	1.19	7.33
flaD	A368	1.05	1.73

5.3.5. Coreceptor binding:

Often in host cell, intracellular response upon pathogenic recognition is activated in presence of series of coreceptor GPCR proteins. The immunological experiment³⁴ indicate that fliC-TLR5 does not involve GPCR. However, the immune response for flaD-TLR5 system is GPCR dependent. Specifically, both the A₂A adenosine receptor^{54, 55} and VPAC1⁵⁶⁻⁵⁸ are found to contribute to the response in flaD-TLR5 interaction.

Ligand bound crystal structure of A₂A along with mutational studies^{54, 55} reveal that the residues (L167, F168 and E169) of second extracellular loop (ecL2) are known to provide ligand binding specificity associated with intracellular response. On the other hand in case of VPAC1⁵⁶⁻⁵⁸, the large conserved extracellular domain (ecd) is responsible for ligand binding. Mutagenesis and docking studies⁵⁶⁻⁵⁸ show that D107, G116, Cysteine; C122 and K127 of extracellular domain are the key residues that are important for ligand binding.

In our conformational thermodynamics analysis of flagellin-TLR5 complex, we find that there one hydrophobic and one polar residues in D1 region of fliC are destabilized after binding with TLR5. They are more prone to participate hydrophobicity mediated function and polar residue might participate in weak electrostatic interactions. However, for flaD the destabilized residues in D1 domain are acidic, basic, polar and hydrophobic as well. So, they are highly probable to participate in binding govern by strong electrostatic attractions. We find that for A₂A

receptor, the known binding residues are mainly hydrophobic, whereas for VPAC1 the residues are hydrophilic in nature. These observations lead us to perform docking of A₂A and VPAC1 to D1 region of flagellin.

In the absence of crystal structure, we build up mode for the ligand specific domain of A₂A and VPAC1 coreceptors using 4UG2.PDB and 2X57.PDB as the templates respectively. The model structure along with the binding hot spot residues of A₂A and VPAC1 are shown in Figs.5.17.(a)-(b). Model structure of A₂A consists of small helix and loop region, whereas for VPAC1 it is primarily helical along with β sheets. We dock these models to flagellin-receptor complex. The docked complexes for fliC-TLR5-A₂A, fliC-TLR5-VPAC1, flaD-TLR5-A₂A along with flaD-TLR5-VPAC1 are shown in Figs.5.18.(a)-(d).

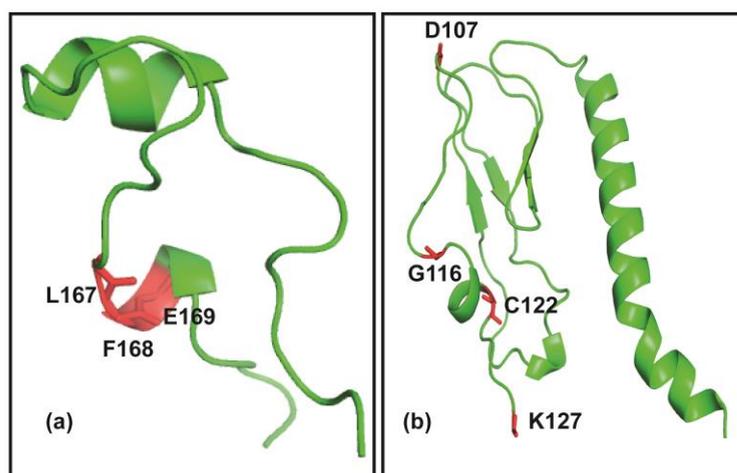


Fig.5.17. Binding hotspots for GPCRs shown in modeled structures; (a) A₂A and (b) VPAC1.

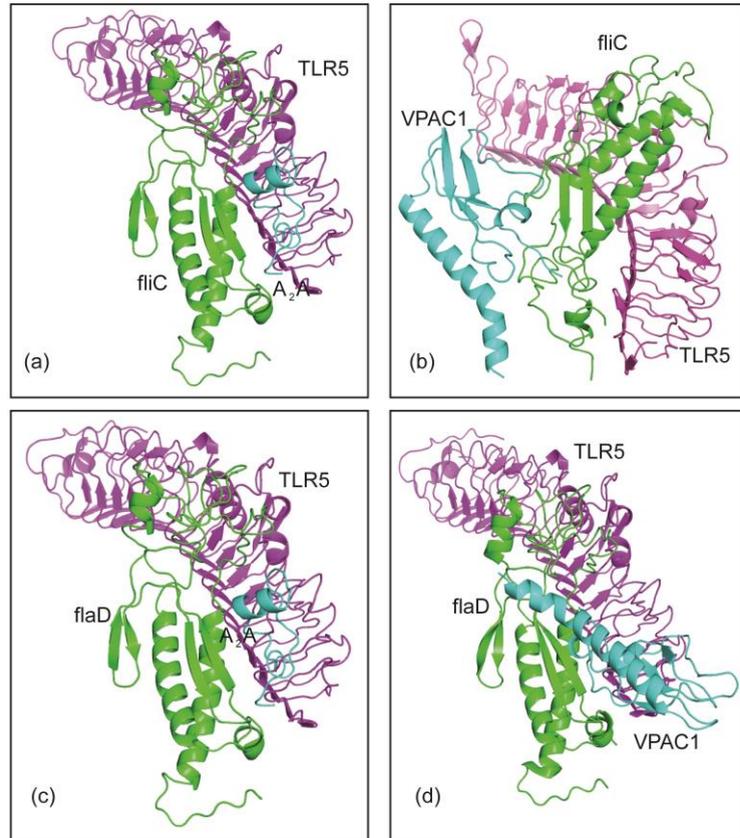


Fig.5.18. Docked structures of the flagellin-receptor-coreceptor; (a) fliC-TLR5-A₂A, (b) fliC-TLR5-VPAC1, (c) flaD-TLR5-A₂A and that for (b) flaD-TLR5-VPAC1 complexes.

Our thermodynamics data show that the receptor protein TLR5 remains stabilized and ordered in the complexes. This leads us to compute changes in free energy and entropy for the flagellin-coreceptor complexes. We calculate changes in free energy ($\Delta G_{A_2A}^{fliC}$) and entropy ($T\Delta S_{A_2A}^{fliC}$) of A₂A in presence of fliC-TLR5 complex as well as in flaD-TLR5 system ($\Delta G_{A_2A}^{flaD}, T\Delta S_{A_2A}^{flaD}$) with respect to its free state. Similarly, we focus on thermodynamic changes for fliC ($\Delta G_{fliC}^{A_2A}$ and $T\Delta S_{fliC}^{A_2A}$) and flaD ($\Delta G_{flaD}^{A_2A}$ and $T\Delta S_{flaD}^{A_2A}$) in presence of A₂A with respect to the coreceptor free state. We follow the same process for VPAC1 and estimate thermodynamic changes of the coreceptor in presence of both fliC and flaD respectively

by ΔG_{VPAC1}^{fliC} , $T\Delta S_{VPAC1}^{fliC}$, ΔG_{VPAC1}^{flaD} , $T\Delta S_{VPAC1}^{flaD}$. The thermodynamic stability and order of both the flagellins after interacting with VPAC1 (ΔG_{fliC}^{VPAC1} , $T\Delta S_{fliC}^{VPAC1}$, ΔG_{flaD}^{VPAC1} , $T\Delta S_{flaD}^{VPAC1}$) are also shown in Table.5.10. The thermodynamic data suggest that fliC-A₂A as well as flaD-A₂A complexes are stabilized. However, in case of VPAC1 only flaD-VPAC1 shows stability. Thus flaD can form stable complexes both with A₂A and VPAC1, while fliC can form complex with A₂A alone.

Table.5.10. Changes in free energy and entropy of fliC and flaD in presence of coreceptors.

System	Change in free energy (K_bT)	Change in entropy (K_bT)
fliC in presence of A ₂ A	$\Delta G_{A_2A}^{fliC} \sim -19.0$	$T\Delta S_{A_2A}^{fliC} \sim -11.4$
A ₂ A in presence of fliC	$\Delta G_{fliC}^{A_2A} \sim -4.0$	$T\Delta S_{fliC}^{A_2A} \sim -49.8$
fliC in presence of VPAC1	$\Delta G_{VPAC1}^{fliC} \sim -1.0$	$T\Delta S_{VPAC1}^{fliC} \sim 75.8$
VPAC1 in presence of fliC	$\Delta G_{fliC}^{VPAC1} \sim 9.6$	$T\Delta S_{fliC}^{VPAC1} \sim 29.9$
flaD in presence of A ₂ A	$\Delta G_{A_2A}^{flaD} \sim -12.3$	$T\Delta S_{A_2A}^{flaD} \sim -110.0$
A ₂ A in presence of flaD	$\Delta G_{flaD}^{A_2A} \sim -6.1$	$T\Delta S_{flaD}^{A_2A} \sim -34.3$
flaD in presence of VPAC1	$\Delta G_{VPAC1}^{flaD} \sim -11.2$	$T\Delta S_{VPAC1}^{flaD} \sim -7.6$
VPAC1 in presence of flaD	$\Delta G_{flaD}^{VPAC1} \sim -2.7$	$T\Delta S_{flaD}^{VPAC1} \sim -21.3$

5.4. Conclusions:

To summarize, we find that bilayer induces thermodynamic and structural stability to the disordered D0 domain of flagellin. The TLR5 binding interfaces with D1 region for fliC and flaD are not identical. However, both the flagellin-receptor complex is stabilized. We also observe coreceptors like A₂A and VPAC1 can interact with flaD-TLR5. The microscopic details for the interaction of flagellin with receptor and coreceptors might be helpful in pharmaceutical applications.

