

# **Microscopic description of protein-ligand and protein-protein interactions**

*Thesis submitted for the degree of*

**Doctor of Philosophy (Sc.)**

*in*

**Physics (Theoretical)**

*by*

**Samapan Sikdar**

**Department of Physics**

**University of Calcutta**

**2016**

*To My Family*

## *Acknowledgements*

*I would start thanking my supervisor Prof. Jaydeb Chakrabarti for his constant support and invaluable guidance over the years which not only enriched my understanding of the subjects, but also was the key to develop myself as a good scientist and human being. His incredible belief and faith in me helped me to pursue diverse problems simultaneously which was very important for my development. I would also thank Dr. Mahua Ghosh who has been a wonderful co-guide. Her contributions as a mentor have been enormous in the field of biomolecules. I would also express my gratitude towards Dr. Molly De Raychaudhury, our collaborator in the works concerning quantum mechanical calculations. A special acknowledgement to my dear friend and senior Dr. Amit Das for his guidance and fruitful discussions during early days of my PhD carrier. Thanks also to my thesis and PhD committee members, Dr. Sanjoy Bandopadhyay, IIT Kharagpur, and Prof. Subinay Dasgupta, Calcutta University, respectively, for helpful comments and suggestions.*

*I thank all of my teachers from undergraduate and master's level, in particular, Prof. Soumen Kumar Roy (Jadavpur University) for a sound introduction to statistical mechanics and thereafter developing my interest towards simulation techniques. Thanks to Prof. Narayan Banerjee, Prof. Ranjan Bhattacharya from Jadavpur University and Prof. C. Vijayan from IIT Madras for their constant help and encouragement.*

*I thank my fellow group-members: Suman Dutta, Sutapa Dutta, Piya Patra, Dr. Manas Mondal and Dr. Laxmi Maganti. In particular, Paramita Saha and Saronath Haldar require special mention, who have provided me great support through scientific discussions and memorable moments of fun all throughout my PhD tenure. I would also thank my seniors and*

*friends Dr. Manik Banik, Saikat Basu from Jadavpur University, Dr. Chiranjib Nayek, Dr. Arindam Roy, Projesh Basu, Sagnik Chakraborty and Abdul Faruque from IIT Madras.*

*My journey could not have sailed so far without the love, support and guidance of my family. No degree of appreciation would be sufficient for my parents, Swapan Sikdar and Arpita Sikdar and my sister Abhipsa Sikdar for their constant support.*

*Finally, I would thank SNBNCBS for providing me a wonderful research environment and facilities. I would also like to express my heartiest gratitude towards Mr. Sunish Kumar Deb, for coordinating an early submission of my thesis. I also thank University Grants Commission, India for a research fellowship.*

**Samapan Sikdar**

S. N. Bose National Centre for Basic Sciences,  
Block JD, Sector III, Salt Lake,  
Kolkata-700106, West Bengal, India.

# *Table of Contents*

CHAPTER 1 Introduction	8
CHAPTER 2 A microscopic insight from conformational thermodynamics to functional ligand binding in proteins	15
2.1 Introduction	15
2.2 Methods	18
2.3 Results	20
A. $\beta$ 4GalT binding site	23
B. Fatty acid binding residues	24
C. $\beta$ 4GalT docking	25
D. Oleic acid docking	27
2.4 Conclusions	29
Appendix I	32
Appendix II	33
Appendix III	35
CHAPTER 3 Conformational thermodynamics guided structural reconstruction of bio-molecular fragments	37
3.1 Introduction	37
3.2 Methods	40
3.3 Results	42
A. Comparison of apo-TnC models	42
B. Functional implications of $\text{Ca}^{2+}$ ion bound TnC	48
C. Comparison with apo-TnC models from different servers	50
D. Conformational thermodynamic comparison of ModLoop model	50
E. Chemical shifts	51

3.4 Conclusions	54
CHAPTER 4 Conformational thermodynamic changes upon protein oligomerization	55
4.1 Introduction	55
4.2 Methods	56
4.3 Results	57
A. Conformational thermodynamic changes	57
B. Comparison with experiments	58
4.4 Conclusions	59
CHAPTER 5 Static and dynamic response of microscopic conformational variables of ligand binding residues in different proteins	60
5.1 Introduction	60
5.2 Methods	61
5.3 Results	64
A. Mathematical Model	64
B. Distribution of barrier heights from MD simulations	66
5.4 Conclusions	78
Appendix I	79
CHAPTER 6 Quantum chemical studies on stability and function of calcium ion bound calmodulin loops	81
6.1 Introduction	81
6.2 Methods	86
6.3 Results	88
A. Vacuum QC calculations	88
B. Implicit solvent QC calculations with neutral capping	97
C. Implicit solvent QC calculations with Zwitterionic capping	102
6.4 Conclusions	107
Appendix I	108

CHAPTER 7 Ab-initio quantum chemical studies on role of hydration water on protein functionality	111
7.1 Introduction	111
7.2 Methods	114
7.3 Results	115
A. Partial density of states of holo-CaM loops	115
B. Partial density of states of apo-CaM loops	121
C. Hydrogen bonding network	122
D. Comparison with implicit solvent QC calculations	124
7.4 Conclusions	125
Appendix I	126
Appendix II	127
REFERENCES	130
LIST of PUBLICATIONS	139

# ***CHAPTER 1***

## **Introduction**

Bio-molecules are complex in nature due to presence of large number of atoms and a variety of inter-atomic interactions, like hydrophobic, electrostatic, van der Waals and hydrogen bonds. Characterization of bio-molecules and their complexes is a major challenge in biophysical and biochemical studies primarily due to involvement of large number of degrees of freedom.<sup>1</sup> These internal degrees of freedom govern the 3-dimensional arrangement of atoms known as bio-molecular conformation. The bio-molecules being intrinsically flexible, conformational fluctuations play significant role in governing myriad of biological processes<sup>2-9</sup>. Recently, conformational flexibility of bio-molecules has also been used to design hybrid electronic devices such as conformational switches<sup>10</sup>. Moreover, electron transfer properties of certain bio-molecules have been explored to fabricate device applications such as bio-sensors<sup>11, 12</sup> and bio-fuel cells<sup>13</sup> with material science aspect.

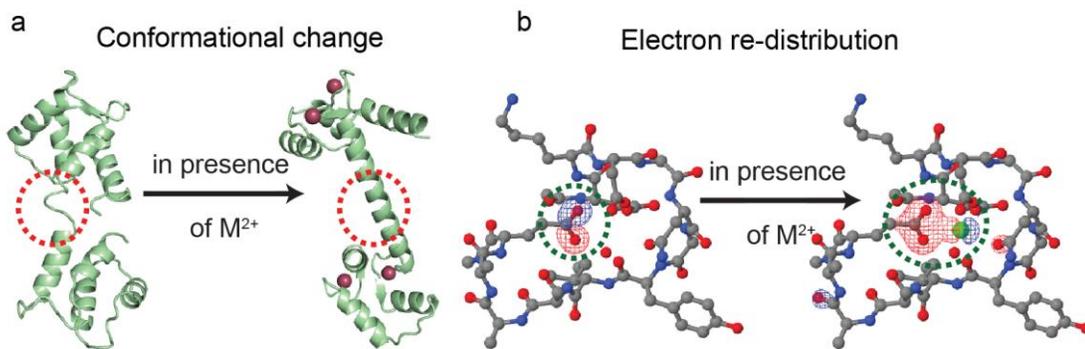
In this thesis, the systems of interest are proteins, its complexes and assemblies. Protein molecules form complexes with various ligands like, metal ions, peptides, proteins and numerous other bio-molecules. Such complexes perform all kinds of functions ranging from enzyme catalysis, photosynthesis to signal transduction to name only a few.<sup>2-4, 14-16</sup> Nearly one third of the known protein structures are capable of binding different metal ions through coordination. Such metal-protein complexes are known as metalloproteins.<sup>17</sup> Under physiological conditions metal ions like Na<sup>+</sup>, K<sup>+</sup>, Mg<sup>2+</sup>, Ca<sup>2+</sup> and Zn<sup>2+</sup> are abundant while, Mn<sup>2+</sup>, Ni<sup>2+</sup>, Cu<sup>+2+</sup>, Fe<sup>2+/3+</sup> and Co<sup>2+/3+</sup> are trace metal ions. All these metal ions can bind to proteins. Certain features of these

metal ions, like small size, positive charge, flexible coordination sphere, specific ligand affinity and varying valence states govern their binding properties to specific proteins.<sup>17</sup> Different ligand atoms like oxygen (*O*), nitrogen (*N*) and sulphur (*S*) from amino acid side chain functional groups and backbone carbonyl *O* generally coordinate with metal ions. Negatively charged amino acids like aspartates and glutamates with  $COO^-$  group and cysteines with  $CH_3S^-$  group are favored ligands due to charge-charge interactions with the cation. The polar ligands within the coordination sphere also show affinity to the metal ions depending on their dipole moment. The vicinal environment of cation gets further complicated due to the presence of both hydrophobic and basic residues. Thus coordination of metal ions to proteins is characterized by a complex interaction between metal ion and diverse amino acids at the metal binding site.

Another class of protein complex is protein assembly, which is a manifestation of protein-protein interactions. Monomeric proteins often form stable aggregates of higher order called oligomeric states, which regulate activity of enzymes, ion-channels and receptors.<sup>18, 19</sup> Aggregation of proteins also occurs due to protein misfolding. This can lead to numerous diseases like Alzheimer's, Parkinson's, Huntington's diseases.<sup>20</sup> Thus microscopic understanding of stability of protein aggregation is important, though not understood well.

The conformational changes in protein complexes can be as small as simple rotation of dihedral angles to large concerted motions like, domain rearrangements (Fig 1.1 a). The conformational changes in a protein have been extensively studied through experiments and computer simulations. Recent works<sup>21-24</sup> show that thermodynamics of changes in protein conformation can be given by histograms of dihedral angles. The peaks of histogram in two conformations are given by Boltzmann factors of effective free energies, while the entropies are calculated using the Gibbs formula. One can estimate the metal ion induced conformational

thermodynamic changes using the histograms of dihedral angles sampled from equilibrium trajectories of all atom molecular dynamics (MD) simulation (see Appendices I and II, chapter 2) of the protein in free and in ligand bound states with explicit water and counter-ions to ensure electro-neutrality. While the static response of dihedral angles is given by change in the histograms, the dynamic response is accessed through dihedral auto-correlation function (DACF)<sup>25-28</sup>. The DACFs indicate how the fluctuations are correlated over time and yields characteristic time-scales of dihedral relaxation. The DACFs,  $C_R^i(\theta, t)$  for dihedral angle fluctuations about its equilibrium value is calculated as the product of the value of  $\delta \cos \theta(\tau)$  at some initial time  $\tau$  and the value after some time interval  $t$ ,  $\delta \cos \theta(t + \tau)$ , averaged over a set of initial conditions.<sup>28</sup> However, the power of dihedral angle based description for microscopic understanding of protein complexes is yet to be explored.



**Fig 1.1. (a) Conformational change of a protein and (b) electronic redistribution upon binding a divalent metal ion,  $M^{2+}$ .**

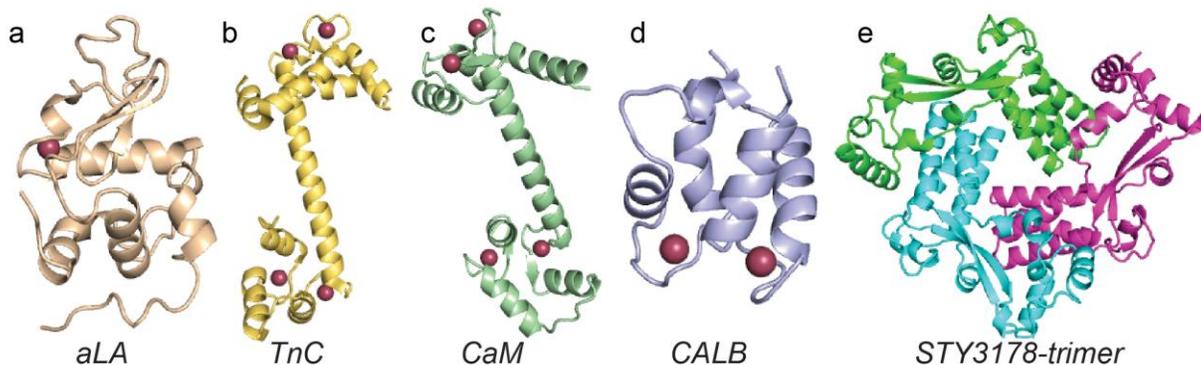
We have applied histogram based method (HBM, Appendix III, chapter 2) to estimate conformational thermodynamic changes to address a number of questions in different protein complexes:

(1) **Prediction of functional residues:**  $\alpha$ -Lactalbumin (aLA)<sup>29</sup>, a divalent metal ion binding protein, comprising of an  $\alpha$ -helical domain and a small  $\beta$ -sheet domain<sup>30</sup> (see Fig 1.2 (a)), which acts as a carrier for fatty acids, like oleic acid, known to possess cytotoxic activity against cancer cell lines<sup>31</sup>. However, the binding sites for fatty acid are largely unknown due to unavailability of crystal structure data.<sup>32, 33</sup> The residues of  $\text{Ca}^{2+}$  bound (holo state) aLA, which undergo a positive change in free energy are destabilized, while residues undergoing positive change in entropy are disordered compared to free (apo state) aLA. We show that such destabilized and disordered basic and hydrophobic residues of  $\text{Ca}^{2+}$ -aLA serve as binding site for oleic acid.<sup>34</sup>

(2) **Prediction of structure from known functions:** Skeletal muscle Troponin C (TnC) is a calcium ( $\text{Ca}^{2+}$ ) binding metallo-protein, exhibiting helix-loop-helix EF-hand motif.<sup>35</sup> The crystal structure<sup>36</sup> of four  $\text{Ca}^{2+}$  bound TnC reveals a dumbbell shaped structure with two globular domains, N-terminal and C-terminal, separated by an extended central helix<sup>37, 38</sup> (see Fig 1.2 (b)). The  $\text{Ca}^{2+}$  bound terminals of TnC bind Troponin I (TnI) forming a complex that regulates muscle contraction and relaxation.<sup>36</sup> In order to understand the molecular basis of function of  $\text{Ca}^{2+}$  ion bound TnC, one needs to have a structure of apo TnC<sup>39</sup>. However, a full molecular structure is not available for apo TnC. We show that conformational thermodynamic changes act as a basis for selecting thermodynamically favorable model of apo-TnC, hence guiding protein structure prediction.<sup>40</sup> Moreover the destabilized and disordered residues of  $\text{Ca}^{2+}$  bound TnC with respect to the apo-TnC model can explain TnI binding.

(3) **Stability of oligomeric protein:** STY3178 from CT18 strain of *Salmonella* Typhi is only biophysically characterized but detailed structural and functional characterizations are absent. STY3178 belongs to yfdX protein family.<sup>41, 42</sup> Recent experiments show that it exists as a

trimer in native state (see Fig 1.2 (e)), which is stable at elevated temperatures and a wide range of pH. We analyze stability of oligomerization of monomeric STY3178 using conformational thermodynamics.<sup>43</sup>



**Fig 1.2. The systems studied in this thesis: (a) aLA, (b) TnC, (c) CaM, (d) CALB and (e) STY3178 trimer.**

We further study the dynamics of dihedral relaxation to understand deviation of rates of bio-molecular processes from Stoke's behavior.<sup>44, 45</sup> Previous simulation studies show that such deviations are accompanied by isomerization of dihedral angles upon ligand binding.<sup>46, 47</sup> We propose a mathematical model (Appendix I, chapter 5) to understand the role of energy barrier in governing dihedral angle transitions between isomeric states.<sup>48</sup> The dihedral angle relaxation is observed to be slow whenever the isomeric states are separated by low energy barrier ( $f_B$ ) with a characteristic time-scale,  $\tau \propto f_B^{-1}$ . The underlying physical picture is such that whenever two isomeric states are separated by a low energy barrier, frequent barrier crossing events occur as a result of which dihedral fluctuations exhibit slow decay profile. Our analysis also shows that participation of several dihedral angles with a distribution of barrier heights account for observed deviation from Stoke's law. We perform MD simulations of different calcium ( $\text{Ca}^{2+}$ ) binding proteins (see Fig 1.2 (a-d)) to extract the barrier height distribution of dihedral angle isomerization states. We find that the distribution shows exponential decay in energy. We also

compute dihedral auto-correlation function from MD simulated trajectories, which yields the  $\tau \propto f_B^{-1}$  dependence consistent with our modeling results.

Metal ion coordination leads to electronic redistribution and hence induces polarization of protein atoms in the binding region (Fig 1.1 b). In general conformational changes and electronic redistribution processes are interdependent. Till date there is no reasonable computational method that incorporates simultaneously both aspects for large systems like proteins. This leads one to consider the two processes separately. The electron redistribution process upon ligand binding is studied through quantum chemical (QC) calculations on a given conformation of protein fragment truncated to include only the region of interest. These QC calculations are primarily based on density functional theory (DFT, Appendix I, chapter 6)<sup>49</sup>. In DFT calculations, the ground state energy of a many-electron system is determined by variation of electron QC calculations have been used to investigate stability and reactivity of such protein fragments.<sup>17, 50-55, 17, 56-60</sup> However, interpretation of such QC results with protein fragments has not been very well documented in literature.

The ligand induced electron redistribution is studied for Ca<sup>2+</sup> bound calmodulin (CaM) through QC calculations. CaM is an EF-hand metallo-protein with four Ca<sup>2+</sup> ion binding loops and participates in signal transduction in eukaryotic cells<sup>9</sup>. The crystal structure<sup>61</sup> of four Ca<sup>2+</sup> bound CaM (see Fig 1.2 (d)) has similar dumbbell shaped structure as in TnC. Several experiments indicate that both ligand and non-ligand residues of CaM stabilize coordination geometry.<sup>62, 63</sup> Moreover, tyrosine residues in loop 3 and loop 4 of CaM are involved in phosphorylation.<sup>16</sup> Earlier QC calculations<sup>17, 52, 59, 64-66</sup> consider only ligand residues due to large size of CaM and do not provide any insight on the role of non-ligands. We perform QC calculations on isolated Ca<sup>2+</sup> bound CaM loops with various capping. Our QC calculations based

on DFT show that coordination levels are independent of capping details and bring out the role of both ligand and non-ligand residues in providing stability to coordination geometry.<sup>67</sup> We also show using neutral capped  $\text{Ca}^{2+}$  bound loops in implicit solvent DFT calculation that electron density at the highest occupied molecular orbital (HOMO) is localized on tyrosine of loop 3, indicating a nucleophilic site, a pre-requisite for phosphorylation.<sup>68, 69</sup> We repeat the QC calculations with charged terminals representing Zwitterionic state. The behavior of charged terminals is similar to the effect of impurities on ground state energy spectrum and may have significance in device applications. We find that the terminal charges do not interfere with the nucleophilic character of Y99. We also perform ab-initio QC calculation with explicit water molecules in vicinity of protein surface to account for nucleophilic character of tyrosines from both loop 3 and loop 4.<sup>70</sup>

The thesis is organized as follows: We discuss oleic acid binding to  $\text{Ca}^{2+}$ -aLA complex in chapter 2.<sup>34</sup> Chapter 3 describes modeling of apo-TnC missing fragment and selecting the thermodynamically favorable conformation.<sup>40</sup> Chapter 4 shows the conformational thermodynamic changes upon protein oligomerization.<sup>43</sup> The role of internal friction in governing dihedral transitions upon  $\text{Ca}^{2+}$  binding is discussed in chapter 5.<sup>48</sup> Chapter 6 describes DFT based QC calculations on  $\text{Ca}^{2+}$  bound loops of CaM in vacuum<sup>67</sup>, implicit solvent<sup>68, 69</sup> and those with explicit hydration layer are given in chapter 7.<sup>70</sup>

## ***CHAPTER 2***

# A microscopic insight from conformational thermodynamics to functional ligand binding in proteins

## **2.1 Introduction**

The function of bio-macromolecules, like proteins, is quite often specific to their conformational states. Understanding protein functions in a given conformational state from microscopic considerations is a major challenge in molecular biology.<sup>1</sup> Such microscopic insight would be a significant advancement not only in understanding variety of bio-molecular phenomena, but also in designing bio-molecules with targeted functionalities.

Ligand binding to a protein changes the protein conformation and hence, modulates its function<sup>6</sup>. Recent studies<sup>21-24</sup> show that thermodynamic contribution due to conformational changes play major role in stabilizing protein-ligand complexes. Dihedral angles have been widely used as microscopic conformational variables to describe protein conformations in theoretical studies<sup>21-24</sup>. The major advantage is that these dihedral angles show only short ranged correlations which can be ignored to estimate conformational entropy. A simple and computationally-efficient histogram based method (HBM) has been proposed to calculate both conformational free energy and entropy changes from the distribution of dihedral angles<sup>21-24</sup>, enabling one to look microscopically into the conformational thermodynamics in different regions of a protein. Here we illustrate through microscopic conformational thermodynamics, the connection between conformational states and functionalities of metalloproteins, where metal ions induce conformational changes in the protein. We consider in particular,  $\alpha$ -Lactalbumin

(aLA), a divalent metal (particularly calcium,  $\text{Ca}^{2+}$ ) ion binding metalloprotein, expressed in mammary glands during lactation<sup>71</sup>. It participates in a variety of cellular functions, including lactose synthesis<sup>72</sup> and fatty acid binding<sup>73</sup>.

aLA comprises of an  $\alpha$ -helical domain and a small  $\beta$ -sheet domain separated by a cleft<sup>30</sup>. The metal ion free native-apo structure of aLA is stable under neutral pH and physiological salt concentration. Calorimetric measurements<sup>74, 75</sup> suggest that apo-aLA can bind to different divalent metal ions including physiologically abundant  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  at its metal binding loop connecting the two domains, resulting in metal ion bound holo structure. Fluorescence<sup>76, 77</sup> and circular dichroism<sup>71</sup> studies indicate that these metal ions provide stability to native-apo aLA against heat or other denaturing agents<sup>74, 78, 79</sup> along with change in conformational state, characterized by narrowing of interfacial cleft<sup>29</sup>. The  $\text{Ca}^{2+}$ -aLA complex is functional in forming regulatory component of lactose synthase enzyme, modulating the activity of catalytic component,  $\beta$ -1,4-galactosyltransferase ( $\beta$ 4GalT), to promote lactose synthesis.<sup>72</sup> The crystal structure of aLA complexed with  $\beta$ 4GalT has been reported.<sup>72, 80</sup> Mutational studies<sup>81</sup> indicate that aLA binding to  $\beta$ 4GalT is mediated through residues from A4 and C-terminal tail  $3_{10}$  helix of aLA.

Another interesting property of aLA is its ability to form molten globule (MG) conformation<sup>30</sup> in acidic pH upon removal of  $\text{Ca}^{2+}$ . The MG state, having native-apo like secondary but fluctuating tertiary structure, is capable of binding to various fatty acids, like oleic acid, stearic acid etc<sup>32, 73</sup>. Recently, MG-aLA-oleic acid complex, popularly known as XAMLET<sup>82</sup> (X stands for human, bovine, goat or other species, Alpha-lactalbumin Made LETHal to Tumor cells), is found to exhibit cytotoxic activity against cancer cell lines<sup>31</sup>. It has been identified that oleic acid is the cytotoxic factor of XAMLET, whereas MG-aLA acts as its

carrier<sup>83, 84</sup>. The binding sites for fatty acid are largely unknown due to unavailability of crystal structure of MG-aLA-oleic acid complex.<sup>32, 33</sup> It is still inconclusive whether divalent metal ion bound native aLA can bind oleic acid similar to MG state. There are studies which indicate that Ca<sup>2+</sup>-aLA can bind to oleic acid<sup>85</sup>, albeit having weaker binding than that in MG state. On other hand, there are reports which rule out the existence of fatty acid binding sites in divalent metal bound aLA<sup>86</sup>. Further, it has also been shown that Ca<sup>2+</sup> binding have no effect on the cytotoxic activity<sup>84, 87, 88</sup> of XAMLET. Since at physiological condition, pH is neutral and there is abundance of divalent metal ions like Ca<sup>2+</sup> and Mg<sup>2+</sup>, it is of paramount interest to understand the role of these ions in fatty acid binding to aLA, particularly in context of cytotoxic activities.

In this chapter<sup>34</sup> we focus on possible functional behaviours of divalent metal ion bound native aLA, highlighting the conformational thermodynamic changes computed from all atom molecular dynamics (see Appendices I and II). The conformational states of aLA are characterized by radius of gyration showing overall shrinkage of the protein upon metal ion binding. We compute free energy changes ( $\Delta G_i^{conf}$ ) and entropy changes ( $T\Delta S_i^{conf}$ ) of (*i-th*) residue in holo state with respect to apo state from histograms of dihedral angles (see Appendix III). We relate the thermodynamics to protein functionality: Residues having instability in conformational free energy ( $\Delta G_i^{conf} > 0$ ) in holo form compared to native-apo state, are likely to participate in further binding event to attain thermodynamic stability by decreasing its free energy. The loss in free energy is further aided by disordering ( $T\Delta S_i^{conf} > 0$ ). Based on these criteria, we compare the functionalities between different metal ion bound conformations of aLA.

## 2.2 Methods

### Preparation of system

The initial configurations are chosen from following protein data bank (PDB) entries: 1F6R (native-apo bovine aLA) and 1F6S (Ca<sup>2+</sup>-bovine aLA). We considered chain A of both 1F6R and 1F6S including bound crystal waters for our calculation. Hydrogen is added to both these crystal structures. The apo and holo proteins are then solvated in rectangular parallelepiped water box of dimensions 78x52x59 Å<sup>3</sup>. Sodium (Na<sup>+</sup>) and chloride (Cl<sup>-</sup>) ions are added to achieve a physiological salt concentration of 0.13 M as well as electroneutrality. Mg<sup>2+</sup>-aLA is prepared by substituting Ca<sup>2+</sup> from the cation binding site due to unavailability of crystal or NMR structures. We perform energy minimization of 1000 steps and then run 100 nanoseconds (ns) long simulations for each of these systems. The details of simulation are given in Appendices I and II.

### Radius of gyration calculation

Radius of gyration ( $R_g$ ) of a molecule is a measure of its dimension. The  $R_g$  is calculated as average distance of C-alpha atoms from their centre of mass ( $\vec{R}_{CM}$ ),  $\vec{R}_{CM} = \sum_i m_i \vec{r}_i / \sum_i m_i$ , where  $m_i$  and  $\vec{r}_i$  are mass and position vectors of i<sup>th</sup> C-alpha atom. The squared  $R_g$  is then calculated as:

$$R_g^2 = \sum_i m_i \left( \vec{r}_i - \vec{R}_{CM} \right)^2 / \sum_i m_i, \quad (2.1)$$

$R_g$  is computed for every structure in the ensemble and a histogram is generated for Ca<sup>2+</sup>-, Mg<sup>2+</sup>- and native-apo aLA.

## Equilibrium correlations and Thermodynamics

The correlation between any two dihedral angles  $\zeta_a$  of  $a$ th residue and  $\zeta'_b$  of  $b$ th residue is defined as<sup>89</sup>

$$C_{\zeta\zeta'}(s) = \langle \cos \zeta_a \cdot \cos \zeta'_b \rangle - \langle \cos \zeta_a \rangle \langle \cos \zeta'_b \rangle, \quad (2.2)$$

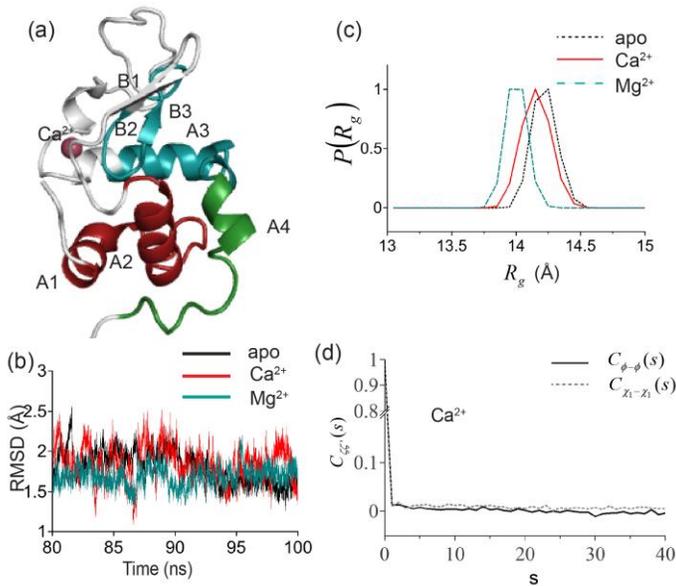
where  $s = |a - b|$  is separation between residue locations. The angular brackets denote ensemble average. If such equilibrium dihedral correlations are insignificant then the dihedrals are independent. The details of thermodynamic calculation are given in Appendix III.

## Docking studies

The average structures of  $\text{Ca}^{2+}$ -aLA and  $\text{Mg}^{2+}$ -aLA are calculated from respective equilibrated trajectories. The structure of  $\beta 4\text{GalT}$  is segregated from aLA- $\beta 4\text{GalT}$  complex (PDB ID: 1NKH). HADDOCK<sup>90, 91</sup> is used for docking  $\text{Ca}^{2+}/\text{Mg}^{2+}$ -aLA complexes with  $\beta 4\text{GalT}$  molecule. The docking protocol<sup>90</sup> consists of three stages, namely, a rigid body energy minimization, torsion angle refinement and finally explicit solvent refinement. We bias the docking procedure by considering destabilized C-terminal residues of  $\text{Ca}^{2+}$ -aLA and F280, Y286, Q288, Y289, F360 and I363 of  $\beta 4\text{GalT}$ <sup>72, 80</sup> (known from crystal structure) as active residues for binding. The resulting docked structures are sorted by minimum energy criteria and root mean square deviation (RMSD) clustering. Interface of the docked complex is analyzed based on  $\text{C}\alpha$  distance (5Å) between residues from binding partners. Oleic acid molecule is segregated from PDB ID: 1LFO. The  $\text{Ca}^{2+}/\text{Mg}^{2+}$ -aLA complexes are used as receptor for oleic acid molecule in docking process performed with PatchDock algorithm<sup>92, 93</sup> based on shape complementarity principle. We bias the docking by considering destabilized basic and hydrophobic residues. The interfaces of docked complexes are analyzed based on distance (5Å) between carbon atoms of oleic acid and  $\text{C}\alpha$  atoms of aLA residues.

## 2.3 Results

The X-ray crystal structure<sup>29</sup> of  $\text{Ca}^{2+}$  ion bound aLA is shown in Fig 2.1 (a). The equilibration of our simulation is ensured through root mean square deviation (RMSD) plots of native-apo,  $\text{Ca}^{2+}$ - and  $\text{Mg}^{2+}$ -aLA as illustrated in Fig. 2.1(b). We characterize the conformational states of aLA through its radius of gyration,  $R_g$  (see Eq. (2.1), Methods). The probability distributions of  $R_g$ ,  $P(R_g)$  are shown in Fig 2.1(c).  $P(R_g)$  of native-apo aLA shows a peak at



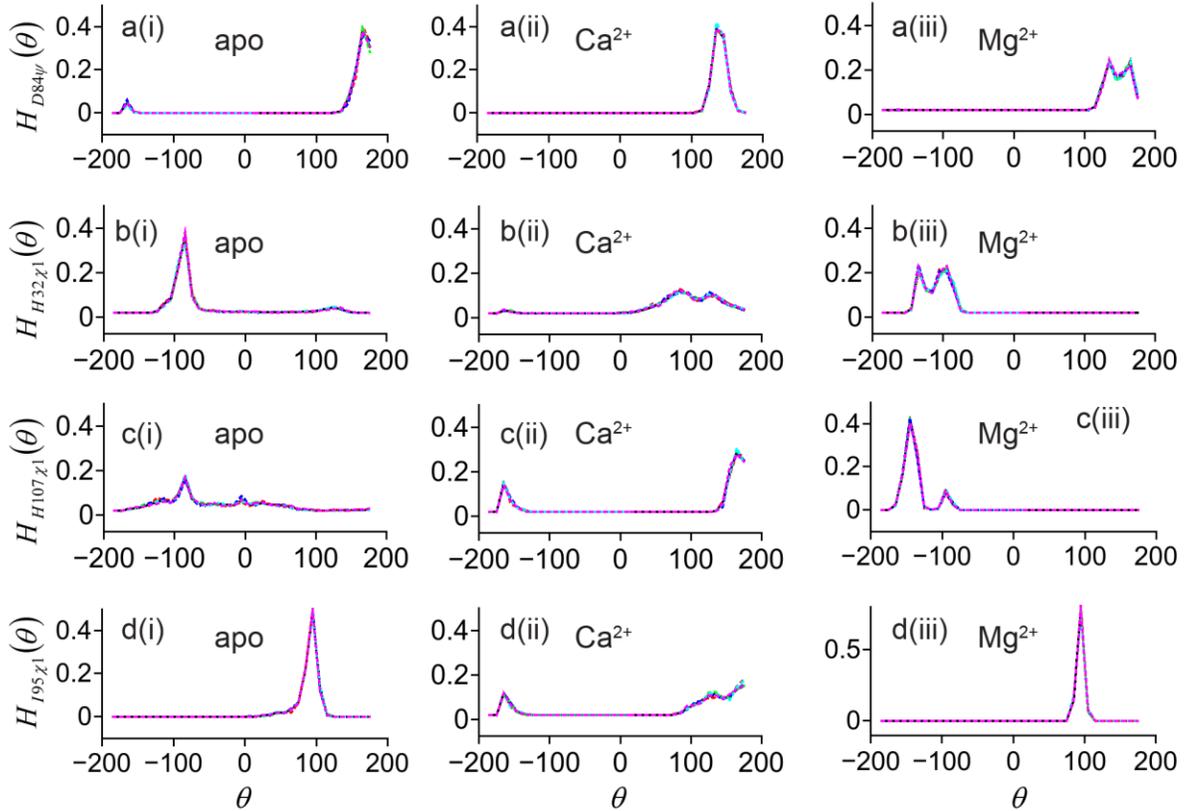
**Fig. 2.1(a)** The crystal structure of  $\alpha$ -Lactalbumin (aLA), showing the  $\alpha$ -helical (A1-A4) and the  $\beta$ -sheet (B1-B3) domains separated by interfacial cleft and connected by the  $\text{Ca}^{2+}$  (red) binding loop. The  $\beta$ 4GalT binding site is shown in green. The putative oleic acid binding sites are formed by the interfacial residues (blue) and the residues from the A1 and A2 helices (red). (b) The backbone rmsd of native-apo,  $\text{Ca}^{2+}$ - and  $\text{Mg}^{2+}$ -aLA. (c) The distribution of radius of gyration ( $P(R_g)$ ) of native-apo,  $\text{Ca}^{2+}$ - and  $\text{Mg}^{2+}$ -aLA. (d) The equilibrium dihedral correlations,  $C_{\phi-\phi}(s)$  and  $C_{\chi_1-\chi_1}(s)$  of  $\text{Ca}^{2+}$ -aLA.

14.3 Å. Metal ion binding leads to an overall shrinking of the metallo-protein complex. The peak of  $P(R_g)$  for  $\text{Ca}^{2+}$ -aLA is 14.1 Å, while that of  $\text{Mg}^{2+}$ -aLA is 14.0 Å. The larger shrinkage of  $\text{Mg}^{2+}$ -aLA complex compared to  $\text{Ca}^{2+}$ -aLA, is due to enhanced electrostatic attraction by higher charge density on smaller  $\text{Mg}^{2+}$  ion.

The dihedral angles,  $\phi, \psi, \chi_1, \chi_2, \chi_3, \chi_4$  and  $\chi_5$ , for all three cases are computed from these equilibrated trajectories. The equilibrium dihedral correlations (Eq. (2.2), Methods) corresponding to these simulated structures of aLA are found to be short ranged as reported in

previous studies<sup>94-96</sup>. For example, backbone dihedrals  $\phi$  of two residues separated by a distance  $s$ ,  $C_{\phi-\phi}(s)$  and that of side chain dihedral  $\chi_1$ ,  $C_{\chi_1-\chi_1}(s)$  are short ranged, as shown for  $\text{Ca}^{2+}$ -aLA (Fig. 2.1(d)). Similar short ranged correlation is also observed in case of  $\text{Mg}^{2+}$ -aLA. We ignore such short range correlations as in the earlier studies<sup>22, 24, 94</sup> and treat the dihedrals as independent conformational variables.

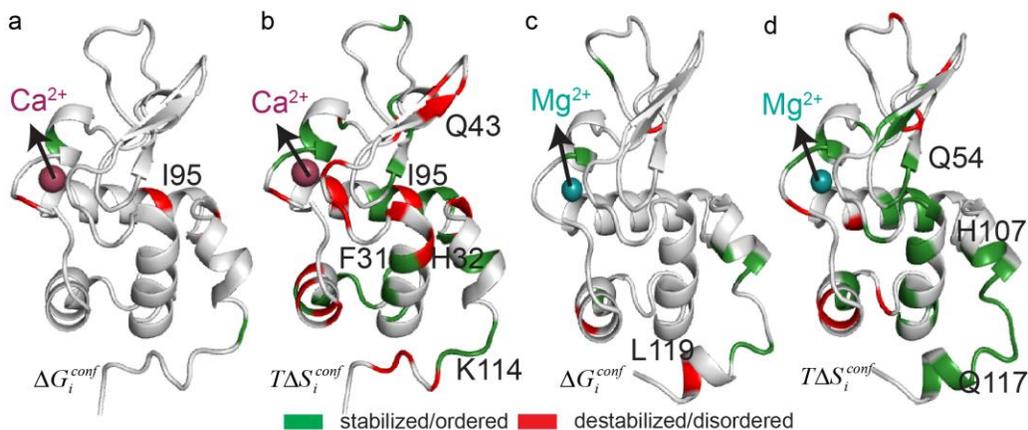
We compute histograms, representing equilibrium distributions for dihedral angles of each residues. The convergence of equilibrium dihedral angle distributions is a pre-requisite to thermodynamic calculations<sup>97, 98</sup> and illustrated in Fig 2.2. We randomly generate five samples



**Fig 2.2. The convergence of representative equilibrium dihedral distributions, (a) D84- $\psi$ , (b) H32- $\chi_1$ , (c) H107- $\chi_1$  and (d) I95- $\chi_1$  corresponding to five samples generated randomly from the 80-100 ns trajectory, for apo- (i),  $\text{Ca}^{2+}$ - (ii) and  $\text{Mg}^{2+}$ -aLA (iii).**

of 1000 conformations each from the equilibrated 80-100 ns trajectory for native-apo and holo complexes. We show a few sample dihedral distributions for D84- $\psi$  (Fig 2.2 a(i)-(iii)), H32- $\chi_1$ (Fig 2.2 b(i)-(iii)), H107- $\chi_1$ (Fig 2.2 c(i)-(iii)) and I95- $\chi_1$ (Fig 2.2 d(i)-(iii)) corresponding to apo-,  $\text{Ca}^{2+}$ -,  $\text{Mg}^{2+}$ -aLA complexes, respectively. The equilibrium dihedral distributions corresponding to these samples show significant overlap with total ensemble. The histograms show only small sample to sample fluctuations ( $\leq 10\%$ ), indicating convergence of dihedral angle distributions.

Conformational free energy change for residue  $i$  ( $\Delta G_i^{conf}$ ) is calculated from the histograms and corresponding entropy change ( $T\Delta S_i^{conf}$ ) (see Appendix III), by adding individual dihedral contributions of a given residue. A negative change in conformational free energy,  $\Delta G_i^{conf}$  with respect to native-apo state indicates conformational stability, whereas that in  $T\Delta S_i^{conf}$  represents conformational ordering. The thermodynamic changes at each protein residue ( $i$ ) for different metal bound forms are illustrated in Fig. 2.3.



**Fig. 2.3.** The conformational thermodynamic changes are illustrated on the structure of aLA. The plots represent (a)  $\Delta G_i^{conf}$ , (b)  $T\Delta S_i^{conf}$  on  $\text{Ca}^{2+}$ -aLA; (c)  $\Delta G_i^{conf}$  and (d)  $T\Delta S_i^{conf}$  on  $\text{Mg}^{2+}$ -aLA complex. Stabilized/ordered residues are illustrated in green and destabilized/disordered residues in red, while residues in white indicate marginal change in conformational thermodynamics.

**Table 2.1. Total conformational thermodynamic changes (kJ/mol) in Ca<sup>2+</sup>- and Mg<sup>2+</sup>-aLA.**

	$\Delta G_{total}^{conf}$	$T\Delta S_{total}^{conf}$	$\Delta H_{total}^{conf}$
<b>Ca<sup>2+</sup>-aLA</b>	-20	-92	-112
<b>Mg<sup>2+</sup>-aLA</b>	-36	-156	-192

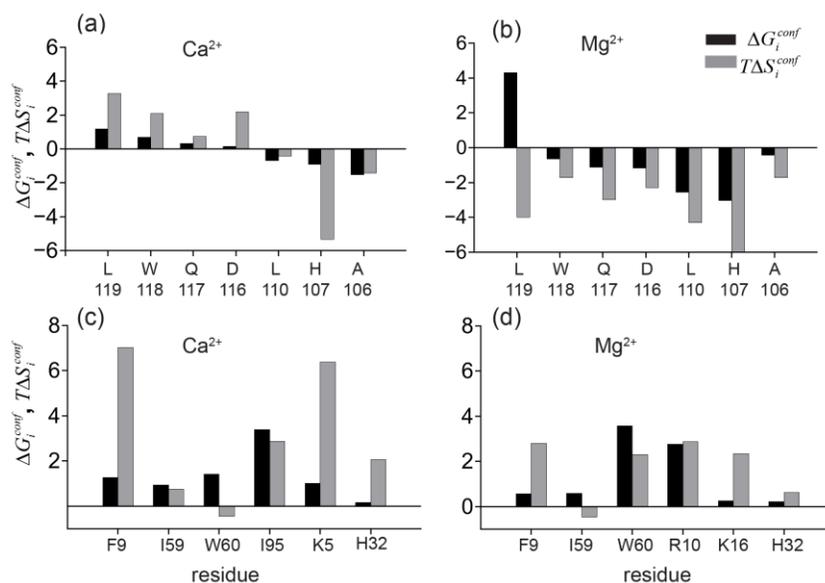
We show free energetically stabilized residues in green and destabilized ones in red. Similarly ordered residues are shown in green and disordered residues in red. Rest of the residues in white indicate marginal change ( $\sim \pm 2$  kJ/mol) in thermodynamic parameters.  $\Delta G_i^{conf}$  and  $T\Delta S_i^{conf}$  plots for Ca<sup>2+</sup>-aLA are shown in Figs. 2.3(a) and 2.3(b) respectively. Figs. 2.3(c) and 2.3(d) represents similar quantities corresponding to Mg<sup>2+</sup>-aLA complex. The overall change in conformational thermodynamics upon metal ion binding with respect to native-apo aLA, obtained by adding all dihedral contributions, is shown in Table. 2.1. Mg<sup>2+</sup> binding to apo-aLA induces larger conformational stability and ordering compared to Ca<sup>2+</sup> binding.

We now examine the residues with destabilization in conformational free energy in the holo states compared to native-apo state. Such residues are likely to participate in further binding events in order to get conformational stability by reducing conformational free energy. Further, if these residues are also disordered i.e. flexible in holo state compared to apo state, being in thermodynamically favourable condition, the stability in free energy can be achieved with low enthalpy cost. We show the data for conformational thermodynamic changes of such residues in Figs. 2.4(a)-(d). We shall first consider  $\beta$ 4GalT binding residues of aLA and then the basic and hydrophobic residues which may pose as putative binding residues for oleic acid.

### **A. $\beta$ 4GalT binding site**

Mutational studies indicate involvement of C-terminal hydrophobic A106, L110 and W118, polar Q117 and basic H107 in the interaction between aLA and  $\beta$ 4GalT<sup>81</sup>. Mutagenesis of these residues leads to complete loss in lactose synthase activity, while partial loss (3 to 30

times) in activity has been reported due to mutations of acidic D116 and hydrophobic L119<sup>81</sup>. The conformational thermodynamic changes of  $\beta$ 4GalT binding residues upon  $\text{Ca}^{2+}$  binding are shown in Fig. 2.4(a) including D116 and L119. L119 shows largest destabilization and disordering followed by W118, Q117 and D116. However, the binding residues from A4 helix, L110, H107 and A106 show increasing stability. Fig. 2.4(b) illustrates the effect of  $\text{Mg}^{2+}$  binding on  $\beta$ 4GalT binding residues. L119 is destabilized but ordered in  $\text{Mg}^{2+}$ -aLA unlike that in  $\text{Ca}^{2+}$ -aLA. W118, Q117 and D116, destabilized and disordered in  $\text{Ca}^{2+}$ -aLA, undergo stabilization and ordering upon  $\text{Mg}^{2+}$  binding. H107 gains more stability and ordering, compared to that induced by  $\text{Ca}^{2+}$  binding. A106, although less stabilized in  $\text{Mg}^{2+}$ -aLA, shows similar ordering for both metal ions.



**Fig. 2.4.** The conformational thermodynamic changes at the  $\beta$ 4GalT binding residues are illustrated for (a)  $\text{Ca}^{2+}$ -aLA and (b)  $\text{Mg}^{2+}$ -aLA. Conformational thermodynamic changes of basic and hydrophobic residues which can aid in oleic acid binding to (c)  $\text{Ca}^{2+}$ -aLA and (d)  $\text{Mg}^{2+}$ -aLA.

## B. Fatty acid binding residues

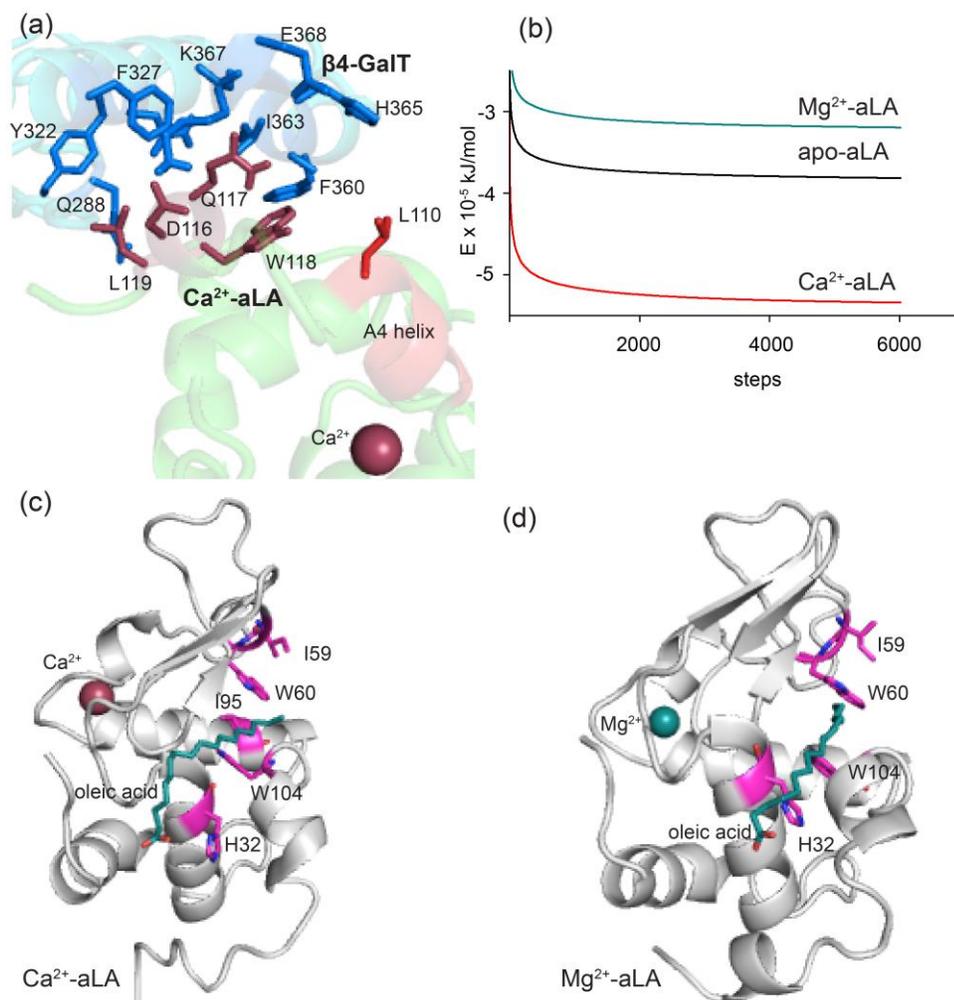
Fatty acids being amphipathic in nature with negatively charged carboxylate ( $\text{COO}^-$ ) head-groups and long hydrophobic tail, complexation of fatty acids with aLA is driven by both

electrostatic and hydrophobic interactions<sup>33, 83</sup>. Experimental studies<sup>33</sup> suggest that putative binding sites for oleic acid lie between A1 and A2 helices and interfacial cleft (see Fig. 2.1(a)). The destabilized residues from Ca<sup>2+</sup>-aLA, which can pose as a binding profile to fatty acids, are illustrated in Fig. 2.4(c). These residues include K5, H32, F9, I59, W60 and I95. Both H32 and K5 are destabilized and disordered. F9 undergoes large disordering in Ca<sup>2+</sup>-aLA. The largest destabilization is observed for I95 along with enhanced flexibility. I59 is also destabilized and disordered, while W60 is destabilized but ordered. Fig. 2.4(d) illustrates the residues of Mg<sup>2+</sup>-aLA which undergo destabilization compared to native-apo aLA. The conformational thermodynamic changes indicate gain in flexibility of H32 albeit less upon Mg<sup>2+</sup> binding compared to Ca<sup>2+</sup>-aLA. Unlike Ca<sup>2+</sup>-aLA, basic R10 and K16 get destabilized and disordered. The hydrophobic F9 undergoes destabilization and disordering, however, to a lesser extent compared to that in Ca<sup>2+</sup>-aLA. I59 is destabilized with marginal ordering compared to Ca<sup>2+</sup>-aLA. W60 undergoes maximum destabilization and enhanced flexibility.

### C. $\beta$ 4GalT docking

A biased docking analysis has been carried out with Ca<sup>2+</sup>-aLA and  $\beta$ 4GalT considering D116, Q117, W118 and L119 as active residues of Ca<sup>2+</sup>-aLA. The interface of docked Ca<sup>2+</sup>-aLA- $\beta$ 4GalT complex is given in Fig. 2.5 (a). The side chains of acidic D116 and that of basic K367 of  $\beta$ 4GalT are within a distance of 4.8Å. The polar Q117 is found to lie in vicinity (~5Å) of polar N323 and T366 and hydrophobic F327, I363, A364 and basic H365 of  $\beta$ 4GalT. Other interfacial residues from Ca<sup>2+</sup>-aLA, namely, W118 and L119 show hydrophobic interactions with F360, I363, A364 and Q288, Y322 respectively. Among the A4 helix residues stabilized upon Ca<sup>2+</sup> binding, L110 is within 6Å distance of H365 from  $\beta$ 4GalT, with its side chain pointing towards the interface, as observed in crystal structure. The side chain of L110 acts like a hinge, which

probably signifies that the distal A4 helix residues get kinetically trapped in binding conformation with  $\beta 4\text{GalT}$  residues. The differences between interfaces of docked and crystal structure are primarily due to kinetically controlled residues, which are not captured entirely in docking analysis. Nonetheless, overall similarities in the two interfaces suggest that destabilized and disordered residues provide thermodynamic stability to  $\text{Ca}^{2+}$ -aLA- $\beta 4\text{GalT}$  complex. Since  $\text{Mg}^{2+}$  binding to native-apo aLA leads to conformational stability and ordering to all residues of C-terminal (see Fig. 2.4(b)), the binding of  $\beta 4\text{GalT}$  is not thermodynamically stabilized.



**Fig. 2.5 (a).** Zoomed view of the interface of the energy minimized docked  $\text{Ca}^{2+}$ -aLA- $\beta 4\text{GalT}$  complex. **(b)** The energy minimization plots corresponding to oleic acid complexes with native-apo aLA,  $\text{Ca}^{2+}$ -aLA and  $\text{Mg}^{2+}$ -aLA. The energy minimized docked **(c)**  $\text{Ca}^{2+}$ -aLA-oleic acid complex and **(d)**  $\text{Mg}^{2+}$ -aLA-oleic acid complex.

#### D. Oleic acid docking

We perform docking analysis of  $\text{Ca}^{2+}$ -aLA with oleic acid. Since electrostatic interaction is energetically predominant, we bias the docking process with destabilized basic residues of  $\text{Ca}^{2+}$ -aLA, namely, K5 and H32. The resulting docked complex is energy minimized in presence of explicit solvent (data shown in Fig. 2.5 (b)), and the optimized structure is shown in Fig. 2.5 (c). We find that negatively charged head group of oleic acid is stabilized through electrostatic interactions with side chain imidazole ring of basic H32 (separation  $\sim 5.0$  Å). The long hydrophobic tail of oleic acid extends towards the interfacial cleft. This long tail is further stabilized through hydrophobic interactions with W104, I95, I59 and W60. Here I95, I59 and W60 are destabilized hydrophobic residues of  $\text{Ca}^{2+}$ -aLA complex (see Fig. 2.4(c)). The aromatic side chain of W104 lies at a distance of 5.4 Å from oleic acid tail. Since W104 is a stabilized residue, it is likely to have kinetic role in binding, similar to that of L110 in case of  $\beta$ 4GalT binding. Although the docking is biased through destabilized basic residues only, docking results are consistent with destabilized hydrophobic residues as well.

Similarly for  $\text{Mg}^{2+}$ -aLA, oleic acid docking is biased through destabilized R10, K16 and H32 (see Fig. 2.4(d)). The energy minimized docked  $\text{Mg}^{2+}$ -aLA-oleic acid complex is illustrated in Fig. 2.5(d). Again the negatively charged head group of oleic acid is stabilized by side chain imidazole ring of basic H32. The hydrophobic tail encounters side chains of W104, I59 and W60. However, number of destabilized hydrophobic residues being less in  $\text{Mg}^{2+}$ -aLA when compared to  $\text{Ca}^{2+}$ -aLA, the stability of  $\text{Mg}^{2+}$ -aLA-oleic acid complex is less, as evident from the energy values (see Fig. 2.5 (b)).

We repeat the docking analysis with all destabilized basic and hydrophobic residues of  $\text{Ca}^{2+}$ -aLA. The docked structures are same as that found by biasing through the basic residues

only. This shows that electrostatic interaction plays a dominant role in oleic acid binding to  $\text{Ca}^{2+}$ -aLA. Our results are in agreement to recent studies on electrostatic interactions playing critical role in functional ligand binding and folding of aLA<sup>99-101</sup>. In spite of biasing the docking with more than one basic residue, we find that the head group of oleic acid stabilizes near H32 and no other destabilized basic residues, like K5. This is probably so, for the hydrophobic residues near K5 are buried and hence cannot stabilize oleic acid tail. The presence of destabilized hydrophobic residues near H32 makes it favourable for both electrostatic and hydrophobic interactions between  $\text{Ca}^{2+}$ -aLA and oleic acid, rendering oleic acid binding specific to H32. Our results are in agreement to reported putative oleic acid binding site in the interfacial cleft<sup>33</sup>.

We further dock the average simulated structure of native-apo aLA to oleic acid without any bias. Top ten docked complexes show that head group of oleic acid stabilizes near different basic residues of aLA. The oleic acid tail gets oriented towards hydrophobic residues in vicinity of the docked basic residues. All the docked complexes show similar energy after minimization irrespective of oleic acid binding region or orientation. We show energy values upon minimization (Fig. 2.5(b)) for one of the docked complexes, in which the head group of oleic acid points towards the side chain of H32. The tail is stabilized through hydrophobic interactions with A109 and L110 of native-apo aLA. The minimum energy for native-apo aLA-oleic acid complex is much higher than  $\text{Ca}^{2+}$ -aLA-oleic acid complex but lower than  $\text{Mg}^{2+}$ -aLA-oleic acid complex. Thus we find that oleic acid binding to  $\text{Ca}^{2+}$ -aLA is more favourable compared to either native-apo or  $\text{Mg}^{2+}$  bound aLA, binding being weakest in  $\text{Mg}^{2+}$ -aLA complex.

Functional sites are often conserved through evolution in a protein family. We use the primary sequence of bovine aLA as query sequence for protein Basic Local Alignment Search Tool (BLAST)<sup>102</sup>. We manually select 36 sequences of aLA from BLAST search corresponding

to different mammalian species and perform a multiple sequence alignment using Clustal Omega<sup>103</sup>. We observe that basic H32 along with hydrophobic residues like W60, I95 and W104 are highly conserved across the aLA sequences. However, the other hydrophobic residue, I59 is weakly conserved (see Fig 2.6).

Existing computational methodologies<sup>104</sup> for identification of functional residues rely on either structure or sequence based data, which do not have strong thermodynamic basis. For instance, a functional site is characterized by a set of distance constraints or existence of pockets and clefts on protein surfaces. A different genre of methods includes graph theoretic approach, that transforms a protein into a graphical network, where amino acids are represented as nodes and interactions as edges<sup>105</sup>. Such networks are then compared to various motifs and graph similarity measures to identify functional sites. It has been observed that high network centrality corresponds to functional sites in proteins.<sup>105</sup> The dynamics of protein governs its functional role and hence functional residues are often identified through larger root-mean-square fluctuations compared to other parts of the protein<sup>106</sup>. Unlike the existing methods, our method for understanding the functionality is based on the sound thermodynamic criteria.

## 2.4 Conclusion

In conclusion, we have related conformational thermodynamic changes with functionality of metal ion bound aLA. We identify the thermodynamically destabilized and disordered residues as putative binding sites of protein with ligands. We verify the role of these residues via docking and force field minimization. Our analysis show that  $\beta$ 4GalT binds thermodynamically to  $\text{Ca}^{2+}$ -aLA complex via C-terminal tail residues like D116, Q117, W118 and L119, as suggested by mutational study. But no such thermodynamically favourable binding can be identified in case of  $\text{Mg}^{2+}$ -aLA complex. We predict that  $\text{Ca}^{2+}$ -aLA can bind to oleic acid through basic H32 of A2

helix and hydrophobic residues, namely, I59, W60 and I95 of interfacial cleft, with H32 playing a specific role. The stabilized residue W104 seems to play kinetic role in binding process. Oleic acid binding is much weaker for  $Mg^{2+}$ -aLA complex. The overall shrinkage of aLA upon  $Mg^{2+}$  binding may lead to reduced activity of the metallo-protein complex. Since equilibrium dihedral angle distributions can be readily extracted from simulated trajectories, our analysis may be easily extended to compare functionalities of bio-macromolecules in different conformational states.

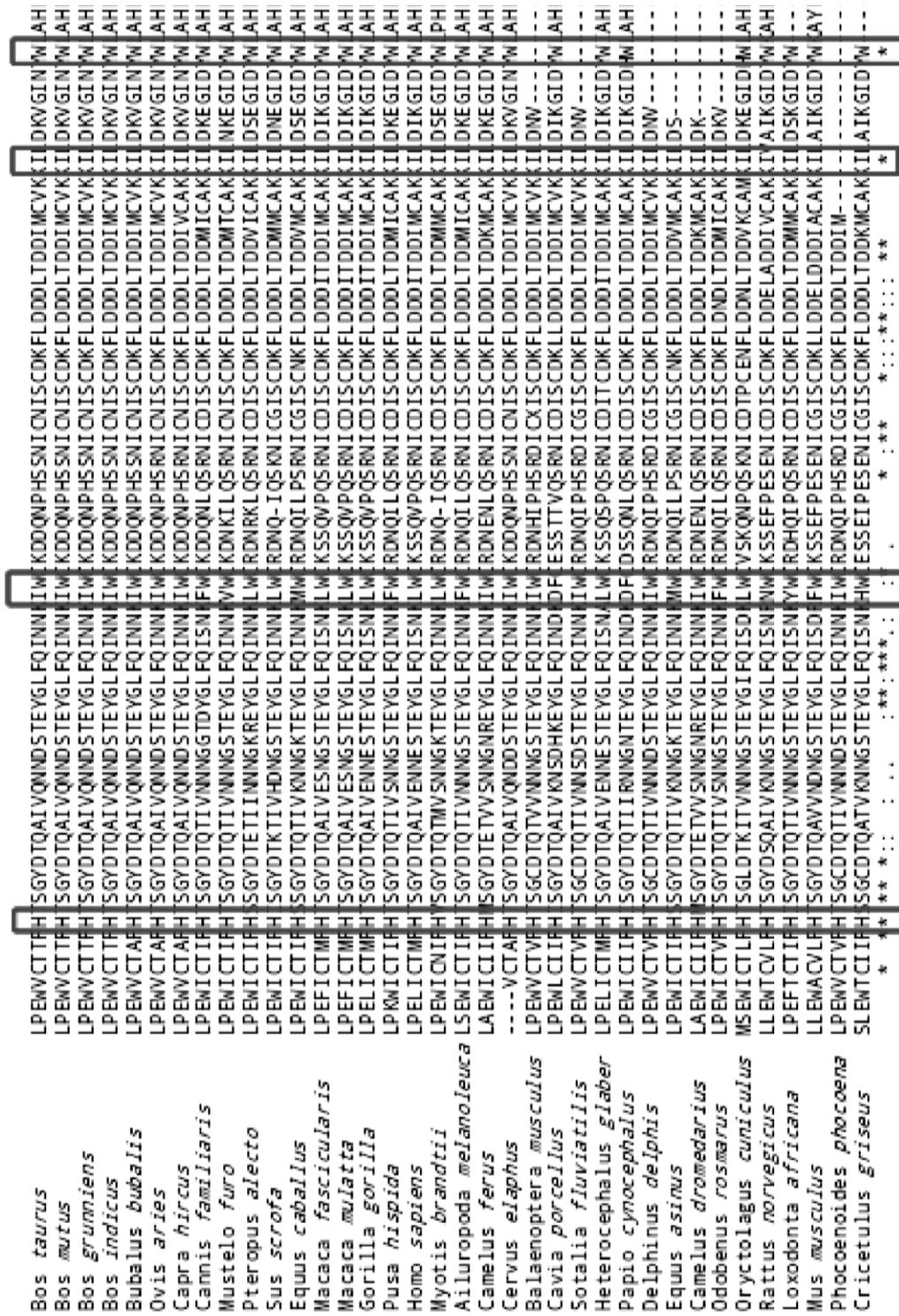


Fig. 2.6. Multiple Sequence Alignment of aLA sequences from different species. The predicted oleic acid binding residues, H32, W60, I95 and W104 are conserved; however, I59 shows weak conservation across the sequences.

# Appendix I

## Molecular dynamics (MD) simulation algorithm<sup>107</sup>

Let  $\vec{r}^N = \{\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N\}$  and  $\vec{p}^N = \{\vec{p}_1, \vec{p}_2, \dots, \vec{p}_N\}$  denote the position and momentum

vectors of the  $i = 1, 2, \dots, N$  atoms/particles in a 3-dimensional system, where the mass of

$i^{\text{th}}$  particle is  $m_i$  and the total force acting on it at time,  $t$  is  $\vec{F}_i$ . Assuming that the  $N$  point

particles interact through a conservative pair potential,  $V(r_{ij})$ , depending only on the pair

separation,  $r_{ij} = |\vec{r}_i - \vec{r}_j|$ , the force acting on particle  $i$  due to another particle,  $j$  is given by

$\vec{f}_{ij} = -\nabla_i V(r_{ij})$ . The total force acting on  $i^{\text{th}}$  particle is  $\vec{F}_i = \sum_{j=1, j \neq i}^N \vec{f}_{ij}$  and the total potential

energy of the system is sum over all pairs,  $V(r_{ij})$ , where each pair interaction is calculated only

once. The total kinetic energy is expressed as,  $K(\vec{p}^N) = \sum_{i=1}^N \frac{|\vec{p}_i|^2}{2m_i}$ . The Newton's equation

of motion for each particle can then be written as  $m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i$ . In MD simulation the new

positions and velocities for each atom is calculated by integrating the equation of motion in

forward time. The integration leads to new position,  $\vec{r}_i(t + \delta t)$  at time,  $t + \delta t$  in terms of known

position,  $\vec{r}_i(t)$  at previous time,  $t$ . For this purpose the Verlet algorithm<sup>108</sup> is used based on

Taylor series expansion about the positions,  $\vec{r}_i(t)$  as  $\vec{r}_i(t + \delta t) \approx 2\vec{r}_i(t) - \vec{r}_i(t - \delta t) + \frac{\vec{F}_i(t)}{m_i} \delta t^2$ . The

velocity is obtained using the formula,  $\vec{v}_i(t) = \frac{\vec{r}_i(t + \delta t) - \vec{r}_i(t - \delta t)}{2\delta t}$ . With the updated positions

the forces are re-calculated and the system evolves with time.

### Periodic boundary condition and minimum image convention

In order to minimize finite size effects, boundary conditions are applied to simulations. This is done by replicating the simulation box in all directions, thereby making the simulated system infinite. The periodic boundary conditions ensure that particles in the central simulation box are surrounded by image particles residing in the periodic replica boxes and as soon as a particle moves out of the central box, its periodic image re-enters through the opposite face into the central box during the course of simulation. The minimum image convention guarantees that each atom interacts only with the nearest atom or image in the periodic array.

## Appendix II

### Force field and simulation details

The interaction between the atoms constitutes the basis of force-fields. The form of the potential energy function is taken directly from CHARMM27 and given by the following equation<sup>109</sup>

$$\begin{aligned}
 V = & \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi - \delta)] \\
 & + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_u (u - u_0)^2 \\
 & + \sum_{\text{nonbonded}} \epsilon \left[ \left( \frac{R_{\text{min}_{ij}}}{r_{ij}} \right)^{12} - \left( \frac{R_{\text{min}_{ij}}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon r_{ij}}
 \end{aligned}$$

The first term in the energy function accounts for the bond stretches where  $k_B$  is the bond force constant and  $b - b_0$  is the distance from equilibrium that the atom has moved. The second term in the equation accounts for the bond angles where  $k_\theta$  is the angle force constant and  $\theta - \theta_0$  is

the angle from equilibrium between 3 bonded atoms. The third term is for the dihedrals where  $k_\phi$  is the dihedral force constant,  $n$  is the multiplicity of the function,  $\phi$  is the dihedral angle and  $\delta$  is the phase shift. The fourth term accounts for the impropers, that is out of plane bending, where  $k_\omega$  is the force constant and  $\omega - \omega_0$  is the out of plane angle. The Urey-Bradley component (cross-term accounting for angle bending using 1,3 nonbonded interactions) comprises the fifth term, where  $k_U$  is the respective force constant and  $U$  is the distance between the 1,3 atoms in the harmonic potential. Nonbonded interactions between pairs of atoms are represented by the last two terms. By definition, the nonbonded forces are only applied to atom pairs separated by at least three bonds. The van Der Waals (VDW) energy is calculated with a standard 12-6 Lennard-Jones potential and the electrostatic energy with a Coulombic potential.

Bonded interactions only involve next neighbours, so they are cheap in terms of computer time. For the non-bonded energy (LJ, electrostatic) it is customary to introduce a cut-off beyond which interactions are ignored. This is a reasonable approximation for the vdW (= LJ) interactions, which decay rapidly for large distances.

### **Particle Mesh Ewald**

For periodic systems Ewald summation method is employed for electrostatic part. Long range forces have spatial interaction that fall off no faster than  $r^{-d}$ ,  $d$  being the dimensionality of the system. Dipole-dipole and charge-charge interactions are such long range forces which pose serious complications in computer simulations. These long range interactions are treated via the Ewald summation<sup>110</sup> method which includes the interaction of the ion or molecule with all its periodic images. It is mandatory to ensure electro-neutrality of the system prior to Ewald summation. During the course of simulation the charge distribution in the central box constitutes

the unit cell for a neutral lattice extending throughout space. Each point charge in the system is surrounded by a Gaussian type charge distribution of equal magnitude and opposite sign spreading out radially from the charge. The distribution screens the interaction between neighboring charges. The screened interactions are short ranged and the total screening potential is computed through sum over all atoms of in the central box and their images in real space lattice of replica boxes. Further a cancelling charge distribution of same shape but with same sign of the original point charge is added such that the overall potential reduces down to that of the original charge.

Prior to MD simulation we solvate the proteins with explicit water and add counter ions like sodium and chloride to ensure electro-neutrality. The MD simulations of different proteins in explicit water are performed with the NAMD<sup>111</sup> program at 310 K and 1 atm pressure in isothermal-isobaric (*NPT*) ensemble using periodic boundary conditions and 1 femto-second time-step. The CHARMM27<sup>109</sup> force field and the TIP3P water model have been used in these simulations. The long ranged electrostatic interactions are treated by particle-mesh Ewald method. The equilibration of simulations is ensured from the root mean squared deviation (RMSD) plots.

## Appendix III

### **Histogram Based Method (HBM) for computing the thermodynamics**

We consider the equilibrated ensemble of structures for further calculation. We compute all the dihedral angles of the protein and generate individual distributions for each of them. A detailed description of the HBM is reported in a recent study<sup>23</sup>. The normalized probability distribution of a protein dihedral  $\zeta$  in a given conformational state  $C_i$  is given by the histogram  $H^{C_i}(\zeta)$ . The peaks of the histograms represent the equilibrium value of the dihedral angles. The

equilibrium free energy change associated with any protein dihedral  $\zeta$  upon transition from conformational state  $C_1$  to  $C_2$  is defined as

$$\Delta G^{conf}(\zeta) = -k_B T \ln[H_{\max}^{C_2}(\zeta) / H_{\max}^{C_1}(\zeta)], \quad (1)$$

where ‘max’ denotes the peak value. The conformational entropy for a dihedral in a given conformational state  $C_i$  is estimated using the Gibbs entropy formula,

$$S_{C_i}^{conf}(\zeta) = -k_B \sum_j H_j^{C_i}(\zeta) \ln H_j^{C_i}(\zeta), \quad (2)$$

where the sum over  $j$  indicates that over the histogram bins. Therefore, the entropy change for the given dihedral between any two conformational states is denoted by the difference,

$$\Delta S^{conf}(\zeta) = S_{C_2}^{conf}(\zeta) - S_{C_1}^{conf}(\zeta). \quad (3)$$

The thermodynamic variables of conformational changes are additive owing to the independent distribution of dihedral angles. The thermodynamic changes of a residue are given by adding the constituent backbone and side chain dihedral contributions and the overall change is calculated by adding the individual residue contributions.

## **CHAPTER 3**

# Conformational thermodynamics guided structural reconstruction of bio-molecular fragments

### **3.1 Introduction**

Quite often bio-molecular structure is only partially resolved due to either experimental limitations or internal flexibility. This results in missing coordinates of fragments in the molecular structure. Such fragments in proteins are commonly long loops or partially structured regions. In depth understanding of bio-molecular function remains elusive in absence of complete structure.

Present methods implemented to model missing fragments can be broadly classified into two categories<sup>112</sup>: (i) Ab-initio (*de-novo*) technique where extensive sampling of conformational space through Monte Carlo and/or molecular dynamics (MD) simulation is performed for a given sequence of missing fragment. (ii) Comparative (homology) modelling based on template identification from structural databases (e.g. PDB) having similarity with the missing target sequence<sup>113</sup>. The template is then used as a restraint to generate a viable conformation of missing fragment, followed by side chain optimization and energy minimization. The *de-novo* method works well for generating loop conformations for large missing fragments<sup>114</sup>. On other hand, success of homology modelling heavily relies on target-template sequence alignment.

Sometimes choice of modelling protocol is not obvious in particular when dealing with partially structured fragments. We consider the example of skeletal Troponin C (TnC). It is a

calcium ( $\text{Ca}^{2+}$ ) binding metallo-protein, exhibiting a helix-loop-helix EF-hand metal ion binding motif<sup>35</sup>, involved in  $\text{Ca}^{2+}$  dependent regulation of muscle contraction and relaxation process. The crystal structure of four  $\text{Ca}^{2+}$  bound TnC (PDB id 1Y TZ)<sup>36</sup> reveals a dumbbell shaped structure with two globular domains, N-terminal (EF hand sites 1 and 2) and C-terminal (EF-hand sites 3 and 4) separated by an extended central helix<sup>37, 38</sup> (F74-D105). TnC *in vivo* forms a part of hetero-trimeric Troponin complex where TnI and TnT are the other binding partners. While TnT interacts only with TnI, the latter binds directly to TnC to both N- and C- terminal domains in presence of four  $\text{Ca}^{2+}$  ions ( $4\text{Ca}^{2+}$ -TnC). However, TnI interacts only with C-terminal domain in the partial (C-terminal) metal ion bound TnC ( $2\text{Ca}^{2+}$ -TnC,  $2\text{Mg}^{2+}$ -TnC). N-terminal residues (1-47) of TnI remain bound to the C-terminal domain of  $2\text{Ca}^{2+}$ -TnC in muscle relaxation phase. During muscle contraction both N-terminal and C-terminal (117-131) of TnI binds to the C-terminal and the N-terminal domains of  $4\text{Ca}^{2+}$ -TnC, respectively.

No molecular structure is available for metal ion free (apo) TnC. The structure of apo-TnC is essential to understand the molecular basis of function of metal ion bound TnC<sup>39</sup>. Previous studies using circular dichroism and fluorescence on isolated domains of metal ion free TnC indicate that isolated N-domain is more structured compared to C-domain<sup>115</sup>. Small angle X-ray scattering (SAXS) and dynamic light scattering (DLS) measurements indicate an extended conformation of  $\text{Ca}^{2+}$  bound TnC compared to apo form. The central linker is hypothesized to be flexible in apo-TnC allowing inter-domain mobility<sup>116</sup>. The flexibility of this linker can be captured by a loop like structure resulting from *de-novo* method. However, flexibility does not necessarily exclude partial structures. For instance, similar proteins in the EF-hand family like calmodulin exhibits a broken central helix linker in apo state<sup>117</sup>. Similarly, nuclear magnetic resonance (NMR) structure of  $\text{Ca}^{2+}$  saturated TnC shows flexible linker with broken helix

conformation. Thus a possible structured model of the linker cannot be ruled out. Hence, it might be important to apply both modelling protocols for apo-TnC and then compare their relative merits to represent the apo-TnC conformation. More specifically, can one select a model which represents the missing part more reliably? This question is particularly subtle in case of large molecules like a protein, where the atoms are subject to strongly heterogeneous interaction. The free energy landscapes for such systems are highly rugged<sup>118</sup> so that the true equilibrium state may not be accessible starting from any arbitrary initial conformation within reasonable computation time.

Although structural fragments modelled via *de-novo* or homology methods can be refined by a variety of existing techniques<sup>119, 120</sup>, yet there is no method to the best of our knowledge to compare the merit of different modelled structures to represent conformations of a missing region. We propose a systematic way<sup>40</sup> for comparing ensembles of conformations generated by different existing modelling protocols. Our method selects the thermodynamically favourable conformation based on differences in conformational thermodynamic quantities. The interplay between conformation and function of a protein has been examined in chapter 2 using HBM<sup>34</sup> (Appendix III, chapter 2). A negative change in conformational free energy indicates stabilization, while that in entropy reflects conformational ordering. Similarly, residues undergoing positive change in free energy are destabilized and that in entropy are disordered. Generally destabilized and disordered residues participate in functional ligand binding<sup>34</sup> in a certain conformational state when compared to another. In the same token performance of different apo conformations generated using various protocols may also be compared from known functional perspectives, using this principle.

## 3.2 Methods

### Preparation of initial models of apo-TnC

In the absence of a full length apo-TnC structure, it can only be modelled based on available partial or metal ion saturated (2Mg<sup>2+</sup>-, 2Mn<sup>2+</sup>-, 2Cd<sup>2+</sup>-, 2Ca<sup>2+</sup>-, 4Ca<sup>2+</sup>-) TnC structures by depleting the metal ions. SAXS and DLS experiments have confirmed a collapsed apo-TnC conformation with flexible central linker<sup>116</sup>. A flexible central linker is indicated in the X-ray crystal structure of Mg<sup>2+</sup> ion bound TnC (PDB id 1YV0)<sup>36</sup> from chicken skeletal muscle. However, residues M85 to E95 corresponding to central helix are missing in the 1YV0 structure. We model a full length apo-TnC structure by depleting the two Mg<sup>2+</sup> ions from C-terminal domain of 1YV0 structure and adding the eleven missing residues of central helix.

The missing fragment of M85 to E95 is constructed via (i) *de-novo* (model A) and (ii) template based homology (model B) modelling. The *de-novo* based model A is prepared in Swiss PDB Viewer<sup>121</sup>, where missing residues are added one-by-one onto the carboxy terminus of Q84 and finally ligated to the amino terminus of E96. In order to optimize the geometry of each added residue, a vacuum energy minimization is performed at every step using the implemented GROMOS96 force field. The template based model B is prepared by ligating the fragment of residues M85-E95 from manganese (2Mn<sup>2+</sup>-) bound TnC (PDB id 1NCY)<sup>122</sup> in place of the missing residues in 1YV0 followed by energy minimization in vacuum. We consider 1YV0 and 1NCY in modelling apo-TnC to avoid any Ca<sup>2+</sup> induced bias.

The initial conformations of both 2Ca<sup>2+</sup>-TnC and 4Ca<sup>2+</sup>-TnC from chicken skeletal muscle are taken from the PDB entries 1TOP<sup>123</sup> and 1YTZ<sup>36</sup>, respectively. Hydrogen atoms are added to each of the four systems and then solvated in a rectangular parallelepiped water box. The systems are neutralized through addition of sodium (Na<sup>+</sup>) and chloride (Cl<sup>-</sup>) ions and MD

simulations are performed using standard protocols defined in Appendices I and II, chapter 2.

### **Docking studies**

The average structures for both  $2\text{Ca}^{2+}$ -TnC and  $4\text{Ca}^{2+}$ -TnC are calculated from their respective equilibrated trajectories. The structure of TnI is obtained from crystal structure PDB id: 1Y TZ. The PDB coordinates of N- and C-fragments of TnI consists of E3-H47 and M116-K131, respectively. The  $2\text{Ca}^{2+}$ -TnC and the  $4\text{Ca}^{2+}$ -TnC are used as receptor for TnI fragments in the docking procedure performed with Autodock4<sup>124</sup>. We bias the docking by considering destabilized and disordered residues in  $2\text{Ca}^{2+}$ -TnC and  $4\text{Ca}^{2+}$ -TnC as flexible. The resulting docked complexes are sorted by minimum energy criteria and root mean square deviation (RMSD) clustering. The interfaces of the docked complexes are analyzed based on distance restraints of  $\sim 5\text{\AA}$  between the binding partners. We further minimize the docked complexes in explicit solvent.