5.5. References:

1. J. Haiko and B. Westerlund-Wikstrom, *Biology*, 2013, **2**, 1242-1267.
2. R. S. Friedlander, N. Vogel and J. Aizenberg, *Langmuir*, 2015, **31**, 6137-6144.
3. Y. Rossez, E. B. Wolfson, A. Holmes, D. L. Gally and N. J. Holden, *PLOS Pathog.*, 2015, **11**.
4. Q. Duan, M. Zhou, L. Zhu and G. Zhu, *J. Basic Microbiol.*, 2013, **53**, 1-8.
5. S. A. Beatson, T. Minamino and M. J. Pallen, *Trends Microbiol*, 2006, **14**, 151-155.
6. A. S. Utada, R. R. Bennett, J. C. N. Fong, M. L. Gibiansky, F. H. Yildiz, R. Golestanian and G. C. L. Wong, *Nat. Commun.*, 2014, **5**.
7. R. Warrington, W. Watson, H. L. Kim and F. R. Antonetti, *Allergy, asthma, and clinical immunology : official journal of the Canadian Society of Allergy and Clinical Immunology*, 2011, **7 Suppl 1**, S1.
8. T. Kawasaki and T. Kawai, *Front Immunol*, 2014, **5**, 461.
9. K. A. Kigerl, J. P. de Rivero Vaccari, W. D. Dietrich, P. G. Popovich and R. W. Keane, *Exp. Neurol.*, 2014, **258**, 5-16.
10. Y. K. Kim, J. S. Shin and M. H. Nahm, *Yonsei Med. J.*, 2016, **57**, 5-14.
11. C. G. M., *The Cell: A Molecular Approach*, 2nd edition, Sinauer Associates, Sunderland (MA), 2000.
12. B. JM, T. JL and S. L., *Biochemistry. 5th edition*, W H Freeman, New York, 2002.

13. S. Akira and K. Takeda, *Nat. Rev. Immunol.*, 2004, **4**, 499-511.
14. D. Gao and W. Li, *Proteins*, 2016, DOI: 10.1002/prot.25179.
15. J. H. Lu and P. D. Sun, *Sci. Signal.*, 2012, **5**.
16. S. I. Yoon, O. Kurnasov, V. Natarajan, M. S. Hong, A. V. Gudkov, A. L. Osterman and I. A. Wilson, *Science*, 2012, **335**, 859-864.
17. M. R. de Zoete, A. M. Kestra, J. A. Wagenaar and J. P. M. van Putten, *J. Biol. Chem.*, 2010, **285**, 12149-12158.
18. K. Zhou, R. Kanai, P. Lee, H. W. Wang and Y. Modis, *J. Struct. Biol.*, 2012, **177**, 402-409.
19. D. D. Chaplin, *J. Allergy Clin. Immunol.*, 2010, **125**, S3-23.
20. S. Y. Kim, X. T. Thanh, K. Jeong, S. B. Kim, S. O. Pan, C. H. Jung, S. H. Hong, S. E. Lee and J. H. Rhee, *Infect. Immun.*, 2014, **82**, 29-42.
21. S. S. Yoon and J. J. Mekalanos, *Infect. Immun.*, 2008, **76**, 1282-1288.
22. K. Yonekura, S. Maki-Yonekura and K. Namba, *Nature*, 2003, **424**, 643-650.
23. F. A. Samatey, H. Matsunami, K. Imada, S. Nagashima, T. R. Shaikh, D. R. Thomas, J. Z. Chen, D. J. DeRosier, A. Kitao and K. Namba, *Nature*, 2004, **431**, 1062-1068.
24. F. A. Samatey, K. Imada, S. Nagashima, F. Vonderviszt, T. Kumasaka, M. Yamamoto and K. Namba, *Nature*, 2001, **410**, 331-337.
25. J. I. Godfroy, 3rd, M. Roostan, Y. S. Moroz, I. V. Korendovych and H. Yin, *Plos One*, 2012, **7**, e48875.
26. B. M. Buchholz and A. J. Bauer, *Neurogastroent Motil*, 2010, **22**, 232-245.
27. N. D. Temperley, S. Berlin, I. R. Paton, D. K. Griffin and D. W. Burt, *BMC genomics*, 2008, **9**.
28. M. S. Hong, S. I. Yoon and I. A. Wilson, *Biochem Bioph Res Co*, 2012, **427**, 119-124.
29. K. G. K. Murthy, A. Deb, S. Goonesekera, C. Szabo and A. L. Salzman, *J. Biol. Chem.*, 2004, **279**, 5667-5675.
30. V. Forstneric, K. Ivicak-Kocjan, T. Plaper, R. Jerala and M. Bencina, *PLOS Pathog.*, 2017, **13**.
31. N. A. Hynes, C. Furnes, B. N. Fredriksen, T. Winther, J. Bogwald, A. N. Larsen and R. A. Dalmo, *Vaccine*, 2011, **29**, 7678-7687.

32. L. M. Harrison, P. Rallabhandi, J. Michalski, X. Zhou, S. R. Steyert, S. N. Vogel and J. B. Kaper, *Infect. Immun.*, 2008, **76**, 5524-5534.
33. S. S. Yoon and J. J. Mekalanos, *Infect. Immun.*, 2008, **76**, 1282-1288.
34. S. das, unpublished.
35. M. C. Peeters, G. J. van Westen, D. Guo, L. E. Wisse, C. E. Muller, M. W. Beukers and A. P. Ijzerman, *FASEB J.*, 2011, **25**, 632-643.
36. M. Wheatley, D. Wootten, M. T. Conner, J. Simms, R. Kendrick, R. T. Logan, D. R. Poyner and J. Barwell, *Br J Pharmacol*, 2012, **165**, 1688-1703.
37. A. Das, J. Chakrabarti and M. Ghosh, *Biophys. J.*, 2013, **104**, 1274-1284.
38. S. Sikdar, J. Chakrabarti and M. Ghosh, *Mol. BioSyst.*, 2014, **10**, 3280-3289.
39. V. K. Vyas, R. D. Ukawala, M. Ghate and C. Chinthia, *Indian J Pharm Sci.*, 2012, **74**, 1-17.
40. L. Feng, P. R. Reeves, R. T. Lan, Y. Ren, C. X. Gao, Z. M. Zhou, Y. Ren, J. S. Cheng, W. Wang, J. M. Wang, W. B. Qian, D. Li and L. Wang, *Plos One*, 2008, **3**.
41. J. Haas, S. Roth, K. Arnold, F. Kiefer, T. Schmidt, L. Bordoli and T. Schwede, *Database : the journal of biological databases and curation*, 2013, **2013**, bat031.
42. K. Arnold, L. Bordoli, J. Kopp and T. Schwede, *Bioinformatics*, 2006, **22**, 195-201.
43. L. J. McGuffin, J. D. Atkins, B. R. Salehe, A. N. Shuid and D. B. Roche, *Nucleic Acids Res.*, 2015, **43**, W169-173.
44. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg, *Nat. Protocols*, 2015, **10**, 845-858.
45. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781-1802.
46. G. C. P. van Zundert, J. P. G. L. M. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastritis, E. Karaca, A. S. J. Melquiond, M. van Dijk, S. J. de Vries and A. M. J. J. Bonvin, *J. Mol. Biol.*, 2016, **428**, 720-725.
47. S. J. de Vries, M. van Dijk and A. M. Bonvin, *Nat. Protoc.*, 2010, **5**, 883-897.
48. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701-1718.
49. C. Kandt, W. L. Ash and D. P. Tieleman, *Methods*, 2007, **41**, 475-488.

50. S. J. Ippolito, K. Kalantar-Zadeh, W. Wlodarski, K. Galatsis, W. J. Fischer, O. Berger and H. Stab, *Commad 2002 Proceedings*, 2002, 165-168.
51. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701-1718.
52. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
53. K. Hinsen, *Proteins*, 1998, **33**, 417-429.
54. V. P. Jaakola, M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. Chien, J. R. Lane, A. P. Ijzerman and R. C. Stevens, *Science*, 2008, **322**, 1211-1217.
55. F. Xu, H. Wu, V. Katritch, G. W. Han, K. A. Jacobson, Z. G. Gao, V. Cherezov and R. C. Stevens, *Science*, 2011, **332**, 322-327.
56. A. Chu, J. S. Caldwell and Y. A. Chen, *Mol. Pharmacol.*, 2010, **77**, 95-101.
57. A. Couvineau, E. Ceraudo, Y. V. Tan, P. Nicole and M. Laburthe, *Front. Endocrinol.*, 2012, **3**, 139.
58. A. Couvineau, E. Ceraudo, Y. V. Tan and M. Laburthe, *Neuropeptides*, 2010, **44**, 127-132.
59. S. A. Hollingsworth and P. A. Karplus, *Biomol Concepts*, 2010, **1**, 271-283.
60. S. A. Smith, O. C. Jann, D. Haig, G. C. Russell, D. Werling, E. J. Glass and R. D. Emes, *BMC Evol. Biol.*, 2012, **12**, 122.