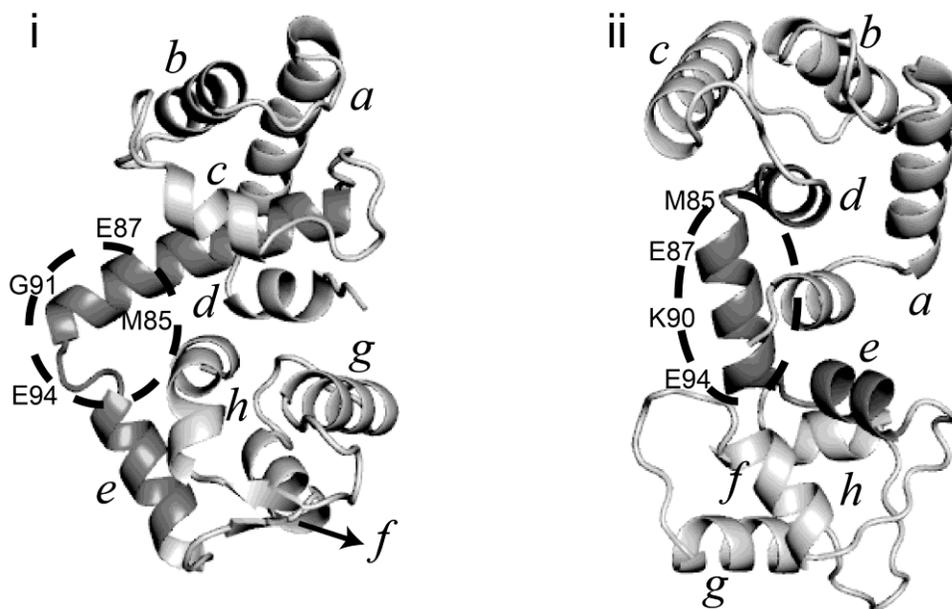
### **BLAST**

Basic Local Alignment Search Tool (BLAST)<sup>102</sup> is an algorithm to compare amino acid sequences of different proteins. A BLAST search enables one to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold. To run the software, BLAST requires a query sequence to search for, and a sequence to search against (also called the target sequence) or a sequence database containing multiple such sequences. BLAST will find sub-sequences in the database which are similar to sub sequences in the query. BLAST searches for high scoring sequence alignments between query sequence and existing sequences in the database using a heuristic approach.

### 3.3 Results

#### A. Comparison of apo-TnC models

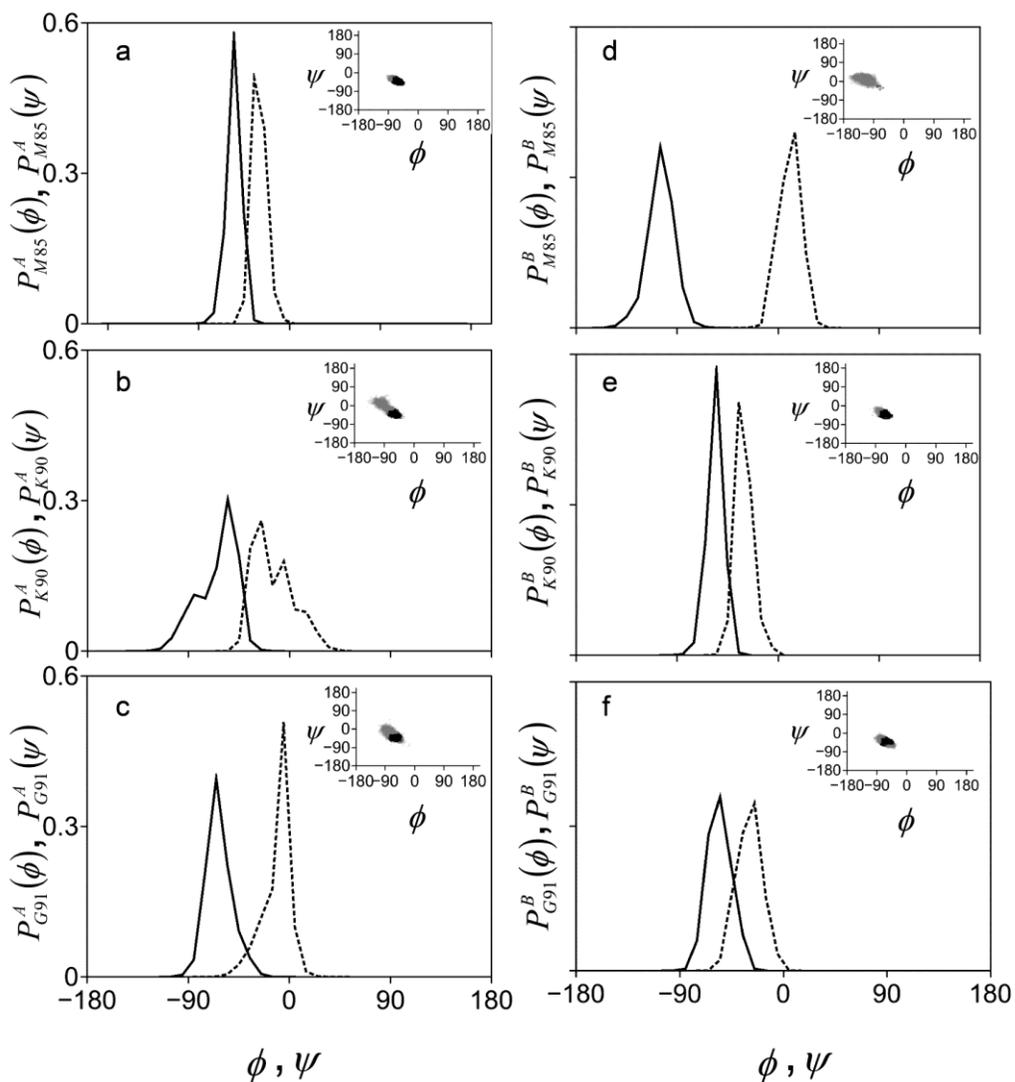
The average conformations from respective ensembles (model A, Fig 3.1 i and model B, Fig 3.1 ii) are generated for each model through structural alignment and averaging the atomic coordinates. The first 6 out of 11 modelled residues in the fragment M85 to E95 are fused with helix *d* in model A. In contrast, the modelled fragment connecting helix *d* in N-terminal and helix *e* in C-terminal shows helical structure in model B except at terminal residues. The calculated radius of gyration,  $R_g \sim 16.7 \text{ \AA}$  for model A and  $\sim 18.7 \text{ \AA}$  for model B, are similar to reported SAXS and DLS values ( $R_g \sim 18.3 \text{ \AA}$ )<sup>116</sup>.



**Fig 3.1. The ensemble averaged conformations of (i) model A and (ii) model B of apo-TnC. The modelled fragment (M85-E95) of central helix is encircled.**

The equilibrium probability distributions of backbone dihedral angles are shown in Fig 3.2 for a representative set of residues belonging to the modelled region of central helix. We denote the equilibrium dihedral distributions as  $P_i^A(\phi)$  (solid line) and  $P_i^A(\psi)$  (dashed line)

corresponding to backbone dihedral angles  $\phi$  and  $\psi$  of the  $i^{\text{th}}$  residue in model A and similarly  $P_i^B(\phi)$  and  $P_i^B(\psi)$  for model B. We use the histograms corresponding to equilibrium distributions of dihedral angles of model A and model B to calculate conformational



**Fig 3.2.** Equilibrium distributions of backbone dihedral angles,  $P_i^A(\phi)$  (black solid line),  $P_i^A(\psi)$  (black dashed line) of (a) M85, (b) K90, (c) G91 in model A and similarly  $P_i^B(\phi)$  (black solid line),  $P_i^B(\psi)$  (black dashed line) of (d) M85, (e) K90, (f) G91 in model B. The insets represent the corresponding  $\phi$ - $\psi$  scatter-plots, where helix conformation is shown in black dots and loop conformation in grey.

thermodynamics. Conformational free energy change of model B with respect to model A

associated with a particular dihedral is calculated as ratio of peaks of equilibrium dihedral angle distribution in model B to that in model A and conformational entropy change using the Gibb's formula (Appendix III, chapter 2). The stabilized residues in model B with respect to model A are shown in green, while the destabilized residues in yellow (Fig 3.3a (i)). Similarly, the ordered residues are coloured green and the disordered ones yellow (Fig 3.3a (ii)). The N-terminal domain comprising of residues 3-80 is observed to be more stabilized and ordered ( $\Delta G_{N-ter}^{conf} = -32.8$  kJ/mol,  $T\Delta S_{N-ter}^{conf} = -65.8$  kJ/mol) compared to C-terminal domain residues 97-161 ( $\Delta G_{C-ter}^{conf} = -19.8$  kJ/mol,  $T\Delta S_{C-ter}^{conf} = -25.5$  kJ/mol) in model B, in agreement to experimental results<sup>115</sup>. Overall we find that the ensemble of model B is more stable and ordered in conformational free energy (-58.5 kJ/mol) and entropy (-119.0 kJ/mol) with respect to the ensemble of model A. The backbone dihedral angles in model B contribute more towards overall stability in free energy (>50%) and the side chains dihedrals predominantly contribute to conformational ordering (>75%). The gain in stability in conformational free energy of model B with respect to model A indicates that model B is a better representative of ensemble of conformations of apo-TnC.

We now probe if secondary and tertiary structural features can make a reasonable distinction between these two models. The secondary structure of the modelled fragment of residues, M85-E95 show characteristic differences between model A and model B. The scatter-plots in the insets of Fig 3.2 represent sampling of backbone dihedral conformational space and differentiate helical conformation (black) from loop-like (gray). The backbone dihedral distributions  $P_{M85}^A(\phi)$  and  $P_{M85}^A(\psi)$  (Fig 3.2a) show single peak at  $\phi = -55^\circ$  and  $\psi = -35^\circ$  respectively, characteristic of helix like conformation. The  $\phi - \psi$  scatter-plot of M85 (inset Fig

3.2a) also confirms helix-like conformation. The distributions  $P_{K90}^A(\phi)$  and  $P_{K90}^A(\psi)$  (Fig 3.2b) are both bimodal. The peak at  $\phi=-55^\circ$  and  $\psi=-25^\circ$  corresponds to helix, while the peak at  $\phi=-85^\circ$  and  $\psi=-5^\circ$  implies loop-like conformational state. The scatter-plot of K90 (inset Fig 3.2b) exhibits two distinct clusters corresponding to both helix and loop conformations in model A. The probability distributions of backbone dihedral angles  $P_{G91}^A(\phi)$  and  $P_{G91}^A(\psi)$  (Fig 3.2c) are characteristic of loop conformation with peaks at  $\phi=-65^\circ$  and  $\psi=-5^\circ$ . The corresponding scatter-plot of G91 (inset Fig 3.2c) in model A indicates large clustering in loop region. The equilibrium distributions  $P_{M85}^B(\phi)$  and  $P_{M85}^B(\psi)$  (Fig 3.2d) show single peak at  $\phi=-105^\circ$  and  $\psi=15^\circ$ . The  $\phi - \psi$  scatter-plot of M85 in model B (inset Fig 3.2d) predominantly shows loop conformation. The backbone dihedral distributions  $P_{K90}^B(\phi)$  and  $P_{K90}^B(\psi)$  (Fig 3.2e) shows characteristic of helix like conformation. The scatter-plot of K90 in model B (inset Fig 3.2e) has clustering in helical region. The peaks of  $P_{G91}^B(\phi)$  and  $P_{G91}^B(\psi)$  (Fig 3.2f) corresponds to helical structure which is also evident from the  $\phi - \psi$  cluster for G91 in model B (inset Fig 3.2f).

We calculate the helical propensity of individual modelled residues, summarized in Table 3.1, from scatter-plots as the ratio of population in helical region to total population. M85, K86, E87, D88 and A89 in model A predominantly acquire helical conformation. K90 samples both helix and loop-like conformations, albeit slightly larger population for helix. On other hand, G91, K92 and S93 show decreasing propensity for helix and preference to loop conformation, while E94 and E95 adopt loop-like conformation. Thus ~45% of the residues in the modelled region prefer loop conformation in model A. On other hand in model B, M85 has exclusively loop-like conformation. K86, E87, D88, A89, K90, G91, K92, and S93 show strong propensity for helix. E94 has nearly 60% while E95 has ~40% helical propensity. The ensemble for model B

shows preference for helical conformation in the modelled region except for terminal residues.

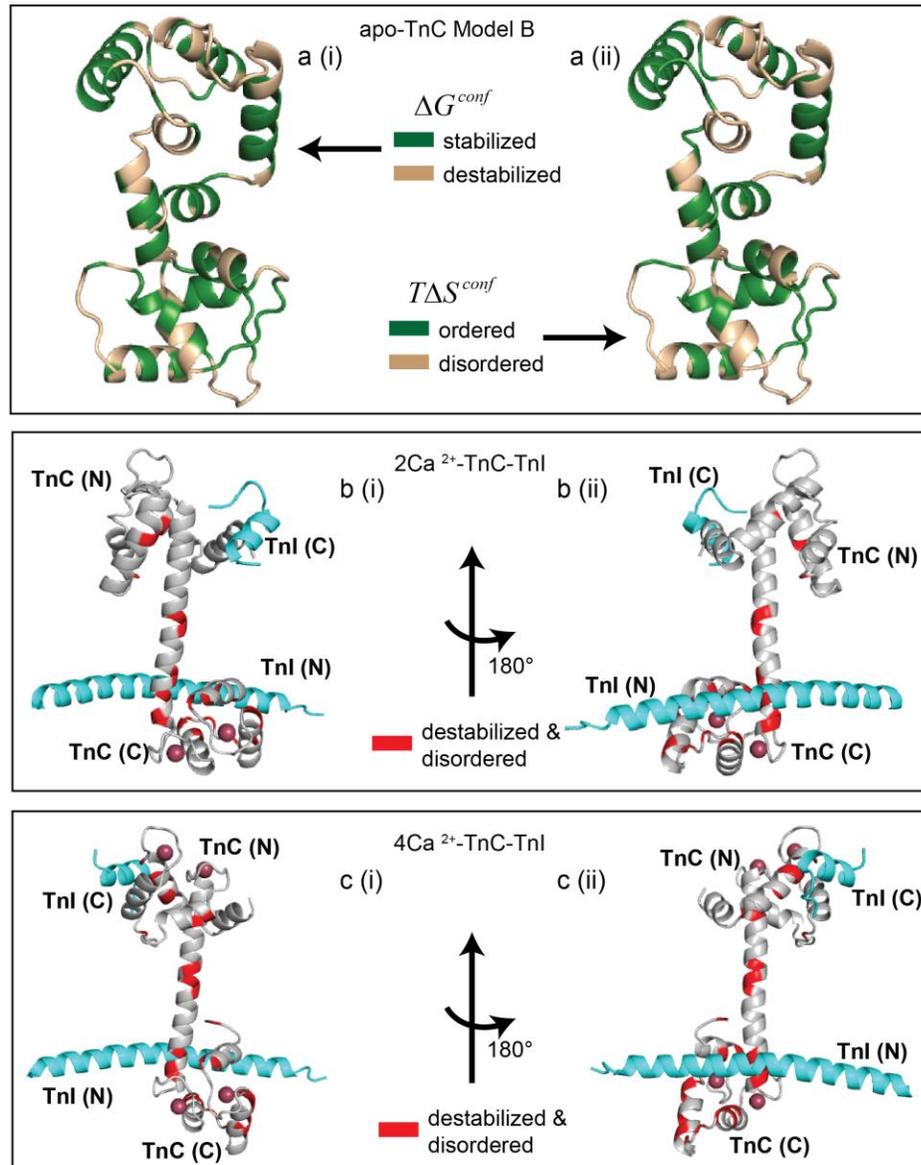


Fig 3.3. (a) Conformational thermodynamic changes in model B with respect to model A, illustrating (i) stabilized (green) and destabilized (yellow) residues (ii) ordered (green) and disordered (yellow) residues. (b) The energy minimized 2Ca<sup>2+</sup>-TnC-TnI docked complex where the destabilized and disordered interfacial residues of 2Ca<sup>2+</sup>-TnC with respect to apo model B are shown in red. (c) The energy minimized 4Ca<sup>2+</sup>-TnC-TnI docked complex where the destabilized and disordered interfacial residues of 4Ca<sup>2+</sup>-TnC with respect to apo model B are shown in red. The N- and C-terminal TnI fragments: TnI (N) and TnI (C) are shown in cyan. (i) and (ii) in (b) and (c) show the 180° rotated view. The N and C domains of TnC are marked by TnC (N) and TnC (C) respectively.

**Table 3.1. Helix propensity (%) of the modelled residues in model A and model B of apo-TnC.**

<b>RESIDUE</b>	<b>Model A</b>	<b>Model B</b>
<b>M85</b>	97.3	0.1
<b>K86</b>	98.1	97.5
<b>E87</b>	98.7	95.4
<b>D88</b>	97.7	96.6
<b>A89</b>	91.3	98.6
<b>K90</b>	59.7	96.3
<b>G91</b>	34.5	88.4
<b>K92</b>	26.0	94.8
<b>S93</b>	19.4	92.8
<b>E94</b>	~0	63.4
<b>E95</b>	~0	43.1

In order to identify any preferred conformation of the modelled fragment residues we compare conformations in other proteins. For this we consider the sequence of modelled residues as a query sequence to search for similar sequences from PDB whose structures are known. We perform a thorough protein Basic Local Alignment Search Tool (BLAST)<sup>102</sup> search with the query sequence “*QMKEDAKGKSEEE*” including immediate flanking residues for both termini of the modelled fragment. The search yields several homologous sequences with variable similarity and length compared to the query sequence including sequences from TnC. We find 55 sequences from PDB having 30%-80% similarity with the query sequence excluding the TnC hits.

We assign the secondary structures from DSSP for these 55 PDB entries and calculate their frequencies for acquiring helical conformation. The data for E87, D88, E94 and E95 are statistically significant. Table 3.2 shows that these residues have high preference for helix like conformation. In the ensemble of model A, E87 and D88 acquire helical conformation but E94 and E95 show loop conformation (Table 3.1). Model B shows that all these residues acquire

predominantly helical structure except E95 which show slightly higher propensity for loop like conformation. Thus model B produces better representation of structural preferences shown by similar sequence of residues in PDB when compared to model A.

**Table 3.2. Frequency percentage of secondary structure.**

<b>RESIDUE</b>	<b>HELIX (%)</b>	<b>BETA (%)</b>	<b>LOOP (%)</b>
<b>E87</b>	65	0	35
<b>D88</b>	58	9	33
<b>E94</b>	85	7.5	7.5
<b>E95</b>	88	8	4
<b>E96</b>	88	8	4

### **B. Functional implications of Ca<sup>2+</sup> ion bound TnC**

We determine TnC-TnI interfacial residues from 1Ytz structure applying a 5Å cut-off atomic distance (backbone and side-chain non-hydrogen atoms) criterion between residues of TnC and TnI.

We compute the thermodynamics of conformational changes in 2Ca<sup>2+</sup>-TnC with respect to both model A and model B of apo-TnC. There are 4 residues both destabilized and disordered at N-terminal and 8 such residues at C-terminal binding interface of 2Ca<sup>2+</sup>-TnC with respect to model A. We denote these residues as set A, which includes 3 hydrophobic residues, V128, M157 and V160 at the C-terminal domain. There are 16 residues both destabilized and disordered at the C-terminal interface and 3 at the N-terminal interface of 2Ca<sup>2+</sup>-TnC when compared to apo model B. We denote these destabilized residues as set B (Fig 3.3b) where majority of the residues are hydrophobic, L97, I103, L121, V128, F150, F153, M157 and V160 and belong to the C-terminal domain.

The residues of both set A and B are used as bias for docking 2Ca<sup>2+</sup>-TnC with N- and C-terminal binding fragments of TnI. The two docked complexes are then energy minimized in explicit solvent. We find that docking biased with set B has a lower energy than that with set A.

Fig 3.3b shows that the N-terminal TnI fragment (R13 to A33), TnI (N), binds to the C-terminal domain of  $2Ca^{2+}$ -TnC, while its C-terminal fragment, TnI (C) remains 15 Å far from the hydrophobic core of N-domain of  $2Ca^{2+}$ -TnC, consistent with experimental observations. In the docked complex, I103 of  $2Ca^{2+}$ -TnC interacts with A25 and I29 of TnI (N); M157, V160 ( $2Ca^{2+}$ -TnC) interacts with M21, L22 of TnI (N). Electrostatic interaction exists between R102, E127, D133 ( $2Ca^{2+}$ -TnC) and E32, R14, K18 of TnI (N). Thus TnI binding to  $2Ca^{2+}$ -TnC can be understood when apo-TnC conformation is represented by the ensemble of model B.

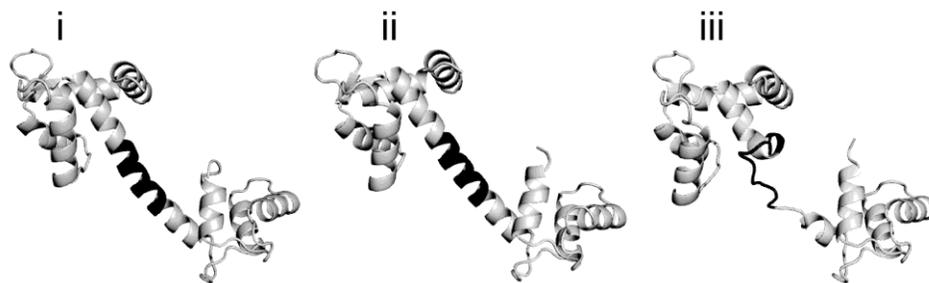
To understand further aspects of  $Ca^{2+}$  bound TnC, the conformational thermodynamic changes of  $4Ca^{2+}$ -TnC are computed with respect to apo model B. The N-terminal and C-terminal domains of  $4Ca^{2+}$ -TnC harbour residues which are both destabilized and disordered (Fig 3.3c). For instance, such residues in N-terminal interface are hydrophobic, F12, F21, A23, A24, F28 and M80, polar, S14, Q50 and acidic E63. Similarly, in C-terminal domain hydrophobic residues I103, I120, L121, A123, V128, F150, F153 and V160, polar C100, T124, S140 and acidic E126, D132, D135, D139 are destabilized as well as disordered. The enhanced disorder in the residues comes predominantly from exposed side-chains of hydrophobic residues in agreement with bio-chemical<sup>125</sup> and NMR<sup>126-129</sup> studies.

These destabilized and disordered residues from the two domains are considered for docking TnI fragments with  $4Ca^{2+}$ -TnC. The docked complex is subsequently energy minimized and shown in Fig 3.3c. We find that TnI (N) docks near C-terminal domain of  $4Ca^{2+}$ -TnC and TnI (C) near N-terminal of  $4Ca^{2+}$ -TnC. Predominant hydrophobic contacts are observed between the binding partners involving destabilized and disordered hydrophobic residues of  $4Ca^{2+}$ -TnC. For instance, the N-terminal residues A23, A24, M27, F28, M47 and L48 of  $4Ca^{2+}$ -TnC show hydrophobic interaction with A118, M121, L122, L125 and L126 of TnI (C). The acidic residues,

E16 and E20 of  $4\text{Ca}^{2+}$ -TnC show electrostatic interaction with basic K129 and K131 of TnI (C). At C-terminal domain of  $4\text{Ca}^{2+}$ -TnC, I103, F104, I120, M156 and M157 make hydrophobic contact with M21, L22, L24, A25, V26 and I29 of TnI (N), while E126 of  $4\text{Ca}^{2+}$ -TnC interacts with R13 of TnI (N).

### C. Comparison with apo-TnC models from different servers

There are several molecular structure modelling servers based on homology or de-novo methods. We model full apo-TnC conformation through popular servers like Rosetta<sup>130</sup>, RaptorX<sup>131</sup> and ModLoop<sup>132</sup> for homology modelling. The predicted model from Rosetta uses the template of  $2\text{Ca}^{2+}$ -TnC (PDB id: 1TOP) while that from RaptorX uses the template  $2\text{Cd}^{2+}$ -TnC (PDB id: 1NCX). The constructed models from Rosetta and RaptorX exhibit an elongated dumb-bell shaped conformation with extended central helix (Fig 3.4) in contrast to collapsed conformation of apo-TnC<sup>116</sup>. We further build the missing fragment of apo-TnC based on the template, PDB id 1YV0 as a loop using the ModLoop<sup>132</sup> server. The resulting structure shows a collapsed conformation.

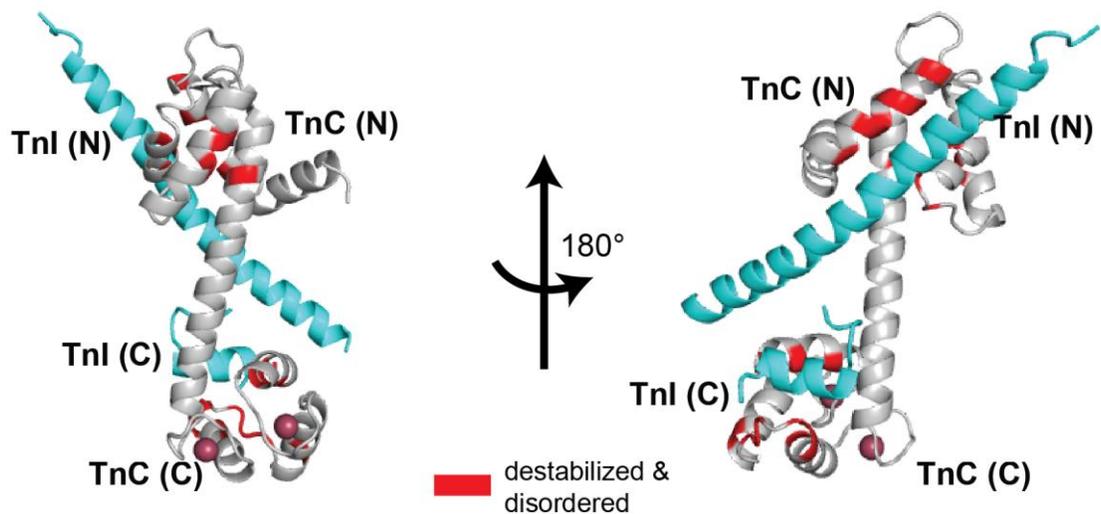


**Fig 3.4. Homology based models of apo-TnC from (i) Rosetta using PDB template 1TOP, (ii) RaptorX using PDB template 1NCX, (iii) ModLoop using PDB template 1YV0. The central linker residues M85-E95 are coloured in black.**

### D. Conformational thermodynamic comparison of ModLoop model

Since the ModLoop model shows a collapsed apo-TnC conformation similar to model A and model B, we perform MD simulation using this model in order to refine the structure and

generate an ensemble for further comparison. The radius of gyration,  $R_g \sim 21.7 \text{ \AA}$  is in contrast to experiment ( $R_g \sim 18.3 \text{ \AA}$ )<sup>116</sup> and that of model B ( $R_g \sim 18.7 \text{ \AA}$ ). The overall conformational thermodynamic changes indicate the ensemble of ModLoop model is destabilized compared to ensemble of model B. We also retrieve functional implications of  $2\text{Ca}^{2+}$ -TnC through its conformational thermodynamic changes with respect to the ensemble of ModLoop model. The number of destabilized and disordered residues in N-domain (12) and C-domain (11) of  $2\text{Ca}^{2+}$ -TnC are similar, with majority being hydrophobic residues. We bias the docking of N- and C-terminal fragments of TnI to  $2\text{Ca}^{2+}$ -TnC through these destabilized and disordered residues. However, the docked complex (see Fig 3.5) shows an improper binding mode in contrast to experiments, where TnI-N binds to the N-domain of TnC and TnI-C to the C-terminal domain of  $2\text{Ca}^{2+}$ -TnC. Thus, model B represents a better apo-TnC conformation than the ModLoop model.



**Fig 3.5.** The energy minimized  $2\text{Ca}^{2+}$ -TnC-TnI docked complex where the destabilized and disordered interfacial residues of  $2\text{Ca}^{2+}$ -TnC with respect to the ModLoop model are shown in red.

### E. Chemical shifts

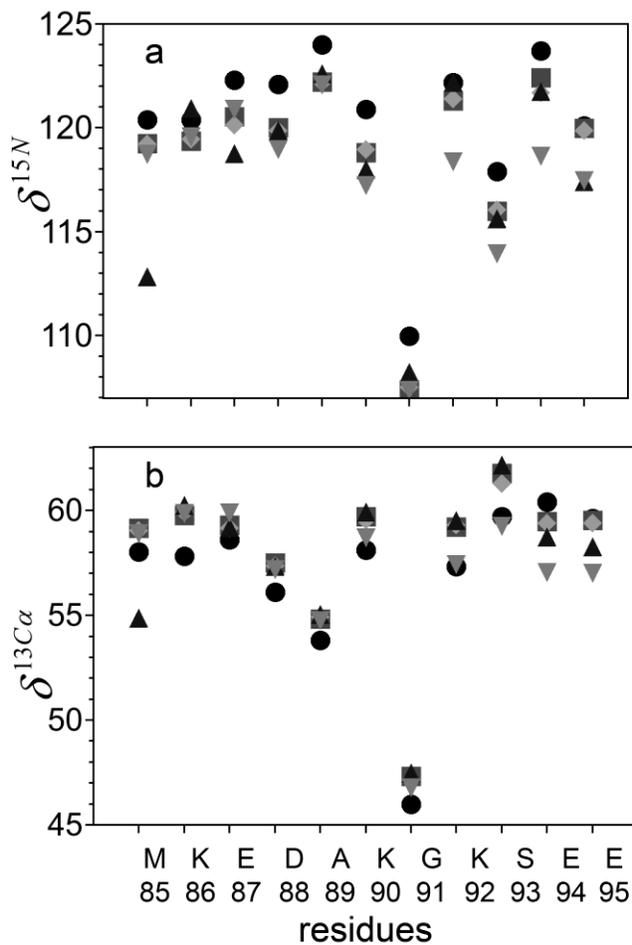
Our calculations using HBM indicates that model B is thermodynamically a favourable conformation for apo-TnC. The central linker in model B, between N and C domains has a

broken helical structure similar to that found in apo-calmodulin<sup>117</sup>. The simulated ensembles of 2Ca<sup>2+</sup>-TnC and 4Ca<sup>2+</sup>-TnC indicate a complete helical structure in that region similar to the PDB crystal structures<sup>36, 123</sup>. The Ca<sup>2+</sup> ion saturated structure of TnC (PDB id 1TNW)<sup>128, 129</sup> derived from NMR, however, show the linker region to have a flexible structure. In order to understand these differences in structures of central linker region of TnC we calculate backbone (N, C $\alpha$ ) chemical shifts for linker residues from SHIFTX2<sup>133</sup> using the ensemble averaged structures of 4Ca<sup>2+</sup>-TnC, 2Ca<sup>2+</sup>-TnC, model A and model B.

The calculated chemical shifts of <sup>15</sup>N amide ( $\delta^{15N}$ ) and <sup>13</sup>C $\alpha$  ( $\delta^{13C\alpha}$ ) are shown in Fig 3.6a and 3.6b respectively. The chemical shifts are plotted residue wise along with experimentally reported values for Ca<sup>2+</sup> saturated TnC<sup>128, 129</sup>. The calculated values  $\delta^{15N}$  and  $\delta^{13C\alpha}$  for central linker residues (M85-E95) of 4Ca<sup>2+</sup>-TnC show reasonable match with experimentally reported chemical shift values. This agreement indicates that central linker residues M85 to E95 share similar environment in the simulated and the NMR structure despite their structural differences. The calculated  $\delta^{15N}$  (Fig 3.6a) and  $\delta^{13C\alpha}$  (Fig 3.6b) values for the central linker residues of 4Ca<sup>2+</sup>-TnC and 2Ca<sup>2+</sup>-TnC are almost identical, as expected from the structural similarity in the region.

We find significant differences (~ 5ppm) in  $\delta^{15N}$  and  $\delta^{13C\alpha}$  between the predicted chemical shifts of the two apo models A and B. Both  $\delta^{15N}$  and  $\delta^{13C\alpha}$  of M85 in model B is upfield shifted (lower chemical shift value) compared to model A. The upfield shifts observed at times are reported due to solvent exposure or random coil like conformation<sup>134-136</sup> of residues. We find that M85 of model B is not only exposed to solvent in the ensemble but also show high propensity for random coil (Table 3.1). On the other hand M85 is buried in a hydrophobic pocket generated by L13, M17 and F21 and also acquire helical conformation in model A. Further,

$\delta^{15N}$  of K92 and  $\delta^{13C\alpha}$  of S93 in model B show significant downfield shift ( $\sim 3-4$  ppm) with respect to model A. These residues are exposed to solvent in model A, while they remain in close proximity to E8, A11 and F12 in model B.



**Fig 3.6. Chemical shift values of (a)  $\delta^{15N}$  (b)  $\delta^{13C\alpha}$  for linker residues: reported from NMR structure 4Ca<sup>2+</sup>-TnC, (PDB id 1TNW) (circle) and calculated from simulated structures of 4Ca<sup>2+</sup>-TnC (diamond), 2Ca<sup>2+</sup>-TnC (square), model A (down-triangle) and model B (up-triangle).**

## 3.4 Conclusion

In conclusion we show based on conformational thermodynamics that the central linker region of  $\text{Ca}^{2+}$  ion free skeletal muscle TnC has a broken helical conformation. This structure explains TnI binding to  $\text{Ca}^{2+}$  ion bound TnC. Our method provides a basis for selecting thermodynamically favourable conformation of a missing fragment in a protein if the modelling protocol for such fragment is not necessarily unique. Such method can be generalized to provide valuable insight into any bio-molecular structure from known functional perspective even in absence of molecular structure in a particular state.

## **CHAPTER 4**

# Conformational thermodynamic changes upon protein oligomerization

### **4.1 Introduction**

Protein aggregation is a manifestation of protein-protein interactions. Monomeric proteins often form stable aggregates of higher order called oligomeric states. Proteins like, actin and glutamate dehydrogenase exist in an aggregated state under native conditions. Such naturally occurring aggregates are essential for their proper functioning in regulating muscle movement and enzymatic activity, respectively.<sup>18, 19</sup> Another class of protein aggregation occurs due to protein misfolding and is often termed as unwanted aggregation state. Such unwanted aggregation as known for several proteins, like,  $\alpha$ -synuclein, amyloid  $\beta$ , polyglutamine and prions results in loss of protein function.<sup>137</sup> This can lead to numerous diseases like Alzheimer's, Parkinson's, Huntington's and prion diseases.<sup>20</sup> Thus the microscopic understanding of protein aggregation is important though not understood well. In this chapter we show how the conformational thermodynamics data may help to understand stability of protein oligomers.

*Salmonella Typhi* (*S. Typhi*) is a pathogenic gram negative bacterium responsible for the cause of typhoid fever.<sup>138-141</sup> Several strains of this bacterium exist with various degrees of virulence, like that of antibiotic resistance.<sup>142-144</sup> These strains of *S. Typhi* possess relatively uncharacterized yet important proteins belonging to yfdX family. Recently one such yfdX protein, STY3178 from multidrug resistant CT18 strain of *S. Typhi*<sup>143, 144</sup> has been characterized through biophysical techniques. This protein exists in a trimeric folded state of oligomerization

in solution<sup>42</sup> as confirmed by dynamic light scattering (DLS) and size exclusion chromatography experiments. STY3178 is reported to be stable at high temperatures and undergoes reversible thermal unfolding as well<sup>41</sup>. Further studies<sup>43</sup> indicate stability of the trimeric conformation of STY3178 for a wide range of pH.

The molecular structure of STY3178 is yet to be resolved. Owing to the absence of experimentally resolved structure of STY3178, a model monomer structure has been proposed based on homology modelling with respect to a template protein from *K. pneumoniae* (PDB id 3DZA). A symmetric trimeric assembly is generated by docking three such monomers, where each subunit shares two interfaces with the other two monomers.<sup>41,42</sup> The interfacial residues are similar across the three subunits and harbours acidic, basic, polar and hydrophobic residues.

In this chapter<sup>43</sup>, we understand the stability of trimeric oligomerization state of STY3178 from conformational thermodynamic changes of trimeric state with respect to monomer, using MD simulated trajectories. The conformational thermodynamic changes indicate oligomerization induced large gain in stability and ordering. The interfacial residues contribute predominantly to these changes, specially the acid-base pairs forming salt bridges between individual monomer subunits. The conformational thermodynamic changes account for observed stability at various pH in agreement to experimental observations.

## 4.2 Methods

MD simulation of 400 ns is performed for the modelled conformations of STY3178 monomer and trimer keeping the total number of atoms in both systems same using standard protocols (see Appendices I and II, chapter 2). Dihedral angle distributions are generated from equilibrated trajectories of both systems and conformational thermodynamic changes are estimated (as in Appendix III, chapter 2).

## 4.3 Results

### A. Conformational thermodynamic changes

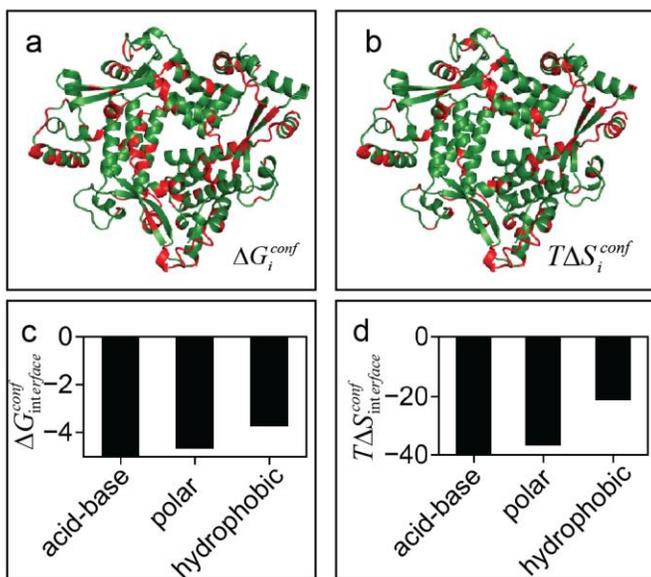
The conformational free energy and entropy changes are calculated from the equilibrium distributions of dihedral angles for each of the three sub-units (denoted by A, B and C) of trimer with respect to the monomer (Table 4.1). We find that the aggregation of STY3178 into the trimeric assembly induces large conformational stability and ordering with respect to the monomer. The three monomeric subunits within the assembly undergo similar gain in conformational stability and ordering. Owing to limited sampling we find small differences in free energy and entropy changes within the three monomeric subunits. Fig 4.1 (a) shows the distribution of conformational free energy changes having a predominance of stabilized residues (green) compared to few destabilized (red) residues. The conformational entropy changes (Fig 4.1 b) indicate that most of the residues undergo conformational ordering (green) with sparsely distributed disordered residues (red).