CHAPTER 6

Model studies on aggregation of misfolded proteins

6.1. Introduction:

We observe that temporal correlations between long distant residues of a protein, electrostatically heterogeneous surface as well as conformational thermodynamics of the protein in different medium play vital role in governing protein functions. Proteins are amphiphilic¹ bio-molecules, having both hydrophobic and hydrophilic parts. The basic building blocks of proteins are amino acids²⁻⁴ (residues) which have different side chains, giving hydrophobic or hydrophilic character. In general, functional proteins expose their hydrophilic groups to aqueous environment⁵, while hydrophobic parts remain buried⁶ in their native folded conformations. However, sometimes due to mutation, physiological stress and changes in physio-chemical conditions proteins undergo deviations from native structures, exposing the hydrophobic parts to water^{7, 8}. These misfolded proteins form supra-molecular assembly structures like other amphiphilic complex assembly of nm size⁹⁻¹². In this chapter, we primarily discuss how the different competing interactions lead formation of such supra-molecules using a model system.

These complex morphologies of misfolded proteins often assemble into nano-tubes, nano-vesicles, nano-spheres, nano-fibers and nano-doughnut structures¹³. They are widely used in bio-medical applications^{14, 15} due to their inherent biocompatibility and biodegradability¹⁶. However, above a critical aggregation concentration¹⁴ peptide amphiphiles can lead to numerous neurodegenerative disorders^{7, 17}. The potency of toxicity^{18, 19} of these misfolded proteins depends on size of the aggregated structures.

Thus, detail knowledge on the stability of nm sized clusters of misfolded proteins has large implications to understand diseases as well as to design biocompatible materials.

Most common end product of protein aggregation is the extended β -amyloid fibril^{20, 21} having β sheets parallel to fibril axis. Dynamic light scattering and fluorescence correlation spectroscopy (FCS) measurements^{18, 19} reveal that three different major isoforms of amyloid assembly exist in aggregation pathway: monomers, soluble intermediate oligomers and the insoluble long β -sheet like aggregated fibrils. Neither the larger aggregated structure of β -amyloid nor the monomer are responsible for toxicity¹⁸, rather smaller oligomers^{22, 23} of 3-10 nm length scale are the main toxic species of β -amyloid. Earlier experimental studies²⁴ indicate that the hydrophobic parts of the proteins are largely responsible for amyloid formation. However, it also depends on salt concentration and pH of the medium which indicates the role of charged interactions^{24, 25} as well. The experimental observations suggest that aggregation of misfolded protein takes place via competing interactions, namely, that between hydrophobic sites and that of electrostatic origin.

Theoretical studies elucidating the role of the competitive interactions in stabilizing nm scale structures are far from complete. Protein aggregations are associated with slow time and large length scales. Theoretical understanding of mechanism behind aggregation from all atom MD simulation in explicit solvent model is involved²⁶⁻²⁹ due to large number of molecules. Hence, coarse grained models²⁶⁻²⁹ are employed to gain mechanistic insights on protein aggregation. Patchy colloids³⁰⁻³³ with anisotropy in size, shape or charge density of their interacting surface to account for variations over protein surface chemistry, are often chosen to model protein assembly. A patchy spherocylinder³⁴ having two hemispheres with anisotropic hard core repulsion and hydrophobic stripes running through the length is

used to study potential dependent morphology of aggregation³⁴. The model emphasizes that the different widths of non-uniformly distributed hydrophobic surface patches lead formation of aggregates with wide variety of shapes. However, effect of electrostatic interaction in governing this assembly is not included in this model. Moreover, it is hard to generalize³⁵ the model for broad class of proteins. Small angle neutron scattering experiment on lysozyme aggregates³⁶ has been explained using short ranged attraction and long range repulsion. Such model has been widely used^{37, 38} to study colloidal aggregation. The short ranged attraction due to depletion of solvent particles between two solutes is known as Asakura-Oosawa model³⁹, but this does not consider repulsion of solvent by the hydrophobic solute surface. A charged-colloid system²⁴ with hydrophobic potential modeled by step function extended up to radius of the particle along with screened electrostatic interaction has been proposed to understand stability of amyloid. However, it is not clear how the competition of exposed hydrophobic surface of misfolded proteins and the charge interaction has been taken into account in these models.

In the present work, we include both interactions into account in a model system. We take the surface of a spherical particle repelling the solvent molecules to mimic the solvent exposed hydrophobic residues and place charge at the center of the spheres to mimic the buried hydrophilic portion of the protein. The schematic diagram of a pair of such particles is shown in Fig.6.1.

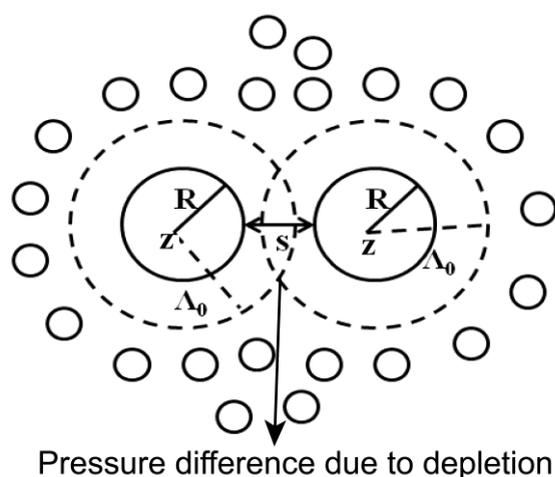


Fig.6.1. Schematic representation for the model system with particles, having solvophobic surface along with charge z at core. The radius of the particle is R . The smaller circles represent solvent molecules repelled away from annular region between R and Δ_0 .

There have been numerous studies⁴⁰⁻⁴² on solvophobic solutes, the solute surface repelling solvent molecules as a model for hydrophobic solutes. These studies show that interaction between two solvophobic solutes is given by a harmonic potential with characteristic spring constant α which depends on the thermodynamic condition of the solvent and strength of repulsion by the solute surface⁴³. The electrostatic interaction between charges at the centers is taken to be screened with Debye screening length⁴⁴ $(\kappa)^{-1}$ by the counter charges and salt ions. One major reason for choosing such model is its simplicity so that analytical treatment is possible. Moreover, surface modification allows synthesizing colloids having various degrees of hydrophobicity with hydrophilic core inside, like ligand capped metal nano particles⁴⁵. We study the system by MC simulations and mean field analysis. Our observations clearly establish competition between solvophobic force mediated attraction and electrostatic repulsion. The particles form finite sized clusters in the presence of electrostatic interaction. The threshold value of κ to form large aggregates show $\alpha \sim \kappa^{-1.5}$ dependence which is qualitatively supported by our mean field analysis.

6.2. Methods:

We consider a model system having electrostatic and hydrophobicity mediated potential. The details of the potential function, simulation protocol and analysis are elaborated in this section.

6.2.1. Details of potential function:

In earlier study⁴³ it is shown that the solute-solvent repulsion generates a gas bubble of radius Λ_0 (shown by dashed line in Fig.6.1.) around the solvophobic particle of radius R , where $\Lambda_0 > R$. When two such solvophobic particles come in proximity, a pressure difference is created at the overlapping zone of the bubbles surrounding the solutes⁴³, which leads to a harmonic attractive force between two solutes of inter-surface separation s . For $s \leq 2\Lambda_0$, the interaction potential, $V_A(s) = -\alpha[s(1 - (\Lambda_0/R)^2) - s^2/2R]$, where $\alpha (= 8\pi\gamma\Lambda_0)$ governs strength of interaction. Here, γ is surface tension of the solvent. We take screened electrostatic repulsion between the core charges by standard Derjaguin–Landau–Verwey–Overbeek (DLVO) potential⁴⁴, $V_C(r) = (z^*e)^2 e^{-\kappa r} / 4\pi\epsilon_0\epsilon r$. Here $z^* = z e^{\kappa R} / (1 + \kappa R)$, r is inter-particle separation, e is electronic charge, z number of charges present in a particle, κ inverse of the Debye screening length, ϵ dielectric constant of solvent and ϵ_0 is electric permittivity of air. $\kappa = (\sum_i z_i^2 c_i e^2 N_A / \epsilon_0 \epsilon k_B T)^{1/2}$, z_i being the valence of each type of i th ion present in the solution, including contribution from salt as well as from the macro-ions, and k_B the Boltzmann constant. Similarly, c_i the ionic concentration in mol/L, $\sum_i z_i^2 c_i$ gives ionic strength (I) and N_A is Avogadro's Number.

6.2.2. Details of MC simulation:

The simulations are performed in a cubic box of dimension $l = 14.0$ nm containing $N = 1000$ model particles with the periodic boundary conditions in all three directions. We consider particle diameter $\sigma (= 1.0$ nm) as the length unit and energy unit $k_B T$ at room temperature ($T = 300$ K). We consider surface tension γ (dimensionless) ~ 0.7 and $\Lambda_0/R \approx 1.27$ as per earlier studies⁴³. Using values of γ and Λ_0 , we obtain that $\alpha^* = \alpha\sigma^2/k_B T = 10.0$. We consider minimum value of $\kappa^* = \kappa\sigma = 2.0$ corresponding to salt concentration in cell^{46, 47}. This minimum value of κ^* is also chosen in such a way that $1/\kappa^* < l^*/2$ to minimize the finite size effects in simulations. We take a fixed value for $z (= 22)$ that is comparable to the number of charged residues of a protein and vary κ which is similar to vary salt concentration in experiments for a given charge at the centre of the colloidal particle.

We update particle position according to Metropolis algorithm⁴⁸. First particles are displaced randomly and then if the change in potential energy in new configuration is less than the old one, then the particles are updated to the new positions. Similarly, the acceptance of the new position is also ensured probabilistically by computing ratio of Maxwell-Boltzmann probability distribution functions of new state with respect to the old one. If the ratio is greater than a random number generated between 0 to 1, then also particles position are updated. After equilibration ($\sim 30,000$ MC step) different quantities are averaged for last 70,000 configurations⁴⁸.

6.2.3. Cluster-size analysis:

In order to identify clusters formed by the particles, we arbitrarily choose i th particle and then calculate distance r_{ij} for all of j th particles ($i \neq j$) with respect to

i th particle over equilibrium trajectories. If r_{ij} is less than a certain distance ($r_{cl} = 1.2\sigma$) then these j th particles are considered to belong to the same cluster as the i th⁴⁸ particle and total number of particles belong to that particular cluster gives cluster size, C_s . The process is repeated for other particles.

6.2.4. Computation of radial distribution functions:

We compute the radial distributions⁴⁸ $g(r)$ of the particles using the standard formula, $g(r) = \langle \frac{n(r, \Delta r)}{N\rho 4\pi r^2 \Delta r} \rangle$. Here r is the separation between two particles, ρ is density of our system and $n(r, \Delta r)$ denotes number of neighbor particles between a spherical shell of radius r and $r + \Delta r$ around the central particle. The angular bracket defines ensemble average over equilibrium configurations.

6.2.5. Mean field analytical treatment:

We use mean field treatment⁴⁹ to understand the region of stability of aggregation structure in our system analytically (Details in Appendix I, Chapter 6). We consider to this end a low wave vector expansion of the direct correlation function of the particles and find the root of the quadratic term.

6.3. Results:

Here, we explain the details analysis on how the strength of electrostatic and solvophobicity mediated attraction in our model potential affect formation of finite size as well as aggregated structure. Similarly, the pair correlation function is also represented in details.

6.3.1. Cluster-size analysis:

Let us first consider the cluster size distribution. We generate distribution of cluster size $P(C_s)$, over equilibrium configurations and compute the mean value (\bar{C}_s). We observe that in presence of solvophobicity mediated attraction only, particles tend to form aggregated structure, along with $\bar{C}_s \sim N$ (data not shown). This is analogous to hydrophobic collapse^{50, 51} known in the literature. Next we consider the effect of the electrostatic repulsion which competes with the solvophobic attraction. We generate $P(C_s)$ for different values of κ^* over a range of α^* . Some representative cases are shown in Figs.6.2.(a)-(c). We find that at $\alpha^* = 3.0$, for low κ^* ($= 2.0$) or at low salt concentration $P(C_s)$ is unimodal (Fig.6.2.(a)) and $\bar{C}_s \sim 2.0$, indicating formation of finite size clusters. If we further increase screening then at $\kappa^* = 6.0$, $P(C_s)$ (Fig.6.2.(b)) is bimodal which signifies the increasing tendency of aggregation. Finally, at $\kappa^* = 13.0$, $P(C_s)$ is primarily unimodal (Fig.6.2.(c)) along with $\bar{C}_s = N$.

Thus, we find that for a fixed value of α^* , above a particular threshold value of κ^* (κ_{th}), $\bar{C}_s \sim N$, indicating aggregation. Below that κ_{th} particles form finite size stable clusters. Thus, finite size clusters are stabilized by electrostatic interaction. For large salt concentration, where the electrostatic interaction is strongly screened, system forms aggregates. The finite clusters are few times particles diameter. Thus, they are of a few nm size for nano particles systems.

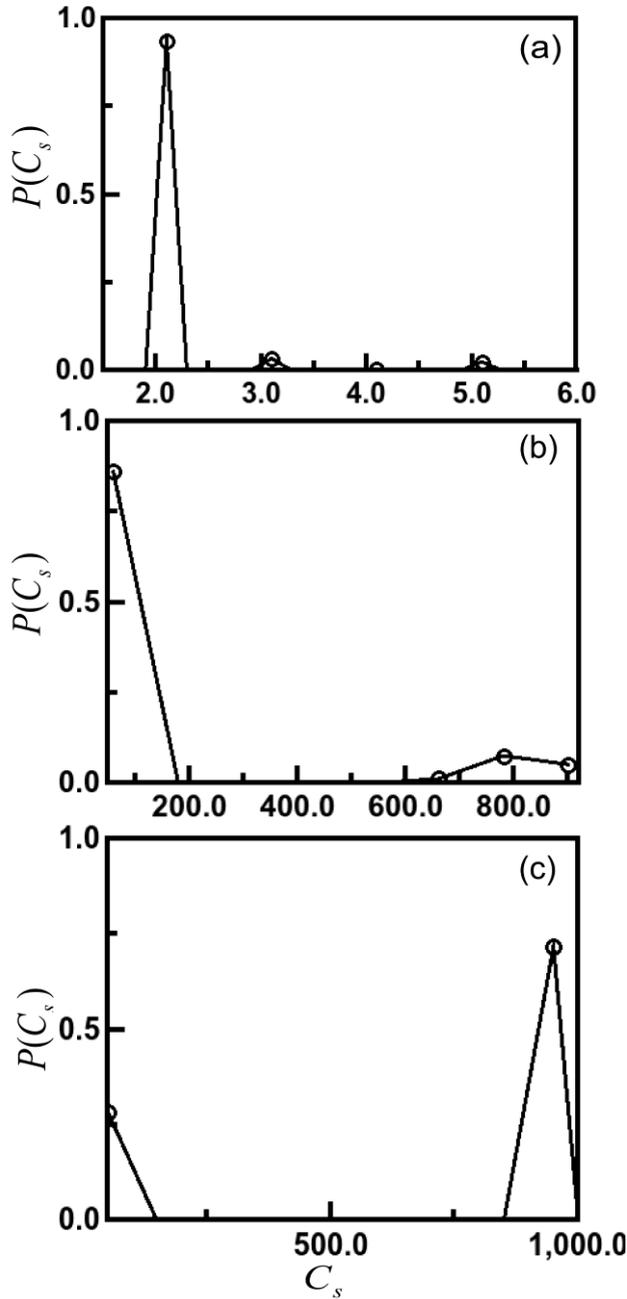


Fig.6.2. Probability distribution of cluster size of the model system, in presence of a fixed solvophobicity mediated attraction ($\alpha^* = 3.0$) and different κ^* . (a) Particles form finite size cluster for $\kappa^* = 2.0$. (b) Coexisting small clusters and aggregation for $\kappa^* = 6.0$. (c) Large aggregation at $\kappa^* = 13.0$.

In order to extract dependence of α^* over κ^* we generate \bar{C}_s with increasing values of κ^* for different α^* . The $\ln\alpha^*$ vs $\ln\kappa^*$ plot in Fig.6.3.(a) shows the aggregation lines. Above the lines the particles form large aggregates, while below it the particles form finite clusters. For lower κ^* , we find from the slope of the log-log plot that $\alpha^* \sim (\kappa_{th}^*)^{-1.5}$. However, for larger κ^* , the dependence is much weaker,

$\alpha^* \sim (\kappa_{th}^*)^{-0.2}$. This is suggestive of the regime dominated by the solvophobicity mediated harmonic potential.

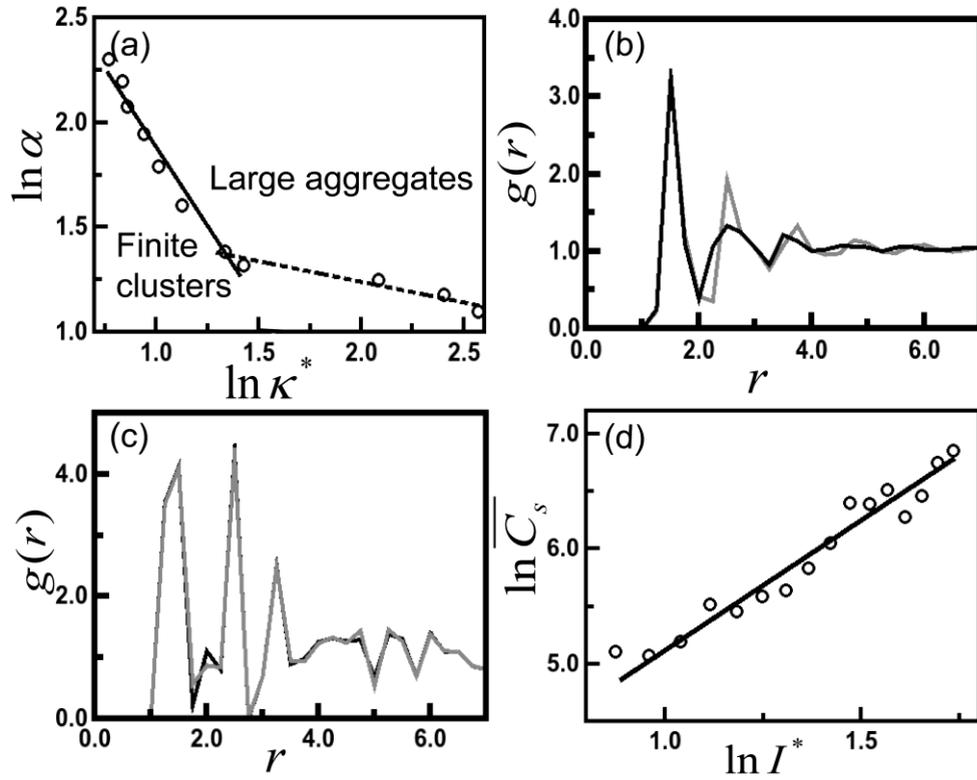


Fig.6.3.(a) Phase diagram in α^* and κ^* in plane showing finite size cluster and large aggregated structure. The solid line shows $\alpha^* \sim (\kappa_{th}^*)^{-1.5}$ dependence, while dashed line represents $\alpha^* \sim (\kappa_{th}^*)^{-0.2}$ for aggregation threshold. (b) Pair correlation function, $g(r)$ vs r plot for different values of α^* and κ^* , Black: $\alpha^* = 10.0, \kappa^* = 2.0$ while Gray: $\alpha^* = 3.0, \kappa^* = 2.0$. and (c) Black: $\alpha^* = 10.0, \kappa^* = 10.0$ where as Gray: $\alpha^* = 3.0, \kappa^* = 13.0$. (d) $\ln \overline{C}_s$ vs $\ln I^*$ plot to show cluster size variation with ionic strength.