**Table 4.1. Conformational free energy ( $\Delta G^{conf}$ ) and entropy ( $T\Delta S^{conf}$ ) changes (in kJ/mol) of STY3178 trimer with respect to monomer.**

Chain id	$\Delta G^{conf}$	$T\Delta S^{conf}$
A	-55.0	-345.4
B	-61.9	-308.6
C	-64.4	-276.6

The conformational thermodynamic changes of the interfacial residues are calculated as an average over the three subunits (see Table 4.2). We find three salt bridges at the interface: D24-K163, D85-K118 and E161-K122. These residues contributing to salt bridge formation show negative changes in both free energy and entropy. The large ordering of these acidic and basic residues indicates loss of flexibility due to salt bridge formation. The trimeric interface shows overall gain in stability (Fig 4.1 c) and ordering (Fig 4.1 d). We find that the acidic and

basic residues at the interface contribute largely to lower both conformational free energy and entropy followed by polar and hydrophobic residues. The interfacial thermodynamics changes indicate that inter-chain salt bridges between acidic and basic residues tend to stabilize the trimeric assembly. The overall conformational thermodynamic changes along with the interfacial thermodynamics complement the experimentally observed stability of the trimeric assembly.



**Fig 4.1.** The conformational thermodynamic changes of STY3178 trimer with respect to monomer showing (a) changes in free energy,  $\Delta G_i^{conf}$  for stabilized (green) and destabilized (red) residues and (b) changes in entropy,  $T\Delta S_i^{conf}$  for ordered (green) and disordered (red) residues. The contributions of interfacial residues to changes in (d) free energy,  $\Delta G_{interface}^{conf}$  and (e) entropy,  $T\Delta S_{interface}^{conf}$  of the trimeric assembly averaged over three chains.

## B. Comparison with experiments

The experimentally observed stability of STY3178 trimer assembly at high temperatures and a wide range of pH can be understood in the following way: The conformational thermodynamic changes at interfacial residues indicate that inter-chain acid-base interactions predominantly contribute to stability of the trimeric complex. Such acid-base interactions lead to formation of salt bridges, which remain unaffected at high temperatures. The surface exposed acidic and basic residues get affected through changes in their protonation state upon changes in pH. The trimeric assembly harbours only a few such surface exposed acidic (D99, E138, D176 and D181) and basic (K25 and K112) residues. This may be the reason that the trimeric assembly

remains stable across a wide range of pH starting from acidic to alkaline medium.

**Table 4.2. Conformational thermodynamic changes (in kJ/mol) of interfacial residues averaged over three subunits.**

Residue	$\Delta G_i^{conf}$	$T\Delta S_i^{conf}$	Residue	$\Delta G_i^{conf}$	$T\Delta S_i^{conf}$
D24	-1.2	-4.5	Q146	0	-2.3
W31	-0.06	-0.7	Q147	-0.2	-0.7
R35	-0.09	-1.1	R149	0	-2.3
M39	-1.4	-0.7	K150	-0.75	-0.9
R41	-1.05	-0.4	T153	-0.2	-2
F42	-0.9	-4.2	T154	-0.3	-0.2
F45	-1.6	-3	E161	-0.5	-5.2
N46	-1	-6.3	G162	-0.04	-0.2
D85	-0.6	-5.8	K163	-0.8	-8
I90	0.07	0.2	Y164	-1	-4.4
N91	-0.2	-2.25	Y165	-1.3	-8.7
S92	0	-0.3	Q166	-1	-7.5
S93	-0.08	-0.2	Q175	0.06	0
N114	0.09	-0.4	S184	0.8	0.6
M117	-0.7	-4.1	V185	0.7	1.2
K118	-0.9	-7.4	E187	0.3	0.04
K122	-1.25	-9.1	S188	-0.4	-1.4

## 4.4 Conclusions

We perform MD simulation of trimer and monomer conformations of STY3178 and estimate the conformational free energy and entropy changes of the former with respect to the latter. Computational results indicate aggregation induced stability and ordering of STY3178 trimer through salt bridge interactions. This accounts for stability of aggregation state in presence of chaotropes and under different pH conditions. Our method can be immediately applied to understand stability of bio-molecular aggregation comparing histograms of dihedral angles.

## ***CHAPTER 5***

# Static and dynamic response of microscopic conformational variables of ligand binding residues in different proteins

### **5.1 Introduction**

The rate of chemical processes in a medium generally shows inverse dependence on medium viscosity,  $\eta^{-1}$  known as the Stoke's behaviour.<sup>44, 45</sup> However, marked deviations from the Stoke's law have been reported in many experiments with bio-molecular systems.<sup>145-161</sup> For instance, rate of folding of many peptides show  $\eta^{-\alpha}$  dependence with  $\alpha < 1$ . The rates of catalysis<sup>146, 154, 159</sup> of enzymes and ligand binding rates<sup>151</sup> report both  $\alpha < 1$  and  $\alpha > 1$ . For peptide folding cases where deviations from Stoke's behaviour have been observed, detailed simulation studies reveal characteristic involvement of internal degrees of freedom, like dihedral angles<sup>22-24, 34</sup>. Dihedral angles undergo transitions between isomeric states as conformational state of the protein changes.<sup>6, 23</sup> These isomeric states are separated by energy barriers, also termed as internal friction<sup>46, 47</sup>. Since protein function is modulated by conformational changes<sup>162</sup>, it is pertinent to ask: Can the diverse experimental reports on deviations of rates from Stoke's behaviour be unified in terms of dihedral transitions? Here we address this question which is expected to throw light to microscopic mechanisms of kinetics of variety of bio-molecular processes.

In this chapter<sup>48</sup>, we propose a mathematical model (see Methods and Appendix I) to understand the role of energy barrier in governing dihedral angle transitions between isomeric states. We find that dihedral angle relaxation is slow whenever the isomeric states are separated by low energy barrier ( $f_B$ ) with a characteristic time-scale,  $\tau \propto f_B^{-1}$ . We further show that participation of several dihedral angles with a distribution of barrier heights account for observed deviation from Stoke's law. We perform MD simulations of different calcium ( $\text{Ca}^{2+}$ ) binding proteins to extract the barrier height distribution of the dihedral angle isomerisation states. We find that the distribution shows exponential decay in energy. We also compute dihedral auto-correlation function from the MD simulated trajectories, which yields the  $\tau \propto f_B^{-1}$  dependence consistent with our modelling results.

## 5.2 Methods

We perform explicit solvent MD simulation using standard protocols (see Appendix I and II, chapter 2) for (i) calmodulin (CaM)<sup>22, 23</sup>, (ii) troponin C (TnC)<sup>40</sup>, (iii)  $\alpha$ -Lactalbumin (aLA)<sup>34</sup> and (iv) calbindin (CALB).

### Radius of gyration

The radius of gyration,  $R_g$  is calculated as the average distance of the C- $\alpha$  atoms from their centre of mass ( $\vec{R}_{CM}$ ),  $\vec{R}_{CM} = \sum_i m_i \vec{r}_i / \sum_i m_i$ , where  $m_i$  and  $\vec{r}_i$  are the mass and position vectors of the  $i^{\text{th}}$  C- $\alpha$  atom. The squared  $R_g$  is then calculated as:

$$R_g^2 = \sum_i m_i \left( \vec{r}_i - \vec{R}_{CM} \right)^2 / \sum_i m_i, \quad (1)$$

The  $R_g$  corresponding to the calcium binding loops is computed for every structure in the

ensemble and a histogram is generated for each of the loops in the apo and the holo states.

### Dihedral autocorrelation function

We compute the histograms of the dihedral angles,  $\phi$ ,  $\psi$  and  $\chi_1$  of the loop residues, from the ensemble. The dihedral auto-correlation function (DACF) in time,

$$C_{\theta}(t) = \frac{\langle (\cos \theta(\tau + t) - \langle \cos \theta \rangle)(\cos \theta(\tau) - \langle \cos \theta \rangle) \rangle}{\langle (\cos \theta(\tau) - \langle \cos \theta \rangle)^2 \rangle} \quad (2)$$

of fluctuations of any dihedral  $\theta$  is calculated for a set of initial conditions  $\tau$  with  $t = 0.01$  ns. The  $\cos \theta(\tau)$  and  $\cos \theta(\tau + t)$  terms are the cosine values of the dihedral  $\theta$  at some initial time  $\tau$  and after some time interval  $t$ , while  $\langle \cos \theta \rangle$  is the time-averaged cosine value of the dihedral angle<sup>28, 163</sup>. The outer angular brackets indicate averaging over the initial conditions chosen from the equilibrated trajectory. The DACFs from MD trajectory are computed as follows:

For any given time-difference  $t = |t_2 - t_1|$ , we compute the product from the series of conformations starting with the  $i^{\text{th}}$  initial condition  $A_i(t_2, t_1) = (\cos \theta(t_2) - \langle \cos \theta \rangle)(\cos \theta(t_1) - \langle \cos \theta \rangle)$ . The average  $\langle A_i(t_2, t_1) \rangle$  is computed over all possible combinations of  $t_1$  and  $t_2$  in the given time series giving rise to same  $t$ . We then choose several initial conditions from various parts of the trajectory to calculate

$$C_{\theta}(t_2, t_1) = \frac{1}{n} \sum_i \langle A_i(t_2, t_1) \rangle \quad (3)$$

where  $n$  is the number of independent initial conditions. The stationarity of the system in the equilibrium allows considering  $t_1 = 0$  and  $t_2 = t$ , leading to  $C_{\theta}(t)$ . The computation is done till  $C_{\theta}(t)$  approaches zero for sufficiently large  $t$ .

## Mathematical modelling

We formulate a mathematical model to relate the equilibrium bimodal dihedral distributions to the corresponding DACFs. The dynamics of dihedral angles can be modelled through the Langevin equation of a damped oscillator<sup>28, 163</sup>.

$$I\ddot{\theta} + \frac{f_B}{\omega_0}\dot{\theta} + k\theta(t) = f(t) \quad (4)$$

where,  $\theta(t)$  is the torsional displacement,  $I$  is the moment of inertia about the torsional axis,  $f_B$  is the energy barrier between transition from one peak to other of a bimodal distribution,  $\omega_0 = \sqrt{\frac{k}{I}}$  is the vibrational frequency and  $k$  is the associated force constant. The  $f(t)$  is random torque due to environmental fluctuations and considered as Gaussian noise. The detailed algebra is shown in Appendix I.

## 5.3 Results

We work out a mathematical model to understand the role of energy barrier in governing dihedral angle transitions from one isomeric state to another. We also perform MD simulations of different calcium ( $\text{Ca}^{2+}$ ) binding proteins in apo and holo state to support the mathematical model.

### A. Mathematical Model

We characterize the dynamics of dihedral angles in the presence of an energy barrier. We consider a free energy profile of a dihedral  $\theta$  having two minima, one primary and the other secondary, separated by a barrier,  $f_B$  (see Fig 5.1 for schematic). Let us consider the situation that the dihedral angle undergoes isomerisation by transition from the primary minimum to the secondary one. The barrier hinders the motion which can be considered as frictional force, given by  $\frac{f_B}{\omega_0}$  where  $\omega_0^{-1}$  is a typical time-scale associated with the attempts to cross the barrier in presence of thermal noise  $f(t)$ . The noise,  $f(t)$  is taken to have Gaussian statistics.

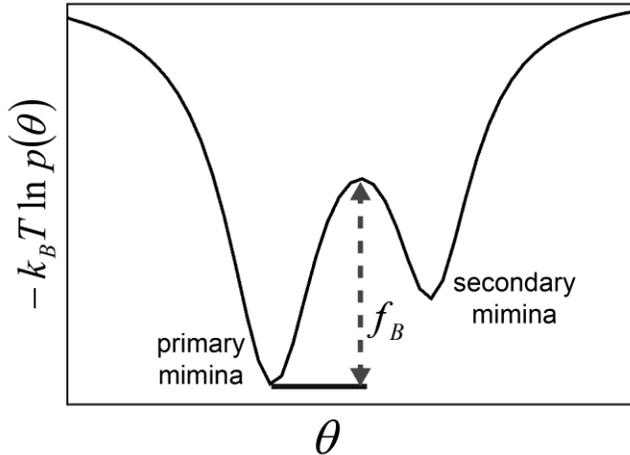


Fig 5.1. Schematic representation of free energy profile of a dihedral,  $\theta$  having a primary and a secondary minima separated by an energy barrier,  $f_B$ .

The dynamical characteristic is given in terms of auto-correlation function<sup>28</sup>,  $\langle \theta(t)\theta(0) \rangle$  in time. The function reflects how the values of a dihedral angle at time ‘ $t$ ’ is correlated to its

initial value. Physically, it describes the decay behaviour of dihedral angle dynamics, namely, for a given dihedral angle at its primary minima, it reflects the time-scale of transition to its secondary minima across an energy barrier. We calculate the auto-correlation function (see Appendix I) from the Langevin equation,  $\langle \theta(t)\theta(0) \rangle = A \exp(-t/\tau) \sin(\omega_n t)$ . Here  $\tau \left( = \frac{2k}{f_B \omega_0} \right)$  denotes the overall decay time of dihedral fluctuations and  $\omega_n = \sqrt{1 - \left( \frac{f_B}{2k} \right)^2} \omega_0$ ,  $k$  being the force constant given by the inverse width at the primary peak. Here,  $\omega_n$  is a finite number if  $f_B < 2k$  and becomes imaginary for large barrier ( $f_B > 2k$ ). Thus the sinusoidal oscillation with frequency,  $\omega_n$  represents the barrier crossing events between two isomeric states.

Now we consider a binding process involving a pair of bio-molecules with internal degrees of freedom given by their respective dihedral angles. Let the binding event be accompanied by isomerisation of dihedral angles at the region of binding. These two bio-molecules approach each other in presence of a solvent, governed by diffusion controlled rate,  $\Gamma$  inversely proportional to solvent viscosity ( $\sim \eta^{-1}$ ). We have shown that the dihedral relaxation time-scale across barrier separated isomerisation states is given by  $\tau$ . Hence, the rate of dihedral angle isomerisation will be  $\tau^{-1}$ . The total time for overall process is given by,  $t_{tot} = \left( \Gamma + \tau^{-1} \right)^{-1}$ .

Suppose a large number of dihedral angles at the binding region undergo isomerisation in the process via various energy barriers, which contributes to the overall time scale of binding process. Let the energy barriers have a probability distribution,  $p(f_B)$ . The overall time scale is then given by integrating over the distribution of barrier heights,  $t_{tot} = \int df_B p(f_B) \left( \Gamma + \frac{f_B \omega_0}{2k} \right)^{-1}$ .

The integration yields a dependence of overall time-scale,  $t_{tot} \sim \eta^\alpha$  (see Appendix I), if the

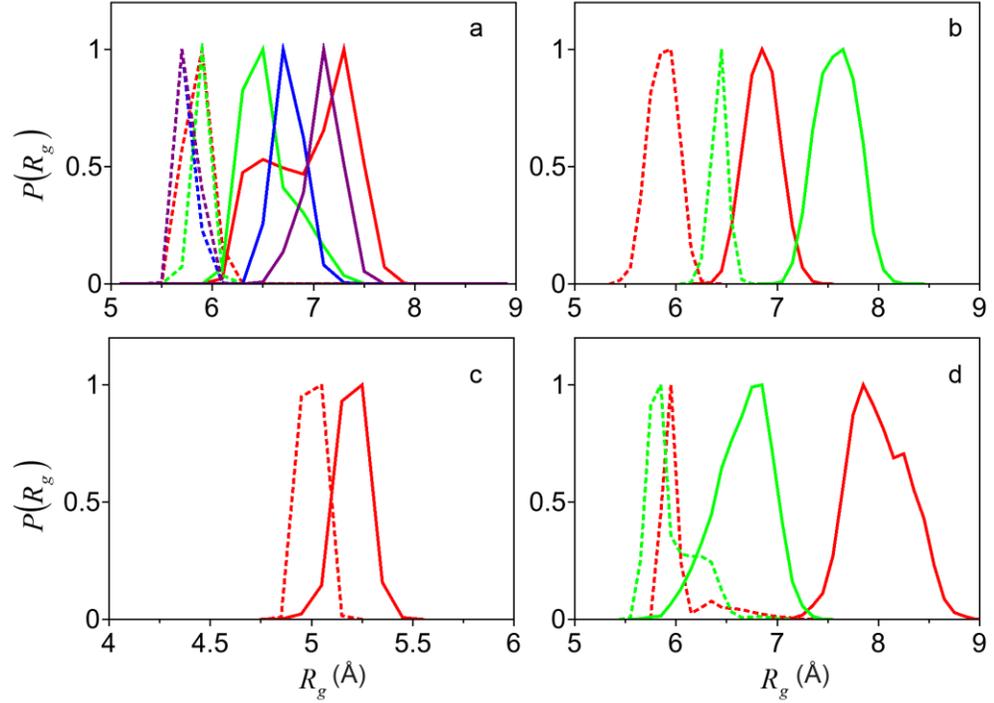
barrier height distribution shows a power law behaviour which falls off algebraically,  $p(f_B) \propto f_B^{-\alpha}$ . Another interesting situation arises for exponential distribution,  $p(f_B) \propto \exp\left(\frac{-f_B}{\lambda}\right)$  where  $\lambda$  is a decay constant in energy. This yields a time-scale  $t_{tot} \sim \eta^2$  in the high viscosity regime (see Appendix I). Physically, the overall time scale is governed by the convolution over all the individual dihedral angle relaxation in combination with solvent viscosity mediated diffusion.

## **B. Distribution of barrier heights from MD simulations**

Our mathematical analysis shows that the distribution of barrier heights separating the isomeric states of dihedrals decide the overall dependence of binding rate on viscosity. We examine the barrier heights between isomeric states of dihedral angles of binding residues in an important class of biomolecular process called molecular recognition of ligand binding to proteins<sup>8, 164-167</sup>. The different models for molecular recognition of protein-ligand systems include Fischer's 'lock and key'<sup>168</sup> (LK), Koshland's 'induced fit'<sup>169</sup> (IF) and conformational selection<sup>170-172</sup> (CS). The LK model states that the ligand-free conformation of the protein is capable of binding to the ligand without undergoing any conformational change. According to the IF mechanism, ligand binding induces conformational changes in the protein leading to stability of the complex. However, experimental evidences<sup>173</sup> suggest that proteins intrinsically span a wide range of conformations in the ligand free state and the ligand chooses a conformational state compatible to its binding through the CS mechanism. Recent developments in this field indicate a hybrid recognition mechanism where CS is followed by an adjustment through IF (CS-IF)<sup>174</sup>.

*Radius of gyration distribution:* We consider a simple case of ligand binding, specially calcium ( $\text{Ca}^{2+}$ ) ion binding to different metallo-proteins, like calmodulin (CaM)<sup>67</sup>, calbindin

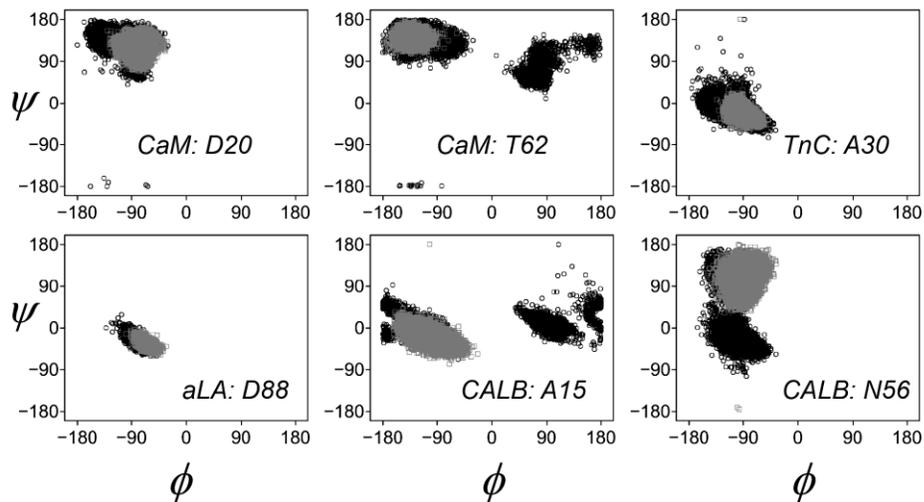
(CALB)<sup>175, 176</sup>, skeletal muscle Troponin C (TnC)<sup>36</sup> and alpha-Lactalbumin (aLA)<sup>29</sup>. The nature of Ca<sup>2+</sup> ion binding residues in all these proteins is similar, either acidic or polar. All atom MD simulations are performed for both Ca<sup>2+</sup> ion free (apo) and bound (holo) proteins. We specify the



**Fig 5.2.** The equilibrium distributions of radius of gyration  $P(R_g^i)$  of Ca<sup>2+</sup> binding loops in apo (solid line) and holo (dotted line) states of proteins "i", namely CaM, TnC, aLA and CALB. (a)  $P(R_g^{CaM-L1})$  in red,  $P(R_g^{CaM-L2})$  in green,  $P(R_g^{CaM-L3})$  in blue and  $P(R_g^{CaM-L4})$  in brown. (b)  $P(R_g^{TnC-L1})$  in red and  $P(R_g^{TnC-L2})$  in green. (c)  $P(R_g^{aLA-L1})$  in red (d).  $P(R_g^{CALB-L1})$  in red and  $P(R_g^{CALB-L2})$  in green.

conformational states through equilibrium distribution of radius of gyration ( $P(R_g)$ ) of the Ca<sup>2+</sup> ion binding loops in both apo and holo states. We find that the mean  $\langle R_g \rangle$  in holo state is always smaller compared to that in apo state due to strong electrostatic interaction between the Ca<sup>2+</sup> ion and acidic residues of these loops.  $P(R_g^{CaM-L1})$  (Fig 5.2 a) shows a bimodal apo state distribution with a major peak at 7.3 Å and a minor peak at 6.4 Å. The holo state distribution, in contrast, shows a single sharp peak at 5.8 Å in the vicinity of the less populated apo peak, indicating CS-

IF. On other hand, the distributions,  $P(R_g^{CaM-L2})$ ,  $P(R_g^{CaM-L3})$  and  $P(R_g^{CaM-L4})$  show a new conformational state upon  $Ca^{2+}$  ion binding, indicative of the IF mechanism. The distribution of radius of gyration of both loop 1 and loop 2 in TnC (Fig 5.2 b) indicate conformational changes governed by IF mechanism upon  $Ca^{2+}$  ion binding. In contrast,  $P(R_g^{aLA-L1})$  exhibits sharp single peak in both apo and holo-aLA with minor shift of 0.2 Å, indicating LK mode of recognition (Fig 5.2 c). The  $P(R_g^{CALB-L1})$  (Fig 5.2 d) in apo state shows maxima around 7.8 Å with a shoulder at 8.2 Å indicating presence of two populations. The holo state population shows a new peak at 5.8 Å with a tail extending towards the apo population. The apo state distribution of the other CALB loop,  $P(R_g^{CALB-L2})$  shows a single peak around 6.8 Å. In the holo state we find a primary maxima at about 5.6 Å and another secondary peak at 6.4 Å. The holo state distributions  $P(R_g^{CALB-L1})$  and  $P(R_g^{CALB-L2})$  indicate new conformation through IF mechanism (Fig 5.2 d). Although the conformational states of the loops change upon metal ion binding, the secondary structural elements remain largely unchanged (Fig 5.3 for representative Ramachandran plots).



**Fig 5.3. Ramachandran plot of different calcium binding residues in apo (black) and holo (grey) state of proteins: CaM, TnC, aLA and CALB.**

*Dihedral angle distribution:* However, we observe changes in dihedral angles of the binding residues, while the conformational state changes from apo to holo. The peaks of equilibrium distribution  $p_R^i(\theta)$  of dihedral  $\theta$  for residue  $R$  in protein “ $i$ ” represent different isomeric states (Fig 5.4). There are four patterns of isomeric transitions between apo and holo states. For instance,  $p_{A103}^{CaM}(\phi)$  (Fig 5.4 a) exhibits overlapping unimodal distributions in both apo- and holo state. We denote such overlapping single peaked distributions as SP1.  $p_{D65}^{TnC}(\psi)$  (Fig 5.4 b) is also unimodal in both apo and holo states but having large amount of peak shift which we denote as SP2. On the other hand,  $p_{D20}^{CaM}(\phi)$  (Fig 5.4 c) is bimodal with peaks at  $-117^\circ$  and  $-62^\circ$  in apo state, while in holo-CaM the population shifts towards the minor peak at  $-62^\circ$ . We denote the cases where the holo state population grows in vicinity of an apo population corresponding to a multimodal apo distribution as MP1. The side-chain dihedrals however show different behaviour, where both distributions in ligand free and bound state are flat.

The free energy profiles,  $-k_B T \ln(p(\theta))$  corresponding to the equilibrium distributions of the dihedrals in apo state are also shown in Fig 5.4 d-f. For instance, the free energy profile of the unimodal distribution of  $p_{A103}^{CaM}(\phi)$  (Fig 5.4 d) shows a single minimum. Similar free energy profile is also observed for  $p_{D65}^{TnC}(\psi)$  (Fig 5.4 e). Let us now consider the MP1 cases. For instance,  $p_{D20}^{CaM}(\phi)$  (see Fig 5.4 f) in apo state has a primary peak ( $\theta_{\max}$ ) at  $-117^\circ$  and a secondary peak at  $-62^\circ$  separated by a minima ( $\theta_{\min}$ ) at  $-102^\circ$ . The secondary apo peak shows population growth in holo-CaM through population shift from  $\theta_{\max}$  across the minima,  $\theta_{\min}$ . So we define

the barrier separating two apo peaks by  $f_B = -k_B T \ln\left(\frac{p(\theta_{\max})}{p(\theta_{\min})}\right)$ .

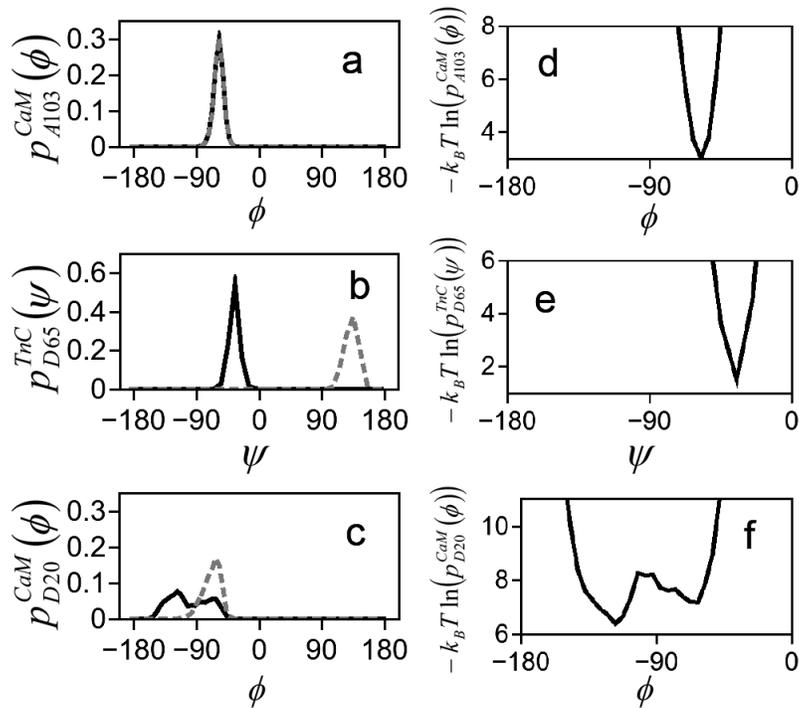


Fig 5.4. The equilibrium distributions,  $p_R^i(\theta)$  of dihedral angles “ $\theta$ ” of residue “ $R$ ” in protein “ $i$ ” for apo (solid line) and holo (dashed line). (a)  $p_{A103}^{CaM}(\phi)$ , (b)  $p_{D65}^{TnC}(\psi)$  and (c)  $p_{D20}^{CaM}(\phi)$ . The corresponding free energy profiles of (d)  $p_{A103}^{CaM}(\phi)$ , (e)  $p_{D65}^{TnC}(\psi)$  and (f)  $p_{D20}^{CaM}(\phi)$ .

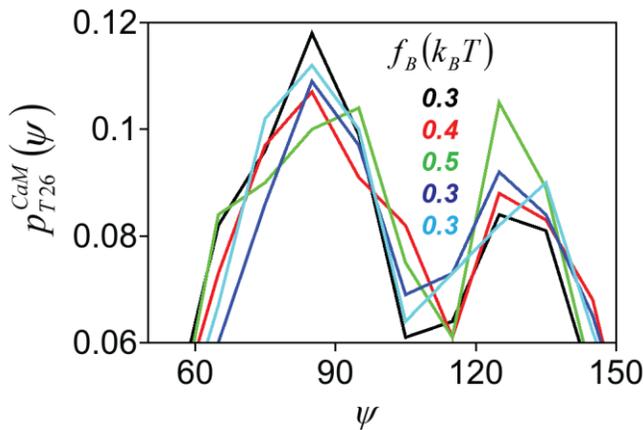


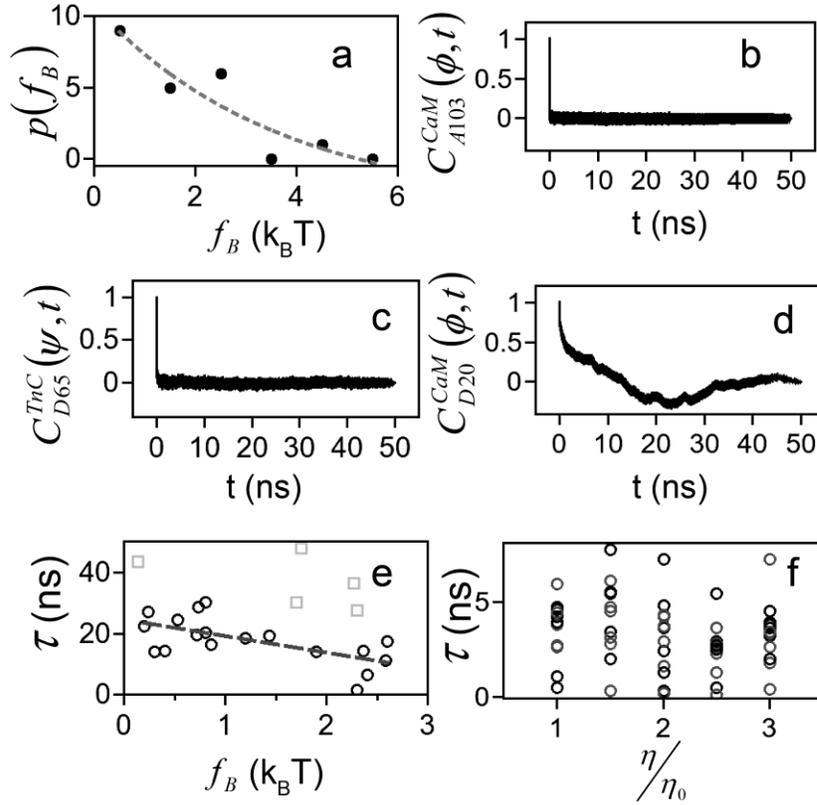
Fig 5.5. The equilibrium distributions,  $p_{T26}^{CaM}(\psi)$  generated from five different ensembles corresponding to different starting configurations of apo-CaM loop 1. The variation in barrier height,  $f_B$  for different ensembles reflects the stochasticity in distribution.

The experimental determination of rate of any bio-molecular process represents an ensemble averaged measurement. In order to mimic this situation we perform MD simulation of apo-CaM loop1 with different starting structures and generate separate ensembles representing

different molecules of the same species. We calculate the barrier heights of backbone dihedral angles from each of these ensembles. For instance, the barrier heights calculated from the dihedral angle distributions,  $p_{T26}^{CaM}(\psi)$  for different ensembles show statistical variation (Fig 5.5). These statistical variations account for stochasticity in the distribution of barrier heights for a bio-molecular system. We calculate the  $f_B$  between dihedral isomeric states for different MP1 cases from all the loops we have simulated. The distribution of barrier height,  $p(f_B)$  for these cases, shown in Fig 5.6 a, exhibits an exponential behaviour  $\exp\left(\frac{-f_B}{\lambda}\right)$  with  $\lambda = 3.6 k_B T$ .

*Dihedral autocorrelation functions:* We compute the dynamic response of the dihedral angles in apo state using the MD simulation trajectories. We compute the dihedral autocorrelation function,  $C_R^i(\theta, t)$  for dihedral angle fluctuations about its equilibrium value as the product of the value of  $\delta \cos \theta(t_0)$  at some initial time  $t_0$  and the value after some time interval  $t$ ,  $\delta \cos \theta(t + t_0)$ , averaged over a set of initial conditions  $t_0$ .<sup>28</sup>

The DACFs for SP1 cases, like  $C_{A103}^{CaM}(\phi, t)$  (Fig 5.6 b) and SP2 cases, like  $C_{D65}^{TnC}(\psi, t)$  (Fig 5.6 c) show fast decay in apo state. However, for MP1 and MP2 cases, the apo state autocorrelation amplitude is characterized by initial slow decay along with significant anti-correlation at long time. This is shown for ligand free DACF,  $C_{D20}^{CaM}(\phi, t)$  (Fig 5.6 d) as a representative for MP1 case. Such apo state DACFs bear signature of damped oscillations characterized by an oscillation frequency,  $\omega_n$  having an overall decay time constant,  $\tau$ .



**Fig 5.6.** (a) The distribution of barrier height,  $p(f_B)$  separating dihedral isomerisation states for MP1 cases. The DACF response of the dihedral angles corresponding to apo (black) states illustrating different time-scales for (b)  $C_{A103}^{CaM}(\phi, t)$ , (c)  $C_{D65}^{TnC}(\psi, t)$  and (d)  $C_{D20}^{CaM}(\phi, t)$ . (e) Correlation plot of barrier height,  $f_B$  with overall decay time constant  $\tau$ . (f) Correlation plot of overall decay time constant  $\tau$  of backbone dihedral angles with viscosity,  $\frac{\eta}{\eta_0}$ .

We fit the DACFs from simulation with the functional form,  $\exp\left(\frac{-t}{\tau}\right)\sin(\omega_n t)$  considering  $\omega_n$  and  $\tau$  as parameters for best fit. The fitted DACFs are shown in Fig 5.7. Similar slow decaying DACFs are observed for multimodal holo state distributions as in the apo cases.

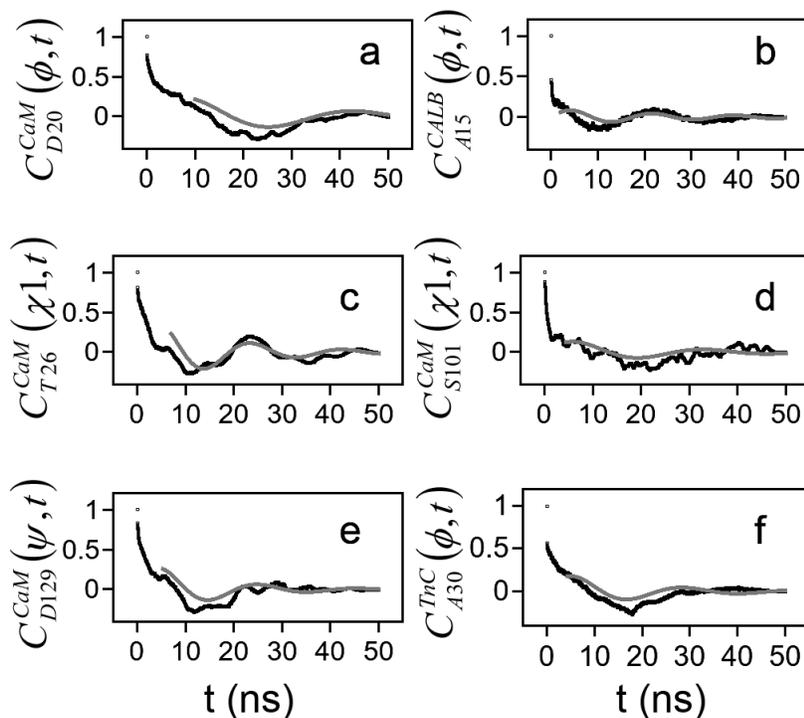
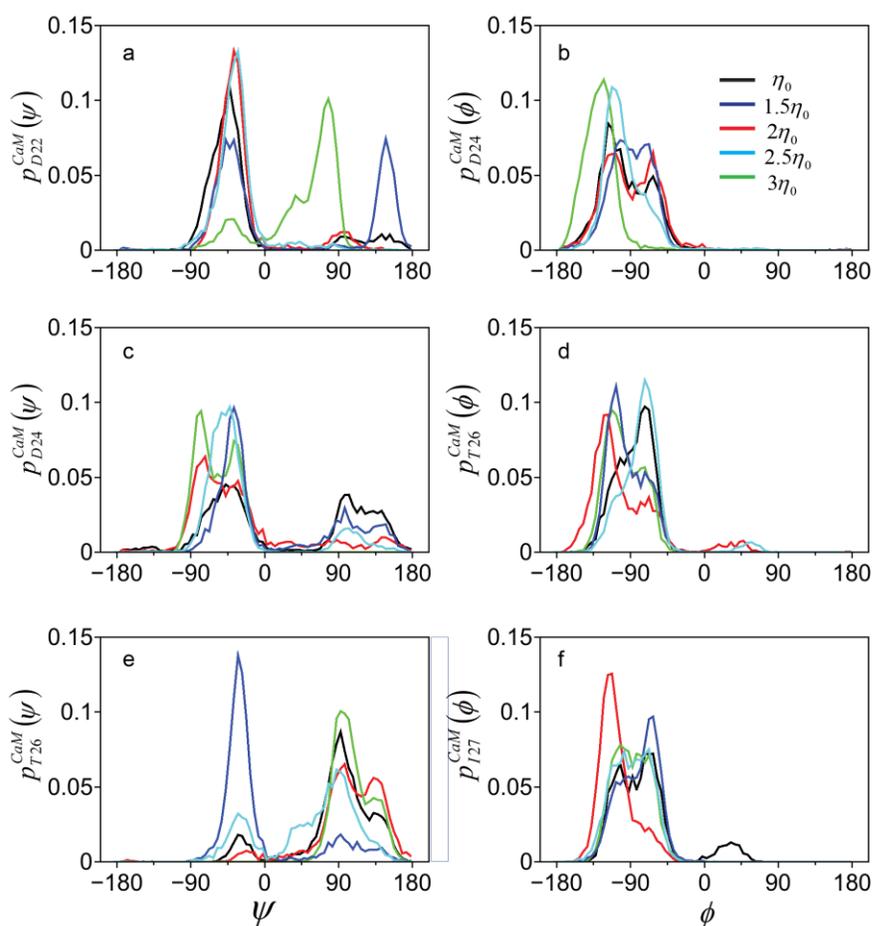


Fig 5.7. The apo state DACFs (black) fitted (grey) with the expression,  $\exp\left(\frac{-t}{\tau}\right)\sin(\omega_n t)$ .

We show in Fig 5.6 e that the overall decay time constant,  $\tau$  decreases linearly with increasing  $f_B$ . This in agreement to the results of mathematical analysis. The dihedral relaxation between isomeric states occurs at slower time-scales when the barrier between them is less. The underlying physical picture is that whenever two isomeric states are separated by a low energy barrier, frequent barrier crossing events occur as a result of which the dihedral fluctuations exhibit slow decay profile.

*Dependence on solvent viscosity:* We examine the dependence of  $\tau$  of dihedral angle fluctuations of the  $\text{Ca}^{2+}$  binding residues on solvent viscosity. We choose MP1 predominant apo-CaM loop 1 for various solvent viscosities through rescaling the mass of explicit water molecules.<sup>177, 178</sup> We consider individual backbone dihedral angles of apo-CaM loop 1 to estimate the respective decay time constants  $\tau$  at different viscosities. The correlation plot of  $\tau$

versus  $\eta/\eta_0$  (Fig 5.6 f) shows little correlation ( $R^2 \approx 0.046$ ). The decay time constant  $\tau$ , governed by the barrier between isomeric states of dihedral angles, is thus observed to be independent of solvent viscosity. This is in agreement to earlier studies<sup>177, 178</sup>, where the time constant of auto-correlation function averaged over backbone dihedral angles representing internal friction time-scale is reported to be independent of solvent viscosity. We show representative dihedral angle distributions of residues from apo-CaM loops under different viscosity conditions in Fig 5.8.



**Fig 5.8.** The equilibrium distributions,  $p_R^{CaM}(\theta)$  of dihedral angles “ $\theta$ ” of residue “ $R$ ” in apo-CaM loop 1 at different viscosities.

*Connection to experimental results:* Table 5.1 summarizes different bio-molecular processes with rates deviating from Stoke's law. We consider a diversity of bio-molecular rate processes, which include ligand binding and dissociation, protein folding-unfolding, electron transfer, proton exchange and enzyme catalysis. The binding and dissociation rates of carbon mono-oxide and oxygen to heme proteins like haemoglobin, hemerythrin and myoglobin exhibit fractional power law in viscosity.<sup>145, 151, 152</sup> The folding rates<sup>150, 156-158</sup> of various proteins and peptides and enzyme catalysed reaction rates<sup>146, 148, 153, 154, 159</sup> also show fractional power dependence,  $\alpha \leq 1$  on solvent viscosity. Such dependence on solvent viscosity has been observed for intra-electron transfer rates in sulphite oxidase<sup>147</sup> and proton exchange rates of Cod PA III<sup>160</sup> as well. Our mathematical analysis accounts for such behaviour when the distribution of barrier heights exhibits power law dependence as  $p(f_B) \propto f_B^{-\alpha}$ , where  $\alpha \leq 1$ . Table 5.1 further shows exponent  $\alpha > 1$  in many cases. For instance, unfolding<sup>161</sup> and electron transfer rates<sup>155</sup> of Cytochrome C, amide proton exchange rates<sup>149</sup> of lysozyme and kinetics of human S4 enzyme<sup>161</sup> show such behaviour at very high viscosities. In these cases the exponent is close to  $\sim 2$ . Our mathematical modelling shows that exponents  $\sim 2$  is possible for high solvent viscosity with an exponential distribution of barrier heights. Our analysis suggests that Stoke's behaviour arises when there is no contribution from internal degrees of freedom or in absence of internal degrees of freedom. However, even in the presence of internal degrees of freedom, Stoke's law can still hold, if  $p(f_B) \propto f_B^{-1}$  with  $\alpha = 1$ .

**Table 5.1. The power law exponent in viscosity,  $\eta^{-\alpha}$  dependence of rates of bio-molecular processes,**

<b>Biomolecular processes</b>	$\alpha$
<i>Enzyme catalysis</i>	
Carboxypeptidase A	0.4-0.6
Subtilisin BPN	0.65
Glutathionine transferase	0.8
Dihydrofolate reductase	0.3
Human S4	2.1
<i>Electron transfer/Proton exchange</i>	
Cytochrome C	2.4
Sulfite oxidase	0.7
Cod PA III	0.63
Lysozyme	1.4
<i>Ligand binding/dissociation</i>	
Myoglobin (O2 diss)	0.4
Myoglobin (CO binding)	0.45
Hemethyrin (O2 diss)	0.5
Haemoglobin (O2/CO exchange)	0.26
<i>Protein folding/unfolding</i>	
Trp Cage	0.84
alpha-helix	0.53
beta-hairpin	0.93
Ig-G (protein L)	0.58
Cytochrome C (unfolding)	2.4
A3D (73 residue apha-helical protein) folding	0.19

*Relation with molecular recognition:* The instances of different isomerisation patterns, SP1, SP2, MP1 and MP2 for different proteins are summarized in Table 5.2. We now relate these isomerisation patterns to the molecular recognition processes: CS, IF and LK as determined from the  $P(R_g)$  of the  $\text{Ca}^{2+}$  binding loops. The molecular recognition process in CaM loop 1 is a hybrid mechanism of CS and IF. We find that MP1 (~ 49 %) is predominant in CaM loop 1. In

CaM loop 2 and loop 3 SP1 (~37 %) cases are predominant, while CaM loop 4 shows preference for SP2 (~53 %) patterns. The two loops of TnC indicating an IF mechanism show a predominance of SP2 isomerisation pattern (>40 %). Similarly, the SP2 pattern is also predominant (~40 %) for the Ca<sup>2+</sup> binding loop of aLA, which is governed by LK. On other hand molecular recognition is governed by IF mechanism in CALB loops, where MP1 (~40 %) pattern is prevalent in loop 1 and SP1 (~35 %) in loop 2.

Since MP1 cases are more abundant in the Ca<sup>2+</sup> ion free CaM loops 1 and 2 than the other loops (see Table 5.2), the relaxation of the corresponding dihedral angles are slower compared to that in loops 3 and 4. Our data on time-correlation thus suggests that Ca<sup>2+</sup> ion binding occurs first at the C-terminal loops and then the N-terminal loops get saturated, which is consistent with experimental results.<sup>179, 180</sup> On a similar note, one would expect slow Ca<sup>2+</sup> ion binding to the CALB loops compared to the kinetics of Ca<sup>2+</sup> ion binding to TnC or aLA.

**Table 5.2. The percentage (%) of SP1, SP2, MP1 and MP2 observed for dihedral angles in different loops of CaM, TnC, CALB and aLA.**

	SP1 %	SP2 %	MP1 %	MP2 %
<i>CaM L1</i>	20	22.8	48.6	8.6
<i>CaM L2</i>	36.7	20	33.3	10
<i>CaM L3</i>	37.5	21.9	21.9	18.7
<i>CaM L4</i>	14.7	53	11.7	20.6
<i>TnC L1</i>	28.1	43.75	21.9	6.25
<i>TnC L2</i>	26.5	47.0	14.7	11.8
<i>CALB L1</i>	30.5	13.9	41.7	13.9
<i>CALB L2</i>	35.3	20.6	20.6	23.5
<i>aLA</i>	36.7	40	13.3	10

## 5.4 Conclusion

We provide a theory to account for the anomalous dependence of different bio-molecular rate processes on solvent viscosity. Our mathematical modelling explains the deviation from Stoke's law and provides a way to understand power law dependence with  $\alpha \neq 1$  for rates of different bio-molecular processes on solvent viscosity based on the dihedral relaxation time-scales. The decay time constant,  $\tau$  is observed to decrease linearly with increasing barrier height. We show that the distribution of barrier heights accounts for the observed deviation from Stoke's law.

We further study a particular class of bio-molecular process, namely  $\text{Ca}^{2+}$  ion binding to different proteins to explore the behaviour of dihedral transitions. To this end we perform MD simulation of  $\text{Ca}^{2+}$  binding proteins in apo and holo state. We compute the dihedral angle distributions and generate a probability distribution of barrier heights separating the dihedral angle isomerisation states of  $\text{Ca}^{2+}$  binding residues. The distribution of barrier heights shows an exponential decay with increasing energy. We also estimate the time-scales of dihedral angle relaxation and find that  $\tau \propto f_B^{-1}$  from the simulation data consistent with mathematical analysis. This time-scale is further observed to be independent of solvent viscosity. The most important aspect of our work is to provide a unified understanding of rate of various bio-molecular processes in terms of microscopic variables.

## Appendix I

Taking the Fourier transform of  $I\ddot{\theta} + \frac{f_B}{\omega_0}\dot{\theta} + k\theta(t) = f(t)$  yields  $\theta(\omega)$  and the DACF in

the frequency domain

$$\langle \theta(\omega)\theta(-\omega) \rangle = \frac{\langle f(\omega)f(-\omega) \rangle}{(k - I\omega^2)^2 + \left(\frac{\omega f_B}{\omega_0}\right)^2} \quad (1)$$

The inverse Fourier transform of equation (2) yields

$$\langle \theta(t)\theta(0) \rangle = \int_{-\infty}^{+\infty} \frac{\exp(-i\omega t)}{(k - I\omega^2)^2 + \left(\frac{\omega f_B}{\omega_0}\right)^2} d\omega \quad (2)$$

Solving the integral we find that

$$\langle \theta(t)\theta(0) \rangle = A \sin\left(\sqrt{1 - \left(\frac{f_B}{2I\omega_0}\right)^2} \omega_0 t\right) \exp\left(-\frac{f_B}{2I\omega_0} t\right) = A \sin(\omega_n t) \exp\left(\frac{-t}{\tau}\right) \quad (3)$$

We use equation (3) for fitting the DACFs.

The rate process of bio-molecular function is governed by both contribution from external friction arising out of solvent viscosity and internal friction of dihedral transitions. Let the rate associated with external friction be  $\Gamma$  and that with internal friction is  $\tau^{-1}$ . The overall time scale is then calculated by integrating over the distribution of barrier

heights,  $\int df_B p(f_B) (\Gamma + \tau^{-1})^{-1}$  We assume a power law dependence of barrier height distribution

given by  $p(f_B) \propto f_B^{-\alpha}$  and replace  $\tau^{-1} = \frac{f_B \omega_0}{2k}$  in the above equation. Taking the inverse Laplace

transform we get  $\int_0^{\infty} ds \exp(-s\Gamma) \int_0^{\infty} df_B \exp\left(\frac{-sf_B \omega_0}{2k}\right) f_B^{-\alpha}$ . Solving the inner integral within the

limits yields  $\int_0^\infty ds \exp(-s\Gamma) \left(\frac{s\omega_0}{2k}\right)^{\alpha-1}$ . This finally gives  $\left(\frac{\omega_0}{2k}\right)^{\alpha-1} \frac{1}{\Gamma^\alpha} \int \exp(-z) z^{\alpha-1} dz$ , where

$z = s\Gamma$ . Our mathematical modelling thus indicate  $\Gamma^\alpha$  dependence of rate processes if  $p(f_B) \propto f_B^{-\alpha}$ .

Now if  $p(f_B) = \exp\left(-\frac{f_B}{\lambda}\right)$ , the inner integral  $\int_0^\infty df_B \exp\left[-f_B \left(\frac{s\omega_0}{2k} + \frac{1}{\lambda}\right)\right]$  yields

$\left(\frac{s\omega_0}{2k} + \frac{1}{\lambda}\right)^{-1}$ . Solving the full integral, we get  $\frac{\omega_0}{2k\Gamma^2} + \frac{1}{\Gamma\lambda}$ . At large  $\Gamma$ , the second term

predominates, while at low  $\Gamma$  ( $0 < \Gamma < 1$ ) the first term is predominant.

## **CHAPTER 6**

# Quantum chemical studies on stability and function of calcium ion bound calmodulin loops

### **6.1 Introduction**

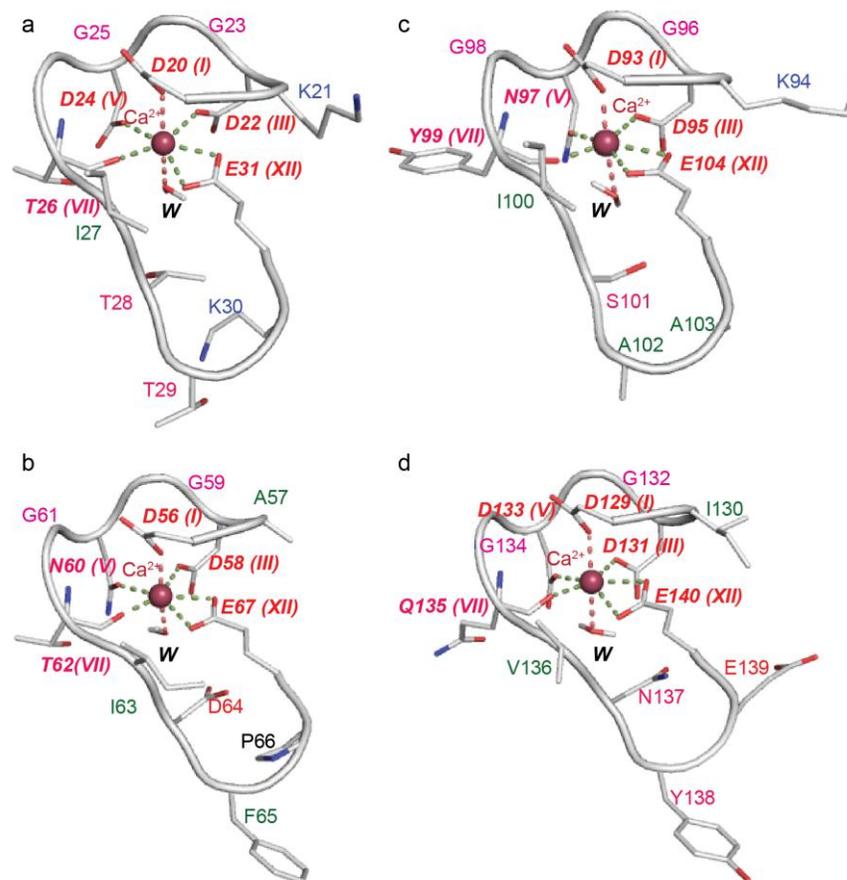
The coordination of metal ion to the active sites of a protein form metal-protein complexes, known as metalloproteins.<sup>17</sup> The metalloproteins perform a plethora of functions ranging from enzyme catalysis, signal transduction to acting as carriers. For instance, calcium binding proteins like calmodulin<sup>181</sup> is responsible for signal transduction; iron binding protein, haemoglobin<sup>182</sup> acts as oxygen carrier, while carbonic anhydrase<sup>183</sup> and endonuclease<sup>184</sup> act as metalloenzymes. On other hand, metalloproteins are often designed with desired properties and find wide range of applications in material science and electronics. For instance, copper binding azurin and iron binding cytochrome c proteins possess electron transfer property owing to redox reactions at their active site. Such electron transfer activity has been exploited for designing electronic devices like molecular switch, bio-sensors and bio-transistors.<sup>10, 11</sup> Iron binding metalloprotein ferritin being semi-conducting in nature, its band gap is known to be tuned by substitution with other metal ions like copper, cobalt and manganese.<sup>12</sup>

The metal ion induced distribution of electronic charges within the protein leading to structural stability as well functionality, can be accounted for by quantum chemical (QC) calculations. The force-field based calculations are non-polarizable in nature and hence, cannot capture electron redistribution effects. The microscopic QC calculation of bio-macromolecular

binding is a formidable challenge due to large molecular sizes, comprising of thousands of atoms. The QC calculations are carried out by truncating the bio-molecule to include only the binding region, usually having about 10-100 atoms. Different electronegative ligand atoms like oxygen, nitrogen and sulphur from amino acid side chain functional groups or backbone generally coordinate to metal ions.<sup>66</sup> The binding region in general possesses a variety of interactions including charged, polar and hydrophobic residues. Since truncation of protein structure may affect some of these interactions, it is non-trivial to assess the effects of using such truncated systems on the results of the quantum calculations. For instance, the Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular Orbital (LUMO) are affected by boundary effects depending on the choice of terminal capping. Moreover, treatment of the solvent medium is necessary for accurate calculation of these levels.

The metal ion binding sites of different proteins has been studied by QC-DFT methods (see Appendix I). These studies include interaction of physiologically abundant divalent metal ions, like calcium and magnesium with different functional groups like carboxylate and carbonyl ligands.<sup>17, 64-66, 185</sup> Mechanisms of enzymatic reactions like phosphorylation, photosynthesis and electron transport involving redox active metal ions like manganese, copper and iron have also been studied using DFT methods.<sup>7, 56, 57, 186-188</sup> These DFT calculations on bio-molecular fragments use a hybrid electron exchange-correlation functional, B3LYP (Becke, three parameter, Lee-Yang-Parr)<sup>189</sup> along with 6-31G(d,p)<sup>190</sup> basis set. The choice of B3LYP functional for DFT calculations yields remarkable performance in calculating energetic and structural properties of bio- and organic molecules in close agreement to experimentally derived values.<sup>191</sup> However, effects of truncation of the region of interest have not been carefully estimated in the literature.

In this chapter<sup>67, 68, 69</sup>, we perform QC calculations on metal ion bound peptides representing truncated binding region of a large protein to systematically explore coordination stability and functionality. In particular we are interested on how truncation affects ground state energy levels. We consider calcium ( $\text{Ca}^{2+}$ ) ion coordination to Calmodulin (CaM), a well studied  $\text{Ca}^{2+}$  coordinating EF-hand<sup>62, 192</sup> metalloprotein, which participates in calcium sensing and signal transduction in eukaryotic cells<sup>9</sup>. The EF-hands exhibit a helix-loop-helix structure with  $\text{Ca}^{2+}$  ions binding to the loop region. CaM has four such  $\text{Ca}^{2+}$  binding sites: The loops 1 and 2 form the N-terminal domain (NTD) and the loops 3 and 4 form the C-terminal domain (CTD), the domains being connected by a linker region. The EF hand loops of CaM are twelve amino acids long (see Fig 6.1), where the I<sup>st</sup>, III<sup>rd</sup>, V<sup>th</sup>, VII<sup>th</sup> and the XII<sup>th</sup> residue of the EF hand loops are involved in  $\text{Ca}^{2+}$  coordination along with a water molecule, leading to pentagonal bipyramidal geometry. Mutational studies highlight the roles of both ligand and non-ligand residues in maintaining the coordination geometry and structural stability of the  $\text{Ca}^{2+}$  bound loops of CaM. The mutation of the coordinating bidentate acidic glutamate (E) leads to drastic loss in  $\text{Ca}^{2+}$  binding affinity.<sup>63</sup> The reduced sensitivity for  $\text{Ca}^{2+}$  is also reported from the mutation of the coordinating acidic aspartates (D).<sup>62, 63</sup> The importance of the conserved hydrophobic isoleucine (I) in maintaining structural stability of the loop, has also been reported through mutagenesis.<sup>63</sup> A recent DFT calculations on isolated CaM loops<sup>52</sup> reports the stabilization energy associated with electron delocalization from coordinating residues to  $\text{Ca}^{2+}$  ion. It sheds light on roles of coordinating residues, but those of non-coordinating residues remain obscured.



**Fig. 6.1** The four  $\text{Ca}^{2+}$  binding loops of CaM: (a) loop 1, (b) loop 2, (c) loop 3 and (d) loop 4. The loop residues are color coded according to their nature, acidic (red), basic (blue), polar (magenta) and hydrophobic (green). The coordinating residues (in italics) and their loop positions (I, III, V, VII and XII) are indicated in brackets.

In general, the polar amino acids, like tyrosine (Y), serine (S) and threonine (T) are functionally significant through participation in phosphorylation, a post-translational chemical modification (PTM), at their side chain hydroxyl  $O_h$  atom<sup>193</sup>. Phosphorylation of  $\text{Ca}^{2+}$  bound CaM leads to differential regulation of CaM-dependent enzymes<sup>16</sup>. It involves a reversible covalent linkage of phosphate group from adenosine-triphosphate (ATP) to specific protein sites, catalyzed by specific enzymes like kinases. The side chain  $O_h$  atom from polar amino acids acts as a nucleophile attacking the phosphorous ( $P_\gamma$ ) of the ATP in presence of the kinase<sup>193</sup> to get

reduced to ADP. Mass spectrometric analysis yields Y99 (loop 3) and Y138 (loop 4) as potential *in vivo* phosphorylation sites<sup>16</sup> in CaM involving their phenyl ring  $O_h$  atom. A recent DFT study shows that the  $O_h$  atom can act as a nucleophile for transfer of phosphoryl group from ATP to tyrosine in insulin receptor tyrosine kinase via a transition state with barrier height  $\sim 14$  kcal/mol<sup>194</sup>. Although serine, threonine and tyrosine are all present in CaM loops, neither serine nor threonine undergoes phosphorylation.

In this chapter we report QC-DFT calculations using B3LYP functionals on  $Ca^{2+}$  ion bound CaM loops:

(1) We assess the effect of truncation and choice of terminal capping on the ground state energy spectra through vacuum QC calculations. We elucidate the roles of both coordinating and non-coordinating residues in governing stability of  $Ca^{2+}$  ion coordination geometry. We find that the coordination energy levels having simultaneous contribution from  $Ca^{2+}$  ion and ligands along with contributions from non-ligand residues as well yield robust picture of coordination, independent of the choice of terminal capping.<sup>67</sup>

(2) The vacuum QC calculations yield unreliable HOMO-LUMO gaps. Often the surface exposed polar and charged side chains contribute to these levels. Such surface exposed side chains undergo polarization induced by the surrounding solvent. Hence the treatment of solvent is required for accurate calculation of these levels and the corresponding gap. The surrounding solvent effect is implicitly considered through the polarizable continuum model (PCM)<sup>195</sup>, where the molecule is embedded in a cavity surrounded by infinite dielectric medium of water (dielectric constant  $\sim 78.39$ ) and has been reported from earlier studies to successfully reproduce electronic properties in solution<sup>196</sup>. We perform implicit solvent QC calculations to compute the

HOMO-LUMO levels of  $\text{Ca}^{2+}$  bound CaM loops with neutral terminal capping.<sup>68</sup> These levels provide insight into the functionality of different loop residues, specially the nucleophilic character of Y99. The coordination picture across the loops is robust and remains similar to vacuum calculations.

(3) In an aqueous solution at neutral pH, any protein, peptide or amino acid acquires Zwitterionic character with protonated amide group at the N-terminus and de-protonated carboxyl group at C-terminus.<sup>197</sup> This leads us to cap terminals of the isolated CaM loops with charged groups resembling the Zwitterionic state having carboxylate ( $\text{C}_\gamma\text{O1O2}^-$ ) capping at C-terminal and amino ( $\text{NH}_3^+$ ) capping at N-terminal. The motivation is to understand how the functional aspects get modified if at all with biologically relevant Zwitterionic form of the  $\text{Ca}^{2+}$  binding loops. The charged terminals corresponding to Zwitterionic state introduce pure localized states near the HOMO-LUMO levels which decay exponentially in the neighboring levels.<sup>69</sup>

## 6.2 Methods

The 1.7 Å X-ray crystal structure of  $\text{Ca}^{2+}$  bound (holo) CaM (PDB ID: 1CLL)<sup>61</sup> is used in our study. We isolate the the  $\text{Ca}^{2+}$  binding EF-hand loops: loop 1 (D20 to E31), loop 2 (D56 to E67), loop 3 (D93 to E104) and loop 4 (D129 to E140) retaining coordinating waters and  $\text{Ca}^{2+}$  ions, solvate using the TIP3P model and neutralize with counter ions. The systems are energy minimized involving only hydrogen and terminal capping atoms in NAMD<sup>198</sup> using the CHARMM27 parameters<sup>109</sup>.

The QC DFT calculations are done by variationally minimizing the many-body electronic Hamiltonian for fixed nuclear coordinates where the electronic correlations are approximately taken into account. We perform the electronic minimization using the Gaussian03 package<sup>199</sup> with the B3LYP<sup>189</sup> functional for the electron exchange-correlation in the 6-31G(d,p)<sup>190</sup> basis set. B3LYP (Becke, three parameter, Lee-Yang-Parr) is a hybrid functional constructed as a linear combination of Hartree-Fock exchange functional and DFT exchange-correlation functional. The basis sets represent a set of functions, linear combinations of which yield molecular orbitals (MOs). The 6-31G basis set is an atom centred Gaussian basis function with 6 primitive Gaussians representing the core, while the valence orbitals are composed of two basis functions: the first one is a linear combination of 3 primitive Gaussians and the second one as 1 primitive Gaussian. The basis set is polarized by adding ‘*d*’ orbitals to heavy atoms like nitrogen, carbon and oxygen and ‘*p*’ orbitals for hydrogen.

We perform vacuum QC calculations on CaM loop 1 with different truncation length and terminal capping. We cap the two truncated terminals with either charged Zwitterionic or neutral groups. The Zwitterionic capping involves amino ( $NH_3^+$ ) and carboxylate ( $COO^-$ ) groups at N- and C-terminals respectively, while acetyl ( $COCH_3$ ) group at N-terminus and N-methylamide ( $CH_3NH$ ) group C-terminus resemble neutral capping. We also perform QC calculations for all the  $Ca^{2+}$  binding loops of CaM in implicit solvent considered through the polarizable continuum model (PCM)<sup>195</sup>.

The atom condensed Fukui index have the following form:  $f_k^+ = q_k^{N+1} - q_k^N$  and  $f_k^- = q_k^N - q_k^{N-1}$ , where  $q_k^N$  represents the charge on atom  $k$  for a system of  $N$  electrons.  $q_k^{N-1}$  denotes the same for a system of  $N - 1$  electrons upon removal of an electron from HOMO

and similarly  $q_k^{N+1}$  for a system of  $N + 1$  electrons upon addition of an electron to the LUMO. Here,  $f_k^-$  measures the capacity of atom  $k$  to provide a nucleophilic attack. Similarly,  $f_k^+$  denotes the tendency to act as an electrophile<sup>200</sup>.

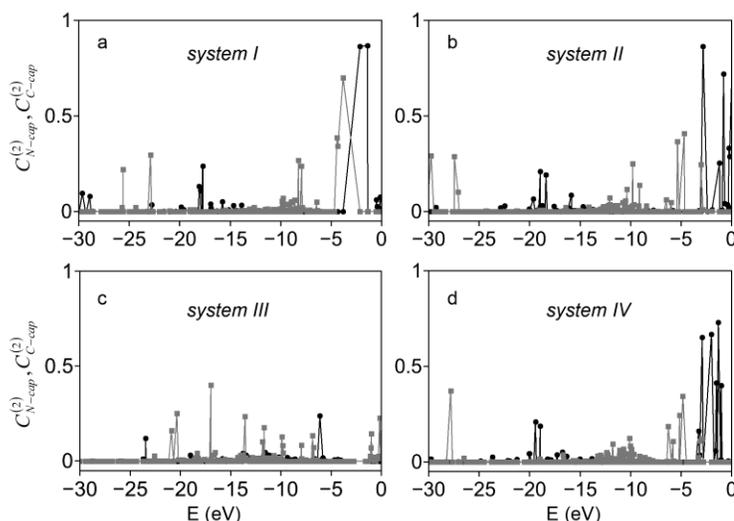
## 6.3 Results

### A. Vacuum QC calculations

We perform vacuum QC calculations on  $\text{Ca}^{2+}$  ion bound CaM loop 1 with different truncation length and terminal capping through charged and neutral moieties to elucidate their effect on ground state electronic energy spectra. We have used amino ( $\text{NH}_3^+$ ) at N-terminal and carboxylate ( $\text{C}_\gamma\text{O}1\text{O}2^-$ ) at C-terminal for charged terminal capping resembling the Zwitterionic state. On other hand, for neutral capping we have used acetyl ( $\text{COCH}_3$ ) group at the N-terminal and N-methylamide ( $\text{CH}_3\text{NH}$ ) group at the C-terminal. To this end we consider the following systems: (a) System I, classical force-field optimized geometry of  $\text{Ca}^{2+}$  bound loop 1 with Zwitterionic capping. (b) System II, crystal structure geometry of  $\text{Ca}^{2+}$ -loop 1 complex with similar capping as the former. (c) System III, similar to system II but with neutral terminal capping. System IV, also similar to system II with two extra residues at the two ends and Zwitterionic capping at terminals.

*Capping contributions:* The capping contribution,  $C_{N\text{-cap}}^{(2)}$  and  $C_{C\text{-cap}}^{(2)}$  is calculated as the sum of squares of MO coefficients of the constituent atoms of respective N- and C-terminal groups and shown in Fig 6.2. The charged Zwitterionic terminals of system I (Fig 6.2 a) indicate predominant effect near the HOMO-LUMO levels. The contribution of N-terminal  $\text{NH}_3^+$  cap ,

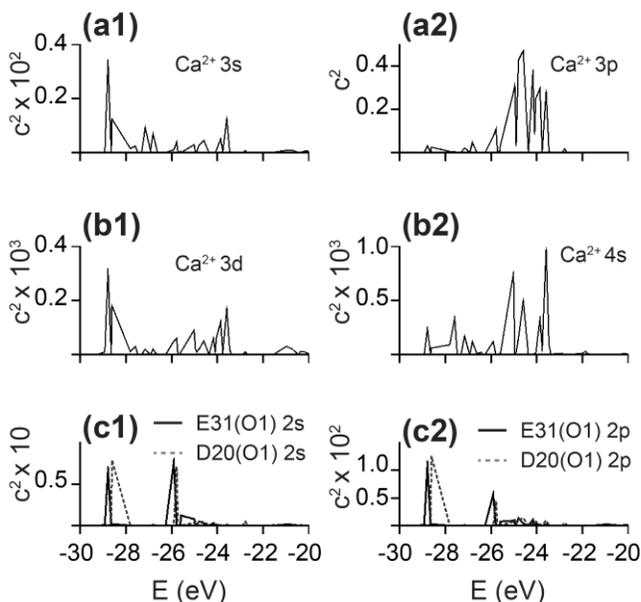
$C_{N-cap}^{(2)}$  is maximum at the LUMO level, while C-terminal  $C_{\gamma O1O2^- cap, C_{C-cap}^{(2)}$  indicates peak at the HOMO. The energy range -20 eV to -5 eV below HOMO ( $\sim -4$  eV) show little contribution from either charged terminals. The Zwitterionic terminals of system II (Fig 6.2 b) show similar predominant effects at the HOMO-LUMO levels but decrease at energy states below the HOMO ( $\sim -2$  eV). Unlike the former systems, the neutral capped system III (Fig 6.2 c) shows minimal capping contribution throughout the energy spectrum. Both  $C_{N-cap}^{(2)}$  and  $C_{C-cap}^{(2)}$  are observed to be negligible near the HOMO-LUMO levels. The terminal capping of system IV (Fig 6.2 d) being Zwitterionic similar effects as that of system I and system II are observed.



**Fig. 6.2** The capping contribution,  $C_{N-cap}^{(2)}$  (black circle) and  $C_{C-cap}^{(2)}$  (grey square) for (a) system I, (b) system II, (c) system III and (d) system IV across the ground state energy spectra.

*Coordination levels:* We examine the energy levels below HOMO having simultaneous contributions from the  $Ca^{2+}$  ion and some of the loop residues, designated as coordination states. The  $Ca^{2+}$  ion coordinates to the loop via the oxygen atoms through hybridization. The hybridization of the valence orbitals of  $Ca^{2+}$  ion is of the spd type, while the oxygen valence states adopt the sp hybrid character. The spd hybridized orbitals of  $Ca^{2+}$  ion further hybridize with the sp hybrid orbitals of ligand oxygen atoms giving rise to the coordination states. We

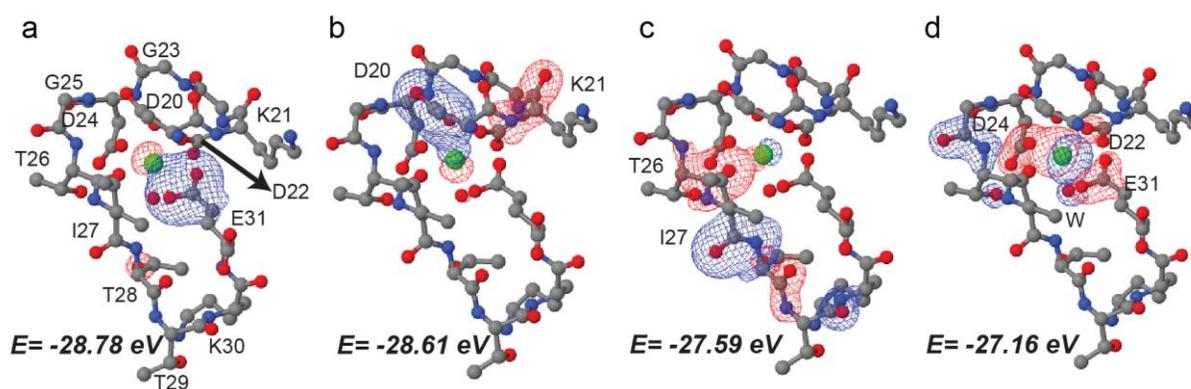
show in Fig 6.3, these contributions for different energies,  $E$ , in the window of -29 eV to -24 eV. Fig 6.3 (a1) and 6.3 (a2) show the squares of the MO coefficients ( $C^{(2)}$ ) of 3s and 3p levels of  $\text{Ca}^{2+}$ . Similarly, the Fig 6.3 (b1) and 6.3 (b2) show  $C^{(2)}$  of 3d and 4s levels of  $\text{Ca}^{2+}$ . Fig 6.3 (c1) and 6.3 (c2) show  $C^{(2)}$  for the valence levels of ligand oxygen atoms for E31 and D20.



**Fig 6.3.** The contributions (square of MO coefficient,  $C^{(2)}$ ) of the valence orbitals (a1) 3s, (a2) 3p, (b1) 3d and (b2) 4s of  $\text{Ca}^{2+}$  ion to coordinating energy levels ( $E$ ) between -29 to -24 eV. The valence orbitals of ligand oxygen atom (O1) of E31 (black solid) and D20 (grey dashed) (c1) 2s, (c2) 2p also contribute to similar energy range.

Fig 6.4 (a) shows the electron densities on different loop atoms and the  $\text{Ca}^{2+}$  ion in system I in its lowest coordination molecular orbital (LCMO) for  $E = -28.78$  eV. It indicates the hybridization between the ligand carboxylate oxygen (O1 and O2) atoms of E31 with the  $\text{Ca}^{2+}$  ion. The electron density at LCMO is highly localized over E31 and the  $\text{Ca}^{2+}$  ion without any contribution from other loop residues. The two carboxylate oxygen atoms of E31 are in the bonding state which further bonds with the  $\text{Ca}^{2+}$  ion in the LCMO. The axial D20 (Fig 6.4 (b)) shows predominant contribution to the coordinating state MO at  $E = -28.61$  eV through the bonding states of its side chain carboxylate oxygens, further bonding with the ion. The backbone atoms of non-ligand K21 are also involved in forming an antibonding state with the D20 carboxylate oxygens. The immediate energetically higher coordination state at  $E = -27.59$  eV

(Fig 6.4 (c)) is characterized by the participation of the bonding state of the backbone carbonyl oxygen of the planar T26 with the ion and antibonding with the amide nitrogen of non-ligand I27. The carboxylate oxygens of the planar aspartates, D22 and D24, and oxygen of the axial water (W) are in bonding states with the ion, while the carboxylate oxygens of E31 antibond with the ion in the coordination state at  $E = -27.16$  eV (Fig 6.4 (d)).



**Fig 6.4.** Electron density contour plot of a few representative coordination state MOs. The loop atoms include N (blue), C (grey) and O (red) and  $\text{Ca}^{2+}$  (green). Red and blue meshes indicate electron rich and deficient regions respectively. (a) LCMO at  $-28.78$  eV and other coordination state MOs at (b)  $-28.61$  eV (c)  $-27.59$  eV and (d)  $-27.16$  eV. The 3d-plots are generated using JMOL by setting an isosurface value (0.01).

We summarize in Table 6.1, the detailed residue-wise picture of mixing of levels of loop atoms along with those of  $\text{Ca}^{2+}$  levels in different coordinating states of system I. The stabilization of the metal ion coordinating states is as follows: The LCMO is formed by E31. The stabilization is further achieved by the participation of the axial D20 along with the non-ligand K21. The plane of the coordination geometry is stabilized via T26 along with the non-ligand, I27. The residues D22 and D24 along with the coordinating water molecule complete the bipyramidal geometry. The participation of the water molecule in almost all the levels indicates that the crystal water has a smeared out contribution to different levels, leading to the overall stability in the ground state. The role of non-coordinating I27 is particularly remarkable in its

participation in a coordinating MO directly. The importances of the residues, like E31, D20, D22, D24 and I27 have been established by the earlier mutational studies.<sup>62, 63, 201</sup>

**Table 6.1. Residues contributing to different coordination state MOs of system I (x indicates no residue contribution).**

$E$ (eV)	Bonding/ Antibonding	$E$ (eV)	Bonding/ Antibonding
-28.78	E31 /x	-25.90	D22 E31 W /D20
-28.61	D20 /K21	-25.78	D20 W /D22
-27.80	T26 I27 /x	-25.63	E31 /x
-27.59	T26 /I27	-25.02	D20 D22 T26 E31 W /x
-27.16	D22 D24 T26 W /E31	-24.96	D20 D24 T26 W /D22 E31
-26.92	D24 T26 /G25	-24.90	E31 W /x
-26.82	D22 D24 W /T26	-24.80	D20 D22 D24 T26 E31 W /x

We check the robustness of the coordinating MOs for systems II, III and IV (Table 6.2). We find that the energy window and the MOs of the coordinating levels in these systems are very similar (-29 eV to -24 eV). The LCMO characterized by E31 is robust for all the systems studied. The higher coordination states find contributions from other planar and axial ligands. The non-coordinating K21 and I27 also contribute to hybridization at these coordination states irrespective of the systems studied. The role of coordinating water molecule is also robust in showing diffused contribution spread across different coordinating levels. Thus the composition of the coordinating MOs, and hence, the stabilization of the coordination geometry are not so sensitive to the truncation of the binding loop and the terminal capping.