6.3.2. Finite correlation in system:

We also analyze the structural correlations in our system using pair correlation function⁴⁹ ($g(r)$). This is the probability distribution of finding a neighbor atom around a central atom within a spherical shell of radii r and $r + \Delta r$. We have shown

some representative cases of $g(r)$ for different values of α^* and κ^* in Figs.6.3.(b) and (c), which indicate that correlations in our systems remain finite.

6.3.3. Analysis from mean field treatment:

Mean field theory⁵² is used to approximate many body interaction of a system by an effective interaction, such that each molecule feels a potential due to a mean field generated by other molecules. Thus, it is an effective mathematical framework to reduce many body problems into one body problem. If the system has negligible spatial fluctuations, then mean field treatment is quite useful. We account for the stability of clusters over aggregations by a simple mean field theory both for electrostatic repulsion and solvophobicity mediated attractions.

Local order or correlation of a liquid can be probed by wave vector (q) dependent static structure factor $s(q)$ using light scattering experiment⁵³, which is Fourier transform of pair correlation function⁴⁹ ($g(r)$). As per Ornstein Zernike equation, direct correlation function⁴⁹ ($c(r)$) between a pair of particles separated by a distance r at a given density ρ depends on all other indirect interactions mediated by a third particle situated at \mathbf{r}' . The equation is $h(r) = c(r) + \int d\mathbf{r}' c(\mathbf{r}' - \mathbf{r})h(r')$, where $h(r) = g(r) - 1$ gives the total correlation between the pair of particles separated by a r . By performing Fourier transform of the equation one can find, $h(q) = c(q)/(1 - \rho c(q))$. In theoretical calculation of liquid system, since $s(q) = 1 + h(q)$, it can be related to $s(q) = 1/(1 - \rho c(q))$. In general, $s(q)$ has peak around ordering wave vector $q_0\sigma = 2\pi$. Clustering is represented by peak at low $q < q_0$, while for aggregation the peak shifts to $q \rightarrow 0$ limit. The peak of $s(q)$ corresponds to that of $c(q)$ as well. At low density, the mean field approximation⁵⁴ represents that $c(r) \approx -\beta u(r)$, where $\beta = 1/k_B T$ and $u(r)$ interaction potential.

For the mean field approximation for our system we consider that $C(q) = C_{electro}(q) + C_{solvo}(q)$. Here $C_{electro}(q)$ is Fourier transform of direct correlation function for DLVO potential and $C_{solvo}(q)$ is that for solvophobic term. This is valid for long-ranged electrostatic potential, namely low κ . Since, the solvophobicity mediated term operates till $s \leq 2\Lambda_0$, we use mean field treatment for this term. By transforming the electrostatic DLVO interaction into Fourier space we get that at long wavelength limit ($q \rightarrow 0$), $C_{electro}(q) = \frac{-4\beta X\sqrt{2\pi}}{\kappa^2}(1 - q^2/\kappa^2)$. Here, $X = (z^2 e^2 e^{2\kappa R}/[1 + \kappa R]^2) 1/4\pi\epsilon_0\epsilon$, the pre-factor of DLVO potential. On the other hand, by performing Fourier transform in the same wave vector range on $C_{solvo}(s)$, $C_{solvo}(q) = \sqrt{2\pi} (-\beta\alpha/R) [\{\frac{32\Lambda_0^5}{5} - \frac{\sigma^5}{5} - A^2[\frac{8\Lambda_0^3}{3} - \frac{\sigma^3}{3}]\} + \{-\frac{128q^2\Lambda_0^7}{42} + \frac{q^2\sigma^7}{42} - A^2[-\frac{32q^2\Lambda_0^5}{30} + \frac{q^2\sigma^5}{30}]\}]$, where $(\Lambda_0^2/R - R) = A$. In $q \rightarrow 0$ limit, $C(q) \approx C_0 + C_2 q^2$, where $C_2 = (p/\kappa^4 - \alpha g)$. $C_0 = \sqrt{2\pi} (-\beta\alpha/R) \{\frac{32\Lambda_0^5}{5} - \frac{\sigma^5}{5} - A^2[\frac{8\Lambda_0^3}{3} - \frac{\sigma^3}{3}]\} - 4\beta X\sqrt{2\pi}/\kappa^2$, $p = 4\beta X\sqrt{2\pi}$ and $g = [\sqrt{2\pi} (-\beta/R) \frac{128\Lambda_0^7}{42}]$ are constants. We find that $C(q)$ has a minimum at $q = 0$ if $C_2 > 0$ and a maximum at $q = 0$, if $C_2 < 0$. Hence, the condition for aggregation is $C_2 = 0$, which yields that $\alpha \sim \kappa^{-4.0}$. The mean field analysis reveals that the stability of the aggregated phase is in qualitative agreement to that obtained from our numerical simulation for low κ^* . This is not surprising for the mean field treatment is valid for longer ranged potential. However, in this analysis we overestimate stability of the aggregated phase which could be due to mean field nature of the analysis which ignores fluctuations.

6.4. Discussions:

Let us now compare our results to existing studies on protein aggregation. FCS study²⁴ on soluble aggregates formed by two Alzheimer's amyloid β peptides ($A\beta_{1-40}$ and $A\beta_{1-42}$) indicates that size of the clusters increases monotonously with increasing ionic strength of the solution by adding NaCl. Scattering measurement on β -amyloid shows that salts of divalent metal ions at micro-molar concentration can prevent²⁵ formation of this oligomeric population by leading them precipitated to higher size assembly. These observations are qualitatively consistent with our results. We find from our simulation that for a fixed value of α^* ($= 3.0$), $\bar{C}_s \sim (I^*)^{2.3}$ below κ_{th}^* . This is shown in $\ln\bar{C}_s$ vs $\ln I^*$ plot in Fig.6.3.(d). This result is in qualitative agreement to the reported experimental observations²⁴ on β amyloid, where mean size of the aggregated cluster exhibits super linear dependence on ionic strength of the solution.

In our model system probability of finding a pair of harmonically bound solvophobic particles at a inter-surface separation s is $\sim e^{-\frac{1}{2}\alpha s^2}$. The length scale of bound pair is $\sim \alpha^{-1/2}$, which can be thought of as length scale of attraction. The range for screened electrostatic repulsion is given by κ^{-1} . We show in Table.6.1. $(\alpha^*)^{-1/2}$ and corresponding κ_{th}^{-1} . We find that $\kappa_{th}^{-1} > (\alpha^*)^{-1/2}$ for large α^* ; otherwise, $\kappa_{th}^{-1} < (\alpha^*)^{-1/2}$. Thus our model fit over a limited range of parameter values with the standard model of colloidal aggregation^{37, 38}, having attraction shorter ranged than repulsion. This difference can be traced to the nature of solvent depletion in our system. Here depletion is governed by the thermodynamics of the system⁴³. The range of depletion can be varied⁴³ by proximity to liquid-vapour transition line where vapour is stabilized near the solvophobic surface. It has been shown in the earlier

work⁴³ that α has nontrivial dependence on temperature with a maximum at $T = T_0$, so that in the vicinity of this temperature, $\alpha^* = \alpha_0^* - \beta(T - T_0)^2$. Our analysis indicates that $\kappa_{th}[\sim (\alpha^*)^{-1/4}]$ will have a minimum at $T = T_0$. This means a nontrivial feature of stability of finite sized clusters: the finite clusters are stable on both ends excluding the neighborhood of the maximum point.

Table.6.1. Range of solvophobicity mediated attraction and electrostatic interactions for different values of α^* and corresponding κ_{th} .

α^*	$\alpha^{*-1/2}$	κ_{th}	κ_{th}^{-1}
10.0	0.32	2.0	0.50
6.0	0.41	3.0	0.33
4.0	0.50	4.0	0.25
3.0	0.58	13.0	0.08

The formation of large aggregates by increasing salt concentration particularly in the low α^* regime is analogous to salting out effect known in many proteins^{55, 56}. Metal nano particles capped with ligands⁴⁵ can be a realization of the model studied here. Such capped nano particles form nm sized clusters which are having immense technological importance^{57, 58}, mainly in the fields of nano medicine, optics, catalysis, sensor. In particular our analysis may be useful guide to tune the cluster sizes. Apart from misfolded proteins, there are other molecular assemblies of soft matter systems where hydrophobic parts are exposed and hydrophilic groups remain buried. Surface modification of halloysite nano-tube clay by proper surfactant leads to fabrication of inorganic reverse micelle with hydrophobic shell and hydrophilic cavity⁵⁹. These tubular self-assembled structures can be used for entrapment of antibacterial agent and in pharmaceutical application⁵⁹. Similarly, amphiphilic organic molecule and clay

minerals nano-platelets (CNP) like montmorillonite and laponite can assemble⁶⁰ into 2D nano-composite hydrogel. This assembly⁶⁰ with hydrophobic surface and hydrophilic interlayer structure has potential application in biomedical and tissue engineering. The critical balance between underlying hydrophobic and electrostatic interaction towards stability of these supramolecular materials is important in governing their functionalities. Our model can be generalized to include molecular shape anisotropy relevant for the nm sized supra-molecular assemblies in the system by taking separate charged site, connected to the hydrophobic site. However, given the interactions between charged and hydrophobic sites are weak, the basic results of our simplified model should hold good.