**Table 6.2. Residues contributing to different coordination state MOs of system II, III and IV (x indicates no residue contribution).**

System II		System III		System IV	
E (eV)	Bonding/Antibonding	E (eV)	Bonding/Antibonding	E (eV)	Bonding/Antibonding
-29.01	E31/x	-29.04	E31/x	-29.08	E31/x
-28.52	T26 I27/x	-28.05	T26/I27	-28.70	T26 I27/x
-27.59	D20/K21	-26.81	D20 W/K21	-28.38	T26/I27
-26.78	D20 W/x	-26.80	D20 W/x	-27.08	D20 W/K21
-26.72	W/D20	-26.23	D20 E31 W/x	-27.01	D20 W/x
-26.33	D22 T26 E31/x	-26.07	D20 D22 E31/T26 I27	-26.89	W/D20
-26.21	D20 G25 T26/W	-25.92	D24 G25 T26 E31 W/x	-26.30	D22 T26 E31/x
-26.12	D20 D24 G25 E31/T26 W	-25.88	D20 D22 D24 T26/G25 E31	-26.19	D24 G25 T26 E31 W/x
-26.09	D22 D24 W/E31	-25.78	E31/D20 W	-26.13	D22 E31/D20 T26 W
-25.73	E31/D22 G23	-25.54	D22 D24/E31	-25.64	D22 E31/D24
-25.51	D22 D24 T26 E31/x	-25.42	D22 E31/x	-25.61	E31/D22 G23 D24
-25.42	D24 T26/D22 G23 E31	-25.40	D24 T26 E31/x	-25.57	E31 T26 D24/D22
-25.29	D22/D24	-24.99	x/D22 T26 E31	-25.30	x/D22 T26 E31
-25.22	E31/D22 T26				

*Comparison with earlier works:* We compare our results to those from the earlier DFT calculations<sup>52</sup> on the isolated CaM loops. The earlier study reports the stabilization energy<sup>202</sup> associated with electron delocalization from the occupied lone pair orbitals of the ligand oxygen atoms to the vacant orbitals of Ca<sup>2+</sup>. For the sake of comparison, we compute the charge gained

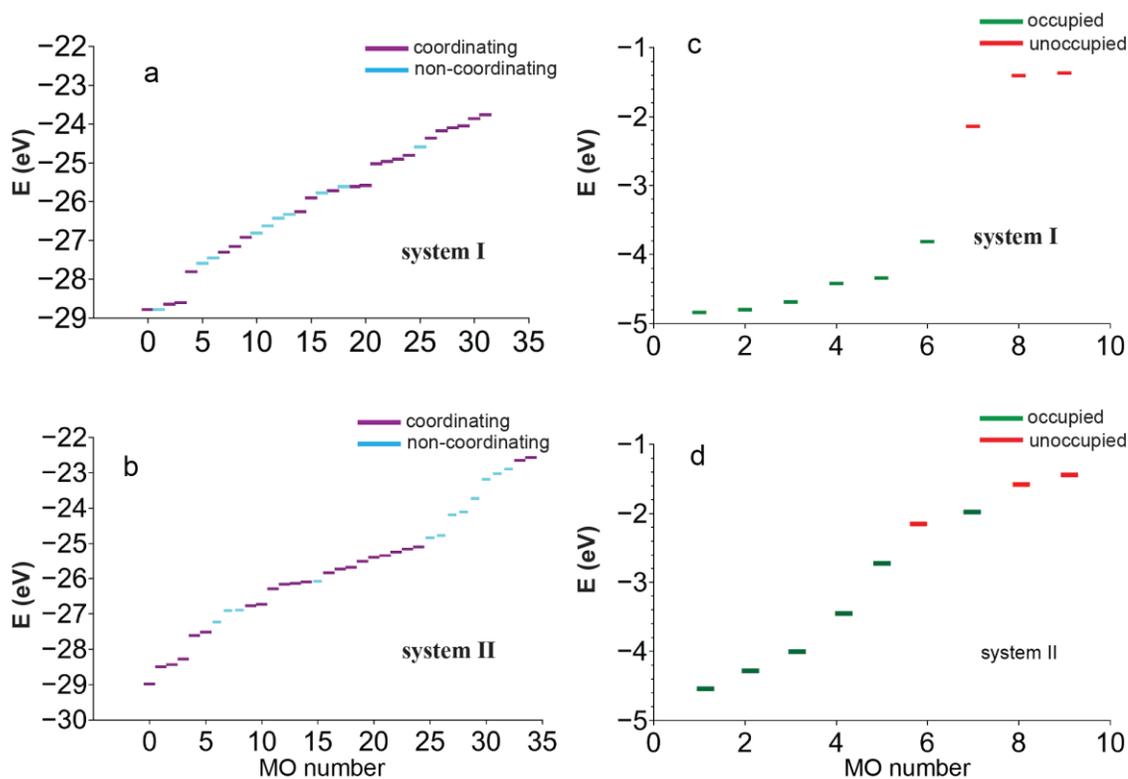
or lost by the ligand atoms with respect to the  $\text{Ca}^{2+}$  ion depleted form of CaM-loop 1 complex as described in system I. The oxidation state of the ion is +1.3, indicating transferring charge of 0.7 from ligand atoms to the ion. The details of the charge transfer ( $\Delta Q$ ) are given in Table 6.3. We consider, in particular, the case of E31. Both the carboxylate oxygen atoms of E31 undergo a small gain in charge which is qualitatively similar to the weak delocalizations observed in the earlier DFT calculations.<sup>52</sup> Our calculations show significant involvement of D22 in charge transfer through its ligand *O1* atom, followed by D24, T26, W and D20. Thus the charge transfer undermines the role of both E31 and D20, contrary to the mutation experiments.

**Table 6.3.** Amount of charge gained ( $\Delta Q = -ve$ ) or lost ( $\Delta Q = +ve$ ) by ligand atoms in system I and system II.

<i>ATOM</i>	$\Delta Q$ (System I)	$\Delta Q$ (System II)	<i>ATOM</i>	$\Delta Q$ (System I)	$\Delta Q$ (System II)
<i>D20 O1</i>	-0.08	-0.13	<i>D24 O2</i>	+0.05	-0.02
<i>D20 O2</i>	+0.03	+0.08	<i>T26 O</i>	-0.14	-0.17
<i>D22 O1</i>	-0.18	-0.22	<i>E31 O1</i>	-0.04	-0.05
<i>D22 O2</i>	+0.04	-0.04	<i>E31 O2</i>	-0.03	-0.05
<i>D24 O1</i>	-0.14	-0.29	<i>Water O</i>	-0.13	-0.06

In order to understand the discrepancies in charge transfer, we compare the energy levels in systems I and II. Fig 6.5 (a) shows the coordination energy spectrum in the range -29 eV to -24 eV that harbours the different coordination states of system I. The coordinating energy levels are interspersed by several non-coordinating levels, having dominant contributions from non-ligand atoms but not the  $\text{Ca}^{2+}$  ion. Since the coordinating states lie far (~20-25 eV) below the HOMO, these levels in system II remain unaffected by system specific details (Fig 6.5 (b)). However, we find significant differences close to the HOMO-LUMO levels between system I (Fig 6.5 (c)) and system II (Fig 6.5 (d)). For instance, the HOMO level of system I occurs at -4

eV in contrast to -2 eV for system II. Moreover, an artefact of vacuum calculation is observed in system II as the LUMO level is slightly below HOMO, thus yielding unphysical HOMO-LUMO gap. Such differences affect the charge transfer calculations, which takes into account all the MOs.



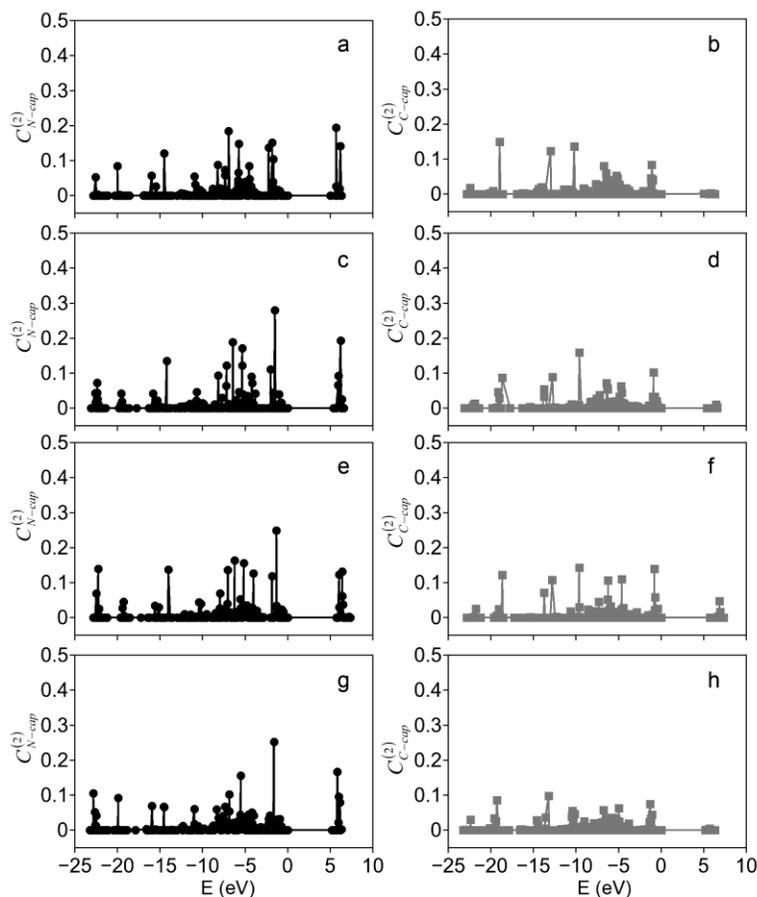
**Fig 6.5.** The coordination energy spectrum (E) plotted against the MO number for different systems: (a) system I, range -29 eV to -22 eV, (b) system II, range -30 eV to -22 eV. The coordinating (violet) levels are interspersed by few non-coordinating (blue) levels. Energy levels near HOMO-LUMO of (c) system I and (d) system II.

We also compare the oxidation states of the ion in systems I and II to account for the observed discrepancies between the charge transfer and the coordinating MOs. We find that  $\text{Ca}^{2+}$  attains a higher oxidation state (+1.5) in system II than in system I. The  $\Delta Q$  values for the ligands are also different in system I and system II (see Table 6.3), although the qualitative trend

is similar in both the cases. These differences are attributed to changes in system specific details. For instance, system I and system II, when aligned, gives a backbone root mean squared displacement of 0.6 Å. Moreover, the distances between the ligands and the ion in system I is less than those in system II. The shorter distances in system I enable stronger hybridization of  $\text{Ca}^{2+}$  with the ligand atoms, resulting in larger charge redistribution than in system II.

## B. Implicit solvent QC calculations with neutral capping

We perform QC calculations in implicit solvent for all the four  $\text{Ca}^{2+}$  bound loops of CaM. We have used the X-ray crystal structure geometries of  $\text{Ca}^{2+}$  bound loops with neutral capping: acetyl ( $\text{COCH}_3$ ) group at N-terminal and N-methylamide ( $\text{CH}_3\text{NH}$ ) group at C-terminal.



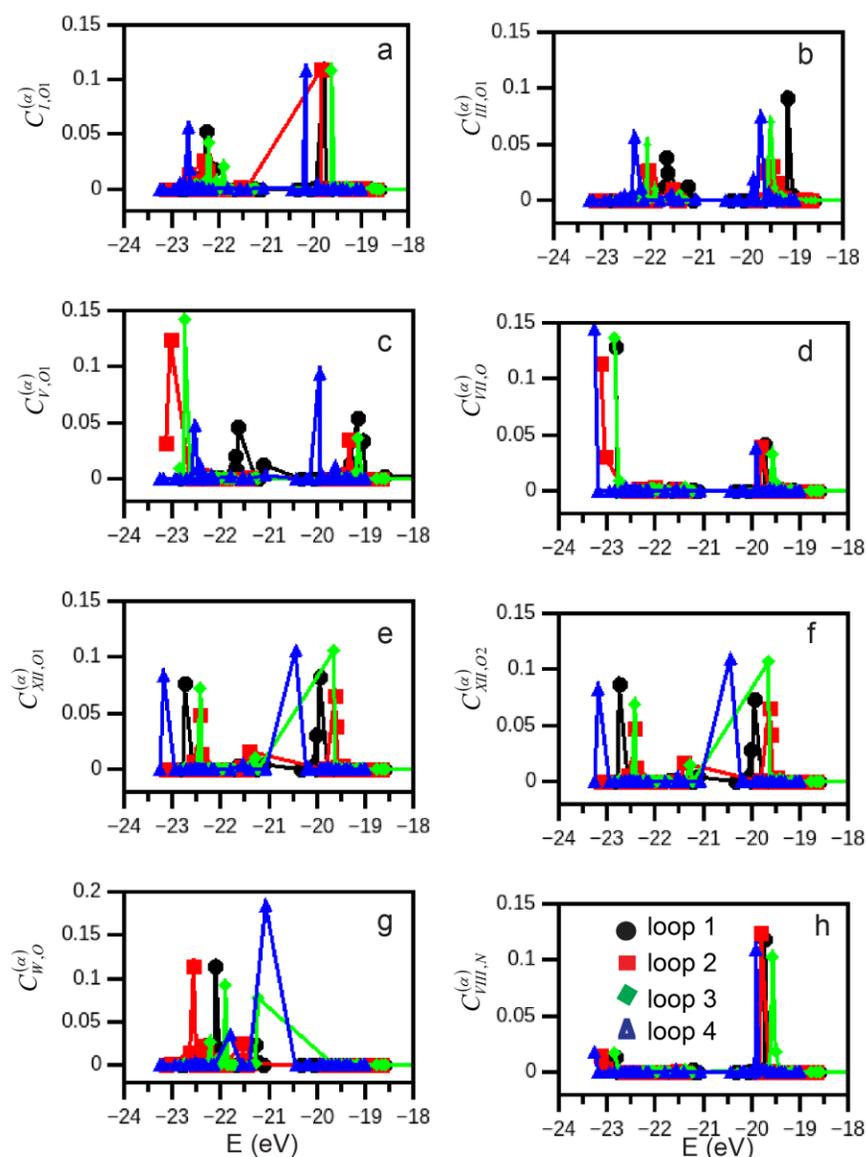
**Fig 6.6. The contribution of terminal neutral capping: (a)  $C_{N-cap}^{(2)}$  and (b)  $C_{C-cap}^{(2)}$  for  $\text{Ca}^{2+}$ -loop 1, (c)  $C_{N-cap}^{(2)}$  and (d)  $C_{C-cap}^{(2)}$  for  $\text{Ca}^{2+}$ -loop 2, (e)  $C_{N-cap}^{(2)}$  and (f)  $C_{C-cap}^{(2)}$  for  $\text{Ca}^{2+}$ -loop 3 and (g)  $C_{N-cap}^{(2)}$  and (h)  $C_{C-cap}^{(2)}$  for  $\text{Ca}^{2+}$ -loop 4.**

*Capping contributions:* The contribution of neutral capping at N-terminal is taken as the sum of squares of MO coefficients,  $C_{N-cap}^{(2)}$  of the constituent atoms of  $\text{COCH}_3$  group and that at

C-terminal,  $C_{C-cap}^{(2)}$  for CH<sub>3</sub>NH group. We show  $C_{N-cap}^{(2)}$  and  $C_{C-cap}^{(2)}$  for the different Ca<sup>2+</sup> bound loops of CaM across the ground state energy range relative to the HOMO level (Fig 6.6). The terminal neutral cap contributions are negligible across the ground state energy spectra including the HOMO and the LUMO levels and consistent for all the four loops. The corresponding HOMO-LUMO gaps,  $\Delta^{neu}$  is given in Table 6.4. The HOMO-LUMO gaps in the N-terminal loops are 5.01 eV (Ca<sup>2+</sup>-loop 1) and 5.37 eV (Ca<sup>2+</sup>-loop 2). Similarly, HOMO-LUMO gaps of 5.71 eV (Ca<sup>2+</sup>-loop 3) and 5.19 eV (Ca<sup>2+</sup>-loop 4) are observed for the C-terminal loops. The coordination levels of the Ca<sup>2+</sup>-loop 1 complex occupy an energy range of -22.82 eV to -19.52 eV. The coordination energy range in implicit solvent is similar to that calculated in vacuum. The other three loops also occupy similar coordination energy range: Ca<sup>2+</sup>-loop 2 (-23.13 eV to -19.34 eV), Ca<sup>2+</sup>-loop 3 (-22.84 eV to -19.42 eV) and Ca<sup>2+</sup>-loop 4 (-23.26 eV to -19.70 eV).

**Table 6.4. The HOMO-LUMO energy gap (in eV) for neutral and Zwitterionic capped CaM loops along with the decay constants (in eV) of Zwitterionic capping groups of CaM loops.**

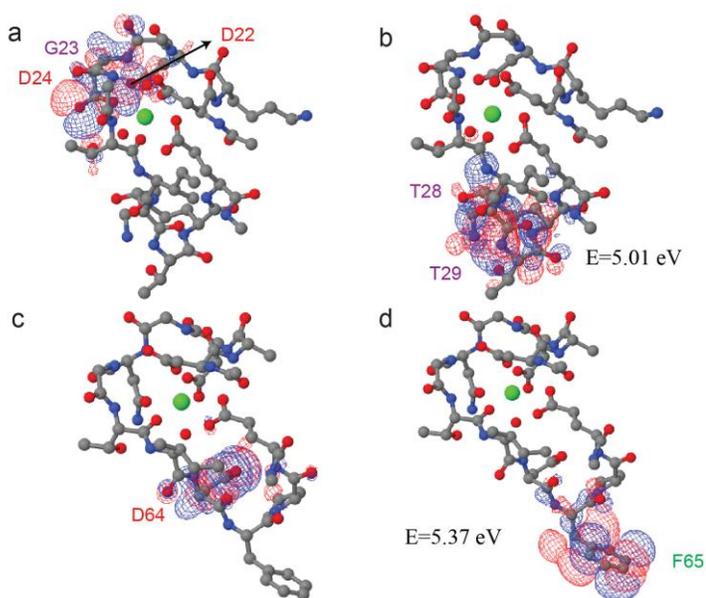
	Neutral	Zwitterionic		
	$\Delta^{neu}$	$\Delta^{Zw}$	$\xi_{C-ter}^{loop}$	$\xi_{N-ter}^{loop}$
Loop 1	5.01	5.10	0.4	0.03
Loop 2	5.37	5.08	0.3	0.01
Loop 3	5.71	4.98	0.5	0.01
Loop 4	5.19	4.62	0.01	0.03



**Fig 6.7.** The coordination energy range for neutral capped systems:  $\text{Ca}^{2+}$ -loop 1 (circle),  $\text{Ca}^{2+}$ -loop 2 (square),  $\text{Ca}^{2+}$ -loop 3 (diamond) and  $\text{Ca}^{2+}$ -loop 4 (triangle) corresponding to the QC calculation on X-ray crystal structure geometry.

*Coordination levels:* The contribution of the ligand and non-ligand atoms from the four different loops across the coordination energy range is illustrated in Fig 6.7. The carboxylate oxygen,  $O1$  of the coordinating aspartates at the  $\text{I}^{\text{st}}$  position of all the loops shows maximum

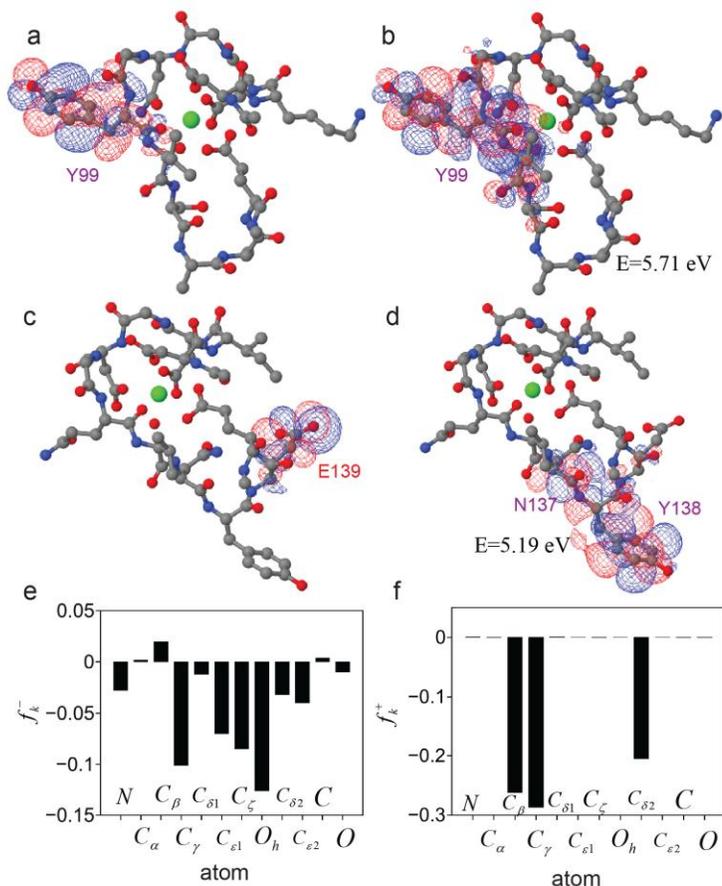
contribution ( $C_{I,O1}^{(\alpha)}$ ) between  $E \sim -20.5$  eV and  $-19.5$  eV along with secondary peaks in a lower energy range  $-23$  eV to  $-21$  eV (Fig 6.5 a). The contributions ( $C_{III,O1}^{(\alpha)}, C_{V,O1}^{(\alpha)}, C_{VII,O}^{(\alpha)}, C_{XII,O1}^{(\alpha)}, C_{XII,O2}^{(\alpha)}$ ) of ligand oxygen from the other residues show similar feature in comparable energy ranges (Fig 6.7 b-f). The coordinating water molecule (Fig 6.7 g) shows diffused contribution ( $C_{W,O}^{(\alpha)}$ ) at  $E \sim -23$  eV to  $-21$  eV. The non-ligand contribution  $C_{VIII,N}^{(\alpha)}$  (Fig 6.7 h) predominates at  $E \sim -20$  eV and  $-19.5$  eV along with small contribution at energetically lower CSMOs with  $E \sim -23$  eV.



**Fig 6.8 (a) HOMO and (b) LUMO of neutral capped  $\text{Ca}^{2+}$ -loop 1 complex. (c) HOMO and (d) LUMO of neutral capped  $\text{Ca}^{2+}$ -loop 2 complex.**

*HOMO-LUMO levels:* The HOMO of neutral capped  $\text{Ca}^{2+}$ -loop 1 (Fig 6.8 (a)) complex is predominantly derived from the side chain carboxylate oxygen,  $O1$  and  $O2$  atoms of D22, D24 and the main chain carbonyl oxygen  $O$  atom of G23. The corresponding LUMO (Fig 6.8 (b)) at  $E=5.01$  eV finds major contribution through both backbone and side chain atoms of T28 and T29. The HOMO in neutral capped  $\text{Ca}^{2+}$ -loop 2 complex (Fig. 6.8 (c)) is characterized by strong contributions of the side chain carboxylate oxygen atoms of D64. The contribution to the

corresponding LUMO (Fig. 6.8 (d)) originates from the strong localized electron density on the side chain phenyl ring of F65.



**Fig 6.9.** (a) HOMO and (b) LUMO of the Ca<sup>2+</sup>-loop 3 complex. (c) HOMO and (d) LUMO of the Ca<sup>2+</sup>-loop 4 complex. (e) The nucleophilicity index,  $f_k^-$  and (f) the electrophilicity index,  $f_k^+$  of Y99 atoms in Ca<sup>2+</sup>-loop 3 complex.

The HOMO of neutral capped Ca<sup>2+</sup> bound loop 3 (Fig 6.9 (a)) shows strong localized electron density on the side chain phenyl ring of Y99. The LUMO (Fig 6.9 (b)) at  $E=5.71$  eV is also characterized through strong contribution from the Y99 side chain. The HOMO of Ca<sup>2+</sup>-loop 4 complex (Fig 6.9 (c)) shows localized charge density on the side chain carboxylate group of E139. On the other hand, the electronic density at the corresponding LUMO (Fig 6.9 (d)) is

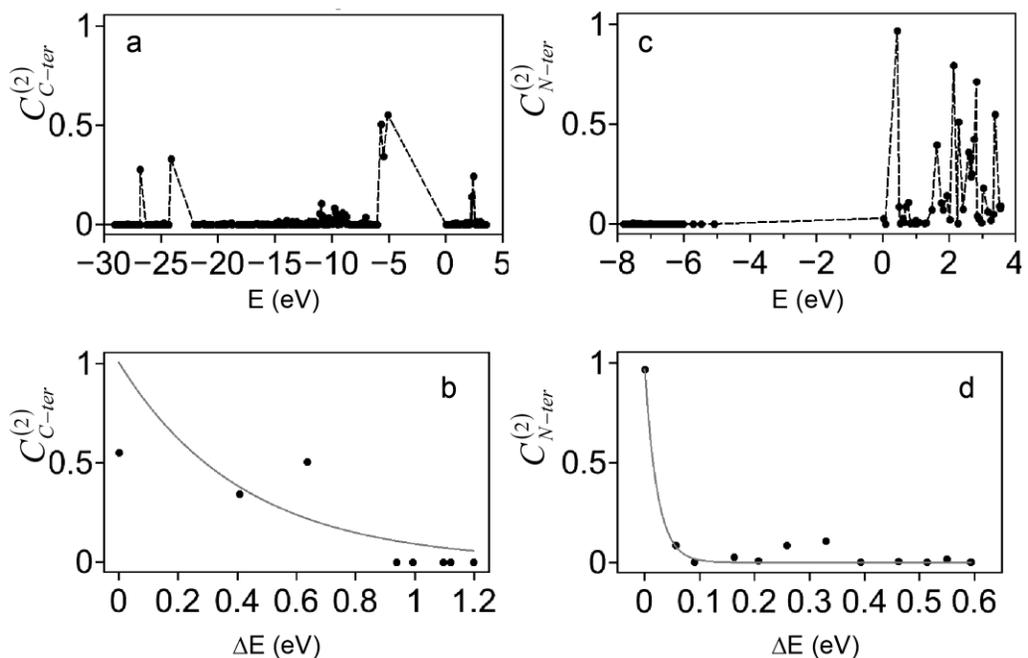
essentially localized on the phenyl ring of Y138 and the side chain carboxy-amide group of N137.

HOMO of neutral capped  $\text{Ca}^{2+}$  bound loop 3 has strong contribution from Y99, where the side chain phenyl ring has strong localized electron density. Such sites being rich in electrons can act as nucleophile. Since earlier QC calculations indicate phosphorylation of tyrosine through a transition state involving  $O_h$  of phenyl ring as a nucleophile, we ascertain the nucleophilic character for all non-hydrogen atoms of Y99 through the computation of atom-condensed Fukui index<sup>49</sup>. Fig 6.9 (e) illustrates the nucleophilicity index,  $f_k^-$  for the atoms of Y99. The  $O_h$  attached to the phenyl ring shows maximum nucleophilicity followed by the  $C_\gamma$  atom of the ring itself. We illustrate the electrophilic index,  $f_k^+$  of the Y99 atoms in Fig 6.9 (f). The phenyl ring carbon atoms  $C_\beta$ ,  $C_\gamma$  and  $C_{\delta 2}$  show significantly large electrophilicity compared to other atoms.

### C. Implicit solvent QC calculation with Zwitterionic capping

*Capping contributions:* The Zwitterionic capping contributions,  $C_{C-ter}^{(2)}$  and  $C_{N-cap}^{(2)}$  of the  $\text{Ca}^{2+}$  bound loop 1 is shown in Fig 6.10.  $C_{C-ter}^{(2)}$  (Fig 6.10 (a)) of  $\text{Ca}^{2+}$ -loop 1 shows maximum at the HOMO with decreasing amplitude towards lower energy levels (-5 eV to -20 eV).  $C_{C-ter}^{(2)}$  is non-zero at -24.1 eV and -26.8 eV, the electron density being solely localized over the capping carboxylate oxygen atoms. Fig 6.10 (b) shows that the contribution,  $C_{C-ter}^{(2)}$  decreases exponentially ( $\approx \exp\left(\frac{-\Delta E}{\xi_{C-ter}^{(L1)}}\right)$ ) from the HOMO towards the lower energy levels ( $\Delta E = |E_i - E_{HOMO}|$ ) with a decay constant,  $\xi_{C-ter}^{(L1)} = 0.4$  eV of loop 1. Similarly,  $C_{N-ter}^{(2)}$  (Fig 6.10 (c)) is predominant at the LUMO but decreases in amplitude at higher energies (0 eV to 4 eV).

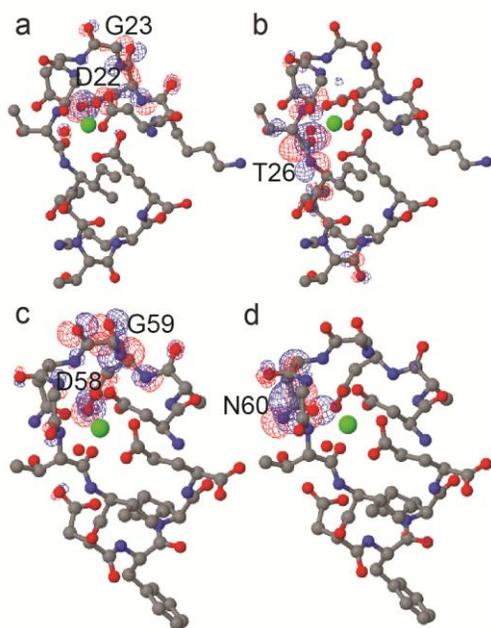
Fig 6.10 (d) shows the amino capping contribution  $C_{N-ter}^{(2)}$  to the LUMO and higher energy levels ( $\Delta E = |E_i - E_{LUMO}|$ ), exhibiting an exponential decay with  $\xi_{N-ter}^{(L1)} = 0.03$  eV in loop 1. Similar exponential decay of terminal capping contributions is observed for the other loops with respective decay constants given in Table 6.4 along with the corresponding HOMO-LUMO energy gaps,  $\Delta^{Zw}$  of Zwitterionic CaM loops.



**Fig 6.10.** The contribution of terminal capping of  $\text{Ca}^{2+}$ -loop 1 complex. (a) C-terminal  $C_{\gamma}O1O2^{-}$  capping contribution and (b) its exponential decay from HOMO to lower energy states (in eV). (c) N-terminal  $NH_3^{+}$  group capping contribution and (d) its decay from LUMO to higher energy states (in eV).

*HOMO-LUMO levels:* We compare the contribution of Zwitterionic charged capping with respect to that of neutral capping at the HOMO-LUMO levels. The maximum contribution of  $C_{\gamma}O1O2^{-}$  capping of loops 1 and 2 occur at energies similar to the HOMO level in corresponding neutral capped systems. However, for both the Zwitterionic capped  $\text{Ca}^{2+}$ -loop 3 and  $\text{Ca}^{2+}$ -loop 4 complexes, the maximum contribution of  $C_{\gamma}O1O2^{-}$  capping is observed at

about 0.5 eV higher than the corresponding HOMO level in the neutral capped cases. The negatively charged C-terminal capping,  $C_\gamma O1O2^-$  with an excess electron induces an additional electronic state within the HOMO-LUMO gap of the neutral capped peptides leading to bridging of the gap. The N-terminal  $NH_3^+$  capping contributes very close to the LUMO of the neutral capped CaM peptides. Thus the gap decreases from 5.71 eV to 4.98 eV in  $Ca^{2+}$  bound loop 3 and from 5.19 eV to 4.62 eV in  $Ca^{2+}$  bound loop 4 upon introducing Zwitterionic charged states at the terminals.

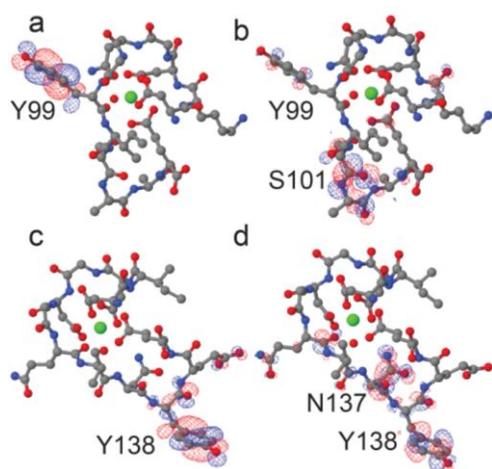


**Fig 6.11.** (a)  $HOMO^-$  and (b)  $LUMO^+$  of Zwitterionic capped  $Ca^{2+}$ -loop 1 complex. (c)  $HOMO^-$  and (d)  $LUMO^+$  of Zwitterionic capped  $Ca^{2+}$ -loop 2 complex.

*HOMO<sup>-</sup> and LUMO<sup>+</sup> levels:* We designate the levels with negligible capping effect closest to HOMO as  $HOMO^-$  ( $E^-$ ) and LUMO as  $LUMO^+$  ( $E^+$ ) of the  $Ca^{2+}$  bound loops of CaM. The redefined  $HOMO^-$  and  $LUMO^+$  bear qualitative similarities with the HOMO-LUMO levels of neutral capped CaM loops. In case of the Zwitterionic capped  $Ca^{2+}$  bound loop 1, the  $HOMO^-$  (Fig 6.11 (a)) finds participation from acidic aspartate, D22 and polar G23, while another polar residue, T26 contributes to the  $LUMO^+$  (Fig 6.11 (b)). The nature of residues contributing

to HOMO<sup>-</sup> and LUMO<sup>+</sup> levels of loop 1 is similar in both studies. The HOMO<sup>-</sup> of loop 2 (Fig 6.11 (c)) with charged capping shows contribution of acidic D58 and polar G59. The corresponding LUMO<sup>+</sup> finds contribution from polar N60 (Fig 6.11 (d)). The nature of contributing residues to HOMO<sup>-</sup> with Zwitterionic caps is similar; however the difference lies in the LUMO<sup>+</sup>, being governed by a polar residue in loop 2.

The HOMO<sup>-</sup> and LUMO<sup>+</sup> (6.12 (a)-(b)) levels of Zwitterionic Ca<sup>2+</sup>-loop 3 complex exhibits strong localized charge density on the polar side chain of Y99 similar to neutral capping. Moreover, the nucleophilic character of Y99 remains preserved with either capping, albeit a quantitative comparison of  $f_k^-$  of  $O_h$  atom reveals a less potential nucleophile ( $f_k^- \sim -0.05$ ) with Zwitterionic capping. We find that both HOMO<sup>-</sup> and LUMO<sup>+</sup> of Zwitterionic loop 4 (Fig 6.12 (c)-(d)) are predominated by polar Y138, unlike E139 and Y138 contributing to HOMO and LUMO levels respectively, of neutral capped loop 4. However, the role of Y138 as a potential nucleophile remains obscured with either of the cappings.



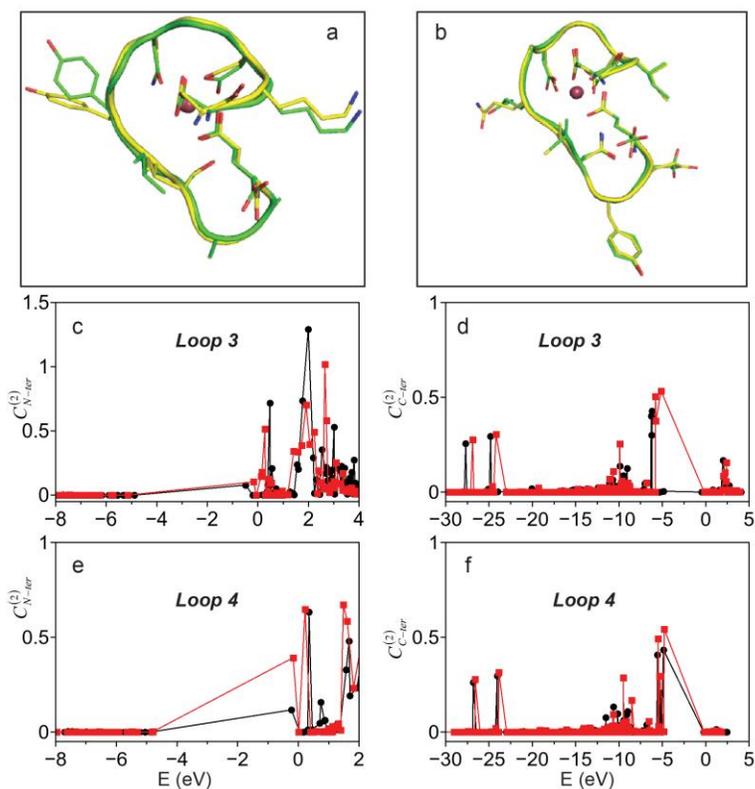
**Fig 6.12. (a) HOMO<sup>-</sup> and (b) LUMO<sup>+</sup> of Zwitterionic capped Ca<sup>2+</sup>-loop 3 complex. (c) HOMO<sup>-</sup> and (d) LUMO<sup>+</sup> of Zwitterionic capped Ca<sup>2+</sup>-loop 4 complex.**

*Localized capping states:* Our calculations on the Zwitterionic state of the isolated CaM loops indicate dominant contributions of terminal capping localized around the HOMO-LUMO

levels. Such behavior is quite unlike the earlier study with neutral terminal capping<sup>12</sup>, where the HOMO-LUMO levels are free from capping effects. This behavior of charged terminals is similar to the effect of impurities on ground state energy spectrum. Within the decay profile of  $C_\gamma O1O2^-$  capping (for instance,  $\xi_{C-ter}^{(L1)} = 0.4$  eV), there are a number of pure capping levels which are isolated and show no contribution from the loop atoms. The maximum of  $C_\gamma O1O2^-$  capping contributions is separated from the adjacent pure capping level by 0.1 eV, which corresponds to approximately 12000 nm. HOMO<sup>-</sup> energy level corresponds to 3000 nm with respect to the maximum capping level. Although both wavelengths belong to infra-red (IR) range, the huge difference in wavelength should be measurable by IR spectroscopic techniques. Since, the  $C_\gamma O1O2^-$  capping contributions are not mixed with other residue contributions; they may act like sensors in the IR range. On the other hand, the  $NH_3^+$  capping levels (for instance,  $\xi_{N-ter}^{(L1)} = 0.03$  eV) are not so well separated from the loop atoms.

*Robustness to choice of conformation:* We also consider a different conformation of  $Ca^{2+}$  bound loop 3 and loop 4 generated from MD simulation. The simulated conformation shows root mean squared deviation (RMSD) of 0.3 Å and 0.2 Å for  $Ca^{2+}$  bound loop 3 (Fig 6.13 (a)) and loop 4 (Fig 6.13 (b)) complexes with respect to their force-field optimized geometries. We perform single point energy calculation on the simulated conformation. The contributions of N-terminal capping,  $C_{N-ter}^{(2)}$  (Fig 6.13 (c)) and C-terminal capping,  $C_{C-ter}^{(2)}$  (Fig 6.13 (d)) of the simulated conformation of  $Ca^{2+}$ -loop 3 complex is similar to that of the optimized conformation. The capping contributions of the  $Ca^{2+}$ -loop 4 complex in both simulated and optimized conformations are also similar (Fig 6.13 (e) and Fig 6.13 (f)). We define the HOMO<sup>-</sup> and the LUMO<sup>+</sup> levels as earlier. The residue contributions to the HOMO<sup>-</sup> and the LUMO<sup>+</sup> levels in the

MD generated conformation also reflect qualitative similarity. Y99 contributes predominantly to both these levels in  $\text{Ca}^{2+}$ -loop 3 along with minor contribution from I100 and S101 to the  $\text{LUMO}^+$  level. In the simulated conformation of  $\text{Ca}^{2+}$  bound loop 4, the  $\text{HOMO}^-$  and the  $\text{LUMO}^+$  levels find predominant contribution from Y138 along with minor contributions from E139 to  $\text{HOMO}^-$  and N137 to  $\text{LUMO}^+$  level.



**Fig 6.13.** Force-field optimized (yellow) and MD simulated (green) conformations of (a)  $\text{Ca}^{2+}$ -loop 3 and (b)  $\text{Ca}^{2+}$ -loop 4. The contribution of terminal capping, (c)  $C_{N-ter}^{(2)}$  and (d)  $C_{C-ter}^{(2)}$  of  $\text{Ca}^{2+}$ -loop 3 complex for optimized (red) and simulated (black) conformations. The contribution of terminal capping, (e)  $C_{N-ter}^{(2)}$  and (f)  $C_{C-ter}^{(2)}$  of  $\text{Ca}^{2+}$ -loop 4 complex for optimized (red) and simulated (black) conformations.

## 6.4 Conclusions

In summary, the vacuum QC calculations indicate that truncation and choice of terminal capping affect the electronic energy states non-trivially. However, the energetically low lying

coordination state MOs remain robust to these system specific details and adequately describe the role of both ligand and non-ligand atoms. These energy levels can provide insight to understand coordination geometry and stability of metalloproteins in general.

The implicit solvent QC calculations with neutral capping indicate a common mechanism of  $\text{Ca}^{2+}$  ion coordination and stability for EF-hand proteins. These calculations yield the presence of nucleophilic site at Y99 of  $\text{Ca}^{2+}$ -loop 3 complex in agreement to experiments. Our calculations can be generalized to understand protein functionality from HOMO-LUMO levels.