6.5. Conclusions:

To conclude, we have studied role of solvophobicity mediated attraction and electrostatic repulsion in governing aggregation of misfolded protein. Our model suggests that by tuning salt concentration of the solution, finite size clusters can be stabilized. We propose that our model can be guide to molecular control over amphiphilic self-assembly of nm size.

Appendix I

Details of mean field treatment:

Using Ornstein-Zernike equation under MSA approximation we obtain that direct correlation function for electrostatic screening is,

$$\begin{aligned}
 C_{electro}(r) &= -\beta V_C(r) \\
 &= -\beta(z^2 e^2 e^{2\kappa R} / [1 + \kappa R]^2) e^{-\kappa r} / 4\pi\epsilon_0\epsilon r \\
 &= -\beta X e^{-\kappa r} / r \quad (1).
 \end{aligned}$$

Assuming, $X = (z^2 e^2 e^{2\kappa R} / [1 + \kappa R]^2) 1/4\pi\epsilon_0\epsilon$.

By performing Fourier Transform of equation (1) we find that,

$$\begin{aligned}
 C_{electro}(q) &= (-\beta X / \sqrt{2\pi}) \int_{-\infty}^{+\infty} e^{-iqr} (e^{-\kappa r} / r) r^2 dr \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi \\
 &= (-\beta X 2\pi / \sqrt{2\pi}) \int_{-\infty}^{+\infty} e^{-iqr \cos \theta} e^{-\kappa r} r dr \int_0^\pi \sin \theta d\theta \quad (2).
 \end{aligned}$$

Assuming , $\cos \theta = z$ in equation (2) we have

$$\begin{aligned}
 C_{electro}(q) &= -\beta X \sqrt{2\pi} \int_{-\infty}^{+\infty} e^{-iqrz} e^{-\kappa r} r dr \int_{-1}^{+1} dz \\
 &= -\beta X \sqrt{2\pi} \int_{-\infty}^{+\infty} e^{-\kappa r} r dr \int_{-1}^{+1} e^{-iqrz} dz \\
 &= -\beta X \sqrt{2\pi} \int_{-\infty}^{+\infty} e^{-\kappa r} r dr \frac{1}{-iqr} [e^{-iqrz}]_{-1}^{+1} \\
 &= -\beta X \sqrt{2\pi} \int_{-\infty}^{+\infty} e^{-\kappa r} r dr \frac{1}{iqr} [e^{iqr} - e^{-iqr}] \\
 &= (-2\beta X \sqrt{2\pi} / iq) \int_0^{+\infty} e^{r(iq - \kappa)} - e^{-r(iq + \kappa)} dr
 \end{aligned}$$

$$\begin{aligned}
&= (-2\beta X\sqrt{2\pi}/iq)[e^{r(iq-\kappa)}/(iq-\kappa) + e^{-r(iq+\kappa)}/(iq+\kappa)]_0^{+\infty} \\
&= (-2\beta X\sqrt{2\pi}/iq)[-1/(iq-\kappa) - 1/(iq+\kappa)] \\
&= (2\beta X\sqrt{2\pi}/iq)[1/(iq-\kappa) + 1/(iq+\kappa)] \\
&= (-2\beta X\sqrt{2\pi}/iq)[2iq/(q^2 + \kappa^2)] \\
&= \frac{-4\beta X\sqrt{2\pi}}{\kappa^2}(1 + q^2/\kappa^2)^{-1} \tag{3}.
\end{aligned}$$

Putting $q \rightarrow 0$ in equation (3) we get,

$$C_{electro}(q) = (-4\beta X\sqrt{2\pi}/\kappa^2)(1 - q^2/\kappa^2) \tag{4}.$$

For solvophobicity mediated attraction we consider,

$$\begin{aligned}
V_A(s) &= -\alpha[s(1 - (\Lambda_0/R)^2) - s^2/2R] \\
&= (\alpha/2R)[s^2 + 2Rs\{(\Lambda_0/R)^2 - 1\}] \\
&= (\alpha/2R)[s^2 + 2s\{\frac{\Lambda_0^2}{R} - R\}] \\
&= \{(\alpha/2R)[s^2 + 2sA + A^2]\} - (\alpha A^2/2R) \text{ (assuming } (\Lambda_0^2/R - R) = A) \\
&= (\alpha/2R)[(s + A)^2] - (\alpha A^2/2R).
\end{aligned}$$

Using, $s' = s + A$, we get,

$$V_A(s) = (\alpha s'^2/2R) - (\alpha A^2/2R) \tag{5}.$$

Using equation (5) we find that direct correlation function for solvophobicity mediated attraction is,

$$C_{solvo}(s) = -\beta V_A(s)$$

$$= (-\beta \alpha s'^2/2R) + (\beta \alpha A^2/2R) \quad (6).$$

By performing Fourier transform of equation (6) we obtain,

$$\begin{aligned} C_{hydro}(q) &= (1/\sqrt{2\pi}) \int_{\sigma}^{2\Lambda_0} ([-\beta \alpha s'^2/2R] + [\beta \alpha A^2/2R]) e^{-iq.s' s'^2} ds' \int_0^{\pi} \sin \theta d\theta \int_0^{2\pi} d\varphi \\ &= [1/\sqrt{2\pi}] \int_{\sigma}^{2\Lambda_0} (-\beta \alpha s'^2/2R) e^{-iq.s' s'^2} ds' \int_0^{\pi} \sin \theta d\theta \int_0^{2\pi} d\varphi \\ &\quad + [1/\sqrt{2\pi}] \int_{\sigma}^{2\Lambda_0} (\beta \alpha A^2/2R) e^{-iq.s' s'^2} ds' \int_0^{\pi} \sin \theta d\theta \int_0^{2\pi} d\varphi \\ &= C_1 + C_2 \quad (7). \end{aligned}$$

Next, we solve equation (7) part wise and find that,

$$\begin{aligned} C_1 &= [1/\sqrt{2\pi}] \int_{\sigma}^{2\Lambda_0} (-\beta \alpha s'^2/2R) e^{-iq.s' s'^2} ds' \int_0^{\pi} \sin \theta d\theta \int_0^{2\pi} d\varphi \\ &= [2\pi/\sqrt{2\pi}] \int_{\sigma}^{2\Lambda_0} (-\beta \alpha s'^2/2R) e^{-iqs' \cos \theta s'^2} ds' \int_0^{\pi} \sin \theta d\theta \end{aligned}$$

Assuming , $\cos \theta = z$ we have

$$\begin{aligned} &= [2\pi/\sqrt{2\pi}] \int_{\sigma}^{2\Lambda_0} (-\beta \alpha s'^2/2R) s'^2 ds' \int_{-1}^{+1} e^{-iqs'z} dz \\ &= [2\pi/\sqrt{2\pi}] \int_{\sigma}^{2\Lambda_0} (-\beta \alpha s'^2/2R) s'^2 ds' \frac{1}{-iqs'} [e^{-iqs'z}]_{-1}^{+1} \\ &= \sqrt{2\pi} \int_{\sigma}^{2\Lambda_0} (-\beta \alpha s'^2/2R) s'^2 ds' \frac{1}{iqs'} [e^{iqs'} - e^{-iqs'}] \\ &= \sqrt{2\pi} \int_{\sigma}^{2\Lambda_0} (-\beta \alpha s'^2/2R) s'^2 ds' \frac{1}{iqs'} \times 2i \sin qs' \end{aligned}$$

$$= \sqrt{2\pi} (-\beta \alpha/R) \int_{\sigma}^{2\Lambda_0} s'^4 (\sin qs'/qs') ds'$$

In $q \rightarrow 0$,

$$\begin{aligned} C_1 &= \sqrt{2\pi} (-\beta \alpha/R) \int_{\sigma}^{2\Lambda_0} s'^4 ([qs' - \frac{q^3 s'^3}{3!}]/qs') ds' \\ &= \sqrt{2\pi} (-\beta \alpha/R) \int_{\sigma}^{2\Lambda_0} s'^4 (1 - q^2 s'^2/3!) ds' \\ &= \sqrt{2\pi} (-\alpha\beta/R) [\frac{s'^5}{5} - \frac{q^2 s'^7}{7 \times 3!}]_{\sigma}^{2\Lambda_0} \quad (8). \end{aligned}$$

$$\begin{aligned} C_2 &= [1/\sqrt{2\pi}] \int_{\sigma}^{2\Lambda_0} (\beta \alpha A^2/2R) e^{-iq.s' s'^2} ds' \int_0^{\pi} \sin \theta d\theta \int_0^{2\pi} d\varphi \\ &= \sqrt{2\pi} \int_{\sigma}^{2\Lambda_0} (\beta \alpha A^2/2R) s'^2 ds' \frac{1}{iqs'} \times 2i \sin qs' \end{aligned}$$

In $q \rightarrow 0$,

$$\begin{aligned} &= \sqrt{2\pi} (\beta \alpha A^2/R) \int_{\sigma}^{2\Lambda_0} s'^2 (1 - q^2 s'^2/3!) ds' \\ &= \sqrt{2\pi} (\beta \alpha A^2/R) [\frac{s'^3}{3} - \frac{q^2 s'^5}{5 \times 3!}]_{\sigma}^{2\Lambda_0} \quad (9). \end{aligned}$$

Using equations (8) and (9) in equation (7) we find,

$$\begin{aligned} C_{solvo}(q) &= \sqrt{2\pi} (-\beta \alpha/R) ([\frac{s'^5}{5} - \frac{q^2 s'^7}{7 \times 3!}]_{\sigma}^{2\Lambda_0} - A^2 [\frac{s'^3}{3} - \frac{q^2 s'^5}{5 \times 3!}]_{\sigma}^{2\Lambda_0}) \\ &= \sqrt{2\pi} (-\beta \alpha/R) \{ [\frac{32\Lambda_0^5}{5} - \frac{128q^2 \Lambda_0^7}{42} - \frac{\sigma^5}{5} + \frac{q^2 \sigma^7}{42}] \\ &\quad - A^2 [\frac{8\Lambda_0^3}{3} - \frac{32q^2 \Lambda_0^5}{30} - \frac{\sigma^3}{3} + \frac{q^2 \sigma^5}{30}] \} \end{aligned}$$

$$\begin{aligned}
&= \sqrt{2\pi} (-\beta\alpha/R) \left\{ \frac{32\Lambda_0^5}{5} - \frac{\sigma^5}{5} - A^2 \left[\frac{8\Lambda_0^3}{3} - \frac{\sigma^3}{3} \right] \right\} + \\
&\sqrt{2\pi} (-\beta\alpha/R) \left\{ -\frac{128q^2\Lambda_0^7}{42} + \frac{q^2\sigma^7}{42} - A^2 \left[-\frac{32q^2\Lambda_0^5}{30} + \frac{q^2\sigma^5}{30} \right] \right\} \quad (10).
\end{aligned}$$

Using equation (4) and equation (10) we obtain that, total component of Fourier transform of direct correlation function is,

$$\begin{aligned}
C(q) &= [-4\beta X\sqrt{2\pi}/\kappa^2 (1 - q^2/\kappa^2)] + \sqrt{2\pi} (-\beta\alpha/R) \left\{ \frac{32\Lambda_0^5}{5} - \frac{\sigma^5}{5} - A^2 \left[\frac{8\Lambda_0^3}{3} - \frac{\sigma^3}{3} \right] \right\} + \\
&\sqrt{2\pi} (-\beta\alpha/R) \left\{ -\frac{128q^2\Lambda_0^7}{42} + \frac{q^2\sigma^7}{42} - A^2 \left[-\frac{32q^2\Lambda_0^5}{30} + \frac{q^2\sigma^5}{30} \right] \right\} \\
&= (-4\beta X\sqrt{2\pi}/\kappa^2) + (4\beta X\sqrt{2\pi} q^2/\kappa^4) + \sqrt{2\pi} (-\beta\alpha/R) \left\{ \frac{32\Lambda_0^5}{5} - \frac{\sigma^5}{5} - A^2 \left[\frac{8\Lambda_0^3}{3} - \frac{\sigma^3}{3} \right] \right\} + \\
&\sqrt{2\pi} (-\beta\alpha/R) \left\{ -\frac{128q^2\Lambda_0^7}{42} + \frac{q^2\sigma^7}{42} - A^2 \left[-\frac{32q^2\Lambda_0^5}{30} + \frac{q^2\sigma^5}{30} \right] \right\}
\end{aligned}$$

Assuming, $C_0 = \sqrt{2\pi} (-\beta\alpha/R) \left\{ \frac{32\Lambda_0^5}{5} - \frac{\sigma^5}{5} - A^2 \left[\frac{8\Lambda_0^3}{3} - \frac{\sigma^3}{3} \right] \right\} - 4\beta X\sqrt{2\pi}/\kappa^2$ we

find

$$\begin{aligned}
C(q) &= \\
C_0 &+ q^2(4\beta X\sqrt{2\pi}/\kappa^4 + [\sqrt{2\pi} (-\beta\alpha/R) \left\{ -\frac{128\Lambda_0^7}{42} + \frac{\sigma^7}{42} - A^2 \left(-\frac{32\Lambda_0^5}{30} + \frac{\sigma^5}{30} \right) \right\}]) \quad (11).
\end{aligned}$$

In our calculation $A = (\Lambda_0^2/R - R) = 0.61R \sim 0.31 \text{ nm}$, $\sigma = 1.0 \text{ nm}$, using these values in equation (11) we obtain,

$$\begin{aligned}
C(q) &= C_0 + q^2(4\beta X\sqrt{2\pi}/\kappa^4 \\
&\quad + [\sqrt{2\pi}(-\beta\alpha/R)\{-\frac{128\Lambda_0^7}{42} + \frac{1}{42} + \frac{3.07\Lambda_0^5}{30} - \frac{0.09}{30}\}]) \\
&= C_0 + q^2(4\beta X\sqrt{2\pi}/\kappa^4 + [\sqrt{2\pi}(-\beta\alpha/R)\{-\frac{128\Lambda_0^7}{42} + \frac{3.07\Lambda_0^5}{30} + 0.02\}])
\end{aligned}
\tag{12}$$

Considering dominating term of Λ_0 of equation (12) we get,

$$C(q) \approx C_0 + q^2(4\beta X\sqrt{2\pi}/\kappa^4 - [\sqrt{2\pi}(-\beta\alpha/R)\frac{128\Lambda_0^7}{42}])$$

Considering constant terms $4\beta X\sqrt{2\pi} = p$ and $[\sqrt{2\pi}(-\beta/R)\frac{128\Lambda_0^7}{42}] = g$, we obtain,

$$C(q) \approx C_0 + q^2(p/\kappa^4 - \alpha g) \tag{13}.$$

We can see from equation (13) that condition of aggregation can be found by putting coefficient of $q^2 = 0$, that leads

$$p/\kappa^4 = \alpha g \tag{14}.$$

Thus, from equation (14), we find that at aggregation $\alpha \sim \kappa^{-4.0}$.

6.6. References:

1. C. Wang, Z. Wang and X. Zhang, *Acc. Chem. Res.*, 2012, **45**, 608-618.
2. B. JM, T. JL and S. L, *Biochemistry. 5th edition*, W H Freeman, New York, 2002.
3. C. G. M, *The Cell: A Molecular Approach, 2nd edition*, Sinauer Associates, Sunderland (MA), 2000.
4. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell, 4th edition*, Garland Science, New York, 2007.

5. I. Gitlin, J. D. Carbeck and G. M. Whitesides, *Angew. Chem. Int. Ed.*, 2006, **45**, 3022-3060.
6. C. Camilloni, D. Bonetti, A. Morrone, R. Giri, C. M. Dobson, M. Brunori, S. Gianni and M. Vendruscolo, *Sci. Rep.*, 2016, **6**, 28285.
7. V. Vetri and V. Fodera, *Febs. Lett.*, 2015, **589**, 2448-2463.
8. J. Beerten, J. Schymkowitz and F. Rousseau, *Curr. Top. Med. Chem.*, 2012, **12**, 2470-2478.
9. D. Lombardo, M. A. Kiselev, S. Magazù and P. Calandra, *Adv. Cond. Matter Phys.*, 2015, **2015**, 22.
10. S. Ma, Y. Hu and R. Wang, *Macromolecules*, 2015, **48**, 3112-3120.
11. P. Calandra, D. Caschera, V. T. Liveri and D. Lombardo, *Colloids Surf. A*, 2015, **484**, 164-183.
12. N. Kovalchuk, V. Starov, P. Langston and N. Hilal, *Adv. Colloid Interface Sci.*, 2009, **147-148**, 144-154.
13. D. Mandal, A. Nasrolahi Shirazi and K. Parang, *Org. Biomol. Chem.*, 2014, **12**, 3544-3561.
14. A. Dehsorkhi, V. Castelletto and I. W. Hamley, *J. Pept. Sci.*, 2014, **20**, 453-467.
15. Y. Chen, H. X. Gan and Y. W. Tong, *Macromolecules*, 2015, **48**, 2647-2653.
16. H. G. Cui, M. J. Webber and S. I. Stupp, *Biopolymers*, 2010, **94**, 1-18.
17. T. K. Karamanos, A. P. Kalverda, G. S. Thompson and S. E. Radford, *Progress in nuclear magnetic resonance spectroscopy*, 2015, **88-89**, 86-104.
18. U. Sengupta, A. N. Nilson and R. Kaye, *Ebiomedicine*, 2016, **6**, 42-49.
19. P. Cizas, R. Budvytyte, R. Morkuniene, R. Moldovan, M. Broccio, M. Losche, G. Niaura, G. Valincius and V. Borutaite, *Arch. Biochem. Biophys.*, 2010, **496**, 84-92.
20. A. Morriss-Andrews and J. E. Shea, *Annu. Rev. Phys. Chem.*, 2015, **66**, 643-666.
21. M. Landreh, M. R. Sawaya, M. S. Hipp, D. S. Eisenberg, K. Wuthrich and F. U. Hartl, *J. Intern. Med.*, 2016, **280**, 164-176.
22. B. O'Nuallain, D. B. Freir, A. J. Nicoll, E. Risse, N. Ferguson, C. E. Herron, J. Collinge and D. M. Walsh, *J. Neurosci.*, 2010, **30**, 14411-14419.
23. J. M. Mc Donald, T. T. O'Malley, W. Liu, A. J. Mably, G. Brinkmalm, E. Portelius, W. M. Wittbold, 3rd, M. P. Frosch and D. M. Walsh, *Alzheimers. Dement.*, 2015, **11**, 1286-1305.
24. B. Sahoo, S. Nag, P. Sengupta and S. Maiti, *Biophys. J.*, 2009, **97**, 1454-1460.