The behavior of charged terminals corresponding to Zwitterionic state is similar to the effect of impurities on ground state energy spectrum and may have significance in device applications in the nanometer length scale. We also find that the terminal charges do not interfere with the nucleophilic character of Y99. Finally, we show a novel method of interpreting the electronic structure with charged terminal capping on an overall neutral polypeptide which forms a more realistic model of biologically active molecule like protein.

## Appendix I

### Density functional theory formalism<sup>203</sup>

The time independent Schrödinger equation of an electron of mass ‘ $m$ ’ in presence of time independent external potential,  $\mathbf{V}$  is given by:  $\mathbf{H}\Psi(r) = E\Psi(r)$ , where the Hamiltonian

operator,  $\mathbf{H} = -\frac{\hbar^2}{2m}\nabla^2 + \mathbf{V}$  acting on the electronic wave function,  $\Psi(r)$  returns the value of total

energy,  $E$  of the electron. An essential part of solving the Schrödinger equation is the Born–Oppenheimer approximation, where the coupling between the nuclei and electronic motion is neglected. This allows the electronic part to be solved with the nuclear positions as parameters,

and the resulting potential energy surface (PES) forms the basis for solving the nuclear motion. The major computational effort is in solving the electronic Schrödinger equation for a given set of nuclear coordinates.

The Schrödinger equation is exactly solvable for one-electron systems. The dynamics of a many-electron system is very complex, and consequently requires elaborate computational methods. Within the Hartree-Fock (HF) model, each electron is described by an orbital, and the total wave function of a many-electron system is given as a product of orbitals conveniently achieved through a Slater determinant. Since the other electrons are described by their respective orbitals, the HF equations depend on their own solutions, and must therefore be solved iteratively. HF theory only accounts for the average electron–electron interactions, and consequently neglects the correlation between electrons.

Density Functional Theory (DFT) based on Kohn–Sham equations<sup>49</sup> can be considered as an improvement on HF theory, where the many-body effect of electron correlation is modeled by a function of the electron density. The Hohenberg-Kohn theorem<sup>204</sup> states that:

- (1) The ground state energy,  $\varepsilon$  of a system of electrons in an external potential, is a unique functional of electron density,  $\rho(r)$ , i.e.  $\varepsilon = \varepsilon[\rho(r)]$ .
- (2) The functional,  $\varepsilon = \varepsilon[\rho(r)]$  has its minimum value when  $\rho(r)$  is the ground state electron density,  $\rho_0(r)$ .

Kohn and Sham wrote the electron density,  $\rho(r)$  for a system of  $N$  electrons as the sum of square moduli of a set of one-electron orthonormal orbitals:  $\rho(r) = \sum_{i=1}^N |\psi_i(r)|^2$  subjected to the restriction that  $\int \rho(r) dr = N$ . The energy functional is written as:

$$E^{KS} = 2 \sum_i^{occ} \int \psi_i^*(\vec{r}) \left( \frac{-\nabla^2}{2} \right) \psi_i(\vec{r}) d\vec{r} + \int V(\vec{r}) \rho(\vec{r}) d\vec{r} + \frac{1}{2} \int \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + E_{xc}[\rho(\vec{r})]$$

The first term in the right-hand side of this expression is the kinetic energy of the non-interacting electrons. The second term corresponds to the interaction of the electrons with the nuclear charges and  $V(\vec{r})$  is the potential as a result of the nuclei. In case only valence electrons are explicitly considered in the calculation,  $V(\vec{r})$  would be a pseudo-potential. The third term corresponds to the classical Coulomb interaction of a density distribution  $\rho$ . The fourth term is a functional of the density that accounts for the remaining contributions to the electron–electron interaction.

## ***CHAPTER 7***

# **Ab-initio quantum chemical studies on role of hydration water on protein functionality**

## **7.1 Introduction**

Bio-molecular functions take place in aqueous medium in cellular environment. The hydrogen bond network gets modified among the water molecules in the vicinity of bio-molecules. Such water molecules form hydration layer around bio-molecules and are also known as biological water<sup>205</sup>. These water molecules participate in hydrogen bonding via charge sharing with solvent exposed hydrophilic charged and polar moieties. Numerous experimental techniques<sup>206</sup> provide detailed insight into important roles of water molecules in cellular processes, like enzymatic reactions<sup>207</sup>, bio-molecular binding<sup>208</sup>, molecular recognition<sup>209</sup> involving nucleic acids and proteins. Due to large number of atoms, the theoretical studies on bio-molecules are either primarily classical force field based atomistic simulations<sup>208, 210</sup> or quantum mechanical calculations<sup>207</sup>, truncating the bio-molecules to the region of interest in water treated as a continuum medium with dielectric constant.<sup>195</sup>

Despite providing insights to bio-molecular functions, none of these approaches adequately capture quantum mechanical nature of charge sharing in hydrogen bonds. One needs to carry out full quantum mechanical calculations on bio-molecules with explicit water to point out the importance of hydrogen bonds through electron correlation effects. Such calculations have been attempted only for pure water<sup>211</sup> and in the presence of small molecules<sup>212</sup>. A hybrid approach, called quantum mechanical/molecular mechanical (QM/MM) calculation is very

popular for bio-molecules to account for the solvent polarization effects.<sup>213-215</sup> In such calculations, the region of interest is treated at QM level, while surrounding solvent and/or protein is treated with classical force-field simulation. QM/MM calculations are routinely performed to study enzymatic catalysis<sup>215</sup>, photo-chemical reactions<sup>216</sup> and drug binding<sup>213</sup> to proteins. However, there are issues regarding the choice of partitioning the whole system into QM and MM regions. A coupling interaction between the two regions also needs to be considered with special precaution at the QM/MM boundary. Both these issues are reported to affect the outcome of QM/MM calculations<sup>213</sup>.

There is no report as yet if the inclusion of relevant water molecules explicitly around a bio-molecular fragment for a full QM calculation may provide adequate description of effects of solvent polarization. Since hydration water molecules form hydrogen bond with hydrophilic groups, it would be tempting to consider only these molecules explicitly. However, as the hydrogen bonds form an extensive network in water, it is not a-priori clear if the inclusion of water molecules in the hydration layer explicitly is sufficient to describe the functionalities of solvent exposed groups. Motivated by this, we carry out quantum calculations on bio-molecules in explicit hydration water molecules to assess how well they capture bio-molecular functions.

Let us consider an example to illustrate the limitation of quantum calculations with continuum models to understand protein functions. We mentioned in earlier chapter that the polar residues such as serine, threonine and tyrosine of  $\text{Ca}^{2+}$  bound CaM<sup>61, 181</sup> can undergo phosphorylation<sup>193</sup> in order to modulate cell signaling<sup>217</sup>. Mass spectrometric analysis have identified tyrosines, Y99 (loop 3) and Y138 (loop 4) of CaM as active nucleophilic sites for phosphorylation<sup>3 [9b]</sup> both in-vivo and in-vitro. Our earlier quantum calculations with continuum medium on isolated  $\text{Ca}^{2+}$  bound CaM loops show that Y99 of loop3 is a nucleophilic site.

However, the nucleophilicity of Y138 has not been explained from these studies in contrast to the experimental observations. It is not clear if the discrepancy between experiments and quantum calculations is due to inherent limitation of continuum models to describe hydrogen bonds. This leads us to consider explicit water molecules in the hydration layer around the CaM loops in quantum calculations.

With this backdrop, we perform quantum calculations on the four  $\text{Ca}^{2+}$  ion free (apo) and bound (holo) EF-hand loops of CaM with neutral capping and water molecules within a cutoff distance of 5 Å from the loop surface representing the hydration water molecules.<sup>70</sup> The calculations are ab-initio in nature starting from the Hamiltonian of a mutually interacting many electron systems in the potential generated by the nuclei, where the ground state energy has been obtained variationally in terms of the electron density.<sup>218-220</sup> Instead of atom-centred localized basis functions<sup>199</sup> as used in earlier chapter with Gaussian03, these DFT (see Appendix I, chapter 6) calculations are performed in the VASP with plane wave (PW)<sup>219, 221</sup> basis and projector augmented-wave (PAW)<sup>218, 220</sup> potentials (see Appendix I). Such DFT-PW calculations have been used to describe systems, like bio-molecules in vacuum<sup>207, 222</sup> and water clusters<sup>223</sup>, where non-bonded interactions are important. We report that the explicit consideration of biological water molecules along with the loop atoms forms an optimal model to capture nucleophilic character of both Y99 and Y138 as in experiments. Moreover, these water molecules contribute significantly to coordination stability. The explicit modeling of biological water around CaM loops takes care of the local interactions of solvent exposed polar tyrosine residues through hydrogen bond formation.

## 7.2 Methods

We use the force-field optimized  $\text{Ca}^{2+}$  free and bound CaM loops along with water molecules within a cutoff distance of 5 Å from loop surface representing hydration water molecules. The number of water molecules is kept fixed at 50 for all the loops in both apo and holo states. The DFT calculations are performed in VASP using the plane wave<sup>219, 221</sup> basis with cutoff energy of 400 eV and periodic box size of 30 X 30 X 30 Å<sup>3</sup>. The basis set consists of 2s, 2p states of nitrogen, (N), carbon, (C), oxygen, (O), 1s state of hydrogen, (H) and 3s, 3p, 3d, 4s states of calcium, (Ca) atoms. We have used PAW<sup>218, 220</sup> potential (see Appendix I) with PBE<sup>224</sup> as exchange-correlational functional and a gamma centred  $k$  mesh of 1 X 1 X 1.

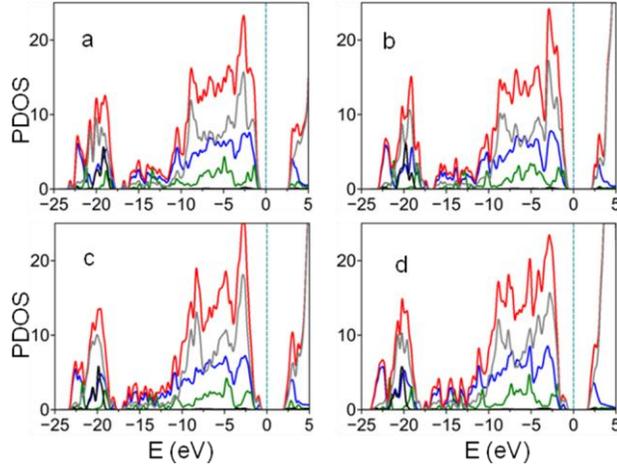
We perform ionic relaxation for all the loops followed by self consistent calculation for relaxing the electronic degrees of freedom and finally the density of states calculation (see Appendix II) is performed. The partial charges are calculated from Bader charge analysis tool.<sup>225</sup> The atom condensed Fukui index for nucleophilicity is calculated as in earlier chapter.

## 7.3 Results

The truncated loops along with biological water are optimized in vacuum and density of states calculations are performed on the optimized geometries. The optimized EF-hand loops of CaM show no significant root-mean-square deviation (RMSD~0.5 Å) from the PDB geometries. The coordination distances of ligand oxygen atoms from Ca<sup>2+</sup> ion in relaxed geometry are also similar compared to that in the PDB structure of 1CLL. We perform explicit solvent single point energy calculation with the relaxed geometries of these loops in both Ca<sup>2+</sup> ion free and bound states and extract density of states representing the number of electronic states per unit of energy (in eV).

### A. Partial density of states (PDOS) of holo-CaM loops

We denote the partial density of states (PDOS) contributed by the  $k^{th}$  atom of protein at energy level  $E$ , as  $\rho_k^{(\alpha)}(E)$  and that by terminal capping atoms as  $\rho_{cap}^{(\alpha)}(E)$  for the  $\alpha^{th}$  loop. Hence, the PDOS of protein loops are given by,  $\rho_{pro}^{(\alpha)}(E) = \sum_k \rho_k^{(\alpha)}(E)$ , which also includes ligand  $O$  atoms,  $\rho_{O-i}^{(\alpha)}(E)$ ,  $i (= I, III, V, VII, XII)$  representing loop position and  $\rho_{O-W}^{(\alpha)}(E)$  of coordinating water molecule ( $W$ ) within the loop. Similarly, we denote PDOS of Ca<sup>2+</sup> ion by  $\rho_{Ca^{2+}}^{(\alpha)}(E)$  and that of biological water by  $\rho_{hyd}^{(\alpha)}(E)$ . The total density of states is given by,  $\rho_{tot}^{(\alpha)}(E) = \rho_{cap}^{(\alpha)}(E) + \rho_{pro}^{(\alpha)}(E) + \rho_{Ca^{2+}}^{(\alpha)}(E) + \rho_{hyd}^{(\alpha)}(E)$ . Fig 7.1 shows  $\rho_{tot}^{(\alpha)}(E)$  (red),  $\rho_{pro}^{(\alpha)}(E)$  (blue),  $\rho_{Ca^{2+}}^{(\alpha)}(E)$  (black),  $\rho_{cap}^{(\alpha)}(E)$  (green) and  $\rho_{hyd}^{(\alpha)}(E)$  (gray) across all the energy levels in ground state relative to Fermi level,  $E_F$  (dotted line) for all the loops. PDOS of capping atoms,  $\rho_{cap}^{(\alpha)}(E)$  is



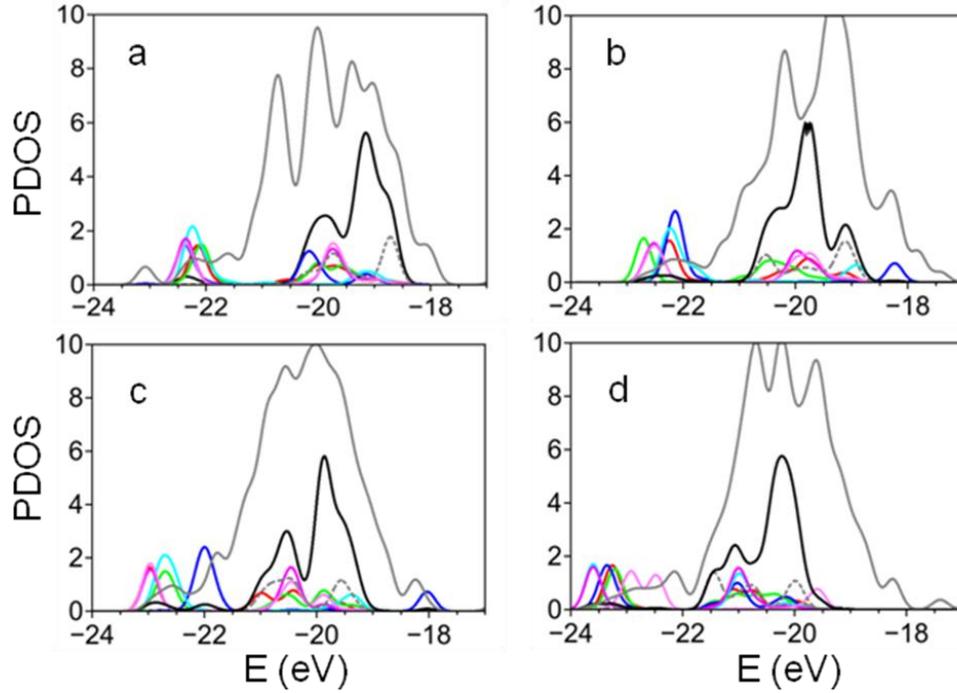
**Fig 7.1.** PDOS:  $\rho_{tot}^{(\alpha)}(E)$  (red),  $\rho_{pro}^{(\alpha)}(E)$  (blue),  $\rho_{Ca^{2+}}^{(\alpha)}(E)$  (black),  $\rho_{cap}^{(\alpha)}(E)$  (green) and  $\rho_{hyd}^{(\alpha)}(E)$  (gray) across the ground state energy range for (a) holo loop 1, (b) holo loop 2, (c) holo loop 3 and (d) holo loop 4.  $\rho_{tot}^{(\alpha)}(E)$ ,  $\rho_{pro}^{(\alpha)}(E)$  and  $\rho_{hyd}^{(\alpha)}(E)$  are scaled down by a factor of 5. The Fermi level,  $E_F$  is shown in dotted line.

found to be negligible compared to  $\rho_{pro}^{(\alpha)}(E)$  across the ground state energy spectra. Simultaneous contribution from protein ( $\rho_{pro}^{(\alpha)}(E)$ ) and bound  $Ca^{2+}$  ion ( $\rho_{Ca^{2+}}^{(\alpha)}(E)$ ) occurs between -23 eV and -18 eV. We define these as coordination energy levels, which have significant contribution from biological water ( $\rho_{hyd}^{(\alpha)}(E)$ ) as well. The contribution of  $Ca^{2+}$  ion falls off beyond this; the following energy range from -16 eV till HOMO has contributions from both  $\rho_{pro}^{(\alpha)}(E)$  and  $\rho_{hyd}^{(\alpha)}(E)$ . This picture is similar for all the four loops. The HOMO level of the CaM loops is defined as the energy level just below  $E_F$  where  $\rho_{tot}^{(\alpha)}(E)$  is maximum. The LUMO is defined as the energy level just above  $E_F$  where  $\rho_{tot}^{(\alpha)}(E)$  has a maximum. The energy gap between HOMO and LUMO levels of  $Ca^{2+}$  ion bound loops,  $\Delta_{holo}^{(\alpha)}$  and the corresponding  $E_F$  are summarized in Table 7.1.

**Table 7.1.** Fermi level,  $E_F$  and HOMO-LUMO gap,  $\Delta_{holo}^{(\alpha)}$ ,  $\Delta_{apo}^{(\alpha)}$  for holo and apo-CaM loops.

$\alpha$	$E_F$ (holo)	$\Delta_{holo}^{(\alpha)}$	$E_F$ (apo)	$\Delta_{apo}^{(\alpha)}$
Loop 1	-4.55	3.5	-5.13	2.5
Loop 2	-3.82	2.75	-3.34	2.25
Loop 3	-4.09	2.5	-3.77	3.0
Loop 4	-3.01	2.25	-3.14	2.75

*Coordination levels:* The coordination energy levels of holo-CaM loops are characterized by simultaneous contributions from ligand oxygen ( $O$ ) atoms and  $\text{Ca}^{2+}$  ion. Fig 7.2 shows  $\rho_{O-i}^{(\alpha)}(E)$ ,  $\rho_{O-w}^{(\alpha)}(E)$ ,  $\rho_{\text{Ca}^{2+}}^{(\alpha)}(E)$  and  $\rho_{\text{hyd}}^{(\alpha)}(E)$  across the coordination energy range.



**Fig 7.2. PDOS:**  $\rho_{O-i}^{(\alpha)}(E)$  of ligand oxygen atoms from loop positions,  $i = \text{I}$  (red),  $\text{III}$  (green),  $\text{V}$  (blue),  $\text{VII}$  (cyan), bi-dentate  $\text{XII}$  (magenta),  $\rho_{O-w}^{(\alpha)}(E)$  of coordinating water (gray dotted line),  $\rho_{\text{Ca}^{2+}}^{(\alpha)}(E)$  of  $\text{Ca}^{2+}$  ion (black) along with  $\rho_{\text{hyd}}^{(\alpha)}(E)$  (gray solid line) in the coordination energy range for (a) holo loop 1, (b) holo loop 2, (c) holo loop 3 and (d) holo loop 4.  $\rho_{\text{hyd}}^{(\alpha)}(E)$  is scaled down by factor of 5.

In  $\text{Ca}^{2+}$  bound loop 1,  $\rho_{\text{Ca}^{2+}}^{(1)}(E)$  (Fig 7.2 (a)), shows maxima between -20.5 eV and -18 eV, while a small peak is observed around -22.5 eV. The ligand contributions  $\rho_{O-i}^{(1)}(E)$  also show presence of two peaks: A narrow peak around -22.5 eV and a broad one between -20.5 eV and -18 eV, overlapping with  $\rho_{\text{Ca}^{2+}}^{(1)}(E)$ . The ligand oxygen atoms of bidentate glutamate, E31 ( $\rho_{O-xII}^{(1)}(E)$ ) and aspartate, D24 ( $\rho_{O-v}^{(1)}(E)$ ) contribute to the energetically lowest coordination

level at -22.5 eV followed by that of threonine, T26 ( $\rho_{O-VII}^{(1)}(E)$ ), D20 ( $\rho_{O-I}^{(1)}(E)$ ) and D22 ( $\rho_{O-III}^{(1)}(E)$ ). Although  $\rho_{Ca^{2+}}^{(1)}(E)$  and  $\rho_{O-i}^{(1)}(E)$  are small between -21.5 eV and -20.5 eV,  $\rho_{hyd}^{(1)}(E)$  is significant in this range.

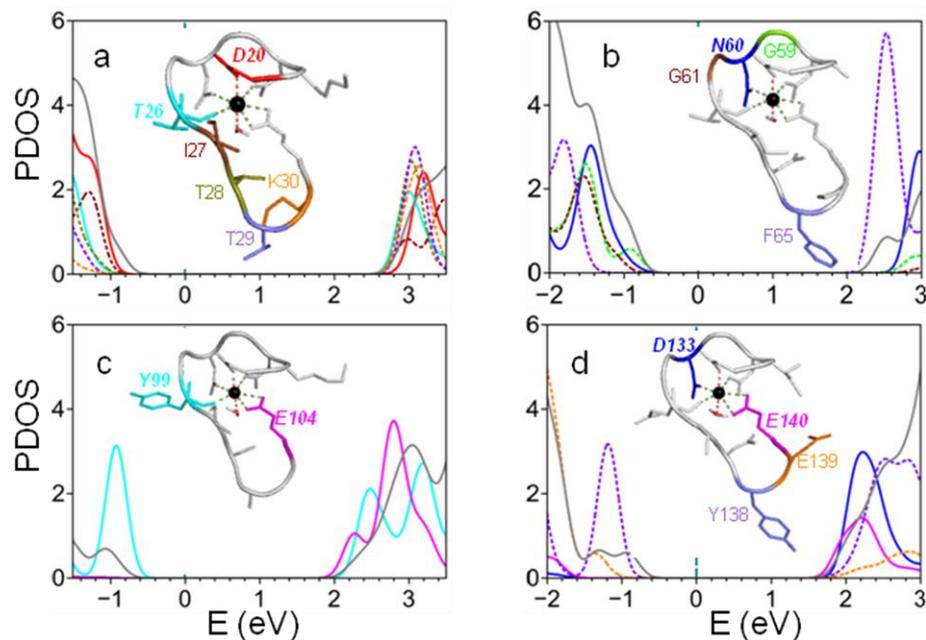
In  $Ca^{2+}$ -loop 2 complex (Fig 7.2 (b)),  $\rho_{Ca^{2+}}^{(2)}(E)$  is maximum between -21 eV and -18.5 eV with a small peak about -22.5 eV. The ligand contributions  $\rho_{O-i}^{(2)}(E)$  are maxima at two energy ranges: One between -23 eV and -21.5 eV and the other between -21 eV and -18.5 eV. The gap between these two maxima finds contribution from  $\rho_{hyd}^{(2)}(E)$ . The energetically lowest coordination level of  $Ca^{2+}$ -loop 2 complex at -23 eV is characterized by D58 ( $\rho_{O-III}^{(2)}(E)$ ), followed by contributions from E67 ( $\rho_{O-XII}^{(2)}(E)$ ), T62 ( $\rho_{O-VII}^{(2)}(E)$ ), D56 ( $\rho_{O-I}^{(2)}(E)$ ) and asparagine, N60 ( $\rho_{O-V}^{(2)}(E)$ ).

Fig 7.2 (c) shows PDOS contribution in coordination energy range of  $Ca^{2+}$  bound loop 3. PDOS of  $Ca^{2+}$  ion,  $\rho_{Ca^{2+}}^{(3)}(E)$  is maximum between -21.5 eV and -19 eV and also shows presence of two smaller peaks between -23.5 eV and -21.5 eV. The PDOS of ligand atoms,  $\rho_{O-i}^{(3)}(E)$  is clustered into two maxima, the intermediate gap finding significant contribution from  $\rho_{hyd}^{(3)}(E)$ . The ligand *O* atoms from bidentate E104 ( $\rho_{O-XII}^{(3)}(E)$ ) and D93 ( $\rho_{O-I}^{(3)}(E)$ ) of  $Ca^{2+}$  bound loop 3 participate in its lowest coordination level ( $\sim$  -23.5 eV) followed by D95 ( $\rho_{O-III}^{(3)}(E)$ ) and tyrosine, Y99 ( $\rho_{O-VII}^{(3)}(E)$ ). The contribution of coordinating *O* of N97 ( $\rho_{O-V}^{(3)}(E)$ ) is at -22 eV.

The PDOS of  $Ca^{2+}$ -loop 4 at the coordination energy range is shown in Fig 7.2 (d). The contribution of  $Ca^{2+}$  ion,  $\rho_{Ca^{2+}}^{(4)}(E)$  is dominant between -22 eV and -19.5 eV with a flat peak between -24 eV and -23 eV. The ligand contributions also show bimodal nature at similar

energies with  $\rho_{hyd}^{(4)}(E)$  contributing to the intermediate gap. The lowest coordination level ( $\sim -24$  eV) of  $Ca^{2+}$ -loop 4 is governed by bidentate E140 ( $\rho_{O-XII}^{(4)}(E)$ ) and glutamine, Q135 ( $\rho_{O-VII}^{(4)}(E)$ ), followed by contributions arising from the ligand atoms of D129 ( $\rho_{O-I}^{(4)}(E)$ ), D131 ( $\rho_{O-III}^{(4)}(E)$ ) and D133 ( $\rho_{O-V}^{(4)}(E)$ ).

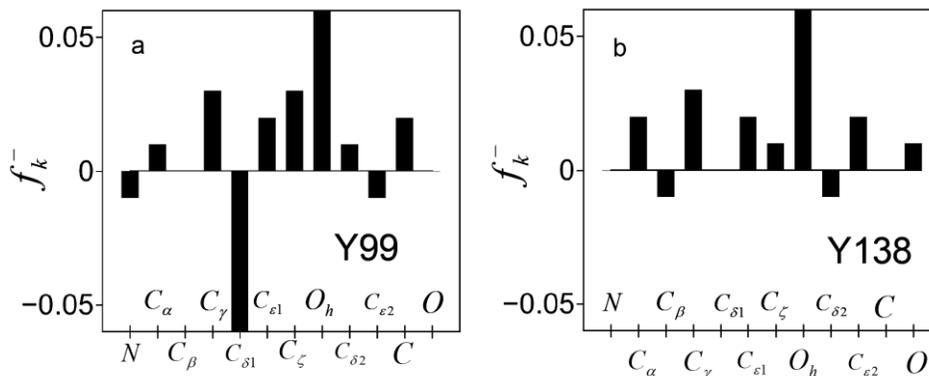
*HOMO-LUMO levels:* The PDOS of different residues along with that of biological water contributing to HOMO-LUMO levels of holo-CaM loops are shown in Fig 7.3. The insets show corresponding loop structure with residues labelled. HOMO of  $Ca^{2+}$  bound loop 1 is having PDOS delocalized over acidic aspartate, D20 and hydrophobic isoleucine, I27, while PDOS at LUMO is delocalized on polar threonines like T26, T28, T29 and basic lysine, K30 (Fig 7.3 (a)). Similar delocalized PDOS at HOMO is observed in holo-loop 2 (Fig 7.3 (b)) over polar residues namely, glycines, G59, G61 and asparagines, N60. In contrast, the PDOS at LUMO is localized on hydrophobic phenylalanine, F65 (Fig 7.3 (b)). HOMO of  $Ca^{2+}$ -loop 3 complex (Fig 7.3 (c)) also exhibits localized PDOS on polar tyrosine, Y99 but at LUMO it is delocalized over both acidic glutamate, E104 and Y99. The picture is very similar in  $Ca^{2+}$  bound loop 4 (Fig 7.3 (d)), where HOMO density of states is localized on polar Y138 with albeit small contribution from acidic E139. On other hand, the PDOS at LUMO is governed by acidic residues D133 and E140.



**Fig 7.3.** PDOS of hydration layer,  $\rho_{hyd}^{(\alpha)}(E)$  (gray solid line) and loop residues contributing to HOMO-LUMO levels of (a) holo loop 1: D20 (red), T26 (cyan solid line), I27 (brown dotted line), T28 (olive dotted line), T29 (violet dotted line) and K30 (orange dotted line). (b) holo loop 2: G59 (green dotted line), N60 (blue solid line), G61 (brown dotted line) and F65 (violet dotted line). (c) holo loop 3: Y99 (cyan) and E104 (magenta). (d) holo loop 4: D133 (blue), Y138 (violet dotted line), E139 (orange dotted line) and E140 (magenta).  $\rho_{hyd}^{(\alpha)}(E)$  is reduced by factor of 10. Insets show loop structure with contributing residues colour coded.

Fig 7.4 (a) shows nucleophilicity index,  $f_k^-$  for Y99 atoms. The phenyl ring hydroxyl oxygen,  $O_h$  shows maximum nucleophilicity ( $\sim 0.06$ ) followed by  $C_\gamma$  and  $C_\xi$  atoms of the ring. Similarly, nucleophilicity index,  $f_k^-$  for atoms of Y138 (Fig 7.4 (b)) is maximum for its  $O_h$  ( $\sim 0.06$ ) followed by ring carbon atom,  $C_\gamma$ . The negative nucleophilicity indices observed for certain atoms are not unusual. Often the electron density in a part of the bio-molecule can increase even though the total number of electrons decreases. This has been observed in case of

redox induced electron re-arrangement, where a functional group may get reduced although the entire molecule is oxidized.<sup>226</sup>



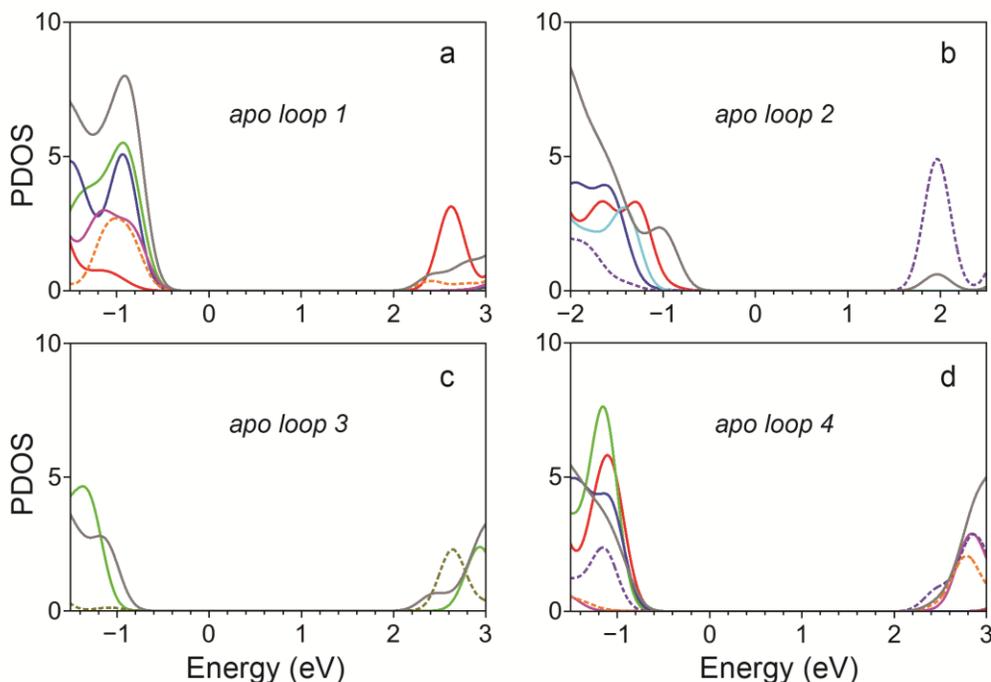
**Fig 7.4.** Atomic nucleophilicity index,  $f_k^-$  of (a) Y99 in holo loop 3 and (b) Y138 in holo loop 4.

## B. PDOS of apo-CaM

The Fermi level,  $E_F$  and the HOMO-LUMO gap,  $\Delta_{apo}^{(\alpha)}$  for different loops are summarized in Table 7.1. The HOMO-LUMO gap widens in the N-domain loops upon  $\text{Ca}^{2+}$  ion binding:  $\Delta_{holo}^{(1)}$  increases by  $\sim 1$  eV and  $\Delta_{holo}^{(2)}$  by 0.5 eV compared to respective apo states. This trend is reversed at the C-domain loops. Both  $\Delta_{holo}^{(3)}$  and  $\Delta_{holo}^{(4)}$  decrease by 0.5 eV compared to  $\Delta_{apo}^{(3)}$  and  $\Delta_{apo}^{(4)}$  respectively. These observations suggest the possibility of tuning HOMO-LUMO gap upon  $\text{Ca}^{2+}$  ion binding.

The residue contribution to HOMO-LUMO levels of apo-CaM loops are shown in Fig 7.5. HOMO of apo-loop 1 (Fig 7.5 (a)) is characterized by contributions from acidic aspartates, D22, D24, E31 and basic lysine, K30, while LUMO is dominated by D20. In apo-loop 2, HOMO finds contribution from acidic D56, polar N60 and T62, while LUMO shows localized PDOS on F65, a hydrophobic residue with phenyl ring (Fig 7.5 (b)). In apo-loop 3 HOMO is dominated by acidic D95 while the LUMO-PDOS is contributed by polar S101 (Fig 7.5 (c)). The PDOS at HOMO of apo-loop 4 is contributed by acidic residues D129, D131 and D133, while that of

LUMO by both polar Y138 and other acidic residues, E139 and E140 (Fig 7.5 (d)). Thus the apo loops of CaM indicate that HOMO is predominantly governed by either acidic or polar residues which are responsible for  $\text{Ca}^{2+}$  ion coordination in holo-CaM. However, the tyrosines, Y99 and Y138 do not contribute to HOMO of apo loops indicating that phosphorylation at such sites is possible only in presence of  $\text{Ca}^{2+}$  ion.

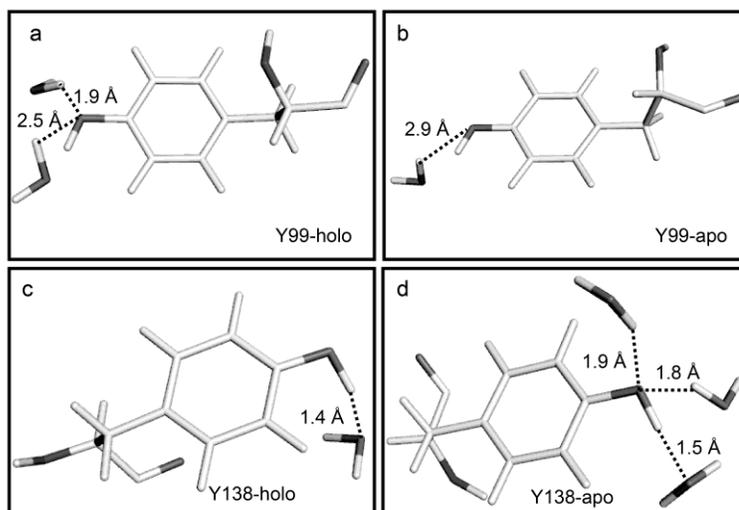


**Fig 7.5.** PDOS of hydration layer,  $\rho_{hyd}^{(\alpha)}(E)$  (gray solid line) and loop residues contributing to HOMO-LUMO levels of (a) apo-L1: D20 (red), D22 (green), D24 (blue), K30 (orange dotted) and E31 (magenta). (b) apo-L2: D56 (red), N60 (blue), T62 (cyan) and F65 (violet-dotted). (c) apo-L3: D95 (green) and S101 (olive dotted). (d) apo-L4: D129 (red), D131 (green), D133 (blue), Y138 (violet-dotted), E139 (orange-dotted) and E140 (magenta). The  $\rho_{hyd}^{(\alpha)}(E)$  is scaled down by a factor of 10.

### C. Hydrogen bonding network

In order to understand the differences in behavior of tyrosines in presence and absence of  $\text{Ca}^{2+}$  ion, we consider the local interactions of solvent exposed phenolic  $O_h$  atom. Y99- $O_h$  is hydrogen bonded to two biological water molecules in holo loop 3 (Fig 7.6 (a)) and one water

molecule in apo state (Fig 7.6 (b)). In both cases, the  $O_h$  atom acts as an acceptor. On other hand, Y138- $O_h$  acts as hydrogen bond donor to a biological water molecule in holo loop 4 (Fig 7.6 (c)). In absence of  $Ca^{2+}$  ion, three water molecules remain hydrogen bonded to Y138- $O_h$  (Fig 7.6 (d)), where the  $O_h$  atom serves as hydrogen bond acceptor for two water molecules and as hydrogen bond donor to another water molecule.



**Fig 7.6. Hydrogen bonding pattern between biological water molecules and phenolic  $O_h$  atom of (a) Y99 in holo state, (b) Y99 in apo state, (c) Y138 in holo state and (d) Y138 in apo state.**

The hydrogen bond strength depends on electro-negativity difference between donor and hydrogen atoms. To this end, we have computed atomic ( $i$ ) partial charges ( $\partial q_i$ ) in apo and holo states. The partial charge of Y99- $O_h$  atoms,  $\partial q_{O_h}$  indicates larger accumulation of charges in holo ( $-1.91 e$ ) state than in apo ( $-1.88 e$ ) state. Similarly,  $\partial q_{O_h}$  is larger for Y138 in holo ( $-2.03 e$ ) case compared to apo ( $-1.87 e$ ). In contrast to apo state, the  $O_h$  atoms of both Y99 and Y138 in presence of  $Ca^{2+}$  ion represent a favourable nucleophilic site owing to larger charge density. The surrounding hydrogen bonded biological water stabilize the accumulated charge density at  $O_h$  atoms and favours its nucleophilic character in holo state.

We calculate the local interaction energies considering dipole-dipole interactions between the phenol hydroxyl group ( $O_hH$ ) of tyrosines and its hydrogen bonded biological water molecules. We calculate the dipole moments ( $\mu_{O_hH}$ ) of the  $O_hH$  group of Y99 and Y138 in both apo and holo CaM with bound  $Ca^{2+}$  ion as origin. The dipole moment,  $\mu_{O_hH}$  in presence of  $Ca^{2+}$  ion is substantially large ( $\sim 150$  D) compared to that in apo ( $\sim 90$  D) CaM. The dipole-dipole interaction energy between tyrosine hydroxyl group,  $O_hH$  and its hydrogen bonded water molecules are also calculated in both apo and holo states. The interaction energy is  $29.8 \times 10^{-19} J$  for apo Y99 compared to  $-27.1 \times 10^{-19} J$  in holo case. Similar positive value is observed for apo Y138 ( $19.3 \times 10^{-19} J$ ) in contrast to that of holo Y138 ( $-0.2 \times 10^{-19} J$ ). Thus the favourable dipole-dipole interactions in holo case stabilize the hydrogen bond network with biological water.

#### **D. Comparison with implicit solvent QC calculations**

Our earlier implicit solvent PCM QC calculation<sup>68</sup> has shown that the HOMO of N-domain loops 1 and 2 find contribution from acidic aspartates, whereas the LUMO are characterized by polar residues in  $Ca^{2+}$ -loop 1 and hydrophobic residue in  $Ca^{2+}$ -loop 2. The HOMO and LUMO levels of holo loop 3 in the C-domain are found to exhibit localized electron density on phenyl ring of Y99, which acts as a potential nucleophile. The HOMO-LUMO picture of loop 1, 2 and 3 remains similar with explicit solvation. However, there is qualitative difference in loop 4 in presence of explicit biological water molecules. The earlier PCM study has indicated strong electron density on acidic E139 at HOMO of  $Ca^{2+}$ -loop 4 complex, while Y138 contributed to LUMO but not HOMO. In the present DFT-PW calculation, we find the PDOS at HOMO of loop 4 is localized on Y138 and that of LUMO finds contribution from acidic residues and further yields both Y99 and Y138 as potential nucleophiles in CaM in

agreement with experiments.<sup>16</sup> Since a different charge partition scheme, natural population analysis has been used in the earlier study to estimate atomic charges, it is not possible to draw a quantitative comparison of nucleophilicity indices between the implicit and explicit solvent QC calculation.

Our DFT-PW calculation indicates the role of biological water in stabilizing coordination geometry of  $\text{Ca}^{2+}$  ions in CaM loops besides that of loop residues. The coordination energy range remains similar across the loops. Our calculation thus yields a robust picture of stabilization of  $\text{Ca}^{2+}$  ion coordination geometry. This picture of stabilization is largely similar to our previous PCM results.<sup>68</sup> In contrast, the novel feature that comes out of our present calculation is that the biological water molecules also contribute to the coordination energy range with maxima around -20 eV, which corresponds to an ultraviolet (UV) frequency regime of  $10^{15}$  Hz. UV-visible spectroscopy has been used to study role of hydration layer on stability of nano-capsules<sup>227</sup>, fullerenes<sup>228</sup> and anions<sup>214</sup>. Moreover isolation of CaM loops is also feasible through grafting approach and has been used to determine site specific  $\text{Ca}^{2+}$  ion binding affinities.<sup>229</sup> Hence the role of biological water molecules in stabilizing the  $\text{Ca}^{2+}$  ion can be verified experimentally through UV-visible spectroscopy techniques on isolated CaM loops.

## 7.4 Conclusions

To conclude, we perform for the first time quantum calculations on protein with explicit hydration layer water molecules. Our calculations capture nucleophilic character of both Y99 and Y138 in holo-CaM, as suggested in in-vivo and in-vitro experiments. We predict that these sites are not nucleophilic in apo-CaM which can be verified experimentally. This difference can be ascribed to the differences in local interactions of solvent exposed polar tyrosine residues through hydrogen bonds with water molecules. Our studies indicate the necessity to go beyond

commonly employed continuum model for microscopic understanding of protein functions. However, explicit consideration of the hydration water molecules alone captures the functionalities at the cost of modest increase in computation time (~5 times for our calculations) compared to continuum calculations. We thus come up with a computationally viable microscopic model which can be applied to various bio-molecular systems, relevant in cellular processes.

## **Appendix I**

### **Pseudopotential**

Valence wavefunctions suffer from rapid oscillations near ion cores due to orthogonality requirement. This requires many Fourier components or very fine mesh to describe the wavefunctions accurately. One way of solving this problem is the use of pseudopotentials<sup>230</sup> in which the collective system of nuclei and core electrons are described by an effective and smooth potential. The Kohn-Sham equations<sup>49</sup> are then solved for the valence electrons only. The pseudopotentials are constructed such that the correct scattering potential is obtained beyond a certain radius from the core. This method reduces the number of wave functions to be calculated, since the pseudo potentials only have to be calculated and tabulated once for each type of atom, so that only calculations on the valence states are needed. A different approach is the augmented-plane-wave method (APW), in which space is divided into atom-centered augmentation spheres inside which the wave functions are taken as some atom-like partial waves, and a bonding region outside the spheres, where some envelope functions are defined. The partial waves and envelope functions are then matched at the boundaries of the spheres. The PAW<sup>218, 220</sup> approach developed by Blochl combines the above two methods.

## Appendix II

### Partial density of states

The probability of finding an electron at specific point  $(x, y, z)$  is given by  $|\psi(x, y, z)|^2$ .

The electrons at bottom of conduction band behave like free particles trapped in a box. The wave function solution for electrons in a rectangular box of volume,  $V = L_x \times L_y \times L_z$  is:

$$\psi(x, y, z) = \sin(k_x x) \sin(k_y y) \sin(k_z z) \quad (1)$$

Where  $k_x, k_y, k_z$  are the wave vectors for an electron in the  $x, y, z$  directions. Owing to the boundary conditions that the wave function goes to zero, the allowed wave vectors satisfy:

$$k_x L_x = \pi n_x, k_y L_y = \pi n_y, k_z L_z = \pi n_z, \text{ for } n_x, n_y, n_z \text{ integers.} \quad (2)$$

The  $k$ -space volume for each allowed state is  $\pi^3 / V$ , the reciprocal being state density. The number of states available for a given magnitude of  $|k|$  is calculated by constructing a sphere of radius,  $|k|$  and thickness,  $dk$ . Volume of this spherical shell in  $k$ -space is  $4\pi k^2 dk$ .

$$\text{The number of } k \text{ states within the spherical shell, } \rho(k)dk = 4\pi k^2 \frac{V}{\pi^3} dk \quad (3)$$

Since, each  $k$  state can hold two electrons of opposite sign and considering positive values of  $n_x, n_y, n_z$  to avoid multiple counting of the same quantum state, we get

$$\rho(k)dk = \frac{Vk^2}{\pi^2} dk \quad (4)$$

We re-write the expression in terms of energy following the simple relationship, where  $m^*$  is the effective mass:

$$E = \frac{\hbar^2 k^2}{2m^*} \quad (5)$$

$$k^2 = \frac{(E - E_f)2m^*}{\hbar^2}, \text{ with respect to Fermi level, } E_f \quad (6)$$

and differentiating:  $2kdk = \frac{2m^*}{\hbar^2} dE \quad (7a)$

Combining (6) and (7a):

$$dk = \frac{m^* dE}{\hbar \sqrt{2m^*(E - E_f)}} \quad (7b)$$

Plugging (6) and (7) into (4):

$$\begin{aligned} \rho(k)dk &= \frac{Vk^2}{\pi^2} \frac{m^* dE}{\hbar \sqrt{2m^*(E - E_f)}} \\ &= \frac{Vm^* [2m^*(E - E_f)]^{1/2}}{\pi^2 \hbar^3} dE \end{aligned} \quad (8)$$

Dividing by  $V$ , the number of electronic states in conduction band per unit volume over an energy range,  $dE$  is

$$\rho(E)dE = \frac{m^* [2m^*(E - E_f)]^{1/2}}{\pi^2 \hbar^3} dE \quad (9)$$

The atomic partial density of states is calculated by considering volume of a particular atom having a cut-off radius, usually defined by the Wigner-Seitz radius.

## REFERENCES

1. J. Wereszczynski and J. A. McCammon, *Q Rev Biophys*, 2012, 45, 1-25.
2. E. L. Baxter, J. A. Zuris, C. Wang, P. L. Vo, H. L. Axelrod, A. E. Cohen, M. L. Paddock, R. Nechushtai, J. N. Onuchic and P. A. Jennings, *Proc Natl Acad Sci U S A*, 2013, 110, 948-953.
3. G. Benaim and A. Villalobo, *Eur J Biochem*, 2002, 269, 3619-3631.
4. W. J. Chazin, *Acc Chem Res*, 2011, 44, 171-179.
5. L. N. Johnson and D. Barford, *Annu Rev Biophys Biomol Struct*, 1993, 22, 199-232.
6. M. S. Marlow and J. Dogan, *Nat Chem Biol*, 2010, 6, 352-358.
7. P. E. Siegbahn and M. R. Blomberg, *J Phys Chem A*, 2008, 112, 12772-12780.
8. A. D. Vogt, N. Pozzi, Z. Chen and E. Di Cera, *Biophys Chem*, 2014, 186, 13-21.
9. Z. Xia and D. R. Storm, *Nat Rev Neurosci*, 2005, 6, 267-276.
10. R. Rinaldi, A. Biasco, G. Maruccio, R. Cingolani, D. Alliata, L. Andolfi, P. Facci, F. De Rienzo, R. Di Felice and E. Molinari, *Advanced Materials*, 2002, 14, 1453-+.
11. S. Prabhulkar, H. Tian, X. Wang, J. J. Zhu and C. Z. Li, *Antioxid Redox Signal*, 2012, 17, 1796-1822.
12. T. Rakshit and R. Mukhopadhyay, *Langmuir*, 2011, 27, 9681-9686.
13. R. Vazquez-Duhalt, S. A. Aguila, A. A. Arrocha and M. Ayala, *Chemelectrochem*, 2014, 1, 496-513.
14. R. D. Brox, M. M. Lopez, H. J. Vogel and G. I. Makhatadze, *J Biol Chem*, 2001, 276, 14083-14091.
15. B. J. Canagarajah, A. Khokhlatchev, M. H. Cobb and E. J. Goldsmith, *Cell*, 1997, 90, 859-869.
16. C. Corti, E. Leclerc L'Hostis, M. Quadroni, H. Schmid, I. Durussel, J. Cox, P. Dainese Hatt, P. James and E. Carafoli, *Eur J Biochem*, 1999, 262, 790-802.
17. T. Dudev and C. Lim, *Annu Rev Biophys*, 2008, 37, 97-116.
18. J. M. Woodcock, J. Murphy, F. C. Stomski, M. C. Berndt and A. F. Lopez, *J Biol Chem*, 2003, 278, 36323-36327.
19. R. Koike, A. Kidera and M. Ota, *Protein Sci*, 2009, 18, 2060-2066.
20. L. J. Lapidus, *Mol Biosyst*, 2013, 9, 29-35.
21. A. Das, S. Sikdar, M. Ghosh and J. Chakrabarti, *Journal of Physics: Conference Series*, 2015, 638, 012013.
22. A. Das, J. Chakrabarti and M. Ghosh, *Chem Phys Lett*, 2013, 581, 91-95.
23. A. Das, J. Chakrabarti and M. Ghosh, *Biophysical Journal*, 2013, 104, 1274-1284.
24. A. Das, J. Chakrabarti and M. Ghosh, *Mol Biosyst*, 2014, 10, 437-445.
25. J. M. Bui, J. Gsponer, M. Vendruscolo and C. M. Dobson, *Biophys J*, 2009, 97, 2513-2520.
26. Y. Cote, P. Senet, P. Delarue, G. G. Maisuradze and H. A. Scheraga, *Proc Natl Acad Sci U S A*, 2012, 109, 10346-10351.
27. J. E. Fitzgerald, A. K. Jha, T. R. Sosnick and K. F. Freed, *Biochemistry*, 2007, 46, 669-682.
28. J. A. McCammon, B. R. Gelin and M. Karplus, *Nature*, 1977, 267, 585-590.
29. E. D. Chrysina, K. Brew and K. R. Acharya, *J Biol Chem*, 2000, 275, 37021-37029.
30. E. A. Permyakov and L. J. Berliner, *FEBS Lett*, 2000, 473, 269-274.

31. J. Pettersson-Kastberg, A. K. Mossberg, M. Trulsson, Y. J. Yong, S. Min, Y. Lim, J. E. O'Brien, C. Svanborg and K. H. Mok, *J Mol Biol*, 2009, 394, 994-1010.
32. M. Svensson, A. K. Mossberg, J. Pettersson, S. Linse and C. Svanborg, *Protein Sci*, 2003, 12, 2805-2814.
33. S. Tolin, G. De Franceschi, B. Spolaore, E. Frare, M. Canton, P. Polverino de Laureto and A. Fontana, *FEBS J*, 2010, 277, 163-173.
34. S. Sikdar, J. Chakrabarti and M. Ghosh, *Mol Biosyst*, 2014, 10, 3280-3289.
35. R. H. Kretsinger and C. E. Nockolds, *J Biol Chem*, 1973, 248, 3313-3326.
36. M. V. Vinogradova, D. B. Stone, G. G. Malanina, C. Karatzaferi, R. Cooke, R. A. Mendelson and R. J. Fletterick, *Proc Natl Acad Sci U S A*, 2005, 102, 5038-5043.
37. O. Herzberg and M. N. James, *Nature*, 1985, 313, 653-659.
38. M. Sundaralingam, R. Bergstrom, G. Strasburg, S. T. Rao, P. Roychowdhury, M. Greaser and B. C. Wang, *Science*, 1985, 227, 945-948.
39. R. S. Fredricksen and C. A. Swenson, *Biochemistry*, 1996, 35, 14012-14026.
40. S. Sikdar, J. Chakrabarti and M. Ghosh, *Mol Biosyst*, 2016, 12, 444-453.
41. P. Saha, C. Manna, J. Chakrabarti and M. Ghosh, *Sci Rep*, 2016, 6, 29541.
42. P. Saha, C. Manna, S. Das and M. Ghosh, *Sci Rep*, 2016, 6, 21305.
43. P Saha, S Sikdar, J Chakrabarti, M Ghosh, *Submitted* 2016.
44. W. Zheng, D. De Sancho, T. Hoppe and R. B. Best, *J Am Chem Soc*, 2015, 137, 3283-3290.
45. D. de Sancho, A. Sirur and R. B. Best, *Nat Commun*, 2014, 5, 4307.
46. B. G. Wensley, S. Batey, F. A. Bone, Z. M. Chan, N. R. Tumelty, A. Steward, L. G. Kwa, A. Borgia and J. Clarke, *Nature*, 2010, 463, 685-688.
47. T. Cellmer, E. R. Henry, J. Hofrichter and W. A. Eaton, *Proc Natl Acad Sci U S A*, 2008, 105, 18320-18325.
48. S Sikdar, M Ghosh, J Chakrabarti, *Pre-print* 2016.
49. W. Kohn, A. D. Becke and R. G. Parr, *J Phys Chem-Us*, 1996, 100, 12974-12980.
50. M. R. Blomberg and P. E. Siegbahn, *J Comput Chem*, 2006, 27, 1373-1384.
51. X. He and J. Z. Zhang, *J Chem Phys*, 2006, 124, 184703.
52. M. Lepsik and M. J. Field, *J Phys Chem B*, 2007, 111, 10012-10022.
53. A. Y. Mehandzhiyski, E. Riccardi, T. S. van Erp, H. Koch, P.-O. Åstrand, T. T. Trinh and B. A. Grimes, *The Journal of Physical Chemistry A*, 2015.
54. P. E. Siegbahn and T. Borowski, *Acc Chem Res*, 2006, 39, 729-738.
55. D. W. Zhang, X. H. Chen and J. Z. Zhang, *J Comput Chem*, 2003, 24, 1846-1852.
56. M. R. Blomberg and P. E. Siegbahn, *Biochim Biophys Acta*, 2012, 1817, 495-505.
57. M. R. A. Blomberg and P. E. M. Siegbahn, *Journal of Physical Chemistry B*, 2001, 105, 9375-9386.
58. N. Dimakis, M. J. Farooqi, E. S. Garza and G. Bunker, *J Chem Phys*, 2008, 128, 115104.
59. T. Dudev, L. Y. Chang and C. Lim, *J Am Chem Soc*, 2005, 127, 4091-4103.
60. G. Lever and D. J. Cole, *J Phys Condens Matter*, 2013, 25, 152101.

61. R. Chattopadhyaya and W. E. Meador, *J Mol Biol*, 1992, 228, 1177-1192.
62. J. L. Gifford, M. P. Walsh and H. J. Vogel, *Biochem J*, 2007, 405, 199-221.
63. J. J. Yang, A. Gawthrop and Y. Ye, *Protein Pept Lett*, 2003, 10, 331-345.
64. T. Dudev and C. Lim, *Chem Rev*, 2003, 103, 773-788.
65. T. Dudev and C. Lim, *J Am Chem Soc*, 2006, 128, 1553-1561.
66. T. Dudev and C. Lim, *Acc Chem Res*, 2007, 40, 85-93.
67. S. Sikdar, M. Ghosh, M. De Raychaudhury and J. Chakrabarti, *Chem Phys Lett*, 2014, 605, 103-107.
68. S. Sikdar, M. Ghosh, M. De Raychaudhury and J. Chakrabarti, *J Phys Chem B*, 2015, 119, 14652-14659.
69. S. Sikdar, M. Ghosh, M. De Raychaudhury and J. Chakrabarti, *Rsc Adv*, 2016, 6, 54608-54614.
70. S Sikdar, M Ghosh, M De Raychaudhury, J Chakrabarti , *Submitted* 2016.
71. P. J. Anderson, C. L. Brooks and L. J. Berliner, *Biochemistry*, 1997, 36, 11648-11654.
72. B. Ramakrishnan, P. S. Shah and P. K. Qasba, *J Biol Chem*, 2001, 276, 37665-37671.
73. K. M. Cawthern, M. Narayan, D. Chaudhuri, E. A. Permyakov and L. J. Berliner, *J Biol Chem*, 1997, 272, 30812-30816.
74. Y. V. Griko and D. P. Remeta, *Protein Sci*, 1999, 8, 554-561.
75. S. E. Permyakov, T. I. Khokhlova, V. N. Uversky and E. A. Permyakov, *Proteins*, 2010, 78, 2609-2624.
76. A. V. Ostrovsky, L. P. Kalinichenko, V. I. Emelyanenko, A. V. Klimanov and E. A. Permyakov, *Biophys Chem*, 1988, 30, 105-112.
77. E. A. Permyakov, L. P. Kalinichenko, L. A. Morozova, V. V. Yarmolenko and E. A. Burstein, *Biochem Biophys Res Commun*, 1981, 102, 1-7.
78. Y. Hiraoka and S. Sugai, *Int J Pept Protein Res*, 1984, 23, 535-542.
79. Y. Hiraoka and S. Sugai, *Int J Pept Protein Res*, 1985, 26, 252-261.
80. B. Ramakrishnan and P. K. Qasba, *J Mol Biol*, 2001, 310, 205-218.
81. V. A. Malinovskii, J. Tian, J. A. Grobler and K. Brew, *Biochemistry*, 1996, 35, 9710-9715.
82. Y. B. Zhang, W. Wu and W. Ding, *Afr J Biotechnol*, 2010, 9, 9270-9276.
83. B. Spolaore, O. Pinato, M. Canton, M. Zambonin, P. Polverino de Laureto and A. Fontana, *Biochemistry*, 2010, 49, 8658-8667.
84. M. Svensson, J. Fast, A. K. Mossberg, C. Durringer, L. Gustafsson, O. Hallgren, C. L. Brooks, L. Berliner, S. Linse and C. Svanborg, *Protein Sci*, 2003, 12, 2794-2804.
85. N. Stanciuc, I. Aprodu, G. Rapeanu and G. Bahrim, *Eur Food Res Technol*, 2013, 236, 257-266.
86. C. Barbana, M. D. Perez, L. Sanchez, M. Dalgalarroondo, J. M. Chobert and T. Haertle, *Int Dairy J*, 2006, 16, 18-25.
87. E. L. Knyazeva, V. M. Grishchenko, R. S. Fadeev, V. S. Akatov, S. E. Permyakov and E. A. Permyakov, *Biochemistry*, 2008, 47, 13127-13137.
88. C. Svanborg, H. Agerstam, A. Aronson, R. Bjerkvig, C. Durringer, W. Fischer, L. Gustafsson, O. Hallgren, I. Leijonhuvud, S. Linse, A. K. Mossberg, H. Nilsson, J. Pettersson and M. Svensson, *Adv Cancer Res*, 2003, 88, 1-29.

89. D. W. Li, S. A. Showalter and R. Bruschweiler, *J Phys Chem B*, 2010, 114, 16036-16044.
90. S. J. De Vries, M. van Dijk and A. M. J. J. Bonvin, *Nat Protoc*, 2010, 5, 883-897.
91. T. A. Wassenaar, M. van Dijk, N. Loureiro-Ferreira, G. van der Schot, S. J. de Vries, C. Schmitz, J. van der Zwan, R. Boelens, A. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H. R. A. Jonker, A. Bagaria, V. Jaravine, P. Guntert, H. Schwalbe, W. F. Vranken, J. F. Doreleijers, G. Vriend, G. W. Vuister, D. Franke, A. Kikhney, D. I. Svergun, R. H. Fogh, J. Ionides, E. D. Laue, C. Spronk, S. Jurksa, M. Verlato, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero and A. M. J. J. Bonvin, *J Grid Comput*, 2012, 10, 743-767.
92. D. Duhovny, R. Nussinov and H. J. Wolfson, *Lect Notes Comput Sc*, 2002, 2452, 185-200.
93. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov and H. J. Wolfson, *Nucleic Acids Research*, 2005, 33, W363-W367.
94. A. Das and J. Chakrabarti, *Biophys J*, 2013, 104, 1274-1284.
95. R. Mendez and U. Bastolla, *Phys Rev Lett*, 2010, 104, 228103.
96. D. W. Li, D. Meng and R. Bruschweiler, *J Am Chem Soc*, 2009, 131, 14610-14611.
97. C. Dalgicdir, O. Sensoy, C. Peter and M. Sayar, *J Chem Phys*, 2013, 139, 234115.
98. B. J. Killian, J. Yundenfreund Kravitz and M. K. Gilson, *J Chem Phys*, 2007, 127, 024107.
99. Y. J. Xie, S. Min, N. P. Harte, H. Kirk, J. E. O'Brien, H. P. Voorheis, C. Svanborg and K. H. Mok, *Proteins-Structure Function and Bioinformatics*, 2013, 81, 1-17.
100. A. N. Naganathan, *J Chem Theory Comput*, 2012, 8, 4646-4656.
101. O. Halskau, Jr., R. Perez-Jimenez, B. Ibarra-Molero, J. Underhaug, V. Munoz, A. Martinez and J. M. Sanchez-Ruiz, *Proc Natl Acad Sci U S A*, 2008, 105, 8625-8630.
102. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden, *Bmc Bioinformatics*, 2009, 10, 421.
103. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins, *Mol Syst Biol*, 2011, 7, 539.
104. F. Xin and P. Radivojac, *Curr Protein Pept Sci*, 2011, 12, 456-469.
105. G. Amitai, A. Shemesh, E. Sitbon, M. Shklar, D. Netanel, I. Venger and S. Pietrokovski, *J Mol Biol*, 2004, 344, 1135-1146.
106. M. Jamroz, A. Kolinski and S. Kmiecik, *Bioinformatics*, 2014, 30, 2150-2154.
107. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids (Oxford Science Publications)*, Clarendon Press, 1989.
108. D. Levesque and L. Verlet, *J Stat Phys*, 1993, 72, 519-537.
109. 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.
110. T. A. Darden, A. Toukmaji and L. G. Pedersen, *J Chim Phys Pcb*, 1997, 94, 1346-1364.

111. 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.
112. D. Petrey and B. Honig, *Mol Cell*, 2005, 20, 811-819.
113. K. A. Rossi, C. A. Weigelt, A. Nayeem and S. R. Krystek, Jr., *Protein Sci*, 2007, 16, 1999-2012.
114. M. Jamroz and A. Kolinski, *Bmc Struct Biol*, 2010, 10, 5.
115. Z. Grabarek, *Biochim Biophys Acta*, 2011, 1813, 913-921.
116. G. A. de Oliveira, C. B. Rocha, A. Marques Mde, Y. Cordeiro, M. M. Sorenson, D. Foguel, J. L. Silva and M. C. Suarez, *Biochemistry*, 2013, 52, 28-40.
117. H. Kuboniwa, N. Tjandra, S. Grzesiek, H. Ren, C. B. Klee and A. Bax, *Nat Struct Biol*, 1995, 2, 768-776.
118. L. L. Chavez, J. N. Onuchic and C. Clementi, *J Am Chem Soc*, 2004, 126, 8426-8432.
119. M. S. Lee and M. A. Olson, *J Chem Theory Comput*, 2007, 3, 312-324.
120. C. Zhang, S. Liu and Y. Q. Zhou, *Protein Science*, 2004, 13, 391-399.
121. N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, 18, 2714-2723.
122. S. T. Rao, K. A. Satyshur, M. L. Greaser and M. Sundaralingam, *Acta Crystallogr D Biol Crystallogr*, 1996, 52, 916-922.
123. K. A. Satyshur, D. Pyzalska, M. Greaser, S. T. Rao and M. Sundaralingam, *Acta Crystallogr D Biol Crystallogr*, 1994, 50, 40-49.
124. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J Comput Chem*, 2009, 30, 2785-2791.
125. J. R. Pearlstone and L. B. Smillie, *Biochemistry*, 1995, 34, 6932-6940.
126. S. M. Gagne, S. Tsuda, M. X. Li, L. B. Smillie and B. D. Sykes, *Nat Struct Biol*, 1995, 2, 784-789.
127. M. X. Li, S. M. Gagne, S. Tsuda, C. M. Kay, L. B. Smillie and B. D. Sykes, *Biochemistry*, 1995, 34, 8330-8340.
128. C. M. Slupsky, F. C. Reinach, L. B. Smillie and B. D. Sykes, *Protein Sci*, 1995, 4, 1279-1290.
129. C. M. Slupsky and B. D. Sykes, *Biochemistry*, 1995, 34, 15953-15964.
130. D. E. Kim, D. Chivian and D. Baker, *Nucleic Acids Res*, 2004, 32, W526-531.
131. M. Kallberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu and J. Xu, *Nat Protoc*, 2012, 7, 1511-1522.
132. A. Fiser and A. Sali, *Bioinformatics*, 2003, 19, 2500-2501.
133. B. Han, Y. Liu, S. W. Ginzinger and D. S. Wishart, *J Biomol NMR*, 2011, 50, 43-57.
134. F. Avbelj, D. Kocjan and R. L. Baldwin, *Proc Natl Acad Sci U S A*, 2004, 101, 17394-17397.
135. W. F. Vranken and W. Rieping, *Bmc Struct Biol*, 2009, 9, 20.
136. D. S. Wishart, B. D. Sykes and F. M. Richards, *J Mol Biol*, 1991, 222, 311-333.
137. A. M. Morris, M. A. Watzky and R. G. Finke, *Bba-Proteins Proteom*, 2009, 1794, 375-397.
138. M. B. Szein, *Clin. Infect. Dis.*, 2007, 45, S15-S19.
139. S. Ghosh, K. Chakraborty, T. Nagaraja, S. Basak, H. Koley, S. Dutta, U. Mitra and S. Das, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, 108, 3348-3353.
140. P. Everest, J. Wain, M. Roberts, G. Rook and G. Dougan, *Trends Microbiol*, 2001, 9, 316-320.

141. J. E. Galan, *Curr Top Microbiol Immunol*, 1996, 209, 43-60.
142. J. Parkhill, G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. G. Holden, M. Sebahia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead and B. G. Barrell, *Nature (London, U. K.)*, 2001, 413, 848-852.
143. J. Parkhill, G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebahia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead and B. G. Barrell, *Nature*, 2001, 413, 848-852.
144. W. Deng, S. R. Liou, G. Plunkett, 3rd, G. F. Mayhew, D. J. Rose, V. Burland, V. Kodoyianni, D. C. Schwartz and F. R. Blattner, *J Bacteriol*, 2003, 185, 2330-2337.
145. D. Beece, L. Eisenstein, H. Frauenfelder, D. Good, M. C. Marden, L. Reinisch, A. H. Reynolds, L. B. Sorensen and K. T. Yue, *Biochemistry*, 1980, 19, 5147-5157.
146. A. P. Demchenko, O. I. Ruskyn and E. A. Saburova, *Biochim Biophys Acta*, 1989, 998, 196-203.
147. C. J. Feng, R. V. Kedia, J. T. Hazzard, J. K. Hurley, G. Tollin and J. H. Enemark, *Biochemistry*, 2002, 41, 5816-5821.
148. B. Gavish and M. M. Werber, *Biochemistry*, 1979, 18, 1269-1275.
149. R. B. Gregory, A. Rosenberg, D. Knox and A. J. Percy, *Biopolymers*, 1990, 29, 1175-1183.
150. G. S. Jas, W. A. Eaton and J. Hofrichter, *Journal of Physical Chemistry B*, 2001, 105, 261-272.
151. T. Kleinert, W. Doster, H. Leyser, W. Petry, V. Schwarz and M. Settles, *Biochemistry*, 1998, 37, 717-733.
152. D. Lavalette and C. Tetreau, *Eur J Biochem*, 1988, 177, 97-108.
153. E. J. Loveridge, L. H. Tey and R. K. Allemann, *J Am Chem Soc*, 2010, 132, 1137-1143.
154. K. Ng and A. Rosenberg, *Biophys Chem*, 1991, 39, 57-68.
155. H. Oh-oka, M. Iwaki and S. Itoh, *Biochemistry*, 1997, 36, 9267-9272.
156. S. A. Pabit, H. Roder and S. J. Hagen, *Biochemistry*, 2004, 43, 12532-12538.
157. K. W. Plaxco and D. Baker, *Proc Natl Acad Sci U S A*, 1998, 95, 13591-13596.
158. L. Qiu and S. J. Hagen, *J Am Chem Soc*, 2004, 126, 3398-3399.
159. G. Ricci, A. M. Caccuri, M. Lo Bello, N. Rosato, G. Mei, M. Nicotra, E. Chiessi, A. P. Mazzetti and G. Federici, *J Biol Chem*, 1996, 271, 16187-16192.
160. A. Rosenberg, K. Ng and M. Punyiczki, *J Mol Liq*, 1989, 42, 31-43.
161. P. Sashi and A. K. Bhuyan, *Biochemistry*, 2015, 54, 4453-4461.
162. E. Papaleo, G. Saladino, M. Lambrughli, K. Lindorff-Larsen, F. L. Gervasio and R. Nussinov, *Chem Rev*, 2016.

163. J. A. McCammon and M. Karplus, *Proc Natl Acad Sci U S A*, 1979, 76, 3585-3589.
164. D. D. Boehr, R. Nussinov and P. E. Wright, *Nat Chem Biol*, 2009, 5, 789-796.
165. G. M. Clore, *Biophys Chem*, 2014, 186, 3-12.
166. N. S. Hatzakis, *Biophys Chem*, 2014, 186, 46-54.
167. R. Nussinov, B. Ma and C. J. Tsai, *Biophys Chem*, 2014, 186, 22-30.
168. E. Fischer, *Berichte der deutschen chemischen Gesellschaft*, 1894, 27, 2985-2993.
169. D. E. Koshland, *Proc Natl Acad Sci U S A*, 1958, 44, 98-104.
170. B. Ma, S. Kumar, C. J. Tsai and R. Nussinov, *Protein Eng*, 1999, 12, 713-720.
171. C. J. Tsai, S. Kumar, B. Ma and R. Nussinov, *Protein Sci*, 1999, 8, 1181-1190.
172. C. J. Tsai, B. Ma and R. Nussinov, *Proc Natl Acad Sci U S A*, 1999, 96, 9970-9972.
173. T. I. Igumenova, K. K. Frederick and A. J. Wand, *Chem Rev*, 2006, 106, 1672-1699.
174. T. Wlodarski and B. Zagrovic, *Proc Natl Acad Sci U S A*, 2009, 106, 19346-19351.
175. L. A. Svensson, E. Thulin and S. Forsen, *J Mol Biol*, 1992, 223, 601-606.
176. N. J. Skelton, J. Kordel and W. J. Chazin, *J Mol Biol*, 1995, 249, 441-462.
177. I. Echeverria, D. E. Makarov and G. A. Papoian, *J Am Chem Soc*, 2014, 136, 8708-8713.
178. J. C. Schulz, L. Schmidt, R. B. Best, J. Dzubiella and R. R. Netz, *J Am Chem Soc*, 2012, 134, 6273-6279.
179. M. A. Shea, B. R. Sorensen, S. Pedigo and A. S. Verhoeven, *Methods Enzymol*, 2000, 323, 254-301.
180. B. R. Sorensen and M. A. Shea, *Biochemistry*, 1998, 37, 4244-4253.
181. Y. S. Babu, C. E. Bugg and W. J. Cook, *J Mol Biol*, 1988, 204, 191-204.
182. J. M. Baldwin, *J Mol Biol*, 1980, 136, 103-128.
183. B. Sjoblom, M. Polentarutti and K. Djinovic-Carugo, *P Natl Acad Sci USA*, 2009, 106, 10609-10613.
184. M. Ghosh, G. Meiss, A. M. Pingoud, R. E. London and L. C. Pedersen, *J Biol Chem*, 2007, 282, 5682-5690.
185. T. Dudev, Y. L. Lin, M. Dudev and C. Lim, *J Am Chem Soc*, 2003, 125, 3168-3180.
186. F. Himo and P. E. Siegbahn, *J Am Chem Soc*, 2001, 123, 10280-10289.
187. P. E. Siegbahn, *Curr Opin Chem Biol*, 2002, 6, 227-235.
188. P. E. Siegbahn and M. R. Blomberg, *Annu Rev Phys Chem*, 1999, 50, 221-249.
189. A. D. Becke, *Journal of Chemical Physics*, 1993, 98, 5648-5652.
190. V. A. Rassolov, J. A. Pople, M. A. Ratner and T. L. Windus, *Journal of Chemical Physics*, 1998, 109, 1223-1229.
191. A. J. Cohen, P. Mori-Sanchez and W. Yang, *Chem Rev*, 2012, 112, 289-320.
192. A. Lewit-Bentley and S. Rety, *Curr Opin Struct Biol*, 2000, 10, 637-643.
193. J. K. Lassila, J. G. Zalatan and D. Herschlag, *Annu Rev Biochem*, 2011, 80, 669-702.
194. B. Zhou and C. F. Wong, *J Phys Chem A*, 2009, 113, 5144-5150.
195. J. Tomasi, B. Mennucci and R. Cammi, *Chem Rev*, 2005, 105, 2999-3093.
196. R. Improta, L. Vitagliano and L. Esposito, *PLoS One*, 2011, 6, e24533.

197. C. N. Pace, S. Trevino, E. Prabhakaran and J. M. Scholtz, *Philos Trans R Soc Lond B Biol Sci*, 2004, 359, 1225-1234; discussion 1234-1225.
198. J. C. Phillips and R. Braun, *J Comput Chem*, 2005, 26, 1781-1802.
199. M. J. Frisch and G. W. Trucks, 2003.
200. P. K. Chattaraj, U. Sarkar and D. R. Roy, *Chem Rev*, 2006, 106, 2065-2091.
201. C. Kobayashi and S. Takada, *Biophys J*, 2006, 90, 3043-3051.
202. A. E. Reed and L. A. Curtiss, *Chemical Reviews*, 1988, 88, 899-926.
203. F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, 2006.
204. P. Hohenberg and W. Kohn, *Physical Review*, 1964, 136, B864-B871.
205. N. Nandi and B. Bagchi, *Journal of Physical Chemistry B*, 1997, 101, 10954-10961.
206. B. Bagchi, *Chem Phys Lett*, 2012, 529, 1-9.
207. G. Henkelman, M. X. LaBute, C. S. Tung, P. W. Fenimore and B. H. McMahon, *Proc Natl Acad Sci U S A*, 2005, 102, 15347-15351.
208. R. Ghosh, S. Banerjee, M. Hazra, S. Roy and B. Bagchi, *J Chem Phys*, 2014, 141, 22D531.
209. P. Setny, R. Baron, P. Michael Kekenes-Huskey, J. A. McCammon and J. Dzubiella, *Proc Natl Acad Sci U S A*, 2013, 110, 1197-1202.
210. K. Chakraborty and S. Bandyopadhyay, *J Phys Chem B*, 2014, 118, 413-422.
211. P. C. do Couto, S. G. Estacio and B. J. C. Cabral, *Journal of Chemical Physics*, 2005, 123.
212. F. Momany and U. Schnupf, *Comput Theor Chem*, 2014, 1029, 57-67.
213. H. M. Senn and W. Thiel, *Angew Chem Int Ed Engl*, 2009, 48, 1198-1229.
214. J. Thogersen, J. Rehault, M. Odellius, T. Ogden, N. K. Jena, S. J. K. Jensen, S. R. Keiding and J. Helbing, *Journal of Physical Chemistry B*, 2013, 117, 3376-3388.
215. M. W. van der Kamp and A. J. Mulholland, *Biochemistry*, 2013, 52, 2708-2728.
216. M. Boggio-Pasqua, M. A. Robb and G. Groenhof, *J Am Chem Soc*, 2009, 131, 13580-13581.
217. L. N. Johnson, *Biochem Soc Trans*, 2009, 37, 627-641.
218. P. E. Blochl, *Phys Rev B Condens Matter*, 1994, 50, 17953-17979.
219. G. Kresse and J. Hafner, *Phys Rev B Condens Matter*, 1993, 48, 13115-13118.
220. G. Kresse and D. Joubert, *Phys Rev B*, 1999, 59, 1758-1775.
221. R. Car and M. Parrinello, *Phys Rev Lett*, 1985, 55, 2471-2474.
222. P. Adhikari, A. M. Wen, R. H. French, V. A. Parsegian, N. F. Steinmetz, R. Podgornik and W. Y. Ching, *Sci Rep*, 2014, 4, 5605.
223. P. L. Silvestrelli and M. Parrinello, *Journal of Chemical Physics*, 1999, 111, 3572-3580.
224. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys Rev Lett*, 1996, 77, 3865-3868.
225. W. Tang, E. Sanville and G. Henkelman, *J Phys-Condens Mat*, 2009, 21.
226. J. Melin, P. W. Ayers and J. V. Ortiz, *J Phys Chem A*, 2007, 111, 10017-10019.
227. X. B. Xu, G. R. Shan, Y. Shang and P. J. Pan, *J Appl Polym Sci*, 2014, 131.
228. R. Rivelino, A. M. Maniero, F. V. Prudente and L. S. Costa, *Carbon*, 2006, 44, 2925-2930.

229. Y. Ye, H. W. Lee, W. Yang, S. Shealy and J. J. Yang, *J Am Chem Soc*, 2005, 127, 3743-3750.
230. D. Vanderbilt, *Phys Rev B*, 1990, 41, 7892-7895.

## ***LIST of PUBLICATIONS***

- [1] **S Sikdar**, M Ghosh, M De Raychaudhury, J Chakrabarti. Quantum chemical studies on the role of residues in calcium ion binding to Calmodulin. *Chem Phys Lett.* 2014; **605**:103-7.
- [2] **S Sikdar**, J Chakrabarti, M Ghosh. A microscopic insight from conformational thermodynamics to functional ligand binding in proteins. *Mol Biosyst.* 2014; **10**:3280-9.
- [3] A Das, **S Sikdar**, M Ghosh, J Chakrabarti. Conformational thermodynamics of biomolecular complexes: The histogram-based method. *Journal of Physics: Conference Series.* 2015; **638**:012013.
- [4] **S Sikdar**, M Ghosh, M De Raychaudhury, J Chakrabarti. Quantum chemical studies on stability and chemical activities in calcium ion bound calmodulin loops. *J Phys Chem B.* 2015; **119**:14652-9.
- [5] **S Sikdar**, J Chakrabarti, M Ghosh. Conformational thermodynamics guided structural reconstruction of bio-molecular fragments. *Mol Biosyst.* 2016; **12**:444-53.
- [6] **S Sikdar**, M Ghosh, M De Raychaudhury, J Chakrabarti. Quantum chemical studies on nucleophilic sites in calcium ion bound Zwitterionic calmodulin loops. *RSC. Adv.* 2016; **6**: 54608-14.
- [7] **S Sikdar**, M Ghosh, M De Raychaudhury, J Chakrabarti. Ab-initio quantum chemical studies on role of hydration water on protein functionality. *Submitted*, 2016.
- [8] **S Sikdar**, M Ghosh, J Chakrabarti. Static and dynamic response of microscopic conformational variables of binding residues in a protein. *Pre-print*, 2016.
- [9] P Saha, **S Sikdar**, J Chakrabarti, M Ghosh. Response to chemical induced changes and their implications in yfdX protein STY3178. *Submitted*, 2016.