25. K. Garai, P. Sengupta, B. Sahoo and S. Maiti, *Biochem. Biophys. Res. Commun.*, 2006, **345**, 210-215.
26. A. Morriss-Andrews and J. E. Shea, *J. Phys. Chem. Lett.*, 2014, **5**, 1899-1908.
27. M. A. Blanco, E. Sahin, A. S. Robinson and C. J. Roberts, *J. Phys. Chem. B*, 2013, **117**, 16013-16028.
28. M. Carballo-Pacheco and B. Strodel, *J. Phys. Chem. B*, 2016, **120**, 2991-2999.
29. C. Wu and J. E. Shea, *Curr. Opin. Struct. Biol.*, 2011, **21**, 209-220.
30. E. Bianchi, R. Blaak and C. N. Likos, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6397-6410.
31. E. Bianchi, B. Capone, I. Coluzza, L. Rovigatti and P. D. J. van Oostrum, *Phys. Chem. Chem. Phys.*, 2017, **19**, 19847-19868.
32. G. R. Yi, D. J. Pine and S. Sacanna, *J. Phys. Condens. Matter*, 2013, **25**.
33. S. James, M. K. Quinn and J. J. McManus, *Phys. Chem. Chem. Phys.*, 2015, **17**, 5413-5420.
34. R. Vacha and D. Frenkel, *Biophys. J.*, 2011, **101**, 1432-1439.
35. J. J. McManus, P. Charbonneau, E. Zaccarelli and N. Asherie, *Curr. Opin. Colloid Interface Sci.*, 2016, **22**, 73-79.
36. A. Stradner, H. Sedgwick, F. Cardinaux, W. C. Poon, S. U. Egelhaaf and P. Schurtenberger, *Nature*, 2004, **432**, 492-495.
37. M. M. van Schooneveld, V. W. de Villeneuve, R. P. Dullens, D. G. Aarts, M. E. Leunissen and W. K. Kegel, *J. Phys. Chem. B*, 2009, **113**, 4560-4564.
38. Z. Varga and J. Swan, *Soft Matter*, 2016, **12**, 7670-7681.
39. S. Asakura and F. Oosawa, *J. Chem. Phys.*, 1954, **22**, 1255.
40. T. Urbic and K. A. Dill, *Phys. Rev. E*, 2017, **96**.
41. V. R. Hande and S. Chakrabarty, *J. Phys. Chem. B*, 2015, **119**, 11346-11357.
42. B. Chacko, R. Evans and A. J. Archer, *J. Chem. Phys.*, 2017, **146**.
43. J. Chakrabarti and S. Dutta, *Chem. Phys. Lett.*, 2015, **620**, 109-113.
44. E. J. W. Verwey and J. T. G. Overbeek, *Theory of the Stability of Lyophobic Colloids*, Elsevier, New York-Amsterdam, 1948.
45. L. R. Baker, G. Kennedy, J. M. Krier, M. Van Spronsen, R. M. Onorato and G. A. Somorjai, *Catal. Lett.*, 2012, **142**, 1286-1294.

46. J. Anastassopoulou and T. Theophanides, *The role of metal ions in biological systems and medicine*, Kluwer Academic Publishers, D.P. Kessissoglou (ed.), Bioinorganic Chemistry edn., 1995.
47. T. Heimburg, *Thermal Biophysics of Membranes*, Wiley, 2007.
48. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, 2nd edition, Oxford Science Publications, New York, 1989.
49. J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, Fourth Edition, Elsevier, 2013.
50. I. T. Li and G. C. Walker, *Acc. Chem. Res.*, 2012, **45**, 2011-2021.
51. J. B. Udgaonkar, *Arch. Biochem. Biophys.*, 2013, **531**, 24-33.
52. Lubensky. T. C and Chaikin. P. M, *Principles of condensed matter physics*, Cambridge University Press, Cambridge, 1995.
53. R. Shinde, G. Balgi, S. Richter, S. Banerjee, J. Reynolds, J. Pierce and E. Sevick-Muraca, *Appl. Opt.*, 1999, **38**, 197-204.
54. J. L. Lebowitz and J. K. Percus, *Phys. Rev.*, 1966, **144**, 251-258.
55. H. N. Xu, Y. Liu and L. F. Zhang, *Soft Matter*, 2015, **11**, 5926-5932.
56. W. F. Li, R. H. Zhou and Y. G. Mu, *J. Phys. Chem. B*, 2012, **116**, 1446-1451.
57. I. Fratoddi, *Nanomaterials*, 2017, **8**.
58. Z. Q. Niu and Y. D. Li, *Chem Mater*, 2014, **26**, 72-83.
59. G. Cavallaro, G. Lazzara, S. Milioto and F. Parisi, *Langmuir*, 2015, **31**, 7472-7478.
60. L. Z. Zhao, C. H. Zhou, J. Wang, D. S. Tong, W. H. Yu and H. Wang, *Soft Matter*, 2015, **11**, 9229-9246.

CHAPTER 7

Conclusion & future direction

Using time dependent dihedral cross correlation functions, we explain how the functionality of a protein is manifested in terms of dynamically correlated distant residue pairs. Thus, this technique might be useful to understand allosteric mechanism in more details and can have impacts on associated applications. The binding of effector molecules at the specific site of an allosteric protein can affect functionality of the distant active site of that particular protein. The biosensors¹ detect the signal resulting from interactions between proteins and stimuli from molecular to macroscopic level. Thus, the allosteric proteins have the potential^{1,2} to not only act as biosensors but also as novel switches in complex circuits, regulators for inducible control of multiple genes along with target for drug design.

Understanding protein-nano particle interactions is essential to stabilize drug delivery system and thus is highly emerging area of research in biomedical applications³. Modified surface of nano particles accordingly, make them biodegradable and biocompatible, so that they can be used for encapsulation, bio conjugation, and therapeutic immobilization techniques. Here, we propose an alternative method on basis of diffusion dynamics of nano sized ligands, to explain how the protein functions are driven by their electrostatic interactions with nano particles. Moreover, due to specific location of the binding pockets, the ligand attached protein may act as directed template for heterogeneous structure at nano meter length scale⁴.

In our study, we use conformational thermodynamics to reveal mechanism behind changes in protein conformations in terms of free energy and entropy. This approach may provide insight to material science application driven by molecular switch using protein conformations as scaffold⁵. Protein conformational switches⁶ signify noticeable structural changes of the proteins upon interacting with external stimuli, like absorbing a photon or binding with a drug. Understanding the switching mechanism can lead to the new development in protein engineering, by which function of a protein can be modulated accordingly using the proper stimuli. Similarly, by using biosensors like biocompatible electronic device or fluorescence sensitive bio-molecules, response of the protein in terms of conformational switches can be detected.

Finally, our study on misfolded proteins based on model with hydrophobic surface along with hydrophilic core can yield idea to gain molecular control over general amphiphilic assembly⁷. This is also useful for designing materials with tuning optical activity in technological applications⁸, bio-diagnostics⁹ and energy conversion¹⁰.

References:

1. A. Villaverde, *Febs. Lett.*, 2003, **554**, 169-172.
2. S. Raman, N. Taylor, N. Genuth, S. Fields and G. M. Church, *Trends Genet.*, 2014, **30**, 521-528.
3. Y. Kim, S. M. Ko and J. M. Nam, *Chem. Asian J.*, 2016, **11**, 1869-1877.
4. S. Bhaskar and S. Lim, *Npg Asia Mater.*, 2017, **9**.
5. M. M. Stratton and S. N. Loh, *Proteins*, 2010, **78**, 3260-3269.
6. J. H. Ha and S. N. Loh, *Chemistry*, 2012, **18**, 7984-7999.

7. A. Sanchez-Iglesias, M. Grzelczak, T. Altantzis, B. Goris, J. Perez-Juste, S. Bals, G. Van Tendeloo, S. H. Donaldson, B. F. Chmelka, J. N. Israelachvili and L. M. Liz-Marzan, *Acs Nano*, 2012, **6**, 11059-11065.
8. M. Gerth and I. K. Voets, *Chem. Commun.*, 2017, **53**, 4414-4428.
9. N. L. Rosi and C. A. Mirkin, *Chem. Rev.*, 2005, **105**, 1547-1562.
10. S. Linic, P. Christopher and D. B. Ingram, *Nat. Mater.*, 2011, **10**, 911-921.

List of Publications

1. **S. Dutta**, M. Ghosh and J. Chakrabarti, Spatio-temporal coordination among functional residues in protein, *Scientific Reports* **7**, 40439, (2017).
2. Lakshmi Maganti, **Sutapa Dutta**, Mahua Ghosh & J. Chakrabarti: Allostery in Orail binding to Calmodulin revealed from conformational thermodynamics, *Journal of Biomolecular Structure and Dynamics*, 1-10, (2018).
3. **S. Dutta**, M. Ghosh and J. Chakrabarti, Anomalous dynamics of metal ions in presence of a bacterial protein, communicated, (2018).
4. **S. Dutta**, and J. Chakrabarti, Model studies on protein aggregation based on hydrophobic and electrostatic interactions, communicated, (2018).
5. **S. Dutta**, M. Ghosh and J. Chakrabarti, Bilayer induced thermodynamic stability of flagellin, communicated, (2018).